

International Land Model Benchmarking (ILAMB) Project

**Forrest M. Hoffman^{†‡}, Pierre Friedlingstein^{*},
Chris D. Jones[§], and James T. Randerson[†]**

[†]University of California - Irvine, [‡]Oak Ridge National Laboratory,
^{*}University of Exeter, and [§]UKMO Hadley Centre

May 19, 2011

**Advancing Land-Use Modeling and
Analysis for Carbon Cycle Studies**

Princeton University, Princeton, New Jersey, USA



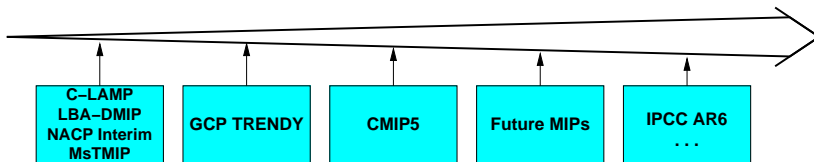
ILAMB Goals

- Develop benchmarks for land model performance, with a focus on carbon cycle, ecosystem, surface energy, and hydrological processes. The benchmarks should be designed and accepted by the community.
- Apply the benchmarks to global models.
- Support the design and development of a new, open-source, benchmarking software system for either diagnostic or model intercomparison purposes.
- Strengthen linkages between experimental, monitoring, remote sensing, and climate modeling communities in the design of new model tests and new measurement programs.

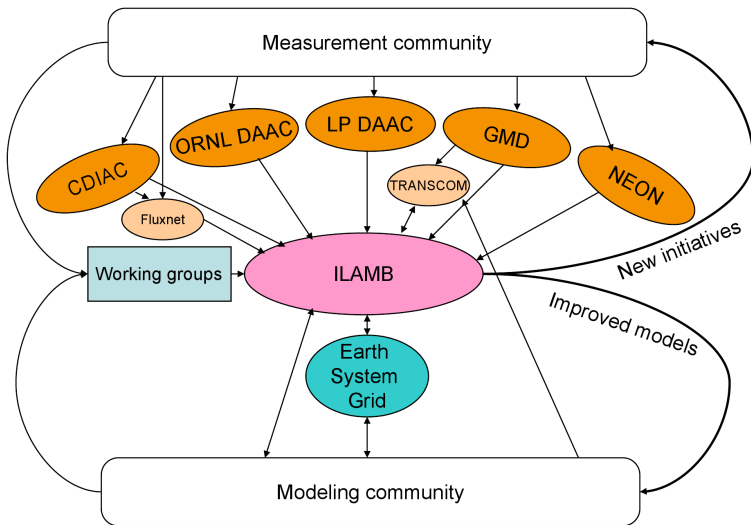
Why Benchmark?

- to show the broader science community and the public that the representation of the carbon cycle in climate models is improving;
- to provide a means, in Earth System models, to quantitatively diagnose impacts of model development in related fields on carbon cycle and land surface processes;
- to guide synthesis efforts, such as the Intergovernmental Panel on Climate Change (IPCC), in the review of mechanisms of global change in models that are broadly consistent with available contemporary observations;
- to increase scrutiny of key datasets used for model evaluation;
- to identify gaps in existing observations needed for model validation;
- to provide a quantitative, application-specific set of minimum criteria for participation in model intercomparison projects (MIPs);
- to provide an optional weighting system for multi-model mean estimates of future changes in the carbon cycle.

An Open Source Benchmarking Software System



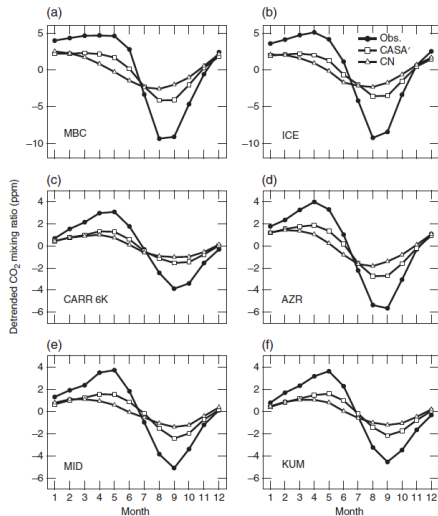
- Human capital costs of making rigorous model-data comparisons is considerable and constrains the scope of individual MIPs.
- Many MIPs spend resources “reinventing the wheel” in terms of variable naming conventions, model simulation protocols, and analysis software.
- **Need for ILAMB:** Each new MIP has access to the model-data comparison modules from past MIPs through ILAMB (e.g., MIPs use one common modular software system). Standardized international naming conventions also increase MIP efficiency.



International Land Model Benchmarking project and diagnostic system

What is a Benchmark?

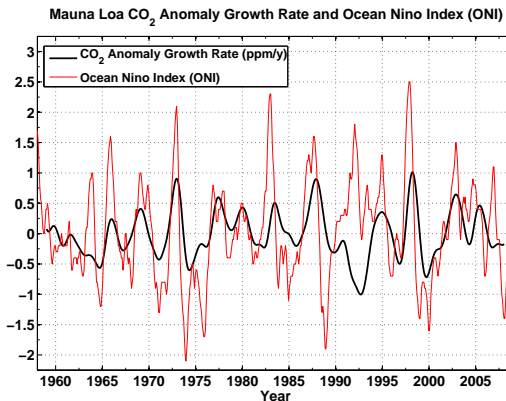
- A benchmark is a quantitative test of model function, for which the uncertainties associated with the observations can be quantified.
- Acceptable performance on benchmarks **is a necessary but not sufficient condition** for a fully functioning model.
- Since all datasets have strengths and weaknesses, an effective benchmark is one that draws upon a broad set of independent observations to evaluate model performance on multiple temporal and spatial scales.



From Randerson et al. (2009)

Example Benchmark – *Interannual to Decadal Time Scale*

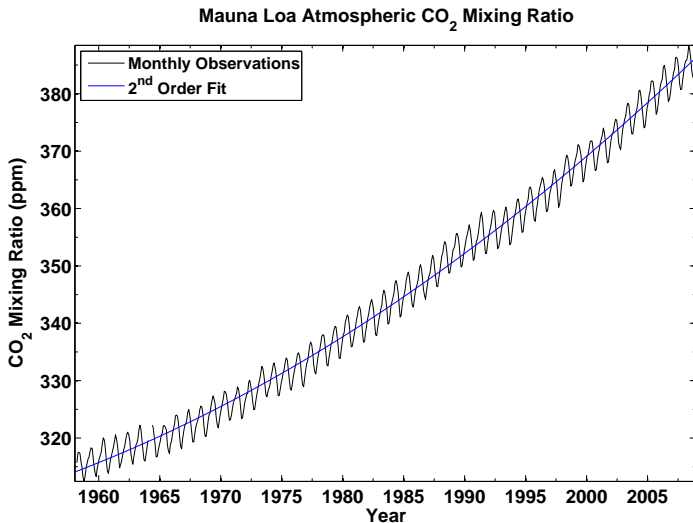
The relationship between El Niño-Southern Oscillation (ENSO) and observed CO₂ anomalies at Mauna Loa may be exploited to evaluate ocean and terrestrial model responses.



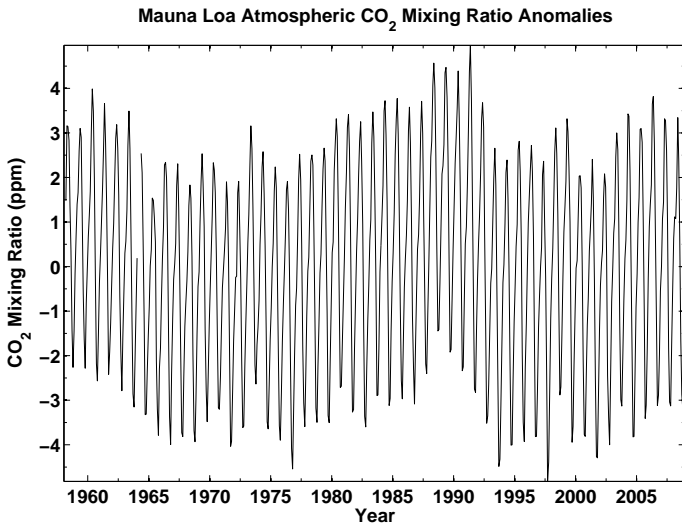
CO₂ Dependence on El Niño-Southern Oscillation (ENSO)

- Keeling and Revelle (1985) described a shutdown in upwelling and biological activity during El Niño years, resulting in a shutdown of CO₂ out-gassing.
- Many others have confirmed this response, including Rayner et al., Feeley et al., Baker et al., and others.
- They suggested the deficiency in CO₂ flux is more than compensated for by widespread forest fires and plant deaths due to drought.
- While the net effect of *natural* processes may once have been a sink, the opposite effect is observed today.
- Opportunistic burning for forest clearing is likely to strengthen the sensitivity of CO₂ to El Niño.

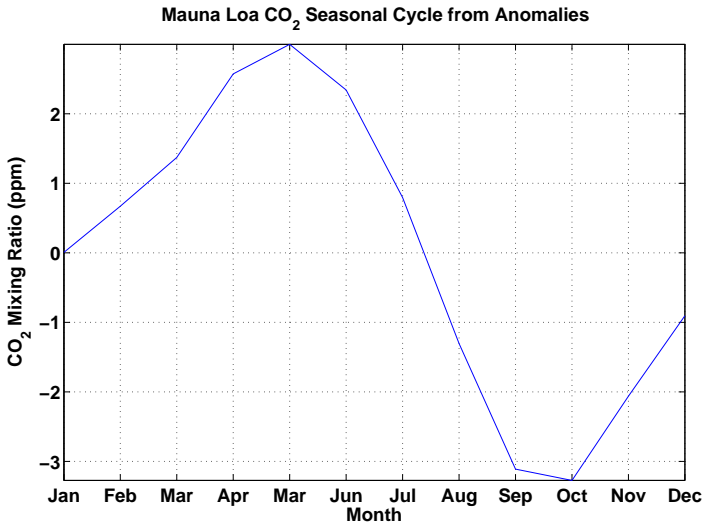
Mauna Loa CO₂ (1957–2008) and Polynomial Curve Fit



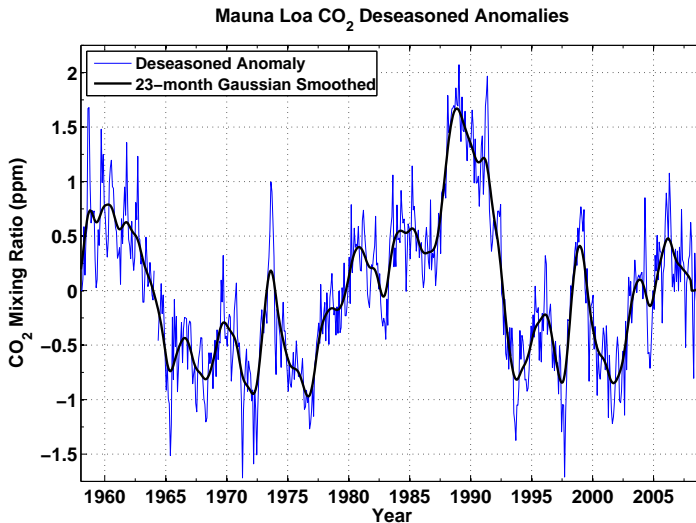
Mauna Loa CO₂ (1957–2008) Minus the Trend



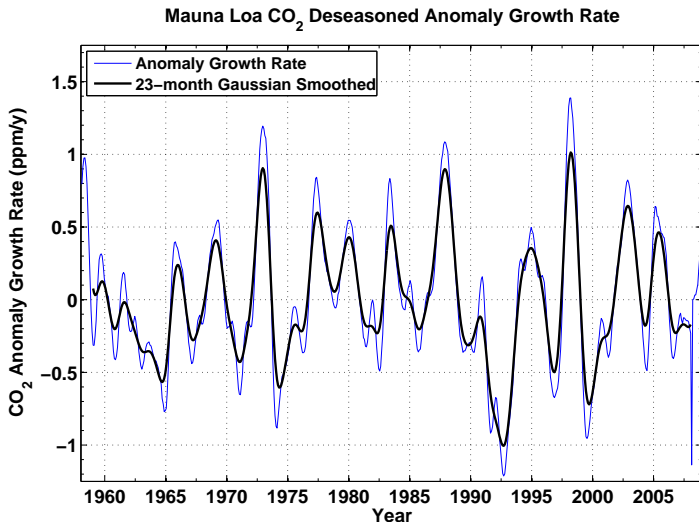
Mauna Loa CO₂ (1957–2008) Mean Seasonal Cycle



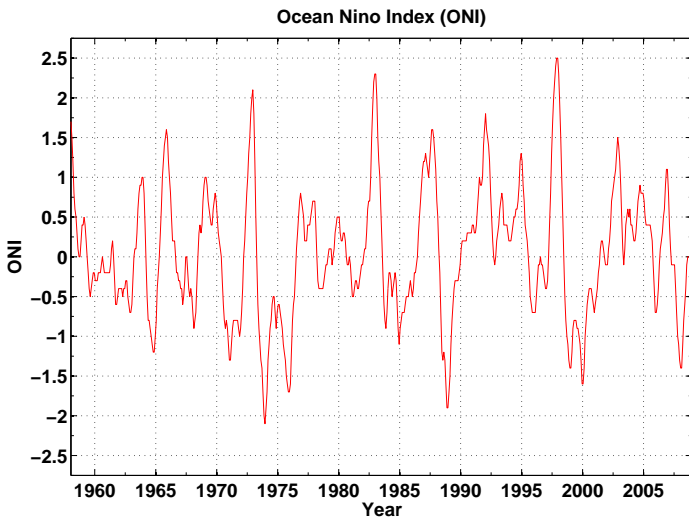
Mauna Loa CO₂ (1957–2008) Deseasoned Anomalies



Mauna Loa CO₂ (1957–2008) Anomaly Growth Rate

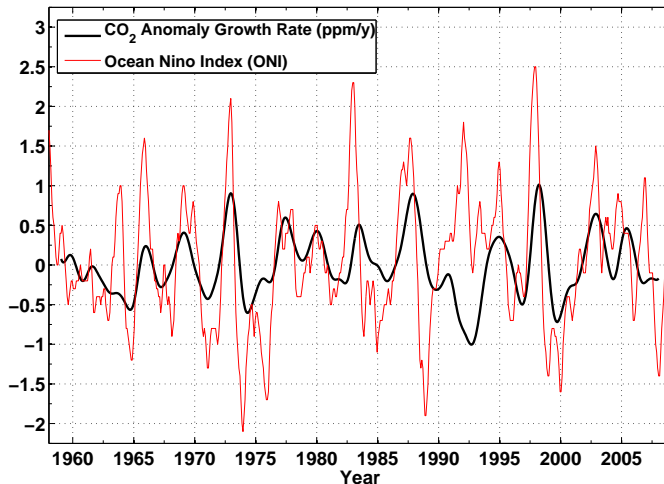


Ocean Niño Index (ONI)



CO₂ Anomaly Growth Rate and Ocean Niño Index

Mauna Loa CO₂ Anomaly Growth Rate and Ocean Niño Index (ONI)

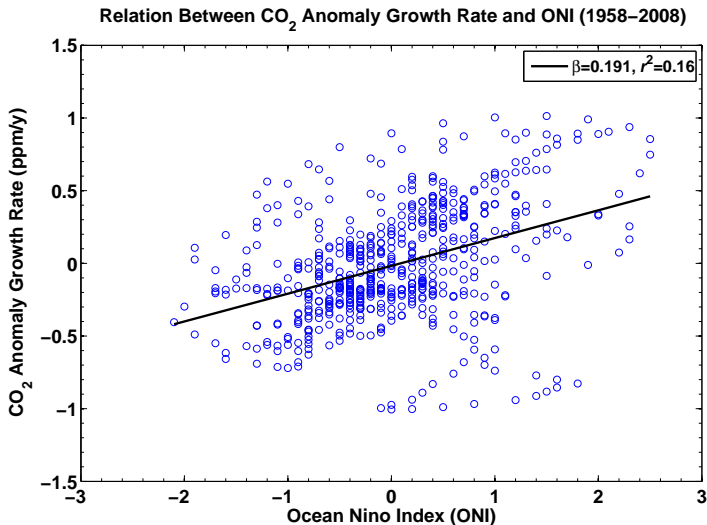


Mount Pinatubo Eruption

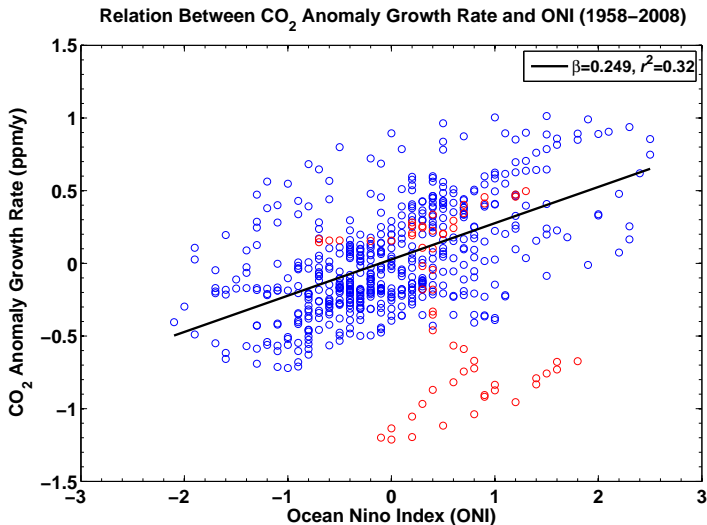
- June 1991 on island of Luzon in the Philippines
- Second largest volcanic eruption of 20th century
- Millions of tons of sulfur dioxide discharged into atmosphere
- Gases and ash reached 34 km high and over 400 km wide
- Largest disturbance of stratosphere since Krakatau in 1883



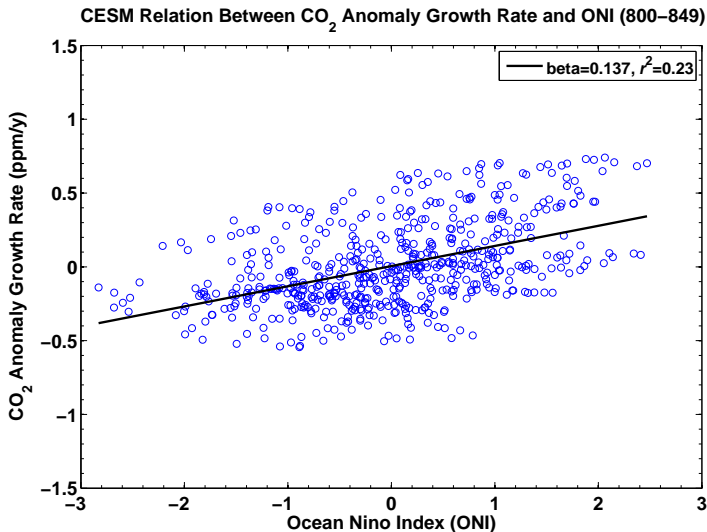
Relation Between CO₂ Anomaly Growth Rate and ONI



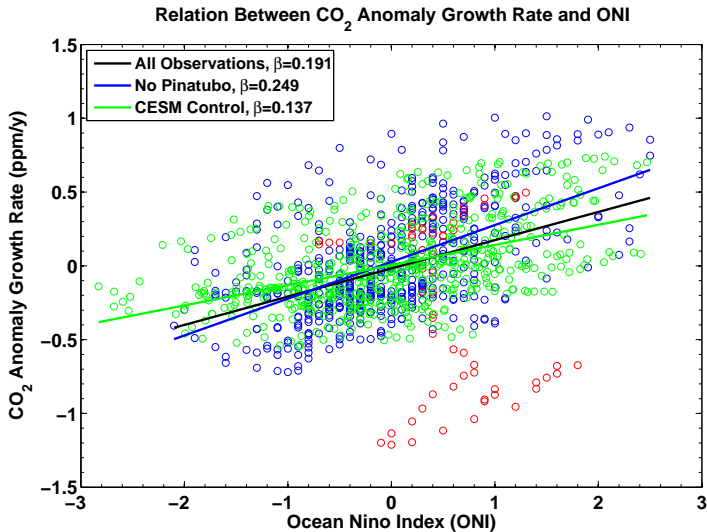
Relation Without 1991–1995 (Pinatubo Period)



Community Earth System Model (CESM) Control Run



CESM vs. Observations



Benchmark Conclusions

- Relationship between Mauna Loa CO₂ anomalies and El Niño are strongly related, except during intervening events.
- Models should capture this relationship *for the right reasons*, so this may be a useful metric for model evaluation.
- More broadly, atmospheric CO₂ is an integrator of terrestrial and ocean fluxes with valuable information for constraining model behavior over a wide range of time scales (see also Cadule et al., 2010).
- For this analysis, time-lag correlation may improve the fit and yield a more accurate slope.
- This slope may change over time as humans exploit El Niño-induced drought for tropical forest clearing.
- The CESM control run does a reasonable job of capturing this relationship.



International Land Model Benchmarking (ILAMB) Meeting The Beckman Center, Irvine, CA, USA January 24-26, 2011



DEPARTMENT OF EARTH SYSTEM SCIENCE
SCHOOL OF PHYSICAL SCIENCES
UNIVERSITY OF CALIFORNIA • IRVINE

- Meeting Co-organized by Forrest Hoffman (UC-Irvine and ORNL), Chris Jones (UK Met Office Hadley Centre), Pierre Friedlingstein (U. Exeter), and Jim Randerson (UC-Irvine).
- About 45 researchers participated from the United States, Canada, the United Kingdom, the Netherlands, France, Germany, Switzerland, China, Japan, and Australia.

ILAMB Meeting Goals

- Design the first set of ILAMB benchmarks for global models.
 - How many flavors (carbon cycle, LUC, hydrology, ...)?
 - What datasets do we include?
 - What graphics and cost functions?
- Coordinate carbon cycle and land model evaluation analyses for TRENDY and CMIP5 results.
- Develop an implementation plan for application of the ILAMB 1.0 benchmarks to TRENDY and CMIP5 output over next year.
- Decide upon the approach for developing ILAMB code.
 - netCDF for datasets? Language for evaluation code?
 - Need to extend variable naming conventions beyond CMIP5.
- Decide upon a future schedule and means to secure funding.
 - Key deadline is July 2012 for submission of manuscripts for IPCC AR5 Working Group 1.
 - Should ILAMB meet once a year until AR6?

Example Benchmark Score Sheet from C-LAMP

Models →

BGC Datasets ↓

Metric	Metric components	Uncertainty of obs.	Scaling mismatch	Total score	Sub-score	CASA'	CN
LAI	Matching MODIS observations			15.0		13.5	12.0
	• Phase (assessed using the month of maximum LAI)	Low	Low		6.0	5.1	4.2
	• Maximum (derived separately for major biome classes)	Moderate	Low		5.0	4.6	4.3
	• Mean (derived separately for major biome classes)	Moderate	Low		4.0	3.8	3.5
NPP	Comparisons with field observations and satellite products			10.0		8.0	8.2
	• Matching EMDI Net Primary Production observations	High	High		2.0	1.5	1.6
	• EMDI comparison, normalized by precipitation	Moderate	Moderate		4.0	3.0	3.4
	• Correlation with MODIS (r^2)	High	Low		2.0	1.6	1.4
	• Latitudinal profile comparison with MODIS (r^2)	High	Low		2.0	1.9	1.8
CO ₂ annual cycle	Matching phase and amplitude at Globalview flash sites			15.0		10.4	7.7
	• 60°–90°N	Low	Low		6.0	4.1	2.8
	• 30°–60°N	Low	Low		6.0	4.2	3.2
	• 0°–30°N	Moderate	Low		3.0	2.1	1.7
Energy & CO ₂ fluxes	Matching eddy covariance monthly mean observations			30.0		17.2	16.6
	• Net ecosystem exchange	Low	High		6.0	2.5	2.1
	• Gross primary production	Moderate	Moderate		6.0	3.4	3.5
	• Latent heat	Low	Moderate		9.0	6.4	6.4
	• Sensible heat	Low	Moderate		9.0	4.9	4.6
Transient dynamics	Evaluating model processes that regulate carbon exchange on decadal to century timescales			30.0		16.8	13.8
	• Aboveground live biomass within the Amazon Basin	Moderate	Moderate		10.0	5.3	5.0
	• Sensitivity of NPP to elevated levels of CO ₂ : comparison to temperate forest FACE sites	Low	Moderate		10.0	7.9	4.1
	• Interannual variability of global carbon fluxes: comparison with TRANSCOM	High	Low		5.0	3.6	3.0
	• Regional and global fire emissions: comparison to GFEDv2	High	Low		5.0	0.0	1.7
				Total:	100.0	65.9	58.3

From Randerson et al. (2009)

	Annual Mean	Seasonal Cycle	Interannual Variability	Trend	Data Source
Atmospheric CO₂					
Flask/conc. + transport		✓	✓	✓	NOAA, SIO, CSIRO
TCCON + transport		✓	✓	✓	Caltech
Fluxnet					
GPP, NEE, TER, LE, H, RN	✓	✓	✓		Fluxnet, MAST-DC
Gridded: GPP	✓	✓	?		MPI-BGC
Hydrology/Energy					
river flow	✓		✓		GRDC, Dai, GFDL
global runoff/ocean balance	✓				Syed/Famiglietti
albedo (multi-band)		✓	✓		MODIS, CERES
soil moisture		✓	✓		de Jeur, SMAP
column water		✓	✓		GRACE
snow cover	✓	✓	✓	✓	AVHRR, GlobSnow
snow depth/SWE	✓	✓	✓	✓	CMC (N. America)
T _{air} & P	✓	✓	✓	✓	CRU, GPCP and TRMM
Gridded: LE, H	✓	✓			MPI-BGC, dedicated ET
Ecosystem Processes & State					
soil C, N	✓				HWSD, MPI-BGC
litter C, N	✓				LIDET
soil respiration	✓	?	✓	✓	Bond-Lamberty
FAPAR	✓	✓			MODIS, SeaWiFS
biomass & change	✓			✓	Saatchi, Pan, Blackard
canopy height	✓				Lefsky, Fisher
NPP	✓				EMDI, Luyssaert
Vegetation Dynamics					
fire — burned area	✓	✓	✓		GFED3
wood harvest	✓			✓	Hurt
land cover	✓				MODIS PFT fraction

Meeting Summary

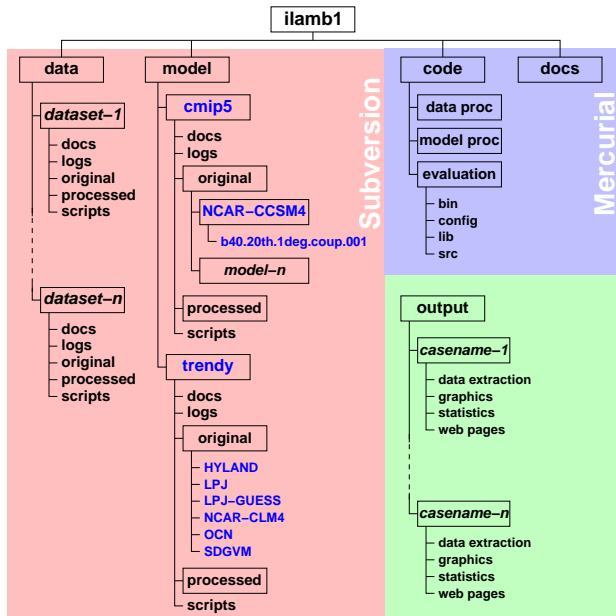
- Five break-out groups met, one for each benchmark category, to identify cost function metrics and graphics.
- Measurement and model uncertainty must be characterized and spatial scaling mismatch considered for evaluation.
- Key objectives are to use publicly available data and freely available software.
- The R package will be used for generating statistical results and diagnostics.
- Five initial benchmarks will be implemented to evaluate existing TRENDY and CMIP5 model results.



A team was identified to begin software architecture design.

A developmental hierarchy for data, model results, code, and docs is established.

Server-based and distributed version control systems will be used for handling data and code, respectively.



Next Steps

- Common model output
 - A draft document proposing additional new netCDF Climate and Forecast (CF) conventions, beyond those created for CMIP5, is available for comment.
 - To assist the modeling community, a translator between ALMA and CF standards may be created.
- Future: New protocols and forcing data comparisons.
- A development Wiki is coming soon.
- ILAMB Town Hall meeting at AGU in December?
- Shilong will host the next ILAMB meeting in Beijing, China, in February–March 2012.

International Land Model Benchmarking (ILAMB) Project

<http://www.ilamb.org/>

References

- P. Cadule, P. Friedlingstein, L. Bopp, S. Sitch, C. D. Jones, P. Ciais, S. L. Piao, and P. Peylin. Benchmarking coupled climate-carbon models against long-term atmospheric CO₂ measurements. *Global Biogeochem. Cycles*, 24(2):GB2016, Oct. 2010.
doi:10.1029/2009GB003556.
- J. T. Randerson, F. M. Hoffman, P. E. Thornton, N. M. Mahowald, K. Lindsay, Y.-H. Lee, C. D. Nevison, S. C. Doney, G. Bonan, R. Stöckli, C. Covey, S. W. Running, and I. Y. Fung. Systematic assessment of terrestrial biogeochemistry in coupled climate-carbon models. *Global Change Biol.*, 15(9):2462–2484, Sept. 2009.
doi:10.1111/j.1365-2486.2009.01912.x.