U.S. DEPARTMENT OF ENERGY

Office of Science

Update on the International Land Model Benchmarking (ILAMB) Package and IOMB

Forrest M. Hoffman
Oak Ridge National Laboratory

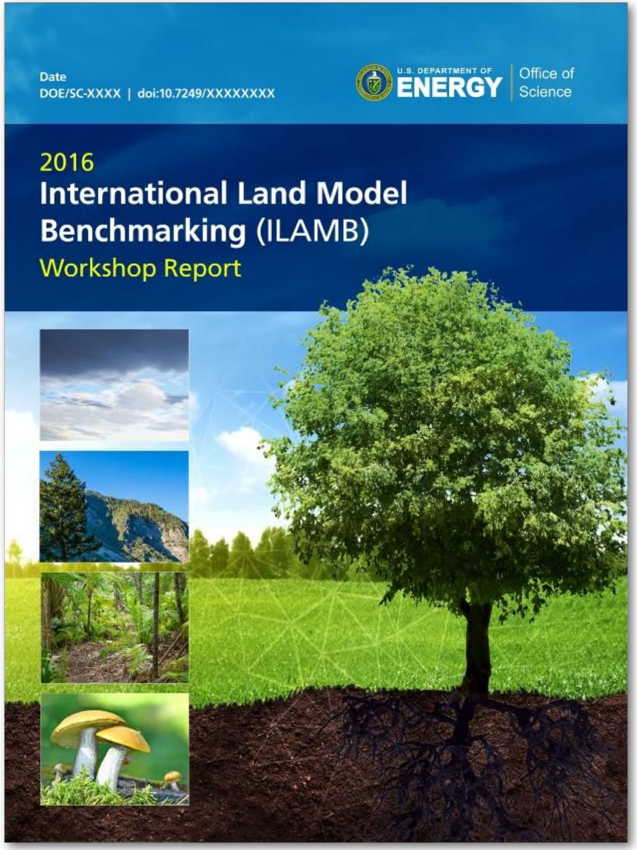
ESS Cyberinfrastructure Working Group Meeting

May 11, 2020

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

The **International Land Model Benchmarking (ILAMB)** community coordination activity was designed to

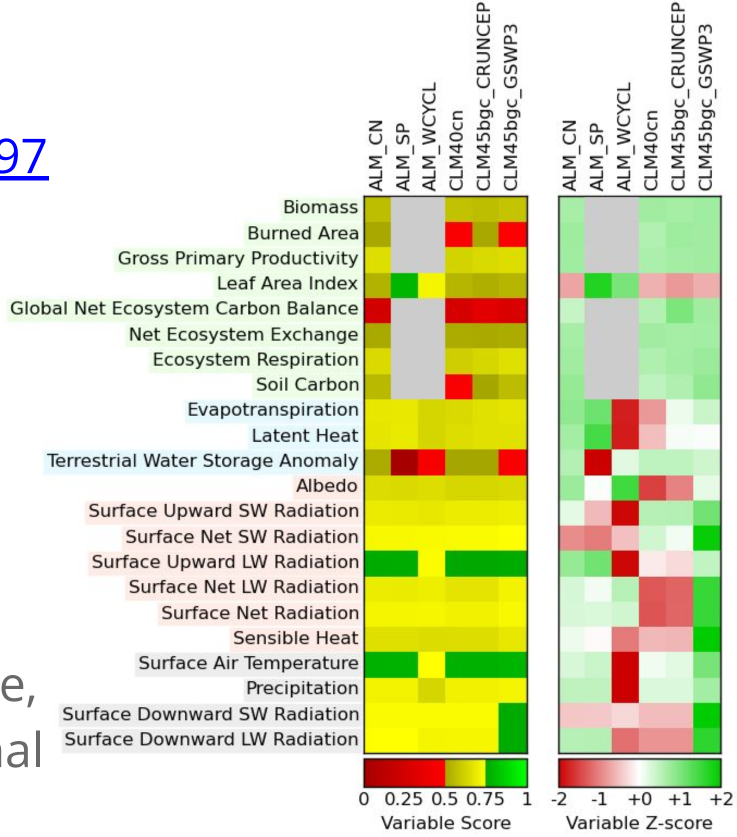
- Develop internationally accepted benchmarks
- Promote the use of these benchmarks
- Strengthen linkages between experimental, remote sensing, and modeling communities
- Support the design and development of open source benchmarking tools (Luo et al., 2012), like the **ILAMB Package** (Collier et al., 2018)



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

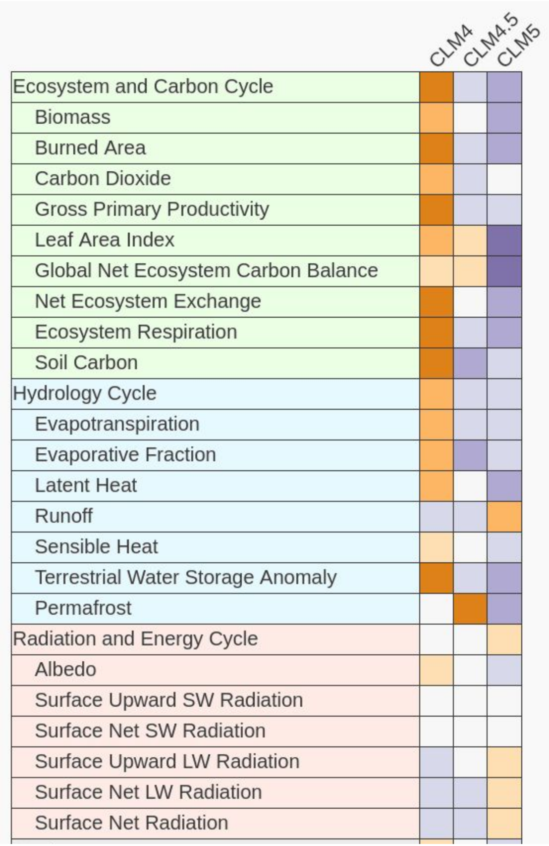
- **ILAMBv1** released at 2015 AGU Fall Meeting Town Hall, doi:10.18139/ILAMB.v001.00/1251597
- **ILAMBv2** released at 2016 ILAMB Workshop, doi:10.18139/ILAMB.v002.00/1251621
- Open Source software freely distributed
- Routinely used for E3SMv1 and CESM2 evaluation during development
- Employed to evaluate CMIP5 models
- Models are scored based on statistical comparisons (bias, RMS error, phase, amplitude, spatial distribution, Taylor scores) and functional response metrics



ILAMBv2.5 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

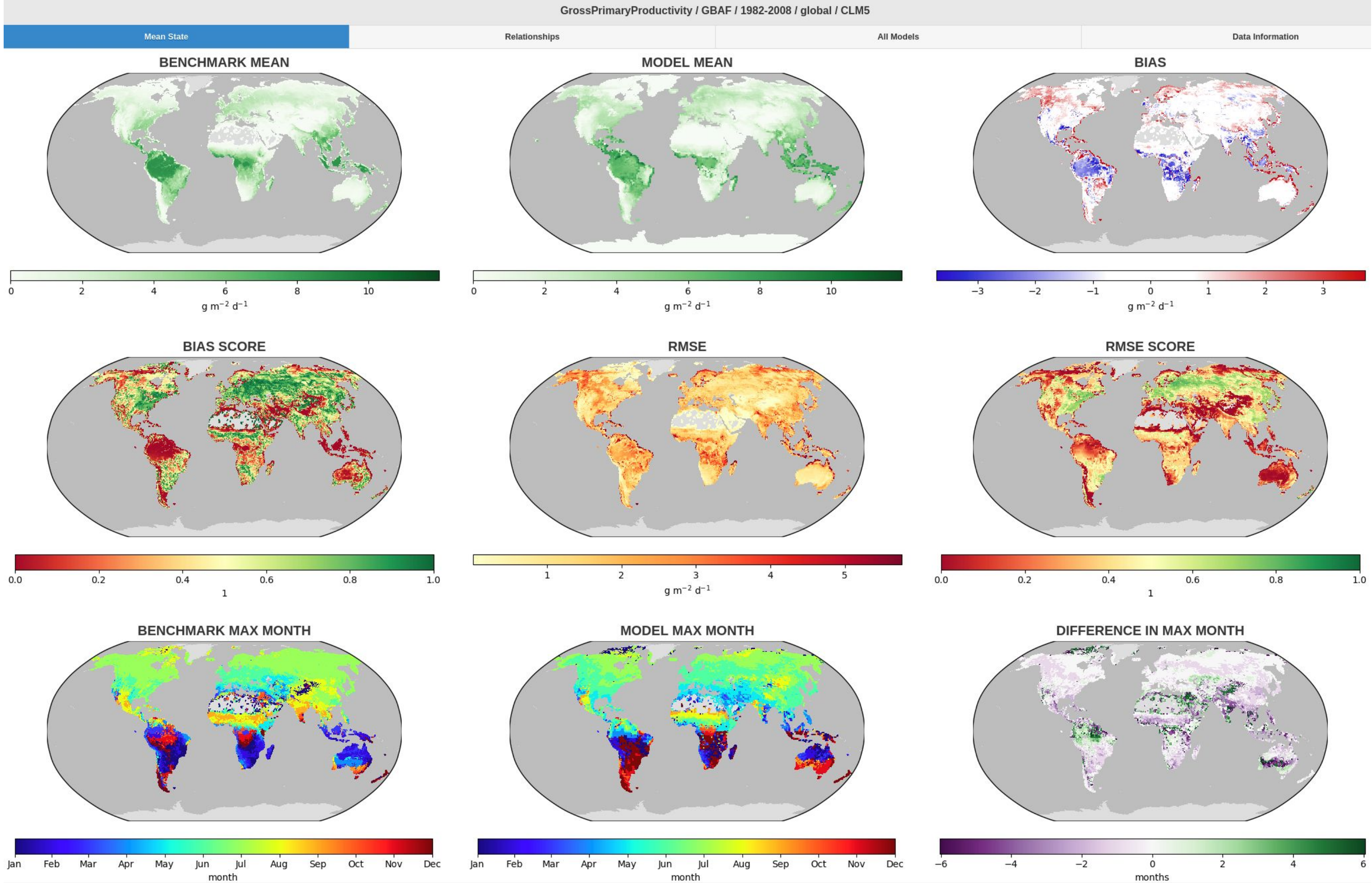


- CLM saw improvements in mechanistic treatment of hydrology, ecology, and land use with many more moving parts
- Simulations improved even with enhanced complexity
- Observational datasets not always self-consistent
- Forcing uncertainty confounds assessment of model development (not shown)

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

(Lawrence et al., 2019)

ILAMB Graphical Diagnostics



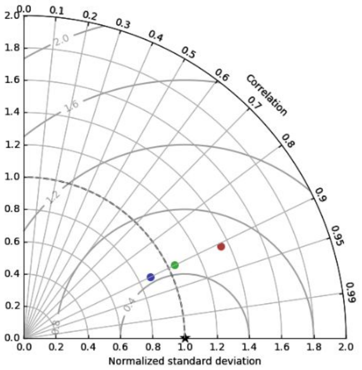
RUBISCO

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

RUBISCO

ILAMB Graphical Diagnostics

SPATIAL TAYLOR DIAGRAM



MODEL COLORS

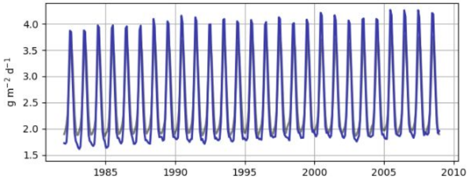
- Benchmark
- CLM4
- CLM4.5
- CLM5

Spatially integrated regional mean

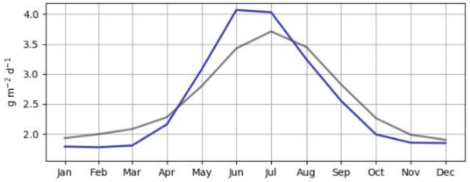
MODEL COLORS

- Benchmark
- CLM4
- CLM4.5
- CLM5

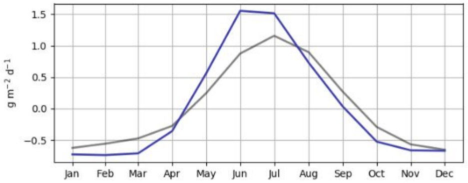
REGIONAL MEAN



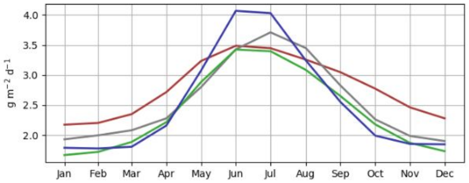
ANNUAL CYCLE



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

New PEcAn-ILAMB site-level diagnostics

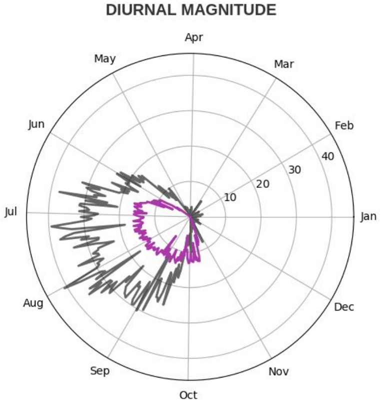
GrossPrimaryProductivity / AMF_US-WCr / global / CLM5

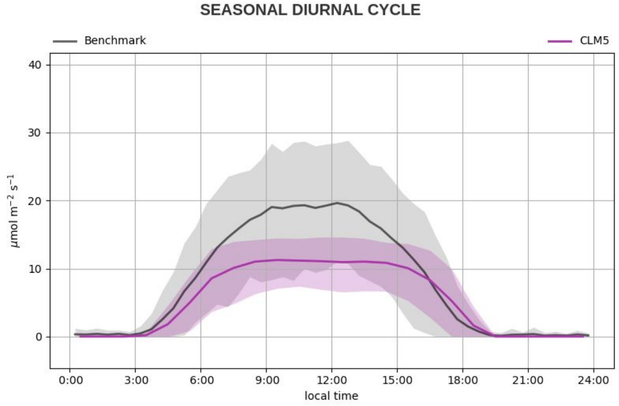
Mean State | All Models | Data Information

	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

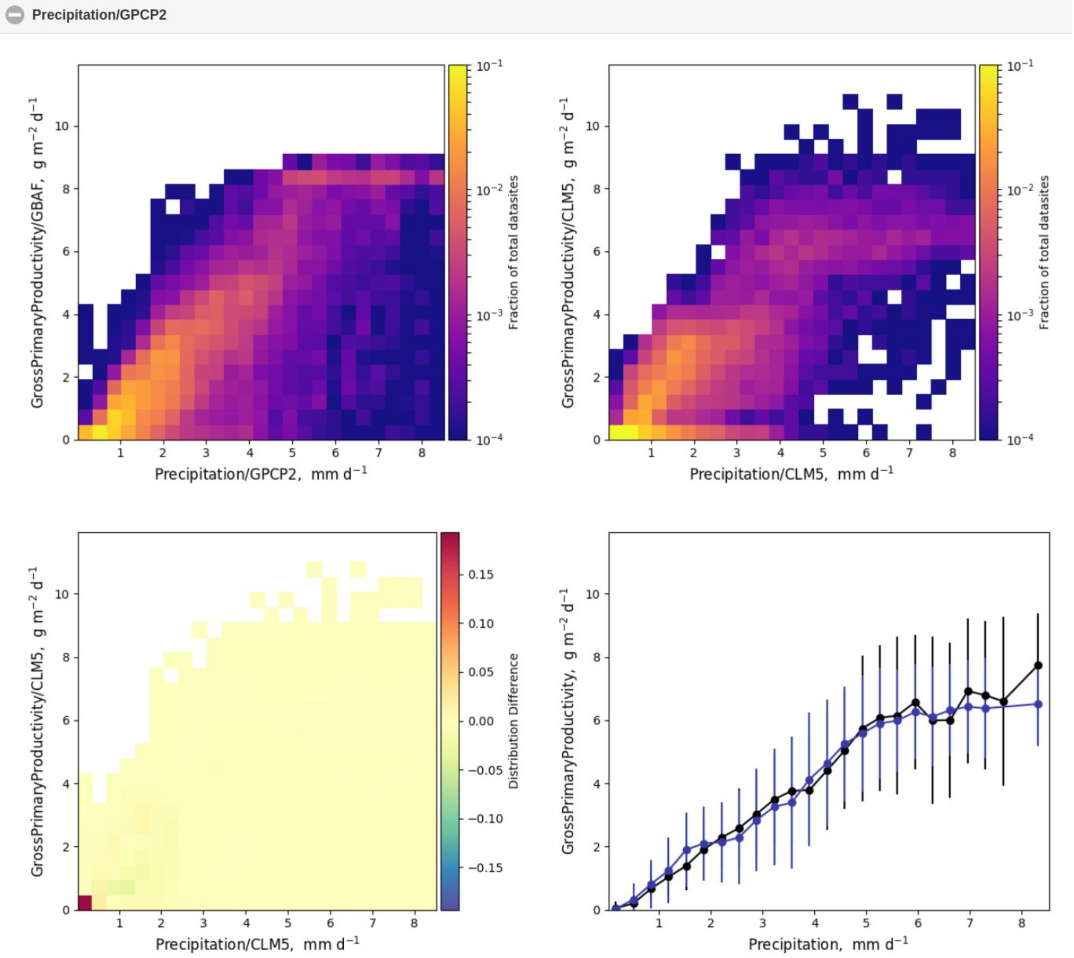
1999

2000

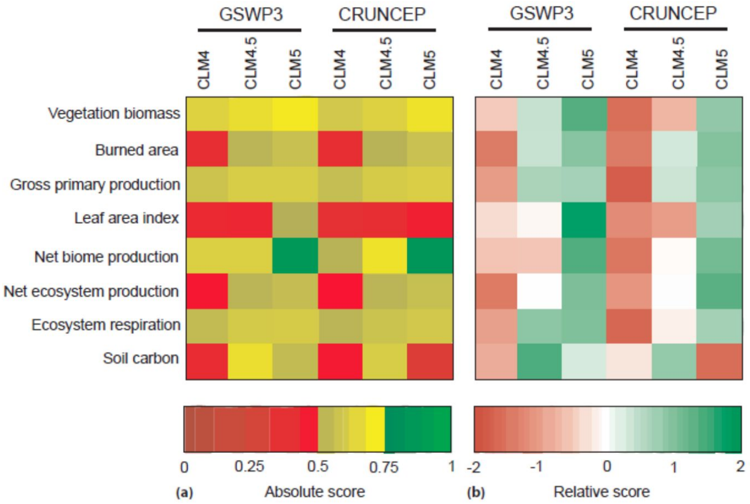


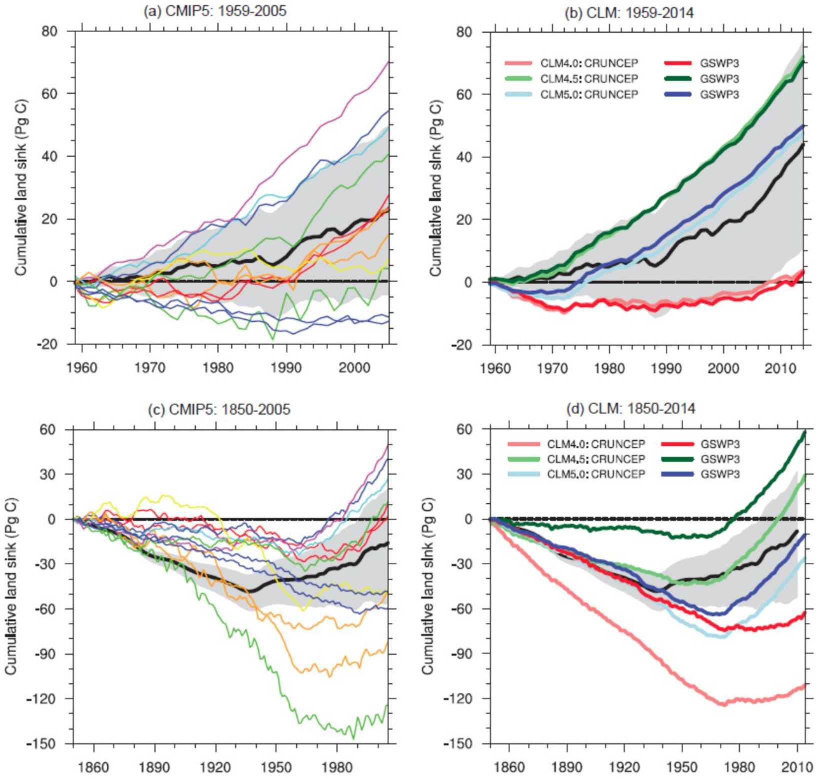


Variable-to-Variable Comparisons



Land Model Performance Depends Strongly on Forcing



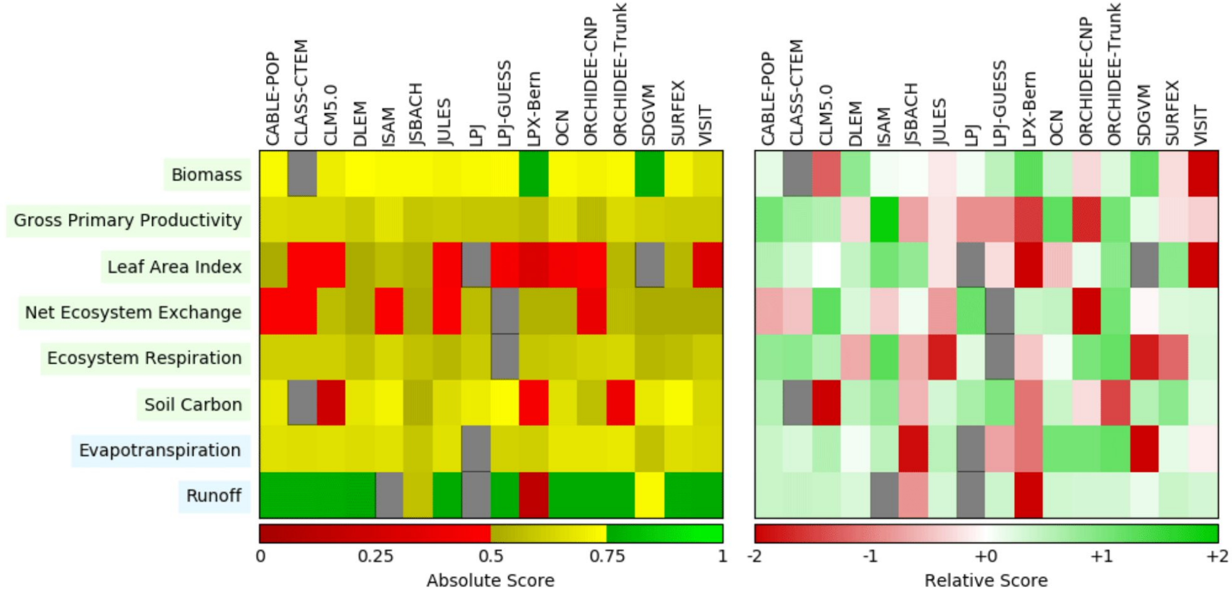


- Depending on the forcing used and the metric selected, different models may perform equally well
- ILAMB scores for CLM4, CLM4.5, and CLM5 forced with GSWP3 vs. CRUNCEP (left) and the cumulative land carbon sink for CMIP5 vs. CLM offline models (right).

Bonan et al. (2019)

Global Carbon Budget 2018 - TRENDY Models

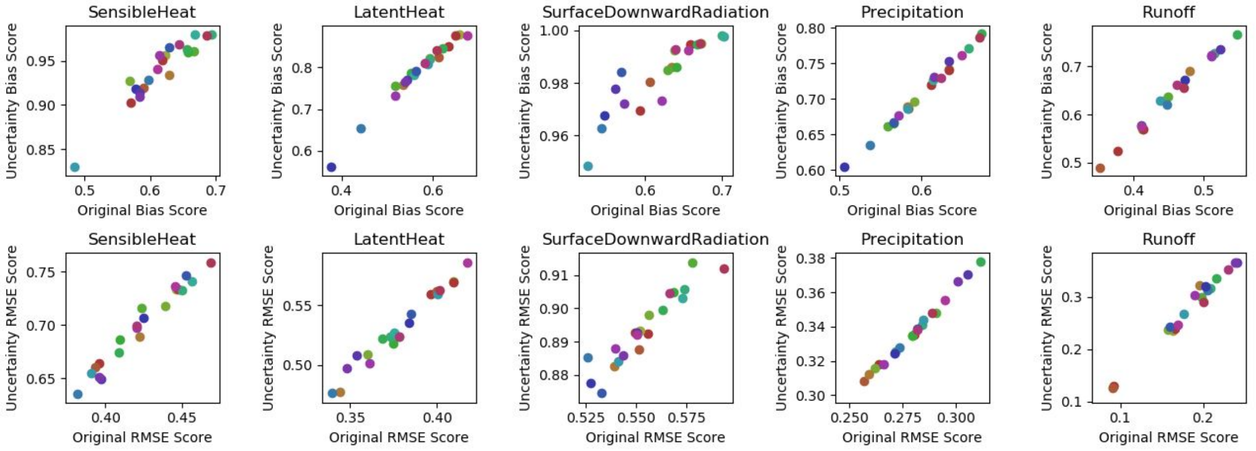
Evaluation of the DGVMs using the International Land Model Benchmarking system (ILAMB; Collier et al., 2018) (left) absolute skill scores and (right) skill scores relative to other models for a subset of ILAMB variables.



Le Quéré et al. (2018)

Addressing Observational Uncertainty

- Few observational datasets provide complete uncertainties
- 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 CLASS self-consistent data (Hobeichi et al., 2020) demonstrates that while scores shift, including uncertainty rarely alters the rank ordering of models (figure)

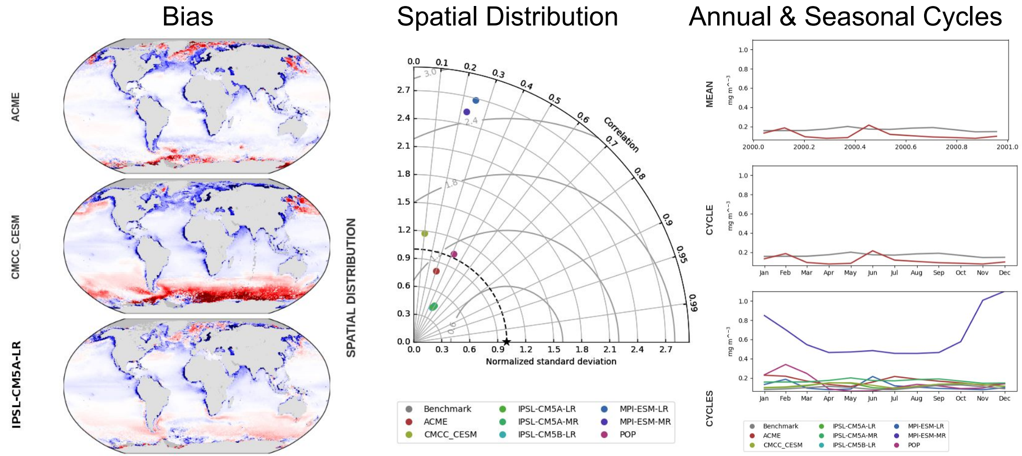
Argonne NATIONAL LABORATORY · BROOKHAVEN NATIONAL LABORATORY · BERKELEY LAB · Los Alamos NATIONAL LABORATORY · NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH · OAK RIDGE National Laboratory · UCI · MICHIGAN

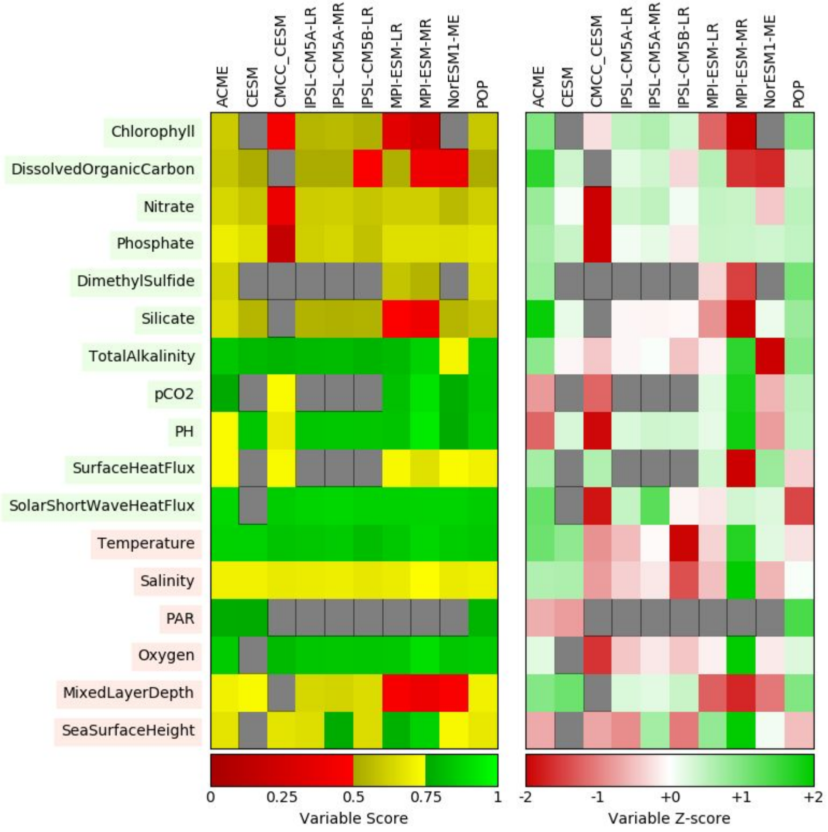
International Ocean Model Benchmarking (IOMB) Package

RUBISCO

- Evaluates ocean biogeochemistry results compared with observations (global, point, ship tracks)
- Scores model performance across a wide range of independent benchmark data
- Leverages ILAMB code base, also runs in parallel
- Built on python and open standards
- Is also open source and will be released soon

Chlorophyll / SeaWIFS



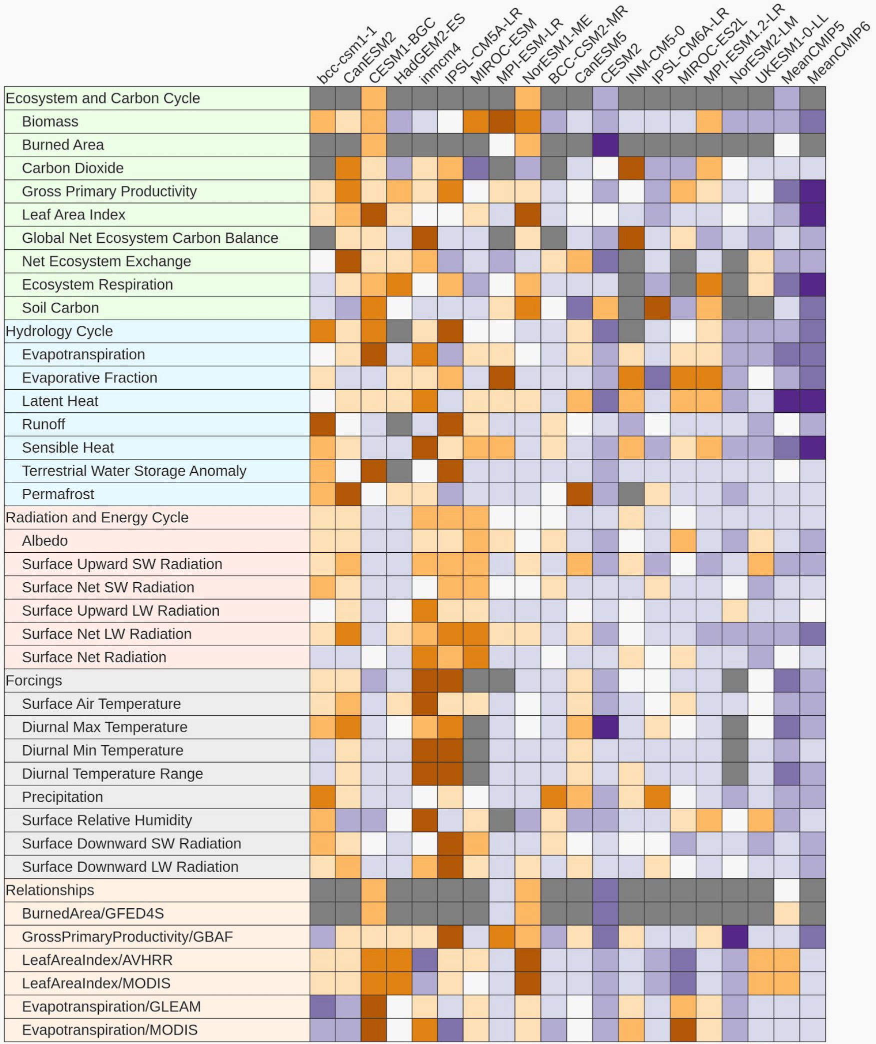


Argonne National Laboratory · Brookhaven National Laboratory · Berkeley Lab · Los Alamos National Laboratory · NCAR National Center for Atmospheric Research · Oak Ridge National Laboratory · UCI · Michigan

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 for CMIP5 outperforms any single CMIP5 model
- The multi-model mean for CMIP6 outperforms any single CMIP6 model
- The multi-model mean CMIP6 land model is the “best model” overall

(Hoffman et al., in prep)

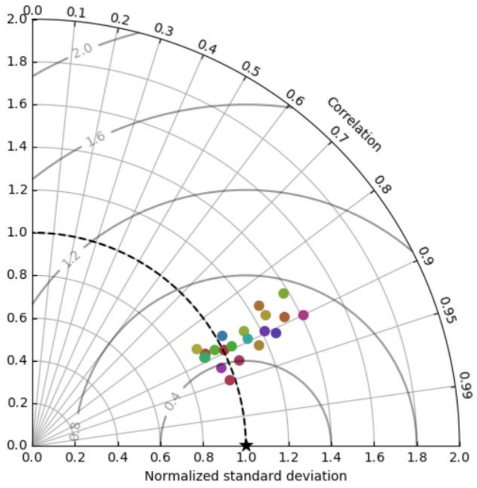


CMIP5 and CMIP6 Land Model Global Gross Primary Productivity

	Download Data	Period Mean (original grids) [Pg yr-1]	Model Period Mean (intersection) [Pg yr-1]	Model Period Mean (complement) [Pg yr-1]	Benchmark Period Mean (intersection) [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	[-]	118.												
bcc-csm1-1	[-]	123.	114.	6.80	118.	0.0600	0.203	1.94	1.27	0.424	0.267	0.809	0.946	0.543
bcc-csm1-1-m	[-]	112.	108.	4.10	118.	0.501	-0.116	1.94	1.38	0.413	0.265	0.794	0.934	0.534
BCC-CSM2-MR	[-]	123.	115.	8.31	118.	0.501	-0.0721	1.68	1.28	0.433	0.326	0.796	0.941	0.564
BCC-ESM1	[-]	157.	133.	21.4	118.	0.0640	0.325	1.84	1.23	0.429	0.302	0.808	0.945	0.557
CanESM5	[-]	141.	131.	8.05	118.		0.675	1.85	1.70	0.427	0.330	0.701	0.934	0.544
CESM1-BGC	[-]	129.	124.	4.32	118.	0.501	0.309	1.74	1.38	0.392	0.350	0.761	0.873	0.545
CESM2	[-]	110.	105.	4.21	118.	0.473	-0.0938	1.72	1.52	0.411	0.364	0.786	0.935	0.572
CESM2-WACCM	[-]	110.	106.	4.28	118.	0.473	-0.0889	1.73	1.50	0.410	0.364	0.788	0.936	0.572
EC-Earth3-Veg	[-]	136.	134.	2.52	118.		0.330	1.99	1.49	0.417	0.312	0.755	0.931	0.545
GFDL-ESM2G	[-]	167.	155.	9.78	118.		1.19	3.18	1.45	0.360	0.185	0.726	0.880	0.467
GISS-E2-1-G	[-]	133.	118.	12.6	117.	1.29	0.0302	1.55	1.23	0.411	0.355	0.741	0.905	0.553
GISS-E2-1-H	[-]	131.	116.	13.8	118.	0.654	-0.0269	1.57	1.19	0.400	0.353	0.760	0.913	0.556
inmcm4	[-]	136.	128.	8.25	113.	5.44	0.351	1.78	1.41	0.451	0.308	0.766	0.935	0.554
IPSL-CM5A-LR	[-]	165.	153.	9.00	118.	0.347	1.10	2.73	1.30	0.318	0.241	0.770	0.889	0.492
IPSL-CM6A-LR	[-]	116.	111.	4.25	118.	0.486	0.0566	1.45	1.32	0.498	0.364	0.751	0.960	0.587
MeanCMIP5	[-]	138.	131.	6.75	118.		0.561	1.44	1.13	0.462	0.408	0.794	0.959	0.606
MeanCMIP6	[-]	121.	116.	5.10	118.		0.159	1.10	1.12	0.522	0.470	0.796	0.973	0.646
MIROC-ESM	[-]	129.	121.	6.01	108.	10.1	0.308	2.06	1.40	0.425	0.322	0.749	0.918	0.547
MPI-ESM-LR	[-]	170.	162.	6.90	110.	8.62	1.22	2.37	1.43	0.378	0.291	0.699	0.926	0.517
NorESM1-ME	[-]	129.	121.	6.29	118.		0.331	1.92	1.46	0.354	0.350	0.759	0.838	0.530
SAM0-UNICON	[-]	131.	126.	4.95	118.	0.501	0.371	1.75	1.39	0.398	0.338	0.764	0.845	0.537

- Most models of the same lineage improved in various characteristics between CMIP6 and CMIP6
- The mean CMIP6 and CMIP5 models perform best

Spatial Taylor Diagram



(Hoffman et al., in prep.)

For more information...

- **International Land Model Benchmarking (ILAMB) Package**
  https://www.ilamb.org/

- **Reducing Uncertainties in Biogeochemical Interactions through Synthesis and Computation (RUBISCO) Science Focus Area**
  https://www.bgc-feedbacks.org/

- **Forrest M. Hoffman**
  Oak Ridge National Laboratory
  forrest@climatemodeling.org

For more information…

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

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

S. Hobeichi, G. Abramowitz, and J. Evans. Conserving land–atmosphere synthesis suite (class). *J. Clim.*, 33(5):1821–1844, Mar. 2020. doi:10.1175/JCLI-D-19-0036.1.

D. M. Lawrence, R. A. Fisher, C. D. Koven, K. W. Oleson, S. C. Swenson, G. 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. Val Martin, and X. Zeng. The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *J. Adv. Model. Earth Sy.*, 11(12):4245–4287, Dec. 2019. doi:10.1029/2018MS001583.

Y. Q. Luo, J. T. Randerson, G. Abramowitz, C. Bacour, E. Blyth, N. Carvalhais, P. Ciais, D. Dalmonech, J. B. Fisher, R. Fisher, P. Friedlingstein, K. Hibbard, F. Hoffman, D. Huntzinger, C. D. Jones, C. Koven, D. Lawrence, D. J. Li, M. Mahecha, S. L. Niu, R. Norby, S. L. Piao, X. Qi, P. Peylin, I. C. Prentice, W. Riley, M. Reichstein, C. Schwalm, Y. P. Wang, J. Y. Xia, S. Zaehle, and X. H. Zhou. A framework for benchmarking land models. *Biogeosci.*, 9(10):3857–3874, Oct. 2012. doi:10.5194/bg-9-3857-2012.