Diagnosing Climate–Carbon Cycle Feedbacks Constrained by ILAMB

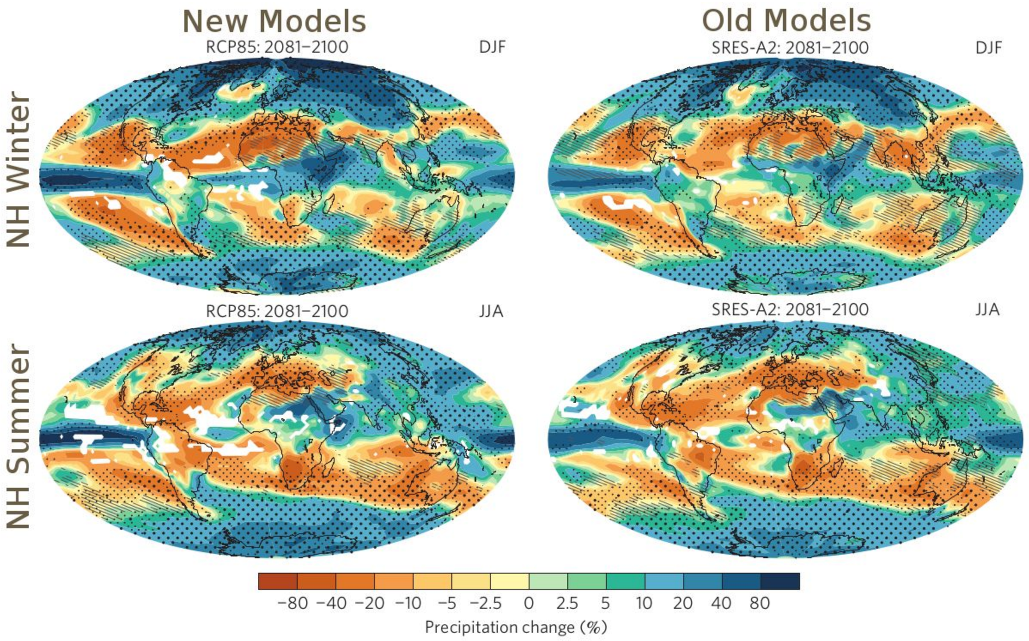
Forrest M. Hoffman[1,2], Nathan Collier[1], Mingquan Mu[3], Min Xu[1], Weiwei Fu[3], Cheng-En Yang[2,1], Gretchen Keppel-Aleks[4], David M. Lawrence[5], Charles D. Koven[6], William J. Riley[6], and James T. Randerson[3]

[1]Oak Ridge National Laboratory, Oak Ridge, TN, USA
[2]University of Tennessee, Knoxville, TN, USA
[3]University of California, Irvine, CA, USA
[4]University of Michigan, Ann Arbor, MI, USA
[5]National Center for Atmospheric Research, Boulder, CO, USA
[6]Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Problem: Model Uncertainty

Model uncertainty is one of the biggest challenges we face in Earth system science, yet comparatively little effort is devoted to fixing it (Carslaw et al., 2018)

- Model complexity is rapidly increasing as detailed process representations are added
- Evidence shows overall model uncertainty is reduced only slowly and sometimes increased (Knutti and Sedláček, 2013)
- Balance must be struck between model “elaboration” and efforts to reduce model uncertainty



Patterns of precipitation change across two generations of models (Adapted from Knutti and Sedláček, 2013)

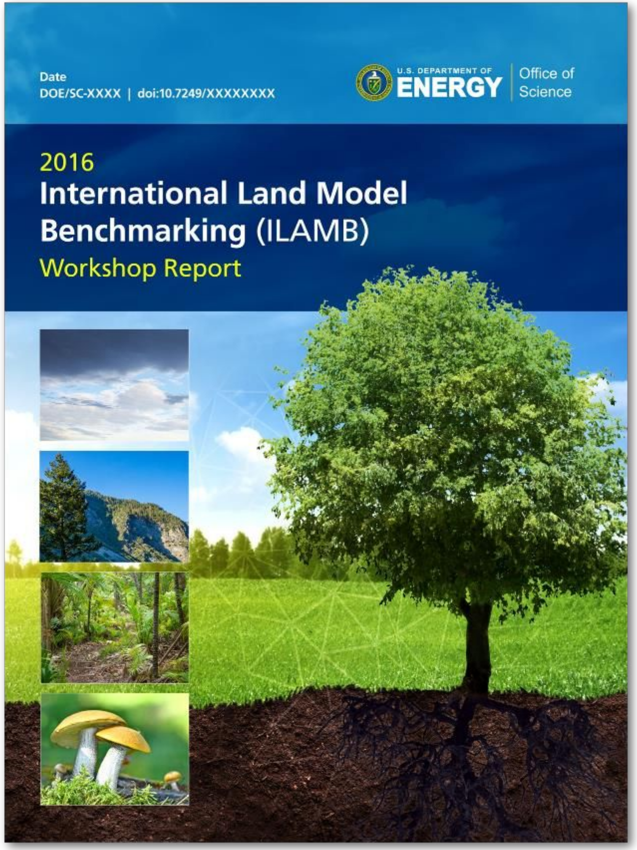
Why is Addressing Uncertainty a Challenge?

- Ecosystems have complex responses to a wide range of forcing factors in heterogeneous spatial environments, requiring highly multivariate approach
- Model uncertainty may increase, even as predictions of states and fluxes improves
- Rigorous confrontation of models with independent observations and hundreds of simulations are required to reduce uncertainty
- Modeling centers have a limited capacity to conduct sensitivity experiments, especially in fully coupled Earth system models, and rely primarily on homegrown methods and tools
- Focus is on adding complexity (e.g., more detailed representations of plant traits, photosynthesis, nutrient limitation, respiration)

2016 International Land Model Benchmarking (ILAMB) Workshop
May 16–18, 2016, Washington, DC

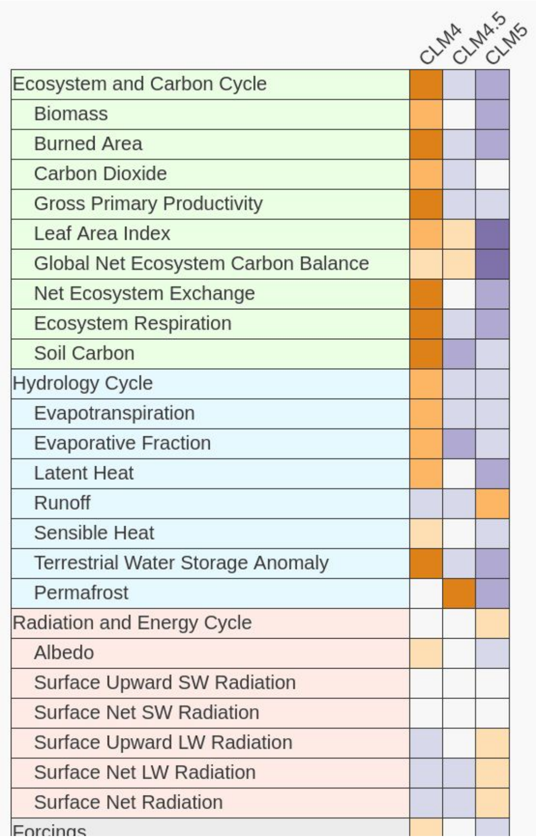
The **International Land Model Benchmarking (ILAMB)** community coordination activity was designed to

- Develop internationally accepted benchmarks
- Promote the use of these benchmarks
- Strengthen linkages between experimental, remote sensing, and modeling communities
- Support the design and development of open source benchmarking tools (Luo et al., 2012), like the **ILAMB Package** (Collier et al., 2018)



Argonne NATIONAL LABORATORY | BROOKHAVEN NATIONAL LABORATORY | BERKELEY LAB | Los Alamos NATIONAL LABORATORY | NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH | OAK RIDGE National Laboratory | UCI | MICHIGAN

ILAMB Assesses Land Model Fidelity

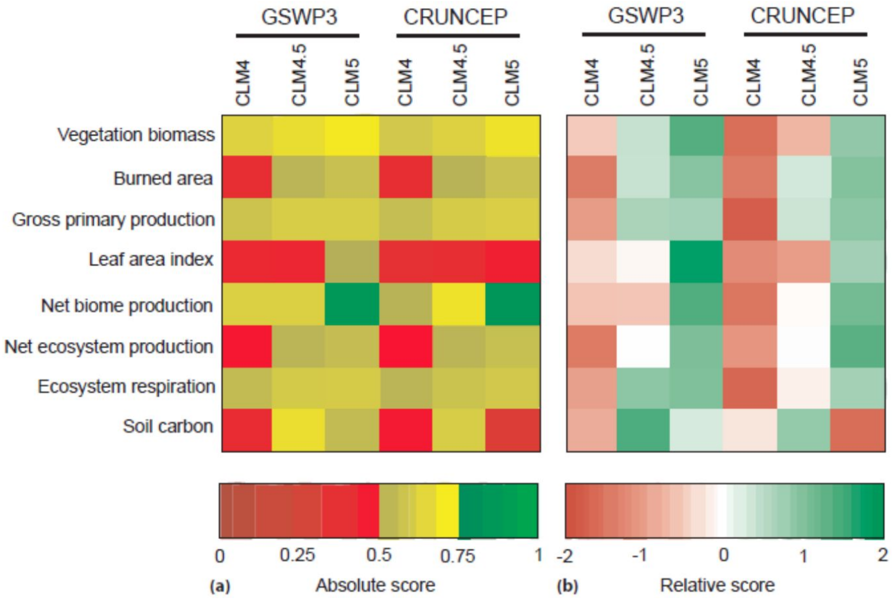


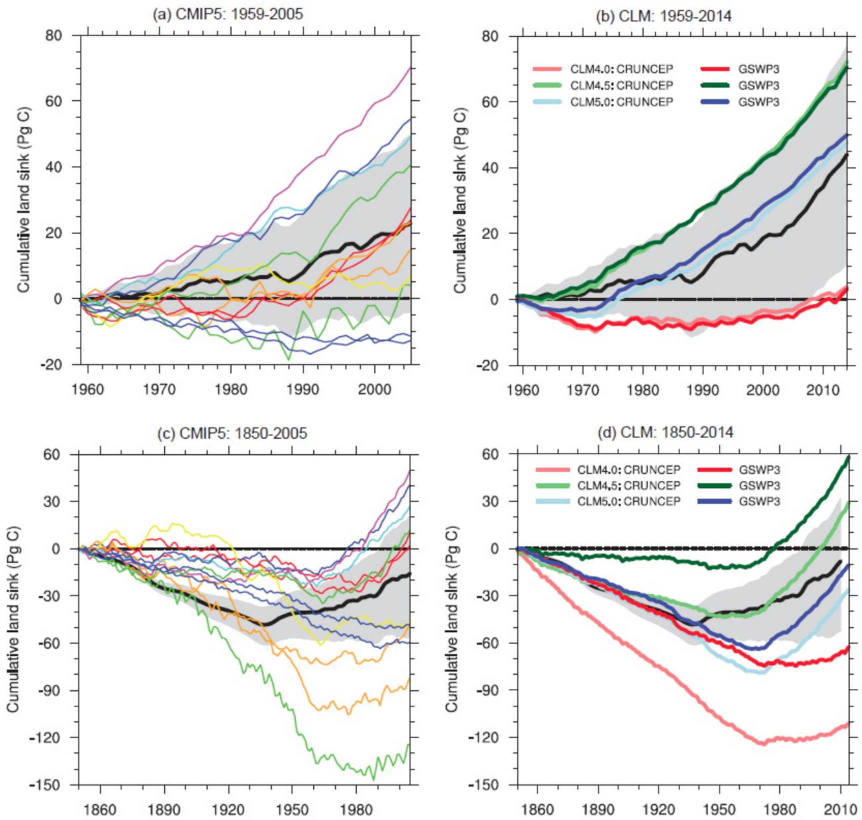
- Improvements in mechanistic treatment of hydrology, ecology, and land use with many more moving parts
- Simulation improved even with enhanced complexity
- Observational datasets not always self-consistent
- Forcing uncertainty confounds assessment of model development (not shown)

http://webext.cgd.ucar.edu/I20TR/_build_set1F/

(Lawrence et al., in press)

Land Model Performance Depends Strongly on Forcing





ILAMB performance for CLM4, CLM4.5, and CLM5 forced with GSWP3 vs. CRUNCEP (left) and the cumulative land carbon sink for CMIP5 vs. CLM offline models (right).

(Bonan et al., 2019)

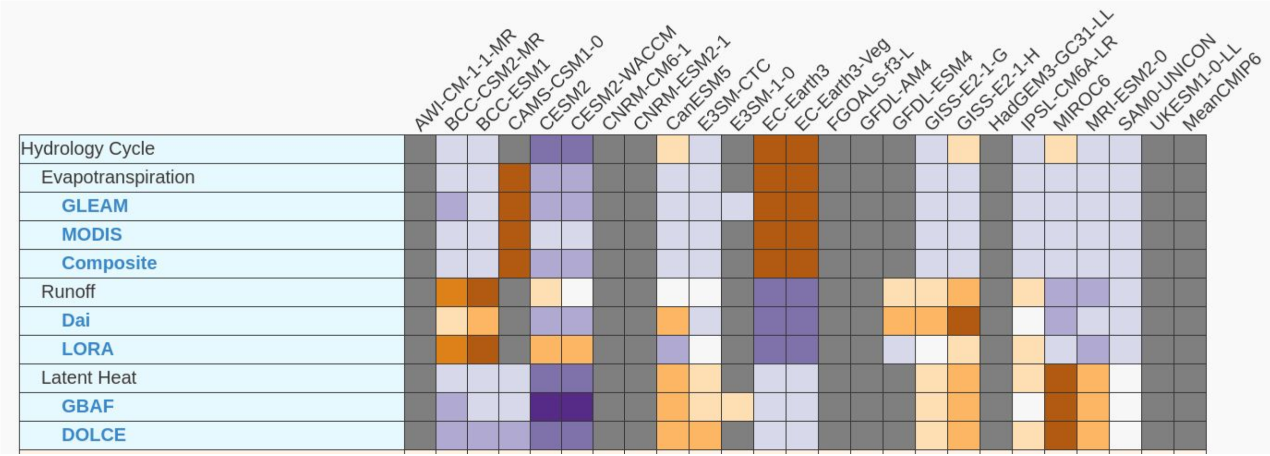
RUBISCO

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

RUBISCO

Addressing Observational Uncertainty

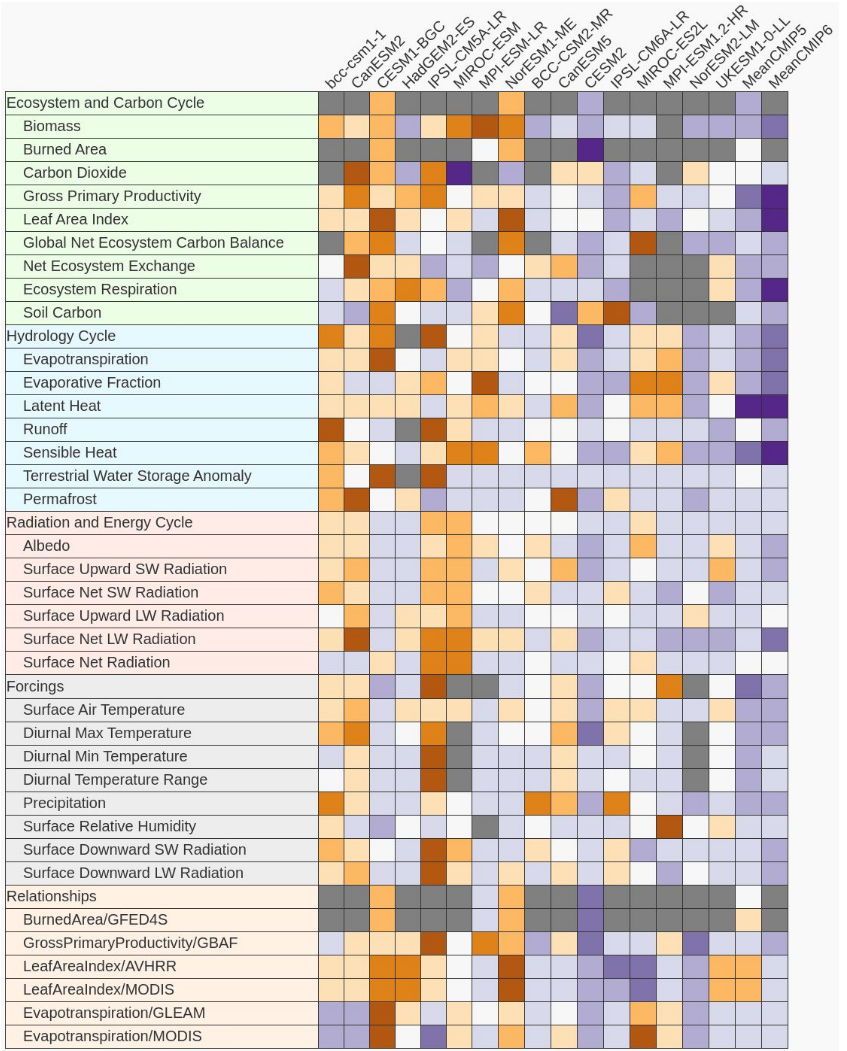
- Few observational datasets provide complete uncertainties
- ILAMB uses multiple datasets for most variables and allows users to weight them according to a rubric of uncertainty, scale mismatch, etc.
- ILAMB can also use:
  - Full spatial/temporal uncertainties provided with the data
  - Fixed, expert-derived uncertainty for a dataset
  - Uncertainties derived from combining multiple datasets



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

CMIP5 vs. CMIP6 Models

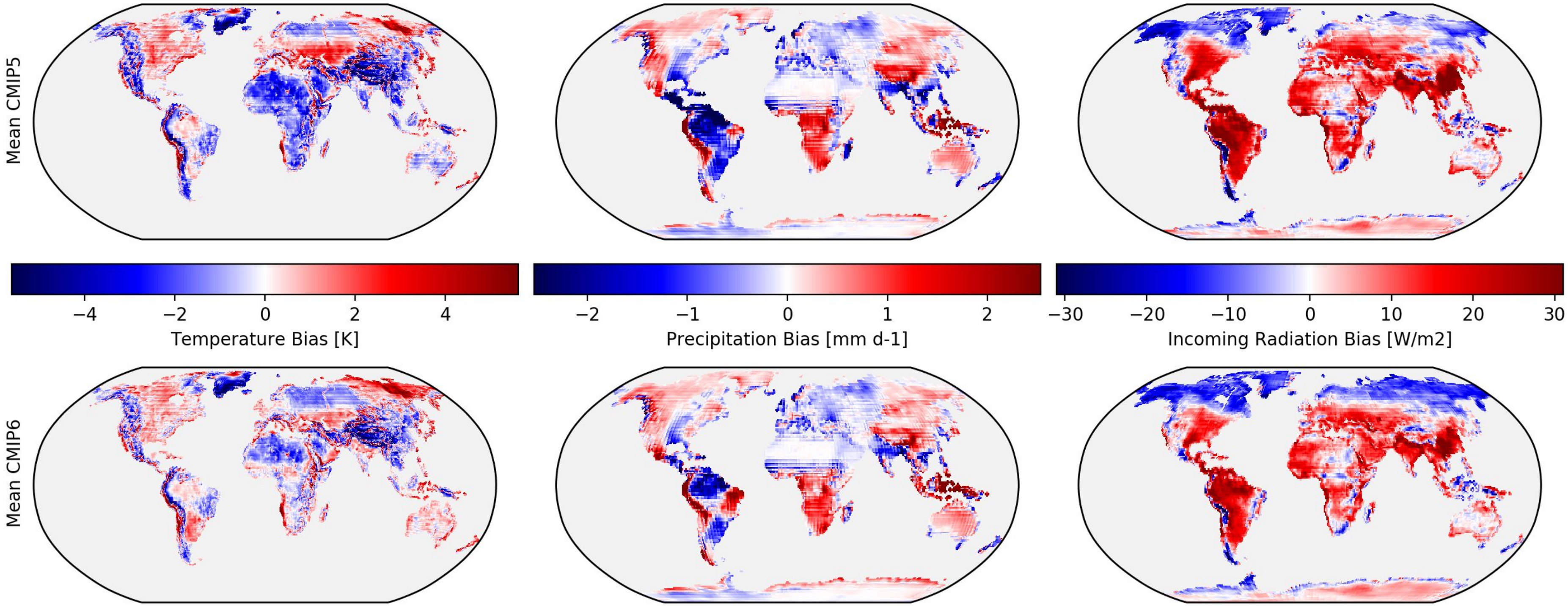
- The CMIP6 suite of land models (right) has improved over the CMIP5 suite of land models (left)
- The multi-model mean outperforms any single model for each suite of models
- The multi-model mean CMIP6 land model is the “best model” overall
- Why did CMIP6 land models improve?



(Hoffman et al., in prep)

Reasons for Land Model Improvements

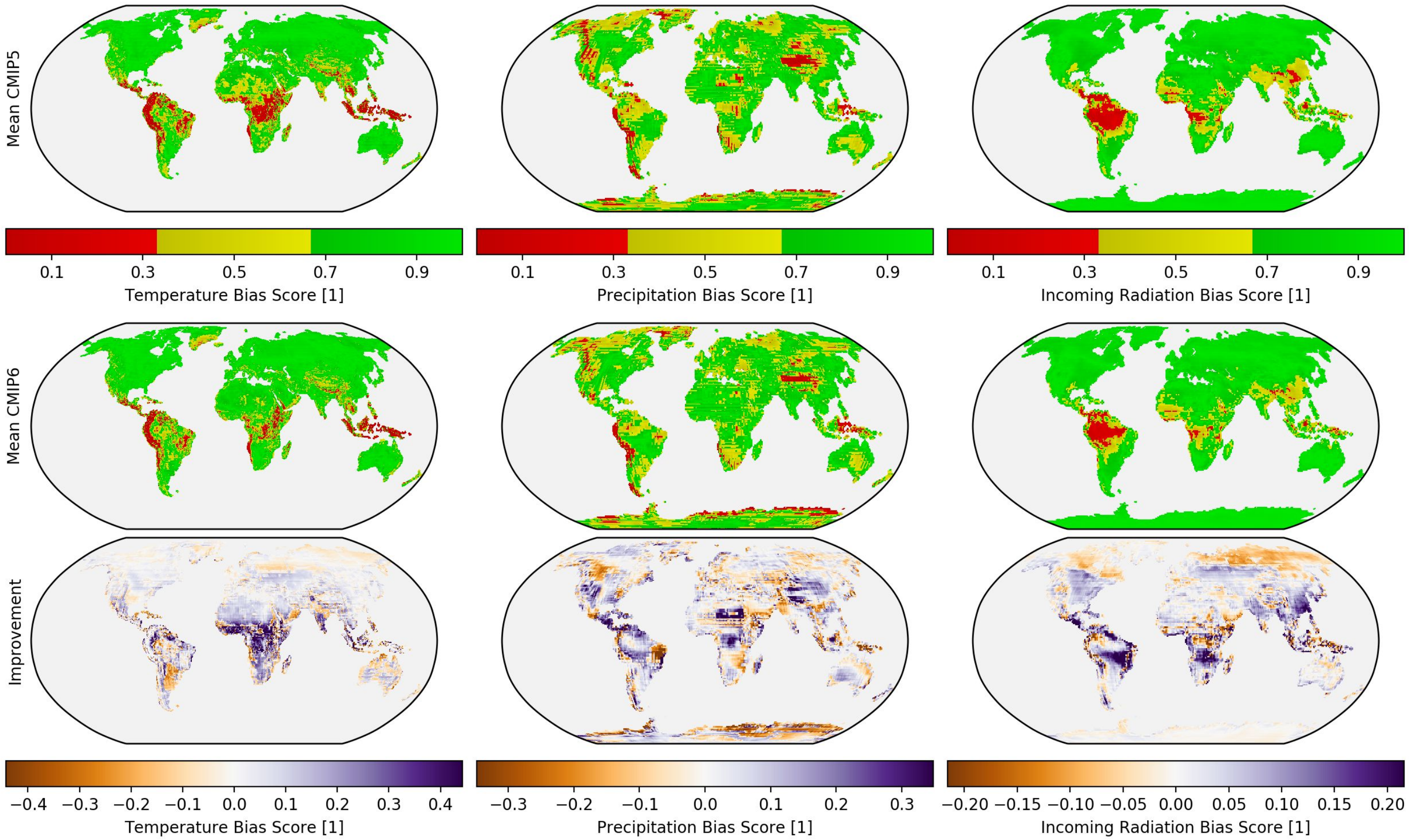
ESM improvements in climate forcings (temperature, precipitation, radiation) likely partially drove improvements exhibited by land carbon cycle models



(Hoffman et al., in prep)

Reasons for Land Model Improvements

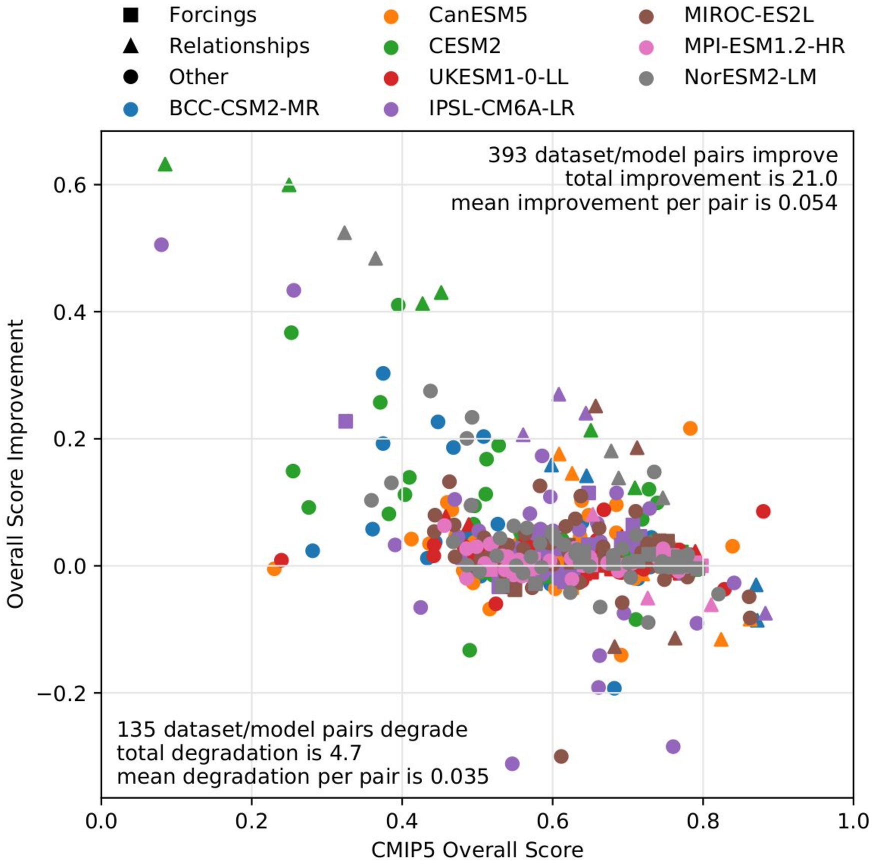
Differences in bias scores for temperature, precipitation, and incoming radiation were primarily positive, further indicating more realistic climate representation



(Hoffman et al., in prep)

Reasons for Land Model Improvements

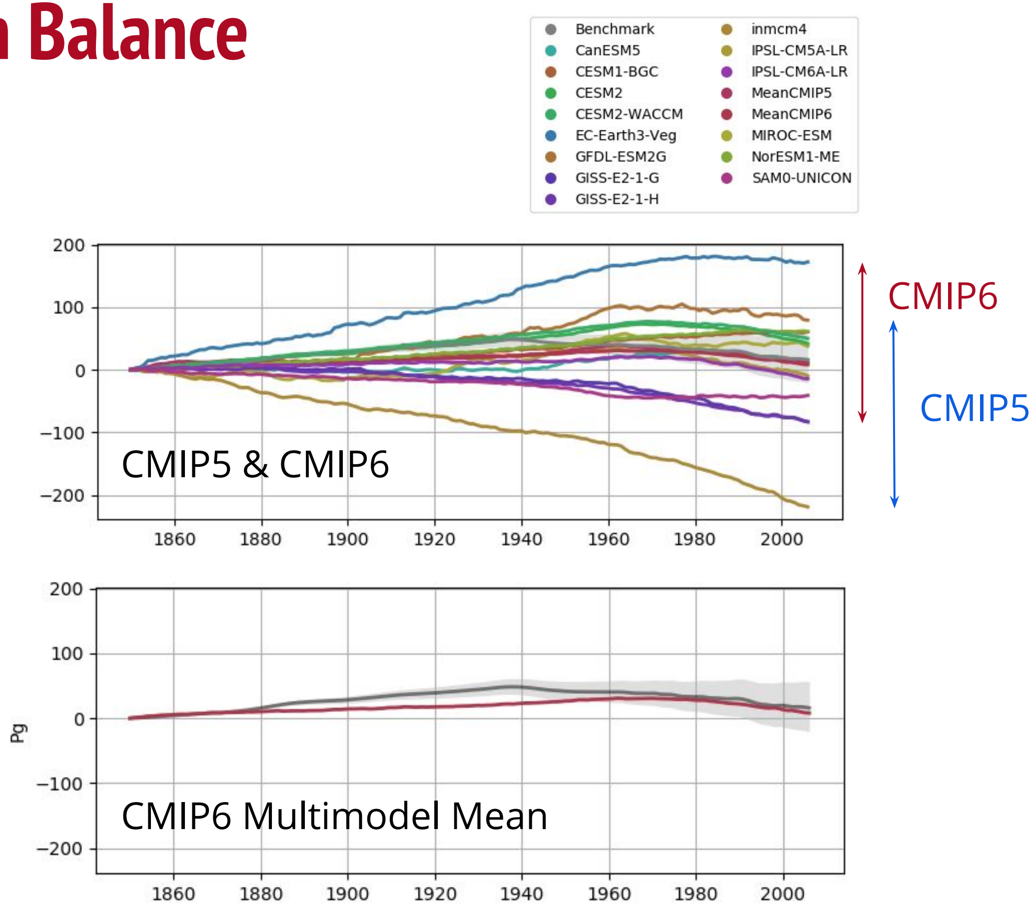
While forcings got better, the largest improvements were in **variable-to-variable relationships**, suggesting that increased land model complexity was also partially responsible for higher CMIP6 model scores



Net Ecosystem Carbon Balance

Initial examination of the range of contemporary accumulated land net carbon loss indicates it has decreased only slightly (or possibly increased?)

Model improvements in mean states and fluxes may not result in reduced uncertainty



RUBISCO

Conclusions

- ILAMB has proven useful for verification during model development and for validation in support of multi-model studies
- Land model performance depends strongly on imposed climate forcing
- CMIP6 land models performed better than CMIP5 land models due to
    - Improved climate forcing
    - Increased land model complexity
- Variable-to-variable relationships exhibited the largest improvements for some models
- Model improvements in mean states and fluxes may not result in reduced uncertainty

Future Science Questions

- Upon further examination, will improved multi-model performance result in reduced spread in feedback sensitivities, projected land carbon storage, and future climate change?
- Can we use ILAMB scores to weight contributions to multi-model means and thereby reduce contemporary biases, reduce future projected uncertainties, and alter expected mitigation targets?

Extra Slides

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

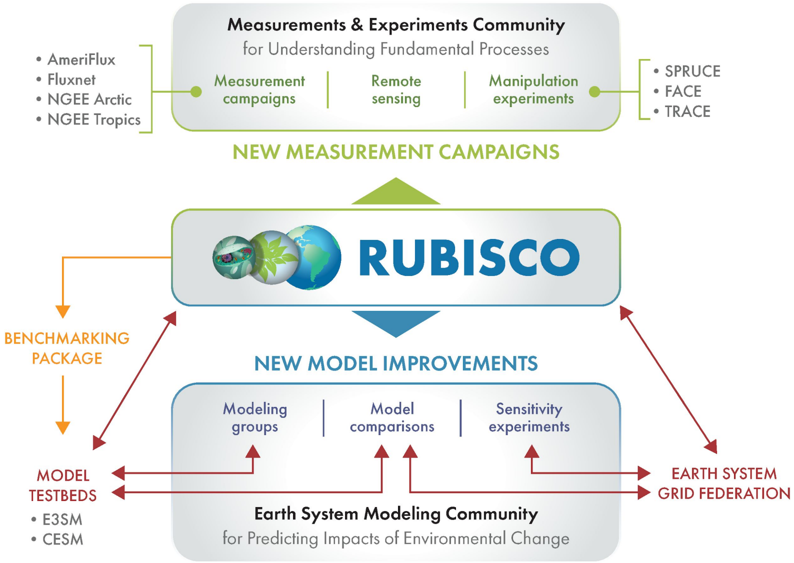
Forrest M. Hoffman (Laboratory Research Manager), William J. Riley (Senior Science Co-Lead), and James T. Randerson (Chief Scientist)

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

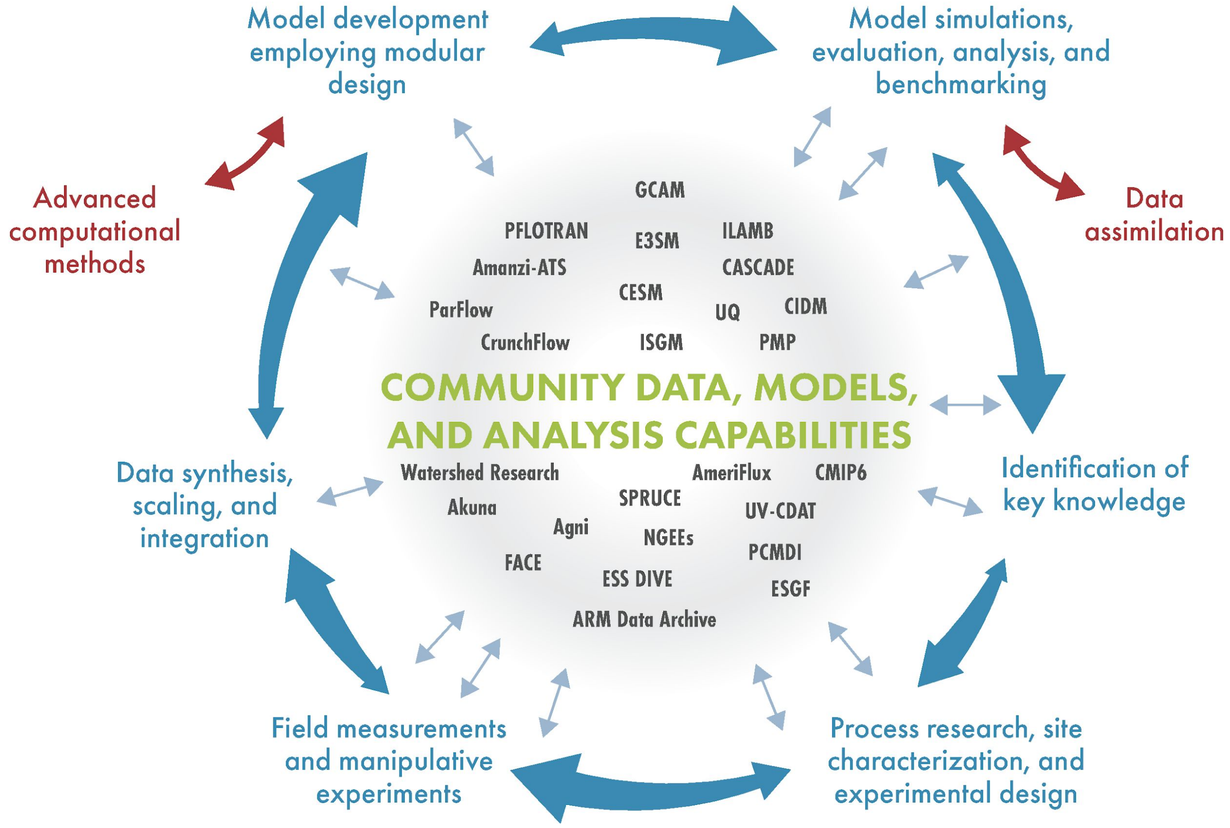
- Perform hypothesis-driven analysis of biogeochemical & hydrological processes and feedbacks in ESMs
- Synthesize in situ and remote sensing data and design metrics for assessing ESM performance
- Design, develop, and release the International Land Model Benchmarking (ILAMB) and International Ocean 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.

RUBISCO

DOE's Model-Data-Experiment Enterprise

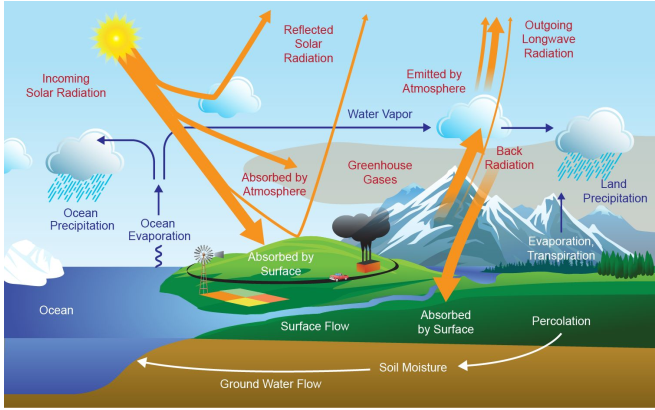


Argonne NATIONAL LABORATORY · BROOKHAVEN NATIONAL LABORATORY · BERKELEY LAB · Los Alamos NATIONAL LABORATORY · NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH · OAK RIDGE National Laboratory · UCI · MICHIGAN

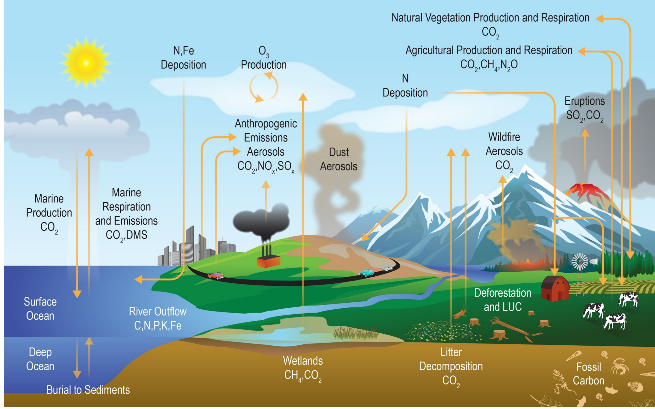
What is ILAMB?

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 Earth system 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

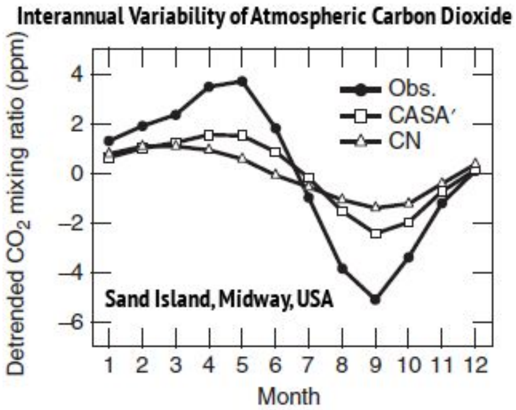


Carbon and Biogeochemical Cycles

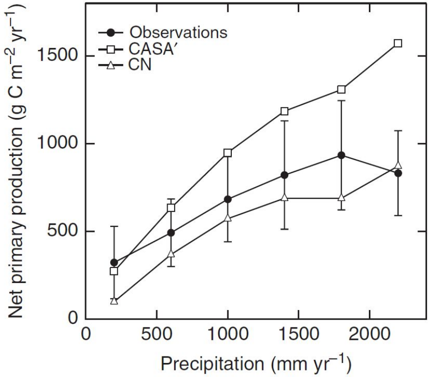
Argonne NATIONAL LABORATORY · BROOKHAVEN NATIONAL LABORATORY · BERKELEY LAB · Los Alamos NATIONAL LABORATORY · NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH · OAK RIDGE National Laboratory · UCI · MICHIGAN

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 a benchmark **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 at multiple 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

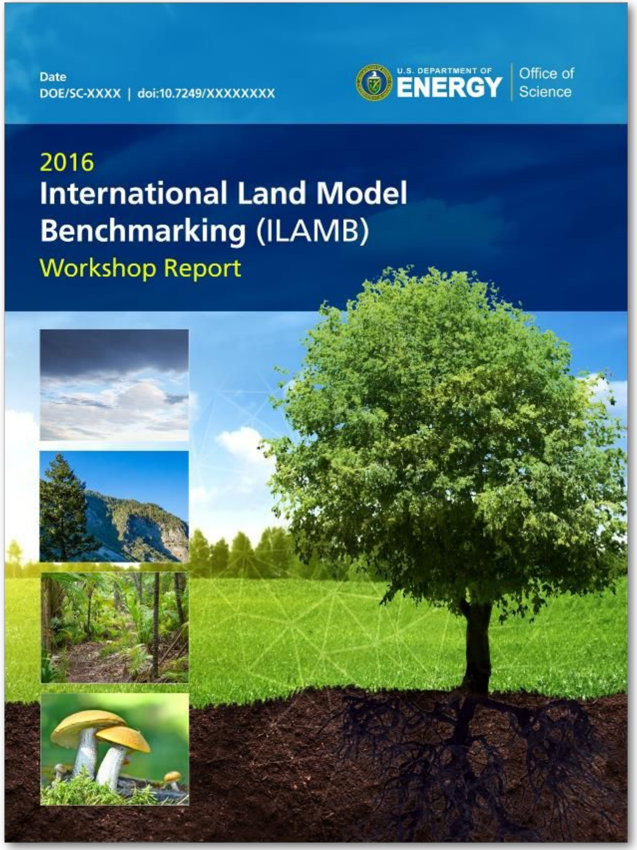
(Randerson et al., 2009)

- **First ILAMB Workshop** was held in Exeter, UK, on June 22–24, 2009
- **Second ILAMB Workshop** was held in Irvine, CA, USA, on January 24–26, 2011
  - ~45 researchers participated from the US, Canada, UK, Netherlands, France, Germany, Switzerland, China, Japan, and Australia
  - Developed methodology for model-data comparison and baseline standard for performance of land model process representations (Luo et al., 2012)

2016 International Land Model Benchmarking (ILAMB) Workshop
May 16–18, 2016, Washington, DC

Third ILAMB Workshop was held May 16–18, 2016

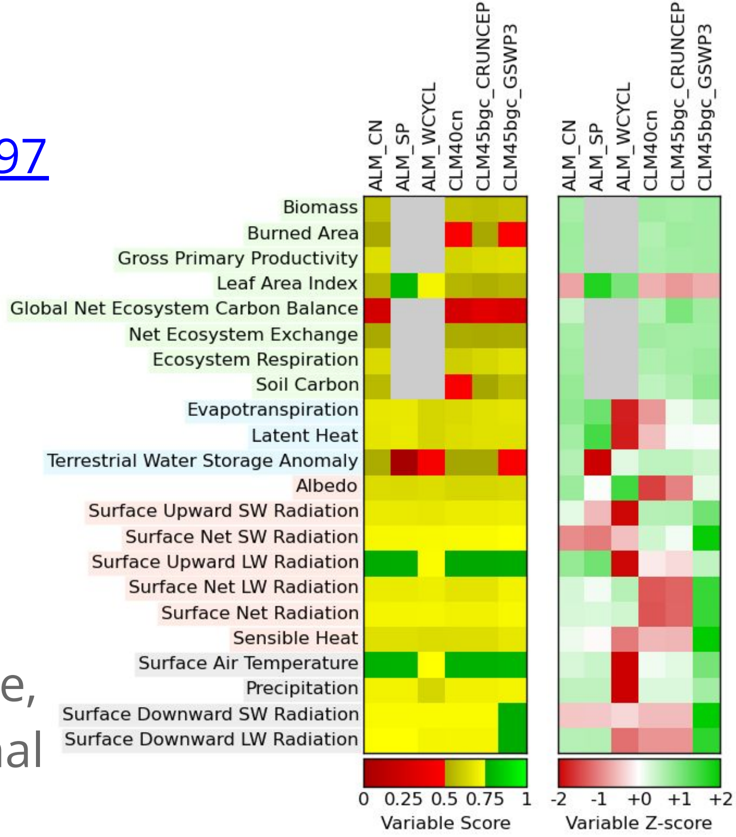
- Workshop Goals
  - Design of new metrics for model benchmarking
  - Model Intercomparison Project (MIP) evaluation needs
  - Model development, testbeds, and workflow processes
  - Observational data sets and needed measurements
- Workshop Attendance
  - 60+ participants from Australia, Japan, China, Germany, Sweden, Netherlands, UK, and US (10 modeling centers)
  - ~25 remote attendees at any time



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

Development of ILAMB Packages

- **ILAMBv1** released at **2015 AGU Fall Meeting** Town Hall, doi:10.18139/ILAMB.v001.00/1251597
- **ILAMBv2** released at **2016 ILAMB Workshop**, doi:10.18139/ILAMB.v002.00/1251621
- Open Source software freely distributed
- Routinely used for E3SM and CESM evaluation during development
- Employed to evaluate CMIP5 models
- Models are scored based on statistical comparisons (bias, RMS error, phase, amplitude, spatial distribution, Taylor scores) and functional response metrics

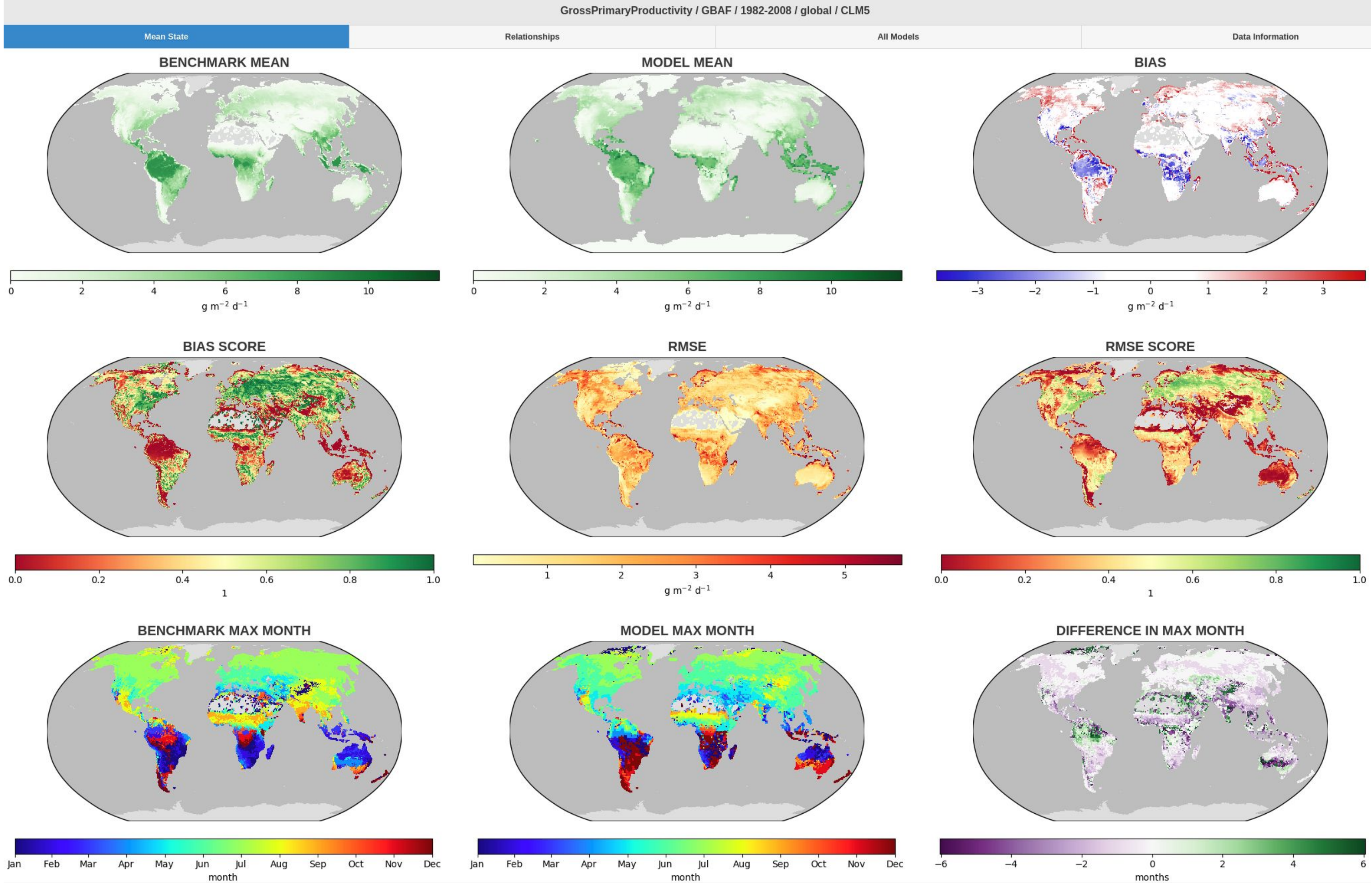


RUBISCO

ILAMBv2.5 Package Current Variables

- **Biogeochemistry:** Biomass (Contiguous US, Pan Tropical Forest), Burned area (GFED3), CO_2 (NOAA GMD, Mauna Loa), Gross primary production (Fluxnet, GBAF), Leaf area index (AVHRR, MODIS), Global net ecosystem carbon balance (GCP, Khatiwala/Hoffman), Net ecosystem exchange (Fluxnet, GBAF), Ecosystem Respiration (Fluxnet, GBAF), Soil C (HWSD, NCSCDv22, Koven)
- **Hydrology:** Evapotranspiration (GLEAM, MODIS), Evaporative fraction (GBAF), Latent heat (Fluxnet, GBAF, DOLCE), Runoff (Dai, LORA), Sensible heat (Fluxnet, GBAF), Terrestrial water storage anomaly (GRACE), Permafrost (NSIDC)
- **Energy:** Albedo (CERES, GEWEX.SRB), Surface upward and net SW/LW radiation (CERES, GEWEX.SRB, WRMC.BSRN), Surface net radiation (CERES, Fluxnet, GEWEX.SRB, WRMC.BSRN)
- **Forcing:** Surface air temperature (CRU, Fluxnet), Diurnal max/min/range temperature (CRU), Precipitation (CMAP, Fluxnet, GPCC, GPCP2), Surface relative humidity (ERA), Surface down SW/LW radiation (CERES, Fluxnet, GEWEX.SRB, WRMC.BSRN)

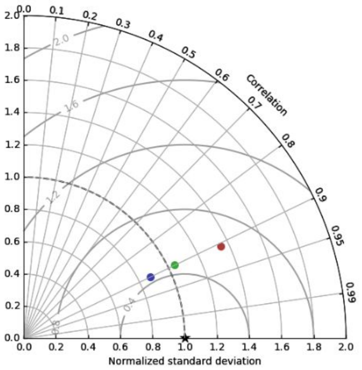
ILAMB Graphical Diagnostics



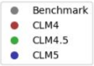
ILAMB Graphical Diagnostics

RUBISCO

SPATIAL TAYLOR DIAGRAM

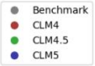


MODEL COLORS

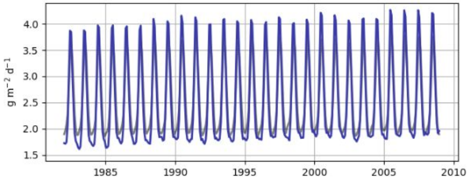


Spatially integrated regional mean

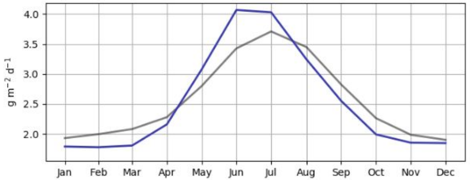
MODEL COLORS



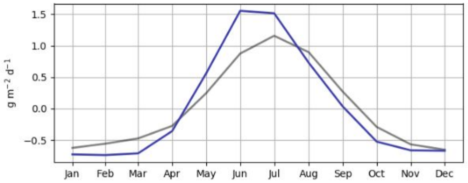
REGIONAL MEAN



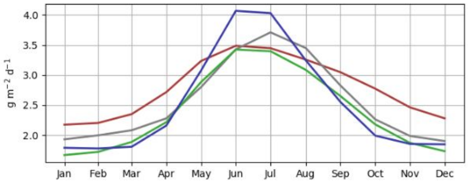
ANNUAL CYCLE



MONTHLY ANOMALY

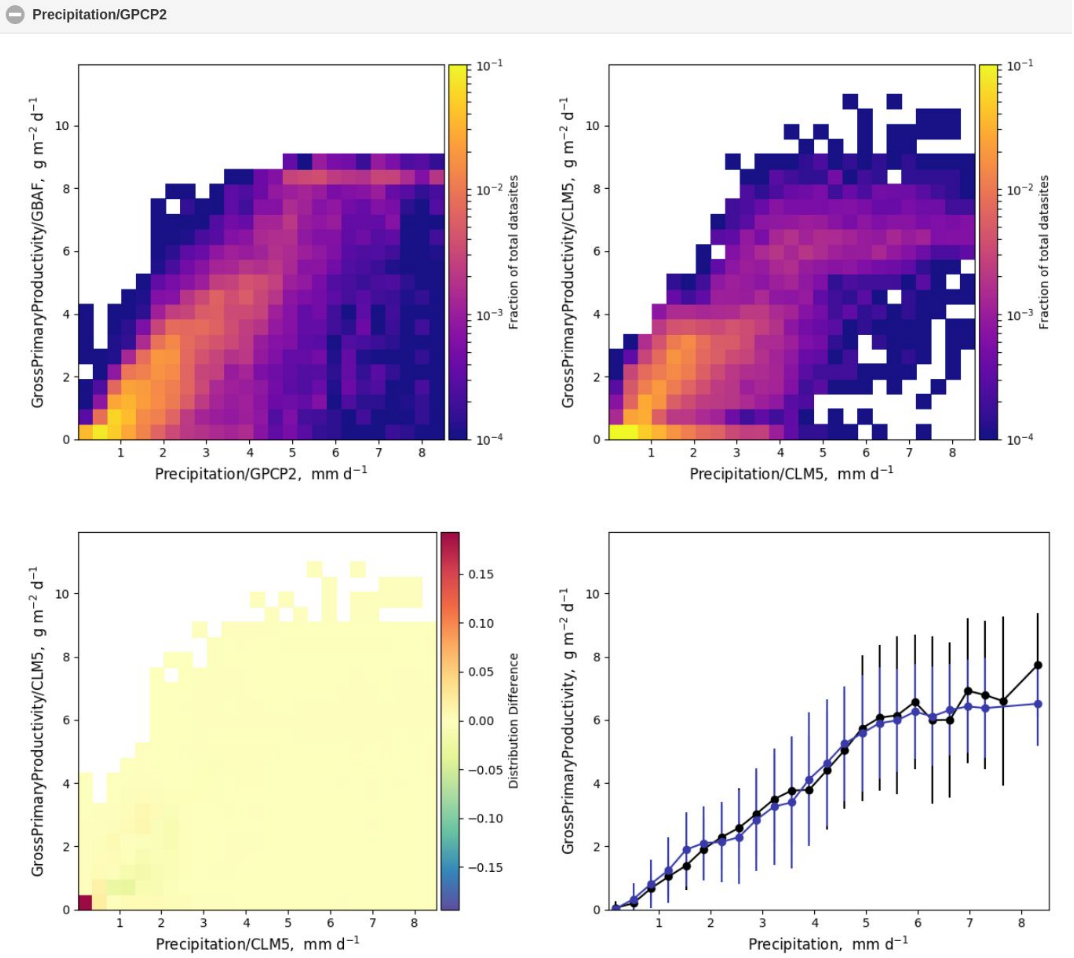


ANNUAL CYCLE



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

Variable-to-Variable Comparisons

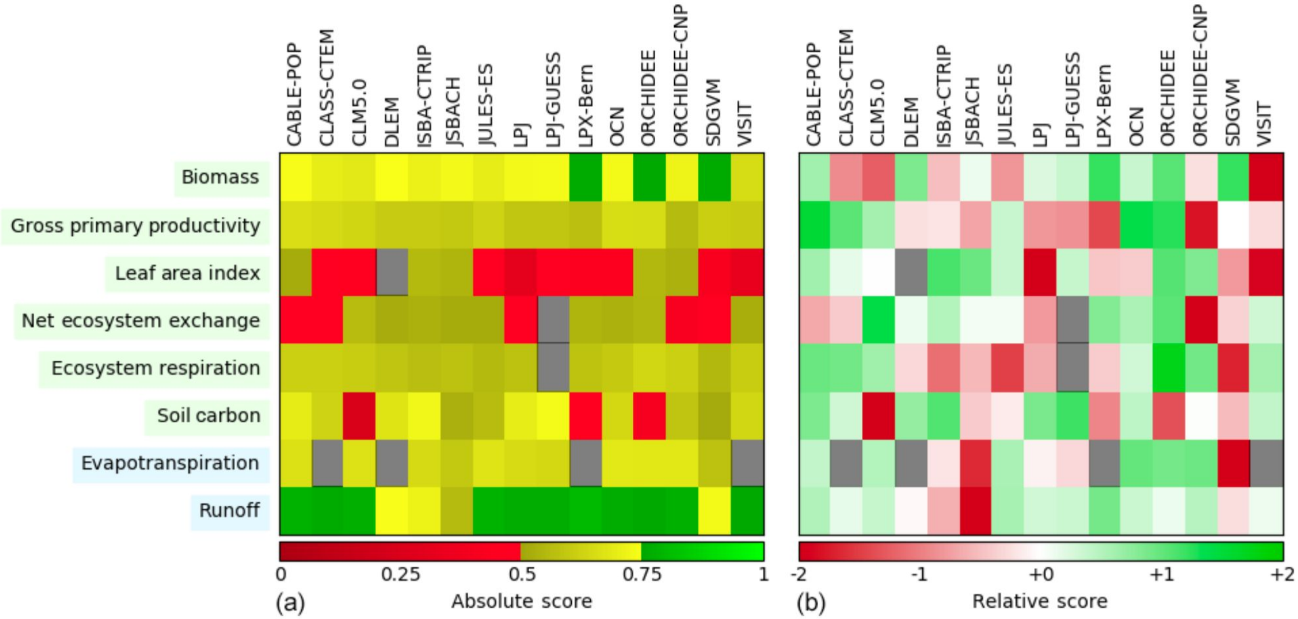


RUBISCO

Argonne NATIONAL LABORATORY | BROOKHAVEN NATIONAL LABORATORY | BERKELEY LAB | Los Alamos NATIONAL LABORATORY | NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH | OAK RIDGE National Laboratory | UCI | MICHIGAN

Global Carbon Budget 2019 - TRENDY Models

Evaluation of the DGVMs using the International Land Model Benchmarking system (ILAMB; Collier et al., 2018) (a) absolute skill scores and (b) skill scores relative to other models for a subset of ILAMB variables.



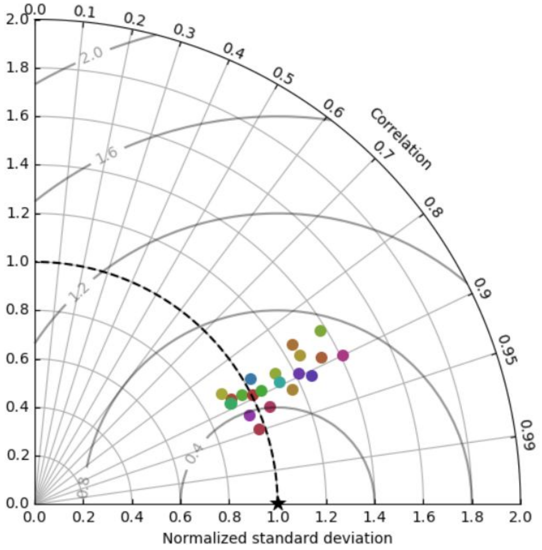
Friedlingstein et al. (2019)

Gross Primary Productivity

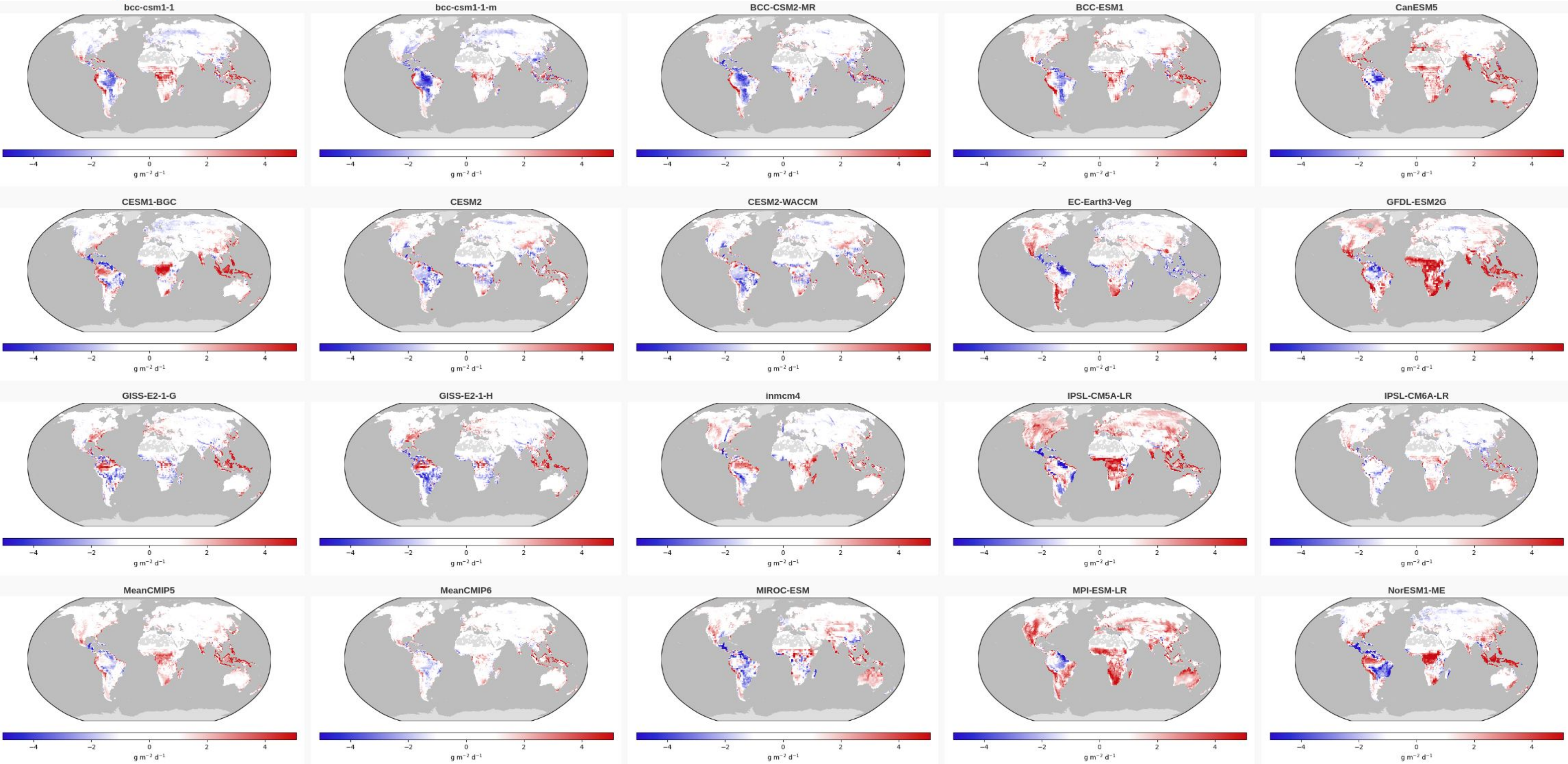
	Download Data	Period Mean (original grids) [Pg yr-1]	Model Period Mean (intersection) [Pg yr-1]	Model Period Mean (complement) [Pg yr-1]	Benchmark Period Mean (intersection) [Pg yr-1]	Benchmark Period Mean (complement) [Pg yr-1]	Bias [g m-2 d-1]	RMSE [g m-2 d-1]	Phase Shift [months]	Bias Score [1]	RMSE Score [1]	Seasonal Cycle Score [1]	Spatial Distribution Score [1]	Overall Score [1]
Benchmark	[-]	118.												
bcc-csm1-1	[-]	123.	114.	6.80	118.	0.0600	0.203	1.94	1.27	0.424	0.267	0.809	0.946	0.543
bcc-csm1-1-m	[-]	112.	108.	4.10	118.	0.501	-0.116	1.94	1.38	0.413	0.265	0.794	0.934	0.534
BCC-CSM2-MR	[-]	123.	115.	8.31	118.	0.501	-0.0721	1.68	1.28	0.433	0.326	0.796	0.941	0.564
BCC-ESM1	[-]	157.	133.	21.4	118.	0.0640	0.325	1.84	1.23	0.429	0.302	0.808	0.945	0.557
CanESM5	[-]	141.	131.	8.05	118.		0.675	1.85	1.70	0.427	0.330	0.701	0.934	0.544
CESM1-BGC	[-]	129.	124.	4.32	118.	0.501	0.309	1.74	1.38	0.392	0.350	0.761	0.873	0.545
CESM2	[-]	110.	105.	4.21	118.	0.473	-0.0938	1.72	1.52	0.411	0.364	0.786	0.935	0.572
CESM2-WACCM	[-]	110.	106.	4.28	118.	0.473	-0.0889	1.73	1.50	0.410	0.364	0.788	0.936	0.572
EC-Earth3-Veg	[-]	136.	134.	2.52	118.		0.330	1.99	1.49	0.417	0.312	0.755	0.931	0.545
GFDL-ESM2G	[-]	167.	155.	9.78	118.		1.19	3.18	1.45	0.360	0.185	0.726	0.880	0.467
GISS-E2-1-G	[-]	133.	118.	12.6	117.	1.29	0.0302	1.55	1.23	0.411	0.355	0.741	0.905	0.553
GISS-E2-1-H	[-]	131.	116.	13.8	118.	0.654	-0.0269	1.57	1.19	0.400	0.353	0.760	0.913	0.556
inmcm4	[-]	136.	128.	8.25	113.	5.44	0.351	1.78	1.41	0.451	0.308	0.766	0.935	0.554
IPSL-CM5A-LR	[-]	165.	153.	9.00	118.	0.347	1.10	2.73	1.30	0.318	0.241	0.770	0.889	0.492
IPSL-CM6A-LR	[-]	116.	111.	4.25	118.	0.486	0.0566	1.45	1.32	0.498	0.364	0.751	0.960	0.587
MeanCMIP5	[-]	138.	131.	6.75	118.		0.561	1.44	1.13	0.462	0.408	0.794	0.959	0.606
MeanCMIP6	[-]	121.	116.	5.10	118.		0.159	1.10	1.12	0.522	0.470	0.796	0.973	0.646
MIROC-ESM	[-]	129.	121.	6.01	108.	10.1	0.308	2.06	1.40	0.425	0.322	0.749	0.918	0.547
MPI-ESM-LR	[-]	170.	162.	6.90	110.	8.62	1.22	2.37	1.43	0.378	0.291	0.699	0.926	0.517
NorESM1-ME	[-]	129.	121.	6.29	118.		0.331	1.92	1.46	0.354	0.350	0.759	0.838	0.530
SAM0-UNICON	[-]	131.	126.	4.95	118.	0.501	0.371	1.75	1.39	0.398	0.338	0.764	0.845	0.537

- Multimodel GPP is compared with global seasonal GBAF estimates
- We can see Improvements across generations of models (e.g., CESM1 vs. CESM2, IPSL-CM5A vs. 6A)
- The mean CMIP6 and CMIP5 models perform best

Spatial Taylor Diagram



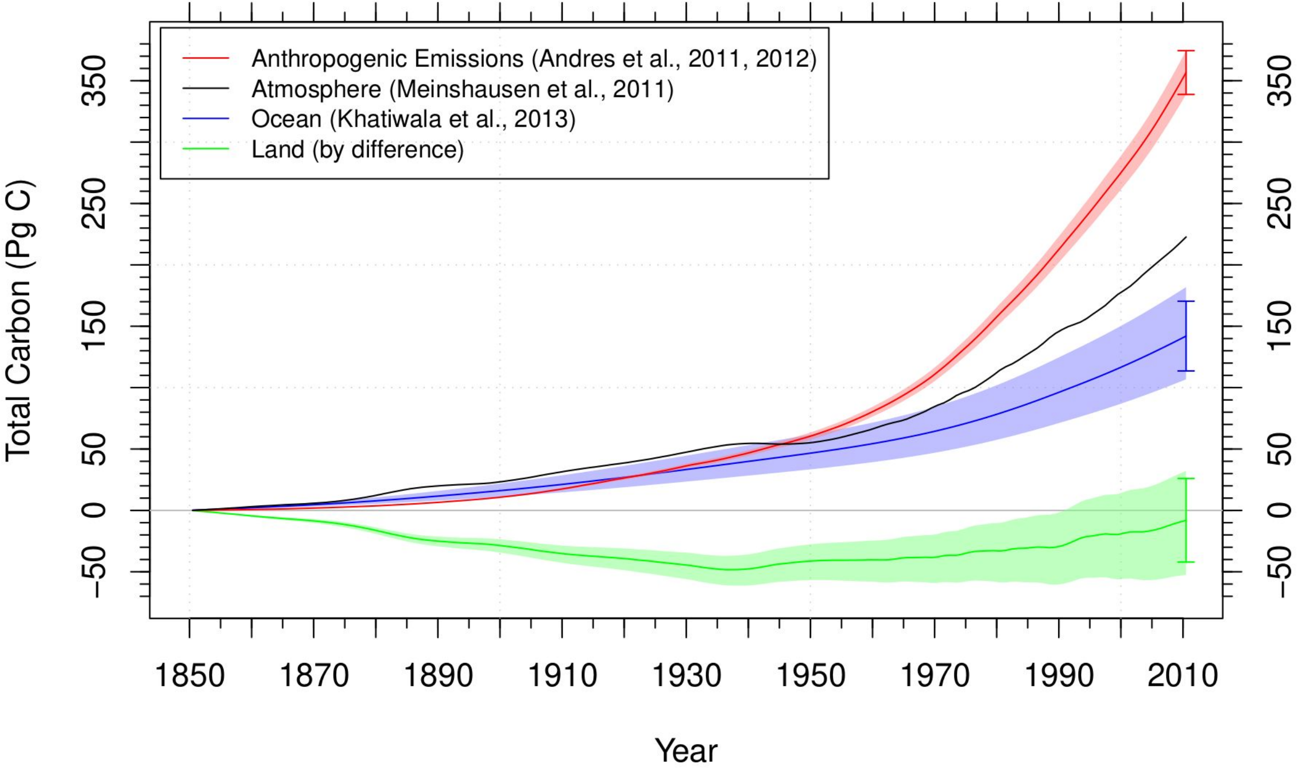
Argonne NATIONAL LABORATORY
BROOKHAVEN NATIONAL LABORATORY
BERKELEY LAB
Los Alamos NATIONAL LABORATORY
NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH
OAK RIDGE National Laboratory
UCI
MICHIGAN



Argonne NATIONAL LABORATORY
BROOKHAVEN NATIONAL LABORATORY
BERKELEY LAB
Los Alamos NATIONAL LABORATORY
NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH
OAK RIDGE National Laboratory
UCI
MICHIGAN

Observed Carbon Accumulation Since 1850

We used fossil fuel emissions estimates, atmospheric CO_2 measurements, and ocean carbon accumulation estimates from Khatiwala et al. (2013) to estimate land carbon accumulation with propagated uncertainties from 1850 to 2010.



RUBISCO

Soil Carbon Dynamics Working Group

U.S. DEPARTMENT OF ENERGY

- Formed after community recommendation from the 2016 International Land Model Benchmarking (ILAMB) Workshop Report
- Objective is to apply data and models to improve predictive understanding
- June and September conference calls led to meeting at ORNL in October

Data to Knowledge
Synthesize existing data from collaborative networks, archives, and publications

+

Knowledge to Data
Perform simulations to test hypotheses and characterize model structural uncertainties

→

Predictive Understanding
Design functional relationship metrics to confront models and apply data-driven approaches to model formulation

Global Data Synthesis Theme

- Combine field observations from collaborative sampling networks and databases, including International Soil Carbon Network (ISCN) and published literature
- Quantify vertical distribution of SOM and responses to controlling mechanisms

Model–Data Integration Theme

- Develop consistent datasets for initializing, forcing, and benchmarking microbially explicit soil carbon models
- Characterize model structural uncertainty through software frameworks to understand controlling mechanisms

For more information, contact Forrest M. Hoffman <forrest@climatemodeling.org> or Umakant Mishra <umishra@anl.gov>

Argonne NATIONAL LABORATORY | BROOKHAVEN NATIONAL LABORATORY | BERKELEY LAB | Los Alamos NATIONAL LABORATORY | NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH | OAK RIDGE National Laboratory | UCI | MICHIGAN

RUBISCO-AmeriFlux Working Group

- Formed after community recommendation from the 2016 International Land Model Benchmarking (ILAMB) Workshop Report
- Several conference calls have occurred, at least one more is scheduled, and **meeting scheduled for mid October**
- More than 40 scientists have registered to attend

- ***Multifactor*** ecosystem responses to climate change, extreme events, and changes in seasonality using e.g., Ameriflux, phenocam observations, remote sensing products, observations from citizen science programs, and others.
- Roles of ***extreme events*** and "return times" on ecosystem resilience.
- ***Long-term*** trends in light use efficiency, water use efficiency, evapotranspiration, and other quantities, some of which may yield new emergent constraints
- ***Advanced mathematical analyses*** of time series of ecosystem dynamics to infer underlying controls across temporal scales.
- Synthesizing ***new observations*** from data sets across spatial and temporal scales (e.g., AmeriFlux, remote sensing, disturbance maps, SIF, etc.)
- ***"Super site" benchmarks*** developed around stable, long-running flux tower sites with a diversity of collocated measurements (e.g., AmeriFlux, CZOs, LTER, NEON)
- ***Spatial scaling methods*** to interpret point measurements, incorporating ancillary databases, to study areas, regions, continents, and the globe.

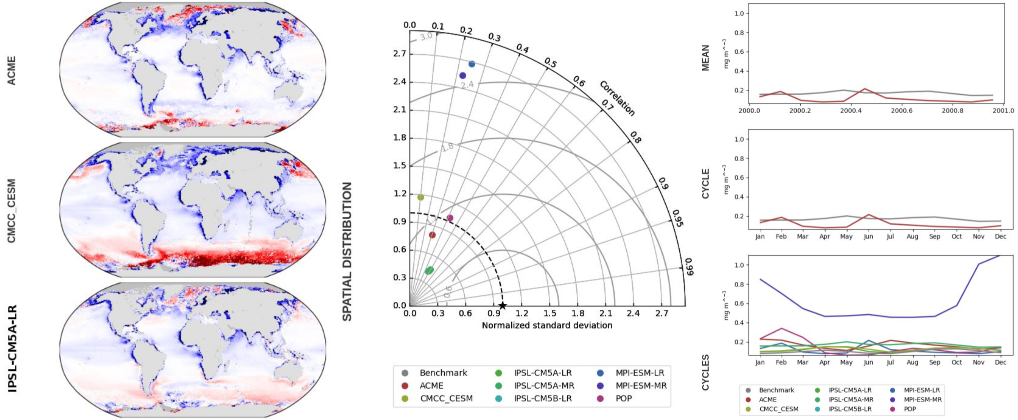
International Ocean Model Benchmarking (IOMB) Package

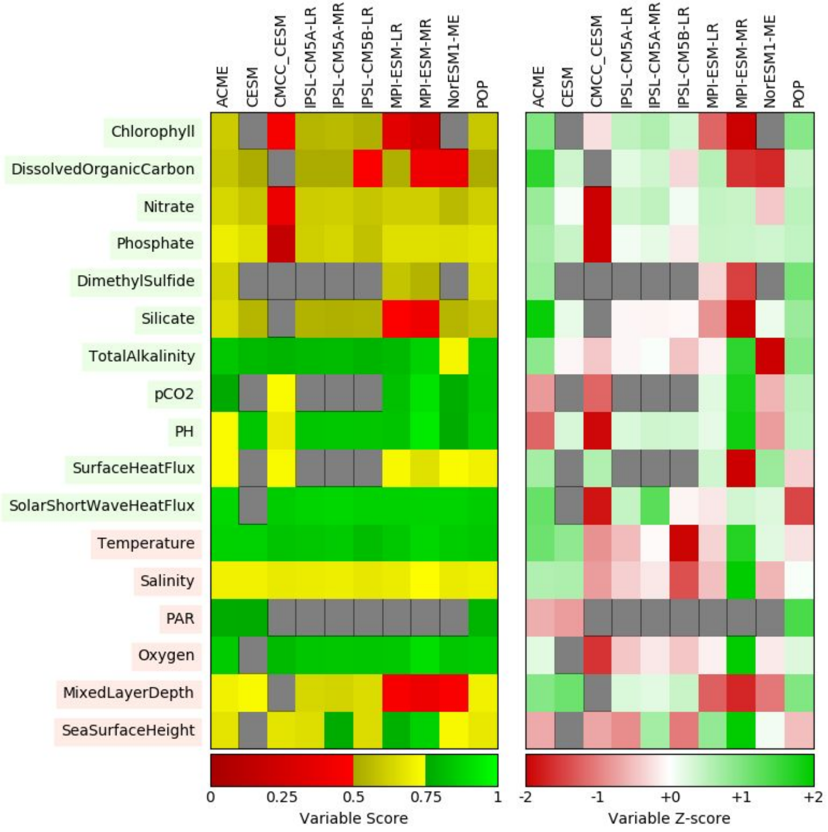
RUBISCO

- Evaluates ocean biogeochemistry results compared with observations (global, point, ship tracks)
- Scores model performance across a wide range of independent benchmark data
- Leverages ILAMB code base, also runs in parallel
- Built on python and open standards
- Is also open source and will be released soon

Chlorophyll / SeaWIFS

Bias | Spatial Distribution | Annual & Seasonal Cycles





For more information...

- **International Land Model Benchmarking (ILAMB) Package**
  https://www.ilamb.org/
- **Reducing Uncertainties in Biogeochemical Interactions through Synthesis and Computation (RUBISCO) Scientific Focus Area**
  https://www.bgc-feedbacks.org/
- **Forrest M. Hoffman**
  Oak Ridge National Laboratory
  forrest@climatemodeling.org