Benchmark Analysis for Improved Prediction

Forrest M. Hoffman[1], William J. Riley[2], Gretchen Keppel-Aleks[3], David M. Lawrence[4], J. Keith Moore[5], and James T. Randerson[5]

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

CCIWG/NACP Workshop on Development of Predictive Carbon Cycle Science

NOAA NCWCP Conference Center, College Park, Maryland, USA

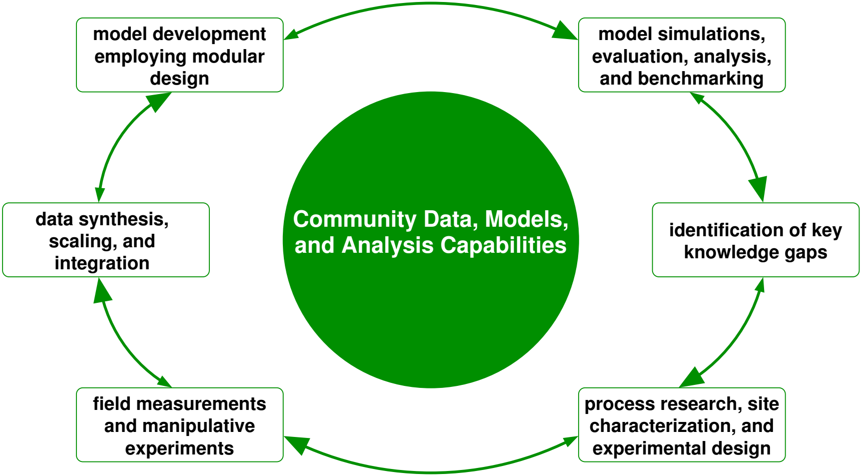
March 7–9, 2016

Climate Change Science Institute

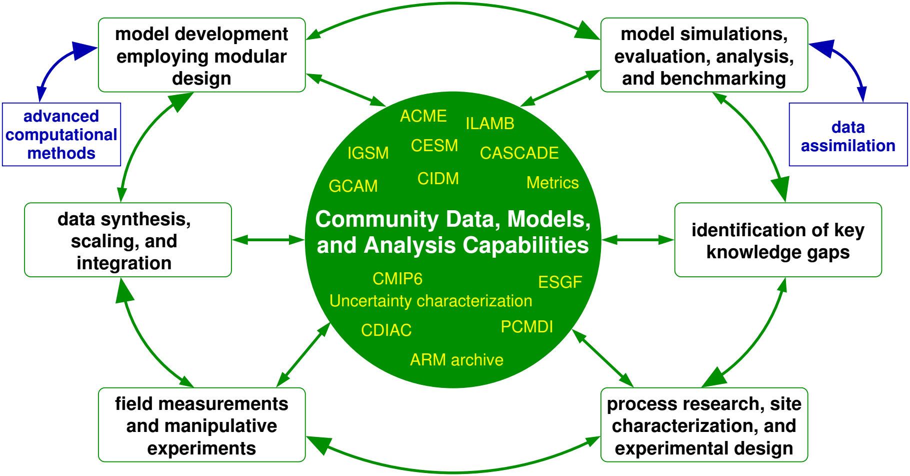
Oak Ridge National Laboratory

Oak Ridge National Laboratory

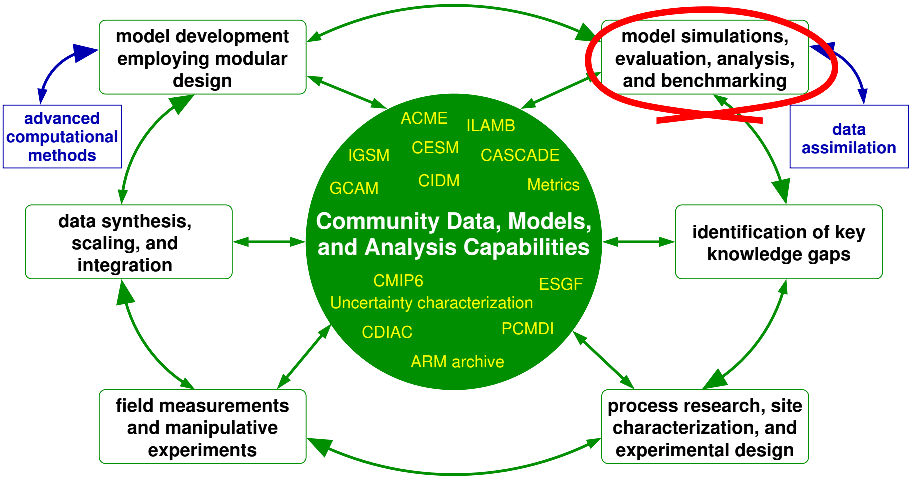
Model, Experiment, and Data Integration Strategy



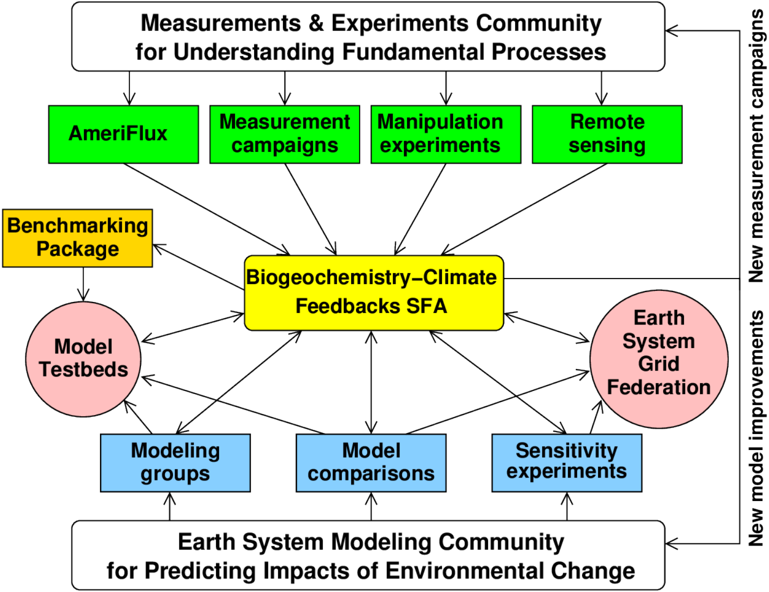
Model, Experiment, and Data Integration Strategy



Model, Experiment, and Data Integration Strategy

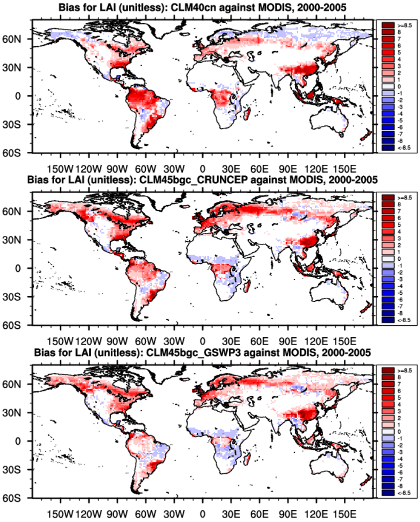


Biogeochemistry–Climate Feedbacks SFA Diagram

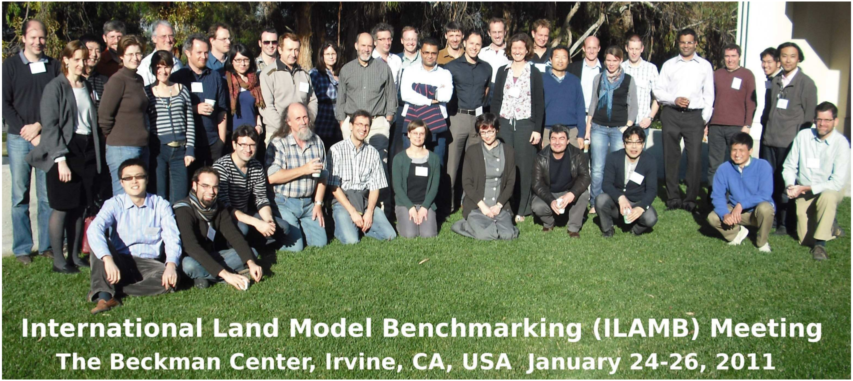


What is ILAMB?

- The **International Land Model Benchmarking (ILAMB)** project seeks to develop internationally accepted standards for land model evaluation.
- Model **benchmarking** can diagnose impacts of model development and guide synthesis efforts like IPCC.
- **Effective benchmarks** must draw upon a broad set of independent observations to evaluate model performance on multiple temporal and spatial scales.
- A free, **open source analysis and diagnostics software package** for community use will enhance model intercomparison projects.



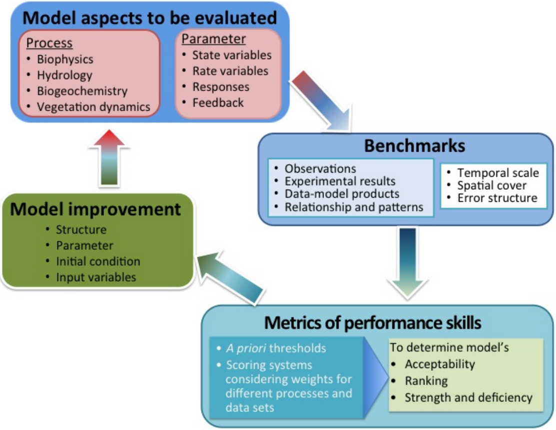
Bias in mean annual leaf area index from comparison of three versions CLM with MODIS.



- 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.
- Methodology for model–data comparison and baseline standard for performance of land model process representations (Luo et al., 2012).

Benchmarking Metholdology (Luo et al., 2012)

- Based on this methodology and prior work in C-LAMP, we developed a new model benchmarking package for ILAMB.
- Prototype is ready for use in NCL and a new version is under development using python.



(Luo et al., 2012)

ILAMB Prototype developed by Mingquan Mu at UCI

- Assesses 24 variables in 4 categories frm ∼45 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, terrestrial water storage anomaly
  - albedo, surface upward SW radiation, surface net SW radiation, surface upward LW radiation, surface net LW radiation, surface net radiation, sensible heat
  - surface air temperature, precipitation, surface relative humidity, surface downward SW radiation, surface downward LW radiation
- Graphics and scoring system
  - annual mean, bias, RMSE, seasonal cycle, spatial distribution, interannual coefficient of variation, spatial distribution, long-term trend
- Software is available at `http://redwood.ess.uci.edu/mingquan/www/ILAMB/index.html`

ILAMB Prototype: Global Variables for 12 Models

Global Variables (Info for Weightings)

	MeanModel	bcc-csm1-1-m	BNU-ESM	CanESM2	CESM1-BGC	GFDL-ESM2G	HadGEM2-ES	inmcm4	IPSL-CM5A-LR	MIROC-ESM	MPI-ESM-LR	MRI-ESM1	NorESM1-ME
Aboveground Live Biomass	0.68	0.52	0.50	0.61	0.65	0.58	0.67	0.54	0.68	0.52	0.51	0.67	0.65
Burned Area	0.38	-	-	-	0.37	-	-	-	-	-	0.38	-	0.38
Carbon Dioxide	0.85	-	0.65	0.65	0.78	0.65	-	-	-	0.79	0.68	0.68	0.75
Gross Primary Productivity	0.77	0.72	0.73	0.64	0.70	0.67	0.68	0.70	0.67	0.69	0.69	0.53	0.70
Leaf Area Index	0.66	0.66	0.41	0.60	0.53	0.49	0.59	0.68	0.66	0.62	0.68	0.43	0.50
Global Net Ecosystem Carbon Balance	0.58	-	0.38	0.27	0.38	0.18	-	0.46	0.25	0.38	0.42	0.27	0.40
Net Ecosystem Exchange	0.49	0.47	0.47	0.39	0.48	0.49	0.46	0.44	0.53	0.48	0.50	0.48	0.48
Ecosystem Respiration	0.75	0.72	0.72	0.65	0.67	0.71	0.66	0.70	0.67	0.68	0.68	0.47	0.66
Soil Carbon	0.55	0.50	0.42	0.56	0.38	0.51	0.51	0.53	0.57	0.53	0.41	0.53	0.39
Summary	0.64	0.59	0.54	0.54	0.55	0.53	0.59	0.57	0.57	0.58	0.54	0.51	0.55
Evapotranspiration	0.75	0.73	0.72	0.72	0.73	0.70	0.74	0.69	0.75	0.70	0.73	0.73	0.72
Latent Heat	0.80	0.76	0.77	0.77	0.78	0.74	0.77	0.72	0.77	0.75	0.76	0.78	0.76
Terrestrial Water Storage Anomaly	0.53	0.45	0.35	0.54	0.48	0.43	-	0.52	0.45	0.52	0.55	0.47	0.45
Summary	0.69	0.65	0.61	0.68	0.66	0.62	0.75	0.64	0.65	0.66	0.68	0.66	0.64
Albedo	0.72	0.71	0.61	0.71	0.73	0.69	0.74	0.67	0.71	0.67	0.73	0.64	0.72
Surface Upward SW Radiation	0.78	0.73	0.67	0.74	0.78	0.74	0.77	0.74	0.74	0.72	0.78	0.67	0.76
Surface Net SW Radiation	0.84	0.86	0.84	0.85	0.85	0.86	0.85	0.84	0.82	0.83	0.87	0.85	0.85
Surface Upward LW Radiation	0.90	0.91	0.91	0.91	0.92	0.91	0.92	0.89	0.90	0.91	0.92	0.92	0.92
Surface Net LW Radiation	0.81	0.82	0.81	0.79	0.82	0.81	0.83	0.79	0.78	0.78	0.81	0.82	0.81
Surface Net Radiation	0.78	0.79	0.76	0.80	0.80	0.80	0.79	0.74	0.77	0.76	0.80	0.78	0.80
Sensible Heat	0.76	0.69	0.70	0.71	0.75	0.69	0.75	0.66	0.69	0.69	0.69	0.72	0.72
Summary	0.79	0.78	0.75	0.78	0.80	0.78	0.80	0.75	0.76	0.76	0.79	0.77	0.79
Surface Air Temperature	0.87	0.87	0.85	0.85	0.88	0.85	0.87	0.85	0.87	0.85	0.88	0.88	0.87
Precipitation	0.70	0.67	0.66	0.67	0.70	0.68	0.72	0.68	0.68	0.68	0.70	0.69	0.69
Surface Relative Humidity	0.81	-	0.80	0.76	0.82	-	-	0.79	0.82	-	-	0.83	0.81
Surface Downward SW Radiation	0.86	0.88	0.87	0.87	0.88	0.87	0.87	0.87	0.83	0.86	0.88	0.86	0.88
Surface Downward LW Radiation	0.90	0.92	0.91	0.91	0.92	0.92	0.92	0.90	0.89	0.91	0.93	0.91	0.91
Summary	0.82	0.82	0.81	0.80	0.83	0.82	0.84	0.81	0.81	0.81	0.84	0.83	0.82
Overall	0.69	0.51	0.59	0.60	0.64	0.56	0.49	0.57	0.57	0.59	0.61	0.59	0.63

ILAMB Prototype: Global Variables for 12 Models

Global Variables (Info for Weightings)

	MeanModel	bcc-csm1-1-m	BNU-ESM	CanESM2	CESM1-BGC	GFDL-ESM2G	HadGE
Aboveground Live Biomass	0.68	0.52	0.50	0.61	0.65	0.58	0.6
Burned Area	0.38	-	-	-	0.37	-	-
Carbon Dioxide	0.85	-	0.65	0.65	0.78	0.65	-
Gross Primary Productivity	0.77	0.72	0.73	0.64	0.70	0.67	0.6
Leaf Area Index	0.66	0.66	0.41	0.60	0.53	0.49	0.5
Global Net Ecosystem Carbon Balance	0.58	-	0.38	0.27	0.38	0.18	-
Net Ecosystem Exchange	0.49	0.47	0.47	0.39	0.48	0.49	0.4
Ecosystem Respiration	0.75	0.72	0.72	0.65	0.67	0.71	0.6
Soil Carbon	0.55	0.50	0.42	0.56	0.38	0.51	0.5
Summary	0.64	0.59	0.54	0.54	0.55	0.53	0.5
Evapotranspiration	0.75	0.73	0.72	0.72	0.73	0.70	0.7
Latent Heat	0.80	0.76	0.77	0.77	0.78	0.74	0.7
Terrestrial Water Storage Anomaly	0.53	0.45	0.35	0.54	0.48	0.43	-
Summary	0.69	0.65	0.61	0.68	0.66	0.62	0.7
Albedo	0.72	0.71	0.61	0.71	0.73	0.69	0.7
Surface Upward SW Radiation	0.78	0.73	0.67	0.74	0.78	0.74	0.7
Surface Net SW	0.84	0.86	0.84	0.85	0.85	0.86	0.8

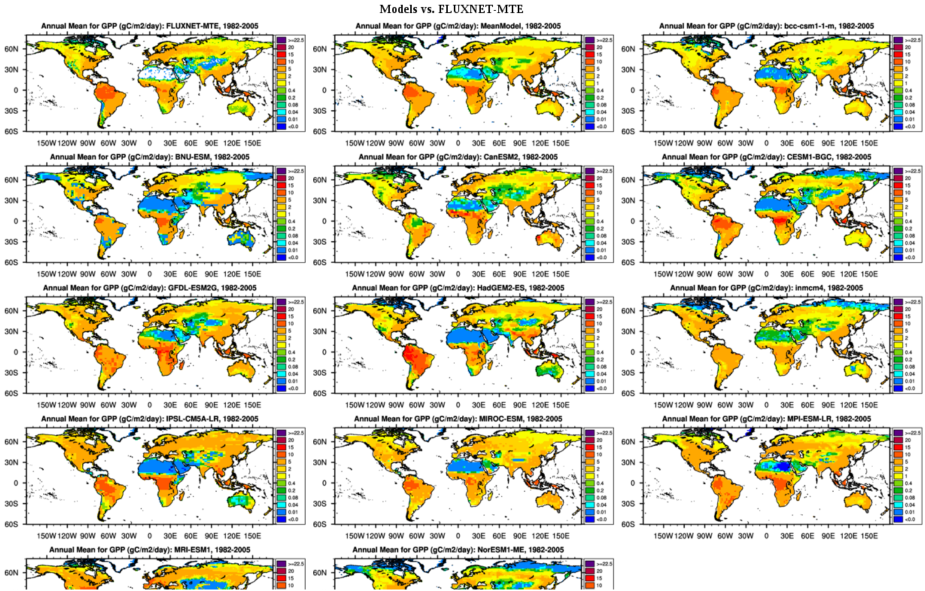
Scoring for Global GPP from Fluxnet-MTE

Diagnostic Summary for Gross Primary Productivity: Model vs. FLUXNET-MTE

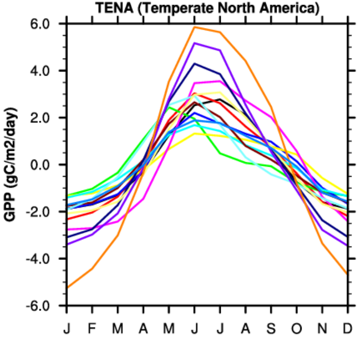
	Global Patterns				Regional and Seasonal Patterns	Scoring (Info)				
	Annual Mean (PgC/yr)	Bias (PgC/yr)	RMSE (PgC/mon)	Phase Difference (months)	Regional Means	Global Bias	RMSE	Seasonal Cycle	Spatial Distribution	Overall
Benchmark [Jung et al. (2009)]	118.4	-	-	0.0	access to plots	-	-	-	-	-
MeanModel	145.3	26.9	4.7	0.6	access to plots	0.77	0.73	0.78	0.94	0.79
bcc-csm1-1-m	114.4	-4.0	6.0	-0.2	access to plots	0.72	0.64	0.80	0.89	0.74
BNU-ESM	102.0	-16.4	6.2	0.1	access to plots	0.69	0.66	0.78	0.84	0.73
CanESM2	129.2	10.8	7.3	0.8	access to plots	0.64	0.60	0.68	0.70	0.64
CESM1-BGC	130.3	11.9	5.8	0.5	access to plots	0.69	0.65	0.76	0.87	0.72
GFDL-ESM2G	175.1	56.7	9.8	0.5	access to plots	0.66	0.54	0.73	0.83	0.66
HadGEM2-ES	145.9	27.5	7.4	0.3	access to plots	0.65	0.58	0.78	0.79	0.68
inmcm4	111.4	-7.0	5.6	0.3	access to plots	0.71	0.66	0.78	0.83	0.73
IPSL-CM5A-LR	166.6	48.2	8.8	0.4	access to plots	0.63	0.56	0.77	0.84	0.67
MIROC-ESM	131.7	13.3	6.2	0.2	access to plots	0.72	0.66	0.74	0.86	0.73
MPI-ESM-LR	169.9	51.5	7.4	0.3	access to plots	0.67	0.62	0.70	0.89	0.70
MRI-ESM1	236.1	117.7	12.5	0.2	access to plots	0.45	0.43	0.79	0.59	0.54
NorESM1-ME	130.4	12.0	6.5	0.5	access to plots	0.66	0.62	0.76	0.84	0.70

Notes: In calculating overall score, rmse score contributes double in comparison with all other scores.

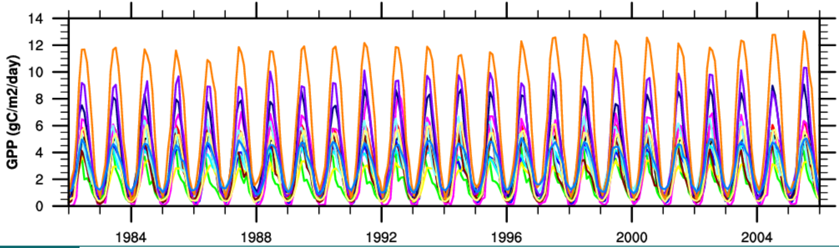
Annual Mean Global GPP



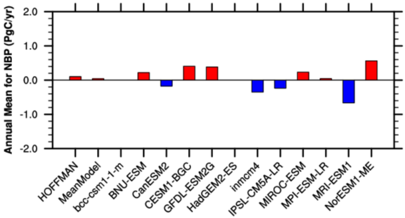
Seasonal Cycle of Regional GPP



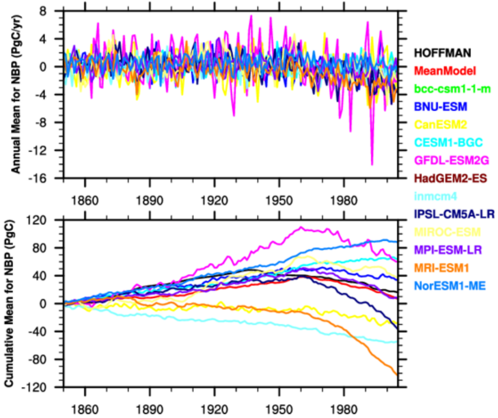
Model	Annual	Bias	RMSE
FLUXNET-MTE	2.36	-999.00	-999.00
MeanModel	2.99	0.63	0.74
bcc-csm1-1-m	1.82	-0.54	1.31
BNU-ESM	2.17	-0.19	0.62
CanESM2	1.76	-0.60	1.08
CESM1-BGC	2.45	0.08	0.78
GFDL-ESM2G	2.85	0.49	1.16
HadGEM2-ES	2.12	-0.24	0.72
inmcm4	3.06	0.70	1.20
IPSL-CM5A-LR	3.95	1.59	1.90
MIROC-ESM	2.48	0.12	0.35
MPI-ESM-LR	4.27	1.91	2.38
MRI-ESM1	6.13	3.76	4.46
NorESM1-ME	2.84	0.48	0.74



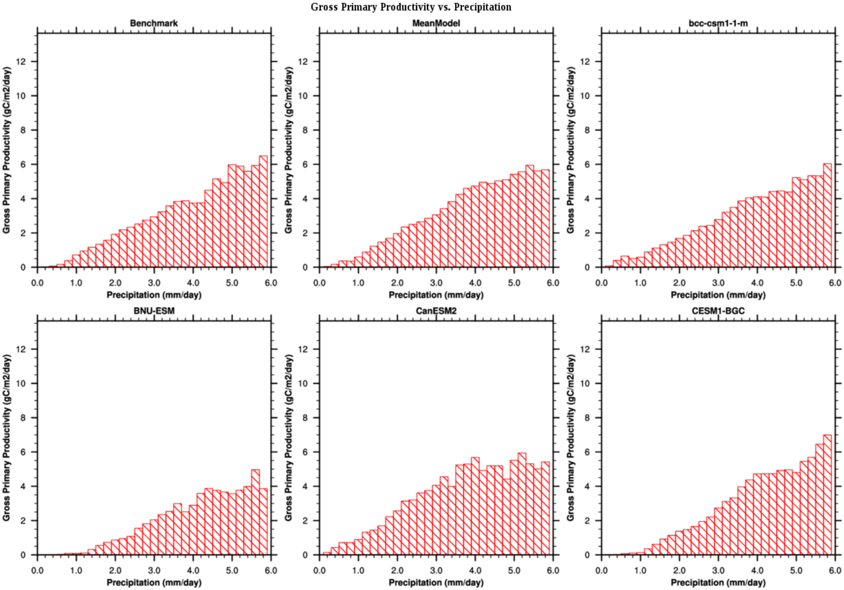
Global Net Ecosystem Carbon Balance



Long term carbon storage



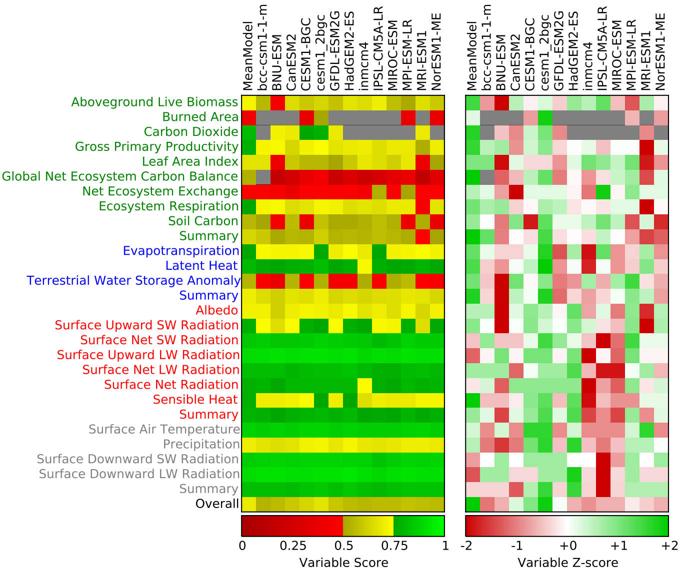
Functional Relationships: GPP vs. Precipitation



BGC Feedbacks

Argonne NATIONAL LABORATORY · Berkeley Lab · Los Alamos NATIONAL LABORATORY · NCAR · Oak Ridge National Laboratory · UCI · Michigan

ILAMB Model Scoring by Variable



ILAMB Next Generation Layout

Ecosystem and Carbon Cycle

	bcc-csm1-1	bcc-csm1-1-m	BNU-ESM	CanESM2	CCSM4	CESM1-BGC	GFDL-ESM2G	HadGEM2-CC	HadGEM2-ES	inmcm4	IPSL-CM5A-LR	IPSL-CM5A-MR	MIROC-ESM	MIROC-ESM-CHEM	MPI-ESM-LR	MRI-ESM1	NorESM1-M	NorESM1-ME
Biomass	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~
Burned Area	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~
Carbon Dioxide	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~
Gross Primary Productivity	0.53	0.57	0.52	0.47	0.52	0.52	0.52	0.51	0.51	0.05	0.50	0.52	0.55	0.55	0.55	0.45	0.54	0.54
Leaf Area Index	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~
Global Net Ecosystem Carbon Balance	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~
Net Ecosystem Exchange	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~
Ecosystem Respiration	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~
Soil Carbon	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~

Hydrology Cycle

	bcc-csm1-1	bcc-csm1-1-m	BNU-ESM	CanESM2	CCSM4	CESM1-BGC	GFDL-ESM2G	HadGEM2-CC	HadGEM2-ES	inmcm4	IPSL-CM5A-LR	IPSL-CM5A-MR	MIROC-ESM	MIROC-ESM-CHEM	MPI-ESM-LR	MRI-ESM1	NorESM1-M	NorESM1-ME
Evapotranspiration	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~
Latent Heat	0.39	0.39	0.43	0.36	0.44	0.44	0.41	0.42	0.42	0.40	0.44	0.42	0.43	0.43	0.40	0.41	0.45	0.45
Fluxnet-MTE (75.0%)	0.27	0.26	0.31	0.28	0.31	0.31	0.29	0.28	0.28	0.28	0.31	0.30	0.34	0.34	0.28	0.27	0.34	0.33
Fluxnet (25.0%)	0.77	0.76	0.78	0.60	0.83	0.83	0.78	0.86	0.85	0.77	0.83	0.78	0.71	0.71	0.76	0.82	0.79	0.78
Terrestrial Water Storage Anomaly	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~

Radiation and Energy Cycle

	bcc-csm1-1	bcc-csm1-1-m	BNU-ESM	CanESM2	CCSM4	CESM1-BGC	GFDL-ESM2G	HadGEM2-CC	HadGEM2-ES	inmcm4	IPSL-CM5A-LR	IPSL-CM5A-MR	MIROC-ESM	MIROC-ESM-CHEM	MPI-ESM-LR	MRI-ESM1	NorESM1-M	NorESM1-ME
Albedo	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~
Surface Upward SW Radiation	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~
Surface Net SW Radiation	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~
Surface Upward LW Radiation	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~
Surface Net LW Radiation	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~
Surface Net Radiation	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~
Sensible Heat	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~

Forcings

	bcc-csm1-1	bcc-csm1-1-m	BNU-ESM	CanESM2	CCSM4	CESM1-BGC	GFDL-ESM2G	HadGEM2-CC	HadGEM2-ES	inmcm4	IPSL-CM5A-LR	IPSL-CM5A-MR	MIROC-ESM	MIROC-ESM-CHEM	MPI-ESM-LR	MRI-ESM1	NorESM1-M	NorESM1-ME
Surface Air Temperature	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~
Precipitation	0.36	0.35	0.36	0.36	0.37	0.37	0.35	0.36	0.36	0.34	0.35	0.35	0.36	0.36	0.35	0.35	0.36	0.36
Surface Downward SW Radiation	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~
Surface Downward LW Radiation	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~

ILAMB Next Generation Layout

GrossPrimaryProductivity / Fluxnet–MTE / global – Google Chrome

GrossPrimaryProduc ×

Forrest M.

www.climatemodeling.org/~nate/ILAMB/EcosystemandCarbonCycle/GrossPrimaryProductivity/Fluxnet-MTE//Fluxnet-MTE.html

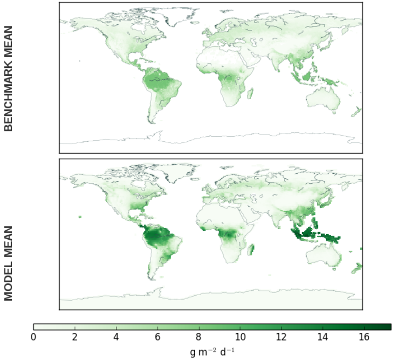
GrossPrimaryProductivity / Fluxnet-MTE / global

Single Model | All Models

global

Model	Data	Period Mean [Pg yr-1]	Bias [Pg yr-1]	RMSE [Pg yr-1]	Phase Shift [d]	Bias Score [-]	RMSE Score [-]	Phase Score [-]	Interannual Variability Score [-]	Spatial Distribution Score [-]	Overall Score [-]
Benchmark	[-]	118.895									
CLM40cn	[-]	134.761	20.969	0.31	22.039	0.402	0.178	0.775	0.333	0.81	0.81
CLM45bgc	[-]	119.838	5.809	0.27	21.843	0.435	0.152	0.788	0.355	0.902	0.902
CLM45bgc_GSWP3	[-]	111.456	-2.765	0.269	21.804	0.423	0.152	0.783	0.331	0.903	0.903

Temporally integrated period mean



Future ILAMB Development and Application

- Current ILAMB Prototype was applied to:
  - Model development of the Community Land Model (CLM)
  - CMIP5 Historical and esmHistorical simulations
  - ACME Land Model evaluation
- 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 (and past) projects where we hope to apply ILAMB:
  - CMIP6, including C^4MIP, LS3MIP, and LUMIP
  - TRENDY
  - MsTMIP, PLUME-MIP
- We will host a second ILAMB Workshop in the U.S. in the Washington, DC, area May 16–18, 2016

Predictive Carbon Cycle Science

- *Routine and systematic* confrontation of models with the growing body of observational data is critical to **identifying model weaknesses**.
- To the extent that models represent the embodiment of our scientific understanding, **model benchmarking helps identify knowledge gaps**.
- **New benchmarks are required** at different space and time scales:
  - Benchmarks for small-scale site-based process studies (e.g., FACE, LiDET, N & P addition and water exclusion experiments)
  - Benchmarks for disturbance and extreme events (e.g., wildfire, insect infestation, land use change)
  - Benchmarks for ecosystem responses on different scales (e.g., El Niño)
- Benchmark data sets could be used to initialize land and ocean carbon models, then consider using models for **ecological forecasting**.
  - To evaluate model fidelity for tropical drought, we are modeling ENSO for comparison with observations in NGEE Tropics.
  - Using NOAA CFS sea surface temperature (SST) predictions to drive the ACME model at $1/4^\circ$ for current ENSO.
- All of these things require **community engagement**!

Acknowledgements

U.S. DEPARTMENT OF ENERGY

Office of Science

NSF

This research was sponsored by 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, the U. S. Department of Agriculture Forest Service, and the National Science Foundation (AGS-1048890). This research used resources of the National Center for Computational Sciences (NCCS) 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.

References

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.