Systematic Evaluation of Land Surface Models Using the International Land Model Benchmarking (ILAMB) Package

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

[1]Oak Ridge National Laboratory, [2]University of California Irvine, [3]Lawrence Berkeley National Laboratory, [4]National Center for Atmospheric Research, and [5]University of Michigan Ann Arbor

97th Annual Meeting of the American Meteorological Society (AMS)
Washington State Convention Center, Seattle, Washington, USA
January 22–26, 2017

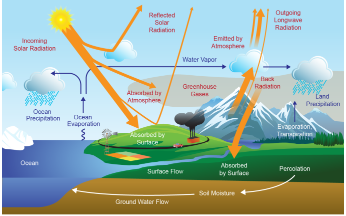
American Meteorological Society 97th ANNUAL MEETING | SEATTLE

BGC Feedbacks

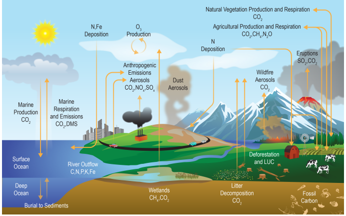
What is ILAMB?

A community coordination activity created to:

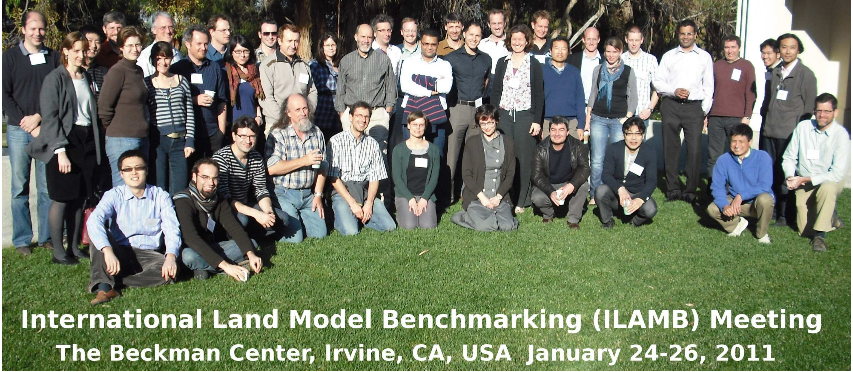
- **Develop internationally accepted benchmarks** for land model performance by drawing upon collaborative expertise
- **Promote the use of these benchmarks** for model intercomparison
- **Strengthen linkages between experimental, remote sensing, and climate modeling communities** in the design of new model tests and new measurement programs
- **Support the design and development of open source benchmarking tools** (Luo et al., 2012)



Energy and Water Cycles

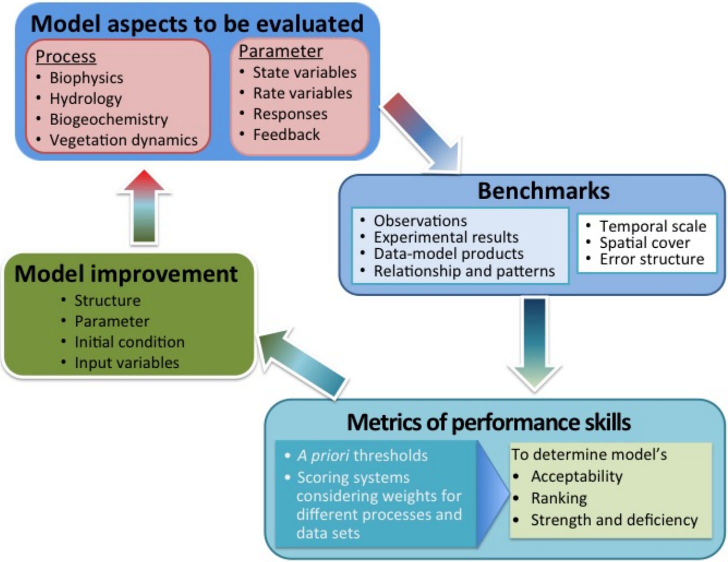


Carbon and Biogeochemical Cycles



- We co-organized inaugural meeting and ∼45 researchers participated from the United States, Canada, the United Kingdom, the Netherlands, France, Germany, Switzerland, China, Japan, and Australia.
- **ILAMB Goals:** Develop internationally accepted benchmarks for model performance, advocate for design of open-source software system, and strengthen linkages between experimental, monitoring, remote sensing, and climate modeling communities. *Initial focus on CMIP5 models.*
- Provides methodology for model–data comparison and baseline standard for performance of land model process representations (Luo et al., 2012).

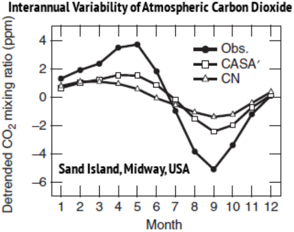
General Benchmarking Procedure



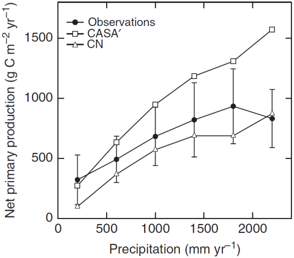
(Luo et al., 2012)

What is a Benchmark?

- A **benchmark** is a quantitative test of model function achieved through comparison of model results with observational data.
- Acceptable performance on benchmarks **is a necessary but not sufficient condition** for a fully functioning model.
- **Functional benchmarks** offer tests of model responses to forcings and yield insights into ecosystem processes.
- Effective benchmarks must draw upon a broad set of independent observations to evaluate model performance on **multiple temporal and spatial scales**.



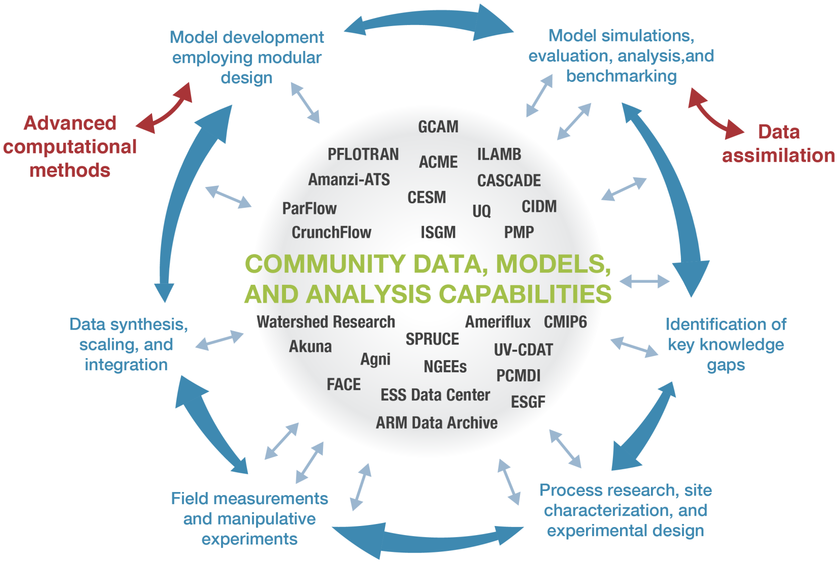
Models often fail to capture the amplitude of the seasonal cycle of atmospheric CO_2.



Models may reproduce correct responses over only a limited range of forcing variables.

(Randerson et al., 2009)

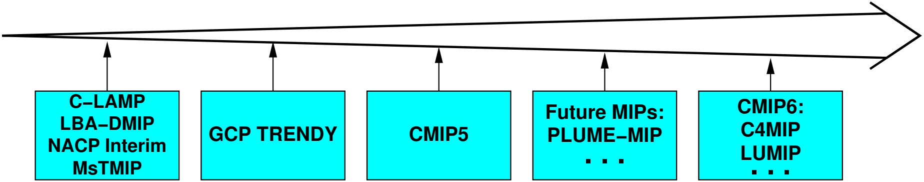
Model–Data–Experimentation Strategy



Why Benchmark?

- **to demonstrate model improvements** in representation of coupled climate and biogeochemical cycles
- **to quantitatively diagnose impacts of model development** in related fields on carbon cycle processes
- **to guide synthesis efforts**, such as the Intergovernmental Panel on Climate Change (IPCC), in assessing model fidelity
- **to increase scrutiny of key datasets** used for model evaluation
- **to identify gaps in existing observations** needed for model validation
- **to accelerate incorporation of new measurements** for rapid and widespread use in model assessment
- **to provide a quantitative, application-specific set of minimum criteria** for participation in model intercomparison projects (MIPs).

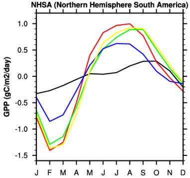
An Open Source Benchmarking Software System



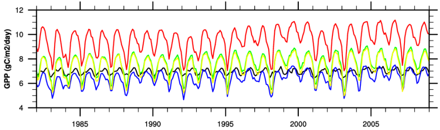
- Human capital costs of making rigorous model-data comparisons is considerable and constrains the scope of individual MIPs.
- Many MIPs spend resources "reinventing the wheel" in terms of variable naming conventions, model simulation protocols, and analysis software.
- **Need for ILAMB:** Each new MIP has access to the model–data comparison modules from past MIPs through ILAMB (*e.g.*, MIPs use one common modular software system). Standardized international naming conventions also increase MIP efficiency.

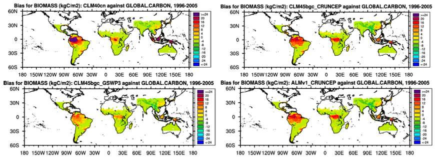
Current Status of the ILAMB Packages

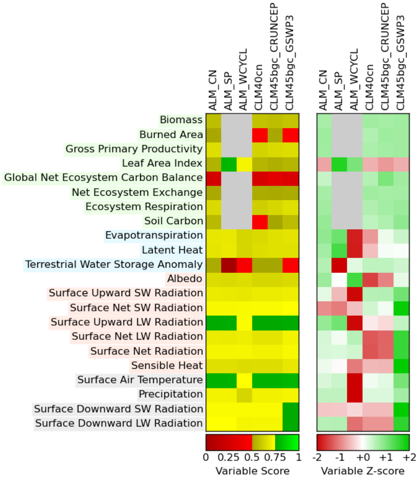
- **ILAMBv1** released at 2015 AGU Town Hall, doi:10.18139/ILAMB.v001.00/1251597
- **ILAMBv2** released at 2016 ILAMB Workshop, doi:10.18139/ILAMB.v002.00/1251621
- Being used for ACME and CESM evaluation



Model	Annual	Bias	RMSE
FLUXNET-MTE	6.95	-999.00	-999.00
CLM40cn	9.55	2.60	2.73
CLM45bgc_CRUNCEP	7.62	0.67	0.96
CLM45bgc_GSWP3	6.42	-0.53	0.71
ALMv1_CRUNCEP	7.43	0.48	0.69







ILAMB Prototype Diagnostics System

- Current variables:

  Aboveground live biomass (Contiguous US, Pan Tropical Forest), Burned area (GFED3), CO_2 (NOAA GMD, Mauna Loa), Gross primary production (Fluxnet, MTE), Leaf area index (AVHRR, MODIS), Global net land flux (GCP, Khatiwala/Hoffman), Net ecosystem exchange (Fluxnet, GBA), Ecosystem Respiration (Fluxnet, GBA), Soil C (HWSD, NCSCDv2), Evapotranspiration (GLEAM, MODIS), Latent heat (Fluxnet, MTE), Soil moisture (ESA), Terrestrial water storage anomaly (GRACE), Albedo (CERES, GEWEX, MODIS), Surface up SW/LW radiation (CERES, GEWEX.SRB, WRMC.BSRN), Sensible heat (Fluxnet, GBA), Surface air temperature (CRU, Fluxnet), Precipitation (Fluxnet, GPCC, GPCP2), Surface down SW/LW radiation (Fluxnet, CERES, GEWEX.SRB, WRMC.BSRN),

- Graphics and scoring systems:

  • Annual mean, Bias, RMSE, seasonal cycle, spatial distribution, interannual coeff. of variation and variability, long-term trend scores

  • Global maps, variable to variable, and time series comparisons

- Software:

  Freely distributed, designed to be user friendly and to enable easy addition of new variables

ILAMBv2 Layout

ILAMB Benchmark Results

Overview | Results Table

Columns...

	CLM40cn	CLM45bgc_CRUNCEP	CLM45bgc_GSWP3
Biomass	0.63	0.65	0.70
Burned Area	0.35	0.49	0.50
Gross Primary Productivity	0.68	0.72	0.74
Leaf Area Index	0.51	0.50	0.56
Global Net Ecosystem Carbon Balance	0.27	0.34	0.30
GCP (50.0%)	0.36	0.48	0.44
Hoffman (50.0%)	0.18	0.21	0.16
Net Ecosystem Exchange	0.49	0.48	0.49
Ecosystem Respiration	0.63	0.68	0.72
Soil Carbon	0.45	0.51	0.65
Evapotranspiration	0.73	0.76	0.78
Latent Heat	0.78	0.80	0.81
Terrestrial Water Storage Anomaly	0.50	0.50	0.48
Albedo	0.74	0.74	0.75
CERES (33.3%)	0.77	0.77	0.78
GEWEX.SRB (33.3%)	0.69	0.70	0.70
MODIS (33.3%)	0.75	0.75	0.76
Surface Upward SW Radiation	0.79	0.79	0.80
Surface Net SW Radiation	0.87	0.87	0.89
Surface Upward LW Radiation	0.94	0.94	0.95
Surface Net LW Radiation	0.78	0.79	0.84

ILAMBv2 Layout

Mean State

climate.ornl.gov/~ncf/ILAMB-next/EcosystemandCarbonCycle/GrossPrimaryProductivity/Fluxnet-MTE//Fluxnet-MTE.html

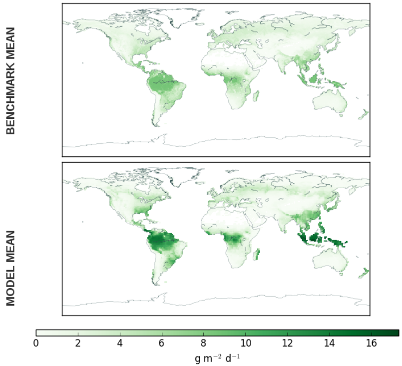
GrossPrimaryProductivity / Fluxnet-MTE / global / CLM40cn

Mean State | Relationships | Data Information

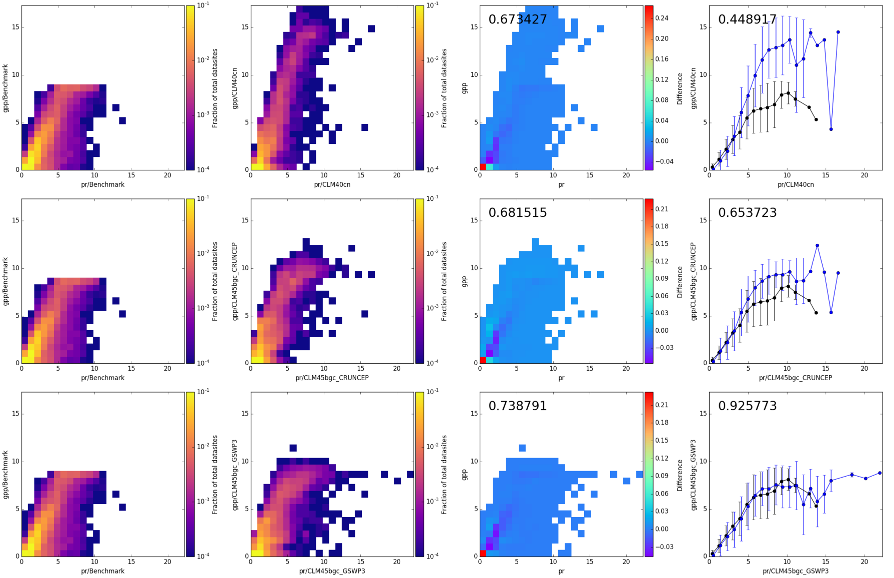
global

Model	Data	Period Mean [Pg yr-1]	Bias [Pg yr-1]	RMSE [Pg yr-1]	Phase Shift [d]	Bias Score [1]	RMSE Score [1]	Seasonal Cycle Score [1]	Spatial Distribution Score [-]	Overall Score [1]
Benchmark	[-]	115.711								
CLM40cn	[-]	130.054	15.099	73.431	-1.661	0.66	0.591	0.783	0.77	0.679
CLM45bgc_CRUNCEP	[-]	117.757	3.041	66.672	-2.555	0.719	0.643	0.789	0.899	0.739
CLM45bgc_GSWP3	[-]	107.227	-6.927	63.287	-1.648	0.729	0.655	0.776	0.922	0.748

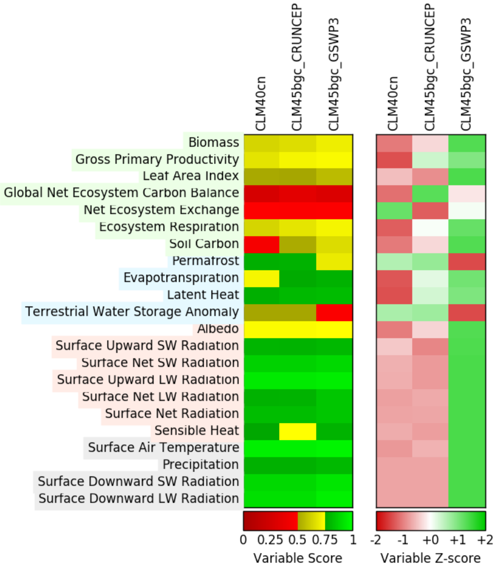
Temporally integrated period mean

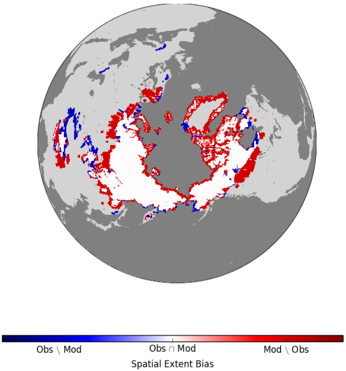


ILAMBv2 Relationships (Under Development)

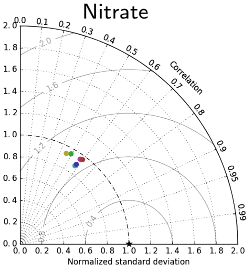


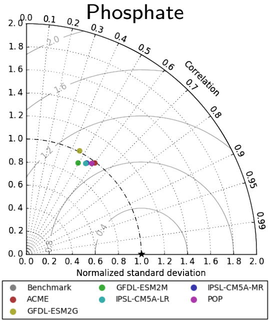
Latest ILAMB Adds Permafrost Extent

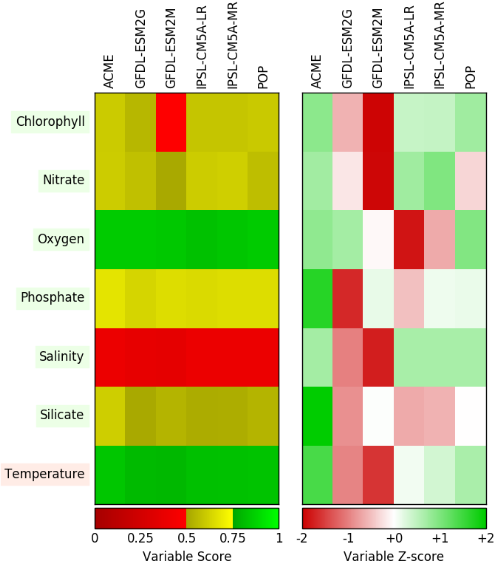




Extending ILAMB for Ocean Model Evaluation







Second US ILAMB Workshop, May 16–18, 2016

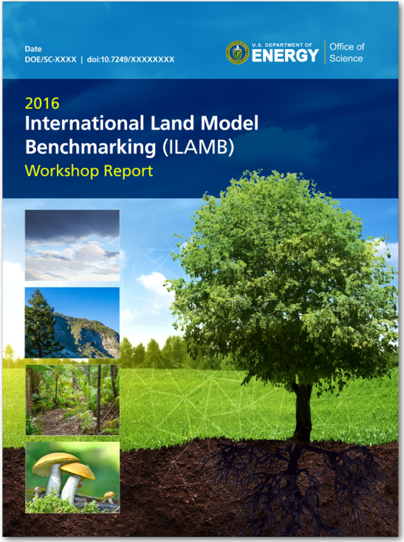
Overarching Workshop Goals

Engage the research community in defining scientific priorities for

- Design of new metrics for model benchmarking
- Model Intercomparison Project (MIP) evaluation needs
- Model development, testbeds, and workflow practices
- Observational data sets and needed measurements

Workshop Attendance

- 60+ participants from Australia, Japan, China, Germany, Sweden, Netherlands, UK, and US
- 10 modeling centers represented
- ∼25 online attendees at any time



doi:10.2172/1330803

2016 ILAMB Workshop Synthesis

Integrating and Cross-cutting Themes

- Process-specific experiments
- Metrics from extreme events
- Design of new perturbation experiments
- High latitude processes
- Tropical processes
- Remote sensing
- Eddy covariance flux networks

Model Intercomparison Projects (MIPs)

- CMIP6 DECK
- Coupled Climate–Carbon Cycle (C4MIP)
- Land Surface, Snow, and Soil Moisture (LS3MIP)
- Multi-scale Synthesis & Terrestrial (MsTMIP)
- Processes Linked to Uncertainties Modeling Ecosystems (PLUME-MIP)

Major Processes

- Ecosystem processes and states
- Hydrology
- Atmospheric CO_2
- Soil carbon and nutrient biogeochemistry
- Surface fluxes
- Vegetation dynamics

Benchmarking Approaches

- Statistical comparisons (bias, RMSE, etc.)
- Functional response or variable-to-variable
- Emergent constraints
- Reduced complexity models & traceability
- Formal uncertainty quantification
- Meta-analyses of perturbation experiments

Benchmarking Challenges and Priorities

- Develop super site 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

Enabling Capabilities

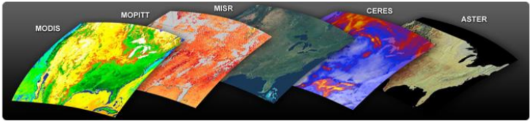
- Model development and new output variables
- Land model testbeds (LMTs)
- Field measurements and monitoring activities
- Perturbation experiments and lab studies
- Observational data archives and repositories
- Computational resources and infrastructure

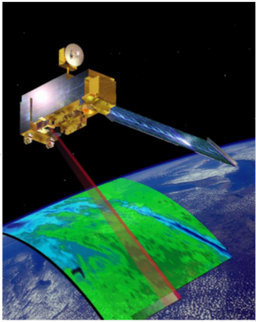
Benchmarking Advances

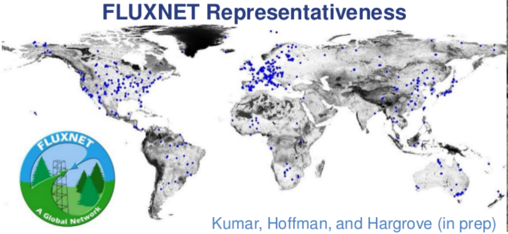
- Process understanding
- Quantified feedbacks
- Reduced uncertainties
- Improved model projections

Benchmarking Challenges and Priorities

- **Super site benchmarks** for AmeriFlux and FLUXNET
- **Benchmarks for soil carbon** turnover, distribution, transport
- **Metrics for extreme events** & response of ecosystems
- **Data for vegetation** recruitment, growth, mortality, phenology, canopy structure
- Benchmarks for critical **high latitude & tropical ecosystems**
- Leverage **field projects & remote sensing methods**







Kumar, Hoffman, and Hargrove (in prep)

Future ILAMB Developent and Application

- ILAMBv1 ILAMBv2 were applied to:
  - CMIP5 Historical and esmHistorical simulations
  - ACME Land Model evaluation
  - Model development of the Community Land Model (CLM)
- Within U.S. Department of Energy projects:
  - NGEE Arctic, NGEE Tropics, and SPRUCE are adopting the framework for evaluating process parameterizations & integrating field observations
  - ACME is developing metrics for evaluation of new land model features
  - BGC Feedbacks is developing the framework and benchmarking MIPs
- Future projects where we hope to apply ILAMB:
  - CMIP6, including C^4MIP, LS3MIP, and LUMIP
  - TRENDY
  - PLUME-MIP
- Others are using and contributing to ILAMB:
  - a NASA-funded Permafrost Benchmarking System
  - in-house model evaluation at Hadley Center, U. Tokyo, MPI-Met

Acknowledgments

U.S. DEPARTMENT OF ENERGY

Office of Science

This research was performed for the Biogeochemistry–Climate Feedbacks Scientific Focus Area, which is sponsored by the Regional and Global Climate Modeling (RGCM) 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. Oak Ridge National Laboratory (ORNL) is managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

References

F. M. Hoffman, 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. D. 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. International Land Model Benchmarking (ILAMB) 2016 workshop report. Technical Report DOE/SC-0186, U.S. Department of Energy, Office of Science, Germantown, Maryland, USA, 2016. doi:10.2172/1330803.

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.

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. *Global Change Biol.*, 15(9):2462–2484, Sept. 2009. doi:10.1111/j.1365-2486.2009.01912.x.