Diagnosing Climate–Carbon Cycle Feedbacks Constrained by ILAMB

Forrest M. Hoffman[1,2], Nathan Collier[1], Mingquan Mu[3], Min Xu[1], Gretchen Keppel-Aleks[4], David M. Lawrence[5], Charles D. Koven[6], Jiafu Mao[1], Qing Zhu[6], William J. Riley[6], and James T. Randerson[3]

[1]Oak Ridge National Laboratory (ORNL), [2]University of Tennessee Knoxville, [3]University of California Irvine, [4]University of Michigan Ann Arbor, [5]National Center for Atmospheric Research (NCAR), and [6]Lawrence Berkeley National Laboratory (LBNL)

North American Carbon Program (NACP)
7th Open Science Meeting

Friday, March 5, 2021

RUBISCO · Argonne National Laboratory · Brookhaven National Laboratory · Berkeley Lab · Los Alamos National Laboratory · NCAR · Oak Ridge National Laboratory · UCI · Michigan

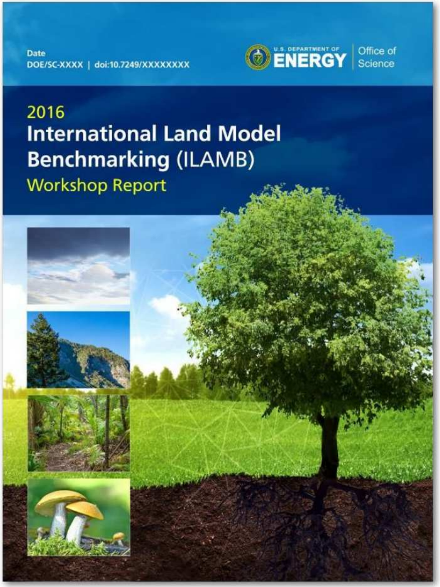
Problem: Model Uncertainty

- Model uncertainty is one of the biggest challenges we face in Earth system science, yet comparatively little effort is devoted to fixing it (Carslaw et al., 2018)
- Ecosystems have complex responses to a wide range of forcing factors in heterogeneous spatial environments, requiring a highly multivariate approach
- The focus is on adding complexity (e.g., more detailed representations of plant traits, photosynthesis, nutrient limitation, respiration), assuming more processes is better
- However, model uncertainty may increase, even as predictions of states and fluxes improve
- Rigorous confrontation of models with observations is required to reduce uncertainty
- Modeling centers have a limited capacity to collect and synthesize the data required to systematically assess all aspects of model fidelity
- Community-developed benchmarking tools are beginning to address some of these problems

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)



RUBISCO · Argonne NATIONAL LABORATORY · BROOKHAVEN NATIONAL LABORATORY · BERKELEY LAB · Los Alamos NATIONAL LABORATORY · NCAR · OAK RIDGE National Laboratory · UCI · MICHIGAN

ILAMB Package Produces Diagnostics and Scores Models

- ILAMB performs model–data comparisons and generates a **portrait plot** of model scores
- For every variable and dataset, ILAMB automatically produces
  - **Tables** containing individual metrics and metric scores (when relevant to the data), including
    - Reference 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})

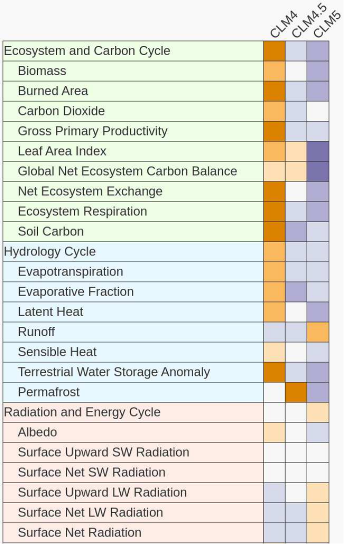
$$\Longrightarrow S_{\text{overall}} = \frac{S_{\text{bias}} + 2S_{\text{rmse}} + S_{\text{phase}} + S_{\text{iav}} + S_{\text{dist}}}{1+2+1+1+1}$$

  - **Graphical diagnostics**
    - Spatial contour maps
    - Time series line plots
    - Spatial Taylor diagrams (Taylor, 2001)
- Similar **tables** and **graphical diagnostics** are produced for functional relationships
- ILAMB design, theory, and implementation are described in Collier et al. (2018)

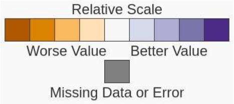
ILAMBv2.5 Package Current Variables

- **Biogeochemistry:** Biomass (Contiguous US, Pan Tropical Forest), Burned area (GFED4.1s), CO_2 (NOAA GMD, Mauna Loa), Gross primary production (Fluxnet, FLUXCOM), Leaf area index (AVHRR, MODIS), Global net ecosystem carbon flux (GCP, Khatiwala/Hoffman), Net ecosystem exchange (Fluxnet, FLUXCOM), Ecosystem respiration (Fluxnet, FLUXCOM), Soil C (HWSD, NCSCDv2, Koven)
- **Hydrology:** Evapotranspiration (GLEAM, MODIS), Evaporative fraction (FLUXCOM), Latent heat (Fluxnet, FLUXCOM, DOLCE), Permafrost (NSIDC), Runoff (Dai, LORA), Sensible heat (Fluxnet, FLUXCOM), Terrestrial water storage anomaly (GRACE)
- **Energy:** Albedo (CERES, GEWEX.SRB), Surface upward and net SW/LW radiation (CERES, GEWEX.SRB, WRMC.BSRN), Surface net radiation (CERES, GEWEX.SRB, WRMC.BSRN)
- **Forcing:** Surface air temperature (CRU, Fluxnet), Dirunal max/min/range temperature (CRU), Precipitation (CMAP, Fluxnet, GPCC, GPCP2), Surface relative humidity (ERA), Surface down SW/LW radiation (Fluxnet, CERES, GEWEX.SRB, WRMC.BSRN)

ILAMB Assessed 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 are not always self-consistent
- Forcing uncertainty confounds assessment of model development



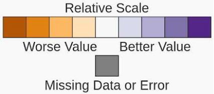
http://webext.cgd.ucar.edu/I20TR/_build_set1F/

(Lawrence et al., 2019)

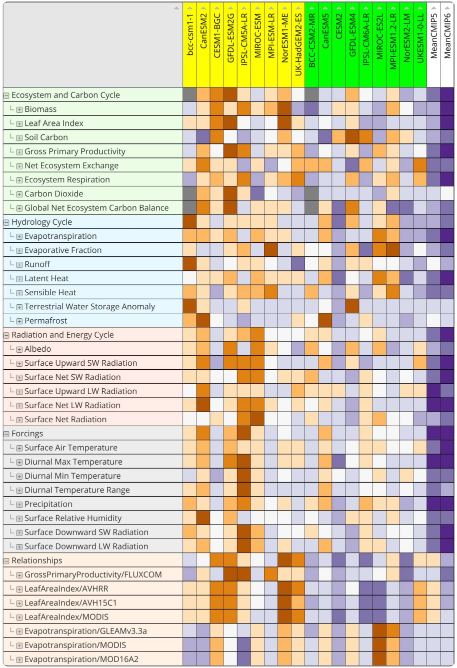
RUBISCO
Argonne NATIONAL LABORATORY
BROOKHAVEN NATIONAL LABORATORY
BERKELEY LAB
Los Alamos NATIONAL LABORATORY
NCAR
OAK RIDGE National Laboratory
UCI
MICHIGAN

CMIP5 vs. CMIP6 Land Models

- The performance of the CMIP6 suite of land models (on right with green headings) has improved over that of the CMIP5 suite of land models (on left with yellow headings)
- The multi-model mean (on far right with white headings) 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 over their CMIP5 progenitors?



(Hoffman et al., in prep.)



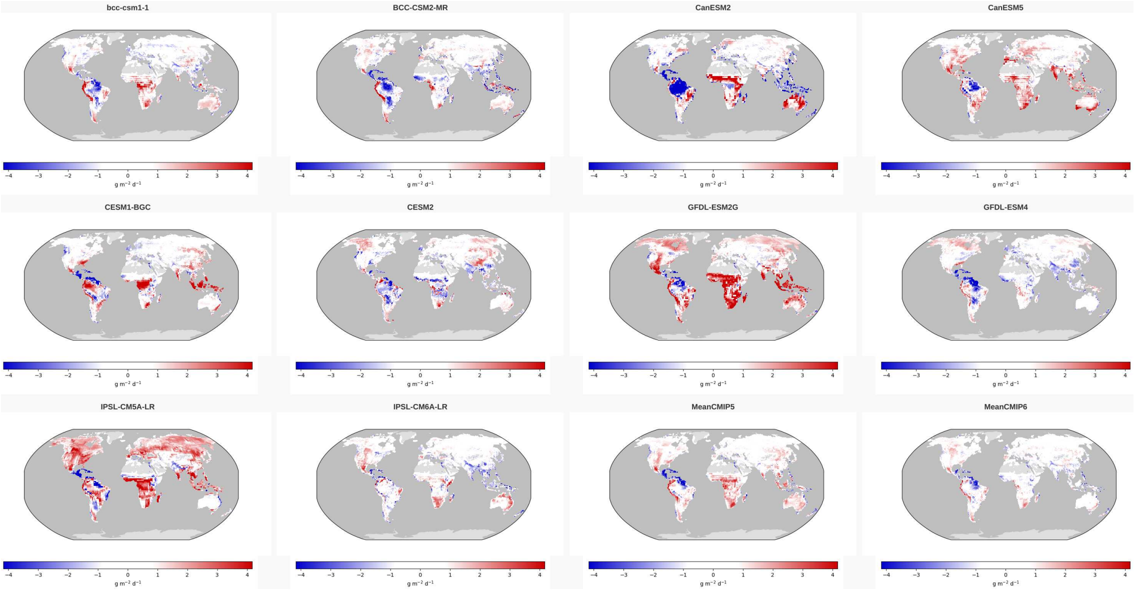
CMIP5 and CMIP6 Land Model Global GPP Compared with FLUXCOM

	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

- Most models of the same lineage improved in various characteristics between CMIP5 and CMIP6
- The MeanCMIP5 and MeanCMIP6 models perform the best

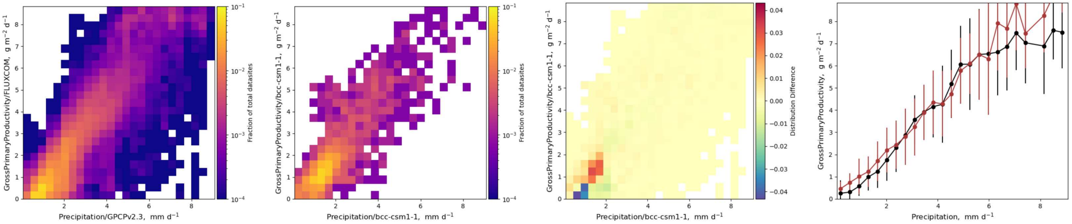
(Hoffman et al., in prep.)

Spatial Distribution of Global GPP Biases



Functional Relationships of GPP with Precipitation and Temperature

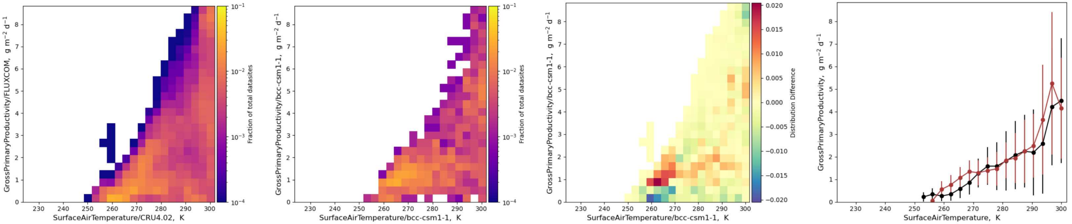
Precipitation/GPCPv2.3



SurfaceDownwardSWRadiation/CERESed4.1

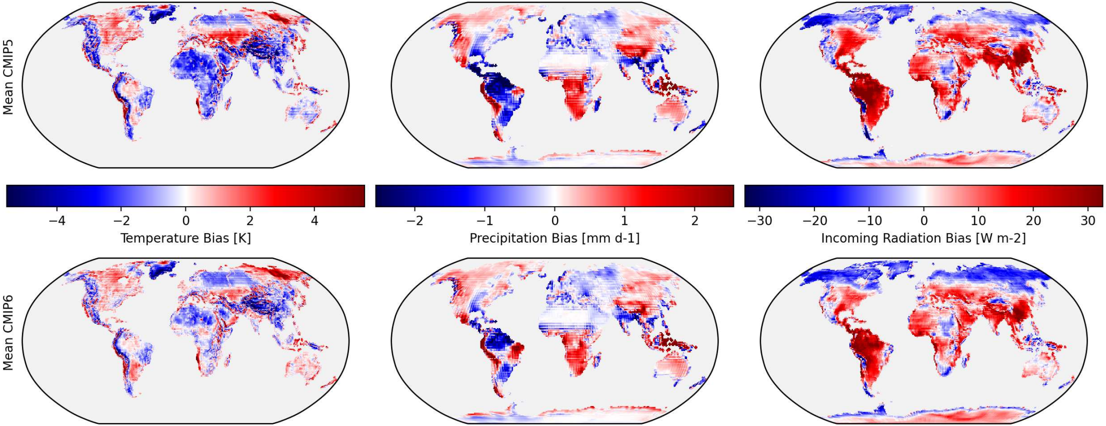
SurfaceNetSWRadiation/CERESed4.1

SurfaceAirTemperature/CRU4.02



Reasons for Land Model Improvements

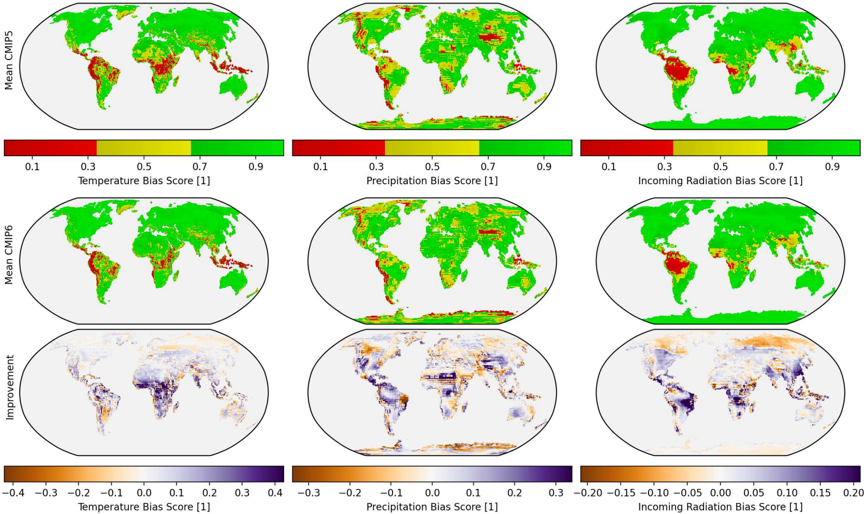
ESM improvements in climate forcings (temperature, precipitation, radiation) likely partially drove improvements exhibited by land carbon cycle models



(Hoffman et al., in prep.)

Reasons for Land Model Improvements

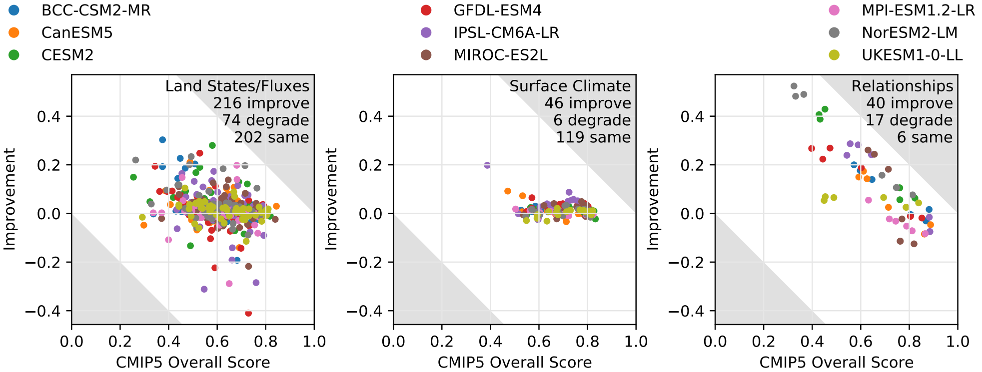
Differences in bias scores for temperature, precipitation, and incoming radiation were primarily positive, further indicating more realistic climate representation by the fully coupled ESMs



(Hoffman et al., in prep.)

RUBISCO Argonne NATIONAL LABORATORY BROOKHAVEN NATIONAL LABORATORY BERKELEY LAB Los Alamos NATIONAL LABORATORY NCAR OAK RIDGE National Laboratory UCI MICHIGAN

Reasons for Land Model Improvements



(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 **functional relationships**, suggesting that increased land model development was also partially responsible for higher CMIP6 land model scores.

Conclusions and Future Research

Summary

- CMIP6 land models performed better than CMIP5 land models due to **(1) improved climate forcing from fully coupled ESMs** and **(2) improved process representation**
- **Functional relationships** exhibited the largest improvements for some models
- Thus, CMIP6 land model results are more valuable for impacts analysis and studies of adaptation and mitigation strategies
- Model improvements in mean states and fluxes may not result in reduced uncertainty or projected model spread

Questions

- Will improved multi-model performance result in reduced spread in feedback sensitivities, projected land carbon storage, and future climate change?
- Can ILAMB scores be used to weight contributions to multi-model means to reduce contemporary biases, reduce projected uncertainties, or alter expected mitigation targets?

Acknowledgments

This research was supported by the *Reducing Uncertainties in Biogeochemical Interactions through Synthesis and Computation Science Focus Area (RUBISCO SFA)*, which is sponsored by the Regional and Global Model Analysis (RGMA) activity of the Earth & Environmental Systems Modeling (EESM) Program in the Climate and Environmental Sciences Division (CESD) of the Office of Biological and Environmental Research (BER) in the US Department of Energy (DOE) Office of Science. Additional support was provided by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC, for the US Department of Energy under contract DE-AC05-00OR22725.

We acknowledge the World Climate Research Programme (WCRP), which, through its Working Group on Coupled Modelling (WGCM), coordinated and promoted the Coupled Model Intercomparison Project phase 6 (CMIP6). We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies that support CMIP6 and ESGF. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the Office of Science of the US Department of Energy under Contract No. DE-AC02-05CH11231. We thank DOE's RGMA activity, Data Management Program, and NERSC for making this coordinated CMIP6 analysis activity possible.

References

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.

K. S. Carslaw, L. A. Lee, L. A. Regayre, and J. S. Johnson. Climate models are uncertain, but we can do something about it. *Eos Trans. AGU*, 99, Feb. 2018. doi:10.1029/2018EO093757.

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.

V. Eyring, P. M. Cox, G. M. Flato, P. J. Gleckler, et al. Taking climate model evaluation to the next level. *Nat. Clim. Change*, 9(2):102–110, Feb. 2019. doi:10.1038/s41558-018-0355-y.

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.

F. M. Hoffman, C. D. Koven, G. Keppel-Aleks, D. M. Lawrence, W. J. Riley, J. T. Randerson, et al. International Land Model Benchmarking (ILAMB) 2016 workshop report. Technical Report DOE/SC-0186, U.S. Department of Energy, Office of Science, Germantown, Maryland, USA, Apr. 2017.

R. Knutti and J. Sedláček. Robustness and uncertainties in the new CMIP5 climate model projections. *Nat. Clim. Change*, 3(4):369–373, Apr. 2013. doi:10.1038/nclimate1716.

D. M. Lawrence, R. A. Fisher, C. D. Koven, K. W. Oleson, S. C. Swenson, et al. 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, et al. A framework for benchmarking land models. *Biogeosci.*, 9(10):3857–3874, Oct. 2012. doi:10.5194/bg-9-3857-2012.

J. T. Randerson, F. M. Hoffman, P. E. Thornton, N. M. Mahowald, K. Lindsay, Y.-H. Lee, C. D. Nevison, S. C. Doney, G. Bonan, R. Stöckli, C. Covey, S. W. Running, and I. Y. Fung. Systematic assessment of terrestrial biogeochemistry in coupled climate-carbon models. *Glob. Change Biol.*, 15(9):2462–2484, Sept. 2009. doi:10.1111/j.1365-2486.2009.01912.x.

K. E. Taylor. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res. Atmos.*, 106(D7):7183–7192, Apr. 2001. doi:10.1029/2000JD900719.

W. R. Wieder, D. M. Lawrence, R. A. Fisher, et al. Beyond static benchmarking: Using experimental manipulations to evaluate land model assumptions. *Global Biogeochem. Cycles*, 33:1289–1309, Oct. 2019. doi:10.1029/2018GB006141.