Using Remotely-sensed Data Sets for Model Evaluation and Benchmarking

Forrest M. Hoffman[†‡], James T. Randerson[‡], and Jiafu Mao[†]

[†]Oak Ridge National Laboratory and [‡]University of California - Irvine

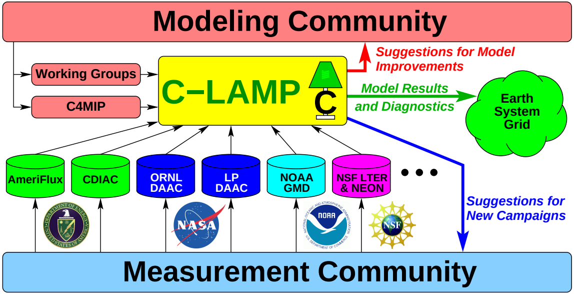
September 14, 2012

ForestSAT 2012 • September 12–14, 2012
Oregon State University, Corvallis, Oregon, USA

Challenges to Using Remote Sensing for Model Evaluation

- Satellites do not measure the properties that land surface models predict.
- Land surface models do not calculate what satellites measure.
- Ecosystem models used to generate satellite products may be similar to those employed by land surface models.
- Fixed timing of satellite overpasses and presence of snow, clouds, and aerosols may bias observations.
- Spatial mismatches between sensors and models are unavoidable and *scale matters*.
- Uncertainties are rarely (usefully) characterized for measurements or models.
- Spatial and temporal averaging may further limit the utility of remote sensing products in constraining models.

- The **Carbon-Land Model Intercomparison Project (C-LAMP)** began as a CCSM Biogeochemistry Working Group project to assess model capabilities in the coupled climate system and to explore processes important for inclusion in the CCSM4 Earth System Model for use in the IPCC Fifth Assessment Report (AR5).
- Unlike traditional MIPs, C-LAMP was designed to confront models with best-available observational datasets, develop metrics for evaluation of biosphere models, and build a general-purpose biogeochemistry diagnostics package for model evaluation.



Model Configurations

- Biosphere models coupled to the Community Climate System Model version 3.1
  - CLM3-CASA′ — Carnegie/Ames/Stanford Approach Model previously run in CSM1.4 (Fung)
  - CLM3-CN — coupled carbon and nitrogen cycles based on the Biome-BGC model (Thornton)
- CCSM3.1 partially coupled ("I" & "F" configurations) run at T42 resolution ($\sim 2.8^\circ \times 2.8^\circ$), spectral Eulerian dycore, $1^\circ \times 0.27^\circ$–$0.53^\circ$ ocean & sea ice data models (T42gx1v3).

C-LAMP Protocol Overview

- Experiment 1: Models forced with an improved NCEP/NCAR reanalysis climate data set (Qian, *et al.* 2006) to examine the influence of climate variability, prescribed atmospheric CO_2, and land cover change on terrestrial carbon fluxes during the 20th century (specifically 1948–2004).
- Experiment 2: Models coupled with an active atmosphere (CAM3), prescribed atmospheric CO_2, prescribed sea surface temperatures and ocean carbon fluxes to examine the effect of a coupled biosphere-atmosphere for carbon fluxes and climate during the 20th century.
- All the forcing and observational datasets are being shared, and model results are available through the Earth System Grid (ESG), just like for CMIP3 (the IPCC AR4 model results).
- Experimental protocol, output fields, and metrics are available at http://www.climatemodeling.org/c-lamp/

Offline Forcing with NCEP/NCAR Reanalysis		
Exp.	**Description**	**Time Period**
1.1	Spin Up	~4,000 y
1.2	Control	1798–2004
1.3	Varying climate	1948–2004
1.4	Varying climate, CO_2, and N deposition	1798–2004
1.5	Varying climate, CO_2, N deposition and land use	1798–2004
1.6	Free Air CO_2 Enrichment (FACE) Control	1997–2100
1.7	Free Air CO_2 Enrichment (FACE) Transient	1997–2100

Coupled Land-Atmosphere Forcing with Hadley SSTs		
Exp.	**Description**	**Time Period**
2.1	Spin Up	~2,600 y
2.2	Control	1800–2004
2.3	Varying climate	1800–2004
2.4	Varying climate, CO_2, and N deposition	1800–2004
2.5	Varying climate, CO_2, N deposition and land use	1800–2004
2.6	Varying climate, CO_2, N deposition, seasonal FFE	1800–2004

All but the land use experiments were run with CCSM3.1 using CLM3-CASA′ and CLM3-CN biogeochemistry models yielding >16,000 y and ~50 TB

C-LAMP Common Model Output - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.climatemodeling.org/c-lamp/protocol/model_output.php

C-LAMP Common Model Output

While all models participating in the Carbon Land Model intercomparison Project (C-LAMP) will output their own "native" fields, a common set of fields is needed to facilitate head-to-head comparison of the models to each other and to available observational datasets. Model results transmitted to the Earth System Grid for redistribution to the community will use common field names, netCDF long names, CF Standard Names and units. Contained below is a table of the common output fields required for the C-LAMP and consistent with the metadata conventions used for CMIP3, formerly called the IPCC 4th Assessment Model Output database. Corrections and suggestions are solicited on this information. Software is available for rewriting model output into netCDF files following the Climate and Forecast (CF) Metadata Convention.

Version 2.1 - Aug 30, 2008

Atmospheric forcing				
Variable Name	**Long Name and CF Standard Name**	**Units**	**Comment**	**Statistics**
husf	Specific humidity at atmospheric forcing height specific_humidity‡	kg kg-1		MHM, MHS, MM
prra	Rainfall precipitation flux rainfall_flux‡	kg m-2 s-1	Rainfall includes all liquid types (rain, large-scale, convective, etc.)	MHM, MHS, MM
prsn†	Snowfall precipitation flux snowfall_flux‡	kg m-2 s-1	Snowfall includes all frozen types (snow, hail, ice, etc.)	MHM, MHS, MM
Biogeochemistry				
Variable Name	**Long Name and CF Standard Name**	**Units**	**Comment**	**Statistics**
agbc*	Above-ground biomass carbon above_ground_biomass_carbon_content	kg m-2	Total carbon content in above-ground live and dead carbon pool(s)	MM
aglbc*	Above-ground live biomass carbon above_ground_live_biomass_carbon_content	kg m-2	Total carbon content in above-ground live carbon pool(s)	MM
agnpp	Above-ground net primary production above_ground_net_primary_productivity_of_carbon	kg m-2 s-1	Component of net primary production attributable to above-ground live biomass	MM
ar	Autotrophic respiration autotrophic_respiration_of_carbon alias(es): plant_respiration_carbon_flux	kg m-2 s-1	Sum of maintenance respiration and growth respiration of vegetation	MHM, MHS, MM
bco	Biogenic carbon monoxide flux biogenic_carbon_monoxide_flux	kg m-2 s-1	Total biogenic carbon monoxide flux out of biosphere	MM

Done

C-LAMP Performance Metrics and Diagnostics

- An evolving document on metrics for model evaluation is available at http://www.climatemodeling.org/c-lamp/
- Each model is scored with respect to its performance on various output fields compared with best-available observational datasets.
- Examples include:
  - leaf area index (LAI): comparison of phase and spatial distribution using MODIS
  - net primary production (NPP): comparison with EMDI and correlation with MODIS
  - CO_2 seasonal cycle: comparison with NOAA/Globalview flask sites after combining fluxes with impulse response functions from TRANSCOM
  - regional carbon stocks (Saatchi et al., 2007)
  - carbon and energy fluxes (Fluxnet sites)
  - other transient dynamics: β factor, fire emissions

Score Sheet for CLAMP - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.climatemodeling.org/c-lamp/results/diagnostics/CN_vs_CA

C-LAMP Score Sheet for Biogeochemical Model Evaluation

Metric	Metric components	Observations & comparison protocol	Model CASA′	Model CN	Score (points)		
					Possible	CASA′	CN
LAI	MODIS Phase	global map	global map model vs obs	global map model vs obs	6.00	5.11	4.24
	MODIS Maximum	global map	global map model vs obs	global map model vs obs	5.00	4.60	4.26
	MODIS Mean	land class obs land class model global map	model vs obs table global map model vs obs	model vs obs table global map model vs obs	4.00	3.75	3.53
NPP	EMDI NPP observations	Class A table	table scatter plot	table scatter plot	1.00	0.68	0.73
		Class B table	table scatter plot	table scatter plot	1.00	0.83	0.82
	EMDI NPP normalized by PPT	Class A histogram	Class A histogram	Class A histogram	2.00	1.50	1.74
		Class B histogram	Class B histogram	Class B histogram	2.00	1.51	1.65
	Correlation with MODIS	global map	model map model vs obs	model map model vs obs	2.00	1.64	1.44
	Correlation with MODIS-zonal mean	zonal mean obs	zonal mean model vs obs plot	zonal mean model vs obs plot	2.00	1.88	1.84
CO_2 Seasonal Cycle — Comparison with Globalview phase and amplitude	60°N–90°N	—	—	—	6.00	4.11	2.77
	30°N–60°N	—	—	—	6.00	4.23	3.23
	0°N–30°N	—	—	—	3.00	2.07	1.71
Energy and C Fluxes from Fluxnet	NEE	line plot	model vs obs	model vs obs	—	—	—
	Net radiation				—	—	—
	Latent heat				—	—	—
	Sensible heat				—	—	—
Energy and C Fluxes from Ameriflux	NEE	line plot	model vs obs timeseries plot	model vs obs timeseries plot	6.00	2.46	2.13
	Shortwave Incoming				—	—	—
	Latent heat				9.00	6.38	6.39
	Sensible heat				9.00	4.90	4.64
	GPP				6.00	3.39	3.46
	ER				—	—	—
	Aboveground live [illegible]	[illegible]	model amazon	amazon map	[illegible]	[illegible]	[illegible]

Done

Score Sheet for CLAMP - Mozilla Firefox

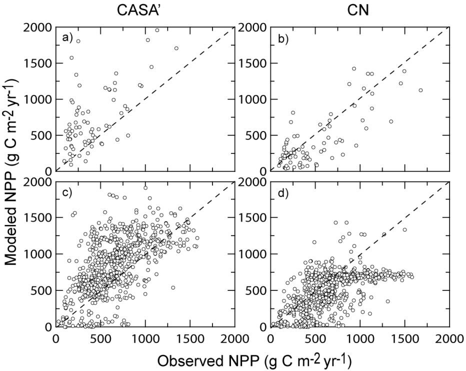
File Edit View History Bookmarks Tools Help

http://www.climatemodeling.org/c-lamp/results/diagnostics/CN_vs_CA

CO_2 Seasonal Cycle — Comparison with Globalview phase and amplitude	60°N–90°N	—	—	—	6.00	4.11	2.77
	30°N–60°N	—	—	—	6.00	4.23	3.23
	0°N–30°N	—	—	—	3.00	2.07	1.71
Energy and C Fluxes from Fluxnet	NEE	line plot	model vs obs	model vs obs	—	—	—
	Net radiation				—	—	—
	Latent heat				—	—	—
	Sensible heat				—	—	—
Energy and C Fluxes from Ameriflux	NEE	line plot	model vs obs timeseries plot	model vs obs timeseries plot	6.00	2.46	2.13
	Shortwave Incoming				—	—	—
	Latent heat				9.00	6.38	6.39
	Sensible heat				9.00	4.90	4.64
	GPP				6.00	3.39	3.46
	ER				—	—	—
Carbon Stocks and Transient Dynamics	Aboveground live biomass in South America	obs amazon	model amazon model vs obs	amazon map model vs obs	10.00	5.28	4.99
	Aboveground live biomass within Amazon Basin (sum within Legal Amazon)	mask obs masked 68.90 (Pg C)	model masked model vs obs 198.87 (Pg C)	model masked model vs obs 160.61 (Pg C)	—	—	—
	NPP Stimulation from elevated CO_2	—	FACE Site table biome table	FACE Site table biome table	10.00	7.87	4.11
	Interannual variability of global carbon fluxes - comparison with TRANSCOM	—	—	—	5.00	3.55	3.00
	Turnover times and pool sizes	—	Leaf Wood Fine Root Litter Coarse Woody Debris Soil	Leaf Wood Fine Root Litter Coarse Woody Debris Soil	—	—	—
	Carbon Sinks (1990–2004)	—	biome mean biome total	biome mean biome total	—	—	—
	Fire Variability (1997–2004)	—	—	global spatial comparison temporal dynamics	5.00	—	1.70
				Total Score	**100.00**	**65.74**	**58.38**

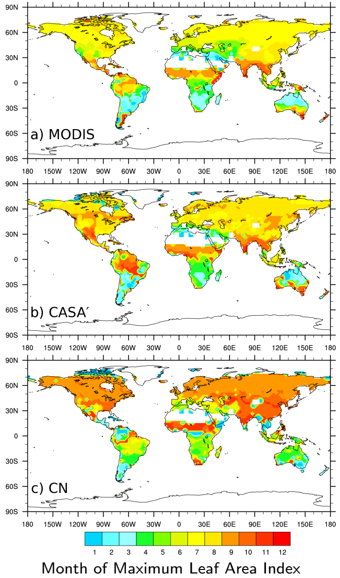
Done

- Comparisons with field observations include net primary production (NPP) from the Ecosystem Model-Data Intercomparison (EMDI).
- Measurements were performed in different ways, at different times, and by different groups for a limited number of field sites.
- Shown here are comparisons of NPP with EMDI Class A observations (Figures a and b) and Class B observations (Figures c and d).



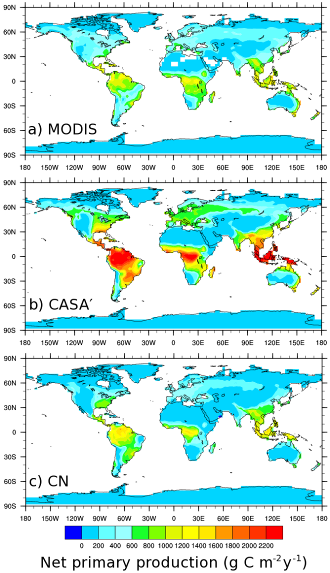
Data provided by NASA Distributed Active Archive Center (DAAC) at ORNL

- Comparisons with satellite "modeled observations" must be made carefully because of high uncertainty.
- This comparison with MODIS leaf area index (LAI) focuses on the month of maximum LAI (phase), a measurement with less uncertainty than the "observed" LAI values.
- C-LAMP accounts for this uncertainty by weighting scores accordingly.
- CLM-CASA′ scored 5.1/6.0 while CLM-CN scored 4.2/6.0 for this metric.

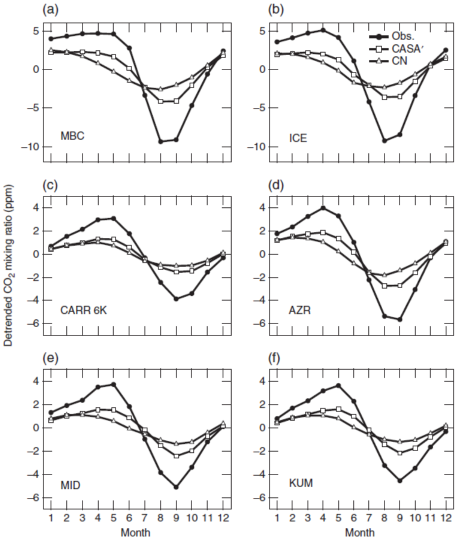


(Myneni et al., 2002; Zhao et al., 2005)

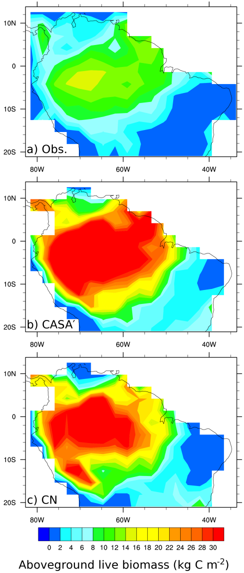
- MODIS net primary production (NPP) "observations" have higher uncertainty.
- Comparison with MODIS NPP (Heinsch et al., 2003) focuses on correlation of spatial patterns.
- CLM-CASA′ scored 1.6/2.0 while CLM-CN scored 1.4/2.0.



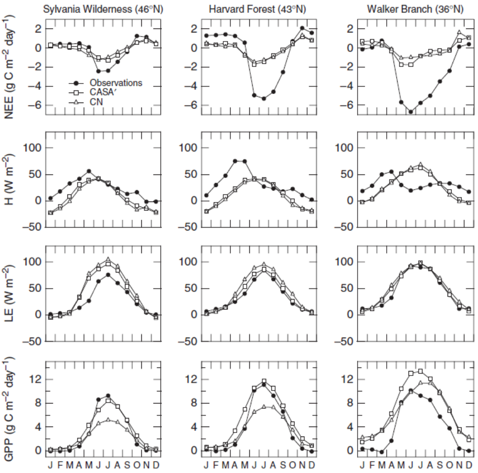
- Comparisons with Globalview flask sites are made by combining model fluxes with impulse response functions from TRANSCOM.
- Shown are the annual cycles of atmospheric CO_2 at (a) Mould Bay, Canada (76°N), (b) Storhofdi, Iceland (63°N), (c) Carr, Colorado (41°N), (d) Azores Islands (39°N), (e) Sand Island, Midway (28°N), and (f) Kumakahi, Hawaii (20°N).
- CLM-CASA′ scored 10.4/15.0 while CLM-CN scored 7.7/15.0 for this metric.



- Estimates of carbon stocks are very difficult to obtain.
- This comparison with estimates of aboveground live biomass in the Amazon by Saatchi et al. (2007) shows that both models are too high by about a factor of 2.
- Using a score based on normalized cell-by-cell differences, CLM-CASA′ scored 5.3/10.0 while CLM-CN scored 5.0/10.0.

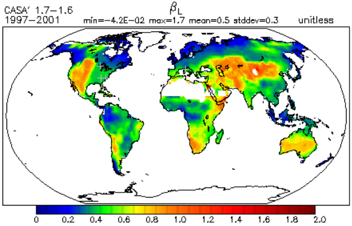


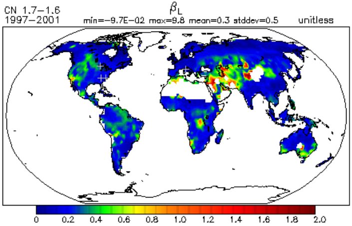
- Comparisons with AmeriFlux eddy correlation CO_2 flux tower sites include net ecosystem exchange (NEE), gross primary production (GPP), respiration, shortwave incoming radiation, and latent and sensible heat.
- Shown here is a comparison of model estimates with eddy covariance measurements from Sylvania Wilderness, Harvard Forest, and Walker Branch.
- Used are the consistent Level 4 data.

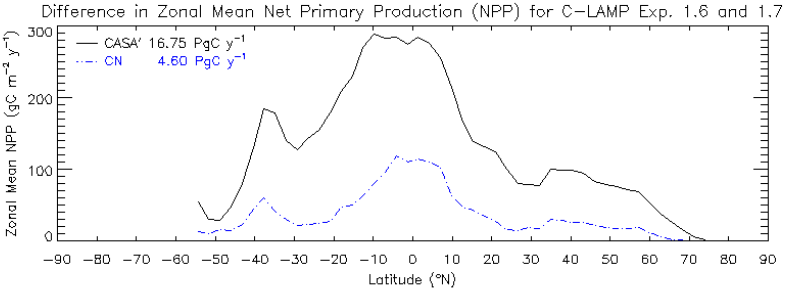


Data provided by ORNL Carbon Dioxide Information Analysis Center (CDIAC).

- Additional field measurement comparisons include the Free Air CO_2 Enrichment (FACE) results, including the ORNL site.
- The Norby *et al.* (2005) synthesis of four FACE site observations suggested "response of forest NPP to elevated $[CO_2]$ is highly conserved across a broad range of productivity, with a stimulation at the median of $23 \pm 2\%$."
- A C-LAMP experiment was added to test this result by increasing $[CO_2]$ to 550 ppmv in 1997.







Site Name	Lon (°E)	Lat (°N)	Observations NPP↑	Observations β_L	CASA′ NPP↑	CASA′ β_L	CASA′ Score	CN NPP↑	CN β_L	CN Score
Duke	-79.08	35.97	28.0%	0.69	16.4%	0.41	0.26	6.2%	0.15	0.65
Aspen	-89.62	45.67	35.2%	0.87	15.6%	0.39	0.39	12.4%	0.31	0.48
ORNL	-84.33	35.90	23.9%	0.59	17.3%	0.43	0.16	5.2%	0.13	0.64
POP-Euro	11.80	42.37	21.8%	0.54	20.0%	0.49	0.04	5.7%	0.14	0.59
4 site mean			27.2%	0.67	17.3%	0.43		7.4%	0.18	
Total M Score							**0.79**			**0.41**

But! Norby is now reporting reduced NPP enhancement at the ORNL FACE site due probably to N limitation!

C-LAMP Score Sheet for CLM3-CASA′ and CLM3-CN

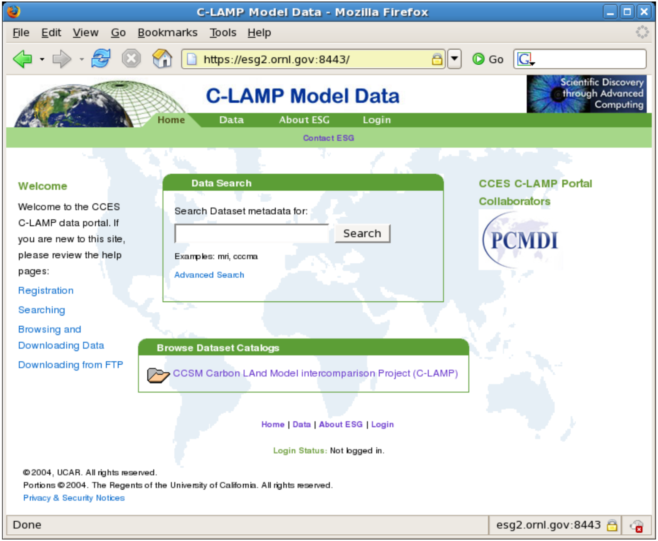
Models

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**

(Randerson et al., 2009)

Earth System Grid (ESG) Node at ORNL for C-LAMP



Global Change Biology

Global Change Biology (2009) **15**, 2462–2484, doi: 10.1111/j.1365-2486.2009.01912.x

Systematic assessment of terrestrial biogeochemistry in coupled climate–carbon models

JAMES T. RANDERSON*, FORREST M. HOFFMAN†, PETER E. THORNTON‡,§, NATALIE M. MAHOWALD¶, KEITH LINDSAY‡, YEN-HUEI LEE‡, CYNTHIA D. NEVISON*,||, SCOTT C. DONEY**, GORDON BONAN‡, RETO STÖCKLI††,‡‡, CURTIS COVEY§§, STEVEN W. RUNNING¶¶ and INEZ Y. FUNG|||

**Department of Earth System Science, Croul Hall, University of California, Irvine, CA 92697, USA, †Oak Ridge National Laboratory, Computational Earth Sciences Group, PO Box 2008, Oak Ridge, TN 37831, USA, ‡Climate and Global Dynamics, National Center for Atmospheric Research, PO Box 3000, Boulder, CO 80307, USA, §Oak Ridge National Laboratory, Environmental Sciences Division, PO Box 2008, Oak Ridge, TN 37831, USA, ¶Department of Earth and Atmospheric Sciences, 2140 Snee Hall, Cornell University, Ithaca, NY 14850, USA, ||Institute for Arctic and Alpine Research (INSTAAR), University of Colorado, Boulder, CO 80309, USA, **Department of Marine Chemistry and Geochemistry, MS 25, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA, ††Department of Atmospheric Sciences, Colorado State University, Ft Collins, CO 80523, USA, ‡‡MeteoSwiss, Climate Service, Federal Office of Meteorology and Climatology, CH-8044 Zurich, Switzerland, §§Program for Climate Model Diagnosis and Intercomparison, 7000 East Avenue, Bldg. 170, L-103, Livermore, CA 94550-9234, USA, ¶¶Numerical Terradynamic Simulation Group, College of Forestry & Conservation, University of Montana, Missoula, MT 59812, USA, |||Department of Earth and Planetary Science and Department of Environmental Science, Policy, and Management, 307 McCone, Mail Code 4767, University of California, Berkeley, CA 94720, USA*

Abstract

With representation of the global carbon cycle becoming increasingly complex in climate models, it is important to develop ways to quantitatively evaluate model performance against *in situ* and remote sensing observations. Here we present a systematic framework, the Carbon-LAnd Model Intercomparison Project (C-LAMP), for assessing terrestrial biogeochemistry models coupled to climate models using observations that span a wide range of temporal and spatial scales. As an example of the value of such comparisons, we used this framework to evaluate two biogeochemistry models that are integrated within the Community Climate System Model (CCSM) – Carnegie-Ames-Stanford Approach′ (CASA′) and carbon–nitrogen (CN). Both models underestimated the magnitude of net carbon uptake during the growing season in temperate and boreal

C-LAMP Outcome

- C-LAMP helped drive the development of model improvements in the terrestrial biogeochemistry models for the Community Land Model version 4 (CLM4).
- Subsequent C-LAMP analyses of six model configurations using CLM3.6 (a pre-release version of CLM4) with CASA′ and CN demonstrated much improved performance by CN.
- CN was fully incorporated into the CLM4 release and used for Fifth Phase Climate Model Intercomparison Project (CMIP5) simulations for IPCC AR5.
- **Next Step:** Entrain the international community to develop *benchmarks* for land model performance focused on carbon cycle, ecosystem, surface energy, and hydrology processes.

ILAMB Goals

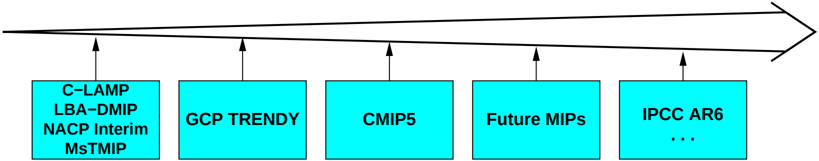
The **International Land Model Benchmarking (ILAMB) Project** will

- develop benchmarks for land model performance that are agreed upon by the international research community;
- apply the benchmarks to global models;
- support the design and development of a new, open-source, benchmarking software system for either diagnostic and model intercomparison purposes; and
- strengthen linkages between experimental, monitoring, remote sensing, and climate modeling communities in the design of new model tests and new measurement programs.

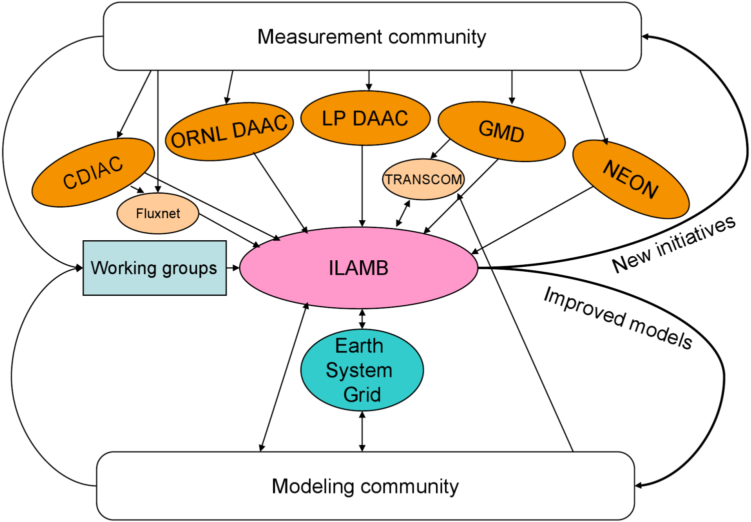
Why Benchmark?

- 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



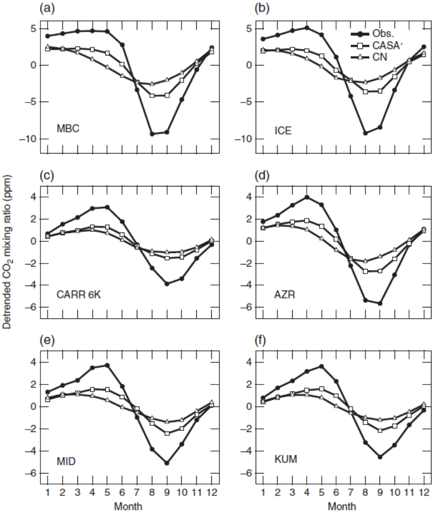
- Human capital costs of making rigorous model-data comparisons is considerable and constrains the scope of individual MIPs.
- Many MIPs spend resources "reinventing the wheel" in terms of variable naming conventions, model simulation protocols, and analysis software.
- **Need for ILAMB:** Each new MIP 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.



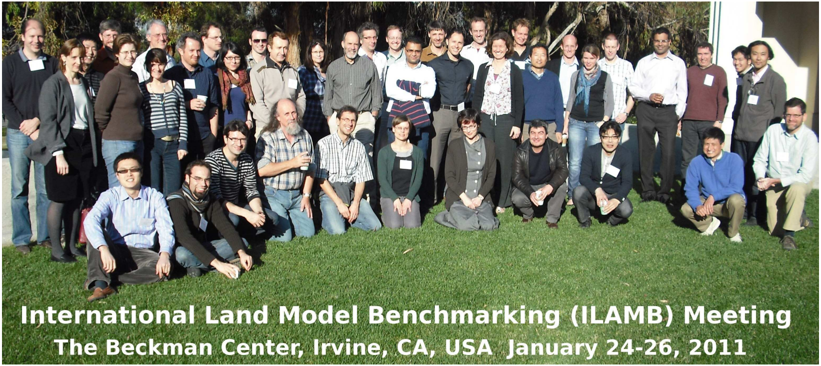
International Land Model Benchmarking project and diagnostic system

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)



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.
- 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.

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 CO_2					
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	✓			✓	Hurtt
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.
- Initial initial benchmarks will be implemented to evaluate existing CMIP5 model results.

Next Steps

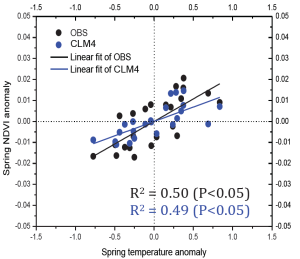
- Common model output
  - A draft document proposing additional new netCDF Climate and Forecast (CF) conventions, beyond those created for CMIP5, is available for comment.
  - To assist the modeling community, a translator between ALMA and CF standards may be created.
- Future: New protocols and forcing data comparisons.
- ILAMB side meeting was held at the 2011 AGU Fall Meeting.
- Another ILAMB meeting is being planned for 2013 to collect benchmarks from individual groups and develop the first release of a benchmarking diagnostics package.

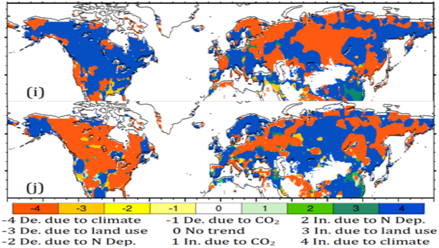
International Land Model Benchmarking (ILAMB) Project
http://www.ilamb.org/

Recent Progress in the Remote Sensing Evaluation of the Community Land Model (CLM4) at Oak Ridge National Laboratory

Forcing Factors – Phenology

- Spatio-temporal patterns of CLM4 spring (April–May) vegetation growth trends over northern mid–high latitudes ($> 25°$N) for 1982 to 2004 were explored.
- Observations and model results show a positive relationship between spring NDVI anomalies and spring temperature anomalies.
- Climate is the dominant factor controlling mid–high latitude simulated NDVI trends, both before (i) and after (j) the temperature turning point.



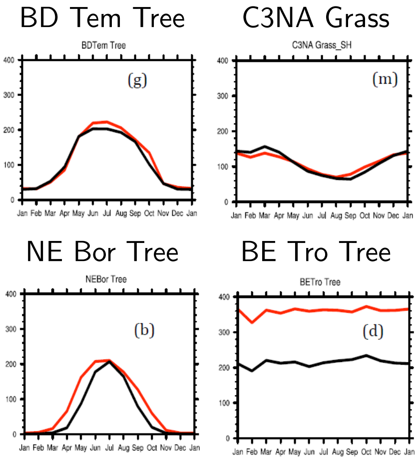


Dominant driving factors for simulated NDVI trends

(Mao et al., 2012a, *Env. Res. Lett.*)

Remote Sensing Evaluation of GPP

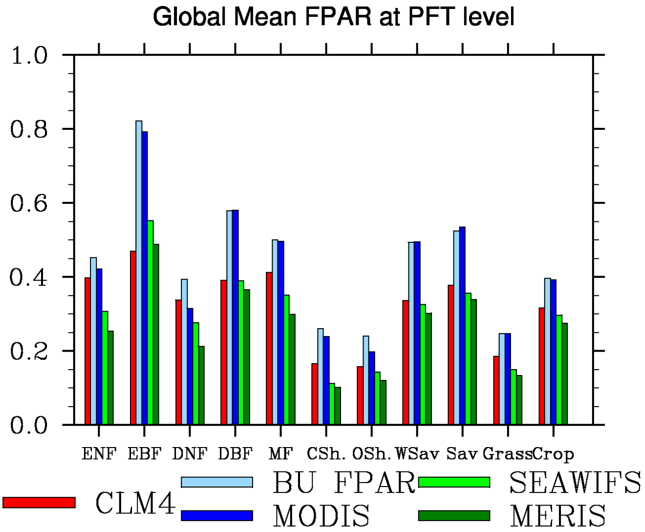
- Comparison of CLM4 PFT-level gross primary production (GPP) with MODIS satellite-based estimates for 2000–2009.
- CLM4 overestimates GPP for tropical evergreen forests and exhibits a longer carbon uptake period for most PFTs.
- CLM4 shows increases in annual averaged GPP over both hemispheres, while MODIS suggests a reduction in the Southern Hemisphere.



— MODIS — CLM4

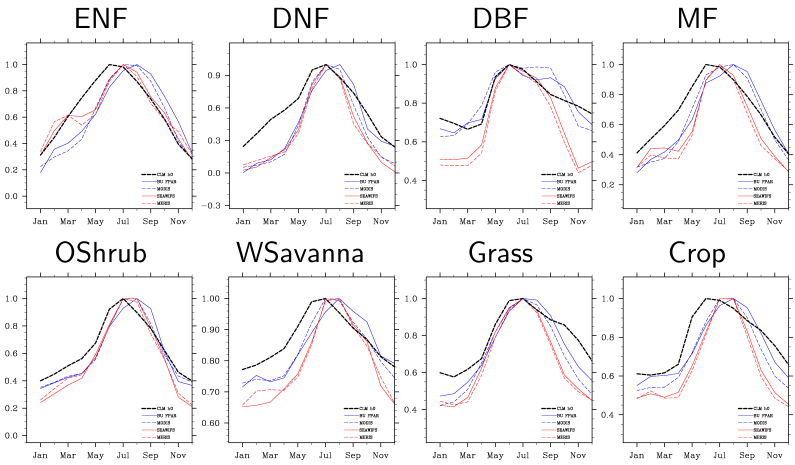
(Mao et al., 2012b, *J. Clim.*)

Spatial Analysis: PFT Level



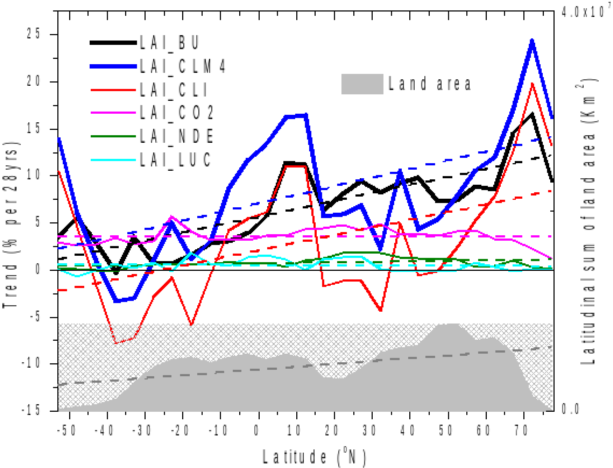
(Wang et al., 2012, in prep.)

Normalized Monthly FPAR



Eight of 11 PFTs (68.82% of land area) show good agreement with FPAR, except for an early peak month (Wang et al., 2012, in prep.).

Latitudinal LAI Trends



Latitudinal LAI trends from remote-sensing estimate (BU LAI, derived from GIMMS-NDVI3g) and factorial simulations of CLM4 between 1982 and 2009 (Mao et al., 2012c, in prep.).

Summary

- Appropriate application of remote sensing data for model evaluation is challenging.
- Carefully crafted model assessment metrics need to be designed cooperatively by the modeling, remote sensing, and *in situ* measurement communities.
- Community developed model benchmarks will help modelers use remote sensing data appropriately and track the evolving performance of their models.
- Systematic uncertainty quantification (UQ) is needed for both remote sensing data and model results, including consideration of scale, regridding, and averaging effects.
- Research is needed to understand how to combine multi-model performance evaluation results to reduce the range of uncertainty for carbon cycle predictions for future IPCC assessments.

References

F. A. Heinsch, M. C. Reeves, P. Votava, S. Kang, C. Milesi, M. Zhao, J. Glassy, W. M. Jolly, R. Loehman, C. F. Bowker, J. S. Kimball, R. R. Nemani, and S. W. Running. User's guide: GPP and NPP (MOD17A2/A3) products, NASA MODIS land algorithm. Technical report, University of Montana, 2003.

J. Mao, X. Shi, P. E. Thornton, S. Piao, and X. Wang. Causes of spring vegetation growth trends in the northern mid–high latitudes from 1982 to 2004. *Environ. Res. Lett.*, 7(1):14010, Jan. 2012a. doi:10.1088/1748-9326/7/1/014010.

J. Mao, P. E. Thornton, X. Shi, M. Zhao, and W. M. Post. Remote sensing evaluation of CLM4 GPP for the period 2000–09. *J. Clim.*, 25(15):5327–5342, Aug. 2012b. doi:10.1175/JCLI-D-11-00401.1.

J. Mao et al. 2012c. In prep.

R. B. Myneni, S. Hoffman, Y. Knyazikhin, J. L. Privette, J. Glassy, Y. Tian, Y. Wang, X. Song, Y. Zhang, G. R. Smith, A. Lotsch, M. Friedl, J. T. Morisette, P. Votava, R. R. Nemani, and S. W. Running. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sens. Environ.*, 83 (1–2):214–231, Nov. 2002. doi:10.1016/S0034-4257(02)00074-3.

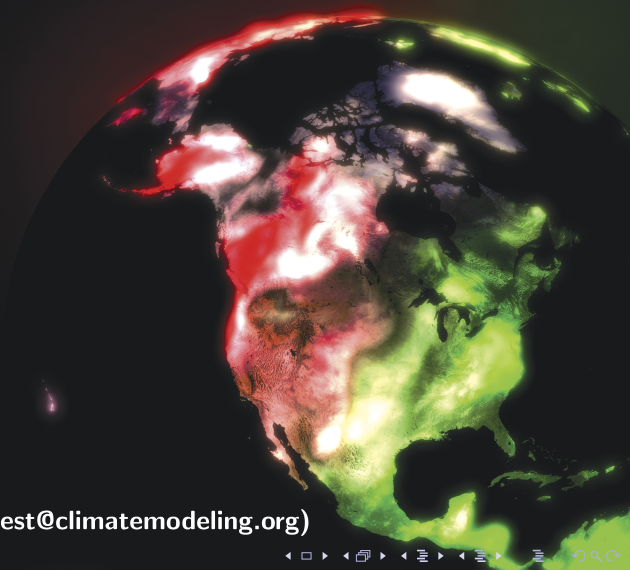
J. T. Randerson, F. M. Hoffman, P. E. Thornton, N. M. Mahowald, K. Lindsay, Y.-H. Lee, C. D. Nevison, S. C. Doney, G. Bonan, R. Stöckli, C. Covey, S. W. Running, and I. Y. Fung. Systematic assessment of terrestrial biogeochemistry in coupled climate-carbon models. *Global Change Biol.*, 15(9):2462–2484, Sept. 2009. doi:10.1111/j.1365-2486.2009.01912.x.

S. S. Saatchi, R. A. Houghton, R. C. D. S. Alvalá, J. V. Soares, and Y. Yu. Distribution of aboveground live biomass in the Amazon basin. *Global Change Biol.*, 13(4):816–837, Apr. 2007. doi:10.1111/j.1365-2486.2007.01323.x.

K. Wang et al. 2012. In prep.

M. Zhao, F. A. Heinsch, R. R. Nemani, and S. W. Running. Improvements of the MODIS terrestrial gross and net primary production global data set. *Remote Sens. Environ.*, 95(2):164–176, Mar. 2005. doi:10.1016/j.rse.2004.12.011.

Questions?



Forrest Hoffman (forrest@climatemodeling.org)