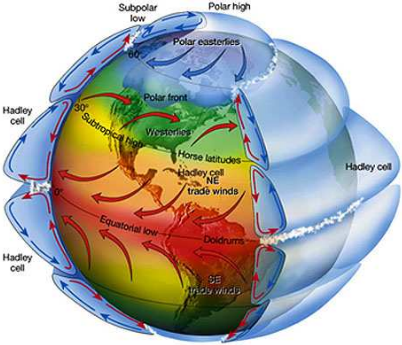
Earth System Modeling and Model Evaluation

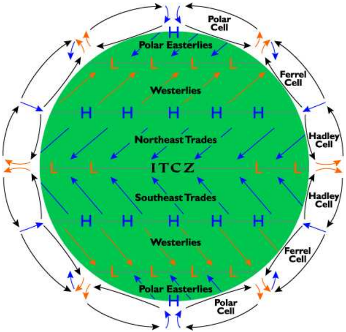
Forrest M. Hoffman

Oak Ridge National Laboratory
Climate Change Science Institute

April 14, 2011
Environmental Fellows Discussion
University of Illinois
Urbana-Champaign, Illinois, USA

Atmospheric Circulation

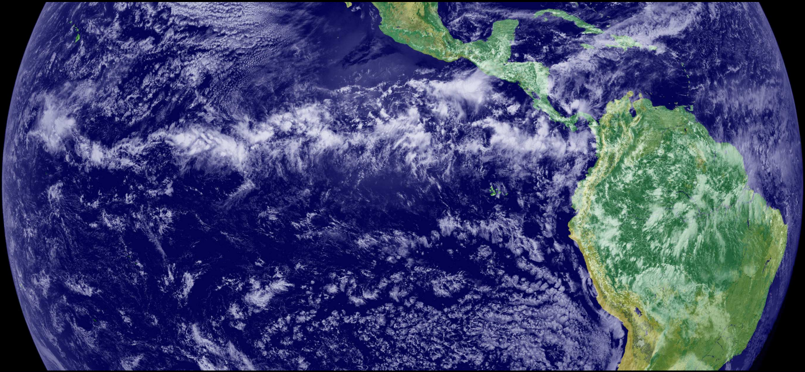




Energy is carried from the tropics poleward via latitudinal circulation cells. Warm, moist air rises along the equator, while cool, dry air sinks at around 30° N/S, forming the Hadley Cells.

Left: `http://serc.carleton.edu/eslabs/hurricanes/1b.html`
Right: `http://www.newmediastudio.org/DataDiscovery/Hurr_ED_Center/Easterly_Waves/Trade_Winds/Trade_Winds.html`

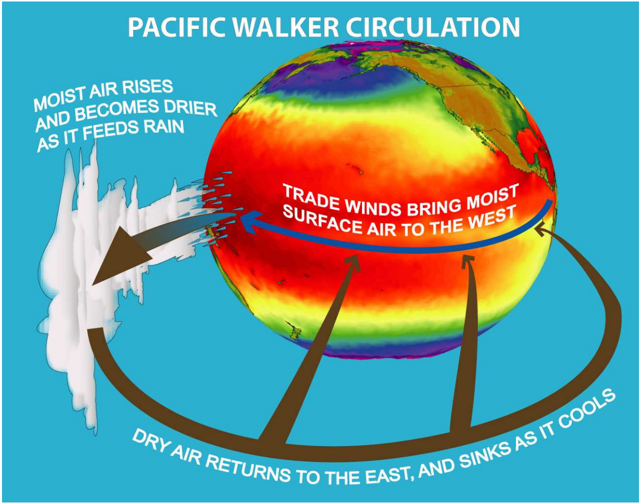
Intertropical Convergence Zone (ITCZ)



This image is a combination of cloud data from NOAAs newest Geostationary Operational Environmental Satellite (GOES-11) and color land cover classification data. The ITCZ is the band of bright white clouds that cuts across the center of the image.

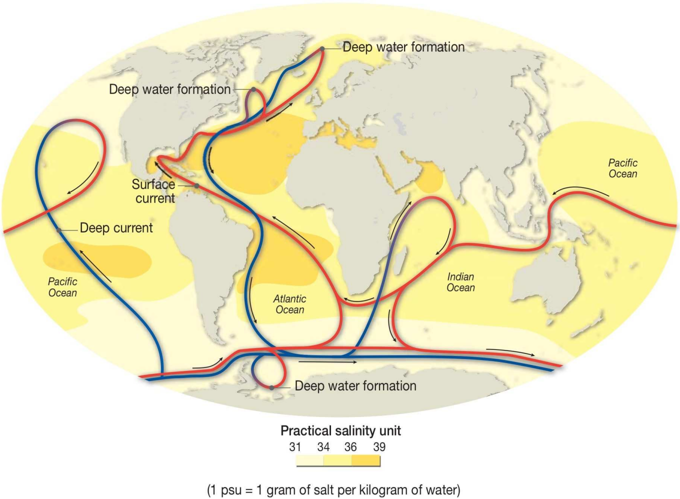
http://earthobservatory.nasa.gov/IOTD/view.php?id=703

Atmospheric Circulation



http://www.noaanews.noaa.gov/stories2007/s2840.htm

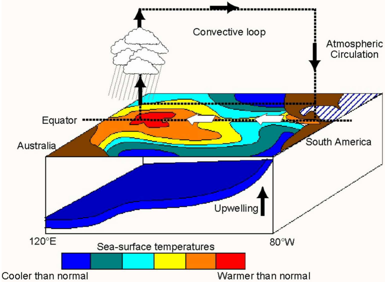
Thermohaline Circulation



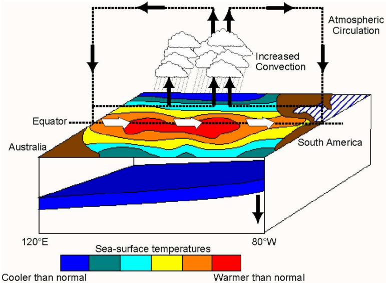
http://maps.grida.no/go/graphic/thermohaline-circulation1

La Niña/El Niño Southern Oscillation

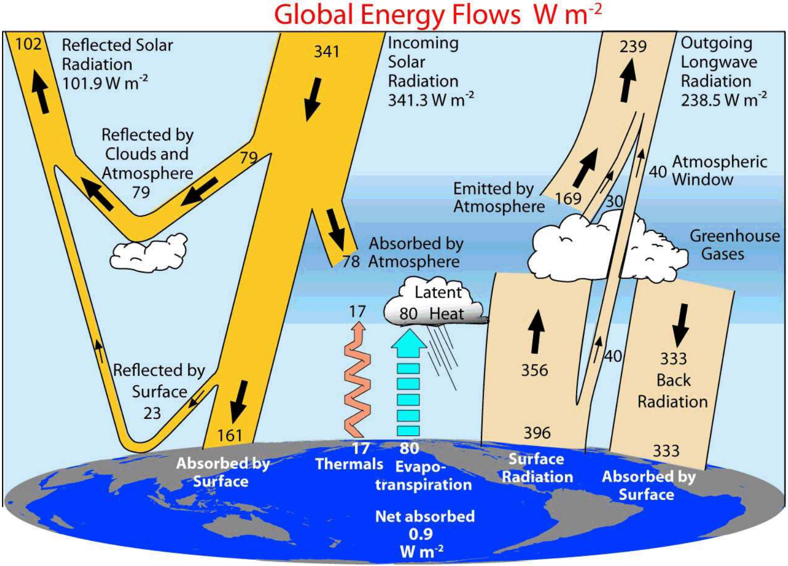
La Niña Conditions



El Niño Conditions

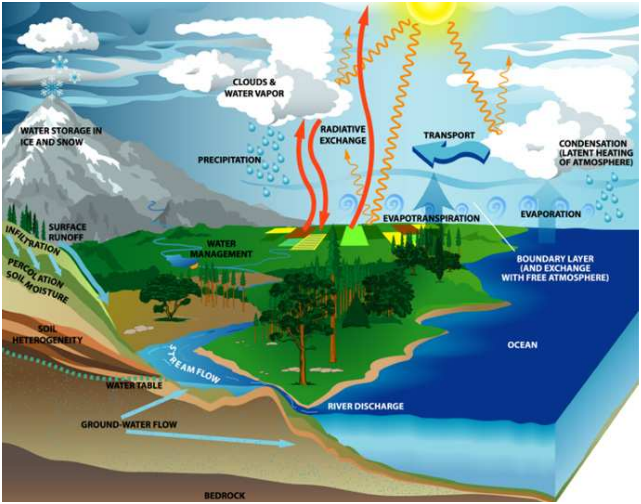


http://old.weathersa.co.za/References/elnino.jsp



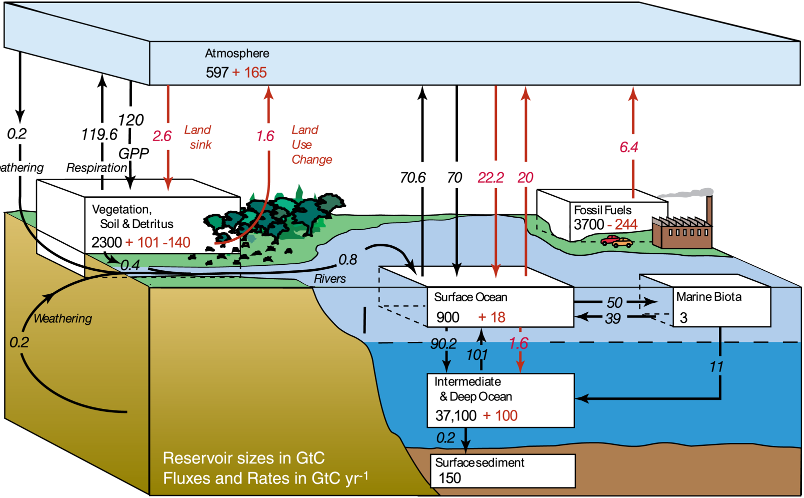
Trenberth et al., 2008

Global Water Cycle



http://ncssmapes.wikispaces.com/Elizabeth+T2

Global Carbon Cycle



http://www.gfdl.noaa.gov/anthropogenic-carbon-cycle

General Circulation Models

General Circulation Models (GCMs) are designed to capture the large-scale circulation of *energy* and *mass* (water) on Earth.

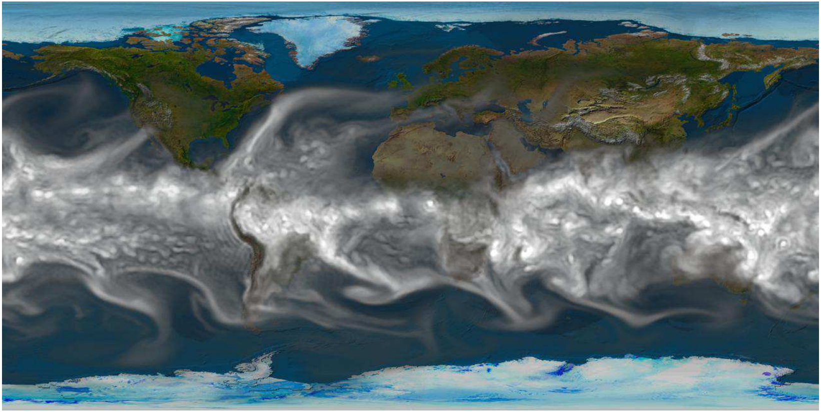


Image courtesy of the National Center for Atmospheric Research

Earth System Models

Earth System Models (ESMs) simulate the large-scale global circulation as well as biogeochemical processes in the ocean, biogeochemical processes and dynamic vegetation on land, reactive atmospheric chemistry and aerosol interactions, and ice sheet dynamics. Humans are now being added to these models.

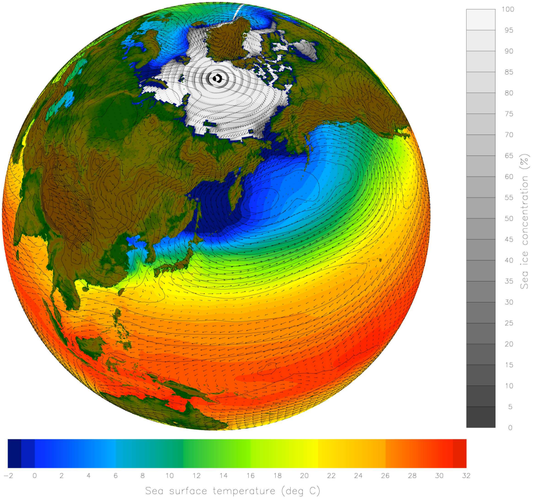
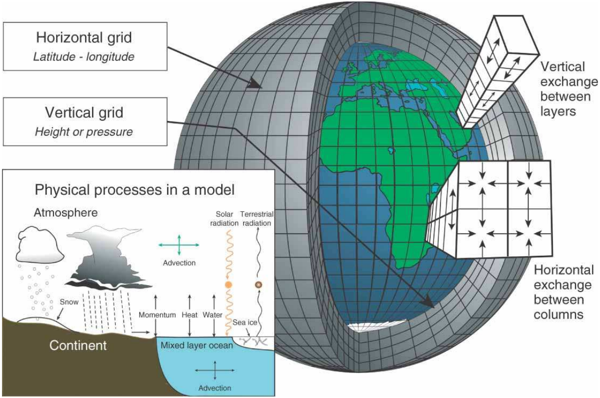


Image courtesy of the National Center for Atmospheric Research

Earth System Models

These models typically exchange fluxes of energy and mass vertically between atmosphere/ocean/snow/soil layers and laterally between grid cells or columns.

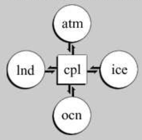


http://wires.wiley.com/WileyCDA/WiresArticle/articles.html?doi=10.1002%2Fwcc.95

ESM Frameworks

- ESMs are typically composed of individual "component models" for simulating ocean, atmosphere, sea ice, and land processes.
- Processes within these component models are further divided into modules (*e.g.*, the atmosphere has multiple dynamical cores that may be coupled to the physics modules).
- These component models exchange states and fluxes through a coupler component.
- The coupler converts, translates, re-grids, and redistributes data from component models at the appropriate frequency.
- Engineering the modeling system to operate efficiently on large supercomputers is hard.

The CCSM Framework

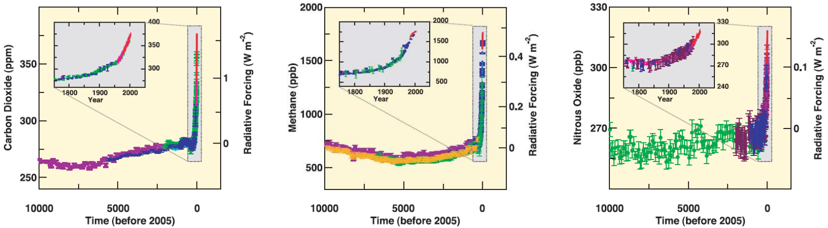


CCSM software is based on a framework that divides the complete climate system into component models connected by a coupler. Individual components -- ocean, atmosphere, land, and sea-ice -- can be exchanged for alternate models, thus allowing different configurations appropriate for different applications.

Intergovernmental Panel on Climate Change (IPCC)

- Every 5–10 years, the United Nations' Intergovernmental Panel on Climate Change (IPCC) produces reports describing the state of knowledge regarding historical and future climate change and its impacts.
- The IPCC forms Working Groups of leading international scientists to author the assessment reports.
  - Working Group I: The Physical Science Basis
  - Working Group II: Impacts, Adaptation, and Vulnerability
  - Working Group III: Mitigation of Climate Change
- The Fourth Assessment Report (AR4) was released in 2007, and its authors were co-recipients of the Nobel Peace Prize, with Al Gore Jr., that same year.
- Climate model results from dozens of models are used to make best estimates of climate impacts under various greenhouse gas emissions scenarios.

Changes in Atmospheric Greenhouse Gas Concentrations

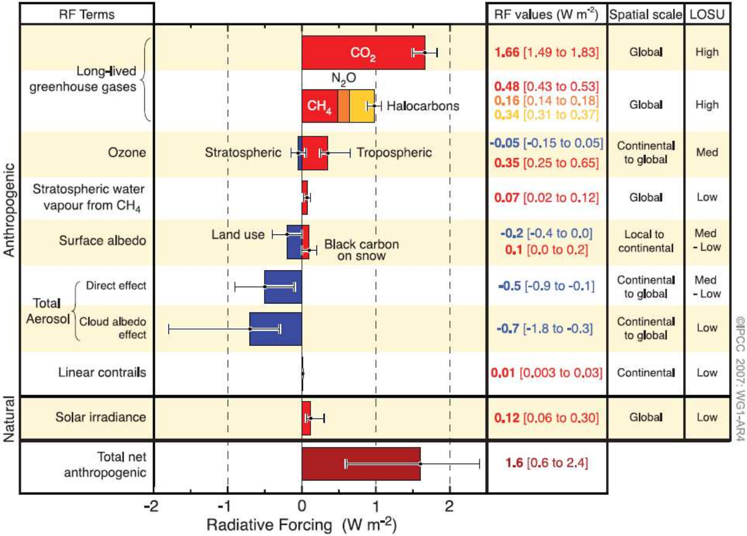


Atmospheric concentrations of carbon dioxide, methane, and nitrous oxide over the 10,000 years (large panels) and since 1750 (inset panels). Measurements are shown from ice cores (symbols with different colours for different studies) and atmospheric samples (red lines). The corresponding radiative forcings are shown on the right hand axes of the large panels. From IPCC AR4 WG1 (2007).

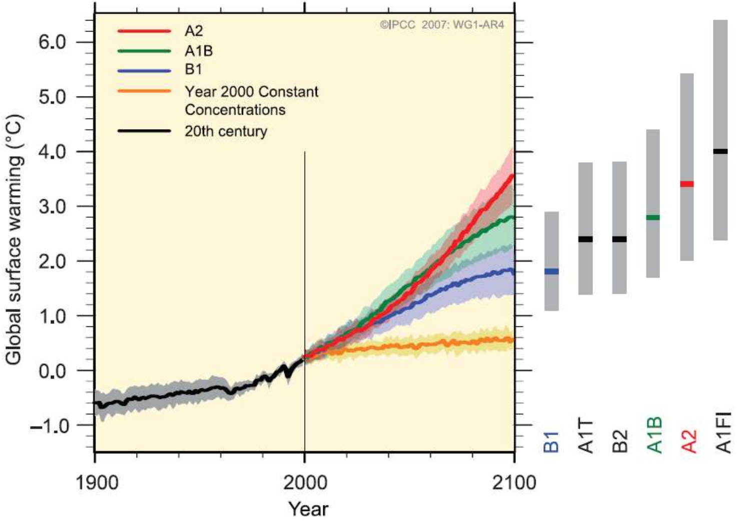
Radiative Forcing

Radiative forcing is a measure of the influence a factor has in altering the balance of incoming and outgoing energy in the Earth-atmosphere system and is an index of the importance of the factor as a potential climate change mechanism. In this report radiative forcing values are for changes relative to pre-industrial conditions defined at 1750 and are expressed in watts per square meter (W/m^2).

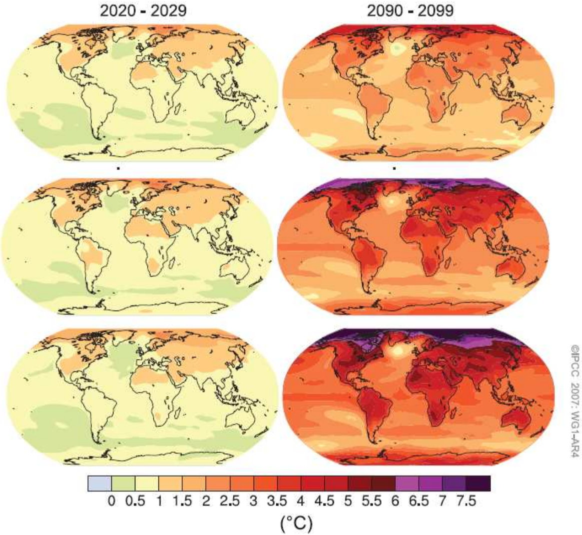
Global Average Radiative Forcing Estimates



Multi-Model Averages and Ranges for Surface Warming



Projected Surface Temperature Changes



The Community Earth System Model (CESM1.0)

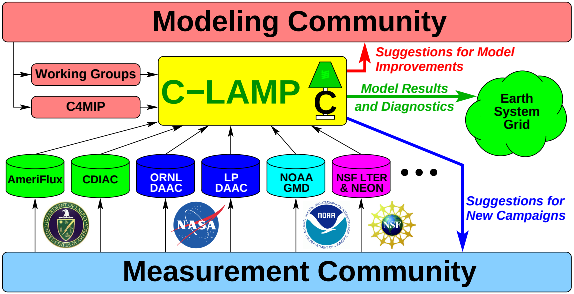
- A first-generation ESM derived from the Community Climate System Model version 4 (CCSM4).
- Developed by the National Center for Atmospheric Research (NCAR) and U.S. Dept. of Energy (DOE) National Labs.
- Consists of the Community Atmosphere Model (CAM5), Community Land Model (CLM4), Parallel Ocean Program (POP2), Community Ice Code (CICE4), Community Ice Sheet Model (Glimmer - CISM), and CESM Coupler (CPL7).
- May be configured with active component models or with "data models" that force simulations using stored data.
- Is being run at NCAR and DOE Labs for Phase 5 of the Coupled Model Intercomparison Project (CMIP5), the results of which will be used for the IPCC Fifth Assessment Report (AR5) expected in 2013.

Essentially, all models are wrong, but some are useful

– George E. P. Box

We need to assess the performance of ESMs over the contemporary observational period in order to improve them and to judge their potential utility for making future projections.

- 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.



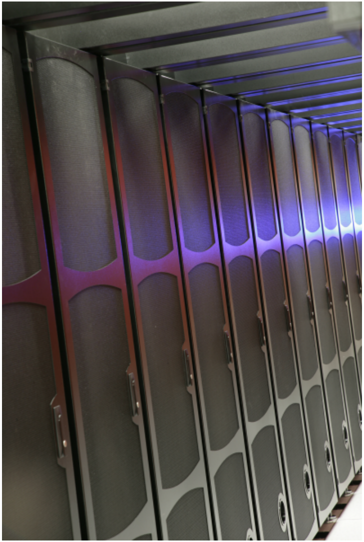
- C-LAMP is a Biogeochemistry Subproject of the **Computational Climate Science End Station** (Warren Washington, PI), a U.S. Dept. of Energy INCITE Project.
- Models were initially run on the Cray X1E vector supercomputer in ORNL's **National Center for Computational Sciences (NCCS)**.

Cray X1E (phoenix)

1024 processors (MSPs), 2048 GB memory, and 18.08 TFlop/s peak

DECOMMISSIONED September 30, 2008

Recently Decommissioned Jaguar: 250 TFlop/s



New Jaguar: Second Fastest in the World at 1.759 PFlop/s

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.*, 2006; Batjes, 2006)
  - 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_C/

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		model amazon	amazon map			

Done

Score Sheet for CLAMP - Mozilla Firefox

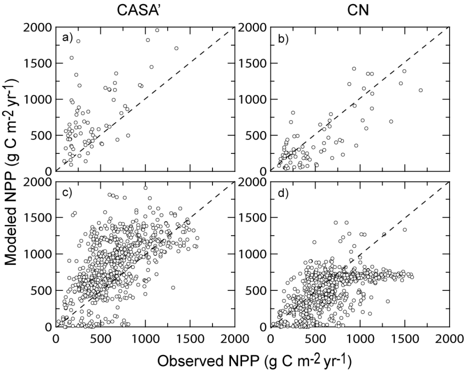
File Edit View History Bookmarks Tools Help

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**

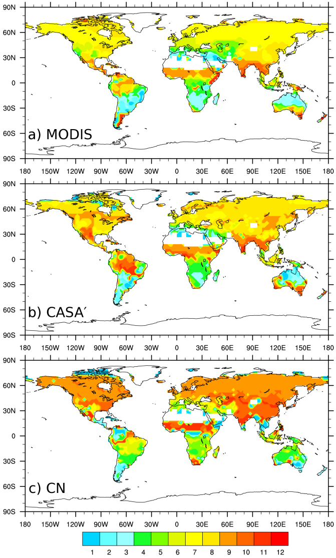
Done

- 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).

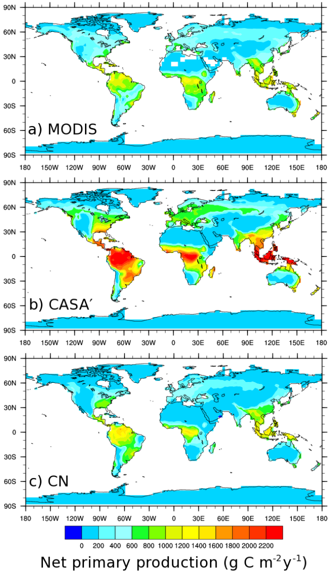


Data provided by NASA Distributed Active Archive Center (DAAC) at ORNL

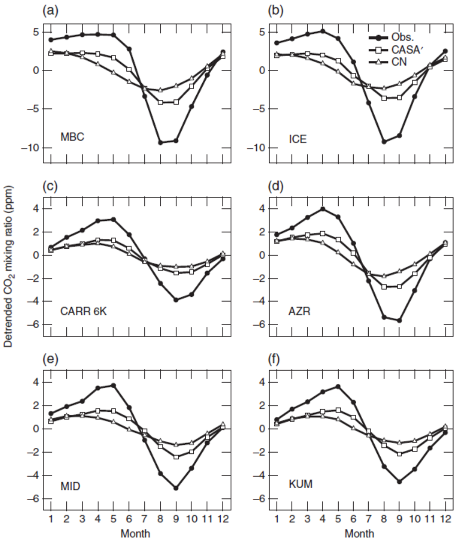
- Comparisons with satellite "modeled observations" must be made carefully because of high uncertainty.
- This comparison with MODIS leaf area index (LAI) focuses on the month of maximum LAI (phase), a measurement with less uncertainty than the "observed" LAI values.
- C-LAMP accounts for this uncertainty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



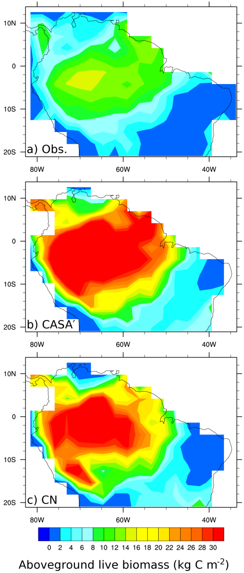
- MODIS net primary production (NPP) "observations" have higher uncertainty.
- Comparison with MODIS NPP focuses on correlation of spatial patterns.
- CLM-CASA′ scored 1.6/2.0 while CLM-CN scored 1.4/2.0.



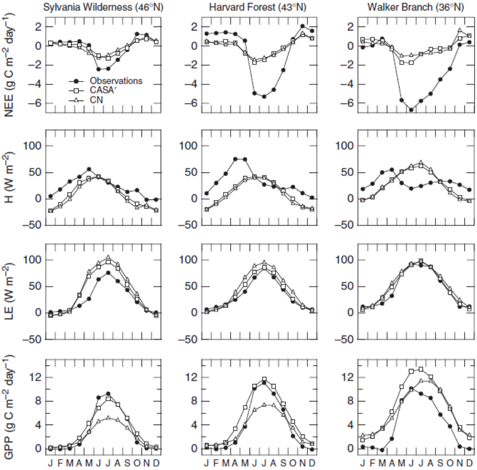
- Comparisons with Globalview flask sites are made by combining model fluxes with impulse response functions from TRANSCOM.
- Shown are the annual cycles of atmospheric CO_2 at (a) Mould Bay, Canada (76°N), (b) Storhofdi, Iceland (63°N), (c) Carr, Colorado (41°N), (d) Azores Islands (39°N), (e) Sand 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.* (2006) 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.

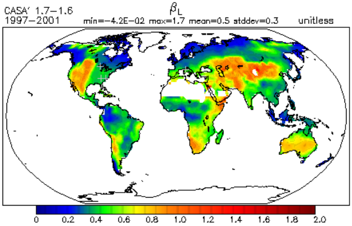


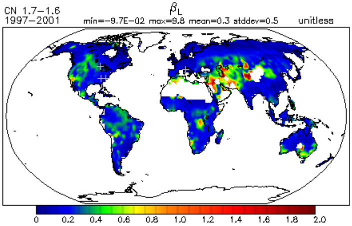
- Comparisons with AmeriFlux eddy correlation CO_2 flux tower sites include net ecosystem exchange (NEE), gross primary production (GPP), respiration, shortwave incoming radiation, and latent and sensible heat.
- Shown here is a comparison of model estimates with eddy covariance measurements from Sylvania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

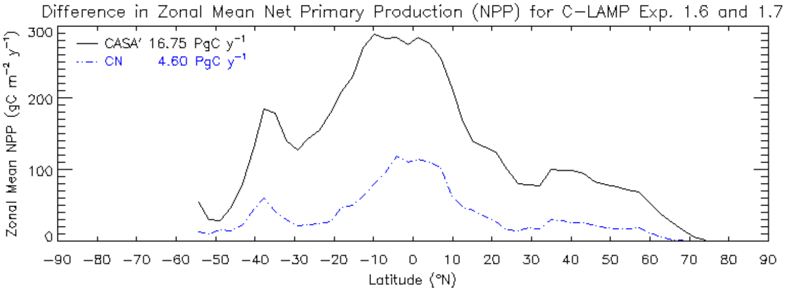


Data provided by ORNL Carbon Dioxide Information Analysis Center (CDIAC).

- 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!

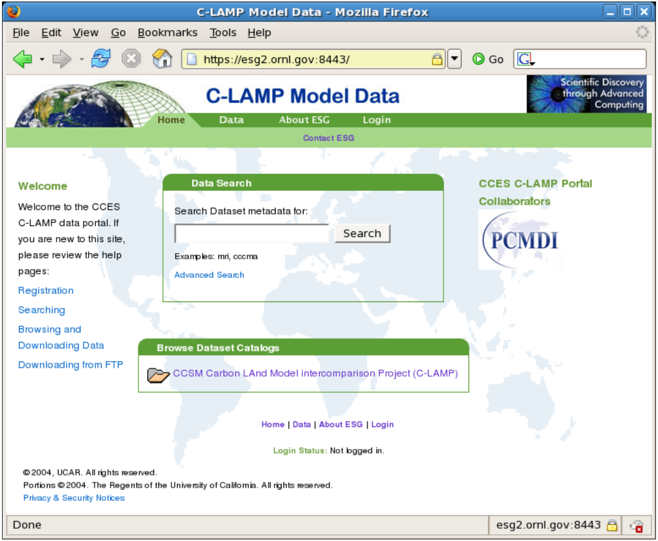
C-LAMP Score Sheet for CLM3-CASA′ and CLM3-CN

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**

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

Recent Progress

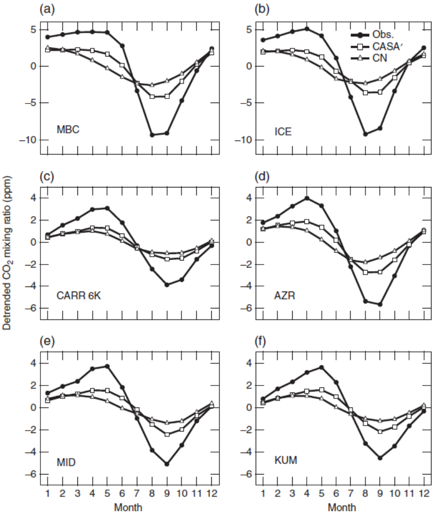
- C-LAMP helped drive the development of model improvements in the terrestrial biogeochemistry models for the Community Land Model version 4 (CLM4).
- Subsequent C-LAMP analyses of six model configurations using CLM3.6 (a pre-release version of CLM4) with CASA′ and CN demonstrated much improved performance by CN.
- It is now recognized that physical model changes must be tested using C-LAMP to ensure that these changes do not have negative impacts on biogeochemistry model performance.
- We are sharing the data and diagnostics package for others to use (*e.g.*, Jena's JEDI model) and hoping to incorporate additional metrics over time.

New International Benchmarking Activity

- We believe that C-LAMP and the initial European ILAMB should serve as a prototype for an international benchmarking activity, the results of which could contribute to AR5.
- Needed are
  1. a well-crafted protocol that exercises model capabilities for simulating energy, hydrological, and biogeochemical cycles;
  2. common model output standards to simplify analyses;
  3. best-available forcing data set; and
  4. best-available observational data sets and diagnostics.
- We should harness various community efforts to develop an open source, modular, extensible, and well documented model evaluation system to support future MIPs, like LBA-MIP, C-LAMP, NACP Syntheses, TRENDY, MsTMIP, and CMIP5.
- Earth System Grid (ESG) is available for sharing model results.

What is a Benchmark?

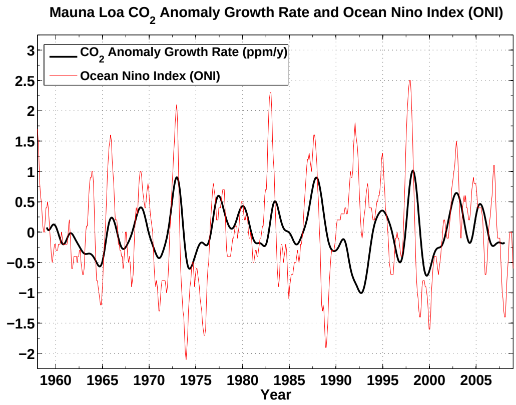
- A benchmark is a quantitative test of model function, for which the uncertainties associated with the observations can be quantified.
- Acceptable performance on benchmarks **is a necessary but not sufficient condition** for a fully functioning model.
- Since all datasets have strengths and weaknesses, 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*

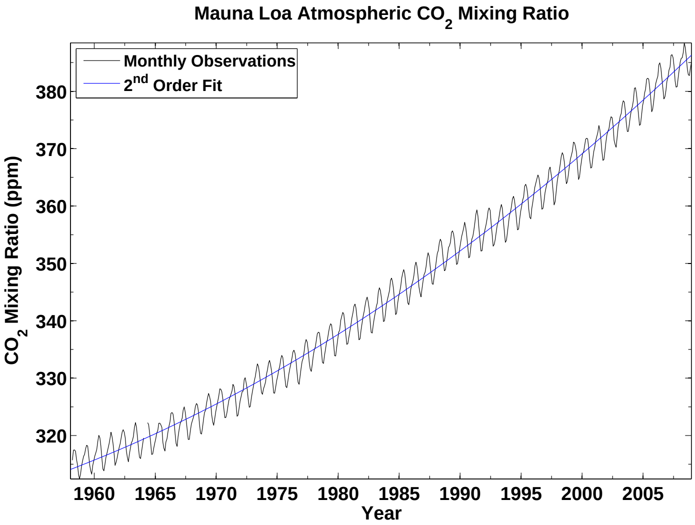
The relationship between El Niño-Southern Oscillation (ENSO) and observed CO_2 anomalies at Mauna Loa may be exploited to evaluate ocean and terrestrial model responses.



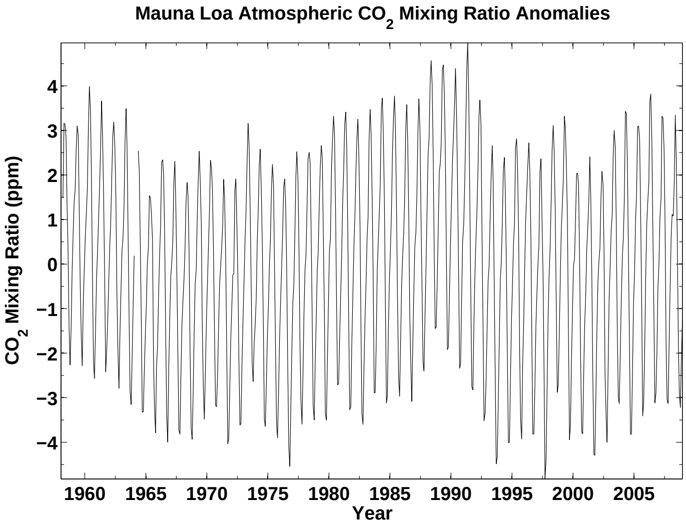
CO_2 Dependence on El Niño-Southern Oscillation (ENSO)

- Keeling and Revelle (1985) described a shutdown in upwelling and biological activity during El Niño years, resulting in a shutdown of CO_2 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 CO_2 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 CO_2 to El Niño.

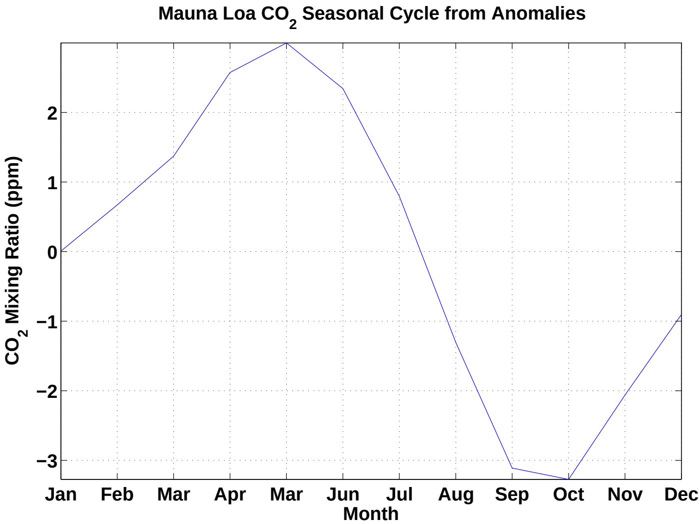
Mauna Loa CO_2 (1957–2008) and Polynomial Curve Fit



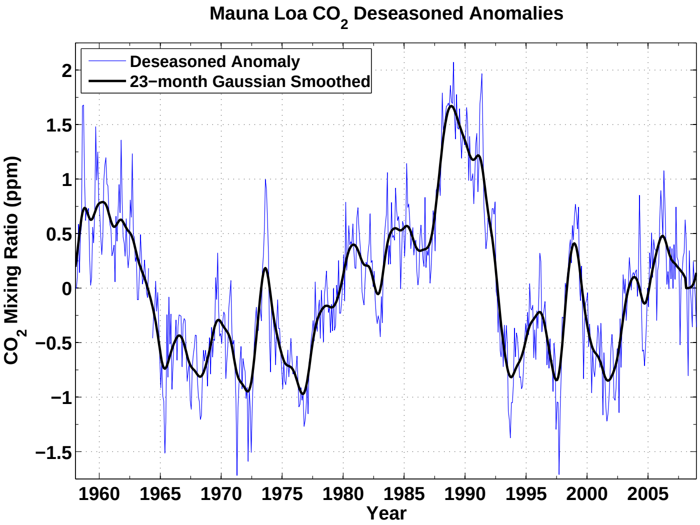
Mauna Loa CO_2 (1957–2008) Minus the Trend



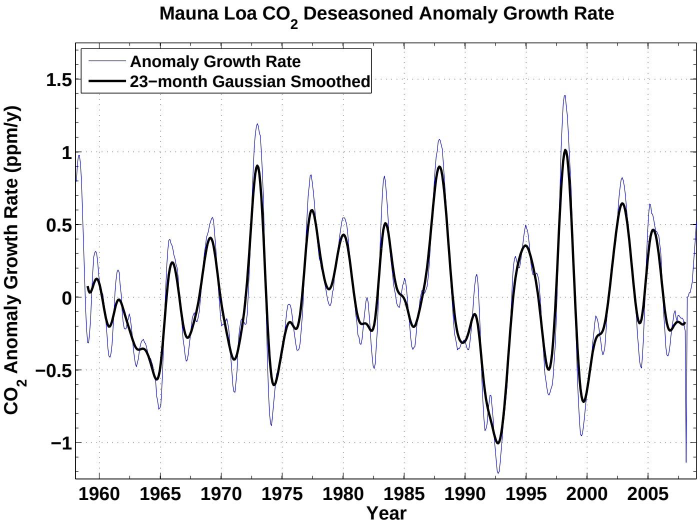
Mauna Loa CO_2 (1957–2008) Mean Seasonal Cycle



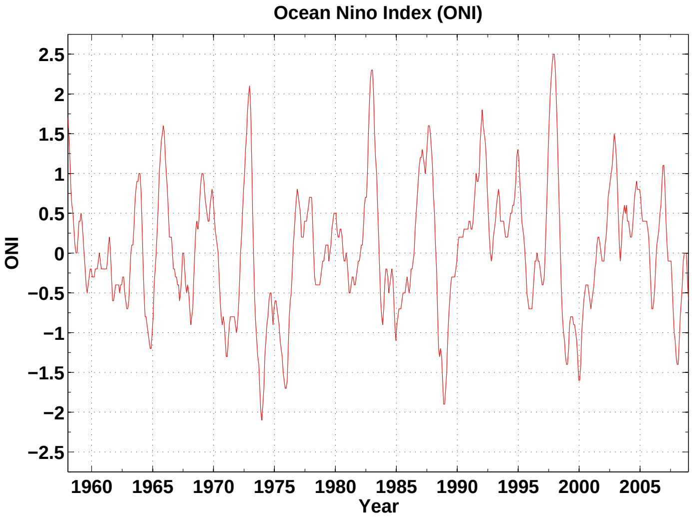
Mauna Loa CO_2 (1957–2008) Deseasoned Anomalies



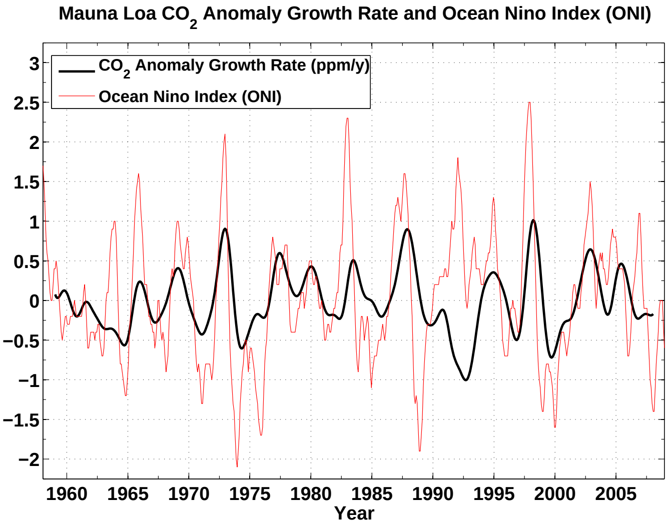
Mauna Loa CO_2 (1957–2008) Anomaly Growth Rate



Ocean Niño Index (ONI)



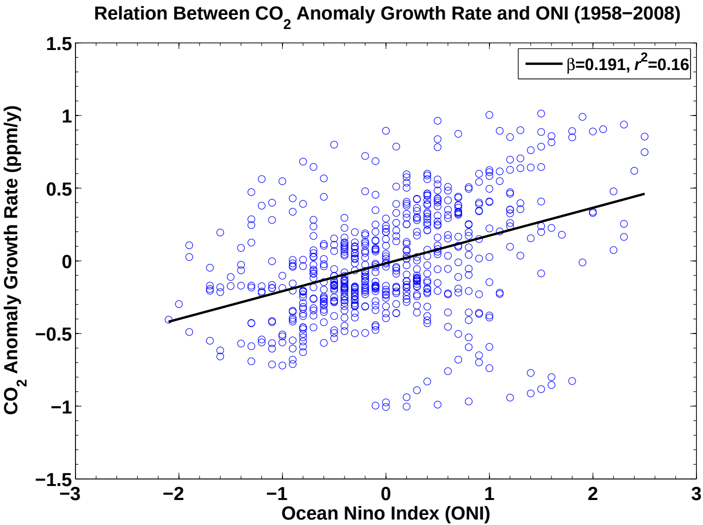
CO_2 Anomaly Growth Rate and Ocean Niño Index



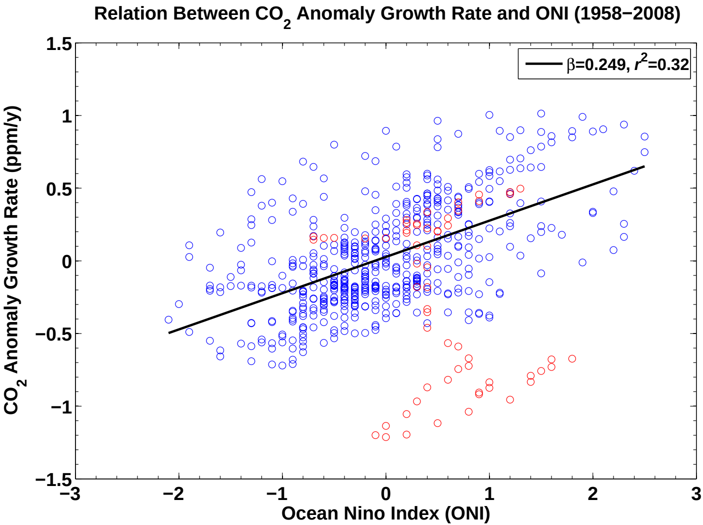
Mount Pinatubo Eruption

- 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

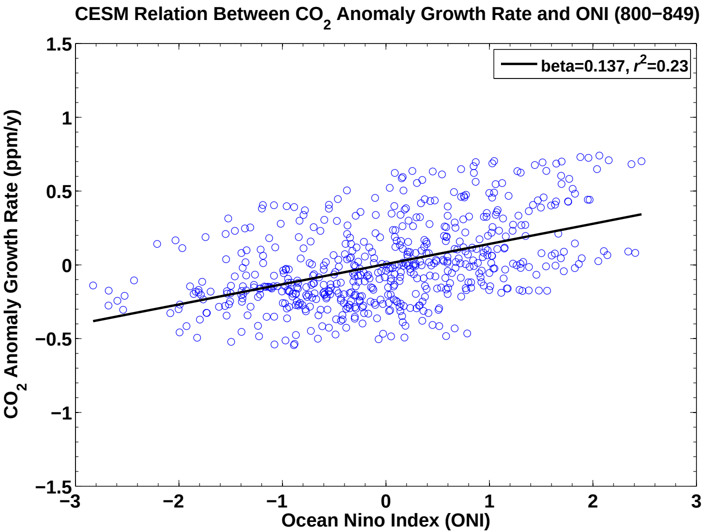
Relation Between CO_2 Anomaly Growth Rate and ONI



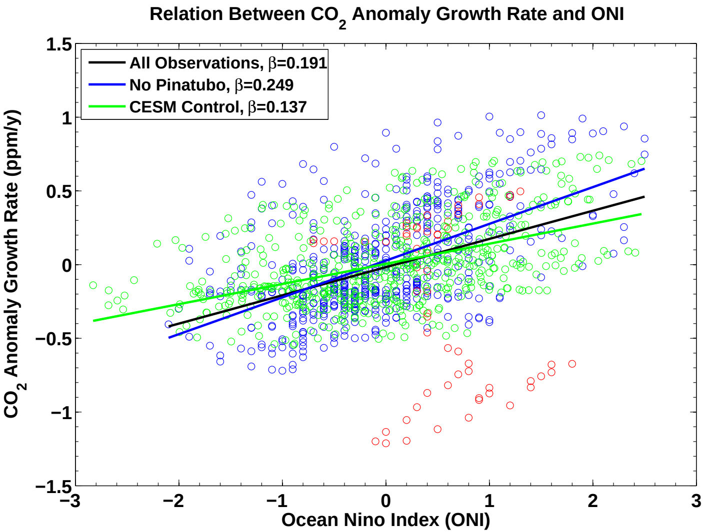
Relation Without 1991–1995 (Pinatubo Period)



Community Earth System Model (CESM) Control Run



CESM vs. Observations



Benchmark Conclusions

- Relationship between Mauna Loa CO_2 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 CO_2 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.

Next Steps

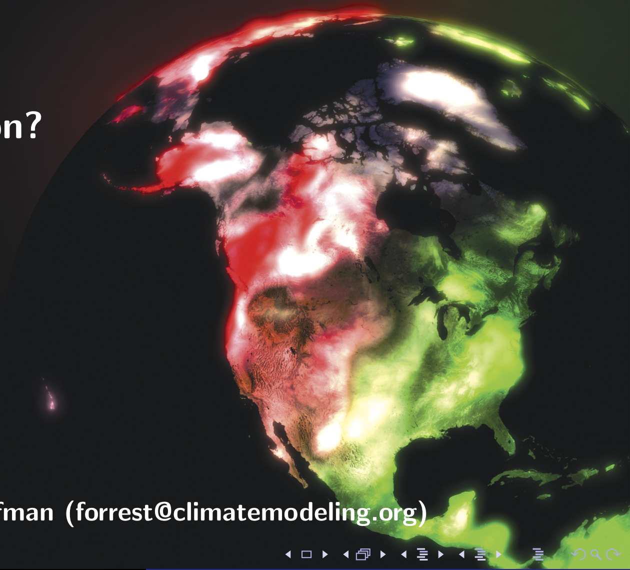
- Common model output
  - A draft document proposing additional new netCDF Climate and Forecast (CF) conventions, beyond those created for CMIP5, is available for comment.
  - To assist the modeling community, a translator between ALMA and CF standards may be created.
- Future: New protocols and forcing data comparisons.
- C-LAMP2 will produce new metrics and diagnostics for CESM1-CLM4 using the ILAMB software architecture.
- Certain C-LAMP2 diagnostics will be contributed to ILAMB.

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

Thank you!

Questions?

More Discussion?

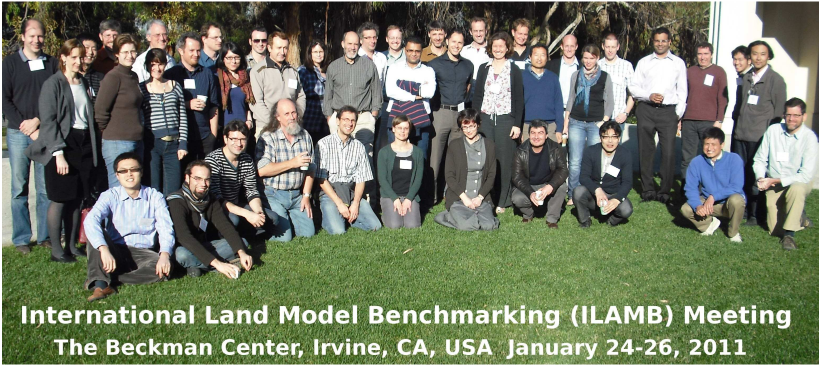


Contact: Forrest Hoffman (forrest@climatemodeling.org)

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 CO_2 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.



- Meeting Co-organized by Forrest Hoffman (UC-Irvine and ORNL), Chris Jones (UK Met Office), Pierre Friedlingstein (U. Exeter and IPSL-LSCE), 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 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.
- 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.