A model-data comparison of gross primary productivity: Results from the North American Carbon Program site synthesis

Kevin Schaefer,[1] Christopher R. Schwalm,[2] Chris Williams,[3] M. Altaf Arain,[4] Alan Barr,[5] Jing M. Chen,[6] Kenneth J. Davis,[7] Dimitre Dimitrov,[8] Timothy W. Hilton,[9] David Y. Hollinger,[10] Elyn Humphreys,[11] Benjamin Poulter,[12] Brett M. Raczka,[7] Andrew D. Richardson,[13] Alok Sahoo,[14] Peter Thornton,[15] Rodrigo Vargas,[16,17] Hans Verbeeck,[18] Ryan Anderson,[19] Ian Baker,[20] T. Andrew Black,[21] Paul Bolstad,[22] Jiquan Chen,[23] Peter S. Curtis,[24] Ankur R. Desai,[22] Michael Dietze,[25] Danilo Dragoni,[26] Christopher Gough,[27] Robert F. Grant,[28] Lianhong Gu,[15] Atul Jain,[29] Chris Kucharik,[30] Beverly Law,[31] Shuguang Liu,[32] Erandathie Lokipitiya,[20] Hank A. Margolis,[33] Roser Matamala,[34] J. Harry McCaughey,[35] Russ Monson,[36] J. William Munger,[37] Walter Oechel,[38,39] Changhui Peng,[40] David T. Price,[8] Dan Ricciuto,[15] William J. Riley,[41] Nigel Roulet,[42] Hanqin Tian,[43] Christina Tonitto,[44] Margaret Torn,[41] Ensheng Weng,[45] and Xiaolu Zhou[40]

Received 20 January 2012; revised 16 May 2012; accepted 26 May 2012; published 19 July 2012.

[1] Accurately simulating gross primary productivity (GPP) in terrestrial ecosystem models is critical because errors in simulated GPP propagate through the model to introduce additional errors in simulated biomass and other fluxes. We evaluated simulated, daily average GPP from 26 models against estimated GPP at 39 eddy covariance flux tower sites across the United States and Canada. None of the models in this study match estimated GPP within observed uncertainty. On average, models overestimate GPP in winter, spring, and fall, and underestimate GPP in summer. Models overpredicted GPP under dry conditions and for temperatures below 0°C. Improvements in simulated soil moisture and ecosystem response to drought or humidity stress will improve simulated GPP under dry conditions. Adding a low-temperature response to shut down GPP for temperatures below 0°C will reduce the positive bias in winter, spring, and fall and improve simulated phenology. The negative bias in summer and poor overall performance resulted from mismatches between simulated and observed light use efficiency (LUE). Improving simulated GPP requires better leaf-to-canopy scaling and better values of model parameters that control the maximum potential GPP, such as $\varepsilon_{\max}$ (LUE), V_{cmax} (unstressed Rubisco catalytic capacity) or $J_{\max}$ (the maximum electron transport rate).

Citation: Schaefer, K., et al. (2012), A model-data comparison of gross primary productivity: Results from the North American Carbon Program site synthesis, *J. Geophys. Res.*, *117*, G03010, doi:10.1029/2012JG001960.

[1]National Snow and Ice Data Center, Cooperative Institute for Research in Environmental Sciences, University of Colorado at Boulder, Boulder, Colorado, USA.
[2]School of Earth Science and Environmental Sustainability, Northern Arizona University, Flagstaff, Arizona, USA.
[3]Graduate School of Geography, Clark University, Worcester, Massachusetts, USA.
[4]School of Geography and Earth Sciences and McMaster Centre for Climate Change, McMaster University, Hamilton, Ontario, Canada.
[5]Climate Research Division, Atmospheric Sciences and Technology Directorate, Environment Canada, Saskatoon, Saskatchewan, Canada.
[6]Department of Geography, University of Toronto, Toronto, Ontario, Canada.
[7]Department of Meteorology, Pennsylvania State University, University Park, Pennsylvania, USA.
[8]Northern Forestry Centre, Canadian Forest Service, Natural Resources Canada, Edmonton, Alberta, Canada.
[9]Department of Biology, University of New Mexico, Albuquerque, New Mexico, USA.
[10]Forest Service, United States Department of Agriculture, Durham, New Hampshire, USA.
[11]Department of Geography and Environmental Studies, Carleton University, Ottawa, Ontario, Canada.
[12]Laboratoire des Sciences du Climat et l'Environnement, Gif-sur-Yvette, France.

Corresponding author: K. Schaefer, National Snow and Ice Data Center, Cooperative Institute for Research in Environmental Sciences, University of Colorado at Boulder, Boulder, Colorado, USA. (kevin.schaefer@nsidc.org)

0148-0227/12/2012JG001960

1. Introduction

[2] Terrestrial gross primary productivity (GPP) is the total photosynthetic uptake or carbon assimilation by plants and is a key component of terrestrial carbon balance. GPP is the main carbon input to terrestrial ecosystems, noting relatively minor inputs by dissolved organic carbon, as well as deposition by rainwater and sedimentation [*Chapin et al.*, 2006]. GPP depends on climate, climate variability, disturbance history, water and nutrient availability, soil type, species composition, and community structure. Understanding how these factors influence GPP remains a challenge due to complex interactions and the difficulty in quantitatively measuring GPP directly at various temporal and spatial scales. Estimates of GPP are only available at eddy covariance flux tower sites for the past decade, so we depend on models to estimate GPP over long periods of time at regional and global scales, and to project future changes in GPP in response to climate change.

[3] Any error in simulated GPP will propagate through the model, introducing errors in simulated biomass and fluxes. If simulated GPP is too low or too high, then predicted leaf area index, wood biomass, crop yield, and soil biomass may also be too low or high [*Schaefer et al.*, 2008]. Net ecosystem exchange (NEE) is total ecosystem respiration (R_{eco}) minus GPP, with a positive NEE indicating a net release of CO_2 to the atmosphere. Autotrophic respiration depends on GPP and heterotrophic respiration depends on soil conditions and dead plant biomass, so errors in GPP readily propagate to errors in R_{eco} and simulated diurnal and seasonal cycles of NEE. Through representation of stomatal control on GPP and transpiration, errors in GPP introduce errors in simulated latent and sensible heat flux, which in turn can introduce error in simulated atmospheric circulation. GPP is a key carbon flux that needs to be simulated as accurately as possible to ensure the most reliable values of simulated biomass and surface fluxes.

[4] We can classify GPP models into enzyme kinetic (EK), light use efficiency (LUE), or empirical models. EK models represent leaf-scale enzyme-kinetics with electron and product transport limits on simulated GPP [*Farquhar et al.*, 1980; *Caemmerer and Farquhar*, 1981; *Collatz et al.*, 1991]. Nearly all EK models include a representation of stomatal conductance balancing GPP against water loss through leaf stomata [*Collatz et al.*, 1991; *Collatz et al.*, 1992]. Most stomatal conductance models are based on empirical correlations between conductance, photosynthesis, and either relative humidity [*Ball et al.*, 1987] or vapor pressure deficit (VPD) [*Wang and Leuning*, 1998]. LUE models estimate either GPP or net primary productivity (NPP) by multiplying incident photosynthetically active radiation (PAR) by a remotely sensed fraction of PAR absorbed by the vegetation (f_{PAR}) and an energy to biomass conversion factor (typically called light use efficiency) [*Monteith*, 1972; *Field*, 1995; *Prince and Goward*, 1995; *Landsberg and Waring*, 1997; *Goetz et al.*, 1999; *Running et al.*, 2000; *Sims et al.*, 2006; *Zhao et al.*, 2007; *Sjöström et al.*, 2009; *As-syakur et al.*, 2010]. Finally, empirical models use statistical relationships between observed environmental conditions and GPP estimated from eddy covariance flux data, which is then expanded to regional or global scales using various reanalysis weather products [*Beer et al.*, 2010].

[13]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, USA.
[14]Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey, USA.
[15]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA.
[16]Departamento de Biología de la Conservación, Centro de Investigación Científica y de Educación Superior de Ensenada, Ensenada, Baja California, Mexico.
[17]Delaware Environmental Institute, University of Delaware, Newark, Delaware, USA.
[18]Laboratory of Plant Ecology, Department of Applied Ecology and Environmental Biology, Ghent University, Ghent, Belgium.
[19]Numerical Terradynamic Simulation Group, University of Montana, Missoula, Montana, USA.
[20]Department of Zoology, University of Colombo, Colombo, Sri Lanka.
[21]Faculty of Land and Food Systems, University of British Columbia, Vancouver, British Columbia, Canada.
[22]Department of Forest Resources, University of Minnesota, Saint Paul, Minnesota, USA.
[23]Department of Environmental Sciences (DES), University of Toledo, Toledo, Ohio, USA.
[24]Department Evolution, Ecology, and Organismal Biology, Ohio State University, Columbus, Ohio, USA.
[25]Department of Earth and Environment, Boston University, Boston, Massachusetts, USA.
[26]Department of Geography, Indiana University, Bloomington, Indiana, USA.
[27]Department of Biology, Virginia Commonwealth University, Richmond, Virginia, USA.
[28]Department of Renewable Resources, University of Alberta, Edmonton, Alberta, Canada.
[29]Department of Atmospheric Science, University of Illinois, Urbana, Illinois, USA.
[30]Department of Agronomy and Nelson Institute Center for Sustainability and the Global Environment, University of Wisconsin–Madison, Madison, Wisconsin, USA.
[31]Department of Forest Ecosystems and Society, Oregon State University, Corvallis, Oregon, USA.
[32]Earth Resources Observation and Science, United States Geological Survey, Sioux Falls, South Dakota, USA.
[33]Centre d'Étude de la Forêt, Laval University, Quebec City, Quebec, Canada.
[34]Biosciences Division, Argonne National Laboratory, Argonne, Illinois, USA.
[35]Department of Geography, Queen's University, Kingston, Ontario, Canada.
[36]Laboratory for Tree Ring Research, University of Arizona, Tucson, Arizona, USA.
[37]School of Engineering and Applied Sciences and Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts, USA.
[38]Department of Biology and the Global Change Research Group, San Diego State University, San Diego, California, USA.
[39]CRI Fondazione Edmund Mach, San Michele all'Adige, Italy.
[40]Instititut of Environment Sciences, Biology Science Department, University of Quebec at Montreal, Montreal, Quebec, Canada.
[41]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA.
[42]Department of Geography, McGill University, Montreal, Quebec, Canada.
[43]International Center for Climate and Global Change Research, School of Forestry and Wildlife Sciences, Auburn University, Auburn, Alabama, USA.
[44]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York, USA.
[45]Department of Botany and Microbiology, University of Oklahoma, Norman, Oklahoma, USA.

These GPP models vary greatly in intended application, complexity, and their representation of physical and biological processes.

[5] The modeling community currently lacks a quantitative evaluation of multiple GPP models to gauge overall performance across different ecosystems and help prioritize long-term model development. Many modeling teams compare simulated NEE to observed NEE measured at single points using eddy covariance techniques [*Baldocchi et al.*, 2001; *Grant et al.*, 2010]. Others compare against NEE for large regions estimated from transport inversions that are optimally consistent with observations of atmospheric CO_2 concentration [*Gurney et al.*, 2002; *Peters et al.*, 2010]. However, comparisons with observed NEE do not distinguish between R_{eco} and GPP and provide little information on model performance relative to GPP. Fortunately, NEE measured by eddy covariance techniques can be partitioned into R_{eco} and GPP: a temperature function is tuned to nighttime R_{eco}, the function is used to calculate daytime R_{eco}, and the estimated GPP is the daytime R_{eco} minus the daytime NEE [*Desai et al.*, 2008; *Lasslop et al.*, 2010]. There are assessments in the literature of how well terrestrial biosphere models simulate GPP, but they focused on a single or small number of models compared to GPP estimated from eddy covariance data at a small number of towers [e.g., *Thornton et al.*, 2002; *Schaefer et al.*, 2008; *Verbeeck et al.*, 2008]. These studies used different techniques to estimate GPP from observed NEE [*Desai et al.*, 2008], making it difficult to differentiate between errors in the partitioning technique and true model-data mismatches. The evaluations were run at different sites and used different input weather, making it difficult to isolate input errors from true model-data mismatches. The performance measures used in these evaluations are difficult to compare because most used qualitative performance criteria while those with quantitative performance measures used different statistical techniques and quantities. None of these model evaluations account for uncertainty in estimated GPP due to uncertainty in the eddy covariance data and partitioning techniques. The actual GPP value lies within the range defined by uncertainty, so the ideal performance target of any model is to match the observed GPP within uncertainty. These studies provide insight into the performance of individual models. However, the differences among evaluations make it very difficult to compare and synthesize the results to identify strengths and weaknesses common to all terrestrial biosphere models and to determine what changes will provide the greatest improvements in simulated GPP.

[6] We hypothesize that model performance depends on 1) model structure and 2) how models simulate GPP response to changing environmental conditions. Model structure refers to differences in how models represent various physical and biological processes, such as LUE versus EK models. To test our hypotheses, we compared simulated GPP from 26 models against estimated GPP at 39 eddy covariance flux tower sites in the North American Carbon Program (NACP) site-level interim synthesis. Our analysis includes observation uncertainty to determine if the models hit the desired performance target: matching observed GPP within uncertainty. The number and variety of models and sites in the NACP site synthesis are sufficient to identify the strengths and weaknesses common to all GPP models and, most importantly, how to improve the models.

2. Methods

2.1. Estimated GPP

[7] Our analysis used daily average GPP estimated at 39 eddy covariance flux tower sites (Table 1). Observed NEE at all towers were processed and partitioned into GPP and R_{eco} using standard techniques [*Barr et al.*, 2004]. The NACP site synthesis included 47 tower sites, but we did not include those sites where the GPP partitioning was not done or the algorithm failed to converge. The chosen sites represented eight major biome types across North America except tundra, with 24 sites from the AmeriFlux network and 15 sites from the Fluxnet Canada Research Network/ Canadian Carbon Program. GPP partitioning was not done for tundra sites due to large data gaps in winter and the lack of nighttime data in summer to train the R_{eco} model. NACP site synthesis used the International Global Biosphere Program (IGBP) biome classifications [*Loveland et al.*, 2000]. Some models were designed for specific biome types, such as forest or agriculture sites only, so not every model simulated all sites, resulting in a total of 627 simulations and an average of 16 simulations per site.

[8] Observed, hourly NEE at each site was gap-filled and decomposed into hourly R_{eco} and GPP using a standard procedure [*Barr et al.*, 2004] and converted into 24-hour daily averages. Before processing, observed NEE was screened to remove outliers [*Papale et al.*, 2006] and exclude data during periods of low turbulence based on a friction velocity threshold (A. G. Barr et al., Use of change-point detection for u*–threshold evaluation for the North American Carbon Program interim synthesis, manuscript in preparation, 2012). R_{eco} was set equal to observed, nighttime NEE and fitted to an empirical model based on air and soil temperature using a moving window approach. The function was then used to calculate the daytime R_{eco}, which was subtracted from daytime NEE to get GPP. Finally, gaps in GPP were filled using an empirical model that was tuned to the estimated GPP values. GPP was set to zero when the soil was frozen and the air temperature was below 0°C, assuming air temperature represents the temperature of the entire system. The partitioning procedure occasionally produced negative GPP values at dawn and dusk at most sites and occasionally for entire days at the dry, grassland sites. The negative GPP values typically greatly exceeded estimates of random uncertainty and probably resulted from errors in the R_{eco} temperature response or the fact that the partitioning algorithm did not account for influences of soil moisture on GPP. We set all negative GPP values to zero and transferred the flux to R_{eco} to maintain the observed NEE, and recalculated the daily average GPP. One third of the models used a daily time step, so we calculated the daily average as the average rate of GPP over a 24-hour period using both estimated and gap-filled values.

[9] Although treated here as "observed" GPP, they are not strictly observed, but rather estimated from observed tower-based NEE. The GPP includes all the strengths, weaknesses, and assumptions of the original NEE observations. The energy budget at flux towers does not balance because the

Table 1. Summary of Site Characteristics

Site	IGBP Class[a]	Description	Longitude (deg)	Latitude (deg)	Start (yr)	Stop (yr)	Number Models	Reference(s)
CA-Ca1	ENF	Campbell River, Mature Douglas-fir	−125.3	49.9	1998	2006	23	*Krishnan et al.* [2009]; *Humphreys et al.* [2006]
CA-Ca2	ENF	Campbell River, Douglas-fir, clearcut	−125.3	49.9	2001	2006	8	*Krishnan et al.* [2009]; *Humphreys et al.* [2006]
CA-Ca3	ENF	Campbell River, Douglas-fir, juvenile	−124.9	49.5	2002	2006	8	*Krishnan et al.* [2009]; *Humphreys et al.* [2006]
CA-Gro	MF	Groundhog River	−82.2	48.2	2004	2006	17	*McCaughey et al.* [2006]
CA-Let	GRA	Lethbridge Grassland	−112.9	49.7	1997	2007	22	*Flanagan and Adkinson* [2011]
CA-Mer	WET	Eastern Peatland, Mer Bleue	−75.5	45.4	1999	2006	18	*Roulet et al.* [2007]
CA-Oas	DBF	BERMS, Old Aspen	−106.2	53.6	1997	2006	23	*Krishnan et al.* [2006]; *Barr et al.* [2006]
CA-Obs	ENF	BERMS, Old Black Spruce	−105.1	54.0	2000	2006	22	*Krishnan et al.* [2008]; *Kljun et al.* [2006]
CA-Ojp	ENF	BERMS, Old Jack Pine	−104.7	53.9	2000	2006	18	*Kljun et al.* [2006]
CA-Qfo	ENF	Quebec, Mature Black Spruce	−74.3	49.7	2004	2006	17	*Bergeron et al.* [2008]; *Bergeron et al.* [2007]
CA-SJ1	ENF	BERMS, Jack Pine, 1994 Harvest	−104.7	53.9	2002	2005	7	*Zha et al.* [2009]
CA-SJ2	ENF	BERMS, Jack Pine, 2002 Harvest	−104.6	53.9	2003	2006	8	*Zha et al.* [2009]
CA-SJ3	ENF	BERMS, Jack Pine, 1995 Harvest	−104.6	53.9	2005	2006	7	*Zha et al.* [2009]
CA-TP4	ENF	Turkey Point, Mature	−80.4	42.7	2002	2007	17	*Arain and Restrepo* [2005]; *Peichl and Arain* [2006]
CA-WP1	WET	Western Peatland, LaBiche River	−112.5	55.0	2003	2007	14	*Flanagan and Syed* [2011]
US-ARM	CRO	ARM, Southern Great Plains	−97.5	36.6	2000	2007	17	*Fischer et al.* [2007]
US-Dk3	ENF	Duke Forest, Loblolly Pine	−79.1	36.0	1998	2005	17	*Oren et al.* [2006]; *Stoy et al.* [2006]
US-Ha1	DBF	Harvard Forest, EMS Tower	−72.2	42.5	1991	2006	23	*Urbanski et al.* [2007]
US-Ho1	ENF	Howland Forest, Main Tower	−68.7	45.2	1996	2004	23	*Richardson et al.* [2009]
US-IB1	CRO	Fermi Lab, agriculture	−88.2	41.9	2005	2007	16	*Post et al.* [2004]
US-IB2	GRA	Fermi, prairie	−88.2	41.8	2004	2007	17	*Post et al.* [2004]
US-Los	WET	Lost Creek	−90.0	46.1	2000	2006	12	*Sulman et al.* [2009]
US-Me2	ENF	Metolius, Intermediate-aged Ponderosa Pine	−121.6	44.5	2002	2007	20	*Thomas et al.* [2009]
US-Me3	ENF	Metolius, Ponderosa Pine, young (2)	−121.6	44.3	2004	2005	8	*Vickers et al.* [2009]
US-Me5	ENF	Metolius, Ponderosa Pine, young (1)	−121.6	44.4	1999	2002	8	*Law et al.* [2001]
US-MMS	DBF	Morgan Monroe State Forest	−86.4	39.3	1999	2006	16	*Schmid et al.* [2000]
US-MOz	DBF	Missouri Ozark	−92.2	38.7	2004	2007	17	*Gu et al.* [2006]
US-Ne1	CRO	Mead, Irrigated maize	−96.5	41.2	2001	2006	17	*Verma et al.* [2005]; *Suyker and Verma* [2010]
US-Ne2	CRO	Mead, Irrigated maize or soybean	−96.5	41.2	2001	2006	17	*Verma et al.* [2005]; *Suyker and Verma* [2010]
US-Ne3	CRO	Mead, Rainfed	−96.4	41.2	2001	2006	23	*Verma et al.* [2005]; *Suyker and Verma* [2010]
US-NR1	ENF	Niwot Ridge	−105.5	40.0	1998	2007	17	*Monson et al.* [2002]; *Monson et al.* [2005]
US-PFa	MF	Park Falls/WLEF	−90.3	45.9	1995	2005	14	*Davis et al.* [2003]
US-Shd	GRA	Shidler	−96.7	36.9	1997	2000	15	*Suyker et al.* [2003]
US-SO2	CSH	Sky Oaks, Old	−116.6	33.4	1998	2006	16	*Luo et al.* [2007]
US-Syv	MF	Sylvania Wilderness Area	−89.3	46.2	2001	2006	15	*Desai et al.* [2005]
US-Ton	WSA	Tonzi Ranch	−121.0	38.4	2001	2007	15	*Ma et al.* [2007]
US-UMB	DBF	UMBS	−84.7	45.6	1999	2006	22	*Gough et al.* [2008]
US-Var	GRA	Vaira Ranch	−121.0	38.4	2000	2007	17	*Ma et al.* [2007]
US-WCr	DBF	Willow Creek	−90.1	45.8	1998	2006	16	*Cook et al.* [2004]

[a]IGBP biome class definitions: CRO, Croplands; CSH, Closed Shrublands; DBF, Deciduous Broadleaf Forest; ENF, Evergreen Needleleaf Forest; GRA, Grasslands; MF, Mixed Forests; WET, Permanent Wetlands; WSA, Woody Savannas.

eddy covariance technique captures small-scale turbulent fluxes less than 1 km, but can underestimate latent and sensible heat fluxes due to large-scale eddies on the order of 10 km [*Foken*, 2008]. Assuming similarity with the lack of surface energy balance closure, the measured NEE might be between 15 and 20% less than the actual values [*Foken*, 2008], indicating a potential underestimate of GPP. Eddy covariance techniques can underestimate nighttime R_{eco}, depending on the threshold used to filter out R_{eco} under stable conditions, also resulting in an underestimate of GPP. The empirical formulation for R_{eco} used to estimate GPP from NEE assumed air temperature represents the temperature of the entire system, underestimating GPP when the canopy temperature is greater than zero and the air temperature is less than zero. Last, the partitioning algorithm did not account for how soil moisture and other factors control R_{eco}, resulting in errors in the estimated GPP (as evidenced by occurrences of negative GPP at some grassland sites).

[10] Total GPP uncertainty included gap filling algorithm uncertainty, partitioning uncertainty, random uncertainty, and threshold friction velocity (u*) uncertainty, summed in quadrature. Summing in quadrature assumes that these sources of error are uncorrelated. We did not correct for potential biases due to the lack of energy closure or underestimates of R_{eco} at night. Random and u* filtering uncertainty was estimated using a Monte Carlo technique

Table 2. Summary of Model Characteristics

Model	Number Sites	Time Step	Soil Layers[a]	Phenology[b]	Nitrogen Cycle	GPP Model[c]	Leaf-to-Canopy	Reference
AgroIBIS	5	Hourly	11	Prognostic	Yes	EK	Big-Leaf	*Kucharik and Twine* [2007]
BEPS	10	Daily	3	Semi-prognostic	Yes	EK	2-Leaf	*Liu et al.* [1999]
Biome-BGC	33	Daily	1	Prognostic	Yes	EK	2-Leaf	*Thornton et al.* [2005]
Can-IBIS	24	Hourly	7	Prognostic	Yes	EK	2-Leaf	*Liu et al.* [2005]
CN-CLASS	28	Hourly	3	Prognostic	Yes	EK	2-Leaf	*Arain et al.* [2006]
DLEM	30	Daily	2	Semi-prognostic	Yes	EK	2-Leaf	*Tian et al.* [2010]
DNDC	5	Daily	10	Prognostic	Yes	LUE	Big-Leaf	*Li et al.* [2010]
Ecosys	35	Hourly	15	Prognostic	Yes	EK	2-Leaf	*Grant et al.* [2009]
ED2	24	Hourly	9	Prognostic	Yes	EK	2-Leaf	*Medvigy et al.* [2009]
EDCM	9	Daily	10	Prognostic	Yes	LUE	Big-Leaf	*Liu et al.* [2003]
ISAM	13	Hourly	10	Prognostic	Yes	LUE	2-Leaf	*Yang et al.* [2009]
ISOLSM	9	Hourly	20	Observed	No	EK	2-Leaf	*Riley et al.* [2002]
LoTEC	10	Hourly	14	Prognostic	No	EK	Big-Leaf	*Hanson et al.* [2004]
LPJ	26	Daily	2	Prognostic	No	EK	Big-Leaf	*Sitch et al.* [2003]
MODIS_5.0	38	Daily	0	Observed	No	LUE	Big-Leaf	*Heinsch et al.* [2003]
MODIS_5.1	37	Daily	0	Observed	No	LUE	Big-Leaf	*Heinsch et al.* [2003]
MODIS_alg	39	Daily	0	Observed	No	LUE	Big-Leaf	*Heinsch et al.* [2003]
ORCHIDEE	32	Hourly	2	Prognostic	No	EK	Big-Leaf	*Krinner et al.* [2005]
SiB3	28	Hourly	10	Observed	No	EK	Big-Leaf	*Baker et al.* [2008]
SiBCASA	32	Hourly	25	Semi-prognostic	No	EK	Big-Leaf	*Schaefer et al.* [2009]
SiBcrop	5	Hourly	10	Prognostic	Yes	EK	Big-Leaf	*Lokupitiya et al.* [2009]
SSiB2	39	Hourly	3	Observed	No	EK	Big-Leaf	*Zhan et al.* [2003]
TECO	32	Hourly	10	Prognostic	No	EK	2-Leaf	*Weng and Luo* [2008]
TRIPLEX	6	Daily	0	Observed	Yes	LUE	Big-Leaf	*Peng et al.* [2002]

[a]Zero soil layers indicate the model does not have a prognostic submodel for soil temperature and moisture.
[b]Observed phenology means the model uses remote sensing data to determine leaf area index (LAI) and gross primary productivity (GPP). Semi-prognostic means that remote sensing data is used to specify either LAI or GPP, but not both.
[c]GPP model types: EK (enzyme kinetic) and LUE (light use efficiency).

[*Richardson and Hollinger*, 2007; A. G. Barr et al., manuscript in preparation, 2012]. A. G. Barr et al. generated a synthetic flux time series using the gap-filling algorithm, randomly introduced artificial gaps, added noise, and then refilled the gaps. Repeating the process 1000 times for each site-year produced probability distribution functions with the 2.5 and 97.5 percentiles representing uncertainty. Assuming the GPP uncertainty due to the gap-filling algorithm was the same as that for NEE, the GPP gap filling uncertainty was the fraction of filled GPP values for each day times the standard deviation of multiple algorithms [*Moffat et al.*, 2007]. Partitioning uncertainty was based on the standard deviation of multiple partitioning algorithms [*Desai et al.*, 2008]. Total uncertainty generally increased with the magnitude of GPP and varied from a minimum of $\sim$1 μmol m^{-2} s^{-1} in winter to 2–4 μmol m^{-2} s^{-1} in summer. Random uncertainty dominated over other sources of uncertainty, ranging from $\sim$90% of total uncertainty in summer months to $\sim$50% of total uncertainty in winter months.

2.2. Modeled GPP

[11] Our analysis used simulated GPP from 24 different models (Table 2) plus two model averages. We used the model characteristics in Table 2 as covariates to determine if the different representations of physical and biological processes produce statistically significant differences in model performance relative to GPP. All the EK models included the effects of stomatal conductance and all but two used the Ball-Berry stomatal conductance model [*Ball et al.*, 1987]. The leaf-to-canopy column indicates whether the strategy for scaling from a single leaf to the entire canopy accounts for the effects of diffuse light on shaded leaves (2-leaf) or not (big-leaf). To determine if resolving the diurnal cycle improved performance, we included the ensemble average of all models and the average of all models that resolved the diurnal cycle. Including the two ensemble averages at each site, we have a total of 627 simulations or 4242 site-years of model output with an average of 23 simulations per model.

[12] We included estimates of MODIS GPP from Collection 5.0 and Collection 5.1, [*Heinsch et al.*, 2003; *Running et al.*, 2004]. We extracted a 3 by 3 pixel window of 8-day maximum composite GPP values at 1 km^2 spatial resolution, with the center pixel containing the tower site [*Distributed Active Archive Center for Biogeochemical Dynamics*, 2010]. We filtered out low quality pixels using the simple binary quality control flag to remove the effects of potential cloud contamination and averaged the rest of the pixels to represent the GPP at each site. We linearly interpolated between 8-day composite values to obtain daily GPP.

[13] The MODIS GPP from Collection 5.0 and Collection 5.1 were not based on observed meteorology, so we also calculated GPP using gap-filled observed weather and the MODIS algorithm [*Heinsch et al.*, 2003; *Running et al.*, 2004]:

$$GPP = 0.45\varepsilon_{\max} F_{SW} f_{PAR} S_{VPD} S_T, \quad (1)$$

where $\varepsilon_{\max}$ is the maximum light use efficiency, F_{SW} is the incident shortwave radiation flux, f_{PAR} is the absorbed fraction of PAR, S_{VPD} is the VPD scaling factor, and S_T is the air temperature scaling factor. S_{VPD} represents the GPP response to drought and humidity stress and S_T represents the GPP response to temperature, with both varying between zero and one. We used the MODIS Biome-Property-Look-Up-Table [*Zhao and Running*, 2010] and daily f_{PAR} values interpolated from the monthly mean GIMMSg NDVI data set [*Tucker et al.*, 2005]. Although the complexity and sophistication varies widely, all the models have a GPP

formulation similar to MODIS: a peak potential rate times the amount of absorbed light, multiplied by a series of scaling factors representing how GPP responds to changing environmental conditions. The scaling factors represent the ratio of actual to a reference or optimal GPP and vary between zero and one.

[14] All models used gap-filled observed weather from each tower site [*Ricciuto et al.*, 2009; http://www.nacarbon.org/nacp/] with input parameters and biophysical characteristics derived from local observations, such as soil texture. Missing air temperature, atmospheric humidity, shortwave radiation, and precipitation data were filled using DAYMET [*Thornton et al.*, 1997] or the nearest available climate station in the National Climatic Data Center's Global Surface Summary of the Day (GSOD) database. Daily GSOD and DAYMET data were temporally downscaled to hourly or half-hourly values using the phasing from observed mean diurnal cycles calculated from a 15-day moving window. When station data were unavailable, a 10-day running mean diurnal cycle was used [*Ricciuto et al.*, 2009; http://nacp.ornl.gov/docs/Site_Synthesis_Protocol_v7.pdf]. The models were run for as many years as required, repeating the gap-filled weather, until they reached steady state initial conditions where R_{eco} balances GPP and the average NEE over the entire simulation is near zero. The steady state assumption influences R_{eco}, but has little or no effect on simulated GPP. All models used their standard values for various biophysical parameters except LoTEC, which used optimized parameter values obtained through data assimilation [*Ricciuto et al.*, 2011].

2.3. Model Performance

[15] We quantified model performance using a statistical analysis of model-data residuals using daily average GPP. We first calculated residuals:

$$r_i = (GPP_{si} - GPP_{oi}), \tag{2}$$

where r_i is the residual for the i^{th} model-data pair, GPP_{si} is simulated daily average GPP, and GPP_{oi} is estimated daily average GPP. Bias is the residual mean and the Root Mean Square Error (RMSE) is the residual standard deviation. X^2 is the mean of residuals normalized by uncertainty:

$$X^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{r_i}{\varepsilon_i}\right)^2, \tag{3}$$

where n is the number of residuals and ε_i is the uncertainty for the i^{th} daily average estimated GPP. We filtered out ~0.1% of daily estimated GPP values with $\varepsilon_i \leq 0.3$ μmol m^{-2} s^{-1} which produced extreme outlier X^2 values that skewed our results. Such unrealistically small ε_i values occasionally occurred when GPP was near zero. We did not filter out daily average GPP values based on the number of filled values per day because these values have higher uncertainty and a proportionally lower influence on X^2.

[16] The ideal target for any model is $X^2 < 1.0$, which means, on average, the residuals are less than uncertainty or the model matches observations within measurement uncertainty. Variations of X^2 within this target range have no meaning relative to model performance. A model with an X^2 value of 0.8, for example is not "better" than a model with an X^2 value of 0.9, since both models show no statistically significant differences with observations. Consequently, we identified performance categories based on ranges of X^2 values. An X^2 value of ≤1.0 indicated good model performance. An X^2 value between 1.0 and 2.0 indicated marginal model performance, where the model-data mismatch is on the order of two times the observation uncertainty. An X^2 value of >2.0 indicated poor model performance, where the model-data mismatch is several times the observation uncertainty. An X^2 value of 9, for example, indicates that the model-data mismatch is, on average, three times the uncertainty.

[17] To test our hypothesis that model structure influence performance, we aggregated the performance measures by model, model characteristic, site, and month-of-year. To identify any statistically significant differences in performance based on how the models represented various physical and biological processes, we aggregated performance measures by the model structural characteristics listed in Table 2. To evaluate seasonal variation in performance parameters, we aggregated by month-of-year, where January is the average of all Januaries, February the average of all Februaries, etc.

[18] To test our hypothesis that model performance depends on how they represent the GPP response to changing environmental conditions, we compared observed and simulated environmental response curves. We sorted the daily average GPP values into bins based on daily average values of input driver variables and calculated the mean, standard deviation, and uncertainty of the daily average GPP for each bin. We focused on downwelling shortwave radiation flux, air temperature, and relative humidity. The relative humidity response function reflects the reduction in GPP due to stomatal closure under drier atmospheric conditions. VPD, the difference between saturated and actual water vapor pressure, also reflects stomatal closure, but varies with temperature such that the range and magnitude varies among sites. Relative humidity always varies between zero and one and greatly simplifies our analysis by allowing easy comparison among sites. More importantly, 16 of the 20 models in this analysis that account for stomatal conductance used the Ball-Berry stomatal conductance model, which is based on relative humidity. Each model's mathematical formulation and associated parameters values determined the shape of the simulated response curves. We compared simulated and observed shape characteristics, such as slope, to isolate those model formulations or parameters that determine model performance. The response function for each driver variable differs slightly from site to site and is only weakly correlated to response functions for other driver variables. For example, the optimal temperature for GPP is different for each site but does not depend on humidity or light. Consequently, we made no attempt to remove covariance between driver variables.

3. Results

3.1. Performance Summary

[19] None of the models in this study achieved a good overall performance of X^2 less than one for all sites (Figure 1). LoTEC, which was optimized against flux data and DLEM achieved marginal performance, while SSIB2 and TECO had large X^2 values due to large biases. Generally speaking, higher RMSE resulted in higher X^2, but the

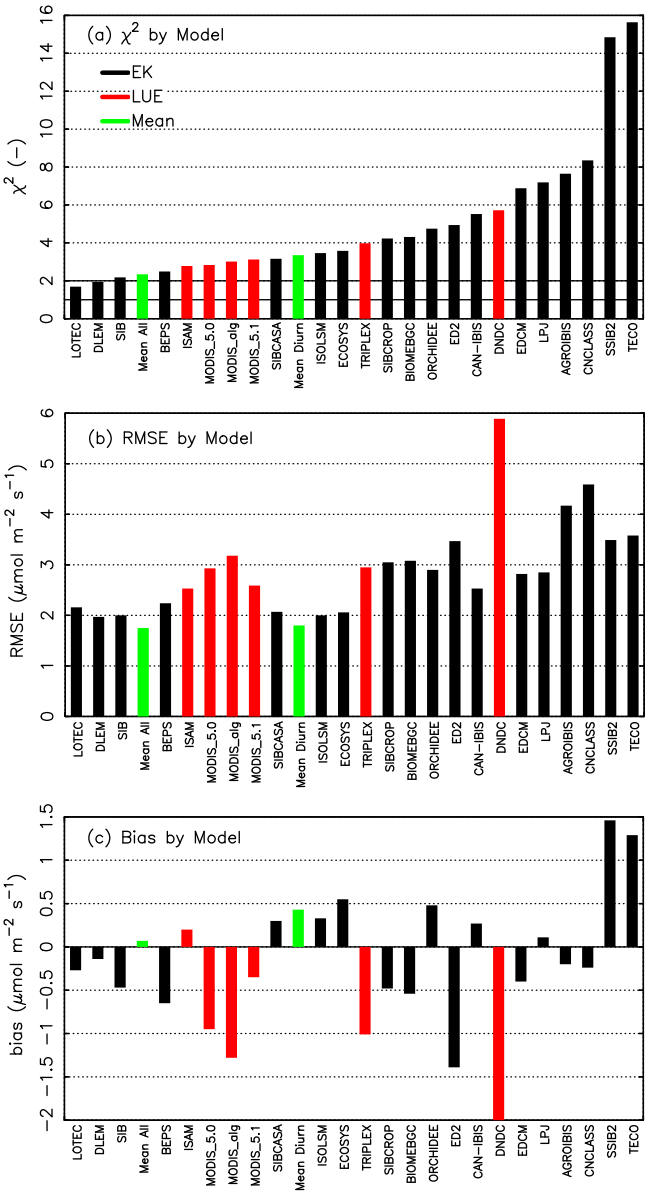

Figure 1. Overall model performance per model for all sites as defined by (a) X^2, (b) root mean squared error (RMSE), and (c) bias. The models are arranged in order of increasing X^2 and color-coded for enzyme kinetic (EK), light use efficiency (LUE) and model means. An X^2 value of <1.0 indicates good performance, an X^2 value of 1.0–2.0 indicates marginal performance, and an X^2 value of >2.0 indicates poor performance.

relationship was weak because X^2 accounts for uncertainty and bias while RMSE does not. There was no relationship between model bias and either RMSE or X^2. The RMSE fell within a narrow performance range, with a mean and standard deviation of 2.8 ± 1.0 μmol m^{-2} s^{-1}. On average, the models underestimated GPP (negative bias) with a mean bias of −0.3 μmol m^{-2} s^{-1}, but the spread between models was very large compared to the mean, with a bias standard deviation of ±0.6 μmol m^{-2} s^{-1}.

[20] Figure 2 shows that models, on average, underestimated GPP in summer (negative bias) and overestimated GPP in winter, spring, and fall (positive bias). These seasonal biases of opposite sign tended to cancel, resulting in the lower overall biases seen in Figure 1. Figure 2 shows the average monthly bias, but every month showed both positive and negative biases for individual models with standard deviations ranging from two to ten times the mean bias. The estimated GPP is smaller in spring and fall, with correspondingly smaller uncertainties, which magnified the relatively small model biases to produce slight peaks in X^2 in spring and fall. The models performed worst in the summer and the best in winter, indicating the models properly shut down GPP during winter, but the real challenge is to capture GPP dynamics during the growing season.

[21] The models generally performed the best at forest sites and the worst at crop, grassland, and savanna sites (Figure 3). The models did not show good overall performance at any site, but did show marginal performance at seven sites. The models performed best for deciduous broadleaf, mixed forest, and evergreen needleleaf biome types and all of the ten sites with the best overall performance were forest. The spread in performance within biome types was broad: two of the ten sites with the worst performance were evergreen needleleaf forest sites. However, seven of the ten sites with the worst model performance were crop, grassland or savanna sites.

[22] The models showed a large spread in both the magnitude and timing of the simulated GPP seasonal cycle, as indicated by the large spread in Figure 4. The models captured the basic observed seasonal pattern in GPP with near-zero values in winter and a peak value in mid-summer, so the standard deviation of the seasonal cycle is an alternative measure of seasonal amplitude. Consequently, the radial

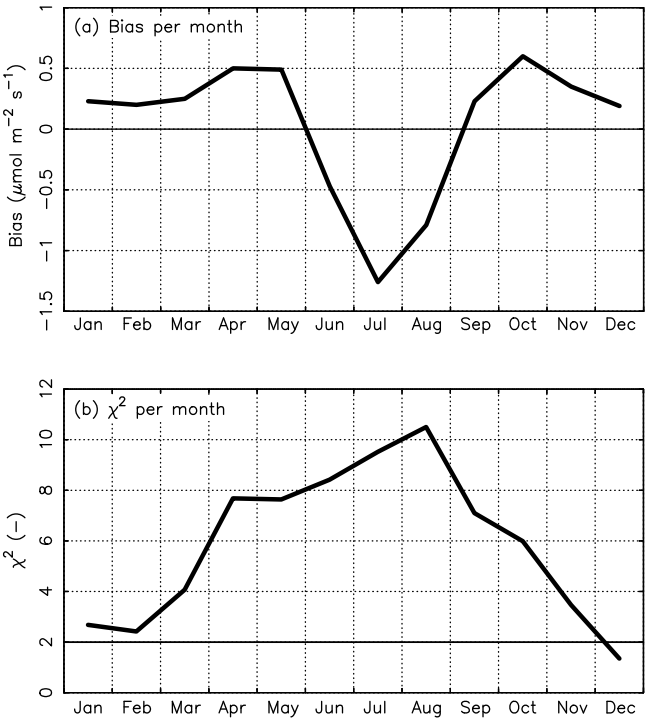

Figure 2. The monthly average bias in (a) simulated GPP and (b) monthly X^2 based on all 627 simulations from all models. An X^2 < 2.0 indicates marginal performance.

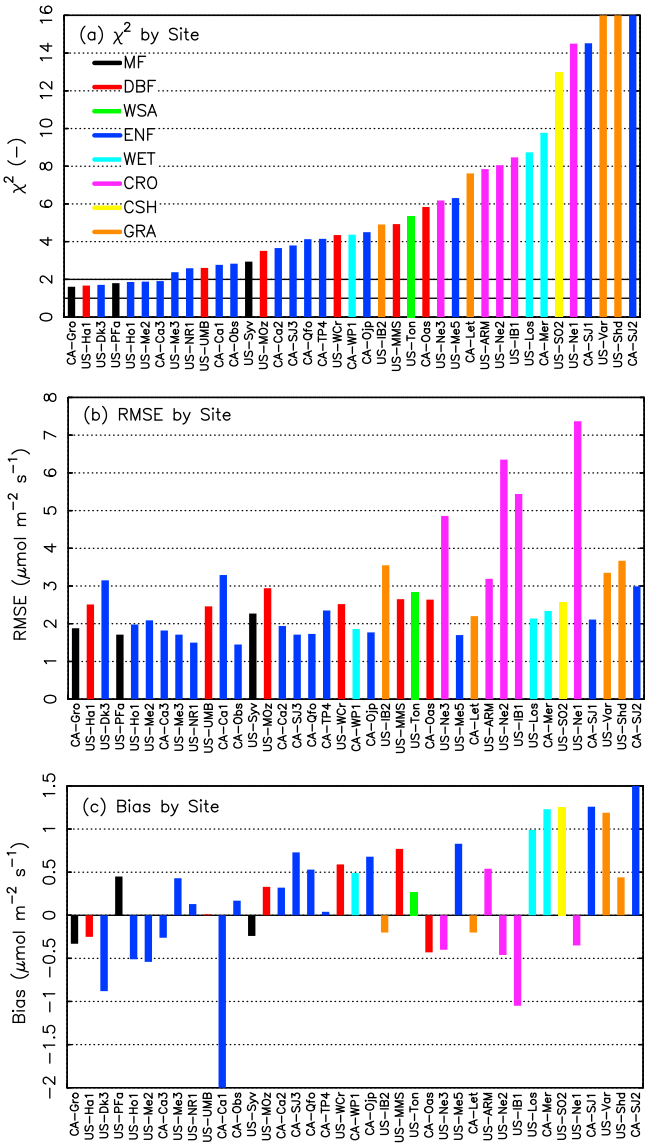

Figure 3. Overall performance per site for all models as defined by (a) X^2, (b) RMSE, and (c) bias. The sites are arranged in order of increasing X^2 and color-coded by biome type.

distance in Figure 4 is effectively the ratio of simulated to estimated seasonal amplitude of GPP, an alternative measure of bias. Ratios less than 1.0 indicate the model underestimates the GPP seasonal amplitude (negative bias). The models had standard deviation ratios ranging from 0.5 to 1.4, which means the simulated GPP ranged from 50% to 140% of estimated GPP. The models showed correlations between 0.6 and 0.9, indicating they varied widely in how well they captured the timing of the estimated GPP seasonal cycle, regardless of how well they captured the magnitude. For example, LOTEC, ISOLSM, and LPJ all had standard deviation ratios near one, consistent with the low biases seen in Figure 1. However, these three models showed progressively smaller correlations and correspondingly larger X^2, indicating that simulated phenology and associated phasing of the GPP seasonal cycle played an important role in determining overall model performance.

[23] The two model means showed the highest correlation with observations. This indicates that model-data mismatches associated with the timing of the GPP seasonal cycle can partly cancel out when averaging the results from multiple models. Essentially, the ensemble mean gave better results than any single model alone. Although errors in timing cancel, an ensemble mean does not eliminate overall bias. The standard deviation ratios for the two model means in Figure 4 reflected the overall, average negative bias of all the models in our analysis. Although there was positive bias in winter, spring, and fall, the negative bias in summer dominated and, on average, the models as a whole underestimate GPP by 20%–30%.

[24] These results complement and extend previous analyses from the NACP site synthesis. *Richardson et al.* [2012] found that overestimation of GPP in spring and fall resulted in models predicting the start of spring uptake about two weeks earlier than observed and the end of uptake in fall about two weeks later than observed. *Schwalm et al.* [2010] found models simulate NEE better at forest sites than grassland sites. The positive biases in spring and fall can help explain the decreased model performance relative to NEE in spring and fall [*Schwalm et al.*, 2010]. Underestimating GPP in summer can explain the peaks in the spectral signature of NEE residuals at the annual time scale [*Dietze et al.*, 2011]. Our results and those from prior studies indicate that seasonal biases in simulated GPP can help explain problems in simulating seasonal changes in NEE.

[25] Overall, there was a very large spread in model performance. On average, a single model showed good performance at 12% of the sites, marginal performance at 26%, and poor performance for the rest. Nearly every model had one "outlier" site where it performed considerably worse than the other sites, with X^2 values often exceeding 20. Conversely, nearly all models showed good or marginal performance at least one site. The spread in performance across sites was equally broad, with three outlier sites where none of the models performed well. The spread among models at a single site was also wide: each site, on average, had two good simulations, four marginal simulations, and two outlier simulations with $X^2 > 20$.

3.2. Model Structure

[26] Model performance did not depend on model structure, as defined by the model characteristics in Table 2. We did not find any statistically significant relationships between performance and how models represent various physical and biological processes (Table 3). In all cases, the difference in mean values between model groups was much smaller than the standard deviation within groups such that none of the differences were statistically significant. For example, LUE models performed better then EK models, but when excluding SSIB2 and TECO, which had large biases, EK models performed better than LUE models. Essentially, EK and LUE models performed equally well in simulating observed GPP. The performance difference for a daily versus hourly time step was consistent with the difference between EK and LUE models, since nearly all EK models use an hourly time step. The same was true for the other model structural characteristics: models that include a nitrogen cycle, a soil model, shaded leaves, or prognostic phenology performed equally well as models that do not.

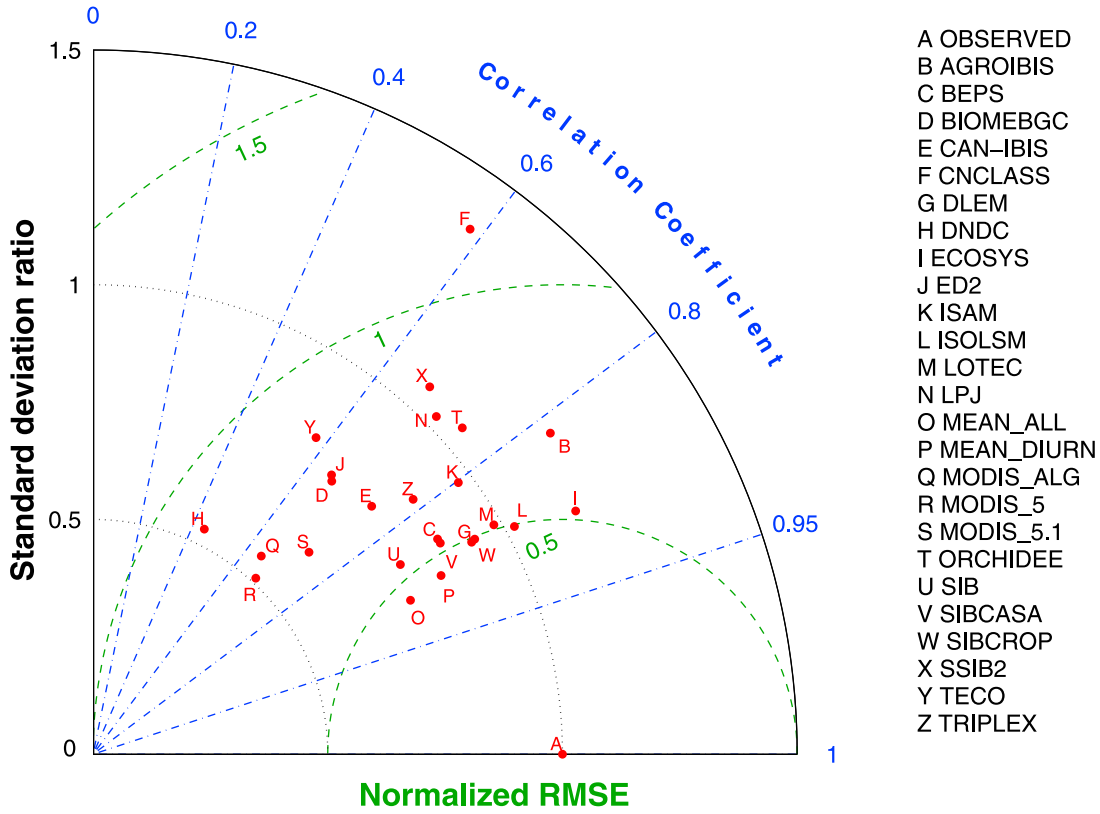

Figure 4. Taylor plot by model for all sites. A Taylor plot is a polar plot where the cosine of the angle from the *x* axis is the correlation coefficient between simulated and observed gross primary productivity (GPP). The correlation coefficient measures how well the simulated GPP captures the phasing and timing of the observed GPP. The radial direction is the ratio of simulated to observed standard deviation. The green lines represent RMSE normalized by standard deviation. An ideal model would have a standard deviation ratio of 1.0 and a correlation of 1.0 (point A).

[27] We found no significant relationships, but this does not mean that model structure does not influence performance. *Sprintsin et al.* [2012], for example, clearly demonstrate that accounting for diffuse light and changing from a big-leaf to a 2-leaf formulation improved BEPS performance. However, SiB is a big-leaf model and performed just as well as BEPS. The lack of significant relationships means that model performance is dominated by some other aspect of model design not represented by the model characteristics in Table 3, such as how models simulate GPP responses to changing environmental conditions.

3.3. Light Response

[28] The poor overall performance and the negative bias in summer resulted from mismatches between simulated and observed LUE. The light response curve is GPP as a function of downwelling shortwave radiation and its slope is the LUE. Figure 5 shows a light response curve based on daily average GPP for US-Me2 which we chose because it had a large number of simulations and was typical of all sites. The uncertainty in Figure 5 was dominated by gap-filling and partitioning uncertainty because the bin averaging tended to greatly reduce the random uncertainty. US-Me2 is an evergreen needleleaf site with simulations from 19 models with a marginal overall performance (X^2 = 1.9). Five models had good performance ($X^2 \leq 1.0$), two showed marginal performance ($1.0 < X^2 \leq 2.0$), and the rest showed poor performance. Four out of the five models with good performance all had LUEs that matched observed values within uncertainty. We saw no clear pattern in bias, with

Table 3. Differences in Performance Based on Model Structural Characteristics (Value ± Standard Deviation) for All 627 Simulations

Characteristic	Value	χ^2 (−)	RMSE (μmol m^{-2} s^{-1})	Bias (μmol m^{-2} s^{-1})
GPP Model Type	EK	4.2 ± 3.9	2.4 ± 1.2	0.1 ± 1.2
GPP Model Type	LUE	3.2 ± 2.3	2.8 ± 1.7	−0.9 ± 1.3
Leaf-to-Canopy	Big-Leaf	3.7 ± 3.2	2.6 ± 1.5	−0.3 ± 1.4
Leaf-to-Canopy	2-leaf	4.2 ± 3.8	2.4 ± 1.2	−0.1 ± 1.3
Phenology	Observed	3.1 ± 2.5	2.3 ± 1.3	−0.3 ± 1.2
Phenology	Prognostic	4.8 ± 4.1	2.7 ± 1.4	−0.1 ± 1.4
Soil Model	No	3.3 ± 2.6	2.6 ± 1.7	−0.6 ± 1.4
Soil Model	Yes	4.2 ± 3.8	2.4 ± 1.2	0.0 ± 1.3
Nitrogen Model	No	4.1 ± 3.6	2.5 ± 1.4	−0.1 ± 1.3
Nitrogen Model	Yes	3.7 ± 3.3	2.5 ± 1.4	−0.3 ± 1.3
Time Step	Daily	3.5 ± 3.0	2.7 ± 1.6	−0.7 ± 1.3
Time Step	Hourly	4.3 ± 3.8	2.4 ± 1.2	0.2 ± 1.2

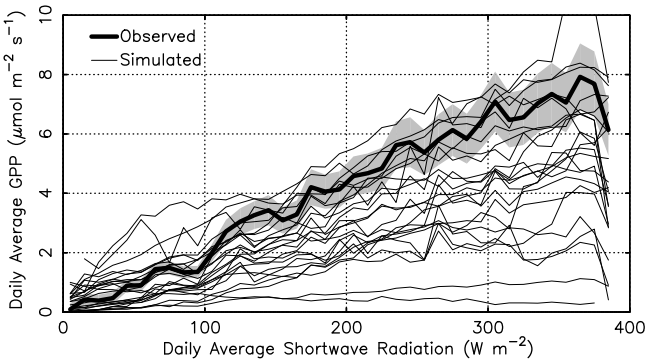

Figure 5. The light response curve for US-Me2 showing simulated and observed GPP as a function of downwelling shortwave radiation. The gray bar indicates uncertainty in the observed GPP response to shortwave radiation.

each model under-predicting GPP at some sites and over-predicting at others, but the spread in simulated GPP among models at US-Me2 was typical of all sites.

[29] Bias decreased as the ratio of simulated to observed LUE approached one (Figure 6). We calculated the observed and simulated LUE as the regression of GPP versus shortwave radiation flux with a Y-intercept forced to be zero. Both the observed and simulated light response curves were noisy, so we forced the Y-intercept to be zero to guarantee that GPP was zero for zero incident shortwave light. The LUE ratio was simulated LUE divided by observed LUE, with a ratio less than one indicating the model underestimated GPP (negative bias). The LUE ratios formed a diagonal line with near zero bias when the LUE ratio was one. Plots of LUE ratios for subgroups defined by the model structural characteristics in Table 2, individual models, and individual sites all showed the same pattern as in Figure 6: when the LUE ratio was one, the bias was zero.

[30] To improve performance in simulated GPP, model developers should focus first on those parameters that determine the simulated LUE. The LUE is determined by the leaf-to-canopy scaling and a small number of parameters that define the maximum potential GPP. For the MODIS algorithm described above, for example, the LUE is determined by $\varepsilon_{\max}$, so a better value of $\varepsilon_{\max}$ will improve performance. For other models, V_{cmax} (the unstressed Rubisco carboxylation rate), α (quantum yield), or $J_{\max}$ (the maximum electron transport rate) determine LUE. These maximum potential GPP parameters are either held constant for all sites, like in ECOSYS [*Grant et al.*, 2009] or vary with biome or plant functional type (PFT), like SiB3 [*Baker et al.*, 2008]. Models account for changing environmental conditions by multiplying the maximum potential GPP by temperature, moisture, and humidity scaling factors that represent the ratio of actual to peak GPP.

[31] How a model scales from a single leaf to an entire canopy also influences the simulated LUE. The maximum potential GPP parameters typically represent peak or optimal values for a single leaf at the canopy top. The leaf-to-canopy scaling factor represents the ratio of GPP for a single leaf to GPP for the entire canopy. A model assumes the distribution of leaf nitrogen and light levels within the canopy, and integrates from canopy top to bottom to calculate a leaf-to-canopy scaling factor. SiB and BEPS, for example, both assume the distribution of light is governed by Beer's law [*Sellers et al.*, 1996; *Sprintsin et al.*, 2012]. Unfortunately, the leaf-to-canopy scaling and the maximum potential GPP parameters are coupled and can compensate, indicating that the model has to get both right to get the correct GPP.

[32] Our results indicate that better LUE parameter values and leaf-to-canopy scaling will improve overall performance in simulated GPP, although we could not delve into individual models to identify the correct parameters and the best values. The number, nomenclature, definition, and units of the various parameters that define LUE differ widely among models. Model developers can use data assimilation of, for example, eddy covariance data to estimate parameter values [*Hanson et al.*, 2004]. The LOTEC model illustrates the potential to improve performance by estimating the maximum potential GPP parameters with data assimilation. However, due to differences among models, parameter values estimated for one model may not work in another. The TRY database of plant characteristics [*Kattge et al.*, 2011] includes observations of V_{cmax}, $J_{\max}$, and α compiled from many studies that could potentially minimize these inter-model differences. The leaf-to-canopy scaling depends on the assumed variation of light levels and parameter values within the canopy, which may require additional field observations. For example, measuring leaf-level nitrogen content, which determines V_{cmax}, is relatively easy, but what models need is canopy-level nitrogen content, which, unfortunately, is much more difficult to measure. Changes to how models treat the distribution of light within the canopy could improve the leaf-to-canopy scaling, such as better canopy radiative transfer models coupled to the GPP models or separating sunlit and shaded leaves, but developers have to demonstrate that such changes improve GPP performance. The observed LUE shows strong variability within PFT classes and the biases may result from the fact that the models assume constant parameter values for each PFT. Although spatially explicit maps of LUE parameter values currently do not exist, remote sensing of canopy nitrogen shows promise [*Ollinger et al.*, 2008]. However, nitrogen in Rubisco is the variable of interest and relating total nitrogen to a canopy V_{cmax} in a way that will work for all models will require new theoretical development.

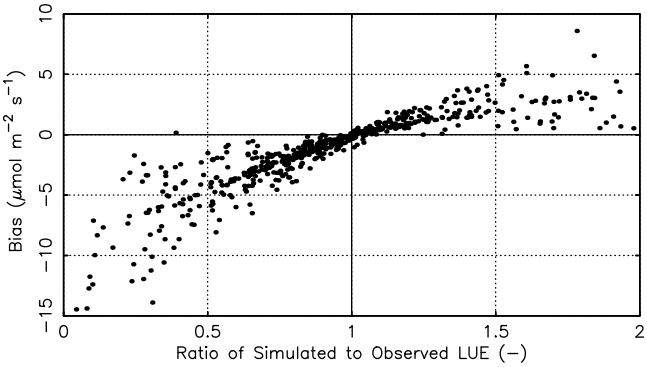

Figure 6. GPP bias as a function of the ratio of simulated to observed LUE for summer (June–July–August) for all 627 simulations.

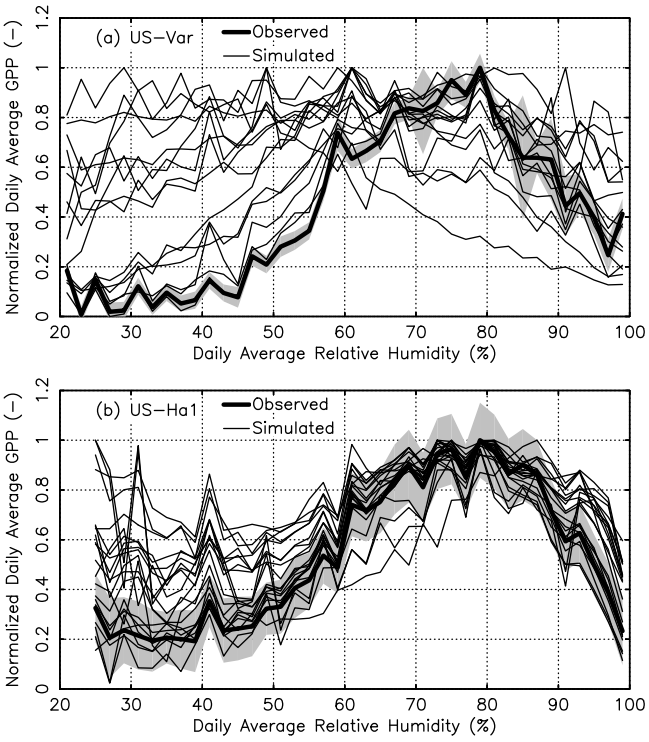

Figure 7. Normalized humidity response curves for (a) US-Var and (b) US-Ha1 showing simulated and observed GPP as a function of relative humidity. The gray bar indicates uncertainty in the observed GPP response. All response curves are normalized such that the maximum value is 1.

3.4. Humidity Response

[33] We found that difficulties in simulating GPP under dry conditions can explain why models performed worse at grassland and savanna sites than forest sites. Figure 7 shows normalized humidity response curves for the US-Var grassland and US-Ha1 deciduous broadleaf forest sites. We normalized the humidity response curves to emphasize the shape of the curves, which were typical of all sites: low GPP under dry conditions, an optimal GPP at 70%–80% relative humidity, and a decrease for higher humidity associated with colder temperatures. Models calculate lower GPP under stressed conditions using scaling factors that represent the ratio of stressed to optimal GPP. The scaling factors determine the shape of the response curves in Figure 7, but the simulated LUE determines the GPP magnitude. Decreased GPP at low relative humidity can be caused either by humidity stress reducing stomatal conductance, high temperature stress, or drier soils with reduced water availability (drought stress). Half of the models overpredicted GPP at both sites under low humidity conditions (relative humidity less than 60%). Such dry conditions occurred only 23% of the time at US-Ha1, but occurred 46% of the time at US-Var, which has dry summers with near zero growth. Even though half the models did not capture GPP under dry conditions at both sites, the effect on performance was much stronger at US-Var because the dry periods occurred twice as often than at US-Ha1. This explains the poor performance at the evergreen needleleaf forest sites CA-SJ1 and CA-SJ2, where the dry periods occurred nearly as often as US-Var. Essentially, the more often the dry periods occurred, the worse the performance.

[34] The NEE partitioning algorithm can partly explain the model-data mismatch at drier sites. The algorithm was not designed for drier sites where soil moisture has greater influence on R_{eco} than temperature. Consequently the algorithm either did not converge or produced negative GPP, which we changed to zero as described above. However, filtering out these zero GPP values did not change model performance at these sites. The partitioning algorithm could be improved to account for moisture, but the models also need to improve simulated GPP under dry conditions.

[35] Determining exactly how to improve simulated GPP under dry conditions was not possible in our analysis because the effects of drought and humidity stress are intertwined. Periods of low rainfall simultaneously reduce both soil moisture and atmospheric humidity, making the associated effects of drought and humidity stress on GPP difficult to separate. Models that account for drought stress typically calculate a GPP scaling factor using either input precipitation or plant water availability from simulated soil moisture [*Schaefer et al.*, 2008; *Potter et al.*, 1993]. Models that account for humidity stress either calculate a GPP scaling factor based on humidity or directly reduce stomatal conductance [*Heinsch et al.*, 2003; *Baker et al.*, 2008]. Thus, overpredicting GPP under dry conditions could result from problems with the simulated soil moisture, the calculation of plant water availability, or the representation of humidity stress.

[36] To complicate matters, a model's representation of humidity stress can compensate for poor representation of drought stress, and vice versa. For example, the MODIS algorithm above does not account for drought stress at all, but the humidity response was strengthened to compensate [*Heinsch et al.*, 2003], such that MODIS reproduced the shape (but not magnitude) of the observed humidity response curves. Determining whether models should improve simulated soil moisture, drought stress, or humidity stress requires a simultaneous analysis of simulated and observed soil moisture, latent heat flux, and GPP, which is beyond the scope of our analysis.

3.5. Temperature Response

[37] Figure 8 shows a typical temperature response function for the evergreen needleleaf forest site, US-Ho1, which had simulations from 23 models and a marginal overall performance (X^2 = 1.8). US-Ho1 was the site closest to the "average" temperature response function for all sites. The observed optimal temperature for GPP was 19°C and average across all sites was 20 ± 5°C. Differences between simulated and observed GPP near the peak or optimal temperature reflected differences in simulated and observed LUE in summer, as described above. For US-Ho1, GPP shut down for daily average temperatures below −6.5°C and the average low temperature cutoff across all sites was −6 ± 3°C. Cutoff temperatures below zero reflected conditions in spring and fall where daytime temperatures were above freezing to allow photosynthesis while the nighttime temperatures were below freezing, resulting in a negative daily average air temperature. The average winter season at US-Ho1, defined as the time with daily average temperature

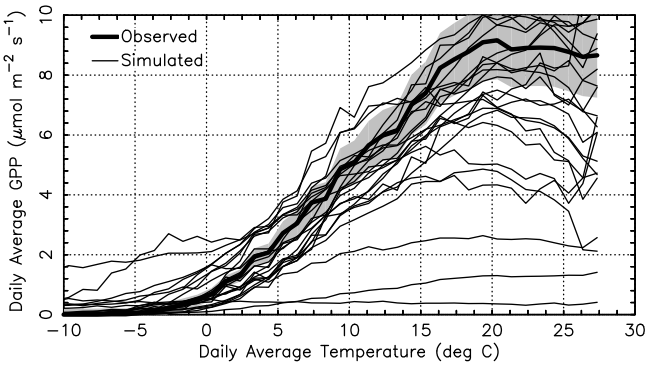

Figure 8. The temperature response curve for US-Ho1 showing simulated and observed GPP as a function of daily mean air temperature. US-Ho1 is the site closest to the average response for all sites. The gray bar indicates uncertainty in the observed GPP response to air temperature.

below 0°C, was 92 days and the average for all sites was 75 ± 47 days.

[38] Temperature was the dominant control of the seasonal variation of GPP at most sites. The simulated start of GPP in spring and the stop of GPP in fall is a representation of phenology. Models primarily use temperature to control phenology, but how this is done varies widely. Some models using growing degree days to predict bud burst in spring while others simply shut down GPP below a lower temperature limit. How models simulate GPP at temperatures near 0°C determined the start and stop of the simulated growing season.

[39] The models tended to overpredict GPP at low temperatures: half of all simulations predicted more than double the observed GPP at temperatures below 0°C. The observations indicated that no more than 0.8 ± 0.6% of total annual GPP occurred in winter, but the models simulated anywhere from 0% to 15% of the annual GPP in winter. Part of this may have resulted from the partitioning algorithm itself, which set GPP to zero when the soil and the air temperature are below zero. However, observations indicate this is realistic since the recovery of photosynthesis after freezing temperatures can be delayed for weeks with repeated exposure to frost and cold and frozen soils limit root uptake of water and stomatal conductance [*Strand and Öquist*, 1985; *Waring and Winner*, 1996]. Overpredicting GPP under cold conditions explained the positive biases in winter, spring, and fall, which in turn resulted in uptake starting earlier than observed in spring and later than observed in fall.

[40] Better low temperature inhibition functions will improve simulated GPP in winter, spring, and fall and improve simulated phenology. Most models use an exponential or "Q_{10}" response function to represent the effects of low temperature on GPP:

$$S_T = (Q_{10})^{\frac{T-T_{ref}}{10}}, \quad (4)$$

where S_T is a temperature scaling factor applied to GPP, T is temperature (°C), and T_{ref} is a reference temperature (°C). All models based on the *Farquhar et al.* [1980] EK model, for example, use this type of formulation. Most models have a second exponential function with separate Q_{10} and T_{ref} values to reflect reduced GPP for high temperatures. Combined, the low and high temperature functions produce an optimal temperature for simulated GPP. The combined temperature scaling factor represents the ratio of actual to optimal GPP and varies between zero and one. The simulated LUE determined GPP magnitude under optimal conditions in midsummer, but the temperature scaling determined the seasonal cycle in simulated GPP and simulated phenology. The exact values of Q_{10} and T_{ref} vary widely between models and we made no attempt to determine which values are correct.

[41] The positive bias in winter, spring, and fall resulted from the fact that the Q_{10} function alone will never reach zero no matter how cold the temperature, so those models using this type of formulation without a frost inhibition function over predicted GPP at low temperatures. The frost inhibition function is an additional scaling factor that shuts down GPP below a specified threshold temperature [*Kucharik and Twine*, 2007; *Li et al.*, 2010]. In addition, some models also include a GPP recovery period after the frost event [*Baker et al.*, 2008; *Schaefer et al.*, 2008]. Data on photosynthesis at low temperatures are relatively scarce, so developing a low-temperature inhibition function to incorporate the effects of nutrient and water availability in partially frozen soils may require more observations. Improving the modeled low temperature inhibition function will improve simulated GPP in spring and fall, and thus simulated phenology.

4. Conclusions

[42] None of the models in this study match estimated GPP within the range of uncertainty of observed fluxes. On average, the models achieved good performance for only 12% of the simulations. Two models achieved overall marginal performance, matching estimated GPP within roughly two times the uncertainty. Our first hypothesis proved false: we found no statistically significant differences in performance due to model structure, mainly due to the large spread in performance among models and across sites. The models in our study reproduced the observed seasonal pattern with little or no GPP in winter and peak GPP in summer, but did not capture the observed GPP magnitude. We found, on average, that models overestimated GPP in spring and fall and underestimated GPP in summer. Our second hypothesis proved true: model performance depended on how models represented the GPP response to changing environmental conditions. We identified three areas of model improvement: simulated LUE, low temperature response function, and GPP response under dry conditions.

[43] The poor overall model performance resulted primarily from inadequate representation of observed LUE. Simulated LUE is controlled by the leaf-to-canopy scaling strategy and a small set of model parameters that define the maximum potential GPP, such as ε_{max} (light use efficiency), V_{cmax} (unstressed Rubisco catalytic capacity) or J_{max} (the maximum electron transport rate). The temperature, humidity, and drought scaling factors determined temporal variability in simulated GPP, but the LUE parameters determined the magnitude of simulated GPP. To improve simulated GPP, model developers should focus first on improving the leaf-to-canopy scaling and the values of those model parameters that control the LUE.

[44] Many models overpredicted GPP under dry conditions, explaining why, on average, models performed worse at grassland and savanna sites than at forest sites. The importance of this to model performance increases at sites where drier conditions occur more frequently. Since dry conditions occur more frequently at grassland and savanna sites than at forest sites, models tended to perform worse at grassland and savanna sites compared to forest sites. Improving how models simulate soil moisture, drought stress, or humidity stress can improve simulated GPP under dry conditions.

[45] Many models overpredicted GPP under cold conditions, partly explaining the positive bias in simulated GPP in winter, spring, and fall. The estimated GPP completely shut down for daily average temperatures less than −6°C, but the Q_{10} formulation used by many models did not shut down GPP under cold or frozen conditions. The simulated GPP started too early in spring and persisted too late in fall, resulting in a positive bias and phasing errors in phenology. Using an ensemble mean can cancel out errors in phenology, but does not cancel out bias. Improving or imposing a low temperature inhibition function in the GPP model will resolve the problem.

[46] **Acknowledgments.** We thank the North American Carbon Program Site-Level Interim Synthesis team and the Oak Ridge National Laboratory Distributed Active Archive Center for collecting, organizing, and distributing the model output and flux observations required for this analysis. We thank Dennis Baldocchi, Lawrence Flanagan, Ni Golaz, Gabrial Katul, Kim Novick, Paul Stoy, and Shashi B. Verma for providing valuable data and advice during the development of this paper. This research was partly funded by NOAA Award NA07OAR4310115 and U.S. National Science Foundation grant ATM-0910766; funding was also provided by the U.S. Department of Energy's Office of Science for AmeriFlux Science Team research to develop measurement and data submission protocols and conduct quality assurance of measurements for AmeriFlux investigators (Grant DE-FG02-04ER63911).

References

Arain, M. A., and N. Restrepo (2005), Net ecosystem production in an eastern white pine plantation in southern Canada, *Agric. For. Meteorol.*, *128*, 223–241, doi:10.1016/j.agrformet.2004.10.003.

Arain, M. A., F. Yaun, and T. A. Black (2006), Soil-plant nitrogen cycling modulated carbon exchanges in a western temperate conifer forest in Canada, *Agric. For. Meteorol.*, *140*, 171–192, doi:10.1016/j.agrformet.2006.03.021.

As-syakur, A. R., T. Osawa, and I. W. S. Adnyana (2010), Medium spatial resolution satellite imagery to estimate gross primary production in an urban area, *Remote Sens.*, *2*, 1496–1507, doi:10.3390/rs2061496.

Baker, I. T., L. Prihodko, A. S. Denning, M. Goulden, S. Milller, and H. da Rocha (2008), Seasonal drought stress in the Amazon: Reconciling models and observations, *J. Geophys. Res.*, *113*, G00B01, doi:10.1029/2007JG000644.

Baldocchi, D., et al. (2001), FLUXNET: A new tool to study the temporal and spatial variability of ecosystem–scale carbon dioxide, water vapor, and energy flux densities, *Bull. Am. Meteorol. Soc.*, *82*(11), 2415–2434, doi:10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2.

Ball, J. T., et al. (1987), A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions, in *Progress in Photosynthesis Research: Proceedings of the VIIth International Congress on Photosynthesis*, vol. 4, edited by J. Biggins, pp. 221–224, Martinus Nijhoff, Dordrecht, Netherlands.

Barr, A. G., T. A. Black, E. H. Hogg, N. Kljun, K. Morgenstern, and Z. Nesic (2004), Inter-annual variability in the leaf area index of a boreal aspen-hazelnut forest in relation to net ecosystem production, *Agric. For. Meteorol.*, *126*, 237–255, doi:10.1016/j.agrformet.2004.06.011.

Barr, A. G., et al. (2006), Climatic controls on the carbon and water balances of a boreal aspen forest, 1994–2003, *Global Change Biol.*, *12*, 1–16.

Beer, C., et al. (2010), Terrestrial gross carbon dioxide uptake: Global distribution and covariation with climate, *Science*, *329*, 834–838, doi:10.1126/science.1184984.

Bergeron, O., H. A. Margolis, T. A. Black, C. Coursolle, A. L. Dunn, A. G. Barr, and S. C. Wofsy (2007), Comparison of CO_2 fluxes over three boreal black spruce forests in Canada, *Global Change Biol.*, *13*, 89–107, doi:10.1111/j.1365-2486.2006.01281.x.

Bergeron, O., H. A. Margolis, C. Coursolle, and M.-A. Giasson (2008), How does forest harvest influence carbon dioxide fluxes of black spruce ecosystems in eastern North America?, *Agric. For. Meteorol.*, *148*, 537–548, doi:10.1016/j.agrformet.2007.10.012.

Caemmerer, S., and G. D. Farquhar (1981), Some relationships between the biochemistry of photosynthesis and the gas-exchange of leaves, *Planta*, *153*(4), 376–387, doi:10.1007/BF00384257.

Chapin, F. S., et al. (2006), Reconciling carbon-cycle concepts, terminology, and methods, *Ecosystems (N. Y.)*, *9*, 1041–1050, doi:10.1007/s10021-005-0105-7.

Collatz, G. J., et al. (1991), Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: A model that includes a laminar boundary layer, *Agric. For. Meteorol.*, *54*, 107–136, doi:10.1016/0168-1923(91)90002-8.

Collatz, G. J., et al. (1992), Coupled photosynthesis-stomatal conductance model for leaves of C4 plants, *Aust. J. Bot.*, *19*, 519–538.

Cook, B. D., et al. (2004), Carbon exchange and venting anomalies in an upland deciduous forest in northern Wisconsin, USA, *Agric. For. Meteorol.*, *126*, 271–295, doi:10.1016/j.agrformet.2004.06.008.

Davis, K. J., P. S. Bakwin, C. X. Yi, B. W. Berger, C. L. Zhao, R. M. Teclaw, and J. G. Isebrands (2003), The annual cycles of CO_2 and H_2O exchange over a northern mixed forest as observed from a very tall tower, *Global Change Biol.*, *9*(9), 1278–1293, doi:10.1046/j.1365-2486.2003.00672.x.

Desai, A. R., P. V. Bolstad, B. D. Cook, K. J. Davis, and E. V. Carey (2005), Comparing net ecosystem exchange of carbon dioxide between an old-growth and mature forest in the upper Midwest, USA, *Agric. For. Meteorol.*, *128*, 33–55, doi:10.1016/j.agrformet.2004.09.005.

Desai, A. R., et al. (2008), Cross-site evaluation of eddy covariance GPP and RE decomposition techniques, *Agric. For. Meteorol.*, *148*, 821–838, doi:10.1016/j.agrformet.2007.11.012.

Dietze, M. C., et al. (2011), Characterizing the performance of ecosystem models across time scales: A spectral analysis of the North American Carbon Program site-level synthesis, *J. Geophys. Res.*, *116*, G04029, doi:10.1029/2011JG001661.

Distributed Active Archive Center for Biogeochemical Dynamics (2010), Modis Land Product Subsets MOD17A2 and MYD17A2, Oak Ridge Natl. Lab., Oak Ridge, Tenn. [Available at http://www.modis.ornl.gov/modis/index.cfm.]

Farquhar, G. D., S. Caemmerer, and J. A. Berry (1980), A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species, *Planta*, *149*(1), 78–90, doi:10.1007/BF00386231.

Field, C. B. (1995), Global net primary production: Combining ecology and remote sensing, *Remote Sens. Environ.*, *51*, 74–88, doi:10.1016/0034-4257(94)00066-V.

Fischer, M. L., D. P. Billesbach, W. J. Riley, J. A. Berry, and M. S. Torn (2007), Spatiotemporal variations in growing season exchanges of CO_2, H_2O, and sensible heat in agricultural fields of the Southern Great Plains, *Earth Interact.*, *11*, 1–21, doi:10.1175/EI231.1.

Flanagan, L. B., and A. C. Adkinson (2011), Interacting controls on productivity in a northern Great Plains grassland and implications for response to ENSO events, *Global Change Biol.*, *17*, 3293–3311, doi:10.1111/j.1365-2486.2011.02461.x.

Flanagan, L. B., and K. H. Syed (2011), Stimulation of both photosynthesis and respiration in response to warmer and drier conditions in a boreal peatland ecosystem, *Global Change Biol.*, *17*, 2271–2287, doi:10.1111/j.1365-2486.2010.02378.x.

Foken, T. (2008), The energy balance closure problem: An overview, *Ecol. Appl.*, *18*(6), 1351–1367, doi:10.1890/06-0922.1.

Goetz, S. J., et al. (1999), Satellite remote sensing of primary production: An improved production efficiency modeling approach, *Ecol. Modell.*, *122*, 239–255, doi:10.1016/S0304-3800(99)00140-4.

Gough, C. M., C. S. Vogel, H. P. Schmid, H.-B. Su, and P. S. Curtis (2008), Multi-year convergence of biometric and meteorological estimates of forest carbon storage, *Agric. For. Meteorol.*, *148*, 158–170, doi:10.1016/j.agrformet.2007.08.004.

Grant, R. F., A. G. Barr, T. A. Black, H. A. Margolis, A. L. Dunn, J. Metsaranta, S. Wang, J. H. McCaughey, and C. P.-A. Bourque (2009), Interannual variation in net ecosystem productivity of Canadian forests as affected by regional weather patterns – A Fluxnet-Canada synthesis, *Agric. For. Meteorol.*, *149*, 2022–2039, doi:10.1016/j.agrformet.2009.07.010.

Grant, R. F., A. G. Barr, T. A. Black, H. A. Margolis, J. H. McCaughey, and J. A. Trofymow (2010), Net ecosystem productivity of temperate and boreal forests after clearcutting – A Fluxnet-Canada measurement and modelling synthesis, *Tellus, Part B*, *62*, 475–496.

Gu, L. H., T. Meyers, S. G. Pallardy, P. J. Hanson, B. Yang, M. Heuer, K. P. Hosman, J. S. Riggs, D. Sluss, and S. D. Wullschleger (2006), Direct and indirect effects of atmospheric conditions and soil moisture on surface energy partitioning revealed by a prolonged drought at a temperate forest site, *J. Geophys. Res.*, *111*(D16), D16102, doi:10.1029/2006JD007161.

Gurney, K. R., et al. (2002), Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models, *Nature*, *415*(6872), 626–630, doi:10.1038/415626a.

Hanson, P. J., et al. (2004), Oak forest carbon and water simulations: Model intercomparisons and evaluations against independent data, *Ecol. Monogr.*, *74*, 443–489, doi:10.1890/03-4049.

Heinsch, F. A., et al. (2003), User's Guide: GPP and NPP (MOD17A2/A3) Products, NASA MODIS Land Algorithm, Version 2.0, Univ. of Montana, Missoula, 57 pp.

Humphreys, E. R., T. A. Black, K. Morgenstern, T. Cai, G. B. Drewitt, and Z. Nesic (2006), Carbon dioxide fluxes in three coastal Douglas-fir stands at different stages of development after harvesting, *Agric. For. Meteorol.*, *140*, 6–22, doi:10.1016/j.agrformet.2006.03.018.

Kattge, J., et al. (2011), TRY: A global database of plant traits, *Global Change Biol.*, *17*(9), 2905–2935, doi:10.1111/j.1365-2486.2011.02451.x.

Kljun, N., T. A. Black, T. J. Griffis, A. G. Barr, D. Gaumont-Guay, K. Morgenstern, J. H. McCaughey, and Z. Nesic (2006), Response of net ecosystem productivity of three boreal forest stands to drought, *Ecosystems (N. Y.)*, *9*, 1128–1144, doi:10.1007/s10021-005-0082-x.

Krinner, G., N. Viovy, N. De Noblet-Ducoudre, J. Ogée, J. Polcher, P. Friedlingstein, P. Ciais, S. Sitch, and I. C. Prentice (2005), A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, *Global Biogeochem. Cycles*, *19*, GB1015, doi:10.1029/2003GB002199.

Krishnan, P., T. A. Black, N. J. Grant, A. G. Barr, E. H. Hogg, R. S. Jassal, and A. N. D. K. Morgenstern (2006), Impact of changing soil moisture distribution on net ecosystem productivity of a boreal aspen forest during and following drought, *Agric. For. Meteorol.*, *139*, 208–223, doi:10.1016/j.agrformet.2006.07.002.

Krishnan, P., T. A. Black, A. G. Barr, D. Gaumont-Guay, N. Grant, and Z. Nesic (2008), Factors controlling the interannual variability in the carbon balance of a southern boreal black spruce forest, *J. Geophys. Res.*, *113*, D09109, doi:10.1029/2007JD008965.

Krishnan, P., T. A. Black, R. S. Jassal, B. Chen, and Z. Nesic (2009), Interannual variability of the carbon balance of three different-aged Douglas-fir stands in the Pacific Northwest, *J. Geophys. Res.*, *114*, G04011, doi:10.1029/2008JG000912.

Kucharik, C. J., and T. E. Twine (2007), Residue, respiration, and residuals: Evaluation of a dynamic agroecosystem model using eddy flux measurements and biometric data, *Agric. For. Meteorol.*, *146*, 134–158, doi:10.1016/j.agrformet.2007.05.011.

Landsberg, J., and R. H. Waring (1997), A generalized model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance, and partitioning, *For. Ecol. Manage.*, *95*, 209–228, doi:10.1016/S0378-1127(97)00026-1.

Lasslop, G., M. Reichstein, D. Papale, A. D. Richardson, A. Arneth, A. Barr, P. Stoy, and G. Wohlfahrt (2010), Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: Critical issues and global evaluation, *Global Change Biol.*, *16*, 187–208, doi:10.1111/j.1365-2486.2009.02041.x.

Law, B. E., P. Thornton, J. Irvine, P. Anthoni, and S. Van Tuyl (2001), Carbon storage and fluxes in ponderosa pine forests at different developmental stages, *Global Change Biol.*, *7*, 755–777, doi:10.1046/j.1354-1013.2001.00439.x.

Li, H., J. Qiu, L. Wang, H. Tang, C. Li, and E. Van Ranst (2010), Modeling impacts of alternative farming management practices on greenhouse gas emissions from a winter wheat-maize rotation system in China, *Agric. Ecosyst. Environ.*, *135*, 24–33, doi:10.1016/j.agee.2009.08.003.

Liu, J., J. M. Chen, J. Cihlar, and W. Chen (1999), Net primary productivity distribution in the BOREAS region from a process model using satellite and surface data, *J. Geophys. Res.*, *104*(D22), 27,735–27,754, doi:10.1029/1999JD900768.

Liu, J., D. T. Price, and J. M. Chen (2005), Nitrogen controls on ecosystem carbon sequestration: A model implementation and application to Saskatchewan, Canada, *Ecol. Modell.*, *186*(2), 178–195, doi:10.1016/j.ecolmodel.2005.01.036.

Liu, S., N. Bliss, E. Sundquist, and T. G. Huntington (2003), Modeling carbon dynamics in vegetation and soil under the impact of soil erosion and deposition, *Global Biogeochem. Cycles*, *17*(2), 1074, doi:10.1029/2002GB002010.

Lokupitiya, E., A. S. Denning, K. Paustian, I. Baker, K. Schaefer, S. Verma, T. Meyers, C. Bernacchi, A. Suyker, and M. Fischer (2009), Incorporation of crop phenology in Simple Biosphere Model (SiBcrop) to improve land-atmosphere carbon exchanges from croplands, *Biogeosciences*, *6*(6), 969–986, doi:10.5194/bg-6-969-2009.

Loveland, T. R., B. C. Reed, J. F. Brown, D. O. Ohlen, J. Zhu, L. Yang, and J. W. Merchant (2000), Development of a global land cover characteristics database and IGBP DISCover from 1-km AVHRR data, *Int. J. Remote Sens.*, *21*, 1303–1330, doi:10.1080/014311600210191.

Luo, H., W. C. Oechel, S. J. Hastings, R. Zulueta, Y. Qian, and H. Kwon (2007), Mature semiarid chaparral ecosystems can be a significant sink, *Global Change Biol.*, *13*, 386–396, doi:10.1111/j.1365-2486.2006.01299.x.

Ma, S., D. D. Baldocchi, L. Xu, and T. Hehn (2007), Inter-annual variability in carbon dioxide exchange of an oak/grass savanna and open grassland in California, *Agric. For. Meteorol.*, *147*, 157–171, doi:10.1016/j.agrformet.2007.07.008.

McCaughey, H., M. R. Pejam, M. A. Arain, and D. A. Cameron (2006), Carbon dioxide and energy fluxes from a boreal mixedwood forest ecosystem in Ontario, Canada, *Agric. For. Meteorol.*, *140*, 79–96, doi:10.1016/j.agrformet.2006.08.010.

Medvigy, D., S. C. Wofsy, J. W. Munger, D. Y. Hollinger, and P. R. Moorcroft (2009), Mechanistic scaling of ecosystem function and dynamics in space and time: Ecosystem demography model version 2, *J. Geophys. Res.*, *114*, G01002, doi:10.1029/2008JG000812.

Moffat, A. M., et al. (2007), Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes, *Agric. For. Meteorol.*, *147*, 209–232, doi:10.1016/j.agrformet.2007.08.011.

Monson, R. K., A. A. Turnipseed, J. P. Sparks, P. C. Harley, L. E. Scott-Denton, K. L. Sparks, and T. E. Huxman (2002), Carbon sequestration in a high-elevation, subalpine forest, *Global Change Biol.*, *8*, 459–478, doi:10.1046/j.1365-2486.2002.00480.x.

Monson, R. K., J. P. Sparks, T. N. Rosenstiel, L. E. Scott-Denton, T. E. Huxman, P. C. Harley, A. A. Turnipseed, S. P. Burns, B. Backlund, and J. Hu (2005), Climatic influences on net ecosystem CO_2 exchange during the transition from wintertime carbon source to springtime carbon sink in a high-elevation, subalpine forest, *Oecologia*, *146*, 130–147, doi:10.1007/s00442-005-0169-2.

Monteith, J. L. (1972), Solar radiation and production in tropical ecosystems, *J. Appl. Ecol.*, *9*, 747–766, doi:10.2307/2401901.

Ollinger, S. V., et al. (2008), Canopy nitrogen, carbon assimilation, and albedo in temperate and boreal forests: Functional relations and potential climate feedbacks, *Proc. Natl. Acad. Sci. U. S. A.*, *105*(49), 19,336–19,341, doi:10.1073/pnas.0810021105.

Oren, R., C. I. Hsieh, P. C. Stoy, J. D. Albertson, H. R. McCarthy, P. Harrell, and G. G. Katul (2006), Estimating the uncertainty in annual net ecosystem carbon exchange: Spatial variation in turbulent fluxes and sampling errors in eddy-covariance measurements, *Global Change Biol.*, *12*, 883–896, doi:10.1111/j.1365-2486.2006.01131.x.

Papale, D., et al. (2006), Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: Algorithms and uncertainty estimation, *Biogeosciences*, *3*(4), 571–583, doi:10.5194/bg-3-571-2006.

Peichl, M., and M. A. Arain (2006), Above- and belowground ecosystem biomass and carbon pools in an age-sequence of temperate pine plantation forests, *Agric. For Meteorol.*, *140*(1–4), 51–63, doi:10.1016/j.agrformet.2006.08.004.

Peng, C. H., J. Liu, Q. Dang, M. J. Apps, and H. Jiang (2002), TRIPLEX: A generic hybrid model for predicting forest growth and carbon and nitrogen dynamics, *Ecol. Modell.*, *153*, 109–130, doi:10.1016/S0304-3800(01)00505-1.

Peters, W., et al. (2010), Seven years of recent European net terrestrial carbon dioxide exchange constrained by atmospheric observations, *Global Change Biol.*, *16*(4), 1317–1337, doi:10.1111/j.1365-2486.2009.02078.x.

Post, W. M., et al. (2004), Enhancement of carbon sequestration in U. S. soils, *BioScience*, *54*(10), 895–908, doi:10.1641/0006-3568(2004)054[0895:EOCSIU]2.0.CO;2.

Potter, C. S., J. T. Randerson, C. B. Field, P. A. Matson, P. M. Vitousek, H. A. Mooney, and S. A. Klooster (1993), Terrestrial ecosystem production: A process-oriented model based on global satellite and surface data, *Global Biogeochem. Cycles*, *7*, 811–842, doi:10.1029/93GB02725.

Prince, S. D., and S. N. Goward (1995), Global primary production: A remote sensing approach, *J. Biogeogr.*, *22*, 815–835, doi:10.2307/2845983.

Ricciuto, D. M., P. E. Thornton, K. Schaefer, R. B. Cook, and K. J. Davis (2009), How uncertainty in gap-filled meteorological input forcing at eddy covariance sites impacts modeled carbon and energy flux, *Eos Trans. AGU*, *90*(52), Fall Meet. Suppl., Abstract B54A-03.

Ricciuto, D. M., A. W. King, D. Dragoni, and W. M. Post (2011), Parameter and prediction uncertainty in an optimized terrestrial carbon cycle model: Effects of constraining variables and data record length, *J. Geophys. Res.*, *116*, G01033, doi:10.1029/2010JG001400.

Richardson, A. D., and D. Y. Hollinger (2007), A method to estimate the additional uncertainty in gap-filled NEE resulting from long gaps in the CO_2 flux record, *Agric. For. Meteorol.*, *147*(3–4), 199–208, doi:10.1016/j.agrformet.2007.06.004.

Richardson, A. D., D. Y. Hollinger, D. B. Dail, J. T. Lee, J. W. Munger, and J. O'Keefe (2009), Influence of spring phenology on seasonal and annual carbon balance in two contrasting New England forests, *Global Change Biol.*, *29*, 321–331.

Richardson, A. D., et al. (2012), Terrestrial biosphere models need better representation of vegetation phenology: Results from the North American Carbon Program site synthesis, *Global Change Biol.*, *18*, 566–584, doi:10.1111/j.1365-2486.2011.02562.x.

Riley, W., C. Still, M. Torn, and J. Berry (2002), A mechanistic model of $H_2^{18}O$ and $C^{18}OO$ fluxes between ecosystems and the atmosphere: Model description and sensitivity analyses, *Global Biogeochem. Cycles*, *16*(4), 1095, doi:10.1029/2002GB001878.

Roulet, N. T., P. M. Lafleur, P. J. H. Richard, T. R. Moore, E. R. Humphreys, and J. Bubier (2007), Contemporary carbon balance and late Holocene carbon accumulation in a northern peatland, *Global Change Biol.*, *13*, 397–411, doi:10.1111/j.1365-2486.2006.01292.x.

Running, S. W., P. E. Thornton, R. Nemani, and J. M. Glassy (2000), Global terrestrial gross and net primary productivity from the Earth observing system, in *Methods in Ecosystem Science*, edited by O. E. Sala et al., pp. 44–57, Springer, New York, doi:10.1007/978-1-4612-1224-9_4.

Running, S. W., R. R. Nemani, F. A. Heinsch, M. Zhao, M. Reeves, and H. Hashimoto (2004), A continuous satellite-derived measure of global terrestrial primary production, *BioScience*, *54*(6), 547–560, doi:10.1641/0006-3568(2004)054[0547:ACSMOG]2.0.CO;2.

Schaefer, K., G. J. Collatz, P. Tans, A. S. Denning, I. Baker, J. Berry, L. Prihodko, N. Suits, and A. Philpott (2008), Combined Simple Biosphere/Carnegie-Ames-Stanford Approach terrestrial carbon cycle model, *J. Geophys. Res.*, *113*, G03034, doi:10.1029/2007JG000603.

Schaefer, K., T. Zhang, A. G. Slater, L. Lu, A. Etringer, and I. Baker (2009), Improving simulated soil temperatures and soil freeze/thaw at high-latitude regions in the Simple Biosphere/Carnegie-Ames-Stanford Approach model, *J. Geophys. Res.*, *114*, F02021, doi:10.1029/2008JF001125.

Schmid, H. P., C. S. B. Grimmond, F. Cropley, B. Offerle, and H. B. Su (2000), Measurements of CO_2 and energy fluxes over a mixed hardwood forest in the mid-western United States, *Agric. For. Meteorol.*, *103*, 357–374, doi:10.1016/S0168-1923(00)00140-4.

Schwalm, C. R., et al. (2010), A model-data intercomparison of CO_2 exchange across North America: Results from the North American Carbon Program site synthesis, *J. Geophys. Res.*, *115*, G00H05, doi:10.1029/2009JG001229.

Sellers, P. J., S. O. Los, C. J. Tucker, C. O. Justice, D. A. Dazlich, G. J. Collatz, and D. A. Randall (1996), A revised land surface parameterization of GCMs, Part II: The generation of global fields of terrestrial biophysical parameters from satellite data, *J. Clim.*, *9*(4), 706–737, doi:10.1175/1520-0442(1996)009<0706:ARLSPF>2.0.CO;2.

Sims, D. A., H. Luo, S. Hastings, W. C. Oechel, A. F. Rahman, and J. A. Gamon (2006), Parallel adjustments in vegetation greenness and ecosystem CO_2 exchange in response to drought in a southern California chaparral ecosystem, *Remote Sens. Environ.*, *103*, 289–303, doi:10.1016/j.rse.2005.01.020.

Sitch, S., et al. (2003), Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model, *Global Change Biol.*, *9*, 161–185, doi:10.1046/j.1365-2486.2003.00569.x.

Sjöström, M., J. Ardö, L. Eklundh, B. A. El-Tahir, H. A. M. El-Khidir, M. Hellström, P. Pilesjö, and J. Seaquist (2009), Evaluation of satellite-based indices for gross primary production estimates in a sparse savanna in the Sudan, *Biogeosciences*, *6*, 129–138, doi:10.5194/bg-6-129-2009.

Sprintsin, M., J. M. Chen, A. Desai, and C. M. Goug (2012), Evaluation of leaf-to-canopy upscaling methodologies against carbon flux data in North America, *J. Geophys. Res.*, *117*, G01023, doi:10.1029/2010JG001407.

Stoy, P. C., G. G. Katul, M. B. S. Siqueira, J.-Y. Juang, K. A. Novick, J. M. Uebelherr, and R. Oren (2006), An evaluation of methods for partitioning eddy covariance-measured net ecosystem exchange into photosynthesis and respiration, *Agric. For. Meteorol.*, *141*, 2–18, doi:10.1016/j.agrformet.2006.09.001.

Strand, M., and G. Öquist (1985), Inhibition of photosynthesis by freezing temperatures and high light levels in cold-acclimated seedlings of Scots pine (*Pinus sylvestris*): I. Effects on the light-limited and light-saturated rates of CO_2 assimilation, *Physiol. Plant.*, *64*, 425–430, doi:10.1111/j.1399-3054.1985.tb08517.x.

Sulman, B. N., A. R. Desai, B. D. Cook, N. Saliendra, and D. S. Mackay (2009), Contrasting carbon dioxide fluxes between a drying shrub wetland in Northern Wisconsin, USA, and nearby forests, *Biogeosciences*, *6*, 1115–1126, doi:10.5194/bg-6-1115-2009.

Suyker, A. E., and S. B. Verma (2010), Coupling of carbon dioxide and water vapor exchanges of irrigated and rainfed maize-soybean cropping systems and water productivity, *Agric. For. Meteorol.*, *150*, 553–563, doi:10.1016/j.agrformet.2010.01.020.

Suyker, A. E., S. B. Verma, and G. G. Burba (2003), Interannual variability in net CO_2 exchange of a native tallgrass prairie, *Global Change Biol.*, *9*, 255–265, doi:10.1046/j.1365-2486.2003.00567.x.

Thomas, C. K., B. E. Law, J. Irvine, J. G. Martin, J. C. Pettijohn, and K. J. Davis (2009), Seasonal hydrology explains interannual and seasonal variation in carbon and water exchange in a semi-arid mature ponderosa pine forest in Central Oregon, *J. Geophys. Res.*, *114*, G04006, doi:10.1029/2009JG001010.

Thornton, P. E., S. W. Running, and M. A. White (1997), Generating surfaces of daily meteorological variables over large regions of complex terrain, *J. Hydrol.*, *190*(3–4), 214–251, doi:10.1016/S0022-1694(96)03128-9.

Thornton, P. E., et al. (2002), Modeling and measuring the effects of disturbance history and climate on carbon and water budgets in evergreen needleleaf forests, *Agric. For. Meteorol.*, *113*, 185–222, doi:10.1016/S0168-1923(02)00108-9.

Thornton, P. E., S. W. Running, and E. R. Hunt (2005), Biome-BGC: Terrestrial Ecosystem Process Model, Version 4.1.1, Oak Ridge Natl. Lab., Oak Ridge, Tenn., doi:10.3334/ORNLDAAC/805. [Available at http://daac.ornl.gov.]

Tian, H., G. Chen, M. Liu, C. Zhang, G. Sun, C. Lu, X. Xu, W. Ren, S. Pan, and A. Chappelka (2010), Model estimates of net primary productivity, evapotranspiration and water use efficiency in the terrestrial ecosystem of the southern United States during 1895–2007, *For. Ecol. Manage.*, *259*, 1311–1327, doi:10.1016/j.foreco.2009.10.009.

Tucker, C. J., J. E. Pinzon, M. E. Brown, D. A. Slayback, E. W. Pak, R. Mahoney, E. F. Vermote, and N. El Saleous (2005), An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data, *Int. J. Remote Sens.*, *26*, 4485–4498, doi:10.1080/01431160500168686.

Urbanski, S., C. Barford, S. Wofsy, C. Kucharik, E. Pyle, J. Budney, K. McKain, D. Fitzjarrald, M. Czikowsky, and J. W. Munger (2007), Factors controlling CO_2 exchange on timescales from hourly to decadal at Harvard Forest, *J. Geophys. Res.*, *112*(G2), G02020, doi:10.1029/2006JG000293.

Verbeeck, H., R. Samson, A. Granier, P. Montpied, and R. Lemeur (2008), Multi-year model analysis of GPP in a temperate beech forest in France, *Ecol. Modell.*, *210*, 85–103, doi:10.1016/j.ecolmodel.2007.07.010.

Verma, S. B., et al. (2005), Annual carbon dioxide exchange in irrigated and rainfed maize-based agroecosystems, *Ag. For. Met*, *131*, 77–96, doi:10.1016/j.agrformet.2005.05.003.

Vickers, D., C. Thomas, C. Pettijohn, J. Martin, and B. E. Law (2009), Influence of disturbance and water stress on sequestration of atmospheric carbon dioxide by ponderosa pine forests, *Agric. For. Meteorol.*, *22*, 111–114.

Wang, Y. P., and R. Leuning (1998), A two-leaf model for canopy conductance, photosynthesis and partitioning of available energy I: Model description and comparison with a multi-layered model, *Agric. For. Meteorol.*, *91*, 89–111, doi:10.1016/S0168-1923(98)00061-6.

Waring, R. H., and W. E. Winner (1996), Constraints on terrestrial primary production along the Pacific Coast of North and South America, in *High Latitude Rain Forests and Associated Ecosystems of the West Coast of the Americas: Climate, Hydrology, Ecology, and Conservation*, edited by R. G. Lawford, P. Alaback, and E. R. Fuentes, pp. 89–102, Springer-Verlag, New York, doi:10.1007/978-1-4612-3970-3_6.

Weng, E. S., and Y. Luo (2008), Soil hydrological properties regulate grassland ecosystem responses to multifactor global change: A modeling analysis, *J. Geophys. Res.*, *113*, G03003, doi:10.1029/2007JG000539.

Yang, X., V. Wittig, A. K. Jain, and W. Post (2009), Integration of nitrogen cycle dynamics into the Integrated Science Assessment Model for the study of terrestrial ecosystem responses to global change, *Global Biogeochem. Cycles*, *23*, GB4029, doi:10.1029/2009GB003474.

Zha, T., et al. (2009), Carbon sequestration in boreal jack pine stands following harvesting, *Global Change Biol.*, *15*, 1475–1487, doi:10.1111/j.1365-2486.2008.01817.x.

Zhan, X., Y. Xue, and G. J. Collaz (2003), An analytical approach for estimating CO_2 and heat fluxes over the Amazonian region, *Ecol. Modell.*, *162*, 97–117, doi:10.1016/S0304-3800(02)00405-2.

Zhao, M., and S. Running (2010), Drought-induced reduction in global terrestrial net primary production from 2000 through 2009, *Science*, *329*, 940–943, doi:10.1126/science.1192666.

Zhao, T., D. G. Brown, and K. M. Bergen (2007), Increasing gross primary production (GPP) in the urbanizing landscapes of southeastern Michigan, *Photogramm. Eng. Remote Sens.*, *73*(10), 1159–1167.