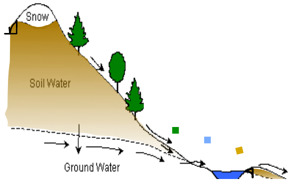
International LAnd Model Benchmarking (ILAMB) Project

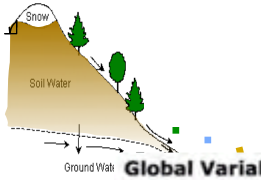
Jim Randerson, Forrest Hoffman, Bill Riley, Dave Lawrence, **Mingquan Mu**, Charlie Koven, Gretchen Keppel-Aleks, Nate Collier

U.S. DEPARTMENT OF ENERGY



International LAnd Model Benchmarking (ILAMB) project
Variables and Visualization within ILAMB

- Currently integrates analysis of 25 variables in 4 categories from ~60 datasets
  - Above ground 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 radiation, surface net SW radiation, surface upward LW radiation, surface net LW radiation, surface net radiation
  - surface air temperature, precipitation, surface relative humidity, surface downward SW radiation, surface downward LW radiation
- Graphics and scoring system
  - annual mean, bias, relative bias, RMSE, seasonal cycle phase, spatial distribution, interannual variability, variable-to-variable
  - Global maps, time series plots averaged over specific regions, individual measurement sites, functional relationships

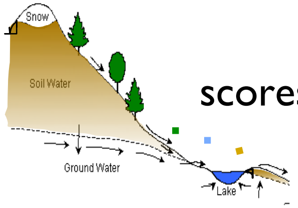


Global Variables

Global Variables (Info for Weightings)

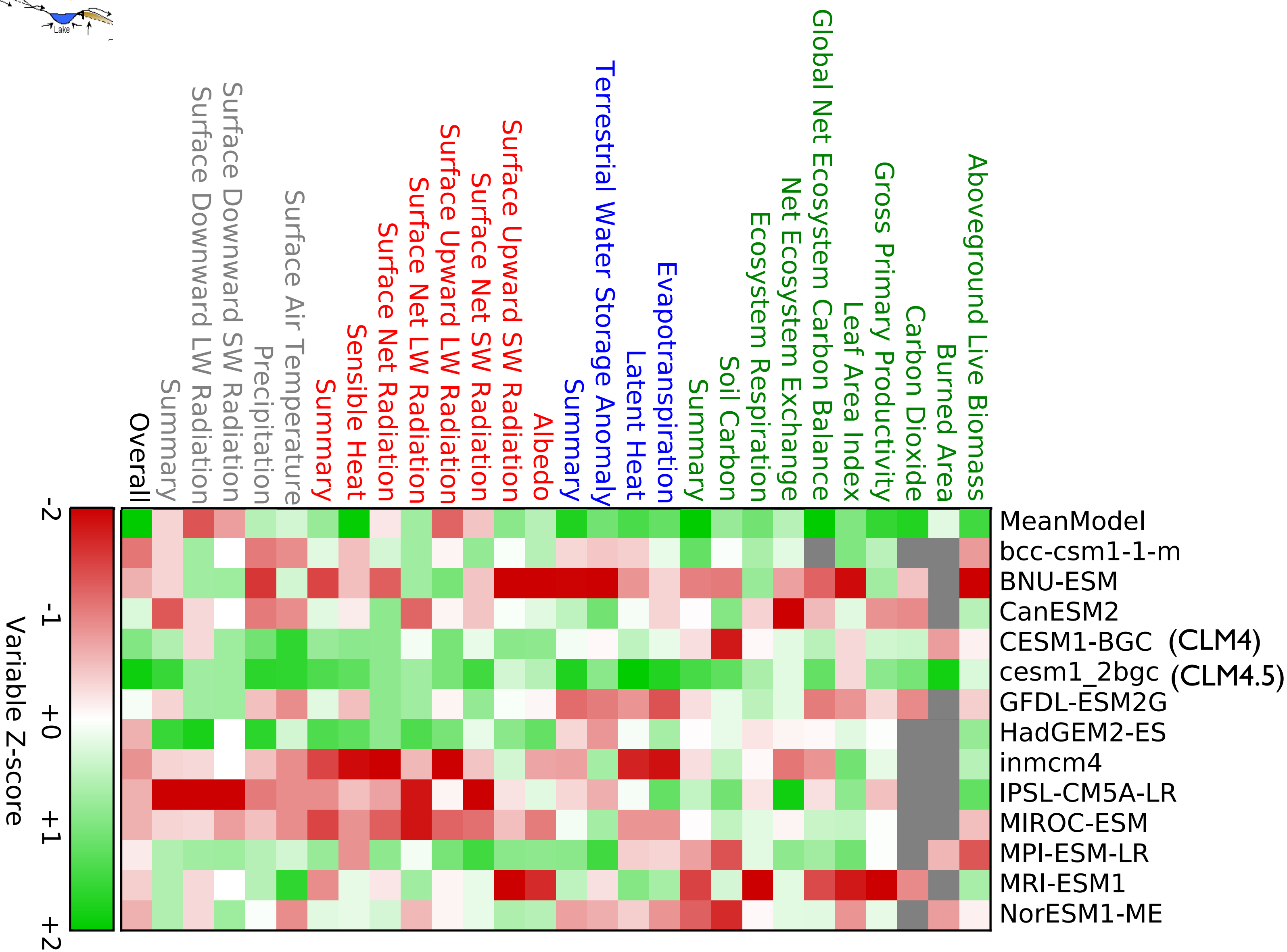
	MeanModel	bcc-csm1-1-m	BNU-ESM	CanESM2	CESM1-BGC	GFDL-ESM2G	HadGEM2-ES	inmcm4	IPSL-CM5A-LR	MIROC-ESM	MPI-ESM-LR	MRI-ESM1	NorESM1-ME
Live Biomass Carbon	0.73	0.68	0.33	0.65	0.60	0.62	0.72	0.50	0.56	0.62	0.58	0.56	0.57
Burned Area	0.38	-	-	-	0.37	-	-	-	-	-	0.38	-	0.38
Carbon Dioxide	0.85	-	0.65	0.65	0.78	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	0.62	0.51	0.55	0.55	0.54	0.60	0.56	0.55	0.59	0.55	0.50	0.54
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
Evaporative Fraction	0.84	0.76	0.77	0.81	0.81	0.75	0.81	0.81	0.72	0.75	0.75	0.80	0.79
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
Runoff	0.61	0.59	0.60	0.58	0.64	0.59	-	0.62	0.57	0.56	0.60	0.70	0.62
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
Terrestrial Water Storage Anomaly	0.38	0.37	0.36	0.38	0.38	0.38	-	0.38	0.37	0.38	0.38	0.38	0.38
Summary	0.68	0.65	0.65	0.66	0.67	0.64	0.77	0.64	0.64	0.63	0.66	0.68	0.66
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	0.77	0.74	0.67	0.74	0.78	0.74	0.77	0.74	0.73	0.72	0.78	0.67	0.76
Surface Net SW Radiation	0.84	0.86	0.84	0.85	0.86	0.86	0.86	0.84	0.82	0.83	0.87	0.85	0.85
Surface Upward LW Radiation	0.89	0.91	0.91	0.91	0.92	0.91	0.92	0.89	0.90	0.91	0.92	0.91	0.91
Surface Net LW Radiation	0.81	0.82	0.81	0.79	0.81	0.81	0.83	0.80	0.78	0.78	0.81	0.81	0.81
Surface Net Radiation	0.78	0.79	0.76	0.80	0.80	0.81	0.80	0.74	0.77	0.77	0.81	0.78	0.80
Summary	0.80	0.80	0.77	0.80	0.81	0.80	0.82	0.77	0.78	0.78	0.82	0.78	0.81
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	0.71	0.69	0.67	0.69	0.72	0.69	0.73	0.69	0.69	0.69	0.72	0.70	0.70
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	0.89	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	0.83	0.81	0.80	0.83	0.82	0.84	0.81	0.81	0.82	0.84	0.83	0.82
Overall	0.69	0.54	0.59	0.61	0.64	0.57	0.48	0.58	0.57	0.59	0.61	0.59	0.63

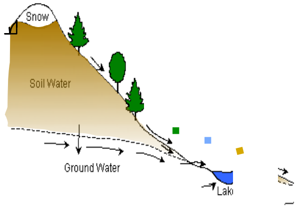
Notes: 4 Categories are divided: Ecosystem and Carbon Cycle, Hydrology and Turbulent Flux, Radiation and Energy Cycle, and Forcings.



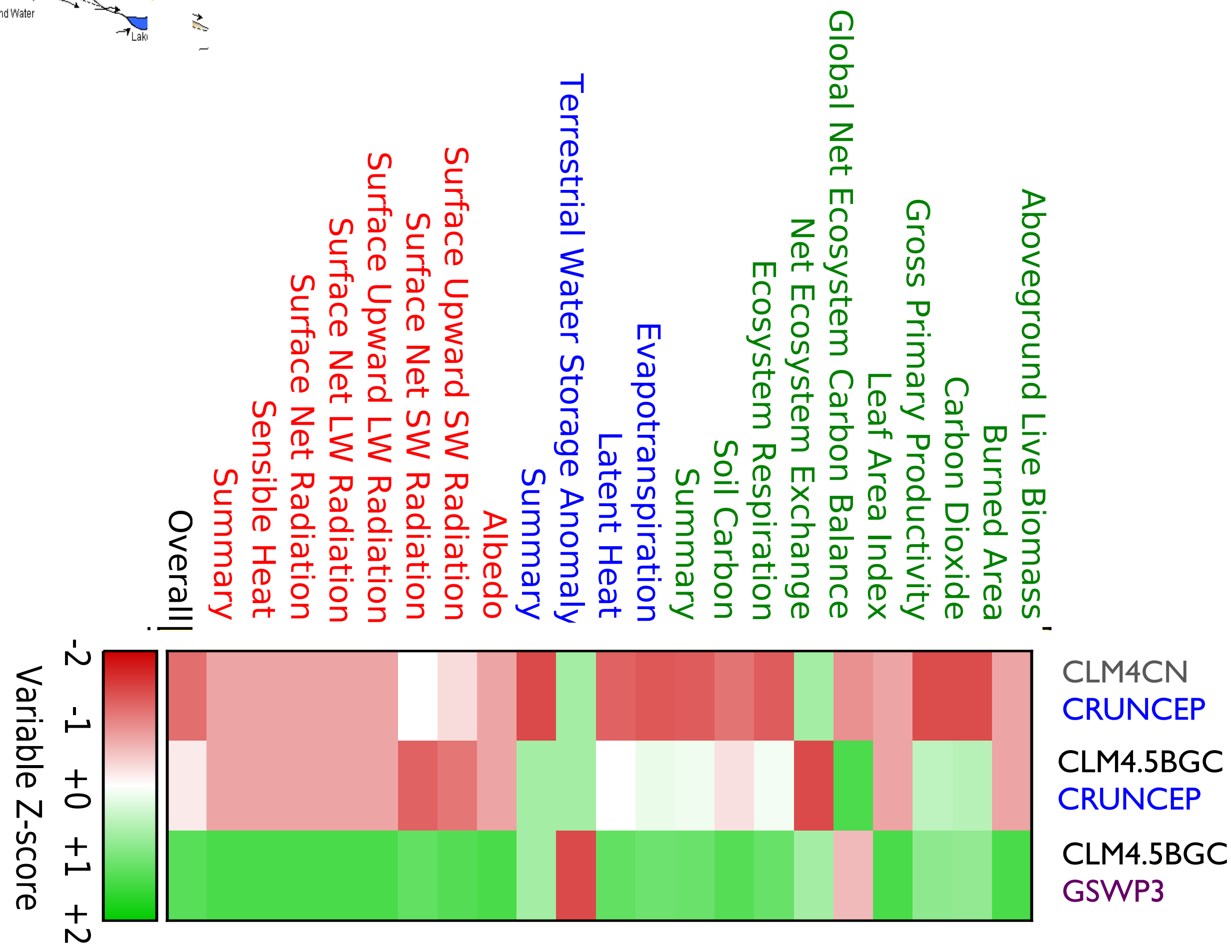
International LAnd Model Benchmarking (ILAMB) package

scores for RMSE, interannual variability, pattern correlation, variable-to-variable comparisons, +





Utilizing ILAMB in model development and assessment process



Impact of model structural and parameter changes

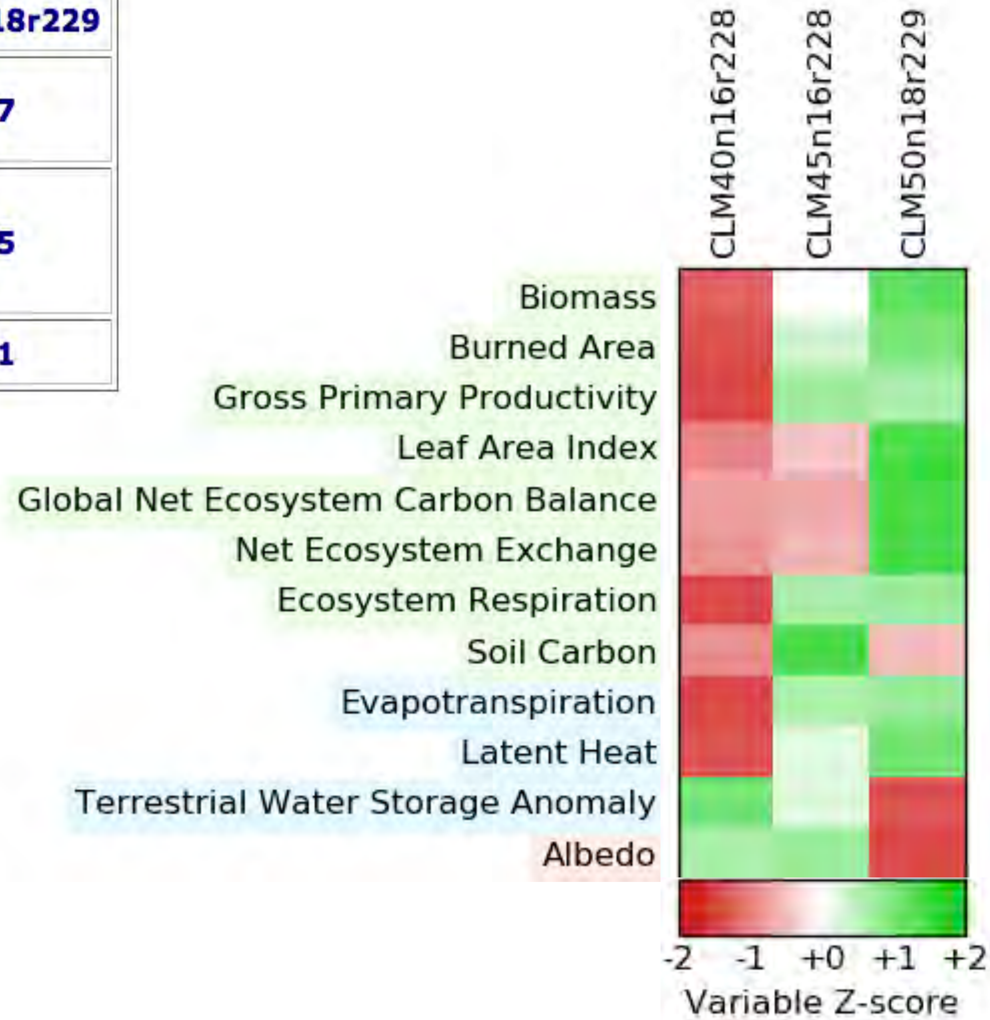
Impact of forcing dataset

Assessment in ILAMB

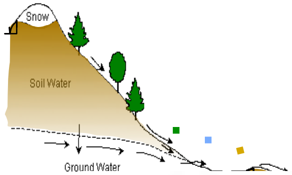
Metrics for RMSE, bias, spatial pattern corr, interannual variability, funct relationships

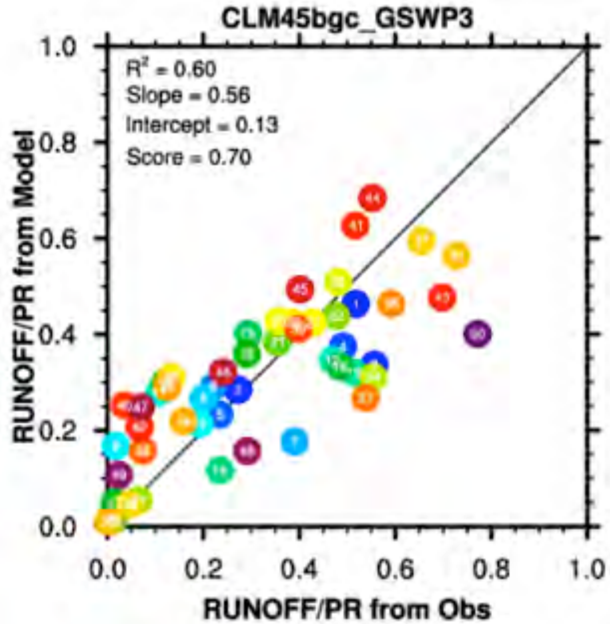
Overall Scores

	CLM40n16r228	CLM45n16r228	CLM50n18r229
Global Variables	0.63	0.66	0.67
Variable to Variable	0.70	0.72	0.75
Overall	0.66	0.69	0.71

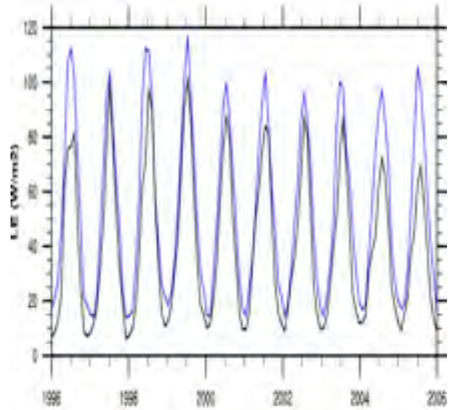


Examples of ILAMB metrics / plots

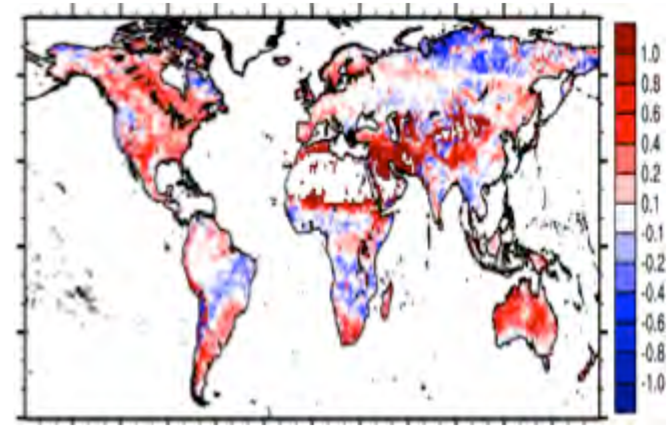




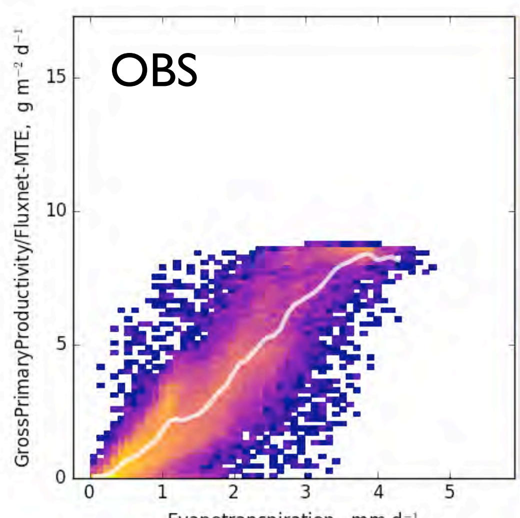
Tower Site

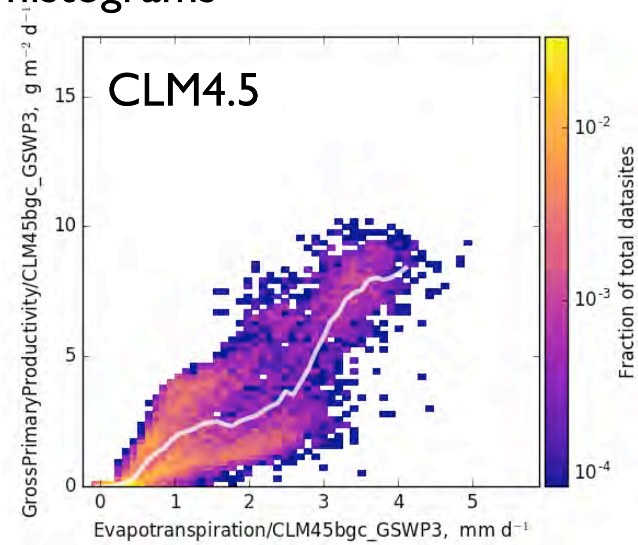


Global bias, relative bias, RMSE

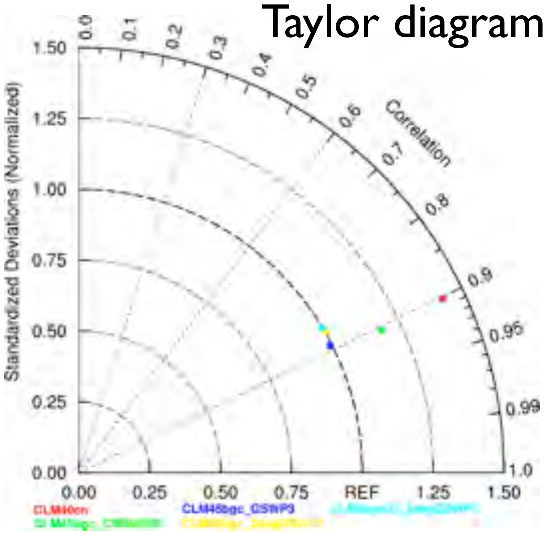


2-d histograms





Taylor diagram



ILAMB versions 1 and 2 are available

- Version 1 – written in NCL
  - http://redwood.ess.uci.edu/mingquan/www/ILAMB/index.html
  - Tuned and vetted versions working with CMIP5 historical, CMIP5 esmHistorical, and CLM development branches

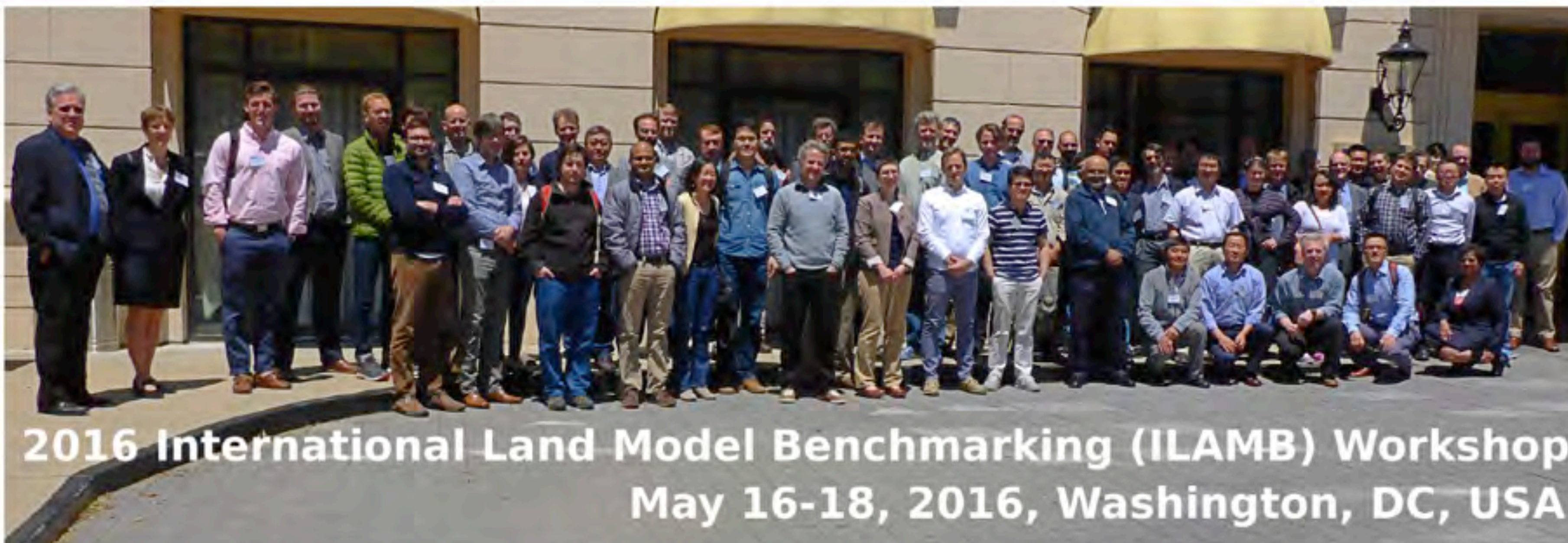
- Version 2 – written in Python and is parallel
  - Hosted in a git repository: https://bitbucket.org/ncollier/ilamb
  - Tutorial:: http://climate.ornl.gov/~ncf/ILAMB/docs/index.html
  - Sample output: http://www.climatemodeling.org/~nate/ILAMB/index.html

- Both versions have the following features:
  - constructed with a modular structure, so that new models, variables or benchmarks can be easily added
  - High quality output files (encapsulated postscript files) can be used directly for publications or proposals.

2016 International Land Model Benchmarking (ILAMB) Workshop

Report of an international workshop held in Washington, DC, USA, May 16–18, 2016

Supported by the US Department of Energy Office of Science, Biological and Environmental Research



http://science.energy.gov/ber/community-resources/

Hoffman, F. M., C. D. Koven, G. Keppel-Aleks, D. M. Lawrence, W. J. Riley, J. T. Randerson, A. Ahlstršm, G. Abramowitz, D. D. Baldocchi, M. Best, B. Bond-Lamberty, M. De 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 (2016), **International Land Model Benchmarking (ILAMB) 2016 Workshop Report**, DOE/SC-0186, U.S. Department of Energy, Office of Science, Germantown, Maryland, USA, 157 pp., doi:10.2172/1330803.

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

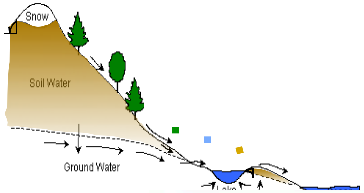
- Develop super site benchmarks integrated with AmeriFlux and FLUXNET
- Create benchmarks for soil carbon turnover and vertical distribution and transport
- Develop benchmark metrics for extreme event statistics and response of ecosystems
- Synthesize data for vegetation recruitment, growth, mortality, 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

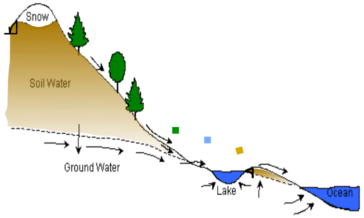
- Process understanding
- Quantified feedbacks
- Reduced uncertainties
- Improved model projections



ILAMB Workshop Report

KEY RECOMMENDATIONS

- Well-established aspects of model assessment should be a routine component of the model development process that over time becomes increasingly comprehensive.
- Evaluation tools should include testing the predictive power of models under a changing climate.
- Benchmarking packages should span a wide range of spatial and temporal scales and extents.
- Integration of a diversity of evaluation tools into a common workflow framework could lead to new insights into climate processes and phenomena.
- Evaluation and benchmarking systems should be open source and freely distributed to leverage the work of many modeling teams and to minimize redundancy.
- Benchmarking tools should be integrated with data repositories that support standardized access through an applications programming interface.

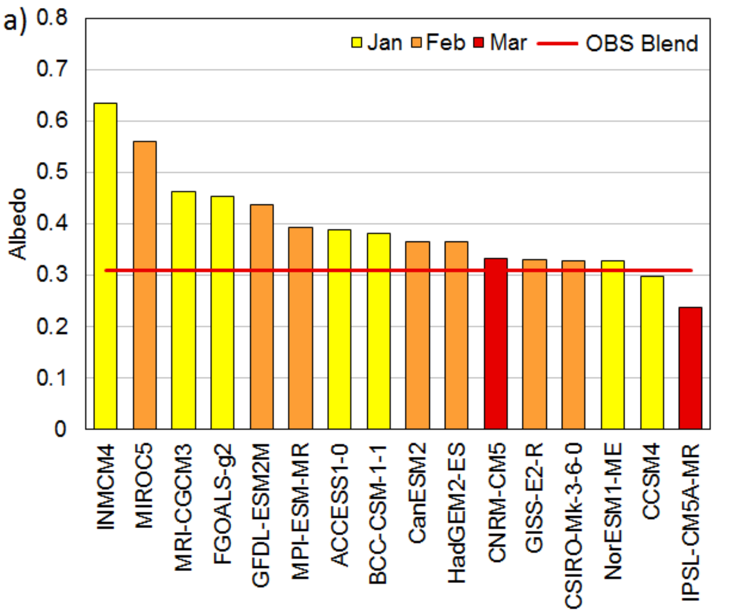


Synergies across land model assessment activities

- Coordination of these distinct and international land model benchmarking/ assessment activities is challenging due to the diversity of approaches and the complexities of the international funding environment.
- Over the longer term, it may be possible and beneficial to integrate existing land diagnostics packages (ILAMB, PALS, LVT, etc) under a loosely coordinated framework (potentially similar to ESMValTool).
- Benefit of a reduction of effort related to the overhead of benchmarking (e.g., workflow processes such as reading in, processing, and reformatting model and observational data), allowing more effort and funding to be devoted to metrics development.
- Should explore potential for a joint benchmarking analysis project, wherein each of the existing packages is applied to a set of multi-model output that would enable direct comparison and evaluation of how each package uniquely contributes to our understanding of model strengths and weaknesses.

Snow Albedo Metric

- Detects model performance related to the seasonality of snow and albedo changes, an important factor for surface albedo feedback strength.
- A multi-observational approach is used for albedo (MODIS/APP-x/GlobAlbedo) and SCF (MODIS/ GlobSnow) to create blended products (OBS Blend) with which to evaluate models.
- Adapt a skill metric from Taylor (2001) to evaluate changes in snow cover and albedo across various regions of interest in relation to correlation (R), and the ratio of standard deviations (α_f).
- Looking at the month-to-month climatological changes in albedo and SCF over the snow (Sep-Jun) or melt (Jan-Jun) seasons.
- This metric does not penalize the mean bias, but this should not have a large impact because we use month-to-month changes in bounded variables (albedo and SCF both start and end the snow season at similar values).



Maximum monthly mean climatological (1980-2005) surface albedo over the boreal forest. Color coding shows when the peak albedo occurs.

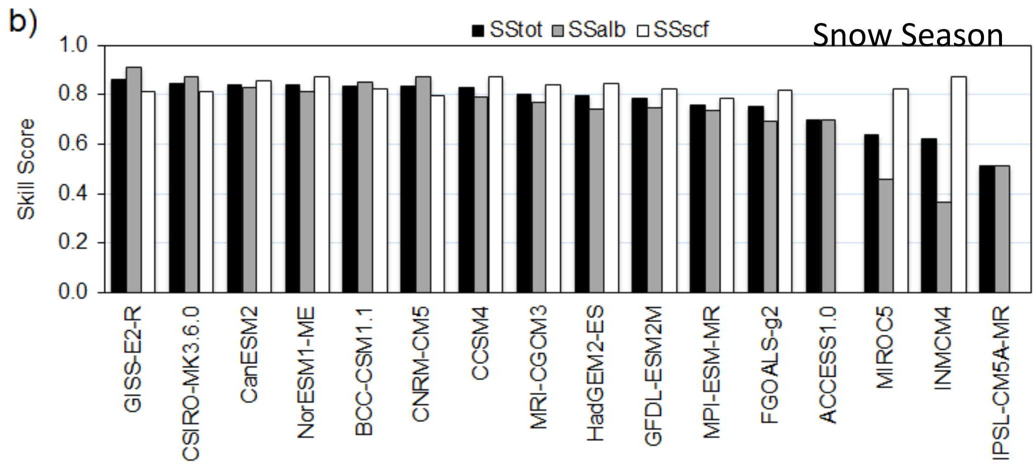
$$SS = \frac{2(1+R)}{(\hat{\sigma}_f + 1/\hat{\sigma}_f)^2}$$

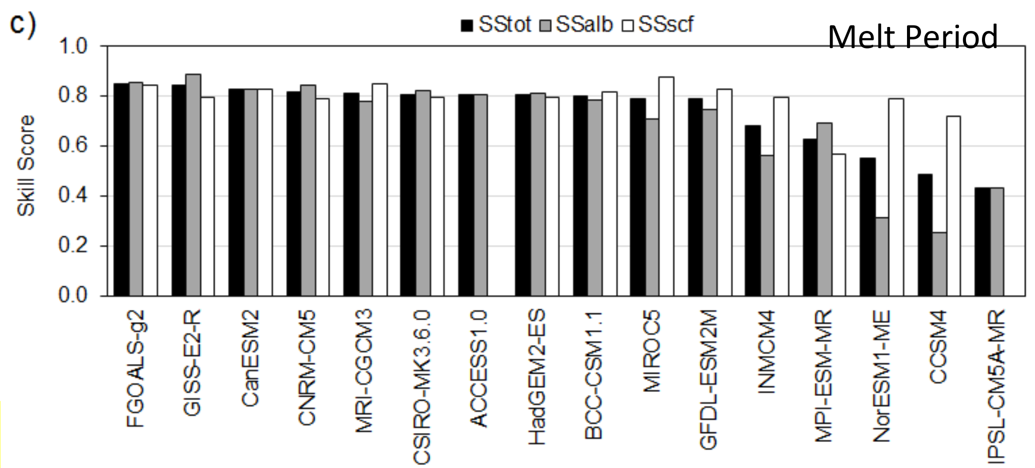
$$SS_{tot} = (SS_{alb} + SS_{scf}) \; / \; 2$$

Thackeray et al. 2015

Boreal Forest Region

- Scores are calculated for two different time periods across the boreal forest: snow season (top) and melt season (bottom).
- The top plot point out models with poor maximum albedo, while the bottom highlights issues with the timing of melt.
- Very low albedo scores in some cases over this region because of complex canopy snow processes.

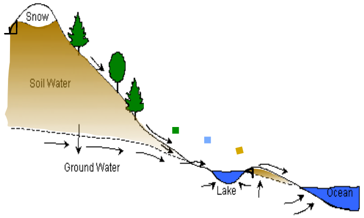


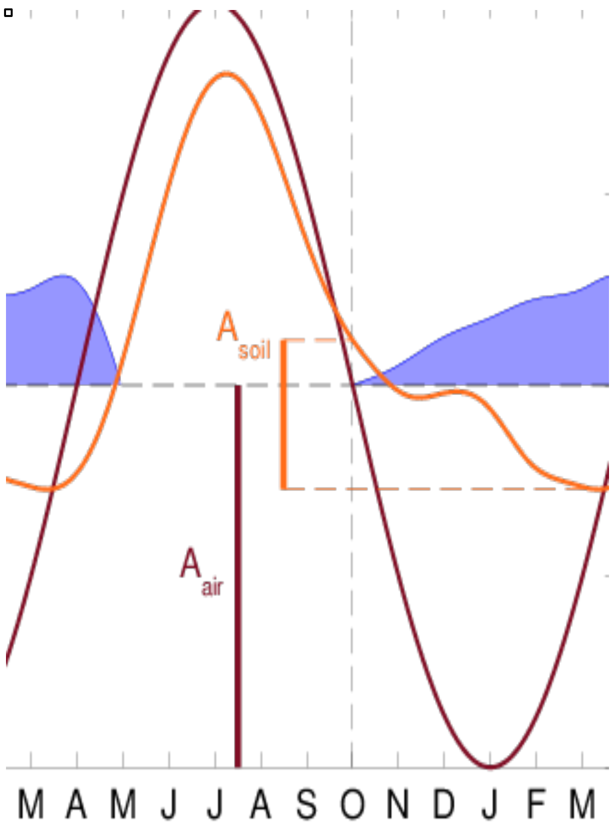


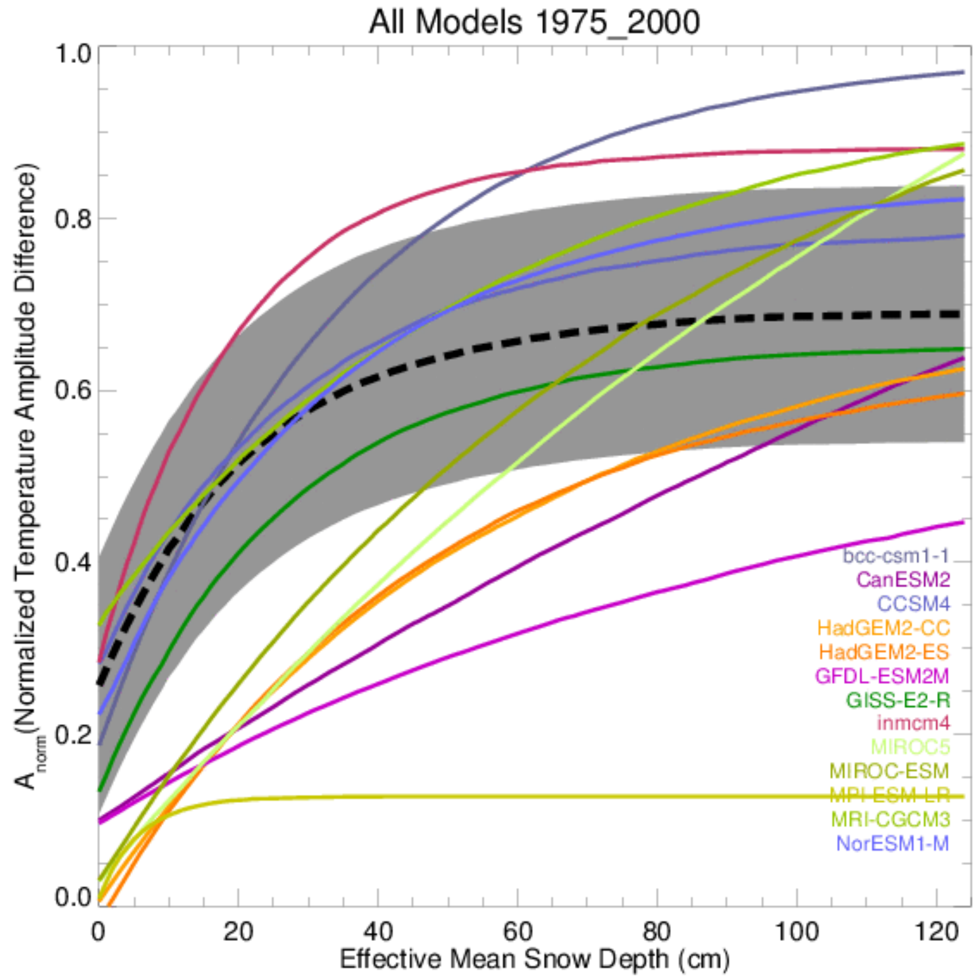
Clustered bar plots of boreal skill scores in descending order of SStot for Sep-Jun (top) and Jan-Jun (bottom).

Thackeray et al. 2015

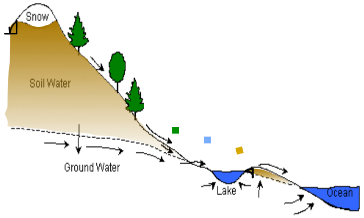
Process-oriented metrics: heat transfer through snow

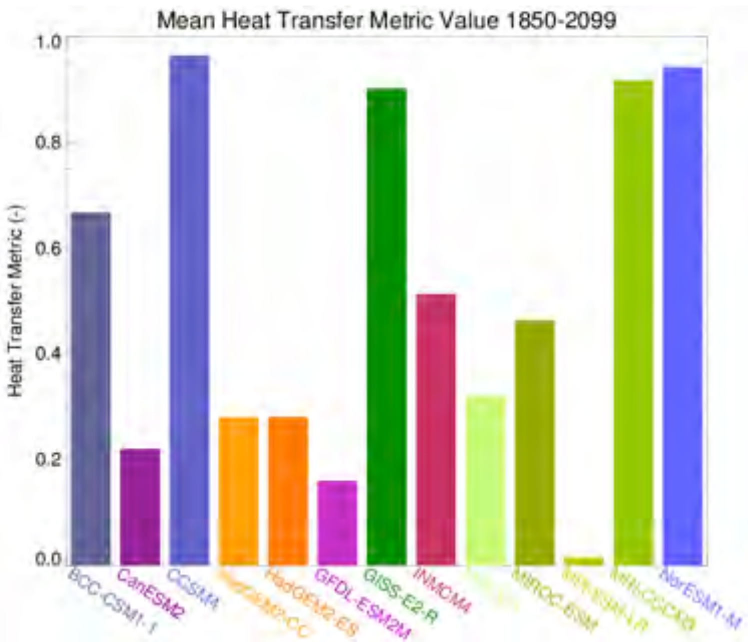


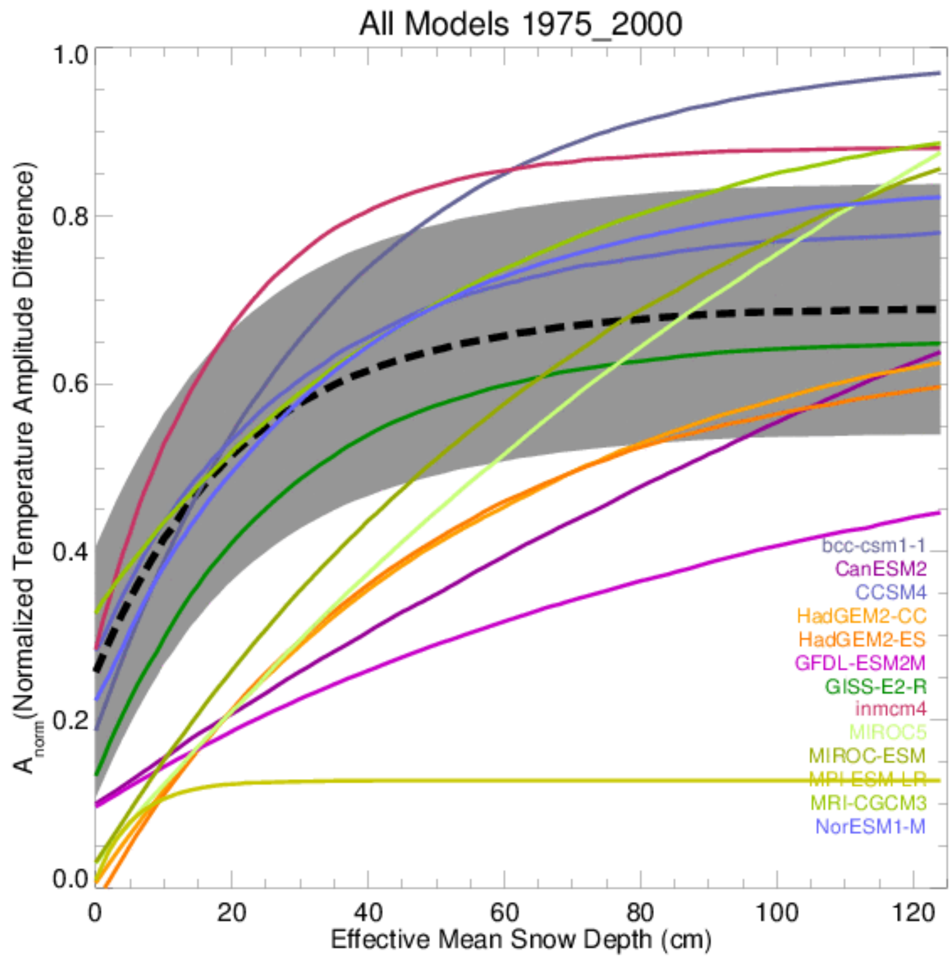




Process-oriented metrics: heat transfer through snow

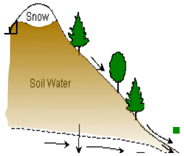






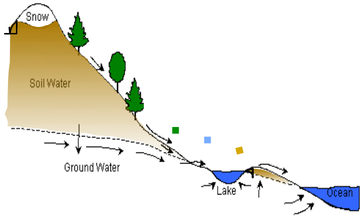
Next steps

- Papers
  - Overview paper assessed against CMIP5 (Randerson)
  - CLM5 overview paper (CLM4, CLM4.5, CLM5, multiple forcing datasets)
- Future development of ILAMB to enhance utility in model development
  - New datasets (WECANN including trans frac, snow, HR, river discharge etc)
  - Diurnal cycle
  - Land-atmosphere coupling metrics
  - Experimental manipulations (N-addition, rainfall exclusion, etc)
  - Develop and integrate arctic and tropical ecosystems modules
  - Emergent constraints
  - Prepare for CMIP6



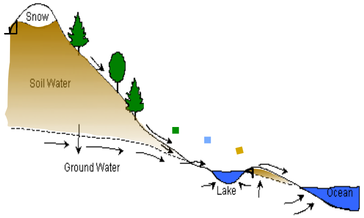
Potential metrics for inclusion in a comprehensive land benchmarking/metric system

- Large-scale state and flux estimates
  - LH, SH, total water storage, albedo, river discharge, SCF, LAI, soil and veg C stocks, GPP, NEE, ER, burnt area, permafrost distribution, T_{2m}, P, …
  - RMSE, spatial pattern corr, interannual variance, annual cycle phase, trends
- Functional relationships and emergent properties
  - soil moisture – ET, soil moisture – runoff, precip – GPP, stomatal response to VPD, precip – burnt area, transient carbon storage trajectory, runoff ratio, spring albedo transition
- Experimental manipulation (testing model functional responses)
  - Nitrogen additions, FACE, artificial warming, rainfall exclusion, ecosystem response to disturbance

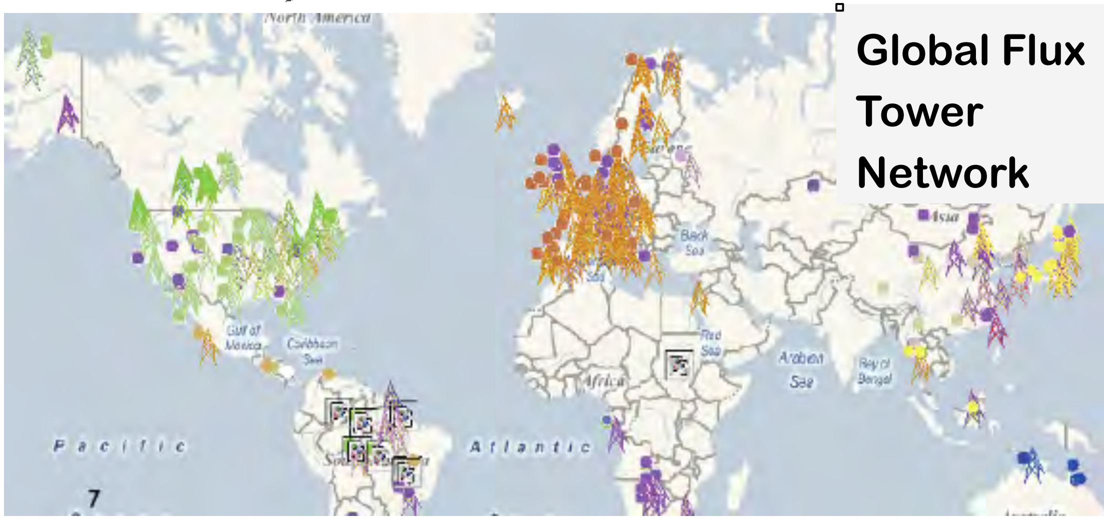


Summary

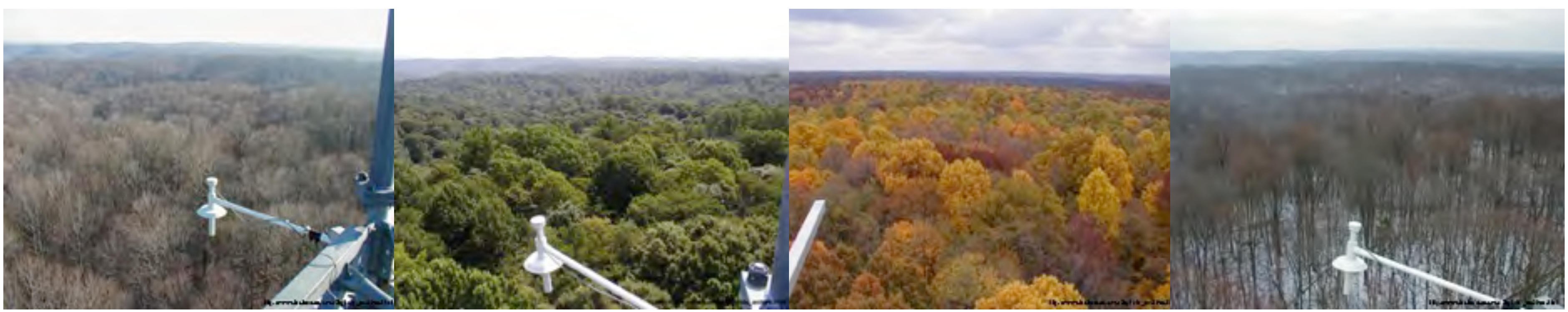
- ILAMB useful tool for model development and assessment
  - Along with tower site simulations, other diagnostics packages, scientific insight and intuition, case studies, etc.
- Provides quick and comprehensive comparison against growing set of observations and metrics
- Future development of ILAMB to enhance utility in model development
  - Parallelization
  - Compare against years outside observational period (e.g. 1850 control)

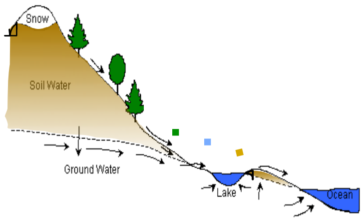


Evaluating and Improving the model with Tower Flux data

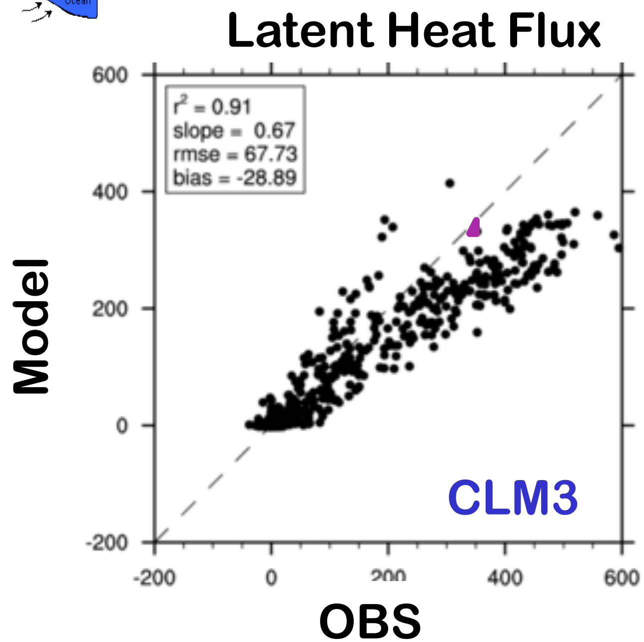


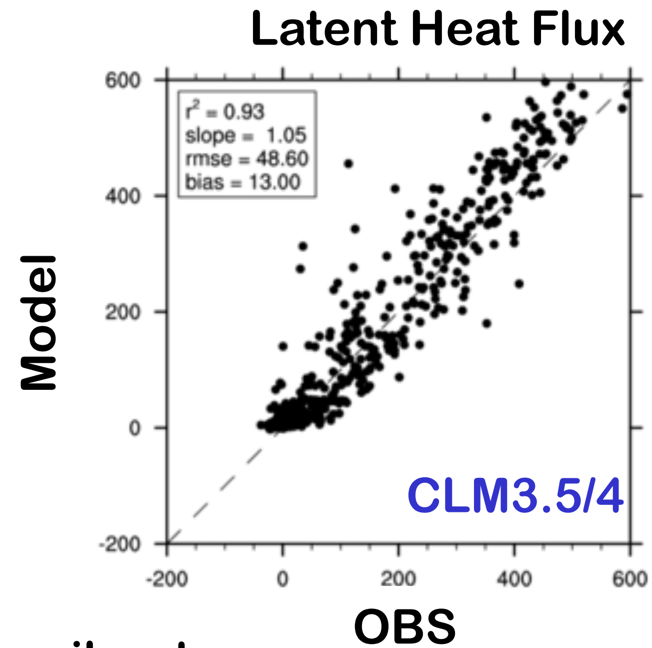
Global Flux Tower Network

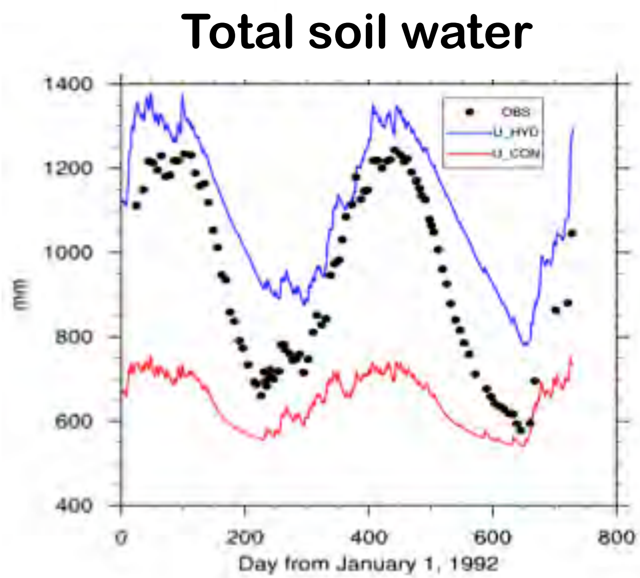




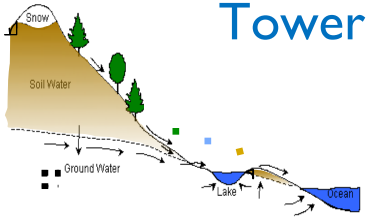
Abracos tower site (Amazon)





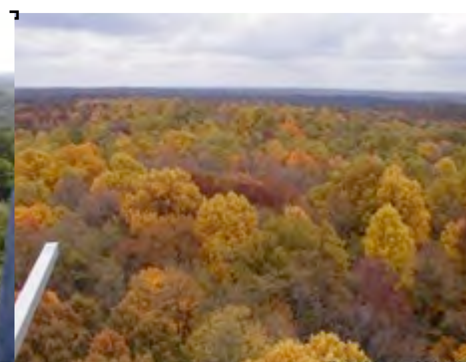


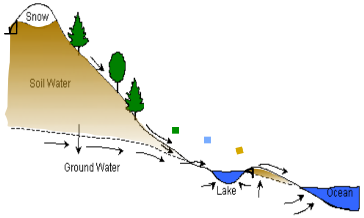
- CLM3.5/4
- CLM3



Tower flux statistics (15 sites incl. tropical, boreal, mediterannean, alpine, temperate; hourly)

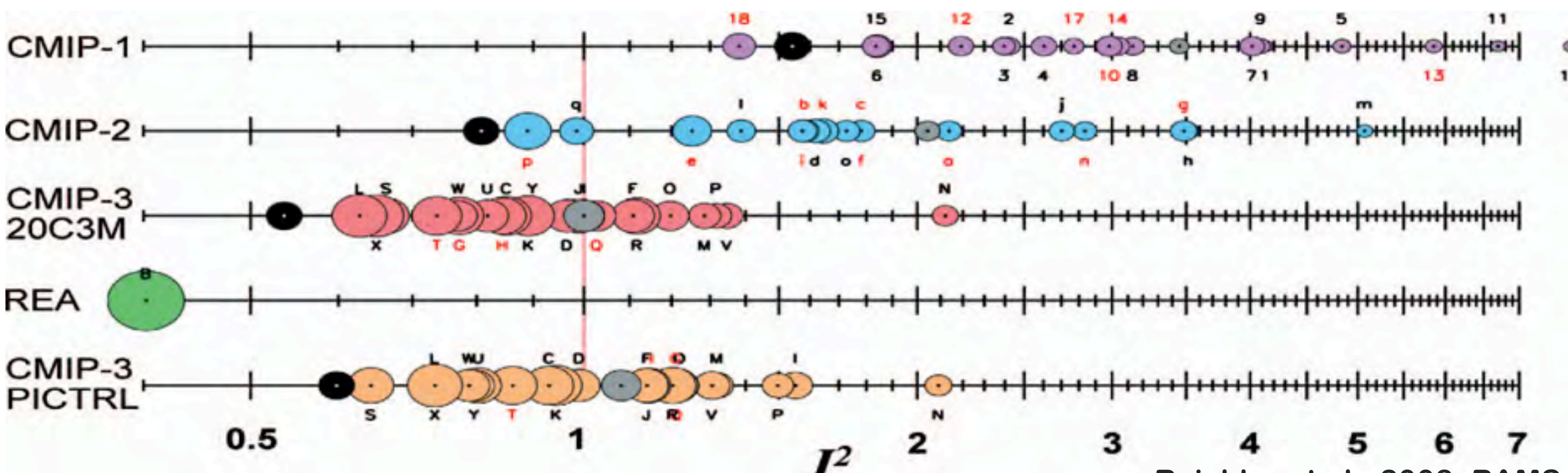
	Latent Heat Flux		Sensible Heat Flux	
	r	RMSE (W/m²)	r	RMSE (W/m²)
CLM3	0.54	72	0.73	91
CLM3.5	**0.80**	**50**	0.79	65
CLM4SP	**0.80**	**48**	**0.84**	**58**





ILAMB Goals

- Develop benchmarks/metrics for land model performance, with emphasis on breadth (carbon cycle, ecosystem, surface energy, and hydrological processes)
- Support the design and development of a new, open-source, benchmarking software system for diagnostics and MIPs
- Strengthen linkages between experimental, monitoring, remote sensing, and climate modeling communities in design of model tests and new measurement programs



Reichler et al., 2008, BAMS