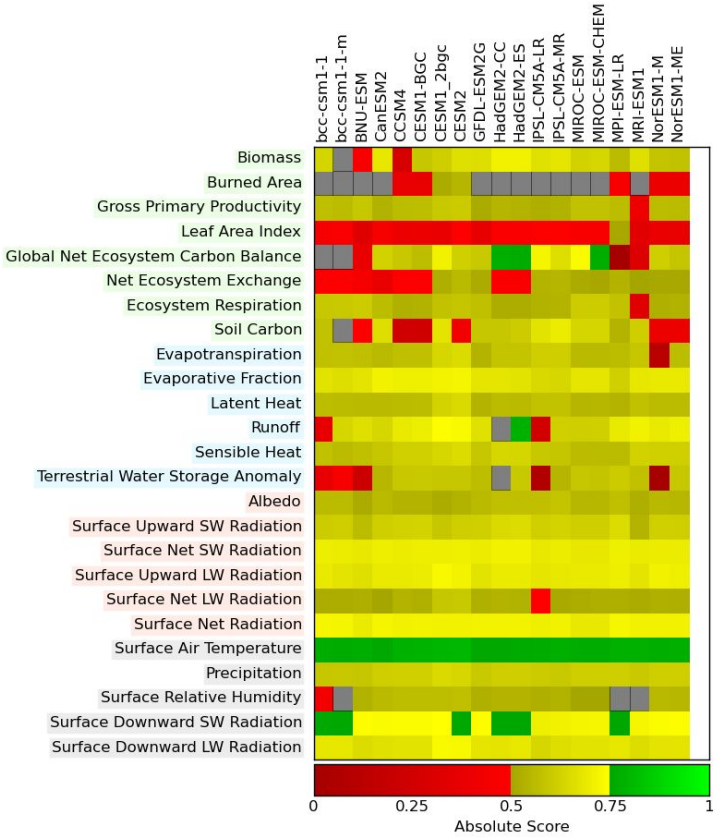
Land Model Benchmarking and Pathways to a Land Model Testbed (LMT)

Forrest M. Hoffman
Oak Ridge National Laboratory

International Land Model Benchmarking (ILAMB) Package

- Provides systematic assessment of land model results compared with observations
- Scores model performance across a wide range of independent benchmark data sets
- Includes comparison of functional relationships (variable-to-variable comparisons)
- Written in Python and runs in parallel
- Produced from an international community coordination effort for designing metrics
- Supported primarily by RUBISCO SFA with support for metrics from E3SM and new observational data from NGEE Arctic & Tropics



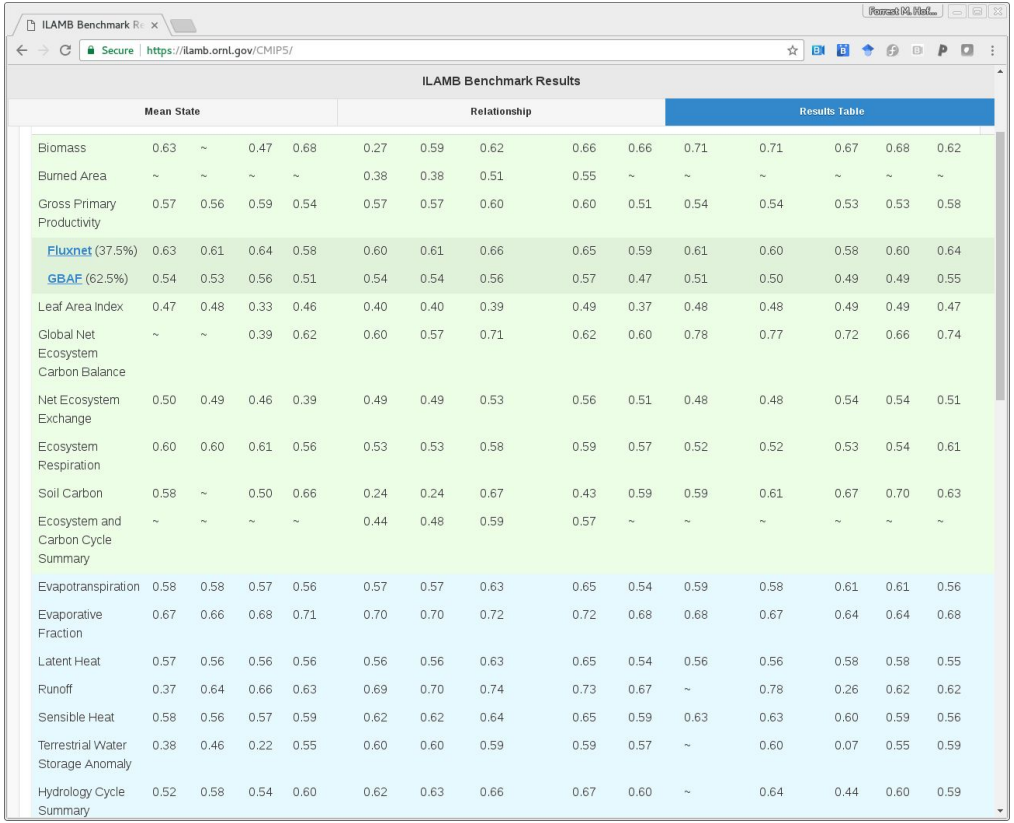
International Land Model Benchmarking (ILAMB) Package

- We invested effort in providing a rich hierarchical user interface
- The top level overview provides “portrait plots” of absolute and relative model scores
- Scores are aggregated from multiple data sets and metrics for each variable

ILAMB Package Features

- Currently integrates analysis of 25 variables in 4 categories from ~60 datasets
  - aboveground live biomass, burned area, carbon dioxide, gross primary production, leaf area index, global net ecosystem carbon balance, net ecosystem exchange, ecosystem respiration, soil carbon
  - evapotranspiration, latent heat, sensible heat, runoff, evaporative fraction, terrestrial water storage anomaly
  - albedo, surface upward SW + LW radiation, surface net SW + LW radiation, surface net radiation
  - surface air temperature, precipitation, surface relative humidity, surface downward SW +LW radiation
- Graphics and scoring system
  - plots and scores model performance for annual mean, bias, relative bias, RMSE, seasonal cycle phase, spatial distribution, interannual variability, variable-to-variable comparisons
  - includes global maps, time series plots averaged over specific regions, individual measurement sites, functional relationship plots, capability to zoom in on specific regions
- Open Source (https://www.ilamb.org/)
  - ILAMBv2.2 is available at https://www.bgc-feedbacks.org/software/

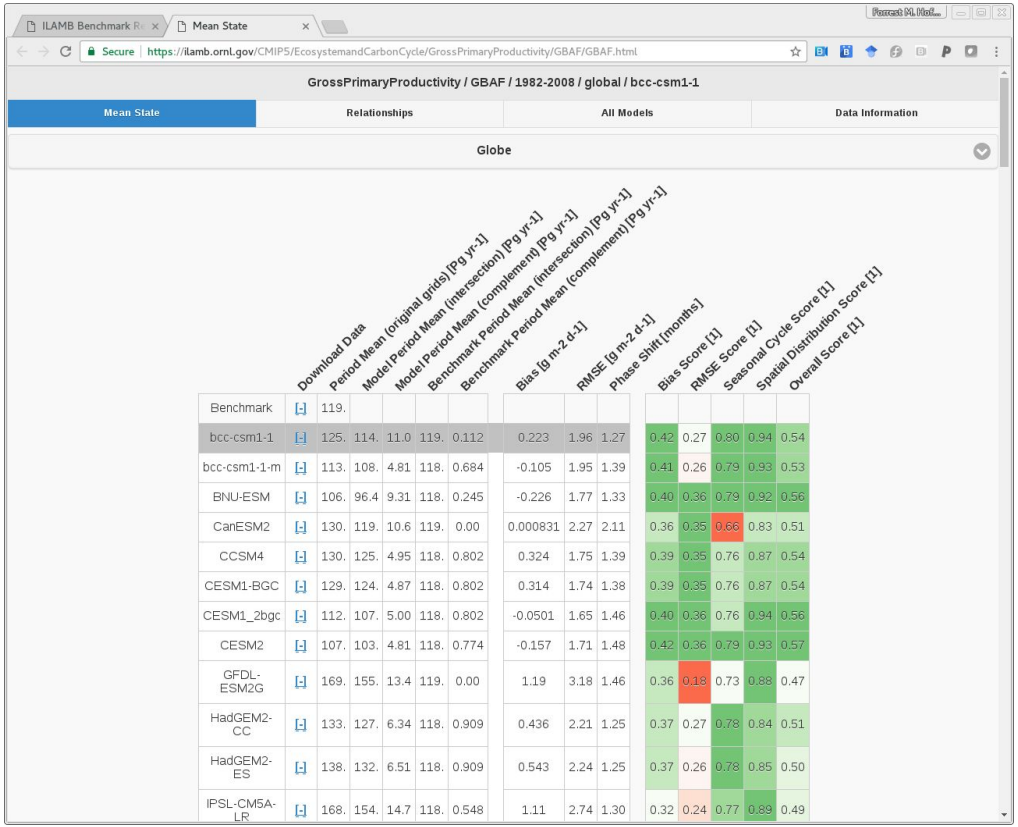
ILAMB Package Results Table



Biomass	0.63	~	0.47	0.68	0.27	0.59	0.62	0.66	0.66	0.71	0.71	0.67	0.68	0.62
Burned Area	~	~	~	~	0.38	0.38	0.51	0.55	~	~	~	~	~	~
Gross Primary Productivity	0.57	0.56	0.59	0.54	0.57	0.57	0.60	0.60	0.51	0.54	0.54	0.53	0.53	0.58
Fluxnet (37.5%)	0.63	0.61	0.64	0.58	0.60	0.61	0.66	0.65	0.59	0.61	0.60	0.58	0.60	0.64
GBAF (62.5%)	0.54	0.53	0.56	0.51	0.54	0.54	0.56	0.57	0.47	0.51	0.50	0.49	0.49	0.55
Leaf Area Index	0.47	0.48	0.33	0.46	0.40	0.40	0.39	0.49	0.37	0.48	0.48	0.49	0.49	0.47
Global Net Ecosystem Carbon Balance	~	~	0.39	0.62	0.60	0.57	0.71	0.62	0.60	0.78	0.77	0.72	0.66	0.74
Net Ecosystem Exchange	0.50	0.49	0.46	0.39	0.49	0.49	0.53	0.56	0.51	0.48	0.48	0.54	0.54	0.51
Ecosystem Respiration	0.60	0.60	0.61	0.56	0.53	0.53	0.58	0.59	0.57	0.52	0.52	0.53	0.54	0.61
Soil Carbon	0.58	~	0.50	0.66	0.24	0.24	0.67	0.43	0.59	0.59	0.61	0.67	0.70	0.63
Ecosystem and Carbon Cycle Summary	~	~	~	~	0.44	0.48	0.59	0.57	~	~	~	~	~	~
Evapotranspiration	0.58	0.58	0.57	0.56	0.57	0.57	0.63	0.65	0.54	0.59	0.58	0.61	0.61	0.56
Evaporative Fraction	0.67	0.66	0.68	0.71	0.70	0.70	0.72	0.72	0.68	0.68	0.67	0.64	0.64	0.68
Latent Heat	0.57	0.56	0.56	0.56	0.56	0.56	0.63	0.65	0.54	0.56	0.56	0.58	0.58	0.55
Runoff	0.37	0.64	0.66	0.63	0.69	0.70	0.74	0.73	0.67	~	0.78	0.26	0.62	0.62
Sensible Heat	0.58	0.56	0.57	0.59	0.62	0.62	0.64	0.65	0.59	0.63	0.63	0.60	0.59	0.56
Terrestrial Water Storage Anomaly	0.38	0.46	0.22	0.55	0.60	0.60	0.59	0.59	0.57	~	0.60	0.07	0.55	0.59
Hydrology Cycle Summary	0.52	0.58	0.54	0.60	0.62	0.63	0.66	0.67	0.60	~	0.64	0.44	0.60	0.59

- Results Table shows scores for each model (columns) by variable (rows)
- Each variable is a “pull-down” for multiple data sets (see GPP for Fluxnet and GBAF)
- Clicking on the data set opens a new browser tab with tabular and graphical diagnostics

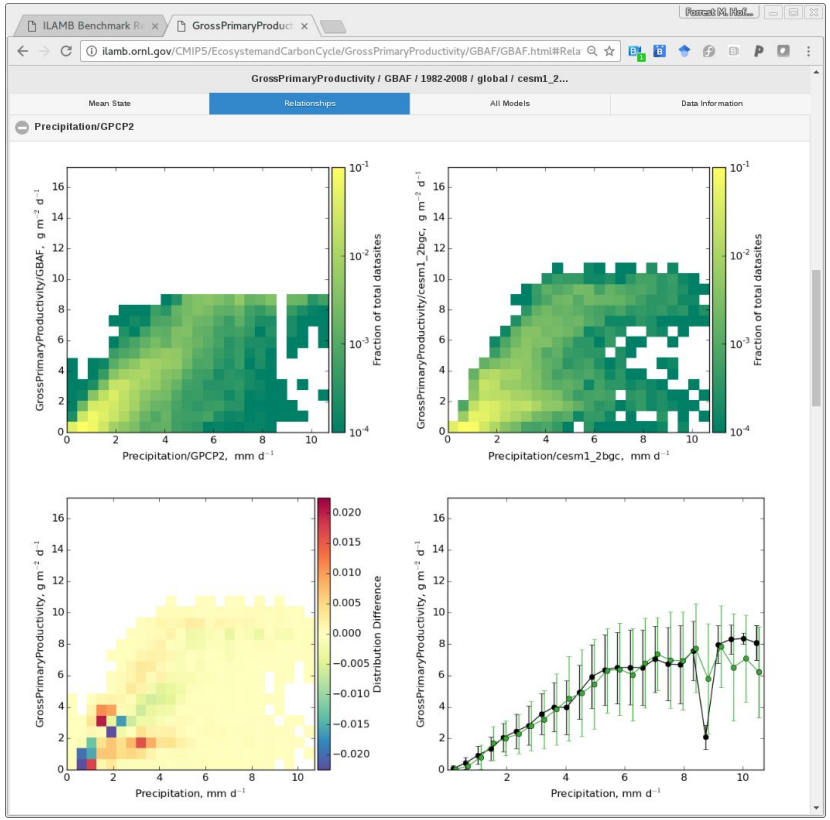
ILAMB Package Results Table



	Download Data	Period Mean (original grids) [Pg yr-1]	Model Period Mean (intersection) [Pg yr-1]	Model Period Mean (complement) [Pg yr-1]	Benchmark Period Mean (intersection) [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	[-]	119.												
bcc-csm1-1	[-]	125.	114.	11.0	119.	0.112	0.223	1.96	1.27	0.42	0.27	0.80	0.94	0.54
bcc-csm1-1-m	[-]	113.	108.	4.81	118.	0.684	-0.105	1.95	1.39	0.41	0.26	0.79	0.93	0.53
BNU-ESM	[-]	106.	96.4	9.31	118.	0.245	-0.226	1.77	1.33	0.40	0.36	0.79	0.92	0.56
CanESM2	[-]	130.	119.	10.6	119.	0.00	0.000831	2.27	2.11	0.36	0.35	0.66	0.83	0.51
CCSM4	[-]	130.	125.	4.95	118.	0.802	0.324	1.75	1.39	0.39	0.35	0.76	0.87	0.54
CESM1-BGC	[-]	129.	124.	4.87	118.	0.802	0.314	1.74	1.38	0.39	0.35	0.76	0.87	0.54
CESM1_2bgc	[-]	112.	107.	5.00	118.	0.802	-0.0501	1.65	1.46	0.40	0.36	0.76	0.94	0.56
CESM2	[-]	107.	103.	4.81	118.	0.774	-0.157	1.71	1.48	0.42	0.36	0.79	0.93	0.57
GFDL-ESM2G	[-]	169.	155.	13.4	119.	0.00	1.19	3.18	1.46	0.36	0.18	0.73	0.88	0.47
HadGEM2-CC	[-]	133.	127.	6.34	118.	0.909	0.436	2.21	1.25	0.37	0.27	0.78	0.84	0.51
HadGEM2-ES	[-]	138.	132.	6.51	118.	0.909	0.543	2.24	1.25	0.37	0.26	0.78	0.85	0.50
IPSL-CM5A-LR	[-]	168.	154.	14.7	118.	0.548	1.11	2.74	1.30	0.32	0.24	0.77	0.89	0.49

- Models can be selected individually and diagnostics update
- Separate statistics and figures are produced for pre-defined regions
- Relationships tab contains variable-to-variable comparisons
- Data provenance provided in Data Information tab

ILAMB Functional Relationships

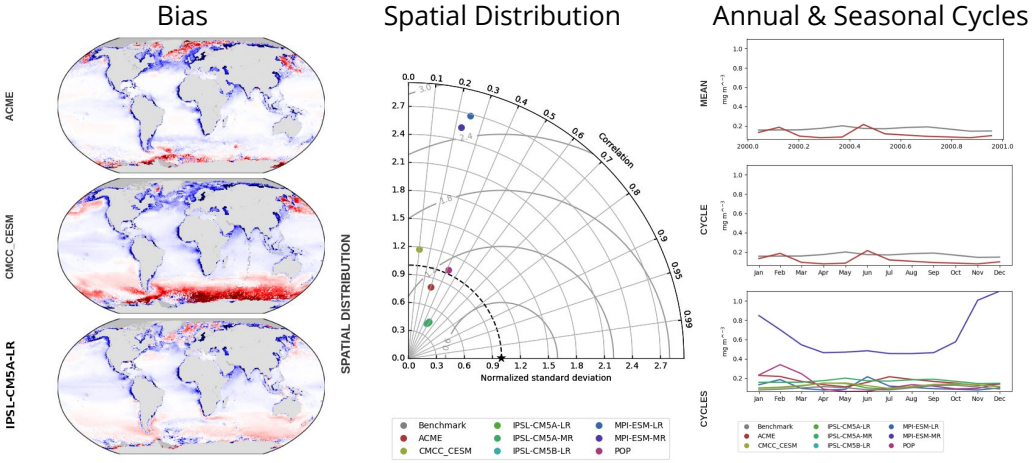


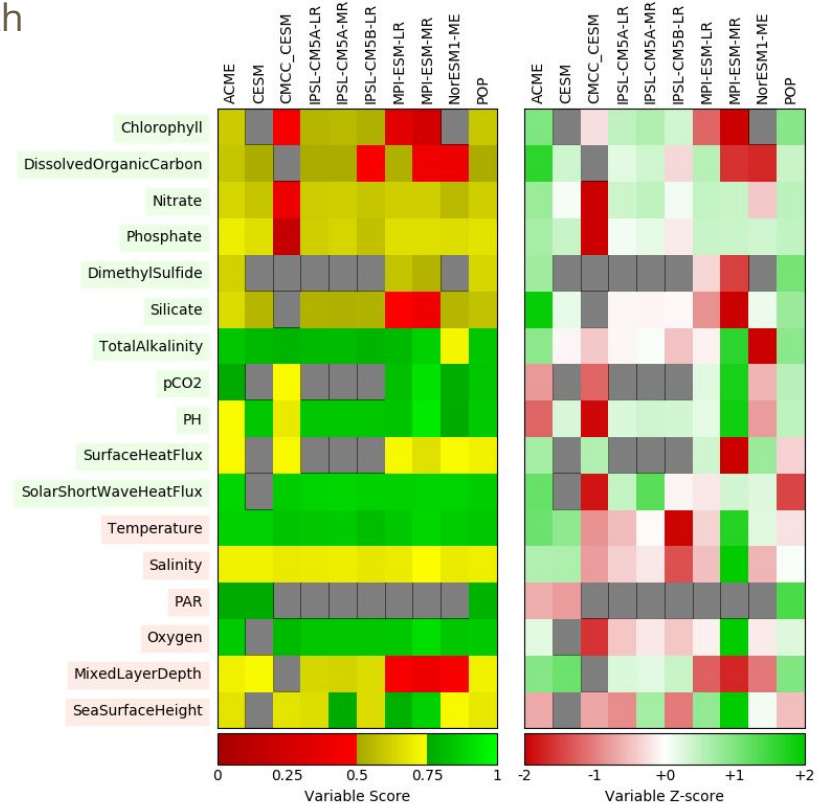
- Variable-to-variable comparisons provide a better way to understand model responses to forcing
- Shown here is GPP vs. Precipitation for a single model compared with observations
- Differences in distribution of points suggests regimes in which model errors are most significant
- Histogram-style line plots indicate if model exhibits overall relationships emerging from observational data

International Ocean Model Benchmarking (IOMB) Package

- Evaluates ocean biogeochemistry results compared with observations (global, regional points, and ship tracks)
- Scores model performance across a wide range of independent benchmark data
- Leverages ILAMB code base; also runs in parallel
- Will be released to the community soon

Chlorophyll / SeaWIFS

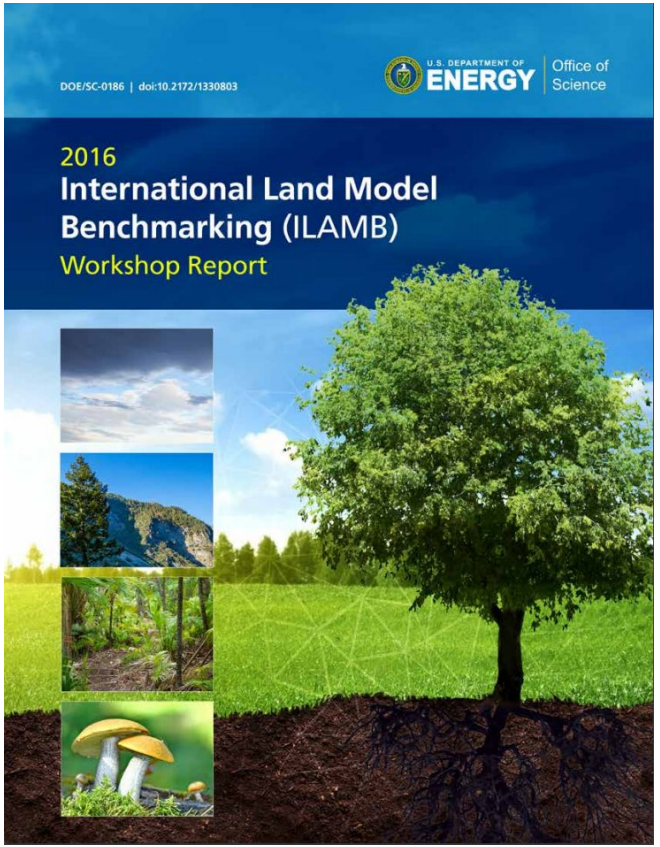




ILAMB and IOMB Target Uses

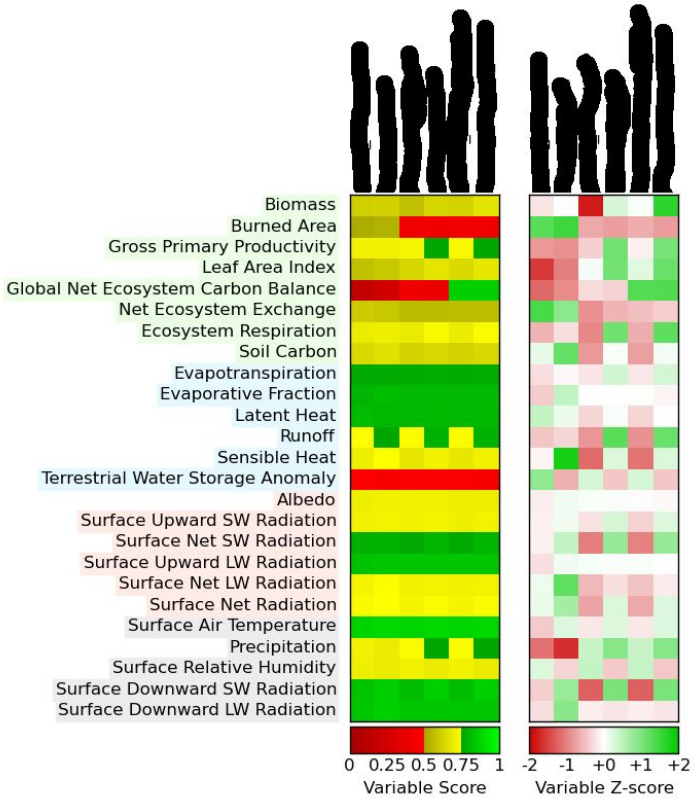
- ILAMB is designed for use by
  - Individual modelers/model developers - for verification
  - Modeling centers - to track model performance evolution
  - Model intercomparison experiments - for multi-model analysis

- ILAMB is being used & developed by the international land model community
  - DOE E3SM - Workflow and Land Model Intercomparison
  - NSF / DOE CESM at NCAR - Workflow (land and ocean)
  - University of New South Wales / PALS / modelevaluation.org - Analysis engine
  - CEH / JULES / EartH2Observe - Published analysis
  - NOAA GFDL - Adding it to their toolkit
  - NASA ABoVE / NOAA NSIDC - Permafrost metrics
  - University of Tokyo / GSWP3 - Runoff metrics and evaluation



E3SM Land Model (ELM) Intercomparison

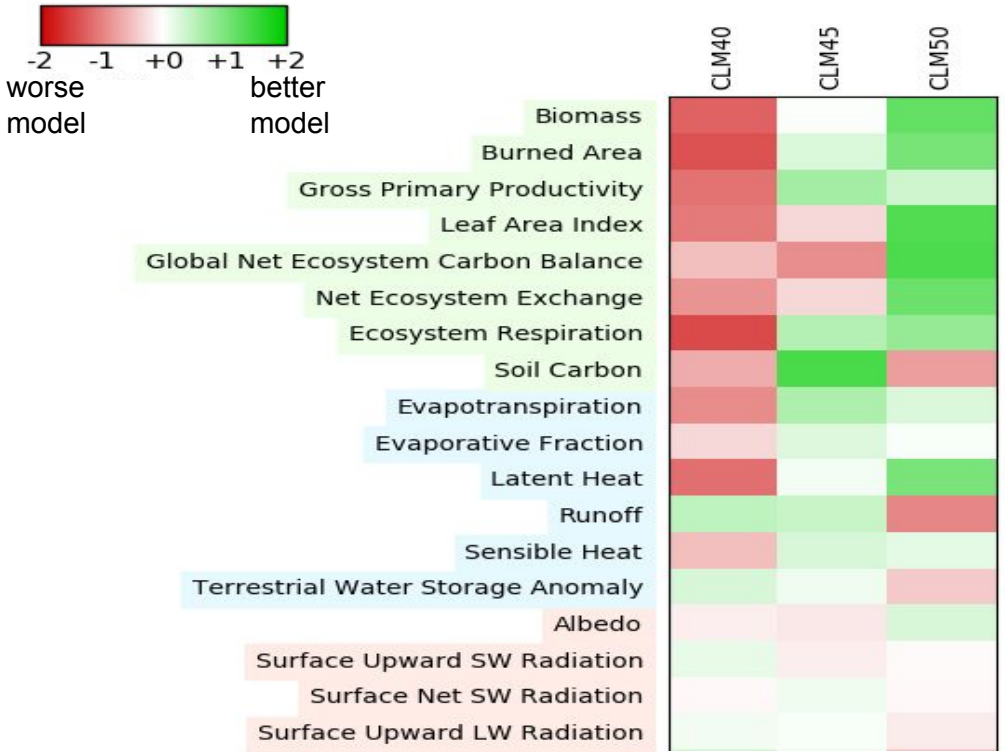
Gross Primary Productivity / GBAF / 1982-2008



	Download Data	Period Mean (original grids) [Pg yr-1]	Model Period Mean (intersection) [Pg yr-1]	Model Period Mean (complement) [Pg yr-1]	Benchmark Period Mean (intersection) [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	[-]	118.810												
[redacted]	[-]	111.448	102.592	8.851	118.383	0.427	-0.037	1.573	1.244	0.727	0.648	0.812	0.862	0.739
[redacted]	[-]	105.181	97.013	8.162	118.383	0.427	-0.185	1.536	1.254	0.737	0.654	0.798	0.865	0.741
[redacted]	[-]	138.154	128.193	9.954	118.383	0.427	0.542	1.614	1.181	0.745	0.644	0.825	0.864	0.744
[redacted]	[-]	137.237	127.339	9.891	118.383	0.427	0.523	1.619	1.139	0.746	0.642	0.825	0.865	0.744
[redacted]	[-]	126.756	117.893	8.856	118.383	0.427	0.281	1.404	1.335	0.790	0.685	0.801	0.899	0.772
[redacted]	[-]	125.544	116.790	8.748	118.383	0.427	0.255	1.412	1.303	0.789	0.681	0.802	0.900	0.771

- An enhanced version of ILAMB is being used to assess multiple land biogeochemistry formulations in ELM
- The ELM Intercomparison, led by Ben Bond-Lamberty, is using ILAMB and other tools and metrics to identify optimal model configurations

ILAMB assessing several generations of CLM



- ILAMB was used as an integral part of CLM5.0 development
- Improvements in mechanistic treatment of hydrology, ecology, and land use with many more moving parts
- Simulation improved even with enhanced complexity
- Observational datasets not always self-consistent
- Forcing uncertainty confounds assessment of model development (not shown)

Lawrence et al., in prep

ILAMB and IOMB Development

- Openly developed in Python using Git repository
  - https://bitbucket.org/ncollier/ilamb
  - Patches welcome! We have had features and bug fixes submitted by users
- Roughly biannual releases
  - v2.0 - May 2016; v2.1 - March 2017; v2.2 - November 2017; v2.3 - March 2018
- Development activity
  - Develop new benchmarks for E3SM and modeling working groups
  - Adapt the ILAMB core to address community needs (ocean, high latitude, diurnal cycle)
  - Address computing environments and performance (laptops, clusters, NERSC, OLCF & ALCF)
  - Hone and improve the current methodology *with research community*
  - Continually improve documentation and tutorials (Provided at major meetings)
- Tracking use through software DOIs, workshop engagement, and interactive website visits — Many users will simply look at results!

Watersheds and Testbeds

- ILAMB architecture was designed to be flexible and extensible
  - Site-level evaluation and diurnal and seasonal cycle assessment are being added
  - New metrics for high latitude/permafrost being implemented
  - Benchmarking philosophy and methodology in Collier et al. (*JAMES*, submitted)
- ILAMB design could be extended to provide verification and validation for watershed models
- New development of models *across scales* is required for improvement of process representations and producing observable quantities or simulating remote sensing observations
- Land Model Testbed (LMT) is needed for execution, calibration, and evaluation of alternative model formulations
- LMT should be incorporated into routine model testing (e.g., nightly or weekly automated testing) and provide diagnostics for understanding model changes

Extra Slides

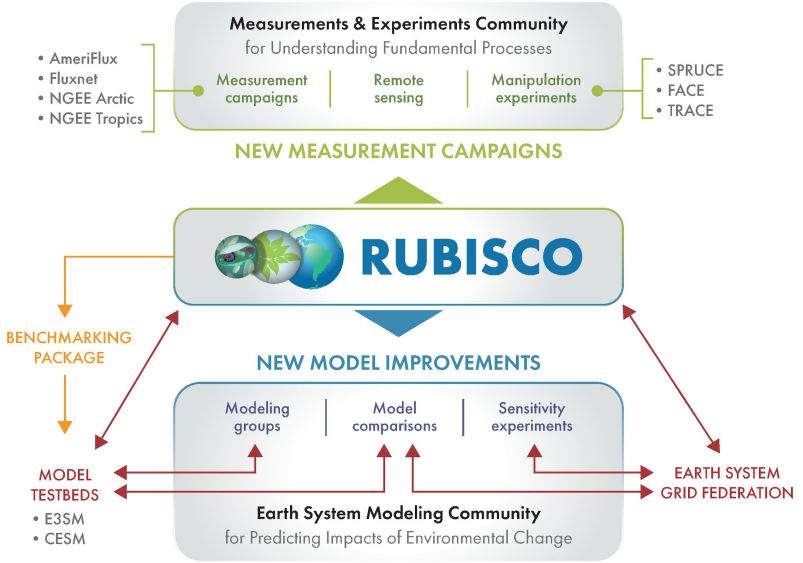
US Dept. of Energy's RUBISCO Scientific Focus Area (SFA)

Research Goals

- Identify and quantify interactions between biogeochemical cycles and the Earth system
- Quantify and reduce uncertainties in Earth system models (ESMs) associated with interactions.

Research Objectives

- Perform hypothesis-driven analysis of biogeochemical & hydrological processes and feedbacks in ESMs
- Synthesize in situ and remote sensing data and design metrics for assessing ESM performance
- Design, develop, and release the International Land Model Benchmarking (ILAMB) and International Ocean Model Benchmarking (IOMB) tools for systematic evaluation of model fidelity
- Conduct and evaluate CMIP6 experiments with ESMs



The RUBISCO SFA works with the measurements and the modeling communities to use best-available data to evaluate the fidelity of ESMs. RUBISCO identifies model gaps and weaknesses, informs new model development efforts, and suggests new measurements and field campaigns.

Argonne NATIONAL LABORATORY · BROOKHAVEN NATIONAL LABORATORY · Berkeley Lab · Los Alamos NATIONAL LABORATORY · NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH · OAK RIDGE National Laboratory · UCI · MICHIGAN