U.S. DEPARTMENT OF ENERGY

Office of Science

Systematic Assessment of Terrestrial and Marine Biogeochemistry in Earth System Models

Forrest M. Hoffman[1,2], Nathan Collier[1], Mingquan Mu[3], Min Xu[1], Weiwei Fu[3], Cheng-En Yang[2,1], Gretchen Keppel-Aleks[4], David M. Lawrence[5], Charles D. Koven[6], William J. Riley[6], and James T. Randerson[3]

[1]Oak Ridge National Laboratory, Oak Ridge, TN, USA
[2]University of Tennessee, Knoxville, TN, USA
[3]University of California, Irvine, CA, USA
[4]University of Michigan, Ann Arbor, MI, USA
[5]National Center for Atmospheric Research, Boulder, CO, USA
[6]Lawrence Berkeley National Laboratory, Berkeley, CA, USA

June 24, 2024

Forrest M. Hoffman, Computational Earth System Scientist

- Group Leader for the ORNL Integrated Computational Earth Sciences Group
- 35 years at ORNL, first in Environmental Sciences Division, then Computer Science and Mathematics Division, and now Computational Sciences and Engineering Division
- Develop and apply Earth system models to study global biogeochemical cycles, including terrestrial & marine carbon cycle
- Investigate methods for reconciling uncertainties in carbon–climate feedbacks through comparison with observations
- Apply artificial intelligence methods (machine learning and data mining) to environmental characterization, simulation, & analysis
- Joint Faculty, University of Tennessee, Knoxville, Department of Civil & Environmental Engineering

Oak Ridge National Laboratory

The University of Tennessee Knoxville

US Dept. of Energy's RUBISCO Science Focus Area (SFA)

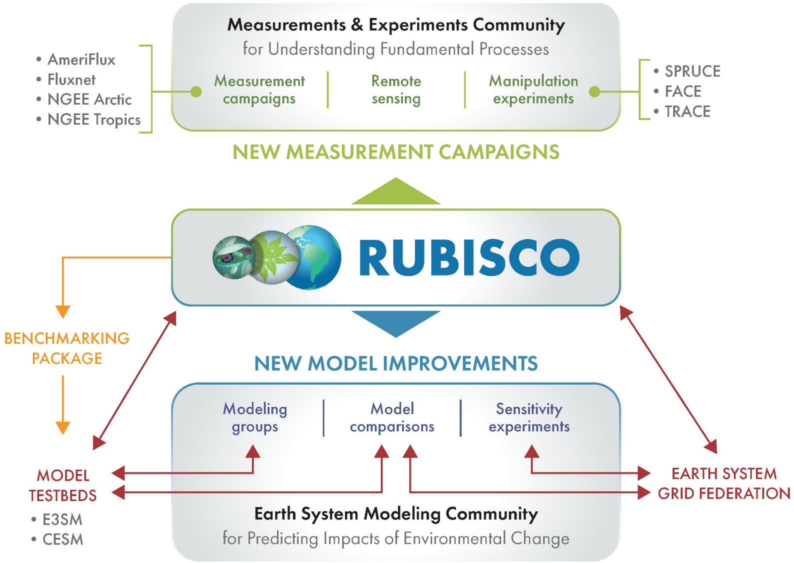
Forrest M. Hoffman (Laboratory Research Manager), William J. Riley (Senior Science Co-Lead), and James T. Randerson (Chief Scientist)

Research Goals

- Identify and quantify feedbacks between biogeochemical cycles and the Earth system
- Quantify and reconcile uncertainties in Earth system models (ESMs) associated with interactions

Research Objectives

- Perform hypothesis-driven analysis of biogeochemical & hydrological processes and feedbacks in ESMs
- Synthesize in situ and remote sensing data and design metrics for assessing ESM performance
- Design, develop, and release the International Land Model Benchmarking (ILAMB) and International Ocean Model Benchmarking (IOMB) tools for systematic evaluation of model fidelity
- Conduct and evaluate CMIP6 experiments with ESMs

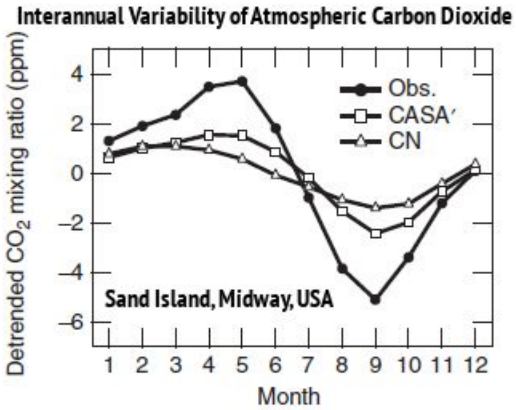


The RUBISCO SFA works with the measurements and the modeling communities to use best-available data to evaluate the fidelity of ESMs. RUBISCO identifies model gaps and weaknesses, informs new model development efforts, and suggests new measurements and field campaigns.

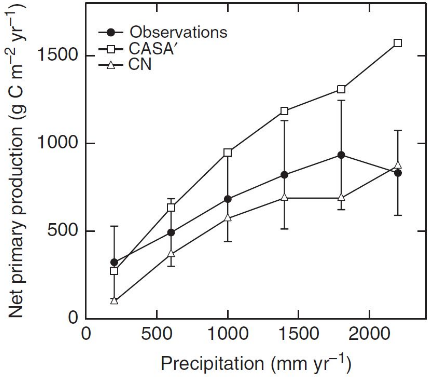
Argonne NATIONAL LABORATORY · BROOKHAVEN NATIONAL LABORATORY · BERKELEY LAB · Los Alamos NATIONAL LABORATORY · NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH · OAK RIDGE National Laboratory · UCI · MICHIGAN

What is a Benchmark?

- A **benchmark** is a quantitative test of model function achieved through comparison of model results with observational data
- Acceptable performance on a benchmark **is a necessary but not sufficient condition** for a fully functioning model
- **Functional relationship benchmarks** offer tests of model responses to forcings and yield insights into ecosystem processes
- Effective benchmarks must draw upon **a broad set of independent observations** to evaluate model performance at multiple scales



Models often fail to capture the amplitude of the seasonal cycle of atmospheric CO_2



Models may reproduce correct responses over only a limited range of forcing variables

(Randerson et al., 2009)

Argonne NATIONAL LABORATORY · BROOKHAVEN NATIONAL LABORATORY · BERKELEY LAB · Los Alamos NATIONAL LABORATORY · NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH · OAK RIDGE National Laboratory · UCI · MICHIGAN

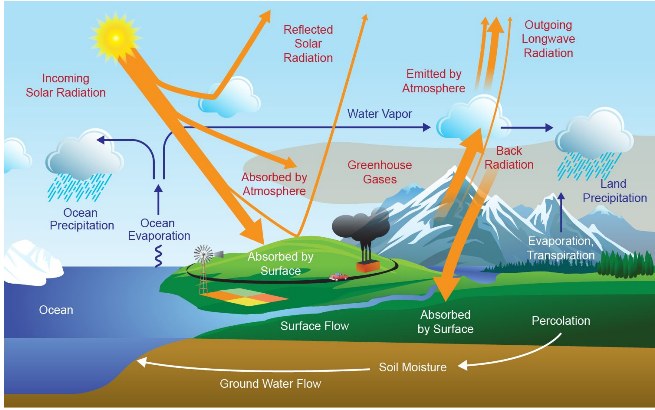
Why Benchmark Models?

- To **quantify and reduce uncertainties** in carbon cycle feedbacks to improve projections of future climate change (Eyring et al., 2019; Collier et al., 2018)
- To **diagnose impacts of *process-based* or *machine learning* model development** on process representations and their interactions
- To **guide synthesis efforts**, such as the Intergovernmental Panel on Climate Change (IPCC), by determining which models are broadly consistent with observations (Eyring et al., 2019)
- To **increase scrutiny of key datasets** used for model evaluation
- To **identify gaps in existing observations** needed to inform model development
- To **accelerate delivery of new measurement datasets** for rapid and widespread use in model assessment

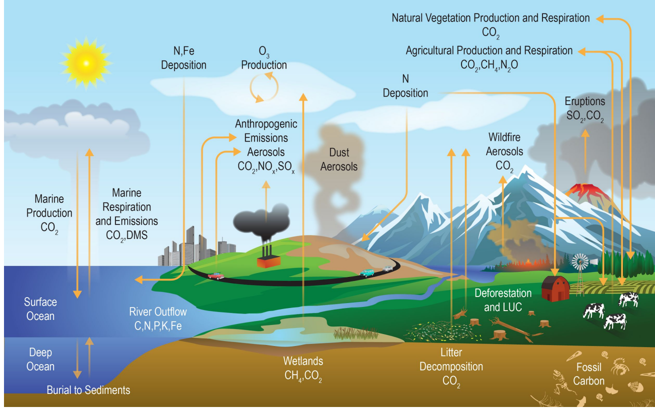
What is ILAMB?

A community coordination activity created to:

- **Develop internationally accepted benchmarks** for land model performance by drawing upon collaborative expertise
- **Promote the use of these benchmarks** for model intercomparison
- **Strengthen linkages between experimental, remote sensing, and Earth system modeling communities** in the design of new model tests and new measurement programs
- **Support the design and development of open source benchmarking tools**



Energy and Water Cycles



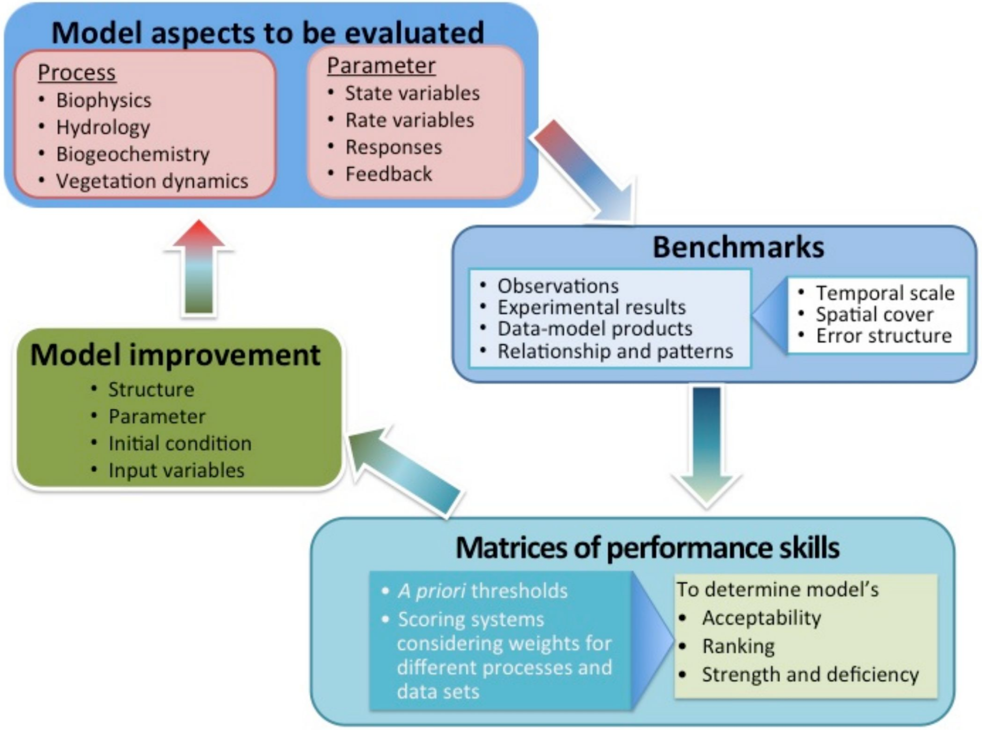
Carbon and Biogeochemical Cycles

Argonne NATIONAL LABORATORY · BROOKHAVEN NATIONAL LABORATORY · BERKELEY LAB · Los Alamos NATIONAL LABORATORY · NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH · OAK RIDGE National Laboratory · UCI · MICHIGAN

- **First ILAMB Workshop** was held in Exeter, UK, on June 22–24, 2009
- **Second ILAMB Workshop** was held in Irvine, CA, USA, on January 24–26, 2011
  - ~45 researchers participated from the US, Canada, UK, Netherlands, France, Germany, Switzerland, China, Japan, and Australia
  - Developed methodology for model-data comparison and baseline standard for performance of land model process representations (Luo et al., 2012)

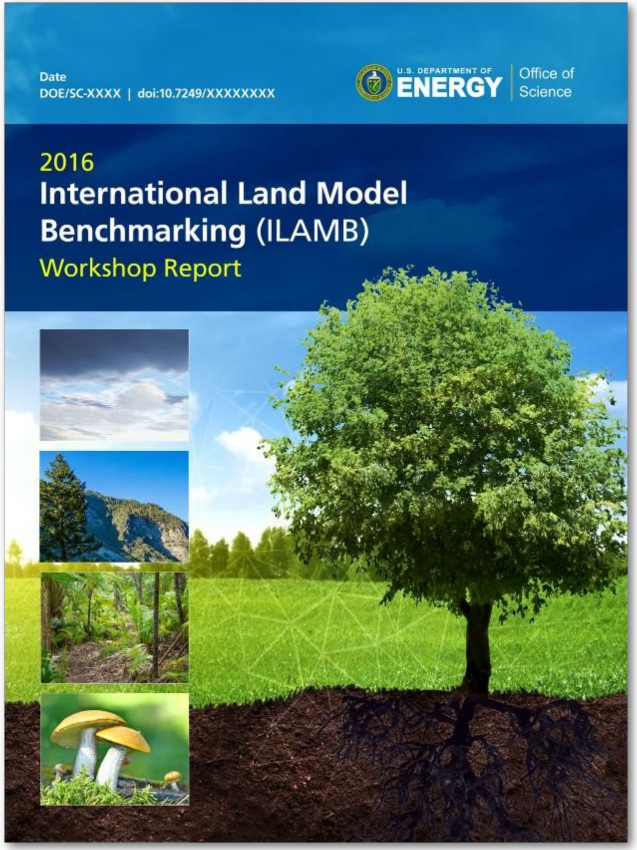
A Framework for Benchmarking Land Models

- A **benchmarking framework for evaluating land models** emerged and included (1) defining model aspects to be evaluated, (2) selecting benchmarks as standardized references, (3) developing a scoring system to measure model performance, and (4) stimulating model improvement
- Based on this methodology and prior work on the **Carbon-LAnd Model Intercomparison Project (C-LAMP)** (Randerson et al., 2009), a prototype model benchmarking package was developed for ILAMB



(Luo et al., 2012)

2016 International Land Model Benchmarking (ILAMB) Workshop
May 16–18, 2016, Washington, DC



Third ILAMB Workshop was held May 16–18, 2016

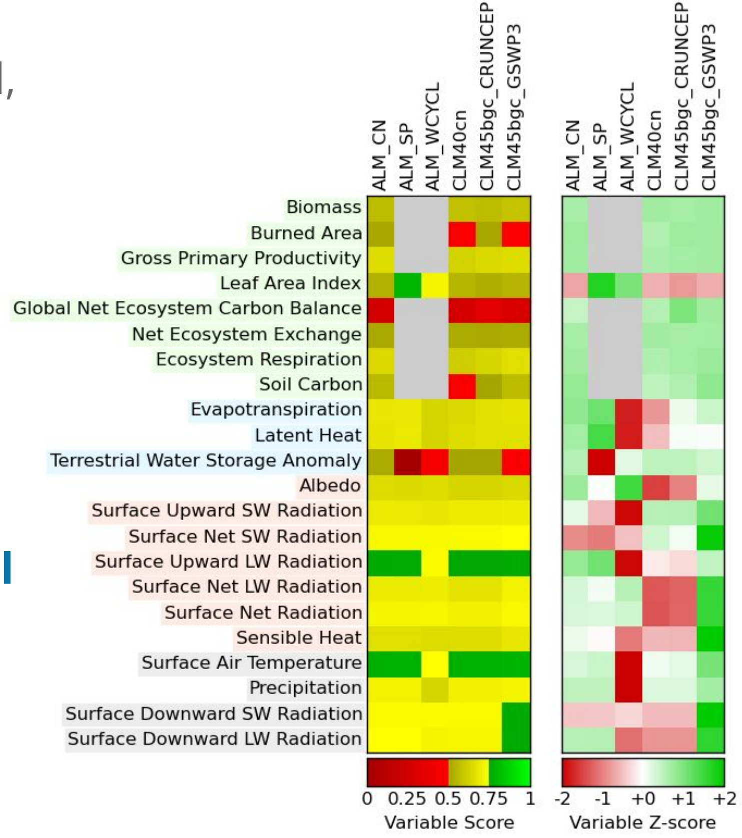
- Workshop Goals
  - Design of new metrics for model benchmarking
  - Model Intercomparison Project (MIP) evaluation needs
  - Model development, testbeds, and workflow processes
  - Observational datasets and needed measurements
- Workshop Attendance
  - 60+ participants from Australia, Japan, China, Germany, Sweden, Netherlands, UK, and US (10 modeling centers)
  - ~25 remote attendees at any time

(Hoffman et al., 2017)

Argonne NATIONAL LABORATORY | BROOKHAVEN NATIONAL LABORATORY | BERKELEY LAB | Los Alamos NATIONAL LABORATORY | NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH | OAK RIDGE National Laboratory | UCI | MICHIGAN

Development of ILAMB Packages

- **ILAMBv1** released at 2015 AGU Fall Meeting Town Hall, doi:[10.18139/ILAMB.v001.00/1251597](https://doi.org/10.18139/ILAMB.v001.00/1251597)
- **ILAMBv2** released at 2016 ILAMB Workshop, doi:[10.18139/ILAMB.v002.00/1251621](https://doi.org/10.18139/ILAMB.v002.00/1251621)
- **Open Source software** written in Python; **runs in parallel** on laptops, clusters, and supercomputers
- Routinely used for land model evaluation during development of ESMs, including the **E3SM Land Model** (Zhu et al., 2019) and the **CESM Community Land Model** (Lawrence et al., 2019)
- **Models are scored** based on statistical comparisons and functional response metrics



ILAMB Produces Diagnostics and Scores Models

- ILAMB generates a top-level **portrait plot** of models scores
- For every variable and dataset, ILAMB can automatically produce
  - **Tables** containing individual metrics and metric scores (when relevant to the data), including
    - Benchmark and model **period mean**
    - **Bias** and **bias score** (S_{bias})
    - **Root-mean-square error (RMSE)** and **RMSE score** (S_{rmse})
    - **Phase shift** and **seasonal cycle score** (S_{phase})
    - **Interannual coefficient of variation** and **IAV score** (S_{iav})
    - **Spatial distribution score** (S_{dist})
    - **Overall score** (S_{overall})

$$S_{\text{overall}} = \frac{S_{\text{bias}} + 2S_{\text{rmse}} + S_{\text{phase}} + S_{\text{iav}} + S_{\text{dist}}}{1+2+1+1+1}$$

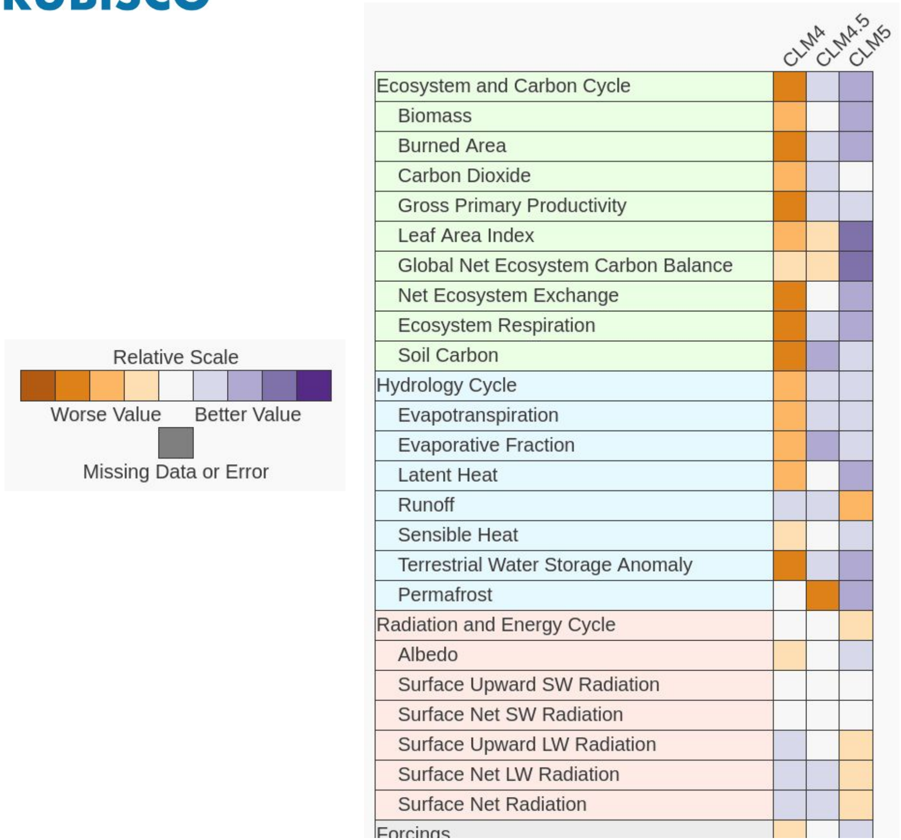
  - **Graphical diagnostics**
    - Spatial contour maps
    - Time series line plots
    - Spatial Taylor diagrams (Taylor, 2001)
- Similar **tables** and **graphical diagnostics** for functional relationships

RUBISCO

ILAMBv2.6 Package Current Variables

- **Biogeochemistry:** Biomass (Contiguous US, Pan Tropical Forest), Burned area (GFED3), CO_2 (NOAA GMD, Mauna Loa), Gross primary production (Fluxnet, GBAF), Leaf area index (AVHRR, MODIS), Global net ecosystem carbon balance (GCP, Khatiwala/Hoffman), Net ecosystem exchange (Fluxnet, GBAF), Ecosystem Respiration (Fluxnet, GBAF), Soil C (HWSD, NCSCDv22, Koven)
- **Hydrology:** Evapotranspiration (GLEAM, MODIS), Evaporative fraction (GBAF), Latent heat (Fluxnet, GBAF, DOLCE), Runoff (Dai, LORA), Sensible heat (Fluxnet, GBAF), Terrestrial water storage anomaly (GRACE), Permafrost (NSIDC)
- **Energy:** Albedo (CERES, GEWEX.SRB), Surface upward and net SW/LW radiation (CERES, GEWEX.SRB, WRMC.BSRN), Surface net radiation (CERES, Fluxnet, GEWEX.SRB, WRMC.BSRN)
- **Forcing:** Surface air temperature (CRU, Fluxnet), Diurnal max/min/range temperature (CRU), Precipitation (CMAP, Fluxnet, GPCC, GPCP2), Surface relative humidity (ERA), Surface down SW/LW radiation (CERES, Fluxnet, GEWEX.SRB, WRMC.BSRN)

ILAMB Assessing Several Generations of CLM

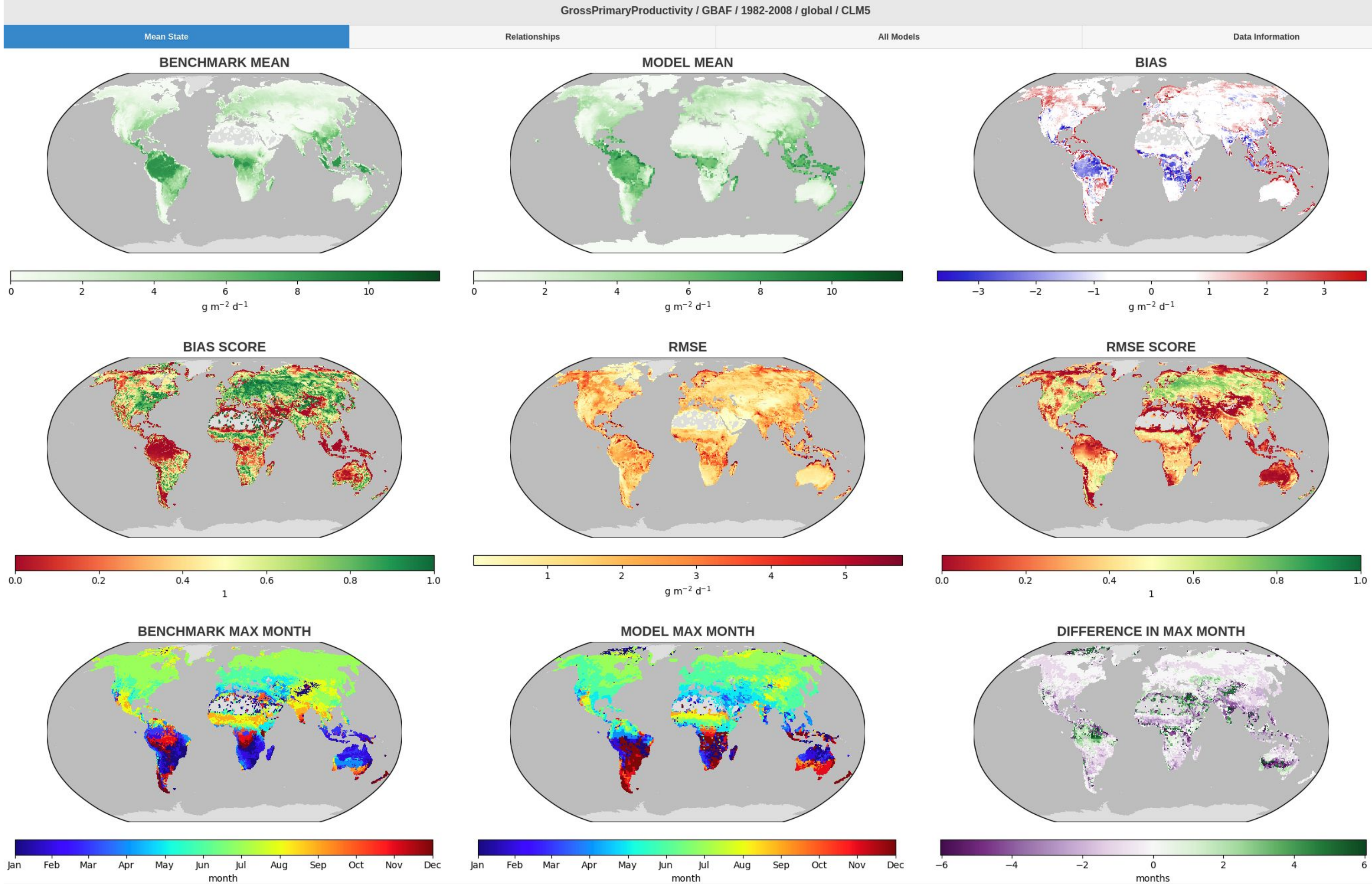


- Improvements in mechanistic treatment of hydrology, ecology, and land use with much more complexity in Community Land Model version 5 (CLM5)
- Simulations improved even with enhanced complexity
- Observational datasets not always self-consistent
- Forcing uncertainty confounds assessment of model development

http://webext.cgd.ucar.edu/I20TR/_build_set1F/

(Lawrence et al., 2019)

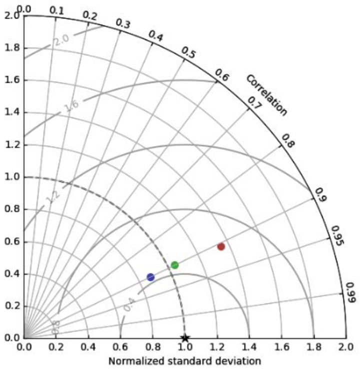
ILAMB Graphical Diagnostics



ILAMB Graphical Diagnostics

RUBISCO

SPATIAL TAYLOR DIAGRAM



MODEL COLORS

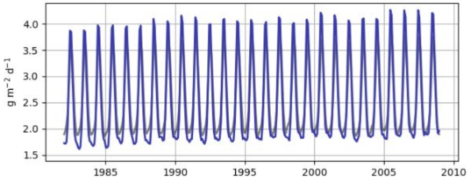
- Benchmark
- CLM4
- CLM4.5
- CLM5

Spatially integrated regional mean

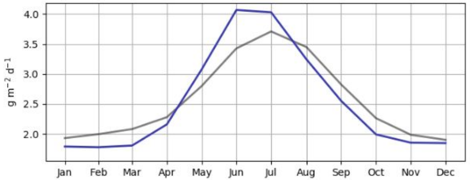
MODEL COLORS

- Benchmark
- CLM4
- CLM4.5
- CLM5

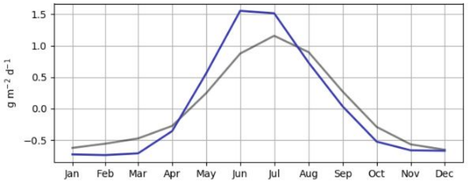
REGIONAL MEAN



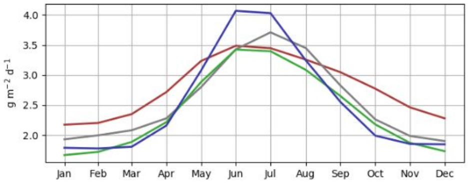
ANNUAL CYCLE



MONTHLY ANOMALY



ANNUAL CYCLE



Argonne NATIONAL LABORATORY
BROOKHAVEN NATIONAL LABORATORY
BERKELEY LAB
Los Alamos NATIONAL LABORATORY
NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH
OAK RIDGE National Laboratory
UCI
MICHIGAN

ILAMB Graphical Diagnostics

Under Construction: ILAMB site-level diagnostics

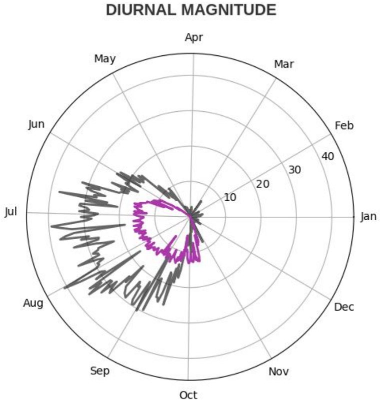
GrossPrimaryProductivity / AMF_US-WCr / global / CLM5

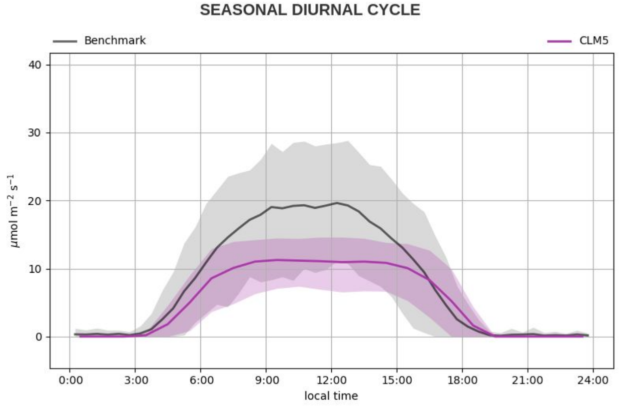
Mean State | All Models | Data Information

	Download Data	Number of Years [1]	Season Length [d]	Diurnal Peak Timing [h]	Mean Season Uptake [1e-6 mol m-2 s-1]	Season Beginning [d]	Season Ending [d]	Diurnal Peak Timing Score [1]	Diurnal Uptake Score [1]	Season Beginning Score [1]	Season Ending Score [1]	Season Strength Score [1]	Overall Score [1]
Benchmark	[-]	14.0	120.	10.6	8.07	144.	264.						
CLM5	[-]	14.0	153.	10.4	5.62	138.	291.	0.896	0.756	0.681	0.361	0.668	0.672
ED2a	[-]	2.00	140.	12.8	3.77	138.	278.	0.750	0.570	0.712	0.616	0.601	0.650
ED2b	[-]	6.00	161.	10.8	4.35	120.	281.	0.910	0.610	0.370	0.517	0.630	0.607
SIPNETa	[-]	7.00	136.	9.79	6.86	145.	281.	0.908	0.818	0.801	0.510	0.835	0.774
SIPNETb	[-]	2.00	178.	4.50	5.76	104.	282.	0.521	0.670	0.205	0.400	0.850	0.529
SIPNETc	[-]	7.00	128.	8.64	8.81	144.	273.	0.830	0.769	0.811	0.716	0.736	0.773

1999

2000



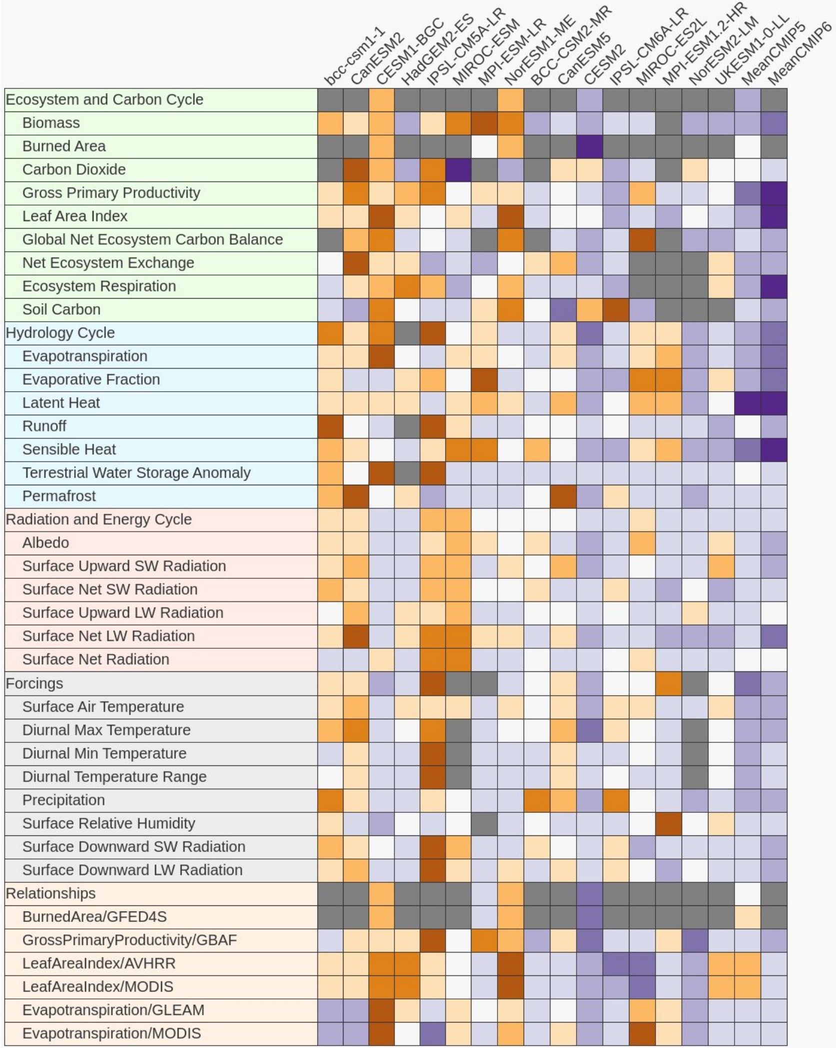


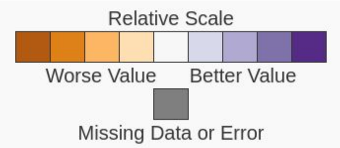
RUBISCO

Argonne National Laboratory · Brookhaven National Laboratory · Berkeley Lab · Los Alamos National Laboratory · NCAR National Center for Atmospheric Research · Oak Ridge National Laboratory · UCI · Michigan

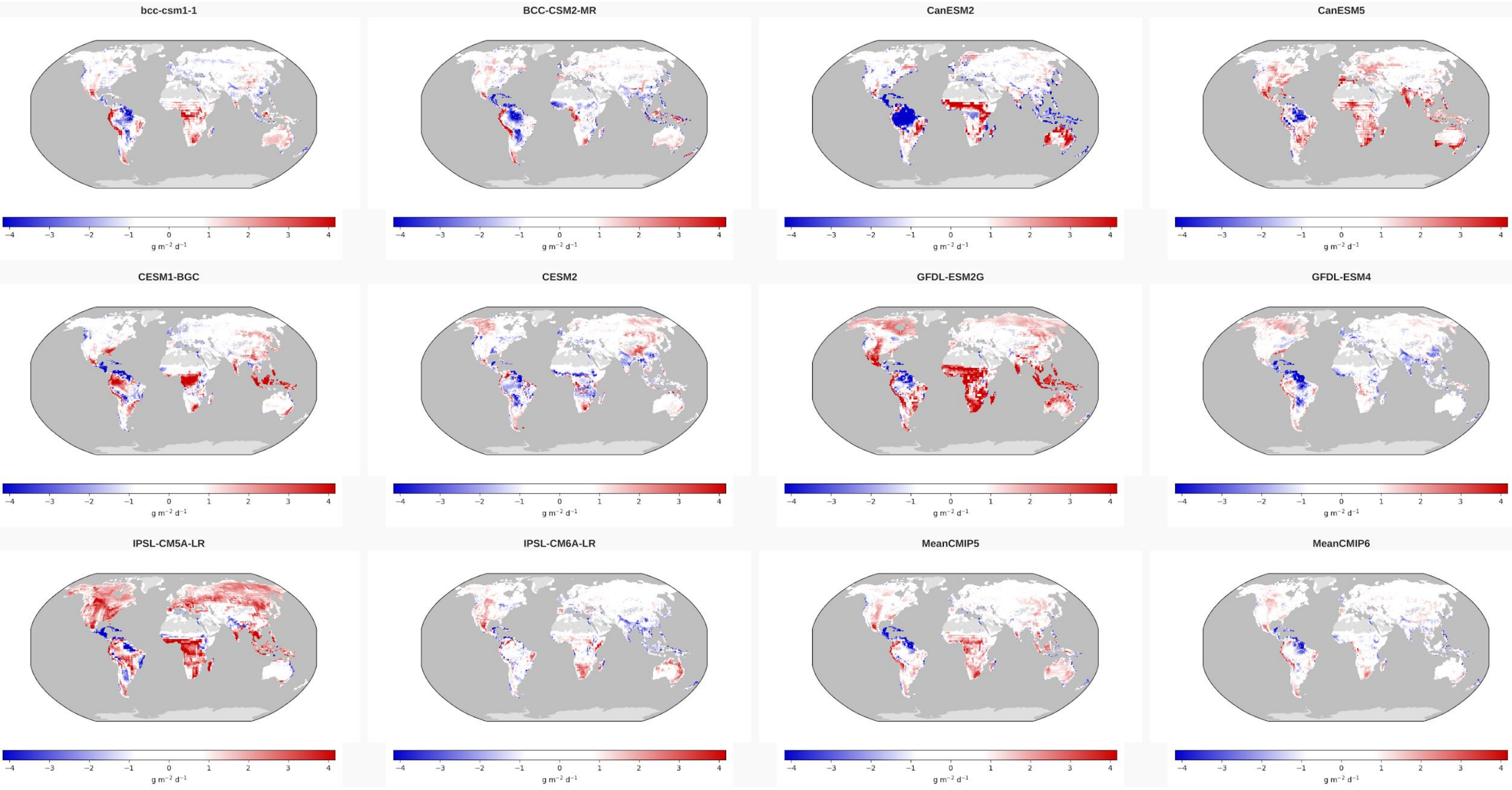
CMIP5 vs. CMIP6 Models

- The CMIP6 suite of land models (right) has improved over the CMIP5 suite of land models (left)
- The multi-model mean outperforms any single model for each suite of models
- The multi-model mean CMIP6 land model is the “best model” overall
- Why did CMIP6 land models improve?





(Hoffman et al., in prep)

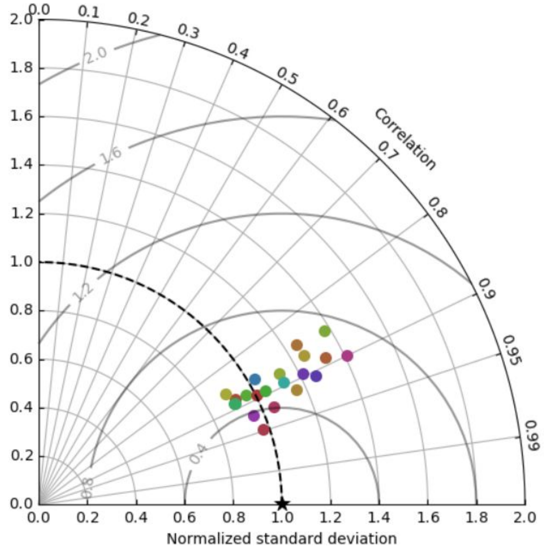

bcc-csm1-1
BCC-CSM2-MR
CanESM2
CanESM5
CESM1-BGC
CESM2
GFDL-ESM2G
GFDL-ESM4
IPSL-CM5A-LR
IPSL-CM6A-LR
MeanCMIP5
MeanCMIP6
−4 −3 −2 −1 0 1 2 3 4
g m^{-2} d^{-1}

Gross Primary Productivity

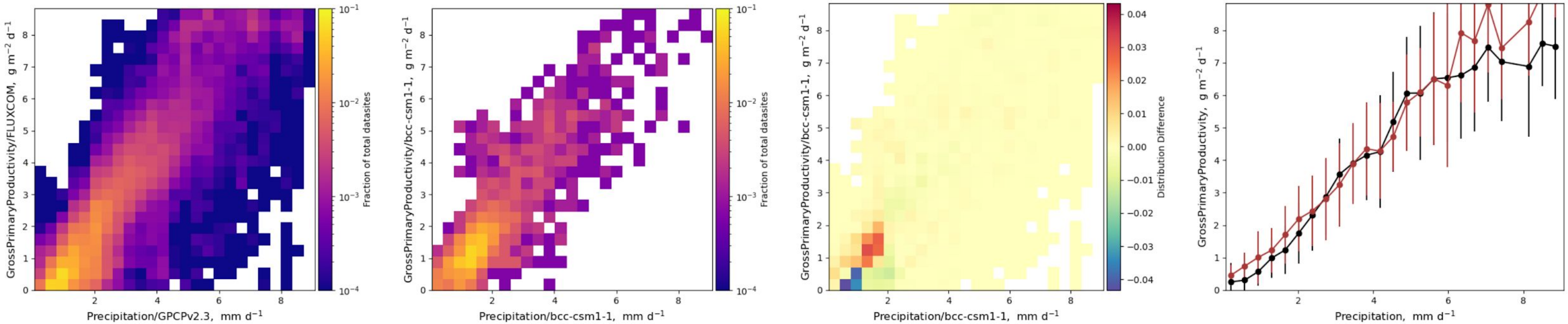
	Download Data	Period Mean (original grids) [Pg yr-1]	Model Period Mean (intersection) [Pg yr-1]	Benchmark Period Mean (intersection) [Pg yr-1]	Model Period Mean (complement) [Pg yr-1]	Benchmark Period Mean (complement) [Pg yr-1]	Bias [g m-2 d-1]	RMSE [g m-2 d-1]	Phase Shift [months]	Bias Score [1]	RMSE Score [1]	Seasonal Cycle Score [1]	Spatial Distribution Score [1]	Overall Score [1]
Benchmark	[-]	114.												
bcc-csm1-1	[-]	123.	112.	114.	8.79	0.0945	0.238	1.51	1.01	0.484	0.435	0.830	0.955	0.628
BCC-CSM2-MR	[-]	114.	107.	113.	5.88	0.671	-0.0233	1.52	1.11	0.479	0.447	0.817	0.941	0.626
CanESM2	[-]	129.	117.	114.	9.54		0.0601	2.31	2.00	0.388	0.437	0.650	0.836	0.549
CanESM5	[-]	141.	128.	114.	10.1		0.730	1.87	1.60	0.449	0.418	0.710	0.948	0.589
CESM1-BGC	[-]	129.	123.	113.	5.55	0.660	0.379	1.66	1.20	0.426	0.468	0.765	0.889	0.603
CESM2	[-]	110.	104.	113.	5.57	0.642	-0.0542	1.62	1.32	0.458	0.466	0.774	0.933	0.619
GFDL-ESM2G	[-]	167.	152.	114.	12.4		1.26	2.78	1.38	0.377	0.288	0.735	0.897	0.517
GFDL-ESM4	[-]	105.	99.0	114.	6.18		-0.177	1.59	1.49	0.495	0.403	0.702	0.939	0.588
IPSL-CM5A-LR	[-]	165.	150.	113.	11.7	0.515	1.18	2.68	1.20	0.327	0.352	0.781	0.896	0.542
IPSL-CM6A-LR	[-]	115.	109.	113.	5.27	0.708	0.111	1.39	1.14	0.547	0.477	0.790	0.961	0.650
MeanCMIP5	[-]	121.	115.	114.	6.65		0.574	1.41	0.981	0.494	0.502	0.799	0.965	0.652
MeanCMIP6	[-]	116.	110.	114.	6.26		0.129	1.17	0.931	0.572	0.522	0.826	0.956	0.679
MIROC-ESM	[-]	129.	118.	102.	9.04	11.4	0.396	1.90	1.27	0.463	0.435	0.767	0.920	0.604
MIROC-ESM2L	[-]	116.	104.	113.	9.90	0.119	-0.0111	1.95	1.99	0.409	0.379	0.628	0.920	0.543
MPI-ESM-LR	[-]	169.	159.	104.	8.91	9.81	1.36	2.36	1.29	0.402	0.371	0.715	0.930	0.558
MPI-ESM1.2-LR	[-]	141.	133.	104.	6.89	9.81	0.725	2.06	1.13	0.409	0.393	0.769	0.925	0.578
NorESM1-ME	[-]	129.	120.	114.	7.82		0.386	1.86	1.25	0.387	0.456	0.761	0.856	0.583
NorESM2-LM	[-]	107.	97.5	114.	7.59		-0.0828	1.63	1.31	0.443	0.472	0.791	0.938	0.623
UK-HadGEM2-ES	[-]	137.	130.	113.	6.93	0.848	0.602	2.01	1.10	0.389	0.388	0.820	0.855	0.568
UKESM1-0-LL	[-]	126.	119.	113.	7.06	0.825	0.387	1.77	1.16	0.436	0.419	0.791	0.924	0.598

- Multimodel GPP is compared with global seasonal GBAF estimates
- We can see improvements across generations of models (e.g., CESM1 vs. CESM2, IPSL-CM5A vs. 6A)
- The mean CMIP6 and CMIP5 models perform best

Spatial Taylor Diagram



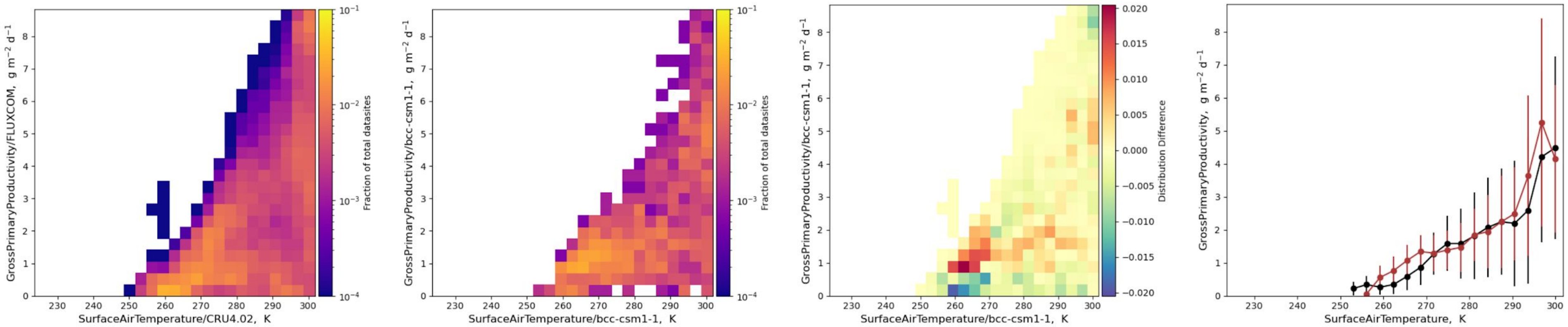
Precipitation/GPCPv2.3



SurfaceDownwardSWRadiation/CERESed4.1

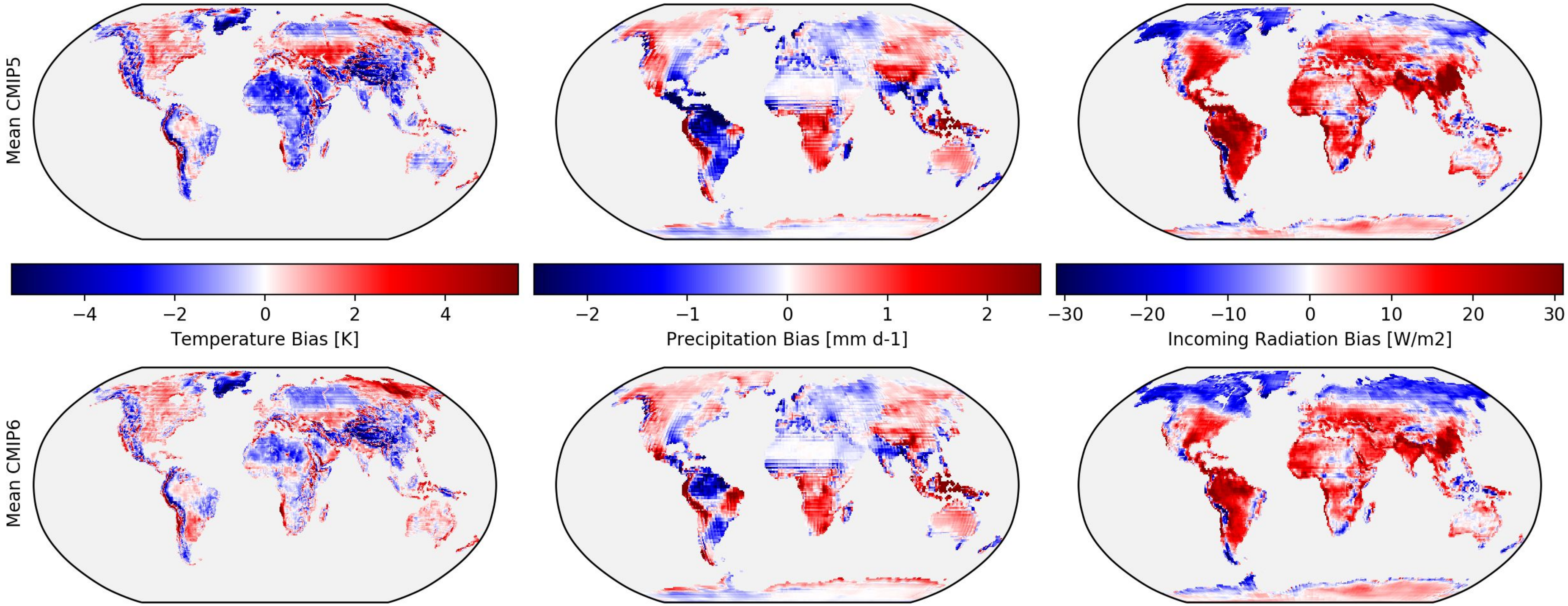
SurfaceNetSWRadiation/CERESed4.1

SurfaceAirTemperature/CRU4.02



Reasons for Land Model Improvements

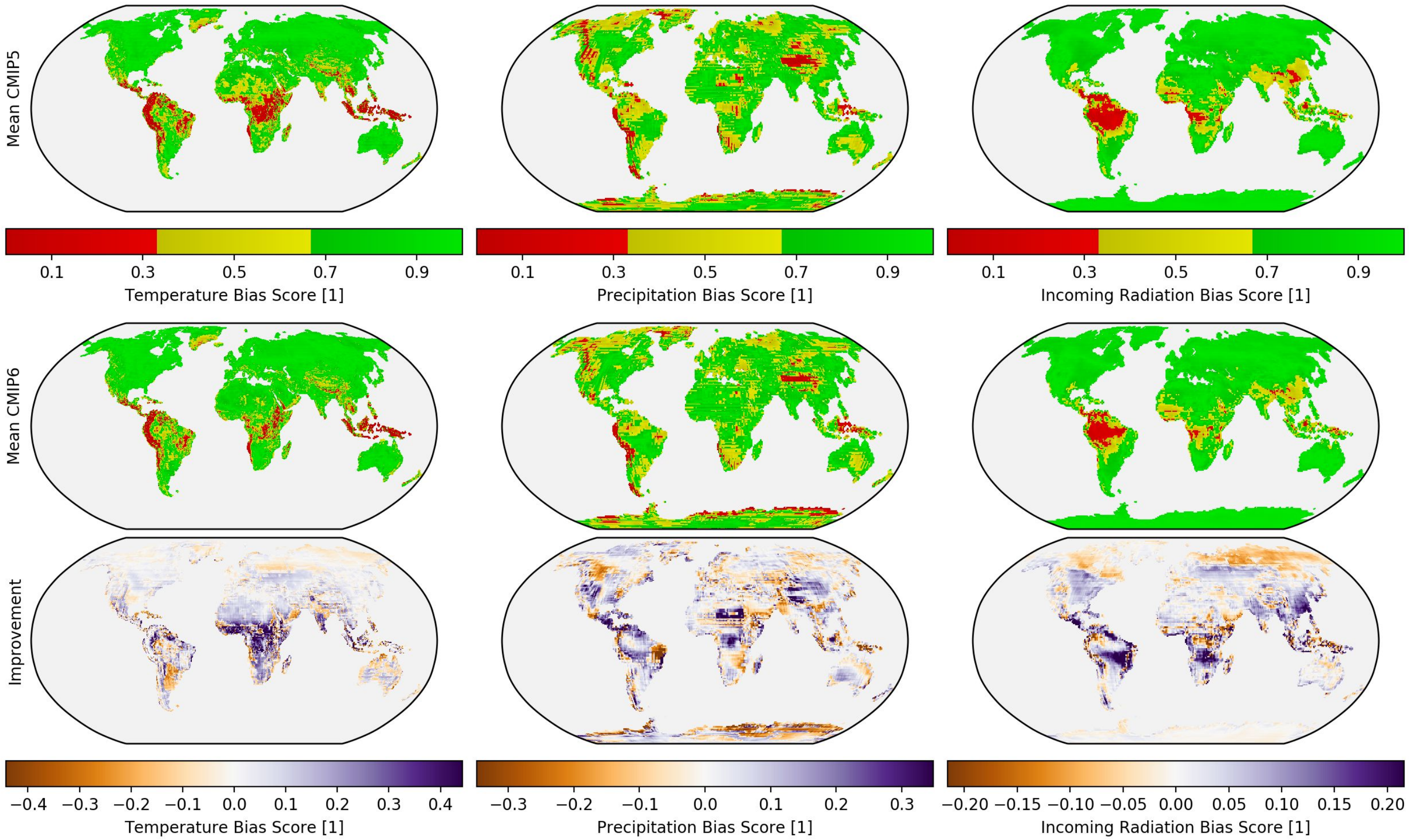
ESM improvements in climate forcings (temperature, precipitation, radiation) likely partially drove improvements exhibited by land carbon cycle models



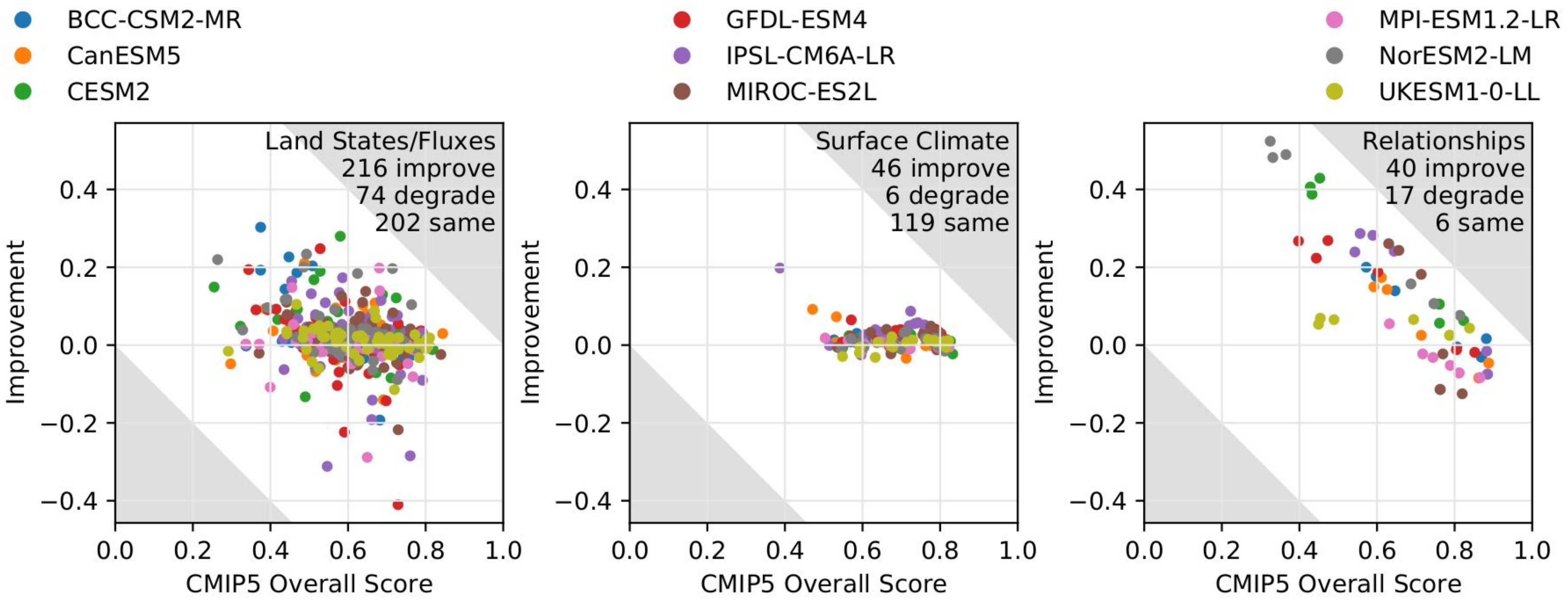
(Hoffman et al., in prep)

Reasons for Land Model Improvements

Differences in bias scores for temperature, precipitation, and incoming radiation were primarily positive, further indicating more realistic climate representation



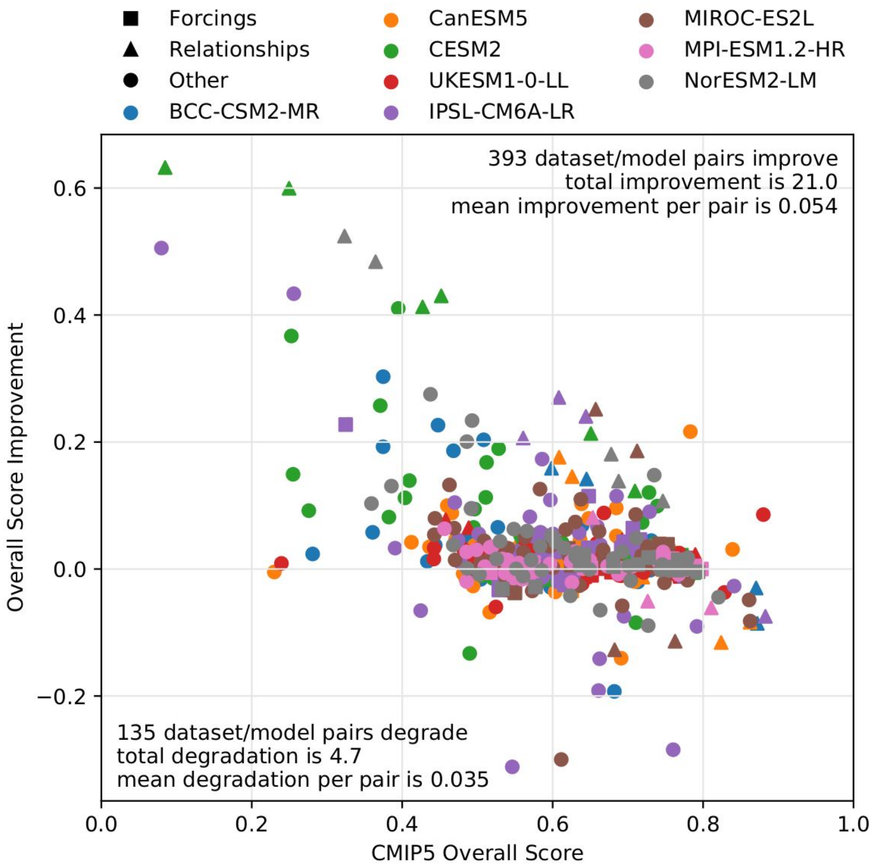
(Hoffman et al., in prep)



Across all land models, scores for most state and flux variables improved (216) or remained nearly the same (202), although some were degraded (74). While atmospheric forcings from CMIP6 ESMs were improved over those from CMIP5 ESMs, the largest improvements were in land model **variable-to-variable relationships**, suggesting that increased land model development was also partially responsible for higher CMIP6 land model scores.

Reasons for Land Model Improvements

While forcings got better, the largest improvements were in **variable-to-variable relationships**, suggesting that increased land model complexity was also partially responsible for higher CMIP6 model scores



International Ocean Model Benchmark (IOMB)

RUBISCO

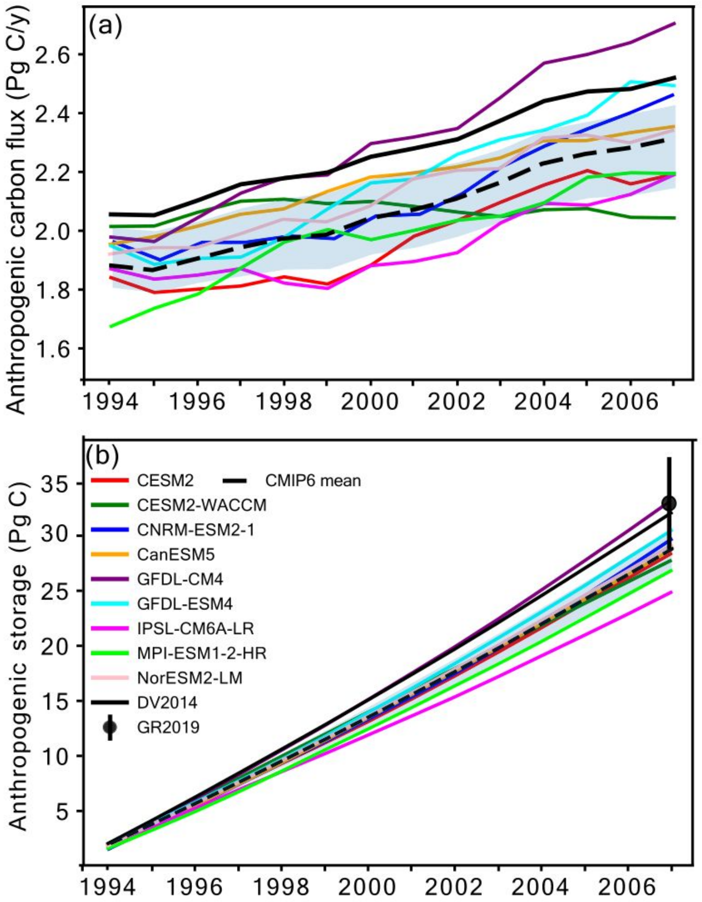
- IOMB is a package for evaluating surface and upper ocean biogeochemistry & carbon uptake
- IOMB software uses the same code base as ILAMB
- We used IOMB to compare CMIP5 vs. CMIP6 models
- We found some improvement from the mean CMIP5 model to the mean CMIP6 model
- Global model estimates of anthropogenic carbon uptake do not change significantly from CMIP5 to CMIP6 models

Relative Scale
Worse Value Better Value
Missing Data or Error

	CMIP5											CMIP6										
	CanESM2	CESM1-BGC	CNRM-CM5	GFDL-ESM2G	GFDL-ESM2M	HadGEM2-ES	IPSL-CM5A-LR	IPSL-CM5A-MR	MPI-ESM-LR	MPI-ESM-MR	NorESM1-ME	CanESM5	CESM2	CNRM-ESM2	GFDL-CM4	GFDL-ESM4	IPSL-CM6A-LR	MPI-ESM1-2-HR	NorESM2-LM	UKESM	meanCMIP5	meanCMIP6
⊟ Ecosystems		1.95	0.10	-0.03	-0.13	0.00	-0.43	0.11	-0.19	-0.13		-0.60	0.60	0.30	1.98	-0.60	-0.48	-0.89	-2.16	0.05	0.08	0.46
└ ⊞ Chlorophyll	-1.81	1.81	0.69	0.12	-0.20	0.24	0.69	1.03	0.16	0.18		-0.98	0.56	0.15	1.93	-0.52	-0.22	-0.73	-2.50	-0.14	-0.19	-0.28
└ ⊞ Oxygen surface		0.83	-0.82	-0.04	0.34	-0.20	-1.89	-1.41	-0.44	-0.36	-1.44	0.82	0.43	0.56	0.67	-0.00	-0.32	-0.21	0.49	0.58	0.73	1.66
⊟ Nutrients		-0.76	-1.41	-0.03	-0.78		0.95	1.21	-0.71	-0.79	-1.15		0.04	0.93		1.04	1.90	-0.82	-1.05	-0.10	-0.10	1.63
└ ⊞ Nitrate surface	0.27	-1.62	-1.17	0.75	-0.68	0.89	1.31	1.47	-0.13	-0.03	-1.69	1.29	-0.87	0.57		0.35	1.29	0.46	-1.77	-0.52	-0.43	0.24
└ ⊞ Phosphate surface		-0.57	-1.32	0.08	-0.17		0.15	0.59	-0.34	-0.44	-0.31		0.50	-0.74	0.17	-0.02	0.28	-0.86	0.12	0.14	1.00	1.75
└ ⊞ Silicate surface		0.49	-0.54	-0.72	-0.54	-2.25	0.28	0.34	-0.82	-0.99	-0.19		0.55	1.33		1.34	1.71	-1.25	-0.17	0.22	-0.28	1.48
⊟ Carbon												1.37	-0.19	-0.68	-1.22	-0.60	-0.68	-0.74	1.44			1.31
└ ⊞ TAlk surface	-0.36	1.05	-0.57	0.07	0.61	-0.30	0.14	0.68	0.29	0.15	-2.60	0.00	-0.46	-2.69	-0.11	0.87	-0.53	0.55	1.34	0.00	1.33	0.53
└ ⊞ DIC surface	0.74	0.00	-1.81	-0.17	0.00	-0.82	-0.25	0.21	-0.20	-0.60	-2.71	0.55	-0.59	0.81	-0.20	0.21	1.16	0.01	1.34	0.97	0.58	0.74
└ ⊞ Anthropogenic DIC 1994-2007												1.47	-0.01	-0.41	-1.24	-0.76	-0.76	-0.83	1.25			1.30
⊟ Physical Drivers		0.19	0.67	-1.22	-0.53				-1.46	-1.24	-0.81	0.41	0.65	0.75		0.10	0.23	-0.18	-1.32	0.53	1.09	2.14
└ ⊞ Mixed Layer Depth		0.28	1.33	-1.36	-0.54				-1.88	-1.72	0.15	0.67	1.30	0.48		0.11	0.39	0.57	-1.27	0.53	-0.35	1.28
└ ⊞ Temperature surface	0.47	0.22	-0.34	-1.20	-0.76	-0.51	-1.43	-1.09	-0.11	-0.03	-0.39	0.30	0.61	0.03	0.10	0.25	-0.27	-0.20	-0.49	1.28	1.63	1.91
└ ⊞ Temperature 200m	0.20	0.41	0.71	-0.89	-0.51	-1.32	-0.04	-0.16	-0.95	-0.67	-1.15	0.33	0.49	1.06	-0.49	-0.69	0.28	-0.36	-1.05	1.03	1.67	2.08
└ ⊞ Temperature 700m	0.74	0.44	0.51	-0.31	-0.43	0.33	0.59	0.52	-2.00	-2.22	0.11	-1.02	0.19	1.32	-0.17	-0.54	0.17	-1.92	0.16	1.06	1.33	1.13
└ ⊞ Vertical temperature gradient	-0.13	0.76	-0.02	-0.09	-0.54	-2.24	-0.27	-0.24	-0.39	0.51	-2.06	0.51	0.74	0.81	1.15	0.46	-0.69	0.11	-1.88	0.89	1.08	1.53
└ ⊞ Salinity surface	-1.25	-0.03	-0.22	-0.72	0.18	-1.81	0.20	0.58	0.01	0.12	-1.45	0.31	0.25	-1.02	1.21	1.59	0.10	-0.33	-0.25	-1.02	1.40	2.16
└ ⊞ Salinity 200m	-0.43	-0.22	-0.02	-0.71	0.07	-0.90	0.39	0.30	-0.41	-0.14	-0.89	0.34	-0.61	0.94	0.29	0.01	0.51	-0.13	-1.12	-0.36	1.26	1.84
└ ⊞ Salinity 700m	-0.33	-1.04	0.00	-0.53	0.10	0.34	0.71	1.00	-0.44	-0.76	-0.79	0.27	-1.15	1.15	-0.12	-0.46	0.56	-0.37	-0.86	-0.52	1.59	1.65
⊟ Relationships		-1.27	-2.65	-0.12	-1.70	0.63	-0.08	0.19	1.31	1.11	-0.16	0.15	0.71	-0.55		1.07	0.29	0.18		-0.70	0.45	1.12
└ ⊞ Oxygen Surface vs Temperature Surface		0.32	-0.87	0.28	0.40	-0.34	-0.58	-0.61	-0.21	-0.25	-0.07	0.35	-0.15	0.53	0.25	0.25	0.23	-0.02		0.08	-0.18	0.58
└ ⊞ Nitrate Surface vs Temperature Surface	-2.02	-1.12	-2.16	-0.06	-1.53	0.75	0.14	0.39	1.34	1.17	-0.03	0.18	0.79	-0.50		1.04	0.33	0.27		-0.55	0.55	1.02

(Fu et al., 2022)

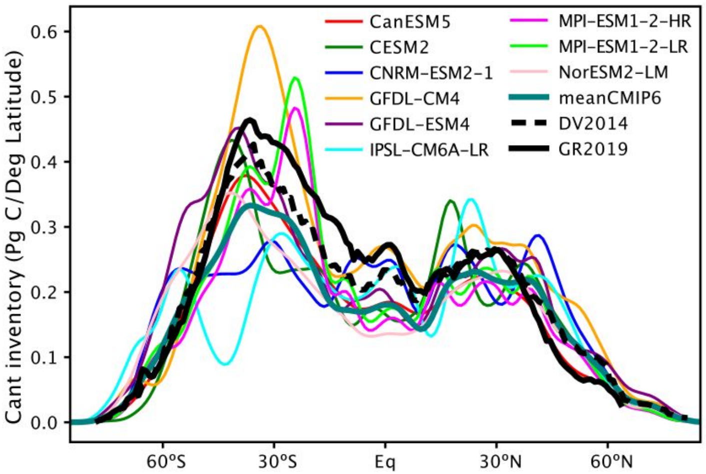
IOMB Evaluation of CMIP5 vs. CMIP6 Models



- Large variation in anthropogenic carbon flux across models [Left (a)]
- Cumulative carbon storage (1994–2007) across models is weak compared to data constraints from Gruber et al. (2019) and DeVries (2014) [Left (b)]
- Ensemble mean uptake is 27.8±0.5 PgC compared with data-derived estimate of 33.0±4.0 PgC

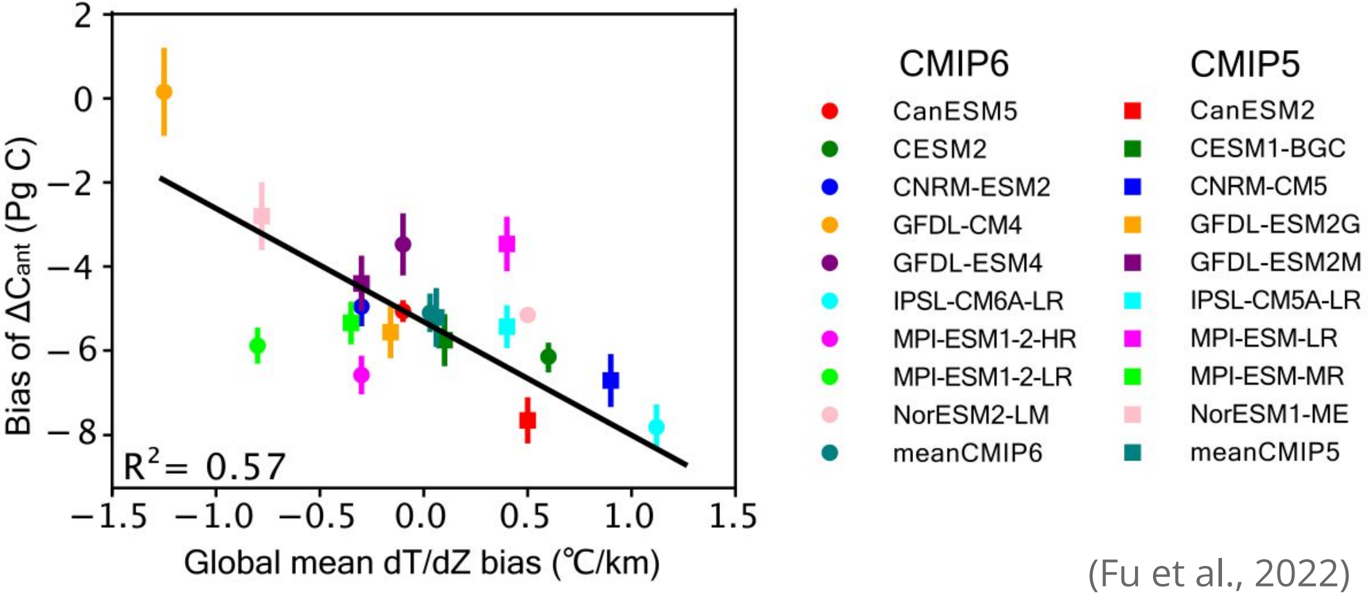
(Fu et al., 2022)

- Many models exhibit the largest negative biases in anthropogenic carbon uptake (1994–2007) in the region of 30°N to 50° N [Right]



IOMB Evaluation of CMIP5 vs. CMIP6 Models

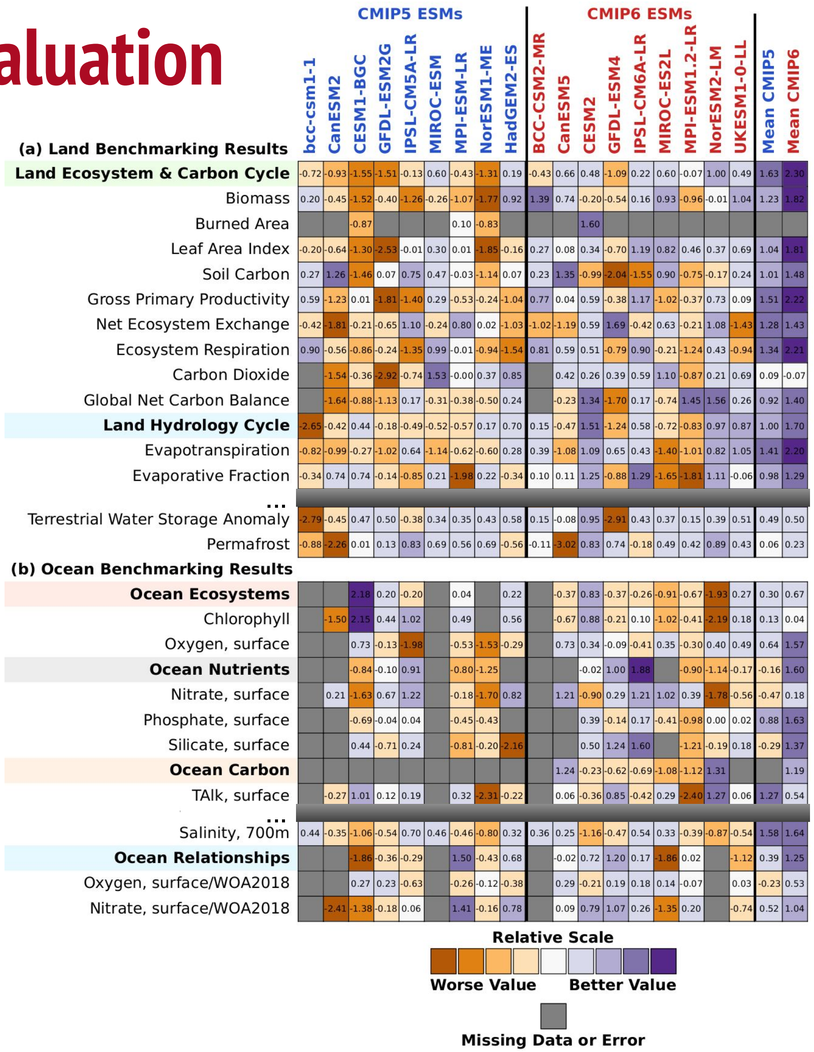
- We found a linear relationship between the bias of the vertical temperature gradient and the bias in global anthropogenic carbon uptake in models
- Consistent with hypothesis that model biases in anthropogenic carbon uptake are related to biases in surface-to-interior exchange by physical processes
- Analysis of CFC11 tracer and DIC bias from four of the models (not shown) is consistent with a weak exchange between surface and interior ocean



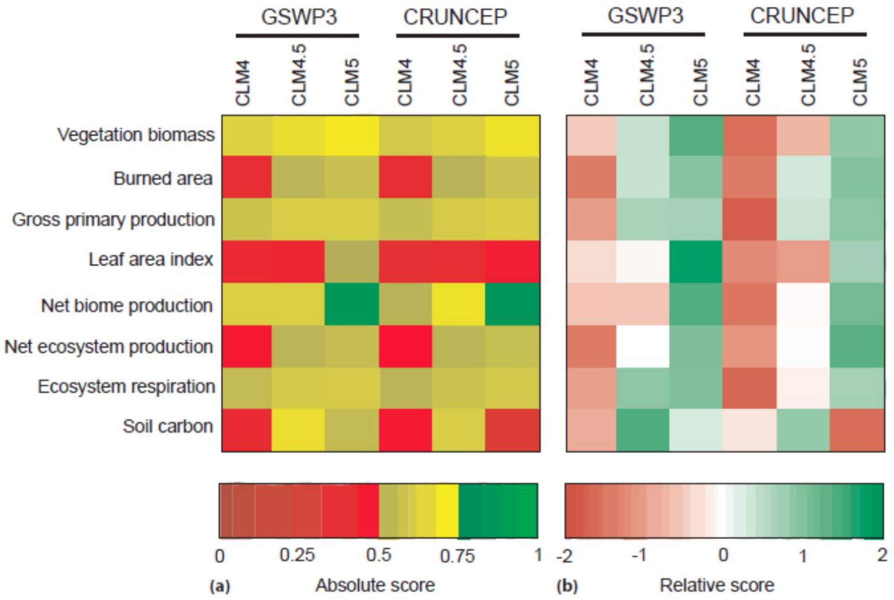
(Fu et al., 2022)

ILAMB & IOMB CMIP5 vs 6 Evaluation

- (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

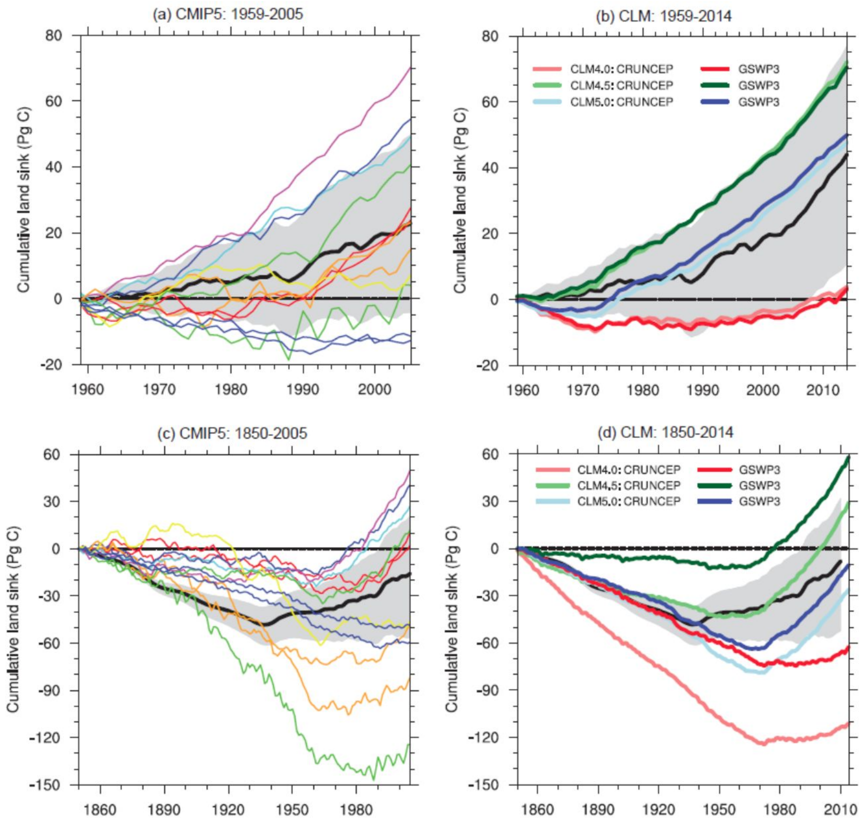


Land Model Performance Depends Strongly on Forcing



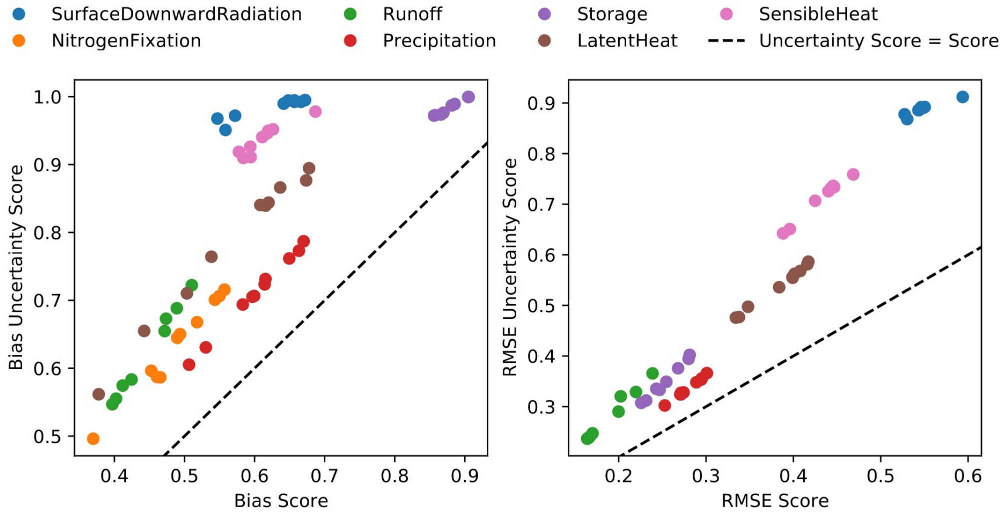
ILAMB performance for CLM4, CLM4.5, and CLM5 forced with GSWP3 vs. CRUNCEP (left) and the cumulative land carbon sink for CMIP5 vs. CLM offline models (right).

Bonan et al. (2019)



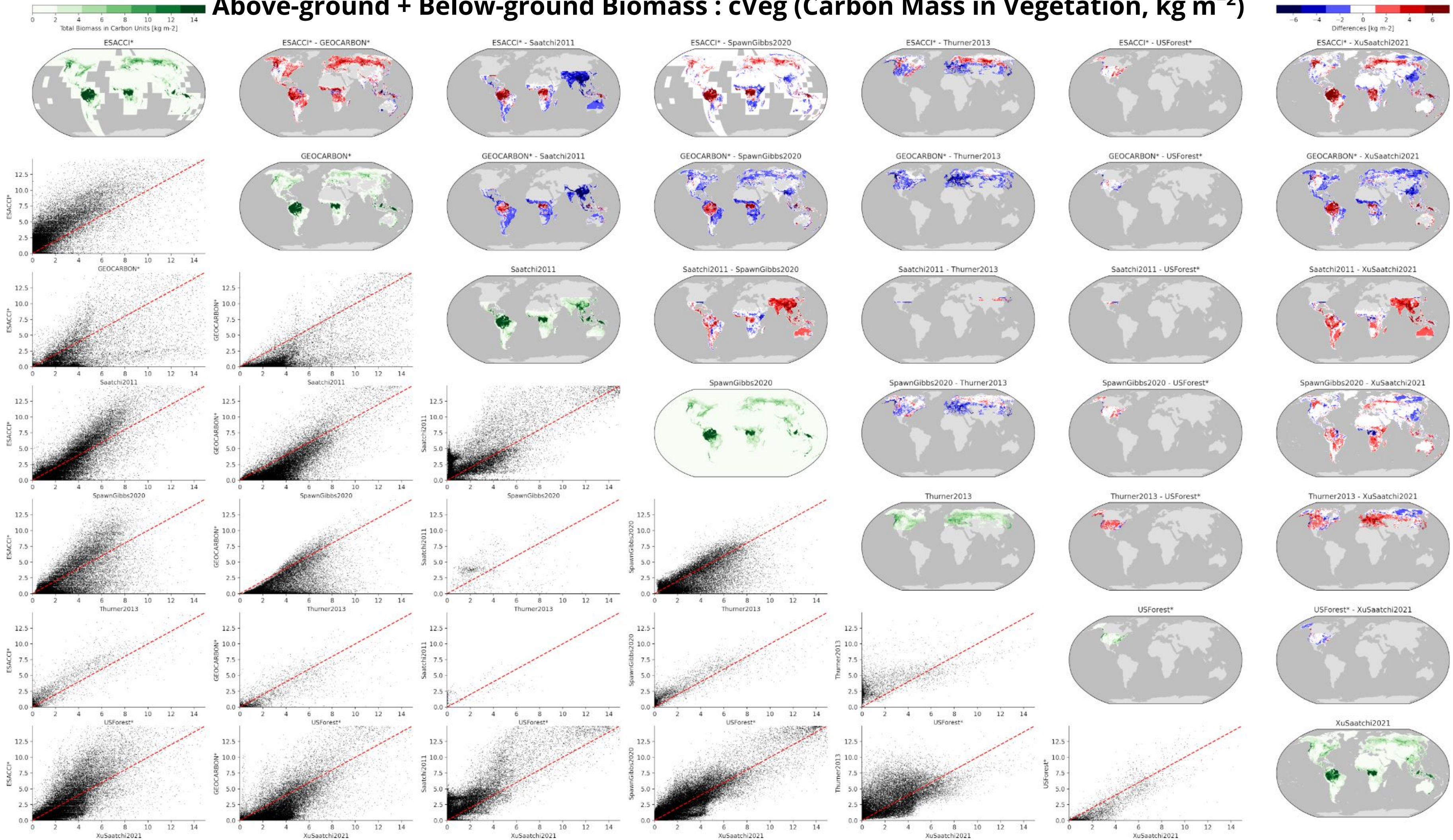
Addressing Observational Uncertainty

- Few observational datasets provide complete uncertainties, but some are appearing
- ILAMB uses multiple datasets for most variables and allows users to weight them according to a rubric of uncertainty, scale mismatch, etc.
- ILAMB can also use:
  - Full spatial/temporal uncertainties provided with the data
  - Fixed, expert-derived uncertainty for a dataset
  - Uncertainties derived from combining multiple datasets



- Experiments with self-consistent CLASS data (Hobeichi et al. 2020) and Barnard's nitrogen fixation data demonstrate that while scores shift, including uncertainty rarely alters the rank ordering of models (figure)

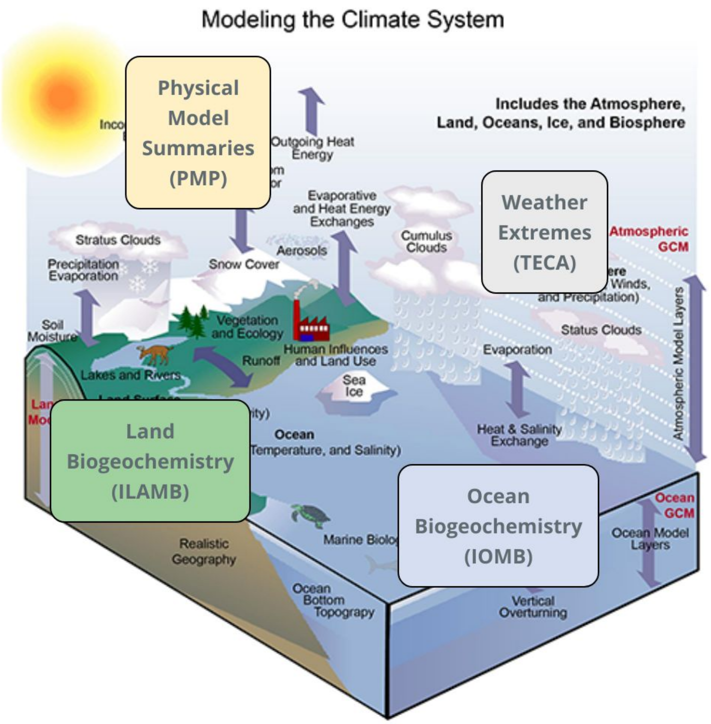
Above-ground + Below-ground Biomass : cVeg (Carbon Mass in Vegetation, kg m^{-2})



Home Capabilities Results Protocols Contacts

Coordinated Model Evaluation Capabilities

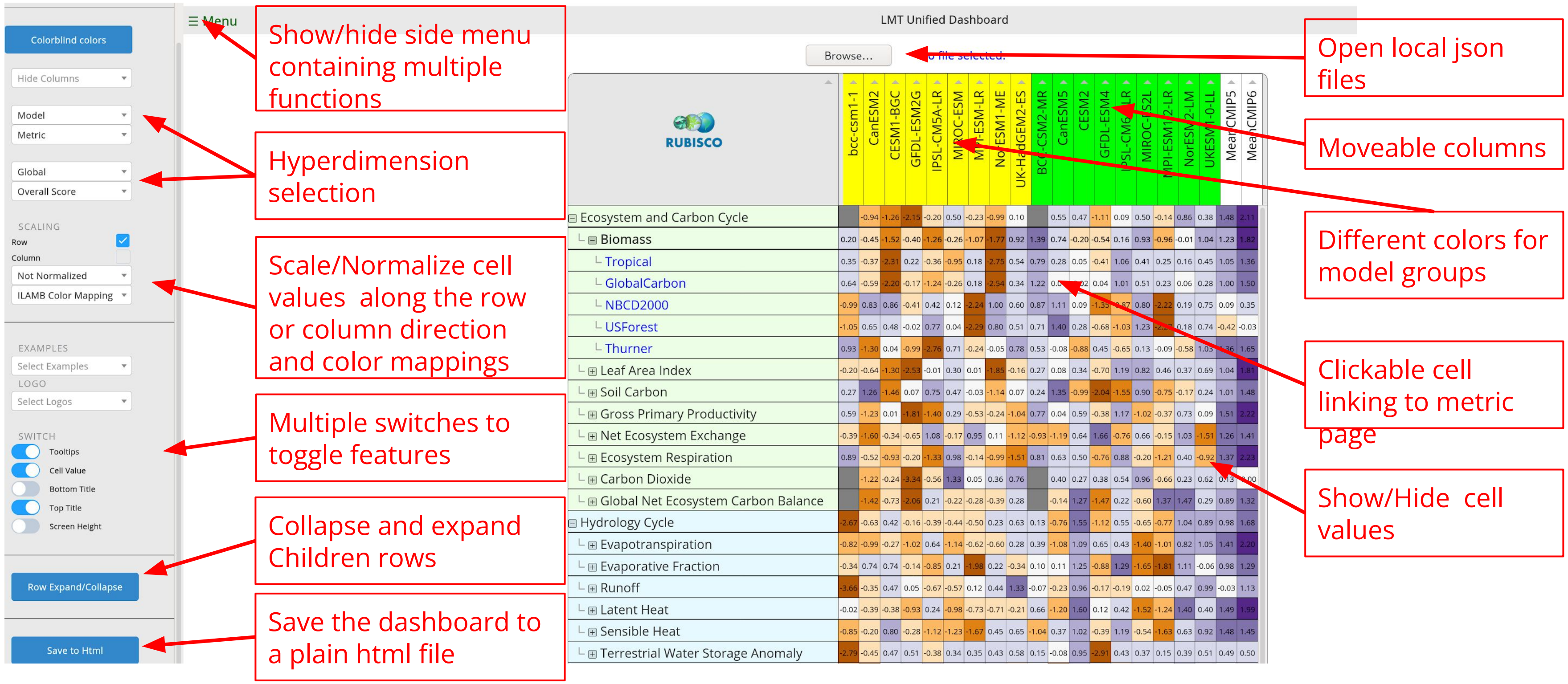
Coordinated Model Evaluation Capabilities (CMEC) is an effort to bring together a diverse set of analysis packages that have been developed to facilitate the systematic evaluation of Earth System Models (ESMs). Currently, CMEC includes three capabilities that are supported by the U.S. Department of Energy, Office of Biological and Environmental Research (BER), Regional and Global Climate Modeling Program (RGCM). As CMEC advances, additional analysis packages will be included from community-based expert teams as well a efforts directly supported by DOE and other US and international agencies.



https://cmec.llnl.gov/

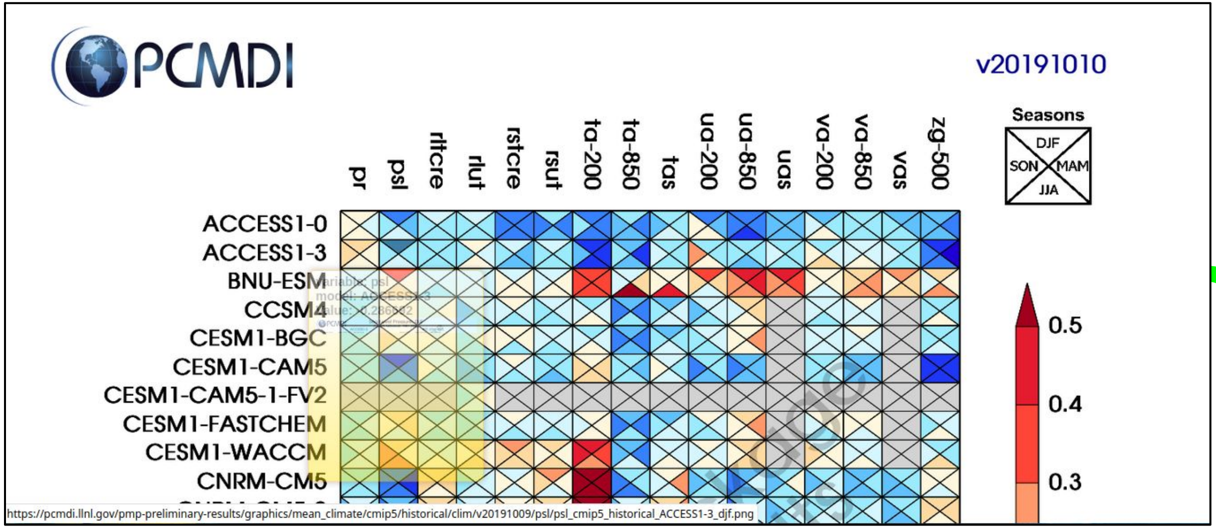
A primary motivation for CMEC is to analyze model simulations that are contributed to the Coupled Model Intercomparison Project (CMIP). Virtually every institution worldwide involved in significant

LMT Dashboard: https://lmt.ornl.gov/unified-dashboard/



- **Tooltips:** show scores when mouse hovers the cells.
- **Column Hiding:** hide some models (columns) to focus into models of interest.
- **Column sorting:** sort the scores along the columns/models to see the best metric for the model.

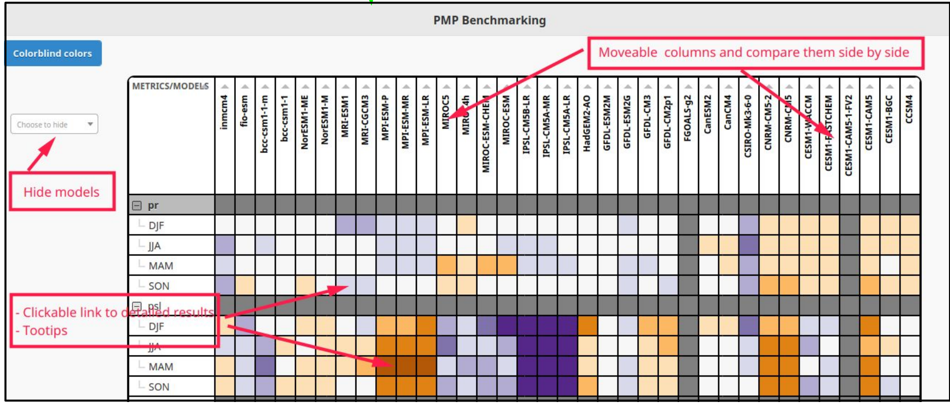
Convert other diagnostic results for use in LMT dashboard



https://lmt.ornl.gov/tab_pmp

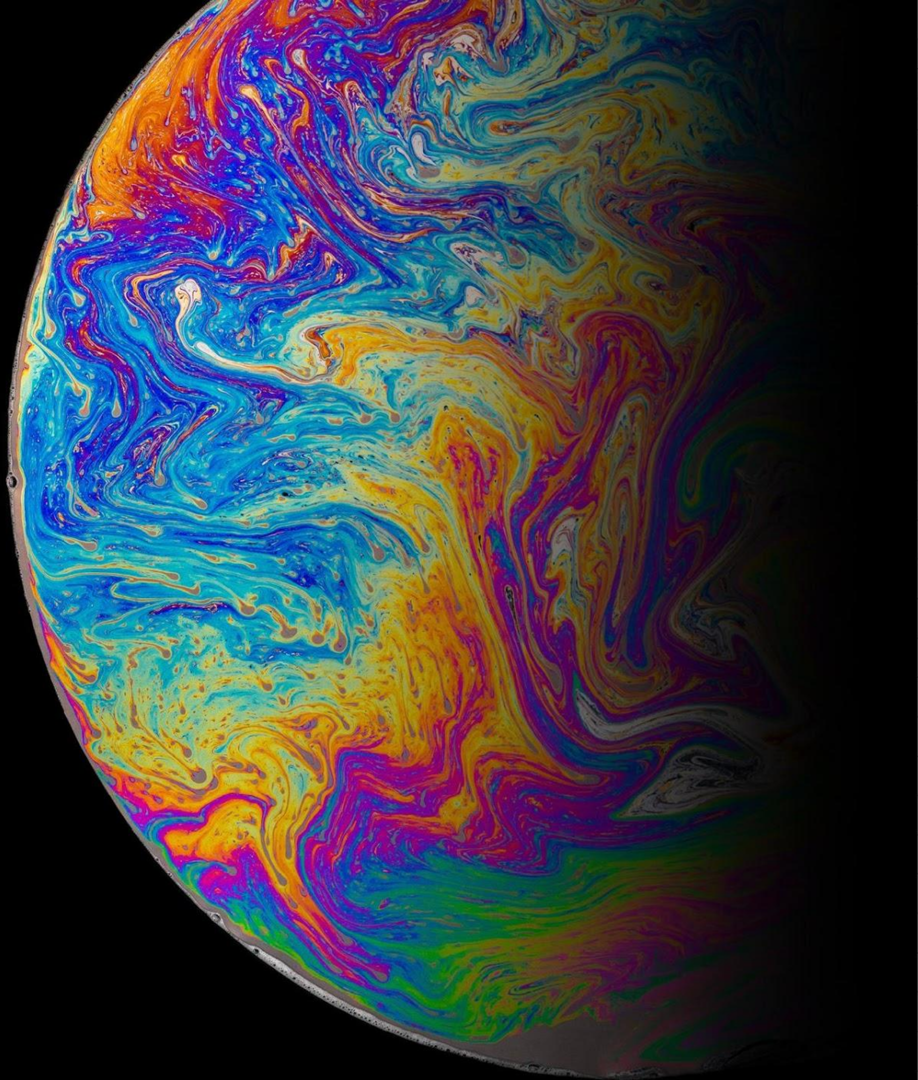
PMP: The Program for Climate Model Diagnostics and Intercomparison (PCMDI) Metrics Package (PMP)

- Clicking cell will go to maps of geographic distributions generated by PMP
- Our LMT dashboard can be used to study science questions like ENSO-BGC feedbacks



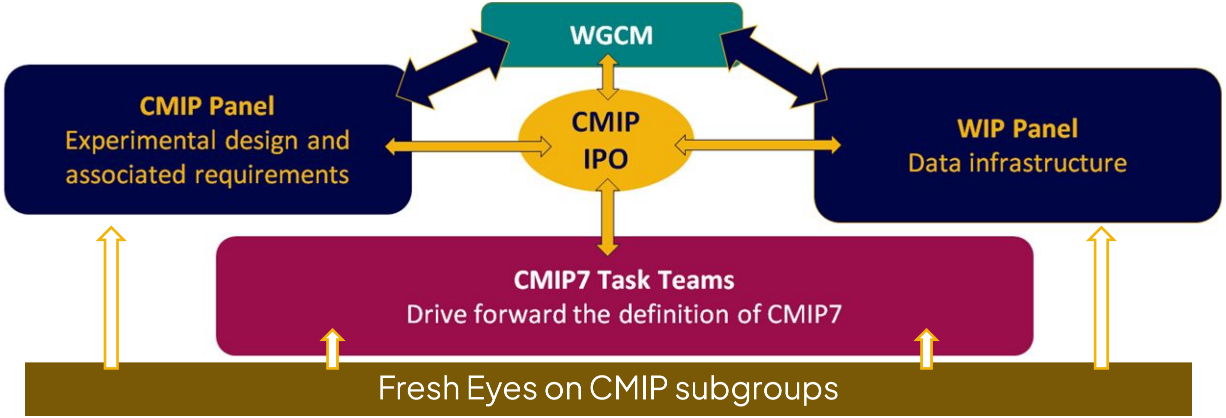
CMIP Coupled Model Intercomparison Project
Model Benchmarking

Climate Model Benchmarking for CMIP7



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

CMIP and Preparations for CMIP7



- CMIP is part of the WCRP's Earth System Modelling and Observation (ESMO) realm, and the Working Group on Coupled Modelling (WGCM)
- CMIP activities are coordinated by the CMIP Panel, the WGCM Infrastructure Panel (WIP), the CMIP IPO and the CMIP7 Task Teams

- Two new CMIP co-chairs were appointed (2022/2023):
  - Helene Hewitt (Met Office, UK)
  - John Dunne (NOAA GFDL, USA)
- The CMIP IPO, newly established in March 2022 and hosted at the European Space Agency's (ESA's) ECSAT site in Harwell, UK, is staffed by five people and is tasked with helping coordinate and support CMIP activities and responsibilities

See **Co-creating the Future of CMIP (Coupled Model Intercomparison Project)** on Thursday, 27 June 2024, at 11:00am – 12:30pm in Luge Hall, Alpensia Convention Center

#CMIP

CMIP Task Teams

- The CMIP Panel and WIP have established a number of Task Teams to support the design, scope, and definition of the next phase of CMIP and evolution of CMIP infrastructure and future operationalization
- Individual Task Teams aim to address specific topics (shown at right) and interact with each other to develop recommendations for the CMIP Panel and WIP

- Data Access
  (Robert Pincus & Atef Ben-Nasser)
- Data Citation
  (Martina Stockhause & Sasha Ames)
- Data Request
  (Martin Juckes & Chloe Mackallah)
- Forcings
  (Paul Durack & Vaishali Naik)
- Model Benchmarking
  (Birgit Hassler & Forrest Hoffman)
- Model Documentation
  (David Hassell & Guillaume Levavasseur)
- Strategic Ensemble Design
  (Ben Sanderson & Isla Simpson)

The Model Benchmarking TT

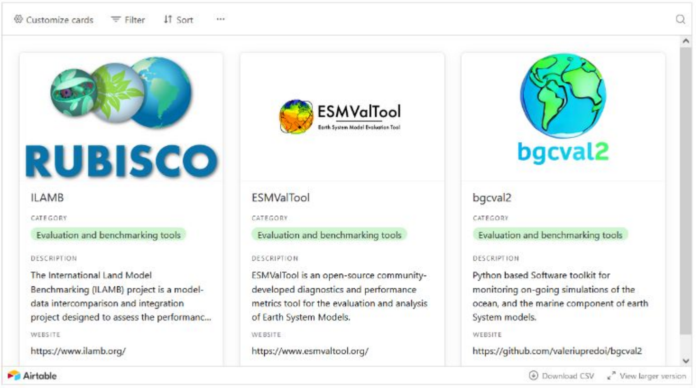
- Rebecca Beadling, *USA*
- Ed Blockley, *UK*
- Jiwoo Lee, *USA*
- Valerio Lembo, *Italy*
- Jared Lewis, *Australia*
- Jianhua Lu, *China*
- Luke Madaus, *USA*
- Elizaveta Malinina, *Canada*
- Brian Medeiros, *USA*
- Wilfried Pokam Mba, *Cameroon*
- Enrico Scoccimarro, *Italy*
- Ranjini Swaminathan,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

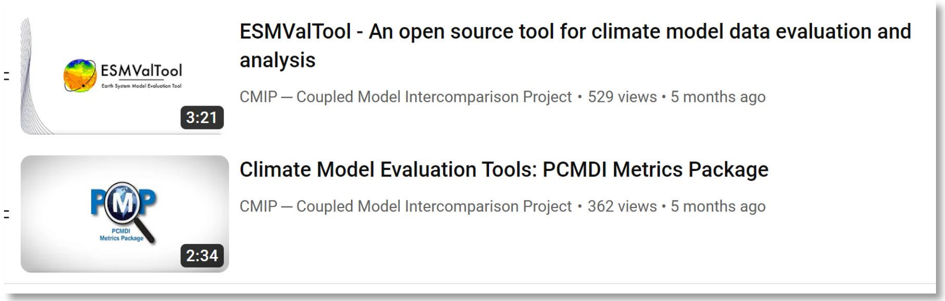
Status: first cards available

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/

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: August 2024

#CMIP

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?

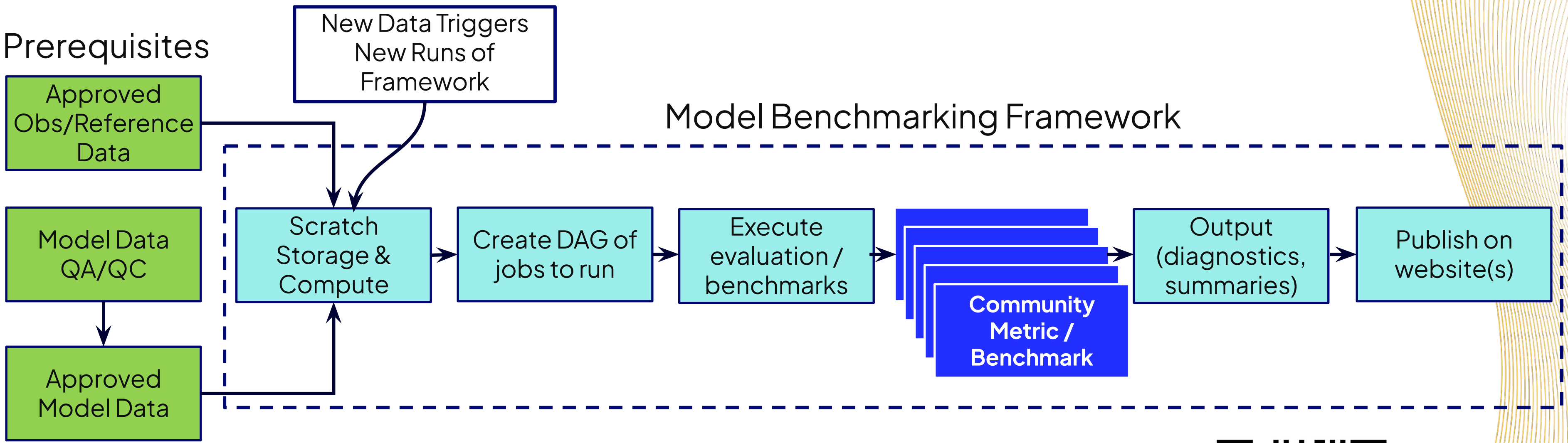
Status: Under development

Planned submission: Summer 2024

Other Planned TT Activities

- In collaboration with Fresh Eyes on CMIP groups
  - Scope out a Rapid Evaluation Framework for automated benchmarking capabilities at the time of AR7 Fast Track data publication
  - Develop scope for better quality assurance / quality control (QA/QC) for CMIP model output
  - Develop a white paper on observational data needs for model benchmarking, including uncertainties

Rapid Evaluation Framework Overview



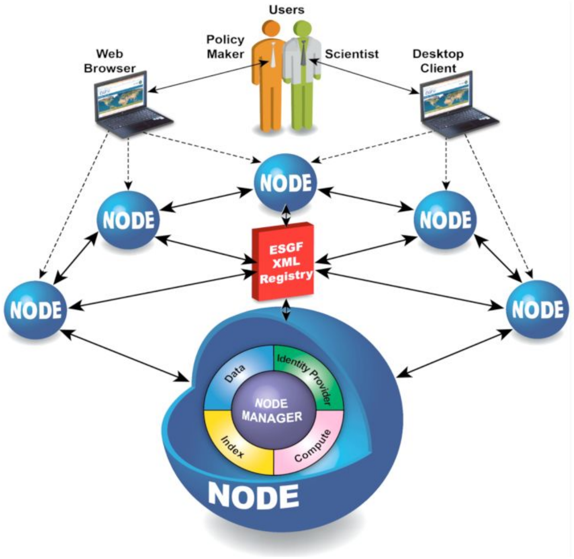
#CMIP

What is the Earth System Grid Federation?

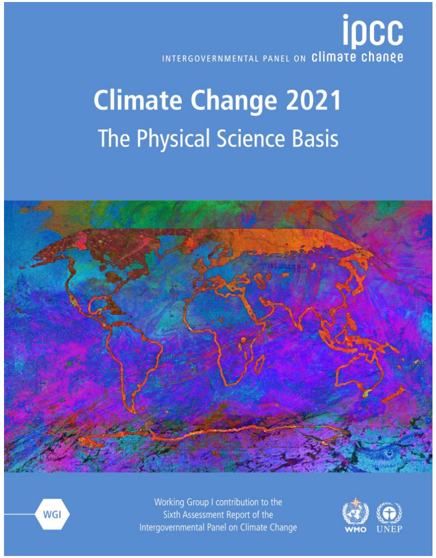
- **Earth System Grid Federation (ESGF)** is an *international consortium* and a *globally distributed peer-to-peer network of data servers* using a common set of protocols & interfaces to archive and distribute climate & Earth system model output and related input, observational, and reanalysis data

- **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



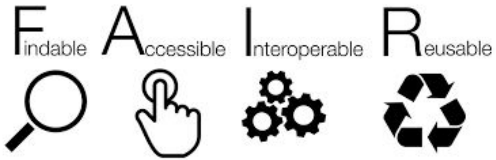
Model output data from ESGF are used for research that underpins IPCC Assessment Reports, like AR6



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

ESGF2 US ESGF Holdings are Open and Large

- CMIP5 totals >1.5 PB (>5 PB including replicas)
- CMIP6 totals >15.9 PB (>27 PB including replicas)
- CMIP7 is expected to have more experiments, high resolution output, and ensembles, totaling ~100 PB
- ESGF is concerned with the *full stack security* and the *integrity of the data*, but we are **not** concerned about controlling *access to the data* (mostly)

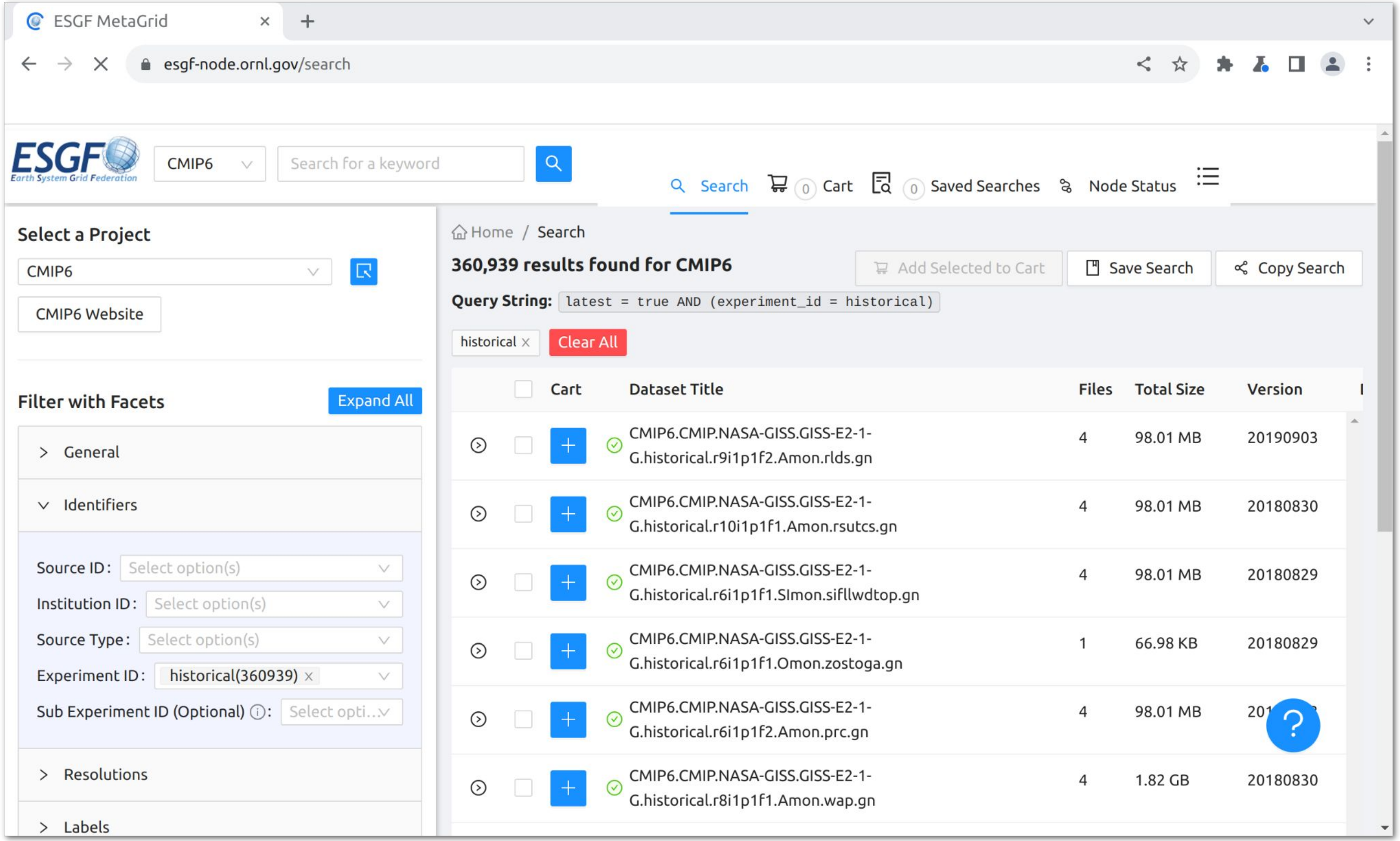


CMIP6: 14,780,267 total datasets, 27,422.79 TB	CMIP6: 7,585,457 distinct datasets, 15,929.59 TB	CMIP6: 7,194,810 replica datasets, 11,493.2 TB
CORDEX: 187,785 total datasets, 1,473.33 TB	CORDEX: 187,513 distinct datasets, 1,472.77 TB	CORDEX: 272 replica datasets, 0.56 TB
CMIP5: 201,130 total datasets, 5,293.61 TB	CMIP5: 52,163 distinct datasets, 1,525.07 TB	CMIP5: 148,967 replica datasets, 3,768.55 TB
INPUT4MIPS: 5,871 total datasets, 10.84 TB	INPUT4MIPS: 0 distinct datasets, 0 TB	INPUT4MIPS: 5,871 replica datasets, 10.84 TB
OBS4MIPS: 126 total datasets, 0.2 TB	OBS4MIPS: 108 distinct datasets, 0.2 TB	OBS4MIPS: 18 replica datasets, 0.01 TB

As of June 20, 2024

Metagrid Data Search Interface

- A redesigned faceted search user interface, called **Metagrid**, replaces the old interface and adds new features
- Offers **shopping cart** and ability to **save & share searches**
- Will soon provide **Globus integration** for fast unattended data transfers

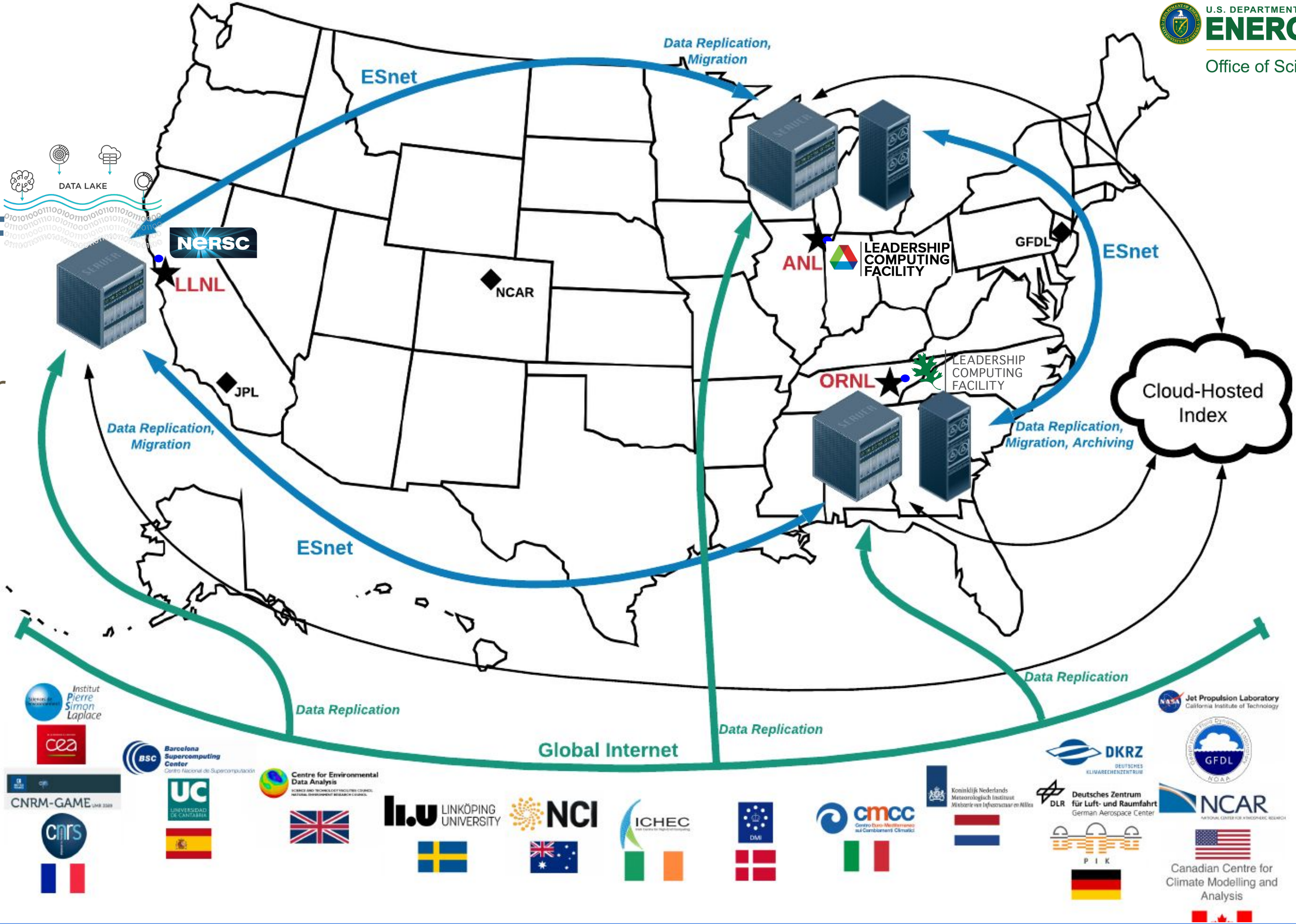


ESGF2 US

U.S. DEPARTMENT OF ENERGY
Office of Science

DOE's Next Generation ESGF

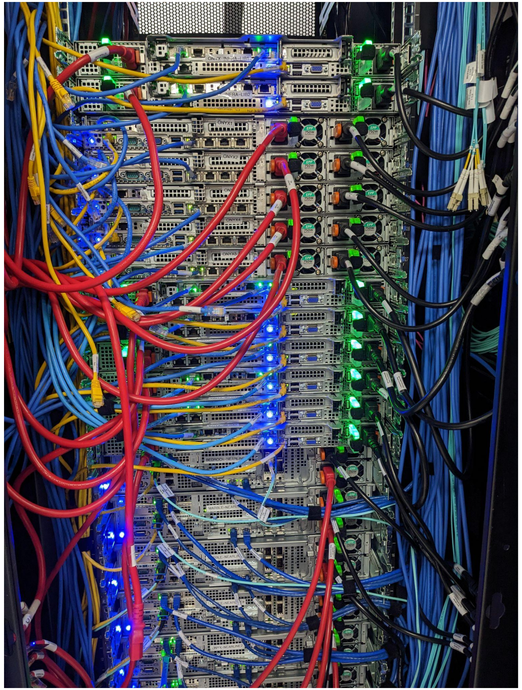
- As many as 3 nodes located at DOE's major computing facilities
- Replicating data from the worldwide Federation
- Providing scalable cloud indexing and tape archiving



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

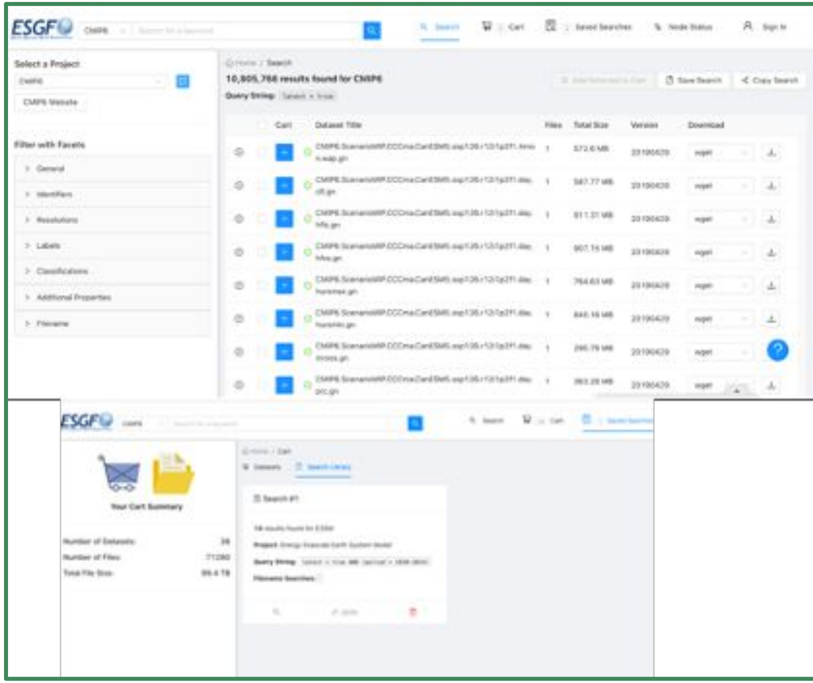
Data and services reside in the Open Network Enclave of NCCS to provide fast and open access to data

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**

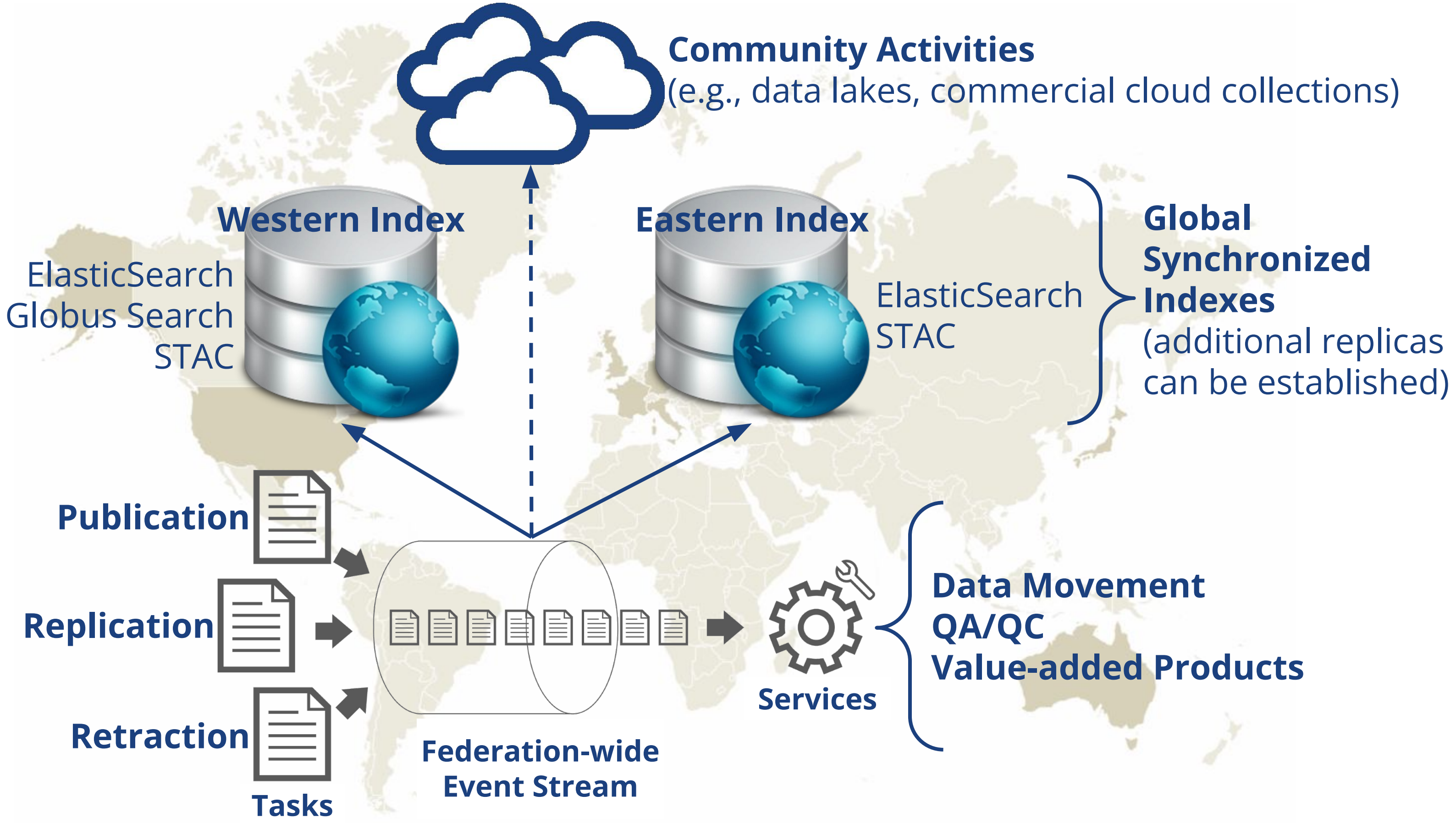
➜ **Tenth ESGF Developer and User Conference scheduled for Rockville, MD, on April 23–26, 2024**

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

Tenth ESGF Conference

- Held 23–26 April 2024 in Rockville, Maryland
- John Dunne joined the meeting to share CMIP priorities and current CMIP 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

Major Accomplishment: New Index Strategy



Other Key Outcomes from the ESGF Conference

- "ESGF1.5" – Current architecture backed by Globus Search instead of SOLR; maintained across the transition to ESGF-NG
- CMIP3, CMIP5, CORDEX-CMIP5 – Not transitioned to ESGF-NG; additional effort required to make it searchable
- CMIP6, CMIP6Plus, input4MIPs, obs4MIPs – Transition to ESGF-NG
- CORDEX-CMIP6 – Data ready to be published soon or now; wait or publish then transition to ESGF-NG?
- "Data Challenges" – Used to demonstrate new technologies across teams
- Draft Timeline for ESGF-NG
- Intermediate face-to-face meeting – In November/December 2024 time frame

Summary

- **Model benchmarking** is increasingly important as model complexity increases
- Systematic model benchmarking is useful for
  - **Verification** – during model development to confirm that new model code improves performance in a targeted area without degrading performance in another area
  - **Validation** – when comparing performance of one model or model version to observations and to other models or other model versions
- The **ILAMB/IOMB package** employs a suite of in situ, remote sensing, and reanalysis datasets to comprehensively evaluate and score model performance, *irrespective of any model structure or set of process representations*
- ILAMB/IOMB is **Open Source**, is written in **Python**, **runs in parallel** on laptops to supercomputers, and has been **adopted in most modeling centers**
- *Usefulness* of packages depends on the **quality of incorporated observational data**, **characterization of uncertainty**, and **selection of relevant metrics**

Recommendations for the Future of Model Benchmarking

- We need better **characterization of uncertainties** in observational and remote sensing data products
  - Do the data help distinguish models from each other?
  - Do the data help inform us about which combination of process representations are important?
- We need to better characterize and understand the **representativeness** of observations
  - Are in situ measurements representative of the data pixels / model grid cells?
  - What additional data are useful for quantifying representativeness and can this inform or direct measurement campaigns or sampling strategies (Matthias' talk, for example)?
- We need to better understand **how processes scale** across space and through time
  - How do we use measurements from stomata to leaves to organisms to inform process representations at the scales of cohorts to canopies to ecosystems to landscapes to watersheds?
  - Can we maintain a constellation of observational systems that produce data at relevant scales over long time periods as the climate changes?
- We need to characterize **plant traits, ecosystem community dynamics, and land use & land cover change** to inform demographic models
  - Do the data help us understand important plant traits and cohort behavior?
  - Can we capture enough data to inform / constrain models of disturbance and recovery?

Recommendations for the Future of Model Benchmarking

- **How many different models** or model configurations are needed to answer science questions?
  - Are models designed to develop mechanistic understanding or address societally relevant questions?
  - What evaluation metrics should be used for models designed for different purposes?
- How can we **combine multisensor observational data** to better inform process representations in models?
  - Can we use AI/ML to derive synthesized or assimilated data products to constrain models?
  - Can we use data-driven AI/ML approaches to produce online parameterizations, hybrid models, surrogate models, and digital twins?
- How can we best **evaluate long timescale processes** with relatively short timescale remote sensing?
  - Can we trade space for time from representativeness analyses with model ensembles?
  - Does contemporary bias removal reduce future model spread?
  - Can we weight models based on ILAMB scores?
- How can we better **organize our communities** to build better (not more?) models, address uncertainties, engage observational community, prepare for CMIP7, 8, 9?
  - **1st Land Surface Modeling Summit** in Oxford (11–15 Sep 2022), Eleanor Blythe & Dave Lawrence
  - **4th Carbon from Space Workshop** in Frascati (25–28 Oct 2022), ESA & NASA
  - **4th ILAMB Workshop** in USA (Late 2024 / Early 2025)

Thank You!

References (1/2)

Bonan, G. B., D. L. Lombardozzi, W. R. Wieder, K. W. Oleson, D. M. Lawrence, F. M. Hoffman, and N. Collier (2019), Model structure and climate data uncertainty in historical simulations of the terrestrial carbon cycle (1850–2014), *Global Biogeochem. Cycles*, 33(10):1310–1326, doi:10.1029/2019GB006175.

Collier, N., F. M. Hoffman, D. M. Lawrence, G. Keppel-Aleks, C. D. Koven, W. J. Riley, M. Mu, and J. T. Randerson (2018), The International Land Model Benchmarking (ILAMB) system: Design, theory, and implementation, *J. Adv. Model. Earth Syst.*, 10(11):2731–2754, doi:10.1029/2018MS001354.

Eyring, V., P. M. Cox, G. M. Flato, P. J. Gleckler, G. Abramowitz, P. Caldwell, W. D. Collins, B. K. Gier, A. D. Hall, F. M. Hoffman, G. C. Hurtt, A. Jahn, C. D. Jones, S. A. Klein, J. Krasting, L. Kwiatkowski, R. Lorenz, E. Maloney, G. A. Meehl, A. Pendergrass, R. Pincus, A. C. Ruane, J. L. Russell, B. M. Sanderson, B. D. Santer, S. C. Sherwood, I. R. Simpson, R. J. Stouffer, and M. S. Williamson (2019), Taking climate model evaluation to the next level, *Nat. Clim. Change*, 9(2):102–110, doi:10.1038/s41558-018-0355-y.

Fu, Weiwei, J. Keith Moore, François Primeau, Nathan Collier, Oluwaseun O. Ogunro, Forrest M. Hoffman, and James T. Randerson (2022), Evaluation of bcean biogeochemistry and carbon cycling in CMIP Earth system models with the International Ocean Model Benchmarking (IOMB) software system, J. Geophys. Res. Oceans, 127(1), e2022JC018965, doi:10.1029/2022JC018965.

References (2/2)

Hoffman, F. M., C. D. Koven, G. Keppel-Aleks, D. M. Lawrence, W. J. Riley, J. T. Randerson, A. Ahlström, G. Abramowitz, D. D. Baldocchi, M. J. Best, B. Bond-Lamberty, M. G. De Kauwe, A. S. Denning, A. R. Desai, V. Eyring, J. B. Fisher, R. A. Fisher, P. J. Gleckler, M. Huang, G. Hugelius, A. K. Jain, N. Y. Kiang, H. Kim, R. D. Koster, S. V. Kumar, H. Li, Y. Luo, J. Mao, N. G. McDowell, U. Mishra, P. R. Moorcroft, G. S. H. Pau, D. M. Ricciuto, K. Schaefer, C. R. Schwalm, S. P. Serbin, E. Shevliakova, A. G. Slater, J. Tang, M. Williams, J. Xia, C. Xu, R. Joseph, and D. Koch (2017), *International Land Model Benchmarking (ILAMB) 2016 Workshop Report*, Technical Report DOE/SC-0186, U.S. Department of Energy, Office of Science, Germantown, Maryland, USA, doi:10.2172/1330803.

Lawrence, D. M., R. A. Fisher, C. D. Koven, K. W. Oleson, S. C. Swenson, G. B. Bonan, N. Collier, B. Ghimire, L. van Kampenhout, D. Kennedy, E. Kluzek, P. J. Lawrence, F. Li, H. Li, D. Lombardozzi, W. J. Riley, W. J. Sacks, M. Shi, M. Vertenstein, W. R. Wieder, C. Xu, A. A. Ali, A. M. Badger, G. Bisht, M. van den Broeke, M. A. Brunke, S. P. Burns, J. Buzan, M. Clark, A. Craig, K. Dahlin, B. Drewniak, J. B. Fisher, M. Flanner, A. M. Fox, P. Gentine, F. M. Hoffman, G. Keppel-Aleks, R. Knox, S. Kumar, J. Lenaerts, L. R. Leung, W. H. Lipscomb, Y. Lu, A. Pandey, J. D. Pelletier, J. Perket, J. T. Randerson, D. M. Ricciuto, B. M. Sanderson, A. Slater, Z. M. Subin, J. Tang, R. Q. Thomas, M. V. Martin, and X. Zeng (2019), The Community Land Model Version 5: Description of new features, benchmarking, and impact of forcing uncertainty, *J. Adv. Model. Earth Syst.*, 11(12):4245–4287, doi:10.1029/2018MS001583.

Zhu, Q., W. J. Riley, J. Tang, N. Collier, F. M. Hoffman, X. Yang, and G. Bisht (2019), Representing nitrogen, phosphorus, and carbon interactions in the E3SM Land Model: Development and global benchmarking, *J. Adv. Model. Earth Syst.*, 11(7):2238–2258, doi:10.1029/2018MS001571.