Benchmarking CMIP Terrestrial Carbon Cycle and Biogeochemistry Models with the ILAMB Package

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

[1]Computational Earth Sciences Group, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA; [2]Department of Civil & Environmental Engineering, University of Tennessee, Knoxville, Tennessee 37996, USA; [3]Department of Earth System Sciences, University of California, Irvine, California 92697, USA; [4]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, Michigan 98195, USA; [5]Climate & Global Dynamics Division, National Center for Atmospheric Research, Boulder, Colorado 80307, USA; and [6]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

RUBISCO

U.S. DEPARTMENT OF ENERGY | Office of Science

Introduction to the ILAMB Package

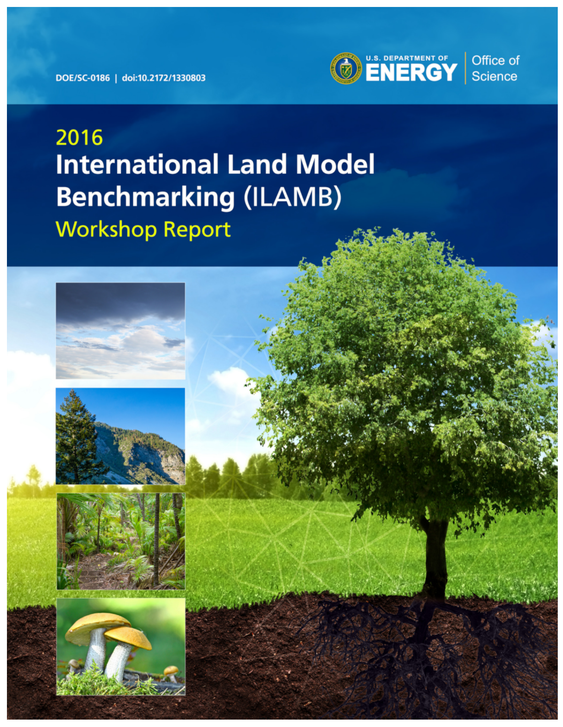
The International Land Model Benchmarking (ILAMB) Project is a community coordination activity to

- Develop internationally accepted benchmarks for scoring land model performance
- Promote the use of these benchmarks in the scientific community
- Strengthen linkages between experimental, remote sensing, and climate modeling communities in developing and applying observational datasets
- Support the design and development of Open Source benchmarking tools

The ILAMB Package (doi:10.18139/ILAMB.v002.00/1251621)

- Is an Open Source toolkit for evaluating land biogeochemistry models through comparisons with observations
- Assesses model fidelity for 29 variables with over 60 observational datasets for biogeochemistry, hydrology, radiation, and climate forcing
- Scores models based on statistical comparisons (period mean, bias, RMSE, phase, amplitude, spatial distribution, Taylor scores) and functional response metrics

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



Model Evaluation and Scoring Methodology

- One or more observational datasets are used to assess model performance for each variable of interest
- For every dataset, ILAMB generates graphical diagnostics (spatial contour maps, time series line plots, and Taylor diagrams)
- Scores are computed for the normalized bias (S_{bias}), normalized central RMSE (S_{rmse}), timing of the maximum of the annual cycle (S_{phase}), interannual variability (S_{iav}), and spatial distribution of the period mean (S_{dist})
- Overall scores for each dataset are calculated from individual scores as follows

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

- Scores for each dataset are averaged to produce an absolute score for each variable.
- Absolute scores are reported in ILAMB and are used to compute relative (Z-values) for each variable across all models included in the analysis.

Relative Performance of CMIP6 Land Models

ILAMB Analysis and Diagnostics available at https://www.ilamb.org/CMIP6/historical/

Relative scores by variable for each model

For CMIP6 land models, the multi-model mean (last column) outperforms any single model for most variables.

Figure 1: *This portrait plot provides an overview of relative scores for available CMIP6 models. Model names are shown across the top and variables are listed down the left side. The last column represents the multi-model mean. For each variable, models that are brown score worse, and models that are purple score better. The multi-model mean outperforms any single model for most variables.* ***WARNING: This is a preliminary analysis; missing data and errors in files and processing are known to exist.***

Relative scores by functional relationship for each model

For CMIP6 land models, the multi-model mean is less likely to outperform any single model for most functional relationships.

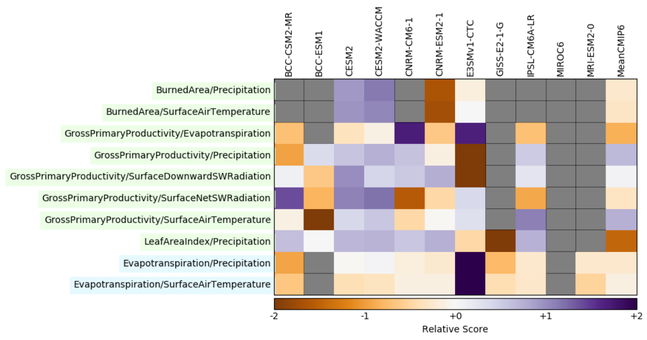


Figure 2: *This portrait plot shows relative scores for functional relationships for available CMIP6 models. Model names are shown across the top and relationships are listed down the left side. The multi-model mean is less likely to outperform any single model for most functional relationships.* ***WARNING: This is a preliminary analysis; missing data and errors in files and processing are known to exist.***

Tracking Relative Performance of CMIP Land Models

ILAMB Analysis and Diagnostics available at https://www.ilamb.org/CMIP5v6/historical/

Relative scores by variable for each CMIP5 and CMIP6 model

- For both CMIP5 and CMIP6 land models, their respective multi-model means (last column in each section) outperforms any single model in their respective collection
- For most variables and models, the CMIP6 land models perform better overall than the CMIP5 land models
- The multi-model mean CMIP6 land model is the "best" model, but for a few variables, individual models may outperform it

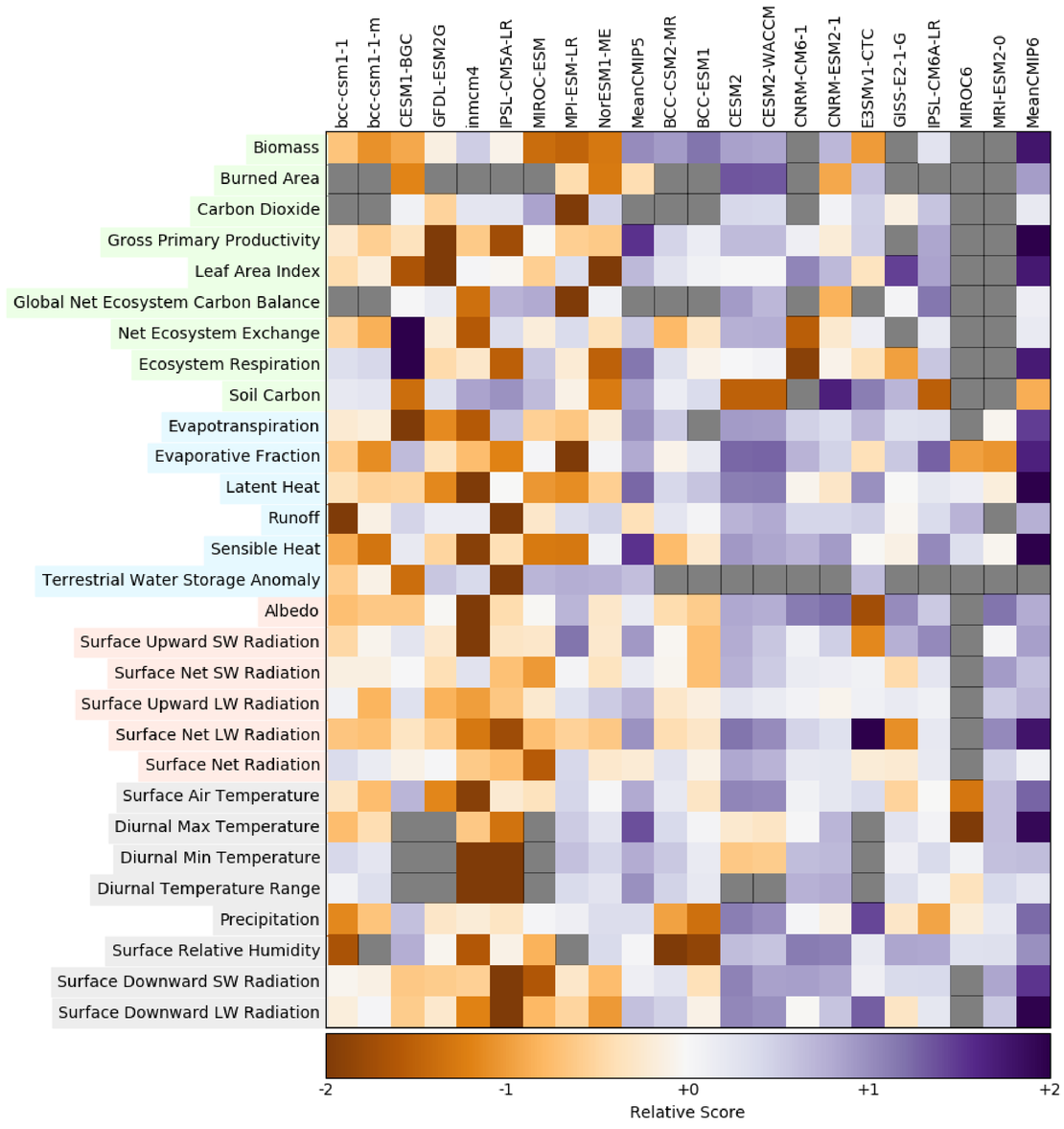


Figure 3: *This portrait plot provides an overview of relative scores for CMIP5 and available CMIP6 models.* ***WARNING: This is a preliminary analysis; missing data and errors in files and processing are known to exist.***

Relative scores by functional relationship for each CMIP5 and CMIP6 model

- For most functional relationships and models, the CMIP6 land models perform better overall than the CMIP5 land models
- However, it is not clear that the multi-model mean CMIP6 land model outperforms all other individual models on functional relationship comparisons

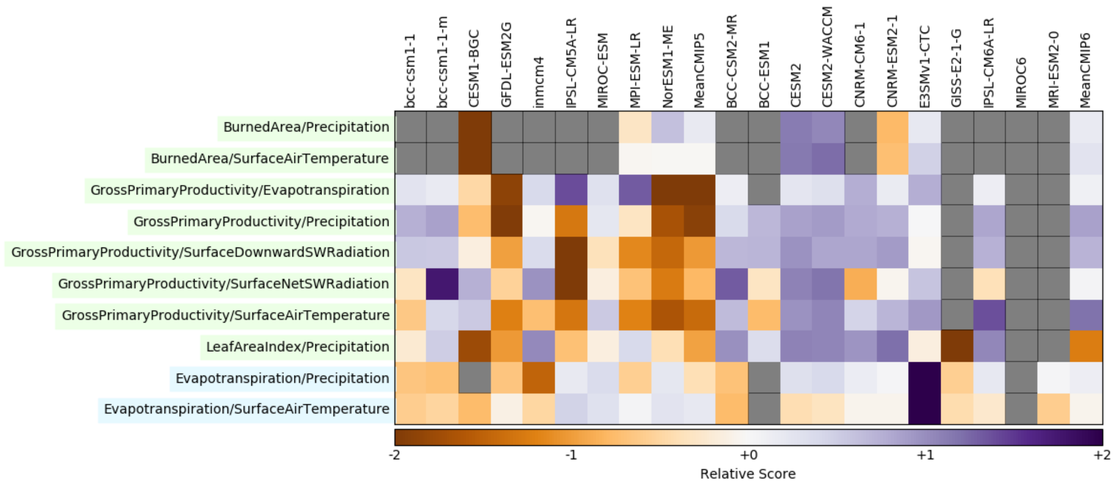


Figure 4: *This portrait plot shows relative scores for functional relationships for CMIP5 and available CMIP6 models.* ***WARNING: This is a preliminary analysis; missing data and errors in files and processing are known to exist.***

Discussion and Conclusions

- Preliminary relative scores from the ILAMB Package suggest that the CMIP6 suite of land models has improved over the CMIP5 suite of land models
- Challenges remain because not all models generate prognostic burned area or have vertically resolved soil carbon
- Few of the observational datasets included in the analysis include uncertainty characterization

Acknowledgments

This research was supported by the Reducing Uncertainties in Biogeochemical Interactions through Synthesis and Computation (RUBISCO) Scientific Focus Area (SFA), which is sponsored by the Regional and Global Model Analysis (RGMA) Program in the Climate and Environmental Sciences Division (CESD) of the Biological and Environmental Research (BER) Program in the U.S. Department of Energy Office of Science. This research used resources of the Oak Ridge Leadership Computing Facility (OLCF) at Oak Ridge National Laboratory (ORNL), which is managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. The Lawrence Berkeley National Laboratory is managed by the University of California for the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The National Center for Atmospheric Research (NCAR) is sponsored primarily by the National Science Foundation. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Figure 3 of this poster) for producing and making available their model output. For CMIP the U. S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.