Evaluating Land Carbon Cycle Processes in Earth System Models: Have Models Improved Over Time?

RUBISCO

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

U.S. DEPARTMENT OF ENERGY Office of Science

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

Introduction to the ILAMB Package

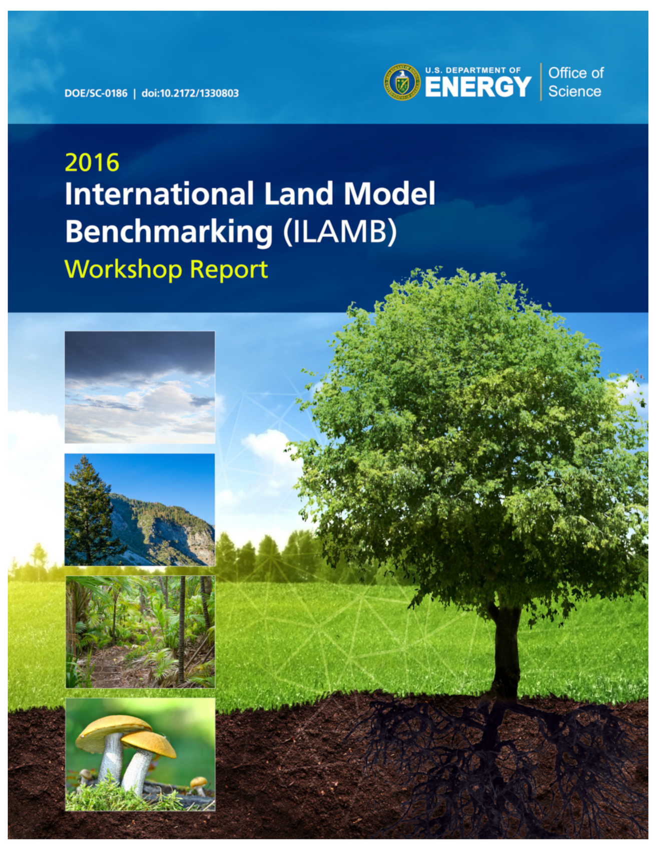
The International Land Model Benchmarking (ILAMB) Project is a community coordination activity to

- 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 ($S_{\rm bias}$), normalized central RMSE ($S_{\rm rmse}$), timing of the maximum of the annual cycle ($S_{\rm phase}$), interannual variability ($S_{\rm iav}$), and spatial distribution of the period mean ($S_{\rm dist}$)
- Overall scores for each dataset are calculated from individual scores as follows

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

- 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

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

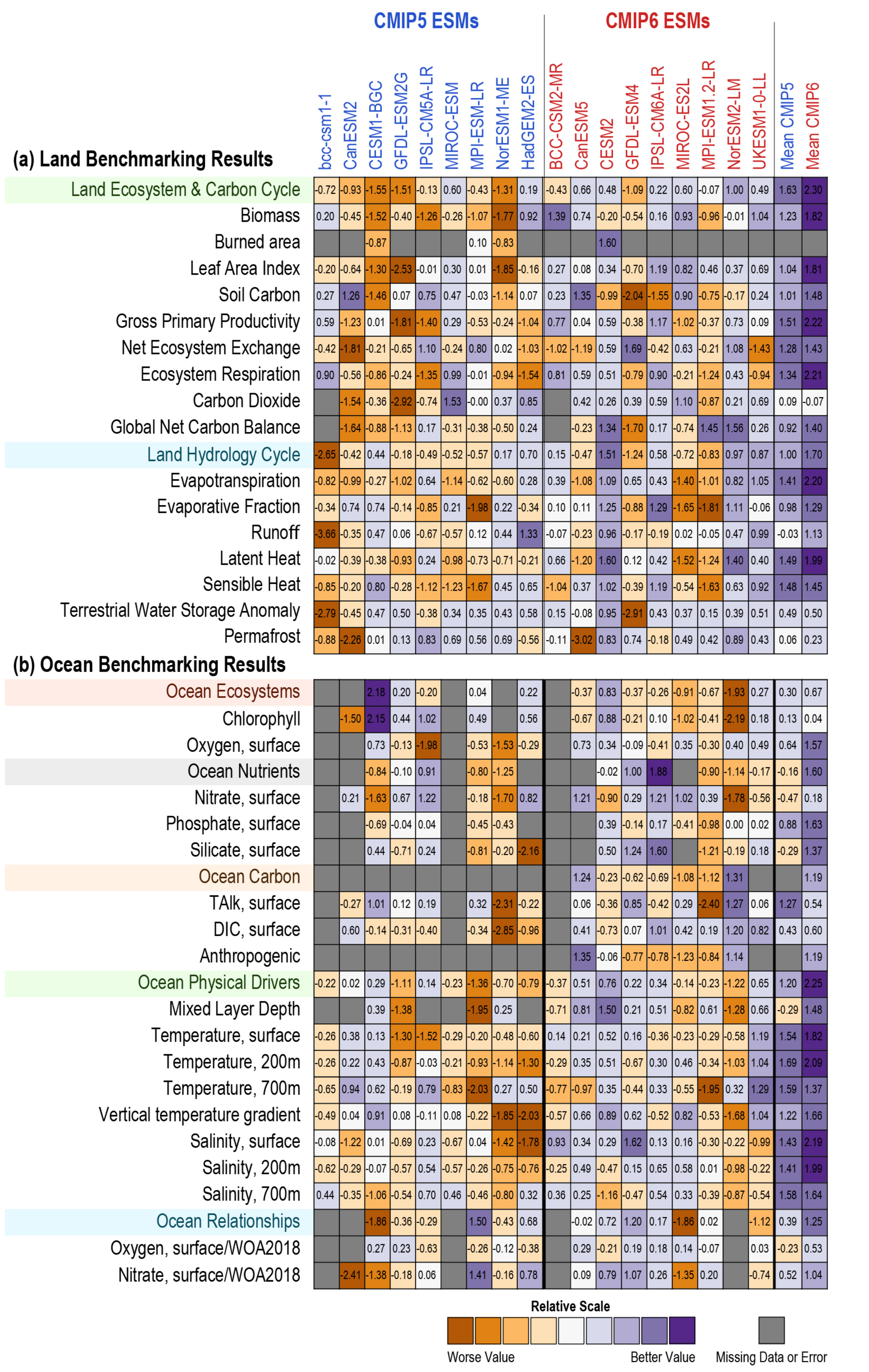


Figure 1: *This portrait plot provides an overview of relative scores for CMIP5 (left-hand side of table) and CMIP6 (right-hand side of table) Earth system models (ESMs), for multiple benchmarks against different datasets. (a) Benchmarking of ESM land models from ILAMB; (b) benchmarking of ocean models from the International Ocean Model Benchmarking (IOMB) package. Scores are relative to other models within each benchmark row, with positive scores indicating a better agreement with observations. Models included are only those from institutions that participated in both CMIP5 and CMIP6 carbon cycle experiments, in order to trace changes from one ensemble to the next. CMIP5 models are labels in blue and CMIP6 in red. This is Figure 5.22 in the IPCC Sixth Assessment Report (AR6), Climate Change 2021: The Physical Science Basis.*

Reasons for Land Model Improvements

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

Climate Forcing Improvements

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

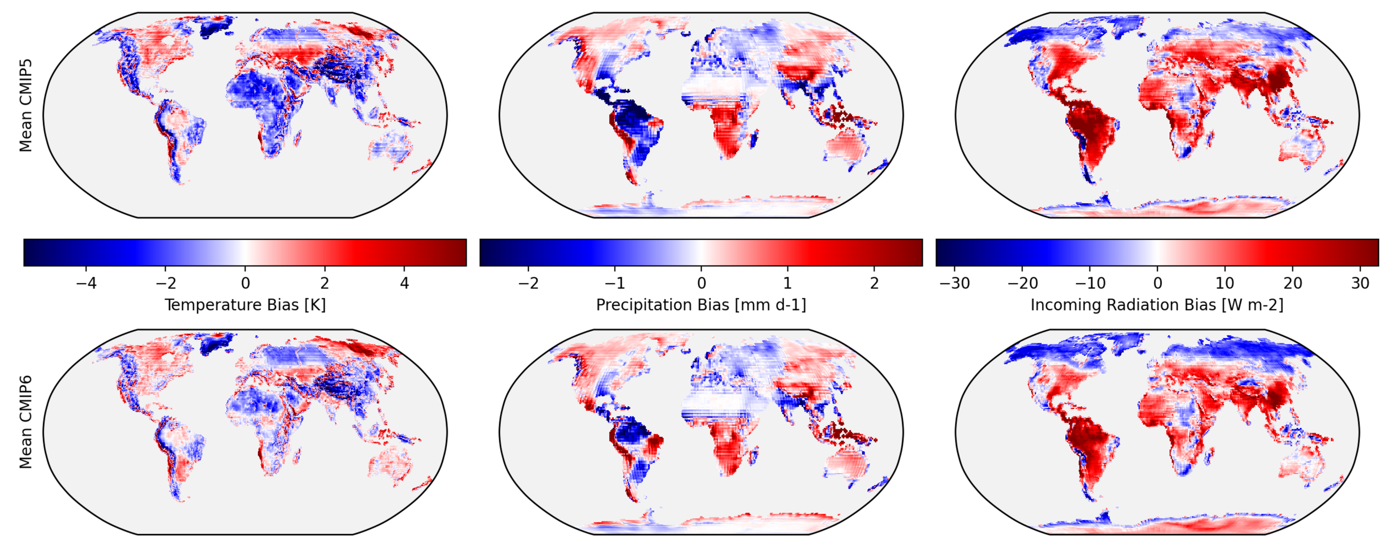


Figure 2: *A comparison of average biases in temperature (first column), precipitation (second column), and incoming radiation (third column) between the mean of the CMIP5 models (top row) and the mean of the CMIP6 models (bottom row) shows a slight reduction in the strengths of those biases in the CMIP6 generation of models.*

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

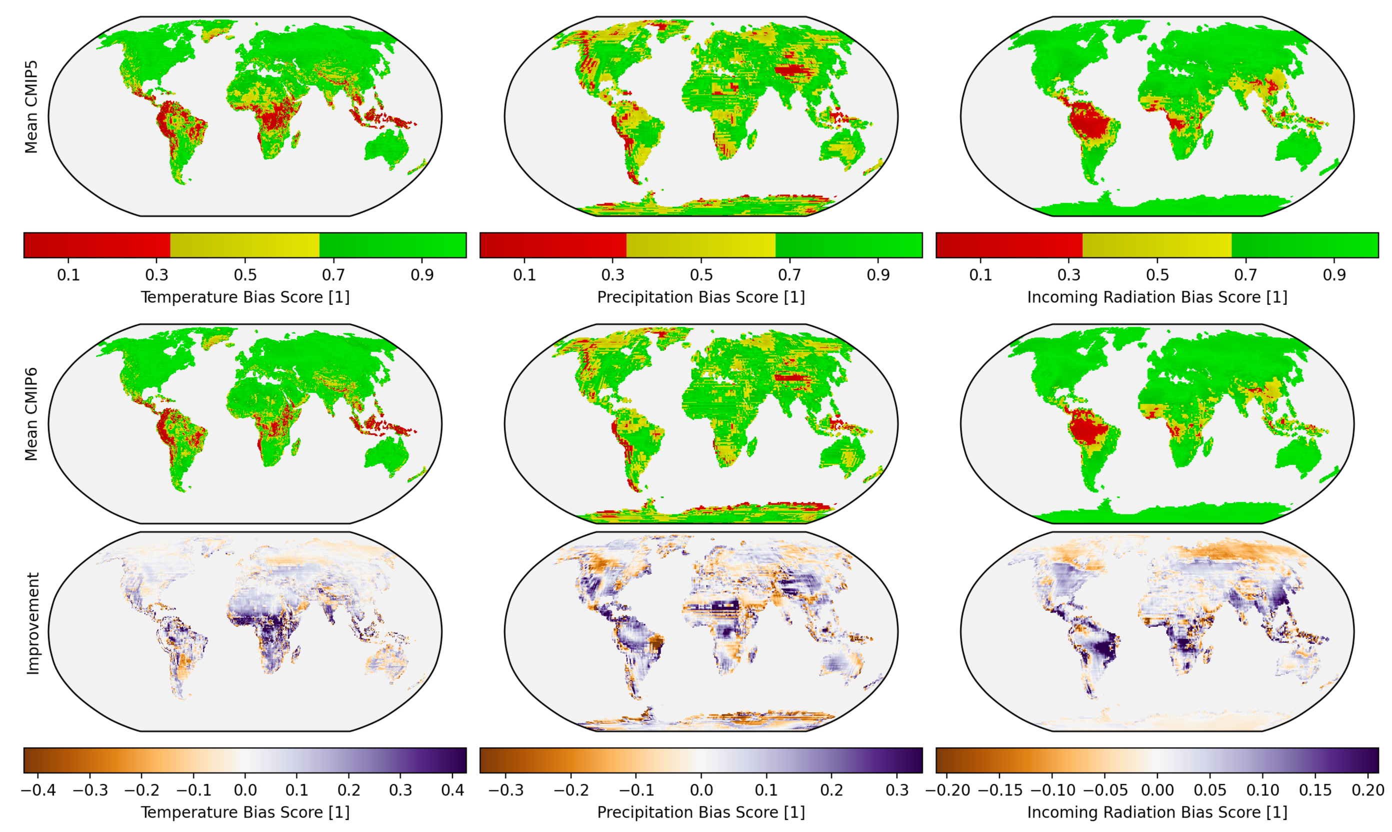


Figure 3: *A comparison of the ILAMB scores of average biases in temperature (first column), precipitation (second column), and incoming radiation (third column) between the mean of the CMIP5 models (top row) and the mean of the CMIP6 models (middle row) shows an increase in bias scores for the CMIP6 generation of models. The bottom row of maps shows the difference between the CMIP6 and CMIP5 bias scores, and the predominance of positive (blue) color indicates an overall improvement in CMIP6 models.*

Land Model Improvements

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

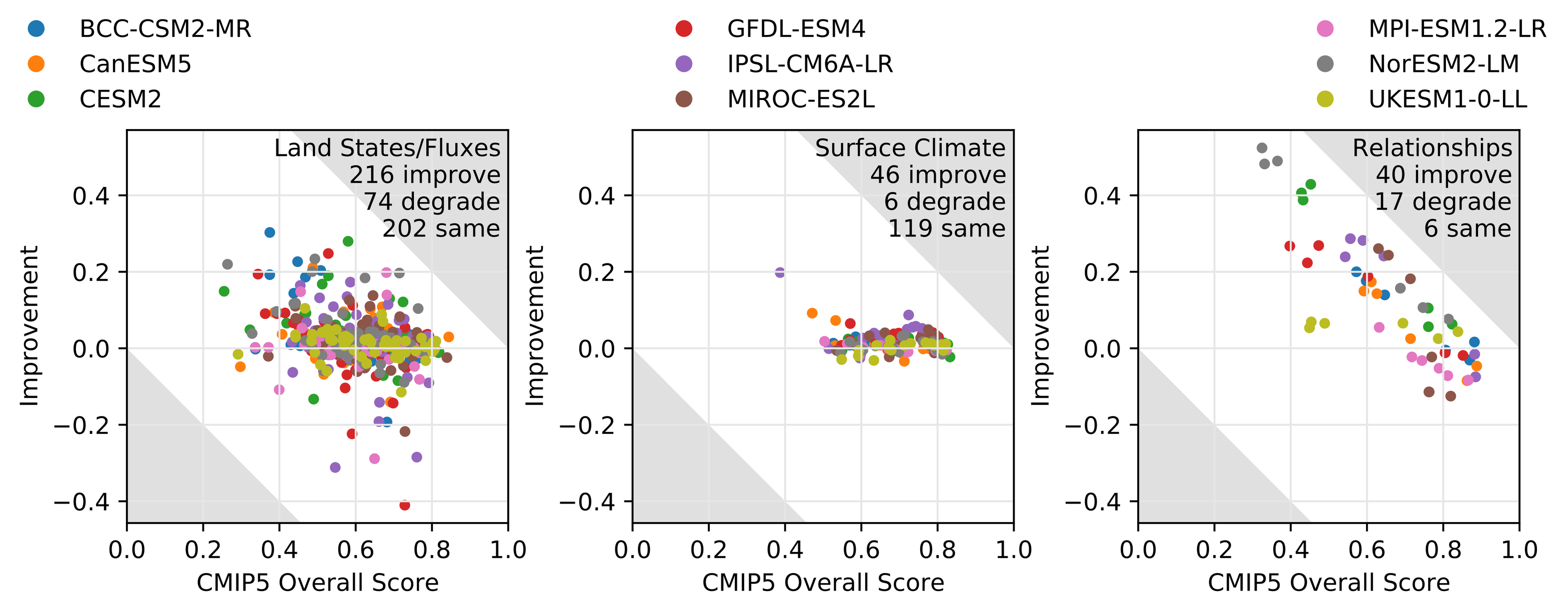


Figure 4: *Improvement in overall ILAMB scores for CMIP6 models (y-axis) versus ILAMB scores for CMIP5 models (x-axis) for nine land models by state and flux variables (first column), surface climate variables (second column), and land variable-to-variable relationships (third column) shows that model results improved in comparison with observations for some land state/flux variables and were degraded for other, but fewer, such variables. The largest improvements were for some of the variable-to-variable relationships.*

Discussion and Conclusions

- **Model benchmarking** is increasingly important as model complexity increases.
- The **ILAMB and IOMB Packages** employ a suite of in situ, remote sensing, and reanalysis datasets to comprehensively evaluation and score land and ocean model performance, irrespective of any model structure or set of process representations.
- Relative scores from the ILAMB Package suggest that **the CMIP6 suite of land models has improved over the CMIP5 suite of land models**.
- Improved land model performance was due to (1) improvements in climate forcings and (2) improvements in land models, **especially in variable-to-variable relationships**.
- **Benchmarking challenges remain** because (1) not all land models represent the same set of processes (e.g., prognostic burned area and vertically resolved soil carbon) and (2) few observational data sets provide quantitative uncertainties.

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) activity of the Earth & Environmental Systems Modeling (EESM) Program in the Earth and Environmental Systems Sciences Division (EESSD) of the Office of Biological and Environmental Research (BER) in the US 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 US Department of Energy under Contract No. DE-AC05-00OR22725. The Lawrence Berkeley National Laboratory is managed by the University of California for the US 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. For CMIP the US 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.

Oak Ridge National Laboratory (ORNL) is managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract DE-AC005-00OR22725.