

Climate–Carbon Cycle Feedbacks Constrained by ILAMB

*Forrest M. Hoffman^{1,2}, Nathan Collier¹, Mingquan Mu³, Min Xu¹, Weiwei Fu³,
Cheng-En Yang^{2,1}, Gretchen Keppel-Aleks⁴, David M. Lawrence⁵, Charles D. Koven⁶,
William J. Riley⁶, and James T. Randerson³*

¹Oak Ridge National Laboratory, Oak Ridge, TN, USA

²University of Tennessee, Knoxville, TN, USA

³University of California, Irvine, CA, USA

⁴University of Michigan, Ann Arbor, MI, USA

⁵National Center for Atmospheric Research, Boulder, CO, USA

⁶Lawrence Berkeley National Laboratory, Berkeley, CA, USA

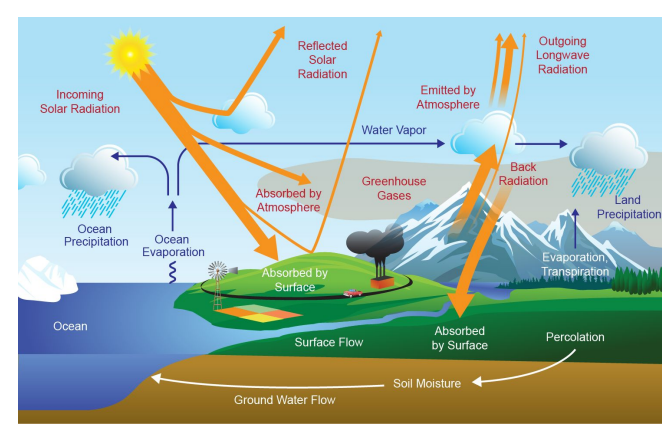
AGU
100
ADVANCING EARTH
AND SPACE SCIENCE

CHAPMAN
CONFERENCE
San Diego, California • 26–29 August 2019

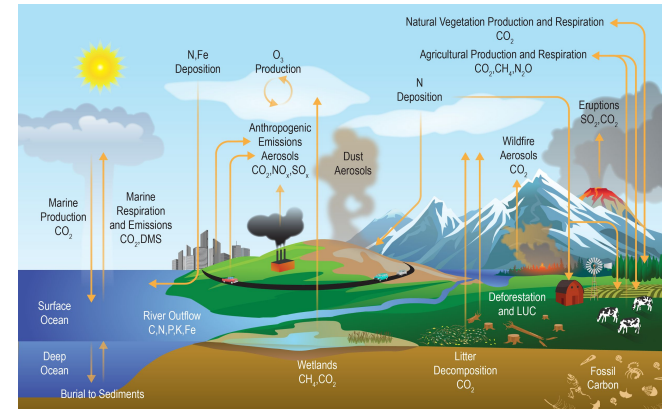
What is ILAMB?

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)



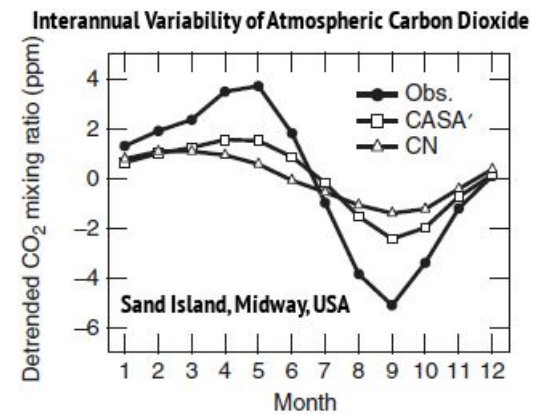
Energy and Water Cycles



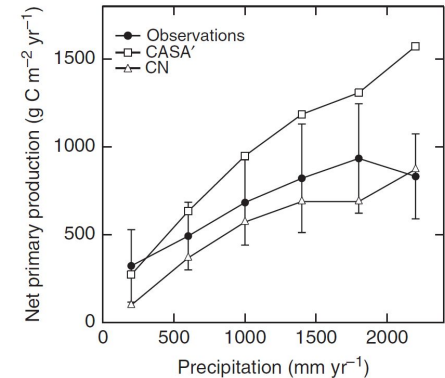
Carbon and Biogeochemical Cycles

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

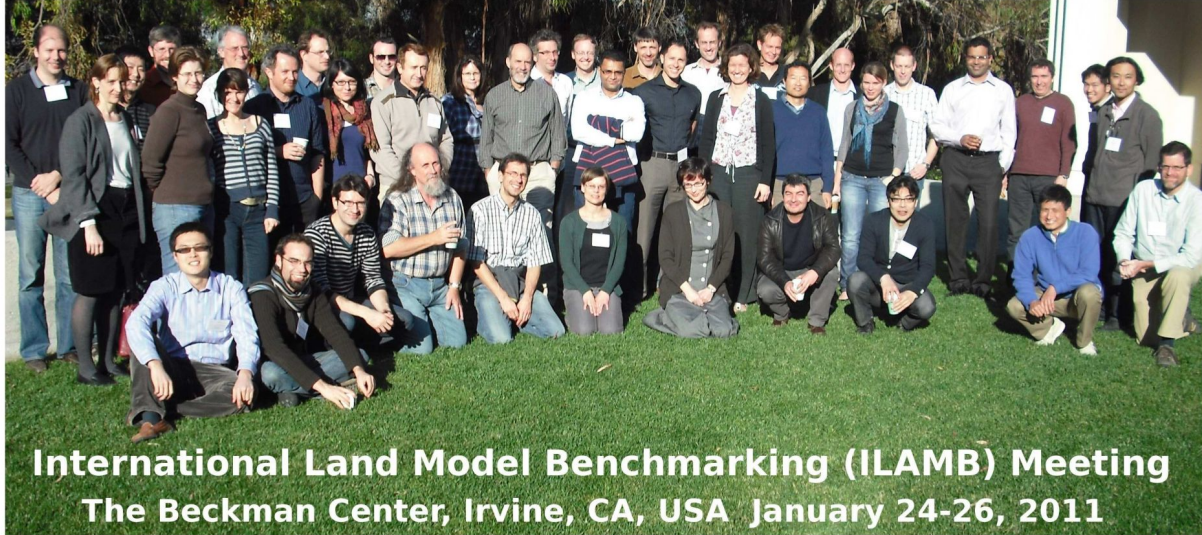


Models often fail to capture the amplitude of the seasonal cycle of atmospheric CO₂



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

(Randerson et al., 2009)



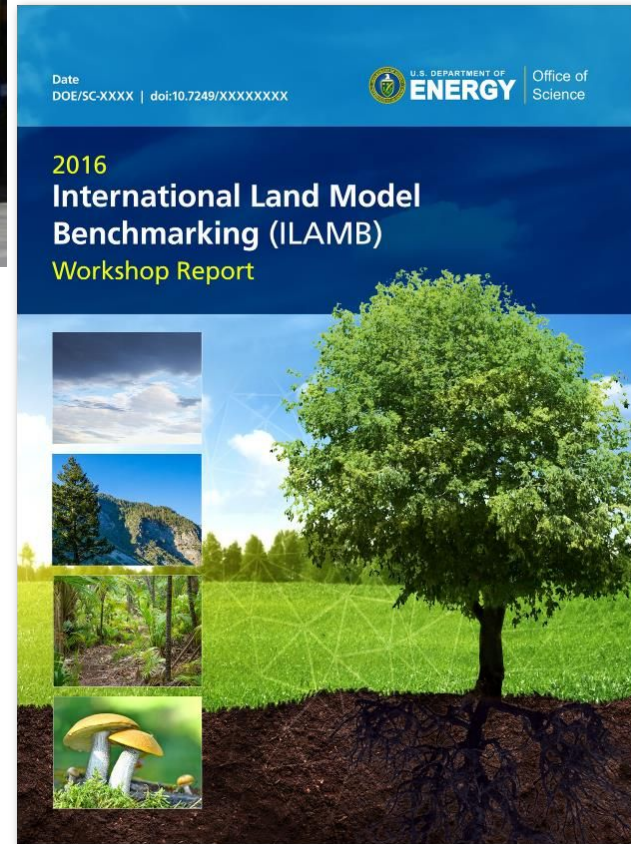
- **First ILAMB Workshop** was held in Exeter, UK, on June 22–24, 2009
- **Second ILAMB Workshop** was held in Irvine, CA, USA, on January 24–26, 2011
 - ~45 researchers participated from the US, Canada, UK, Netherlands, France, Germany, Switzerland, China, Japan, and Australia
 - Developed methodology 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

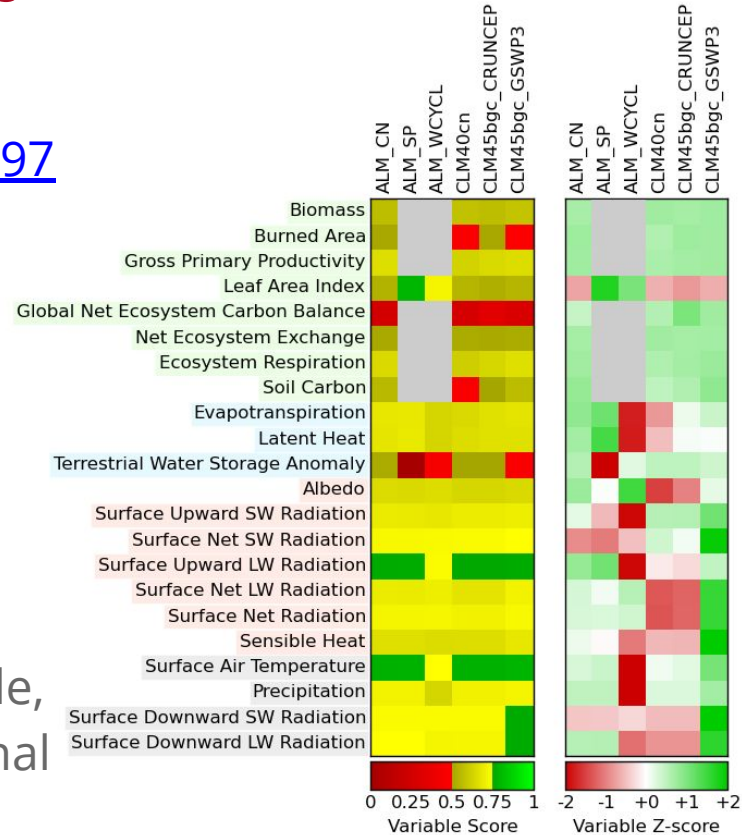
Third ILAMB Workshop was held May 16–18, 2016

- 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



Development of ILAMB Packages

- **ILAMBv1** released at 2015 AGU Fall Meeting Town Hall, doi:[10.18139/ILAMB.v001.00/1251597](https://doi.org/10.18139/ILAMB.v001.00/1251597)
- **ILAMBv2** released at 2016 ILAMB Workshop, doi:[10.18139/ILAMB.v002.00/1251621](https://doi.org/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



ILAMBv2.4 Package Current Variables

- **Biogeochemistry:** Biomass (Contiguous US, Pan Tropical Forest), Burned area (GFED3), CO₂ (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)

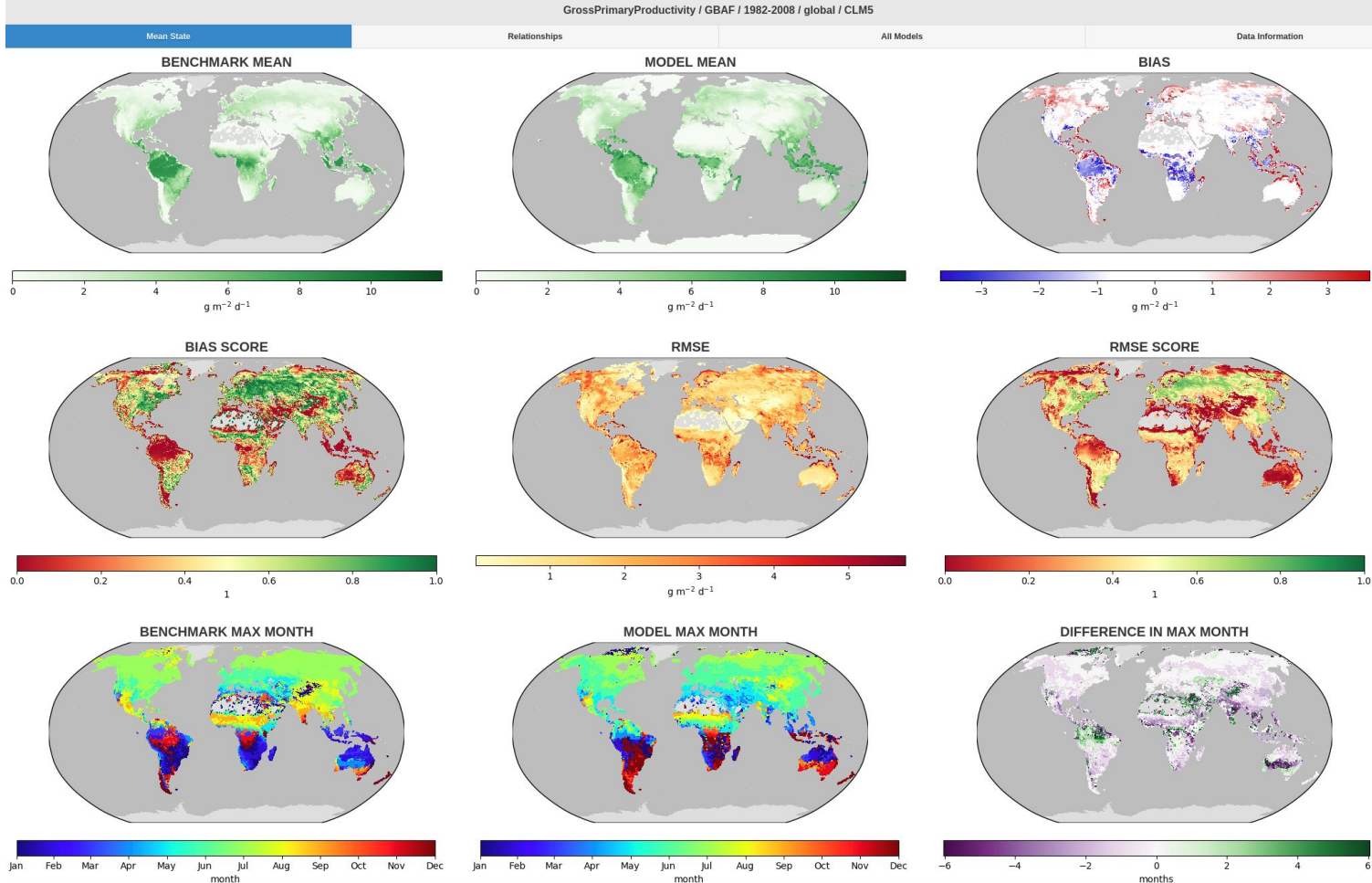
ILAMB Assessing Several Generations of CLM

	CLM4	CLM4.5	CLM5
Ecosystem and Carbon Cycle			
Biomass			
Burned Area			
Carbon Dioxide			
Gross Primary Productivity			
Leaf Area Index			
Global Net Ecosystem Carbon Balance			
Net Ecosystem Exchange			
Ecosystem Respiration			
Soil Carbon			
Hydrology Cycle			
Evapotranspiration			
Evaporative Fraction			
Latent Heat			
Runoff			
Sensible Heat			
Terrestrial Water Storage Anomaly			
Permafrost			
Radiation and Energy Cycle			
Albedo			
Surface Upward SW Radiation			
Surface Net SW Radiation			
Surface Upward LW Radiation			
Surface Net LW Radiation			
Surface Net Radiation			
Forcings			

- 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)

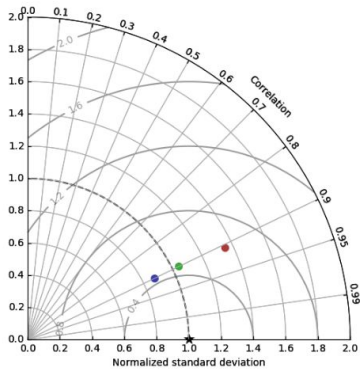
http://webext.cgd.ucar.edu/I20TR/build_set1F/
 (Lawrence et al., in revision)

ILAMB Graphical Diagnostics

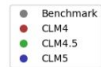




SPATIAL TAYLOR DIAGRAM



MODEL COLORS

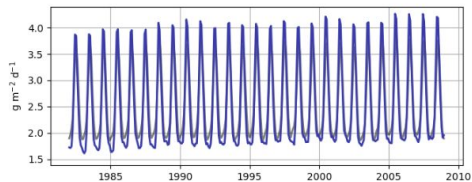


Spatially integrated regional mean

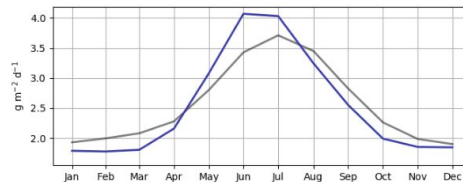
MODEL COLORS



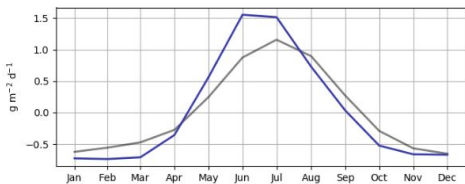
REGIONAL MEAN



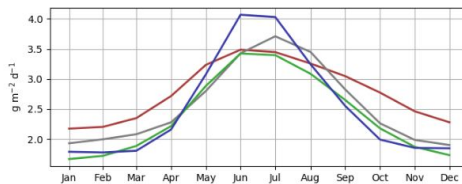
ANNUAL CYCLE



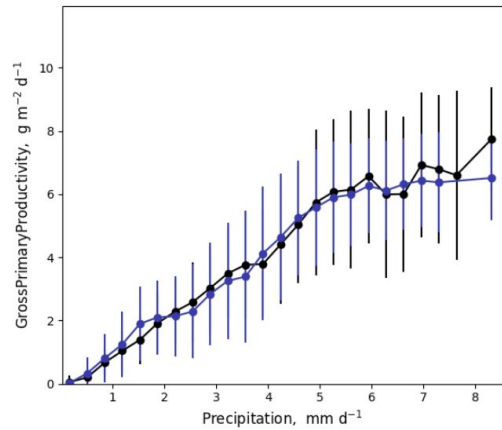
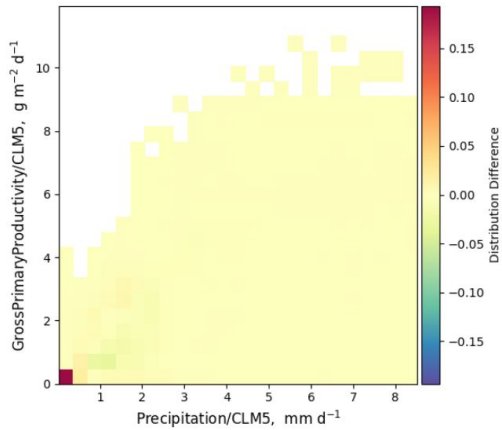
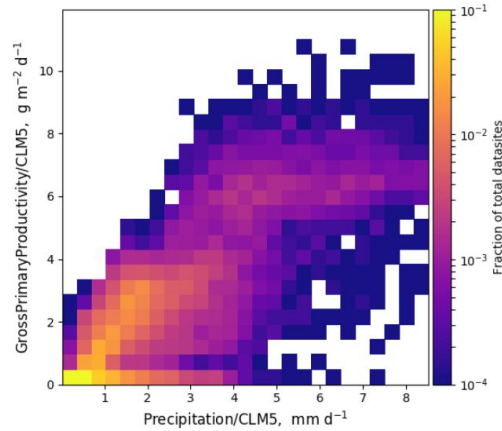
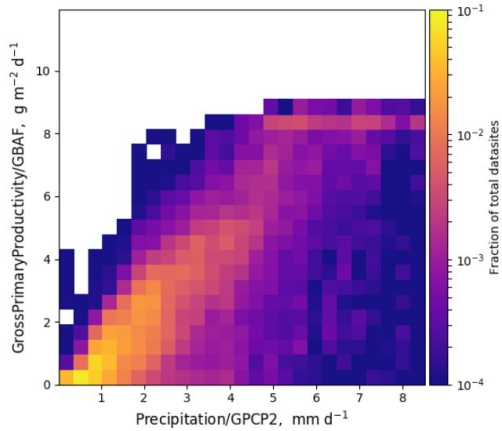
MONTHLY ANOMALY



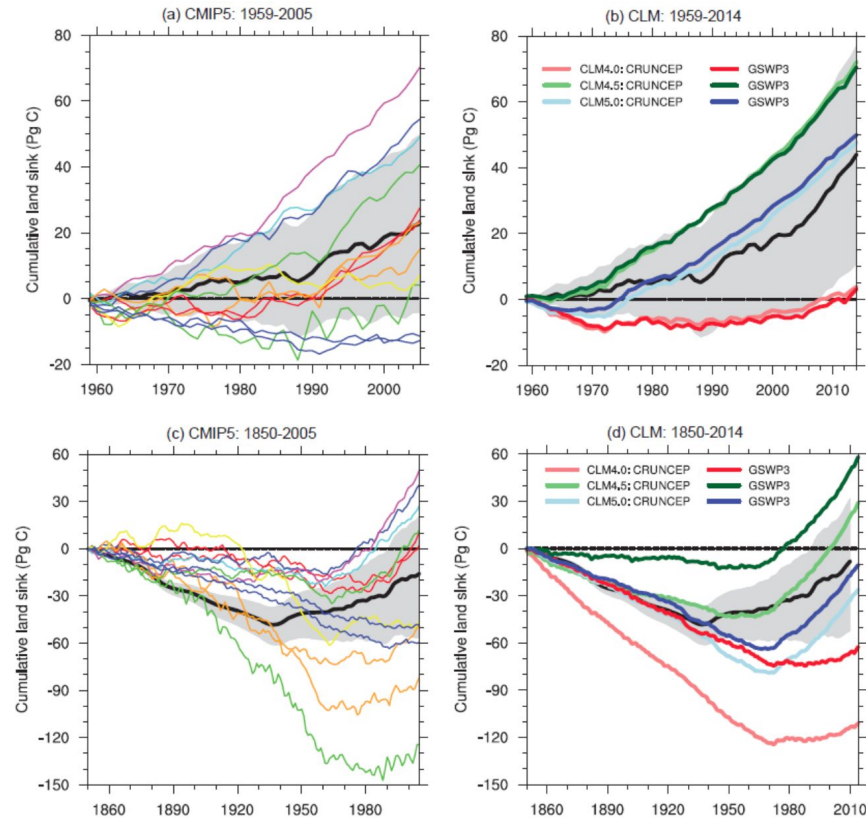
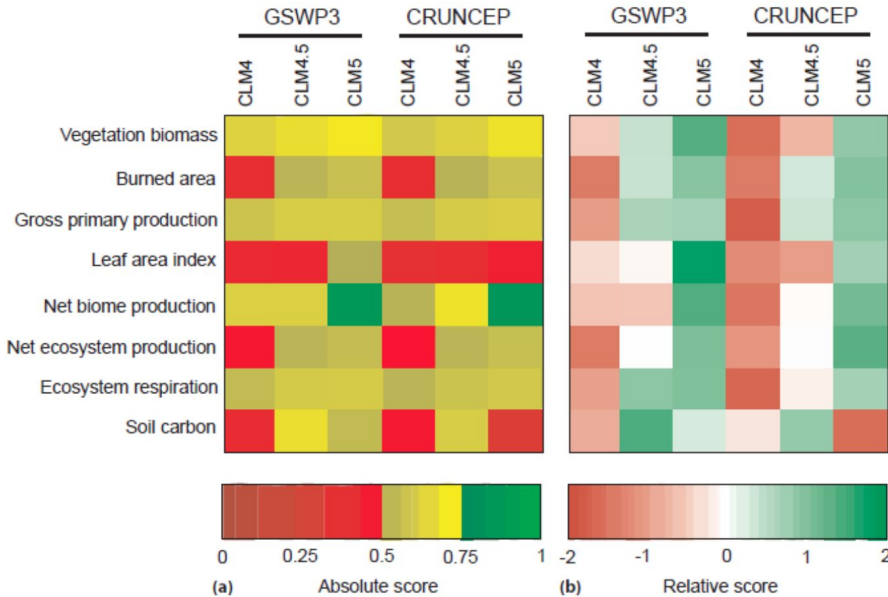
ANNUAL CYCLE



Variable-to-Variable Comparisons



Land Model Performance Depends Strongly on Forcing

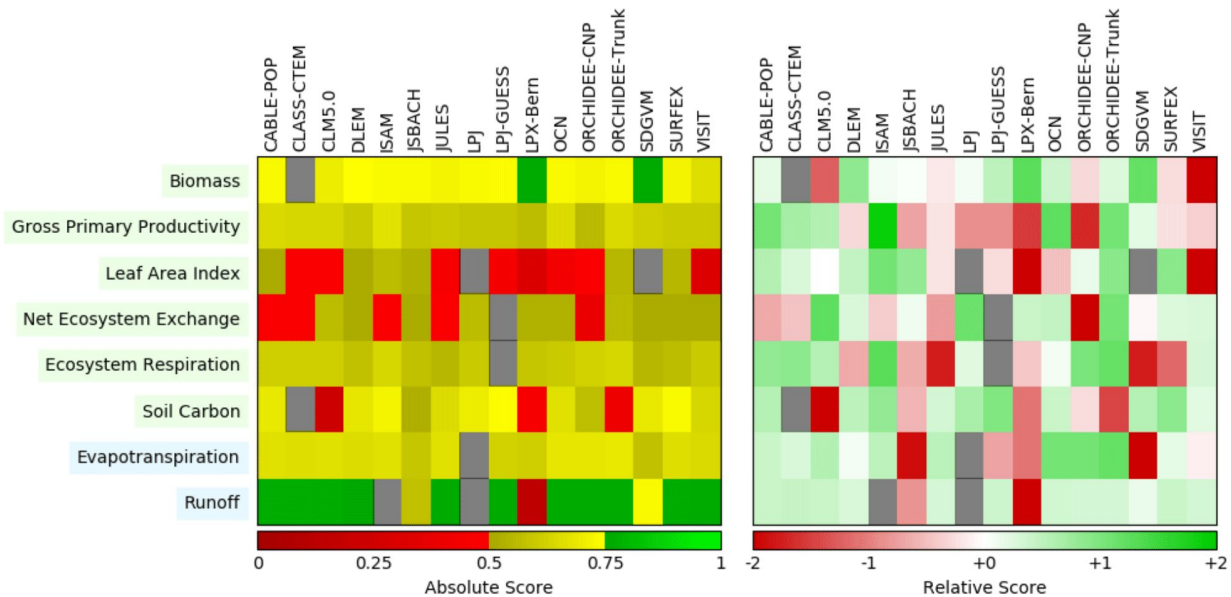


ILAMB performance for CLM4, CLM4.5, and CLM5 forced with GSWP3 vs. CRUNCEP (left) and the cumulative land carbon sink for CMIP5 vs. CLM offline models (right).

Bonan et al. (2019)

Global Carbon Budget 2018 - TRENDY Models

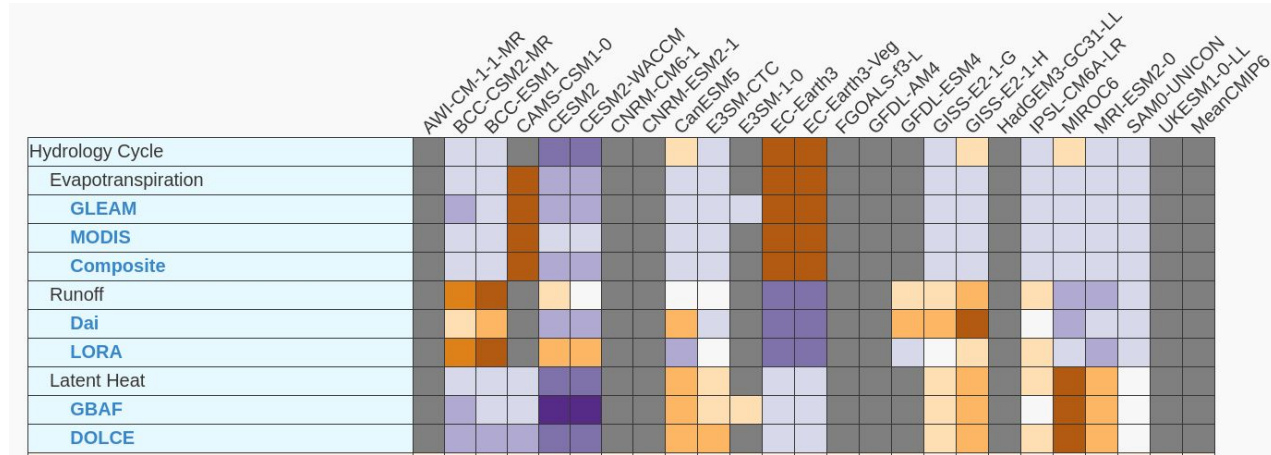
Evaluation of the DGVMs using the International Land Model Benchmarking system (ILAMB; Collier et al., 2018) (left) absolute skill scores and (right) skill scores relative to other models for a subset of ILAMB variables.



Le Quéré et al. (2018)

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

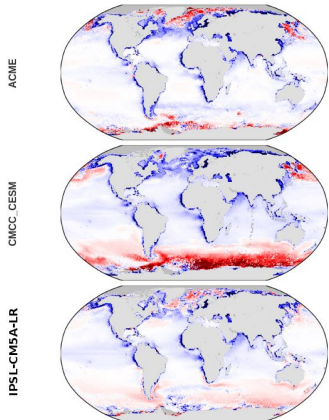


International Ocean Model Benchmarking (IOMB) Package

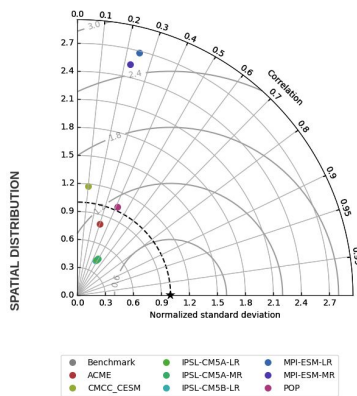
- Evaluates ocean biogeochemistry results compared with observations (global, point, ship tracks)
- Scores model performance across a wide range of independent benchmark data
- Leverages ILAMB code base, also runs in parallel
- Built on python and open standards
- Is also open source and will be released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

Chlorophyll / SeaWiFS

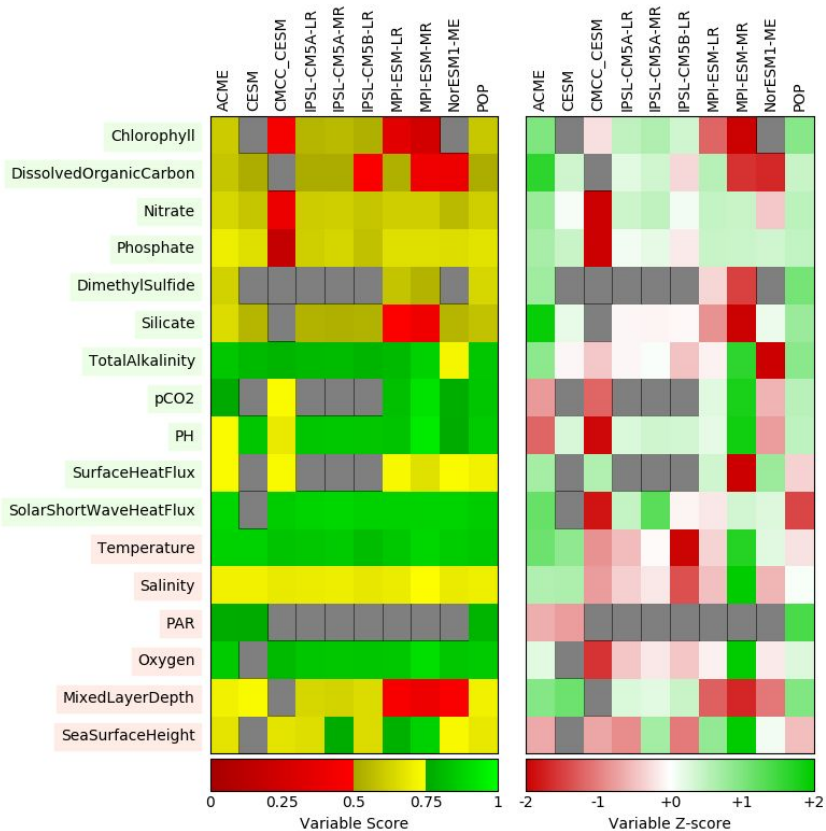
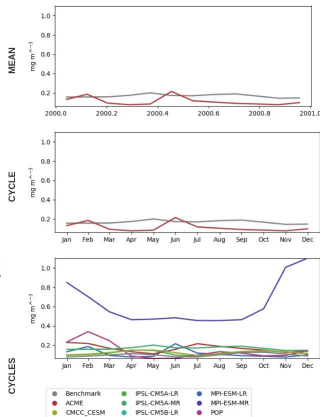
Bias



Spatial Distribution

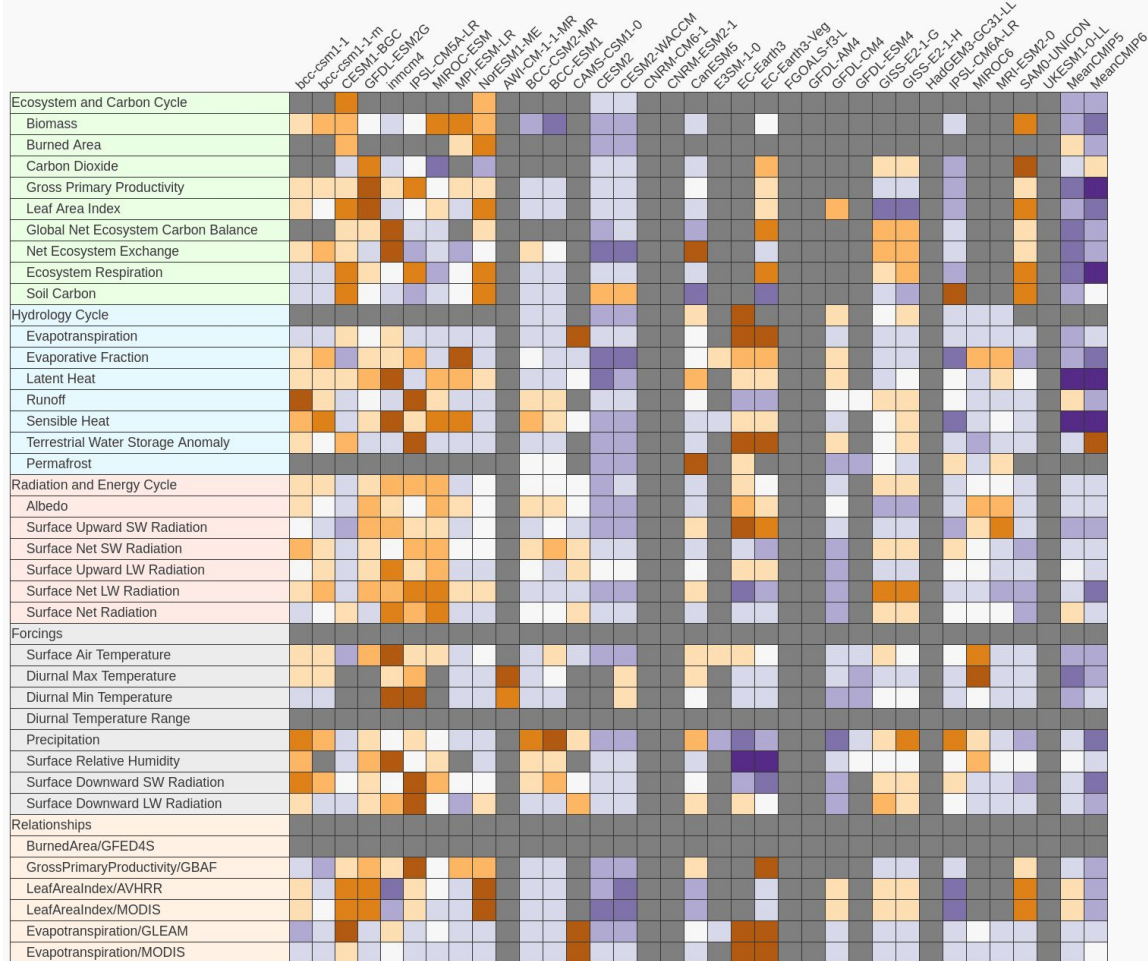
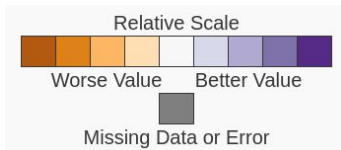


Annual & Seasonal Cycles



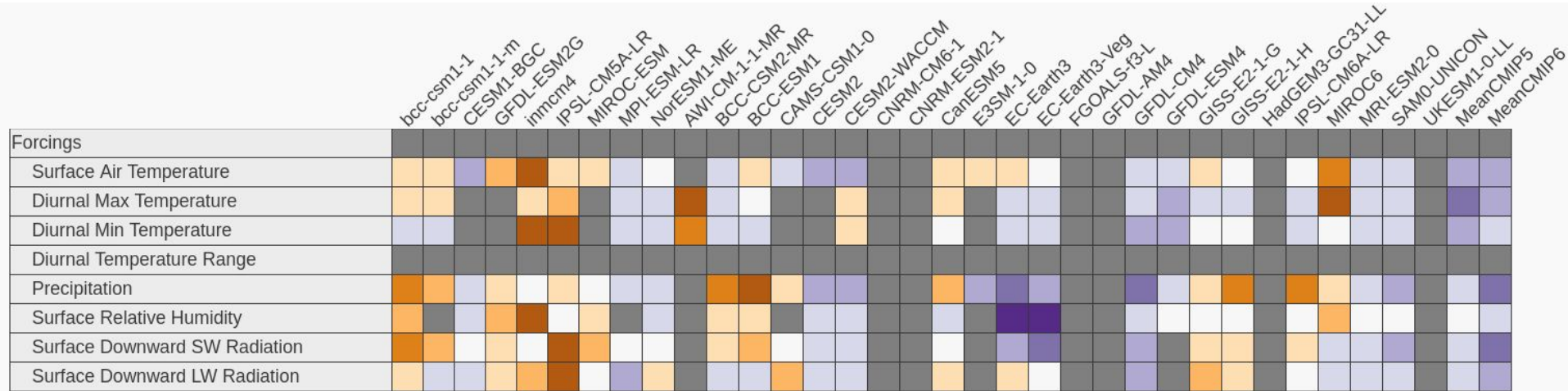
CMIP5 vs. CMIP6 Land Models

- 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



(Hoffman et al., in prep)

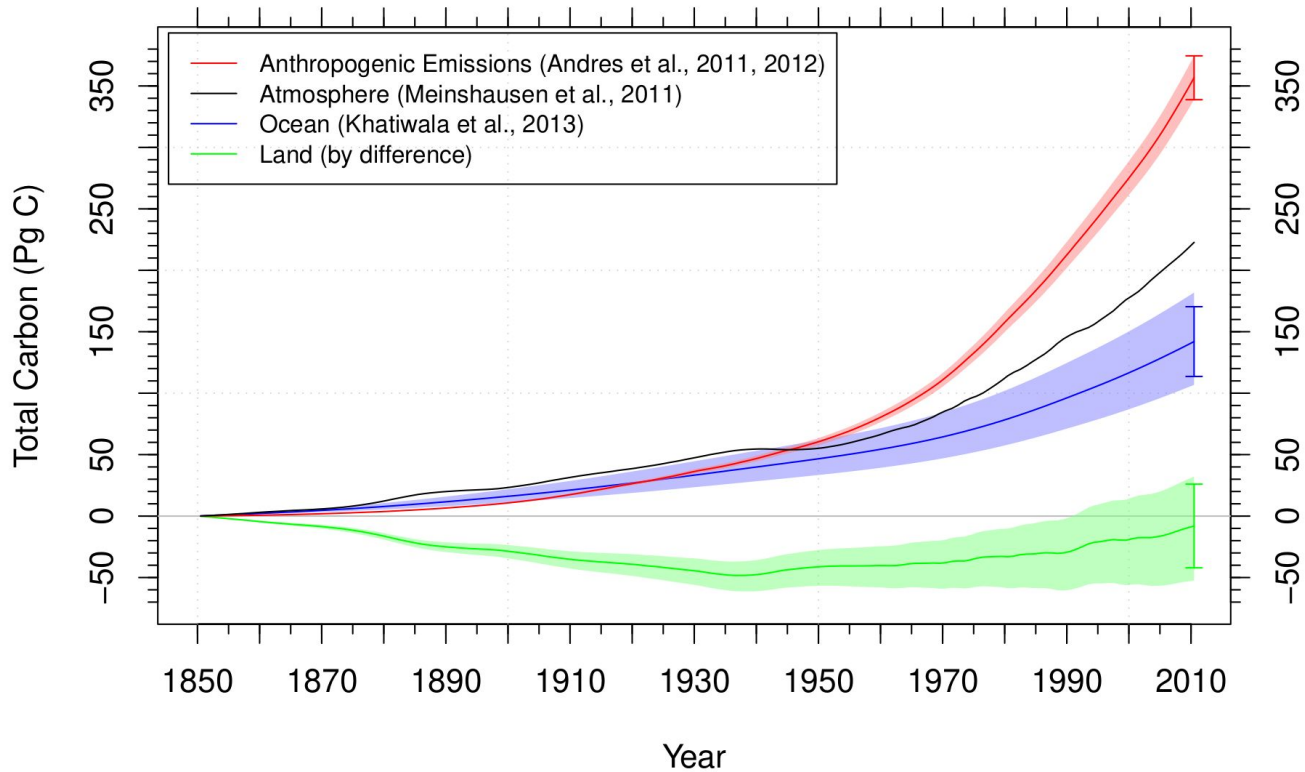
Reasons for Land Model Improvements



Earth system model (ESM) improvements in climate forcing (temperature, precipitation, humidity, and radiation) likely drove some of the improvements exhibited by land carbon cycle models

Observed Carbon Accumulation Since 1850

We used fossil fuel emissions estimates, atmospheric CO₂ measurements, and ocean carbon accumulation estimates from Khatiwala et al. (2013) to estimate land carbon accumulation with propagated uncertainties from 1850 to 2010.



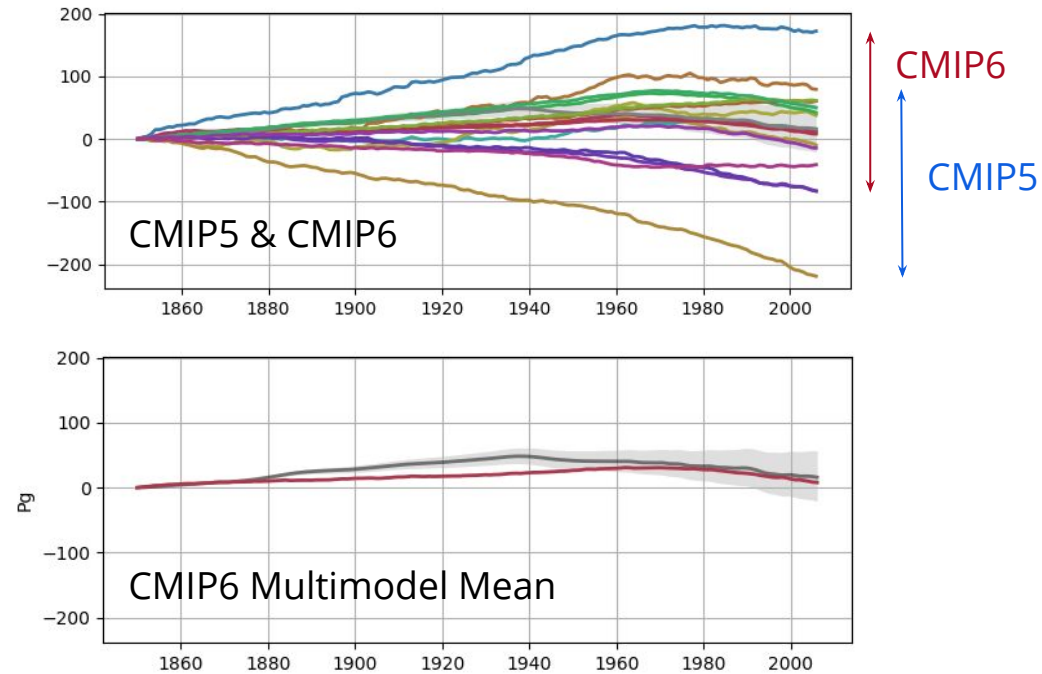


Net Ecosystem Carbon Balance (+□)

- Benchmark
- CanESM5
- CESM1-BGC
- CESM2
- CESM2-WACCM
- EC-Earth3-Veg
- GFDL-ESM2G
- GISS-E2-1-G
- GISS-E2-1-H
- inmcm4
- IPSL-CM5A-LR
- IPSL-CM6A-LR
- MeanCMIP5
- MeanCMIP6
- MIROC-ESM
- NorESM1-ME
- SAM0-UNICON

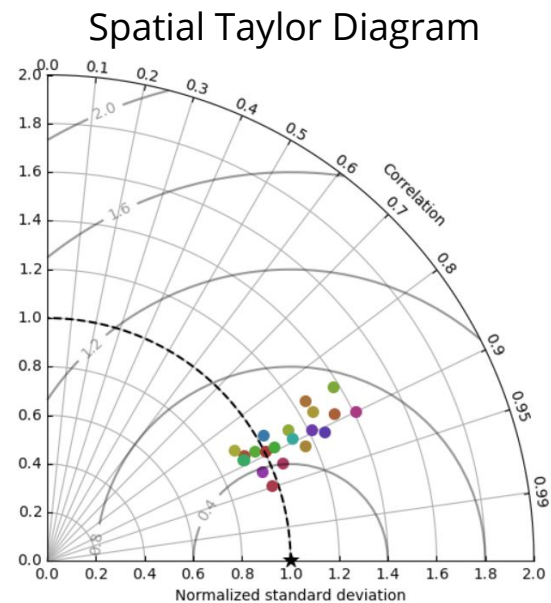
The range of contemporary accumulated land net carbon loss has only slightly decreased (or possibly increased?)

Unit errors probably crept back in on the CMIP5 low-end model, and we have not checked the new CMIP6 high-end model

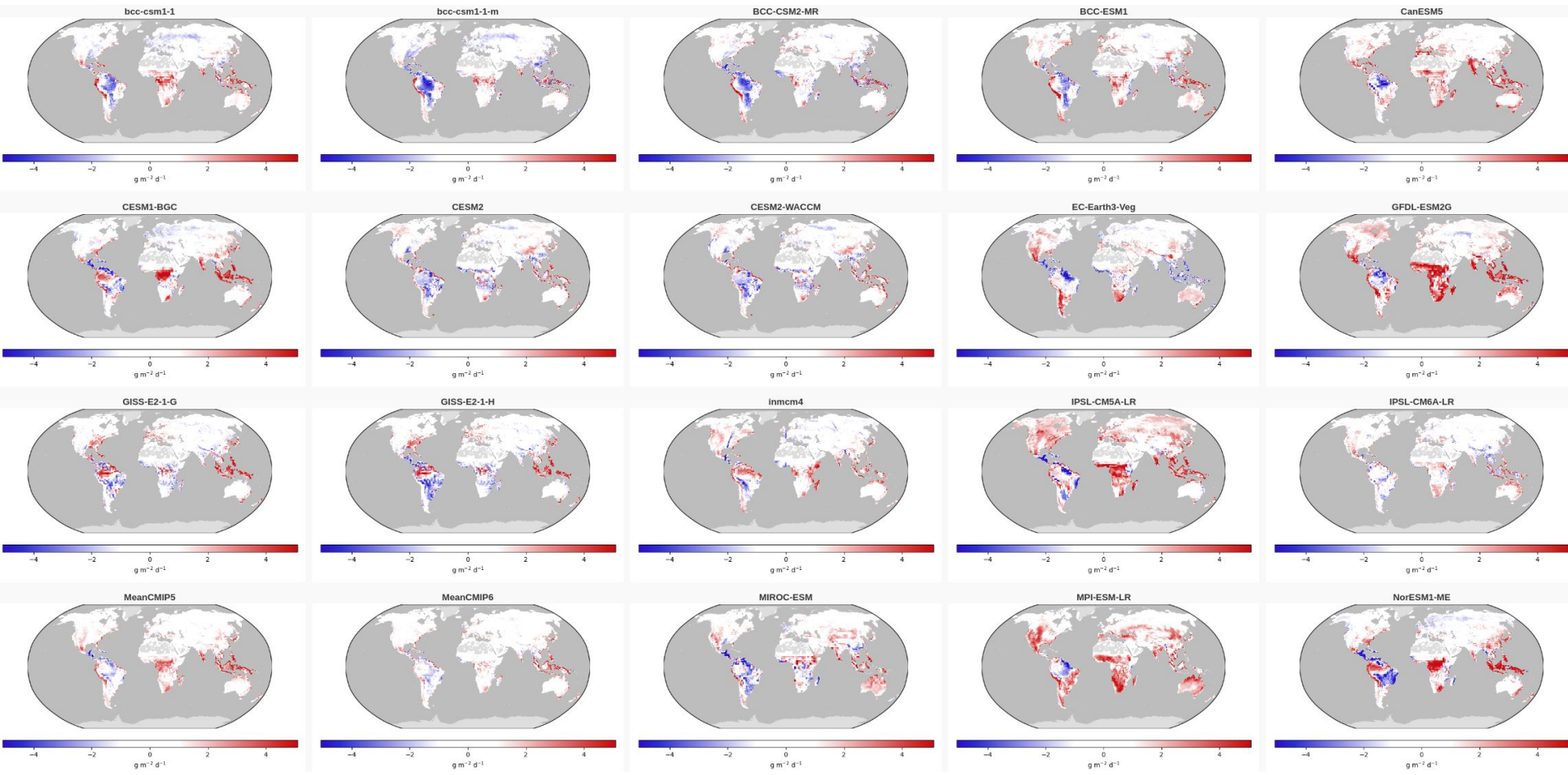


Gross Primary Productivity

- 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



Benchmark	[1]	118.				Download Data	Period Mean (original grids) [Pg yr ⁻¹]	Modal Period Mean (intersection) [Pg yr ⁻¹]	Model Period Mean (complement) [Pg yr ⁻¹]	Benchmark Period Mean (intersection) [Pg yr ⁻¹]	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]
bcc-csm1-1	[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	[1]	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	[1]	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	[1]	157.	133.	21.4	118.	0.0640	0.325	1.84	1.23	0.429	0.302	0.808	0.945	0.557					
CanESM5	[1]	141.	131.	8.05	118.		0.675	1.85	1.70	0.427	0.330	0.761	0.934	0.544					
CESM1-BGC	[1]	129.	124.	4.32	118.	0.501	0.309	1.74	1.38	0.392	0.350	0.761	0.873	0.545					
CESM2	[1]	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	[1]	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	[1]	136.	134.	2.52	118.		0.330	1.99	1.49	0.417	0.312	0.755	0.931	0.545					
GFDL-ESM2G	[1]	167.	155.	9.78	118.		1.19	3.18	1.45	0.360	0.185	0.726	0.880	0.487					
GISS-E2-1-G	[1]	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	[1]	131.	116.	13.8	118.	0.654	-0.0269	1.57	1.19	0.400	0.353	0.760	0.913	0.556					
inmcm4	[1]	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	[1]	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	[1]	116.	111.	4.25	118.	0.486	0.0566	1.45	1.32	0.438	0.364	0.751	0.960	0.587					
MeanCMIP5	[1]	138.	131.	6.75	118.		0.561	1.44	1.13	0.462	0.408	0.794	0.959	0.606					
MeanCMIP6	[1]	121.	116.	5.10	118.		0.159	1.10	1.12	0.522	0.470	0.796	0.973	0.648					
MIROC-ESM	[1]	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	[1]	170.	162.	6.90	110.	8.62	1.22	2.37	1.43	0.378	0.291	0.869	0.926	0.517					
NorESM1-ME	[1]	129.	121.	6.29	118.		0.331	1.92	1.46	0.354	0.350	0.759	0.883	0.530					
SAM0-UNICON	[1]	131.	126.	4.95	118.	0.501	0.371	1.75	1.39	0.398	0.338	0.764	0.845	0.537					



- ILAMB provides a framework for assessing model fidelity in a systematic, comprehensive, and extensible manner
- Routinely confronting models with observations helps identify biases and errors in both models **and data**
- ILAMB has proven useful for verification during model development and for validation in support of multi-model studies
- Better incorporation of observational uncertainty and perturbation experiments, evaluation of diurnal and seasonal behavior, and assessment of responses to extremes are needed

Future Science Questions

- 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?
- **Most relevant here:** Can benchmarking help us resolve the dichotomy between the persistent and growing land sink and evidence for deforestation, browning, and increasing disturbance?

Extra Slides



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

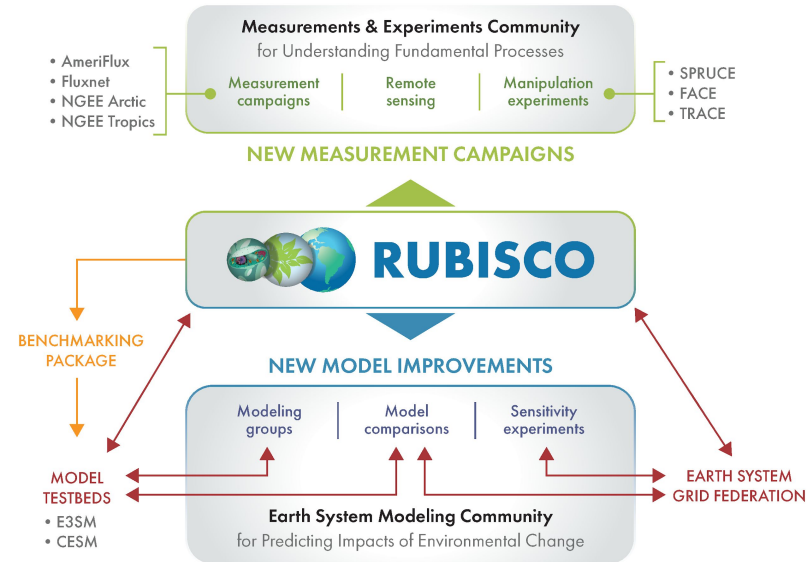
Forrest M. Hoffman (Laboratory Research Manager), William J. Riley (Senior Science Co-Lead), and James T. Randerson (Chief Scientist)

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

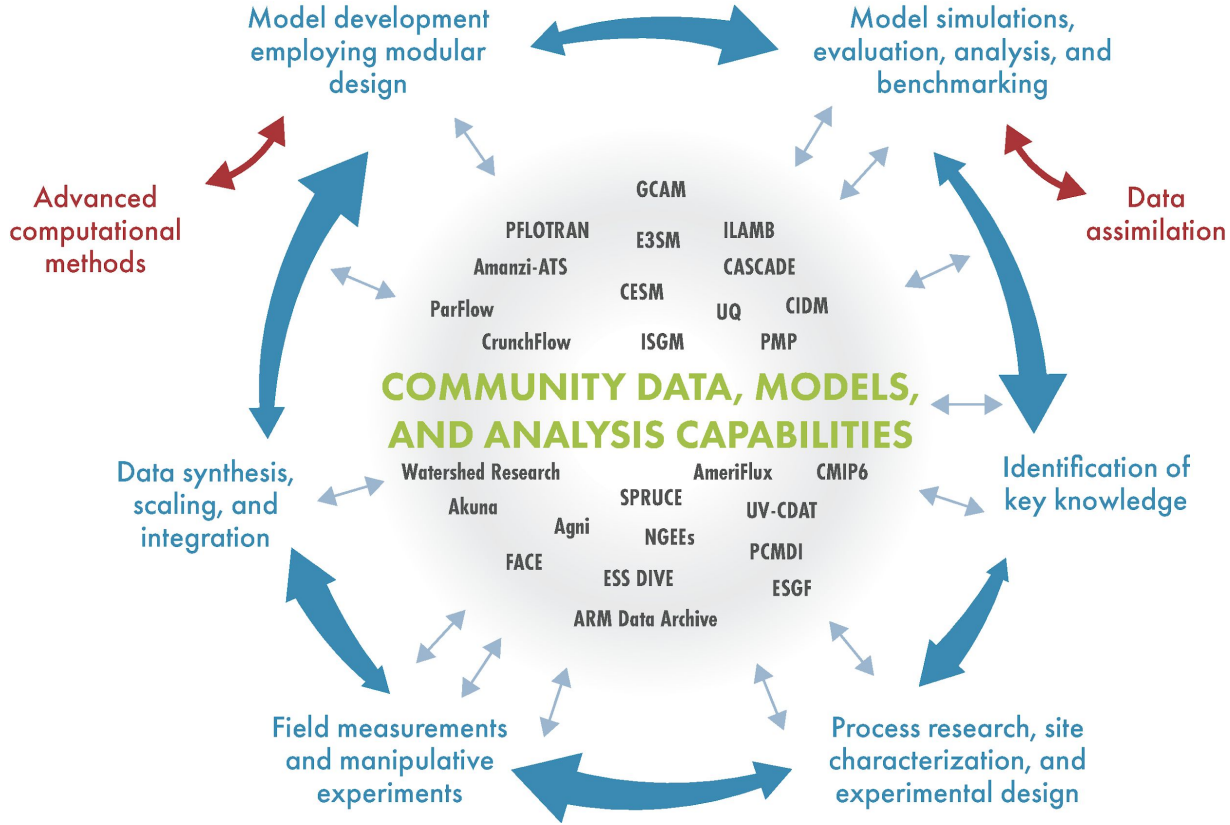
- Perform hypothesis-driven analysis of biogeochemical & hydrological processes and feedbacks in ESMs
- Synthesize in situ and remote sensing data and design metrics for assessing ESM performance
- Design, develop, and release the International Land Model Benchmarking (ILAMB) and International Ocean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



DOE's Model-Data-Experiment Enterprise



- 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>





For more information...

- **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

