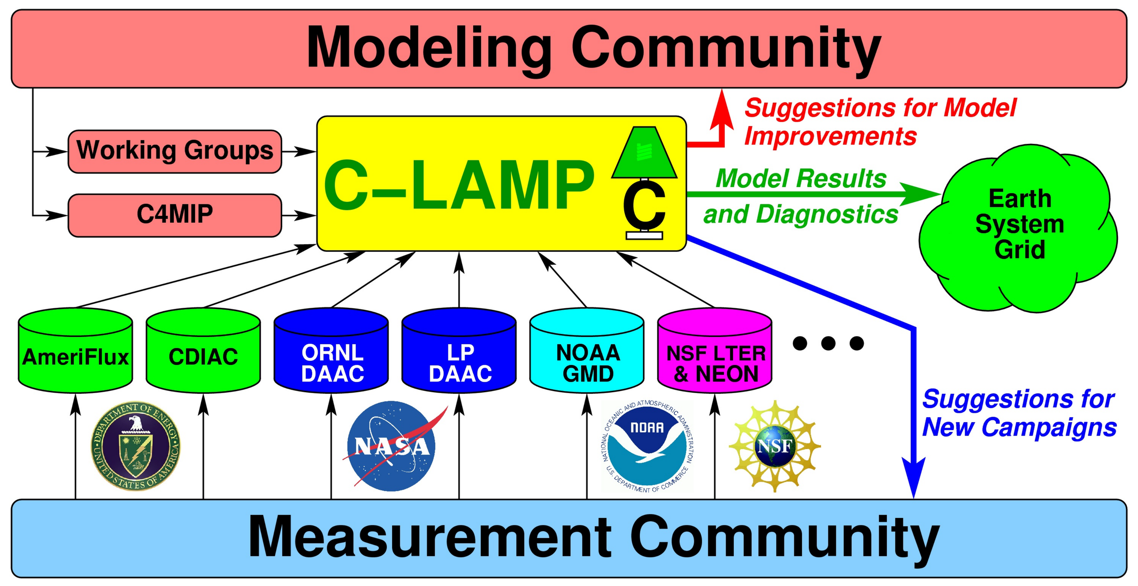
A New Model Evaluation Framework for the International Land Model Benchmarking (ILAMB) Project

Forrest M. Hoffman[1,2] and James T. Randerson[1]

[1]University of California-Irvine and [2]Oak Ridge National Laboratory (ORNL)

Introduction

The need to capture important climate feedbacks in general circulation models (GCMs) has resulted in new efforts to include atmospheric chemistry and land and ocean biogeochemistry into the next generation of production climate models, now often referred to as Earth System Models (ESMs). While many terrestrial and ocean carbon models have been coupled to GCMs, recent work has shown that such models can yield a wide range of results (Friedlingstein *et al.*, 2006), suggesting that a more rigorous set of offline and partially coupled experiments, along with detailed analyses of processes and comparisons with measurements, are warranted. The Carbon-Land Model Intercomparison Project (C-LAMP) provides a simulation protocol and model performance metrics based upon comparisons against best-available satellite- and ground-based measurements (Hoffman *et al.*, 2007). C-LAMP provides feedback to the modeling community regarding model improvements and to the measurement community by suggesting new observational campaigns.



By using the wide variety of measurements made, collected, and distributed by researchers and data centers, C-LAMP identifies areas in which improvements can be made to models as well as identifying needs for new kinds of measurements. In addition, all the C-LAMP model output is distributed via the Earth System Grid (ESG), and model diagnostics are available on the Web for use by the wider scientific community.

Originally designed to test the performance of two such models coupled to the Community Climate System Model Version 3 (CCSM3), C-LAMP established metrics for scoring the performance of models by comparison with best-available observational datasets, from satellite-based to leaf-scale measurements. Shown in the table below are the results of these comparisons. A sampling of individual diagnostics is shown at right. C-LAMP is now serving as a prototype for the International Land Model Benchmarking (ILAMB) Project, an effort to design community-accepted standards for performance benchmarks and an open source diagnostics package.

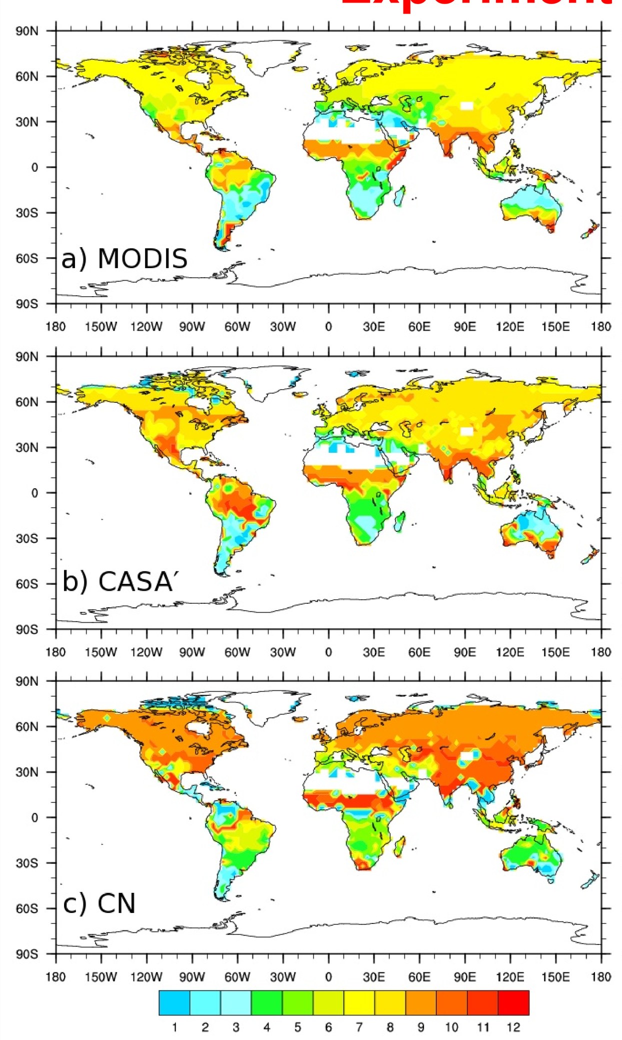
C-LAMP Score Sheet for CASA´ and CN

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**

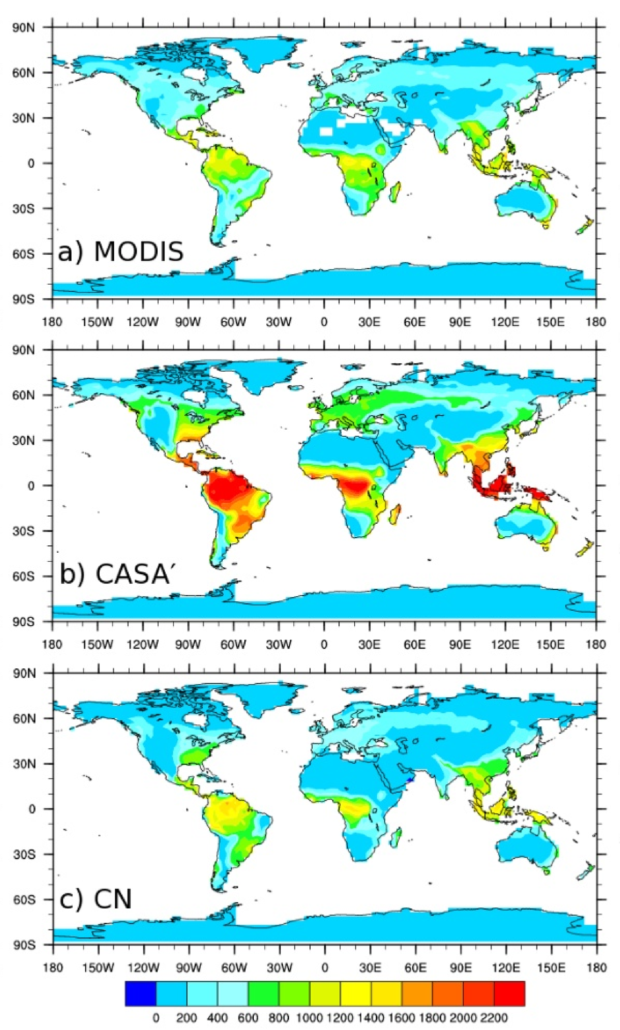
C-LAMP produced a standard set of common output quantities for climate-carbon cycle models and recommendations for carbon accounting. These are being proposed as additions to the NetCDF Climate and Forecast (CF) Metadata Convention for output field names and units in addition to those produced by terrestrial biogeochemistry components of Earth System Models for IPCC AR5.

The complete protocol, metrics for evaluation, and output approach are described at http://www.climatemodeling.org/c-lamp

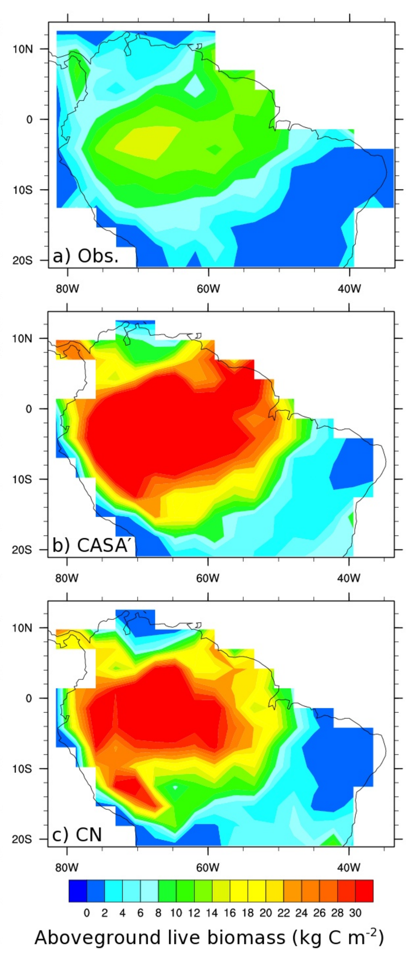
Experiment 1 Results for CASA´ and CN



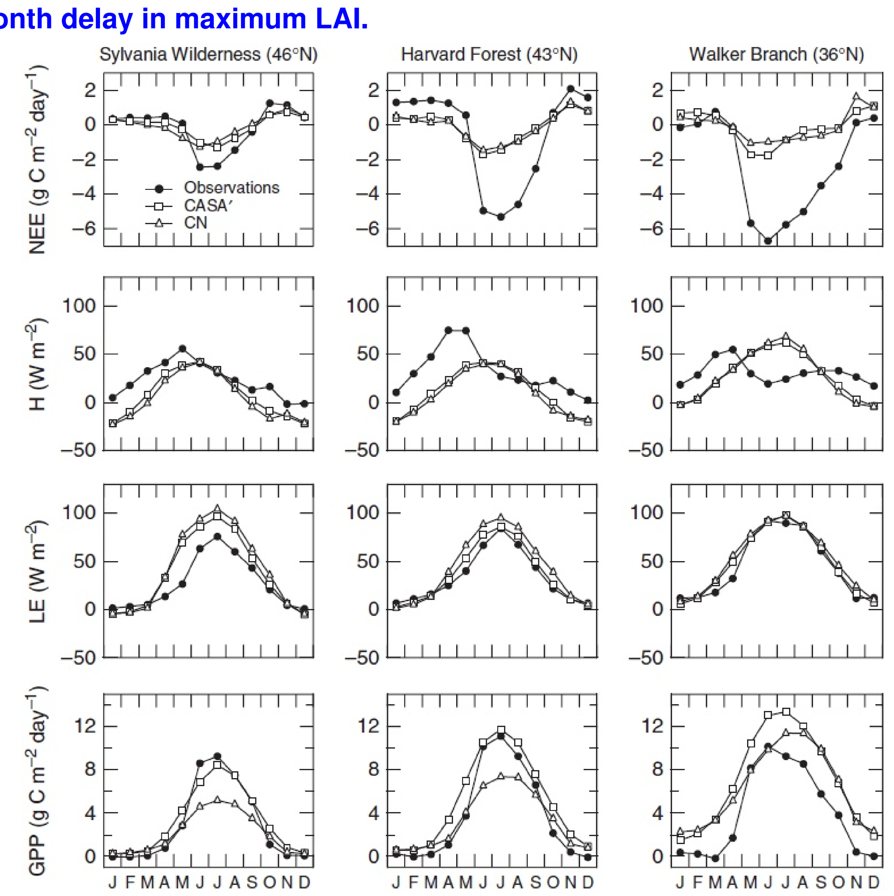
Comparison with MODIS MOD15A2 for month of maximum leaf area index (LAI). While direct comparison of model results with MODIS LAI values is problematic, it is expected that the month of maximum LAI from MODIS has a much lower uncertainty. Both models exhibited a 1–3 month delay in maximum LAI.



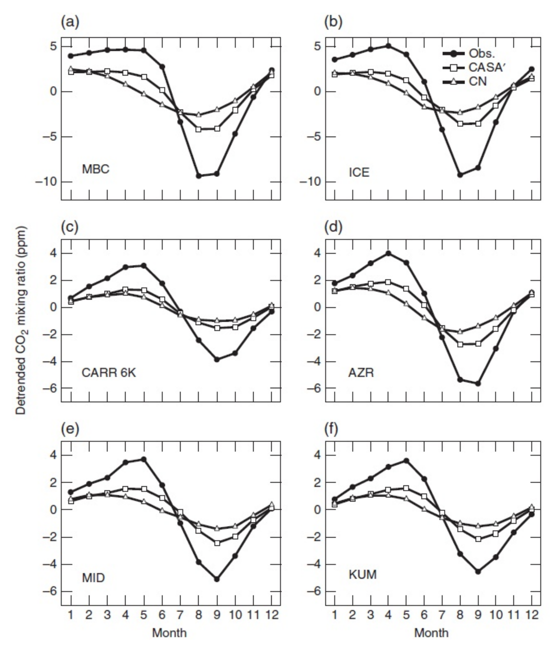
Comparison with MODIS net primary production (NPP) in gC m^{-2} y^{-1}. Models are scored with MODIS NPP, spatial correlation with MODIS NPP, not their actual values. CASA´ had a correlation coefficient of 0.91 while CN had a correlation coefficient of 0.85.



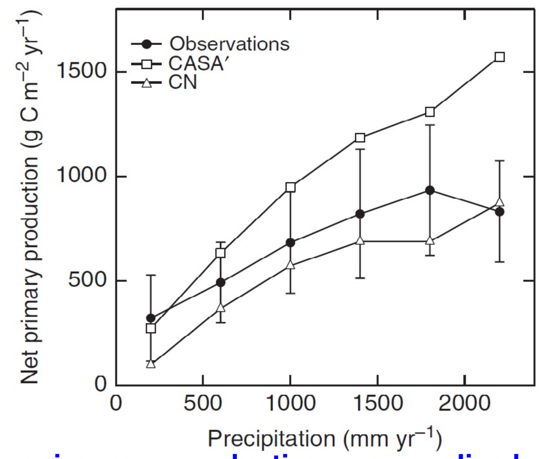
Comparison of above ground live biomass with estimates provided by Saatchi *et al.* (2006). Both models significantly over estimated carbon storage in woody biomass.



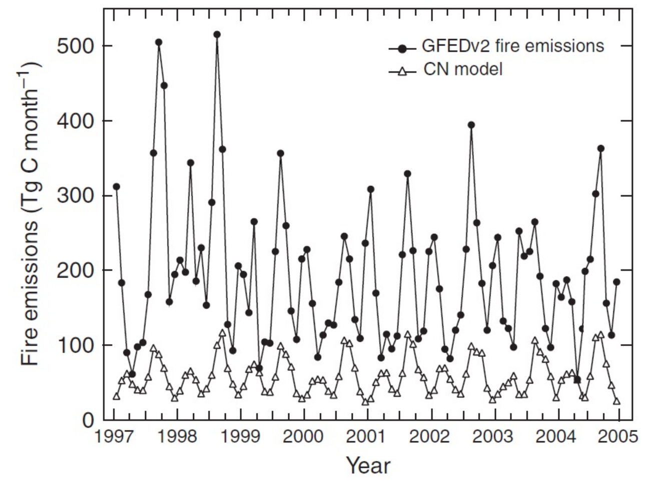
Comparison of model estimates with eddy covariance measurements from Sylvania Wilderness (Desai *et al.*, 2005), Harvard Forest (Barford *et al.*, 2001), and Walker Branch (Wilson & Baldocchi, 2001) sites from the AmeriFlux network. Both models under estimated seasonal variations in NEE and under predicted the rate of GPP increase at the onset of the growing season.



Annual cycle of atmospheric CO_2 at (a) Mould Bay, Canada (76˚N), (b) Storhofdi, Iceland (63˚N), (c) Carr, Colorado (aircraft samples from 6 km masl; 41˚N), (d) Azores Islands (39˚N), (e) Sand Island, Midway (28˚N), and (f) Kumakahi, Hawaii (20˚N). The observations are form Globalview and the model estimates were obtained using model fluxes from Experiment 1.4 and monthly impulse response functions from the TRANSCOM experiment.



Net primary production normalized by precipitation for EMDI NPP measurements and the models. CASA´ exhibits an increasingly high bias while CN exhibits a consistent low bias.

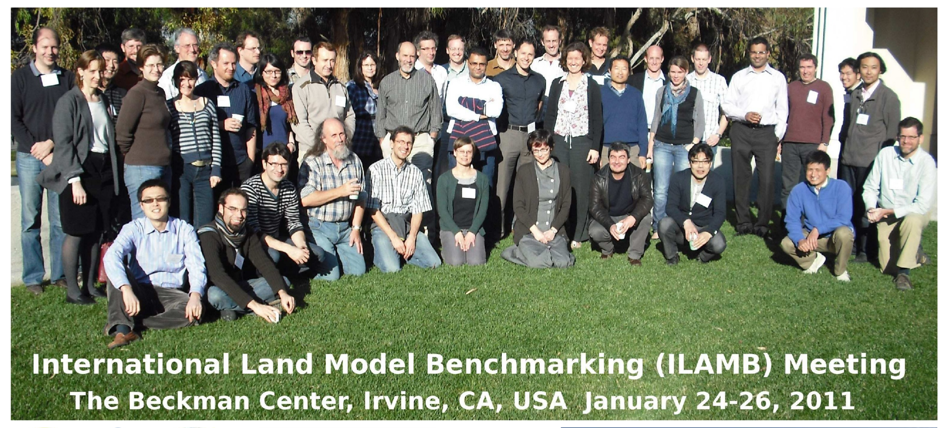


Global fire emissions from CN compared to the Global Fire Emissions Database version 2. The version of CASA´ analyzed here did not simulate fire emissions.

For more results, see

Randerson, James T., 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. September 2009. "Systematic Assessment of Terrestrial Biogeochemistry in Coupled Climate-Carbon Models." *Global Change Biology*, 15(9):2462–2484. doi:10.1111/j.1365-2486.2009.01912.x.

International Land Model Benchmarking (ILAMB) Project



The First ILAMB Meeting was co-organized by Forrest Hoffman, Chris Jones, Pierre Friedlingstein, and Jim Randerson. About 45 researchers participated from the United States, Canada, the United Kingdom, the Netherlands, France, Germany, Switzerland, China, Japan, and Australia.

The goals of the meeting were to:

1) coordinate the design of the first set of benchmarks for global models,
2) coordinate the carbon cycle and land model evaluations for TRENDY and CMIP5 results,
3) develop an implementation plan for application of ILAMB benchmarks to TRENDY and CMIP5 output,
4) decide upon an approach for developing ILAMB software, and
5) decide upon a future schedule and means to secure funding.

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.

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 benchmarks will be implemented to evaluate the existing TRENDY and CMIP5 model results.

For more information, see http://www.ilamb.org/

Initial ILAMB Benchmarks and Datasets

	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

An initial set of benchmarks and available observational data sets identified by the break-out groups is shown in this table.

Depending upon the type of measurements available, the annual mean, seasonal cycle, interannual variability, and long-term trend of the model results will be assessed.

Observational data sets span scales from site/point *in situ* measurements to global remote sensing observations.

Office of Science
U.S. DEPARTMENT OF ENERGY
Oak Ridge National Laboratory
Climate Change Science Institute
Cornell University
The University of Montana
Berkeley

Acknowledgements: Research partially sponsored by the Climate and Environmental Sciences Division (CESD) of the Office of Biological and Environmental Research (OBER) within the U.S. Department of Energy's Office of Science (SC). This research used resources of the National Center for Computational Sciences (NCCS) at Oak Ridge National Laboratory (ORNL), which is managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. The National Center for Atmospheric Research is operated by the University Corporation for Atmospheric Research (UCAR) and receives research funding primarily from the National Science Foundation (NSF).