International Land Model Benchmarking (ILAMB)

Workshop Report

U.S. DEPARTMENT OF ENERGY | Office of Science

Executive Summary

As earth system models (ESMs) become increasingly complex, there is a growing need for comprehensive and multi-faceted evaluation of model projections. To advance understanding of terrestrial biogeochemical processes and their interactions with hydrology and climate under conditions of increasing atmospheric carbon dioxide, new analysis methods are required that use observations to constrain model predictions, inform model development, and identify needed measurements and field experiments. Better representations of biogeochemistry–climate feedbacks and ecosystem processes in these models are essential for reducing the acknowledged substantial uncertainties in 21st century climate change projections.

Building upon past model evaluation studies, the goals of the International Land Model Benchmarking (ILAMB) project are to:

1. Develop internationally accepted benchmarks for land model performance by drawing upon international expertise and collaboration
2. Promote the use of these benchmarks by the international community for model intercomparison
3. Strengthen linkages among experimental, remote sensing, and climate modeling communities in the design of new model tests and new measurement programs
4. Support the design and development of open source benchmarking tools.

The second ILAMB Workshop in the United States was convened on May 16 to 18, 2016, in Washington, District of Columbia, USA. Sponsored by the U.S. Department of Energy's (DOE's) Office of Biological and Environmental Research, the workshop was convened by the Biogeochemistry–Climate Feedbacks Scientific Focus Area (BGC Feedbacks SFA) project. Overarching goals of the workshop were to engage the international research community in defining scientific priorities for (1) design of new metrics, (2) improvement of model development and workflow practices, (3) Coupled Model Intercomparison Project (CMIP) evaluation, and to learn about new observational data sets and measurement campaigns.

The workshop drew more than 60 on-site participants, and between 20 and 30 individuals—including students and postdocs—attended online at any time during the plenary sessions. Participants were from Australia, Canada, China, Germany, Japan, Netherlands, Sweden, United Kingdom, and the United States and represented 10 different major modeling centers. Plenary presentations focused on model benchmarking, emergent constraints, evaluation metrics, uncertainty quantification, and field experiment and measurement networks.

Outcomes of the 2016 ILAMB Workshop

This 2016 ILAMB Workshop Report provides a synopsis of the current state of the science and highlights challenges and opportunities for benchmarking, model development, and field and laboratory measurements needed to advance climate science. The main text of the report provides a synthesis of the ideas, concepts, and scientific priorities presented and discussed at the workshop. The appendix of the report consists of topical white papers that summarize invited presentations, describe breakout group proceedings, and offer recommendations. In addition, the white papers identify critical gaps and opportunities in measurement programs, offer new approaches for model evaluation, and point out synergies among research teams and tools being constructed to support model development, parameter estimation, and model–data integration.

As depicted in the schematic figure below, the topical white papers within the categories of Major Processes and Integrating and Cross-cutting Themes were synthesized with those on the needs of Model Intercomparison Projects (MIPs) to produce a set of next generation Benchmarking Challenges and Priorities resulting from the workshop. Moreover, Benchmarking Approaches for addressing these challenges were identified and Enabling Capabilities needed to facilitate next generation benchmarking and model development were distilled from the white papers. Addressing these challenges will advance climate science by enabling process understanding, quantifying feedbacks, reducing uncertainties, and improving model projections.

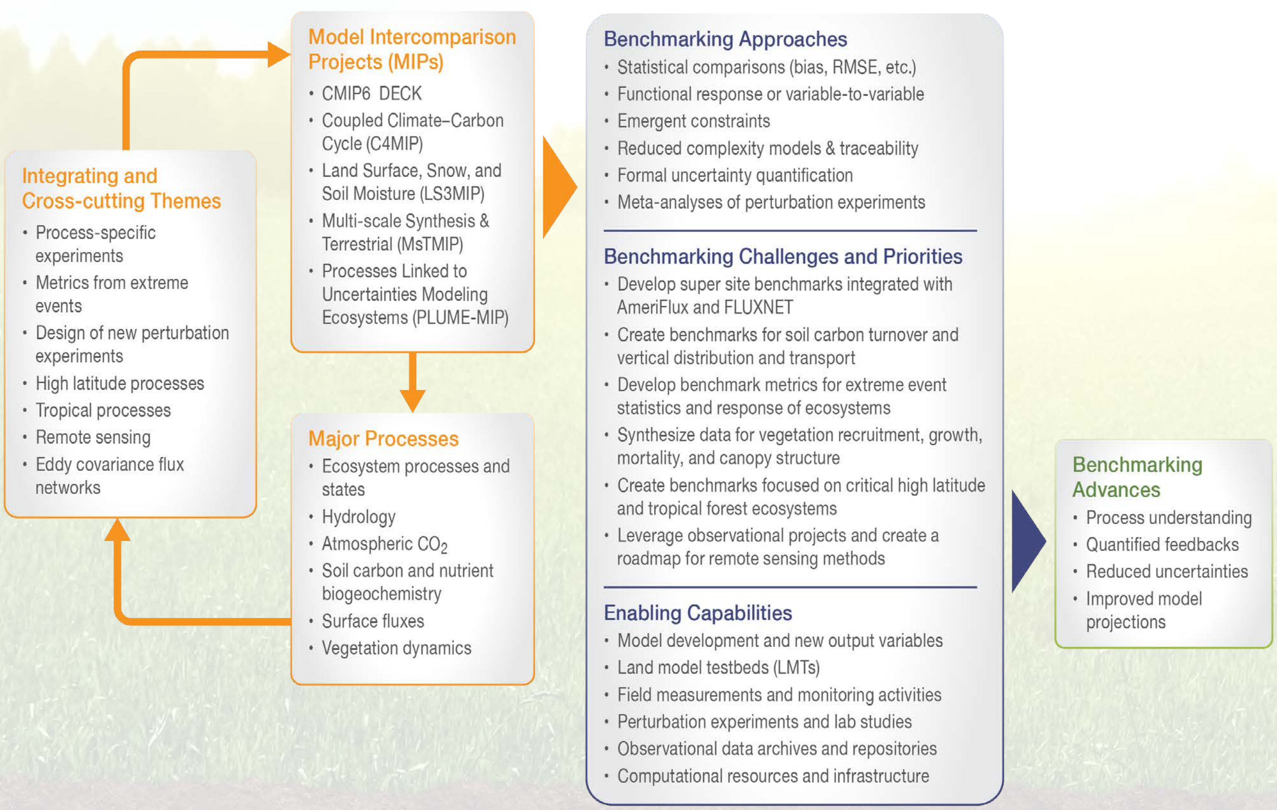
Benchmarking Tools

Model evaluation and benchmarking tools currently employed by international modeling centers were assessed at the workshop. Features of current benchmarking tools—including the Protocol for the Analysis for Land Surface models (PALS), the Program for Climate Model Diagnosis and Intercomparison (PCMDI) Metrics Package (PMP), the Earth System Model Evaluation Tool (ESMValTool), and the Land surface Verification Toolkit (LVT)—were reviewed, and the new ILAMB benchmarking systems were described and demonstrated.

The ILAMB version 1 (v1) and ILAMB version 2 (v2) benchmarking systems compare model results with best-available observational data products, focusing on atmospheric CO_2, surface fluxes, hydrology, soil carbon and nutrient biogeochemistry, ecosystem processes and states, and vegetation dynamics. ILAMBv2 is expected to become an integral part of the workflow for model frameworks, including the Accelerated Climate Modeling for Energy (ACME) model and the Community Earth System Model (CESM). Moreover, ILAMBv2 will contribute model analysis and evaluation capabilities to phase 6 of the Coupled Model Intercomparison Project (CMIP6) and future model and model–data intercomparison projects.

Benchmarking Challenges and Priorities

A variety of statistical approaches have been adopted to evaluate model accuracy through comparison with observations, including calculations of bias, root-mean-square error (RMSE), phase, amplitude, spatial distribution, Taylor diagrams and



scores, functional relationship metrics, and perturbation and sensitivity tests. While many of these statistical measures are not independent, each provides slightly different information about contemporary model performance with respect to observational data and about implications for future projections from ESMs.

However, developing metrics that make appropriate use of observational data remains a scientific challenge because of the spatial and temporal mismatch between models and measurements, poorly characterized uncertainties in observationally constrained data products, biases in reanalysis and forcing data, model simplifications, and structural and parametric uncertainties. A variety of benchmarking challenges and opportunities emerged from workshop breakout group meeting reports. Common themes included the following:

- Need for collocated measurements, particularly around a core set of AmeriFlux and FLUXNET sites with a sustained record of observations for repeated model testing
- Lack of quantified uncertainty information for observational data
- Utility of functional response metrics and variable-to-variable comparisons
- Value of metrics for future projections based on emergent constraints
- Unrealized opportunities for global observational data sets based on satellite remote sensing synthesized with ancillary databases, using new algorithms
- Importance of applying statistical and machine learning methods to upscaling sparse measurements from sites to regions to the globe
- Need for process-level benchmarks and metrics for extreme events
- Opportunities for collaboration with earth system model developers (e.g., ACME, CESM, and others)
- Opportunities for collaboration with important field and laboratory experiments and monitoring activities, including AmeriFlux and FLUXNET, Integrated Carbon Observation System (ICOS), Next Generation Ecosystem Experiments (NGEE) Arctic, Arctic–Boreal Vulnerability Experiment (ABoVE), Spruce and Peatland Responses Under Climatic and Environmental Change (SPRUCE) project, Critical Zone Observatories (CZOs), Long-Term Ecological Research (LTER) sites, National Ecological Observatory Network (NEON), NGEE Tropics, and Tropical Responses to Altered Climate Experiment (TRACE).

Recommendations for next-generation Benchmarking Challenges and Priorities included the following:

- Develop supersite benchmarks integrated with AmeriFlux and FLUXNET
- Create benchmarks for soil carbon turnover and vertical distribution and transport
- Develop benchmark metrics for extreme event statistics and response of ecosystems
- Synthesize data for vegetation recruitment, growth, mortality, and canopy structure
- Create benchmarks focused on critical high latitude and tropical forest ecosystems
- Leverage observational projects and create a roadmap for remote sensing methods.

Model Intercomparison Project (MIPs)

Model Intercomparison Project (MIPs) are important activities for assessing the coherence and reliability of ESMs. Ongoing and future MIPs focused on modeling terrestrial water, energy, and carbon cycles are particularly relevant to ILAMB. Benchmarking needs were evaluated for the CMIP6 historical and Diagnostic, Evaluation, and Characterization of Klima (DECK) experiments; the Coupled Climate–Carbon Cycle MIP (C4MIP); the Land Surface, Snow and Soil Moisture MIP (LS3MIP); and the Land Use MIP (LUMIP). Opportunities for benchmarking model results from other MIPs were also considered.

Key recommendations that emerged on MIP benchmarking needs were the following:

- Develop methods to attribute emergent model behaviors such as carbon feedback parameters to specific processes through emergent constraint and traceability approaches
- Benchmark across coupling and complexity hierarchies—from offline land-only simulations to fully coupled ESMs—to attribute model biases and uncertainties to specific domains and identify feedbacks between domains
- Develop paired site data sets for benchmarking model representations of subgrid scale heterogeneity.

Benchmarking Approaches

New and existing Benchmarking Approaches were identified from the workshop. While traditional statistical comparisons with observations offer a great deal of information about model performance, metrics based on functional responses or variable-to-variable comparisons often suggest why models produce incorrect results. Benchmarking future projections can be accomplished through careful use of emergent constraints. Reduced complexity models and traceability frameworks are usefully applied to enable greater process understanding through more frequent and detailed testing with reduced computational costs. Formal uncertainty quantification (UQ) frameworks and methods, described in papers in the appendix, provide rigorous techniques for understanding model predictions. Finally, meta-analyses of perturbation experiments provide a new approach for constraining model predictions of ecosystem responses under controlled environmental change conditions.

Enabling Capabilities

To address the identified next generation Benchmarking Challenges and Priorities, certain Enabling Capabilities are needed. New model development focused on improving process representations is required, and additional model variables should be saved for comparison with data. A new Land Model Testbed (LMT) capability employing community benchmarks and supporting UQ frameworks would enable more rapid model development and verification, particularly for major ESM frameworks like ACME and CESM.

Additional field measurements and monitoring activities, as well as perturbation experiments and lab studies, could provide valuable observational data for constraining models. High priority measurement needs for developing benchmarks and improving ESMs include the following:

- Long-term energy, carbon, and water flux measurements at AmeriFlux and FLUXNET sites with standardized instrumentation and methods, and additional frequent or continuous ancillary *in situ* measurements of soil moisture, sap flow, tree height and diameter, litterfall, and soil nutrients
- High latitude and tundra soil core measurements of carbon and nutrient distributions, including isotopes and ice/water content, and observations of vegetation growth and expansion of woody vegetation
- Characterization of tropical ecosystem traits and canopy structure and chemistry; observations of tropical ecosystem responses to drought, increased temperatures, and elevated atmospheric CO_2; and measurements of nutrient cycling and hydrology in tropical forests, focusing on land–atmosphere interactions
- Remote sensing algorithms and processing infrastructure for generating data products useful for large-scale ecosystem characterization and monitoring, scaling up *in situ* measurements, and informing future measurement site selection.

Improved observational data archives (e.g., DOE Atmospheric Radiation Measurement (ARM) Climate Research Facility and Environmental System Science (ESS) archives, NASA Distributed Active Archive Centers (DAACs)) and repositories (e.g., Obs4MIPs) are needed that offer data discovery, server-side analysis, and advanced distribution capabilities. Finally, new computational resources and cyber infrastructure will be required to realize the promise of new benchmarking capabilities. This infrastructure needs to offer a balance between pure compute capacity (high core count) and throughput (e.g., cache size, memory size and bandwidth, and input/output bandwidth) to support *in situ* analysis and benchmarking, growing observational data sets, and multi-model comparisons.

Conclusions and Next Steps

The 2016 ILAMB Workshop was successful in bringing together the international community to identify scientific challenges and priorities for future research. The workshop demonstrated broad interest on the part of a vibrant community of scientists spanning many disciplines that are committed to reducing barriers for information flow between the measurement and modeling communities.

To effectively address the individual processes and cross-cutting themes discussed above, small, targeted working groups should be formed to research and publish supporting analyses. A top priority is supporting CMIP6 activities, where additional development of ILAMB functionality could yield powerful automated analyses and model intercomparison capabilities for such national and international assessment efforts.

Over the next decade, the community envisions the ILAMB system to serve as a core capability within a U.S. or international center that will provide a home to focused synthesis working groups, host MIP-related activities, and support expanded use of, and access to, ESMs by a broader cross section of scientists within disciplines of ecosystem ecology, biogeochemistry, and hydrology.

Contacts

Renu Joseph
DOE Program Manager
Regional and Global Climate Modeling
renu.joseph@science.doe.gov

Dorothy Koch
DOE Program Manager
Earth System Modeling
dorothy.koch@science.doe.gov

Forrest M. Hoffman
BGC Feedbacks SFA Laboratory Research Manager
Oak Ridge National Laboratory
forrest@climatemodeling.org

Additional Information

2016 International Land Model Benchmarking (ILAMB) Workshop Report available at

http://science.energy.gov/ber/community-resources/

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. Best, B. Bond-Lamberty, M. De Kauwe, A. S. Denning, A. Desai, V. Eyring, R. 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. Moorcroft, G. S. H. Pau, D. M. Ricciuto, K. Schaefer, C. R. Schwalm, S. Serbin, E. Shevliakova, A. G. Slater, J. Tang, M. Williams, J. Xia, C. Xu, R. Joseph, and D. Koch (2016), International Land Model Benchmarking (ILAMB) 2016 Workshop Report, DOE/SC-0186, U.S. Department of Energy, Office of Science, Germantown, Maryland, USA, 157 pp., doi:10.2172/1330803.