A model-data intercomparison of CO_2 exchange during large scale droughts: Results from the North American Carbon Project Site Synthesis

Corresponding author:

Christopher R. Schwalm[1]

[1] Graduate School of Geography, Clark University, Worcester, MA 01610, USA

cschwalm@clarku.edu, Tel: 508-793-7711

Contributing authors:

Christopher A. Williams[2], Kevin Schaefer[3], Ryan Anderson[4], M. Altaf Arain[5], Ian Baker[6], Alan Barr[7], T. Andrew Black[8], Guangsheng Chen[9], Jing Ming Chen[10], Philippe Ciais[11], Kenneth J. Davis[12], Ankur Desai[13], Michael Dietze[14], Danilo Dragoni[15], Marc L. Fischer[16], Larry Flanagan[17], Robert Grant[18], Lianhong Gu[19], David Hollinger[20], R. César Izaurralde[21], Chris Kucharik[22], Peter Lafleur[23], Beverly E. Law[24], Longhui Li[25], Zhengpeng Li[26], Shuguang Liu[27], Erandathie Lokupitiya[28], Yiqi Luo[29], Siyan Ma[30], Hank Margolis[31], Roser Matamala[32], Harry McCaughey[33], Russell K. Monson[34], Walter C. Oechel[35], Changhui Peng[36], Benjamin Poulter[37], David T. Price[38], Dan M. Riciutto[39], William J. Riley[40], Alok Kumar Sahoo[41], Michael Sprintsin[42], Jianfeng Sun[43], Hanqin Tian[44], Christina Tonitto[45], Hans Verbeeck[46], Shashi B. Verma[47]

[2] Graduate School of Geography, Clark University, Worcester, MA 01610, USA
cwilliams@clarku.edu, Tel: 508-793-7323

[3] National Snow and Ice Data Center, University of Colorado at Boulder, Boulder, CO 80309, USA
kevin.schaefer@nsidc.org, Tel: 303-492-8869

[4] Numerical Terradynamic Simulation Group, University of Montana, Missoula, MT 59812, USA
ryan.anderson@ntsg.umt.edu, Tel: (406) 243-6263

[5] School of Geography and Earth Sciences, McMaster University, Hamilton, ON L8S 4K1, Canada
arainm@mcmaster.ca, Tel: 905-525-9140

[6] Atmospheric Science Department, Colorado State University, Fort Collins, CO 80523, USA

baker@atmos.colostate.edu, Tel: 970-491-4948

[7] Climate Research Division, Atmospheric Science and Technology Directorate, Saskatoon, SK S7N 3H5, Canada

alan.barr@ec.gc.ca, Tel: 306-975-4324

[8] Faculty of Land and Food Systems, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

andrew.black@ubc.ca, Tel: 604-822-2730

[9] School of Forestry and Wildlife Sciences, Auburn University, Auburn, AL 36849, USA

chengu1@auburn.edu, Tel: 334-844-8057

[10] Department of Geography and Program in Planning, University of Toronto, Toronto, ON M5S 3G3, Canada

chenj@geog.utoronto.ca, Tel: 416-978-7085

[11] Laboratoire des Sciences du Climat et de l'Environnement, CE Orme des Merisiers, Gif sur Yvette, 91191 France

philippe.ciais@cea.fr, Tel: +33 1 6908 9506

[12] Department of Meteorology, Pennsylvania State University, University Park, PA 16802, USA

davis@meteo.psu.edu, Tel: 814-863-8601

[13] Center for Climatic Research, University of Wisconsin - Madison, Madison, WI 53706, USA

desai@aos.wisc.edu, Tel: 608-265-9201

[14] Department of Plant Biology, University of Illinois - Urbana Champaign, Urbana, IL 61801, USA

mdietze@life.uiuc.edu, Tel: 217-265-8020

[15] Department of Geography, Indiana University, Bloomington, IN 47405, USA

ddragoni@indiana.edu, Tel: 812-855-5557

[16] Atmospheric Science Department, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

mlfischer@lbl.gov, Tel: 510-486-5539

[17] Department of Biological Sciences, University of Lethbridge, Lethbridge, AB T1K 3M4, Canada

larry.flanagan@uleth.ca. Tel: 403-380-1858.

[18] Department of Renewable Resources, University of Alberta, Edmonton, AB T6G 2E3, Canada

robert.grant@afhe.ualberta.ca, Tel: 780-492-6609

[19] Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

lianhong-gu@ornl.gov, Tel: 865-241-5925

[20] Northern Research Station, USDA Forest Service, NH 03824, USA

davidh@hypatia.unh.edu, Tel: 603-868-7673

[21] Pacific Northwest National Laboratory and University of Maryland, College Park, MD 20740, USA

cesar.izaurralde@pnl.gov, Tel: 301-314-6751

[22] Department of Agronomy & Nelson Institute Center for Sustainability and the Global Environment, University of Wisconsin - Madison, Madison, WI 53706, USA
kucharik@wisc.edu, Tel: 608-263-1859

[23] Department of Geography, Trent University, Peterborough, ON K9J 7B8, Canada
plafleur@trentu.ca, Tel: 705-748-1011

[24] College of Forestry, Oregon State University, Corvallis, OR 97331, USA
bev.law@oregonstate.edu, Tel: 541-737-6111

[25] Laboratoire des Sciences du Climat et de l'Environnement, CE Orme des Merisiers, Gif sur Yvette, 91191 France
longhui.li@lsce.ipsl.fr, Tel: +33 1 6908 4102

[26] ASRC Research and Technology Solutions, Sioux Falls, SD 57198, USA
zli@usgs.gov, Tel: 605-594-6864

[27] Earth Resources Observation and Science, Sioux Falls, SD 57198, USA
sliu@usgs.gov, Tel: 605-594-6168

[28] Department of Atmospheric Science, Colorado State University, Fort Collins, CO 80523, USA
erandi@atmos.colostate.edu, Tel: 970-491-8915

[29] Department of Botany and Microbiology, University of Oklahoma, Norman, OK 73019, USA
yluo@ou.edu, Tel: 405-325-1651

[30] Department of Environmental Science, Policy and Management & Berkeley Atmospheric Science Center, University of California, Berkeley, Berkeley, CA 94720, USA

sma@berkeley.edu, Tel: 510-642-2421

[31] Centre d'études de la forêt, Faculté de foresterie, de géographie et de géomatique, Université Laval, Québec, QC G1V 0A6, Canada

hank.margolis@sbf.ulaval.ca, Tel: 418-656-7120

[32] Argonne National Laboratory, Biosciences Division, Argonne, IL 60439, USA

matamala@anl.gov, Tel: 630-252-9270

[33] Department of Geography, Queen's University Kingston, ON K7L 3N6, Canada

mccaughe@post.queensu.ca, Tel: 613-533-6035

[34] Department of Ecology and Evolutionary Biology, University of Colorado at Boulder, Boulder, CO 80309, USA

russell.monson@colorado.edu, Tel: 303-492-6319

[35] Department of Biology, San Diego State University, San Diego, CA 92182, USA

oechel@sunstroke.sdsu.edu, Tel: 619-594-4818

[36] Department of Biology Sciences, University of Quebec at Montreal, Montreal, QC H3C 3P8, Canada

peng.changhui@uqam.ca, Tel: 514-987-3000

[37] Swiss Federal Institute for Forest, Snow and Landscape, CH-8903 Birmensdorf, Switzerland

benjamin.poulter@wsl.ch, Tel: +41 44 7392 215

[38] Northern Forestry Centre, Canadian Forest Service, Edmonton, AB T6H 3S5, Canada

dprice@nrcan.gc.ca, Tel: 780-435-7249

[39] Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

ricciutodm@ornl.gov, Tel: 865-574-7067

[40] Climate and Carbon Sciences, Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

wjriley@lbl.gov, Tel: 510-486-5036

[41] Center for Research on Environment and Water, Calverton, MD 20705, USA

asahoo@iges.org, Tel: 240-247-1444

[42] Department of Geography and Program in Planning, University of Toronto, Toronto, ON M5S 3G3, Canada

misprin@gmail.com, Tel: 416-946-7715

[43] Department of Biology Sciences, University of Quebec at Montreal, Montreal, QC H3C 3P8, Canada

jianfeng_sun@yahoo.ca, Tel: 514-987-3000

[44] School of Forestry and Wildlife Sciences, Auburn University, Auburn, AL 36849, USA

tianhan@auburn.edu, Tel: 334-844-1059

[45] Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA

ctonitto@cornell.edu, Tel: 607-254-4257

[46] Faculty of Bioscience Engineering, Ghent University, 9000 Ghent, Belgium

hans.verbeeck@ugent.be, Tel: +32 9 264 61 13

[47] School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE 68583, USA

sverma1@unl.edu, Tel: 402-472-6702

Abstract

Our current understanding of terrestrial carbon processes is represented in various models that are routinely used to integrate and scale measurements of CO_2 exchange from remote sensing and other spatiotemporal data. Yet assessments are rarely conducted to determine how well models simulate carbon processes across a range of vegetation types and environmental conditions. Here we compare observed and simulated monthly CO_2 exchange using 44 eddy covariance flux towers in North America and model runs from 22 terrestrial biosphere models. The analysis period spans ~220 site-years, 10 biomes, and includes two large scale drought events in North America, providing a natural experiment to evaluate model skill as a function of drought and seasonality. We also evaluate models' ability to simulate the seasonal cycle of CO_2 exchange using Taylor diagrams and analyze links between model characteristics, site history, and model skill. Overall model performance was poor; the difference between observations and simulations was 10-times observational uncertainty, with forested ecosystems better predicted than non-forested. Model-data agreement was highest in summer and in temperate evergreen forests. In contrast, model performance was poor in spring and fall, especially in ecosystems with large deciduous components, and during wet or dry periods during the growing season. Models used across most biomes, the mean model ensemble, and an optimized model showed high consistency with observations. Overall skill was higher for models with simplified or detailed carbon and soil structure as well as those that estimated net ecosystem exchange as the difference between gross primary productivity and autotrophic respiration.

Keywords: carbon modeling, ecosystem models, model validation, carbon exchange, drought, North American Carbon Program

Introduction

Drought is a reoccurring phenomenon in all climates (Larcher, 1995) and is characterized by a partial loss in plant function due to water limitation and heat stress. For terrestrial CO_2 exchange, drought typically reduces photosynthesis more than respiration (Baldocchi, 2008; Ciais et al., 2005; Schwalm et al., 2009), resulting in decreased net carbon uptake from the atmosphere. In the recent past drought conditions have become more prevalent globally (Dai et al., 2004) and in North America (Cook et al., 2004). Both incidence and severity of drought (Seager et al., 2007b) as well as heatwaves (Meehl et al., 2004) are expected to further increase in conjunction with global warming (Houghton et al., 2001; Huntington, 2006; Sheffield & Wood, 2008; Trenberth et al., 2007).

There is a continued need for models to improve consistency and agreement with observations (Friedlingstein et al., 2006), both overall and under more frequent drought events related to global environmental change. Past validation studies of terrestrial biosphere models focused only on few models and sites, typically in close proximity and primarily in forested biomes, with differing objectives, levels of calibration to site data, and parameter values (e.g., Amthor et al., 2001; Delpierre et al., 2009; Grant et al., 2005; Hanson et al., 2004; Granier et al. 2007; Ichii et al., 2009; Ito, 2008; Siqueira et al., 2006; Zhou et al., 2008). Furthermore, assessing model performance relative to drought

requires high quality observed CO_2 exchange data, a reliable drought metric as well as a natural experiment across sites and drought conditions.

In this study we evaluate model performance using terrestrial CO_2 flux data and simulated fluxes collected from 1991 to 2007. This timeframe included two widespread droughts in North America: 1) the turn-of-the-century drought from 1998 to 2004 that was centered in the interior West of North America (Seager et al., 2007a) and 2) a smaller-scale drought event in the southern continental Untied States from winter of 2005/2006 through October 2007 (Seager et al., 2009). During these events Palmer Drought Severity Index values (Cook et al., 2007; Dai et al., 2004) and precipitation anomalies (Seager et al., 2007a; 2009) were highly negative over broad geographic areas. Ongoing eddy covariance measurements (Baldocchi et al., 2001) under these circumstances provided flux data across gradients of time, space, seasonality, and drought. We use these data to examine model skill relative to site-specific drought severity, climatic season, and time. We also link model behavior to model architecture and site-specific attributes. Specifically, we address the following questions: Are current state-of-the-art terrestrial biosphere models capable of simulating CO_2 exchange subject to gradients in dryness and seasonality? Are these models able to reproduce the seasonal variation of observed CO_2 exchange across sites? Are certain characteristics of model structure coincident with better model-data agreement? Which biomes are simulated poorly/well?

Methods

Modeled and observed net ecosystem exchange (*NEE*, net carbon balance including soils where negative values indicate outgassing of CO_2 to the atmosphere) data were analyzed from 22 terrestrial biosphere models (Table 1) and 44 eddy covariance (EC) sites spanning ≈ 220 site-years and 10 biomes in North America (Table 2). All terrestrial biosphere models analyzed simulated carbon cycling with process based formulations of varying detail for component carbon fluxes. Simulated *NEE* was based on model-specific runs using gap-filled observed weather at each site and locally observed values of soil texture according to a standard protocol (http://isynth-site.pbworks.com). Apart from the mean model ensemble across all models and LoTEC, an assimilated model (Ricciuto et al., 2008), each model was spun up to steady state initial conditions with a target *NEE* of zero integrated over the last five years of the simulation period.

Gaps in the meteorological data record occurred at EC sites due to data screening or instrument failure. Missing values of air temperature, humidity, shortwave radiation, and precipitation data, i.e., key model inputs, were filled using DAYMET (Thornton et al., 1997) before 2003 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 using the phasing from observed mean diurnal cycles calculated from a 15-day moving window. The phasing used a sine wave assuming peak values at 15:00 local standard time (LST) and lowest values at 3:00 LST. In the absence of station data a 10-day running mean diurnal cycle was used (http://nacp.ornl.gov/docs/Site_Synthesis_Protocol_v7.pdf).

EC data were produced by AmeriFlux and Fluxnet Canada investigators and processed as a synthesis product of the North American Carbon Program (NACP) Site Level Interim Synthesis (http://www.nacarbon.org/nacp/). The observed *NEE* were corrected for storage, despiked (i.e., outlying values removed), filtered to remove conditions of low turbulence (friction velocity filtered), and gap-filled to create a continuous time series (Barr et al., 2004). The time series included estimates of random uncertainty and uncertainty due to the friction velocity filtering (Barr et al., 2004; ftp://nacp.ornl.gov/synthesis/2008/firenze/site/observations/uncertainty/). In this analysis *NEE* was aggregated to monthly values using only non-gap-filled data, i.e., observed values deemed spurious and subsequently infilled were not considered. Coincident modeled *NEE* values were similarly excluded. This removed the influence of gap-filling algorithms in the comparison of observed and modeled *NEE*.

Drought level was quantified using the 3-month Standard Precipitation Index (SPI, McKee et al., 1993). Monthly SPI values were taken from the U.S. Drought Monitor (http://drought.unl.edu/DM/) whereby each tower was matched to nearby meteorological station(s) indicative of local drought conditions given proximity, topography, and human impact. This study used three drought levels: dry required SPI < -0.8, wet corresponded to SPI > +0.8, otherwise normal conditions existed. Climatic season was defined by four seasons of three months each with winter given by December, January, and February.

Model skill

Model-data mismatch was evaluated using normalized mean average error (*NMAE*; Medlyn et al., 2005), the reduced χ^2 statistic (χ^2; Taylor, 1996) as well as Taylor diagrams and skill (*S*; Taylor, 2001). The first metric quantifies bias, the "average distance" between observations and simulations in units of observed mean *NEE*:

$$NMAE = \sum_{ijkl} \frac{NEE_{obs} - NEE_{sim}}{n\overline{NEE_{obs}}}, \qquad (1)$$

where the overbar indicates averaging across all values, *n* is sample size, the subscript *obs* is for observations and *sim* is for modeled estimates. The summation is for any arbitrary data group (denoted by subscripts on the summation operator only) where subscript *i* is for site, *j* is for model, *k* is for climatic season, *l* is for drought level.

The second metric used to evaluate model performance was the reduced χ^2 statistic. This is the squared difference between paired model and data points over observational error normalized by degrees of freedom:

$$\chi^2 = \frac{1}{n}\sum_{ijkl}\left(\frac{NEE_{obs} - NEE_{sim}}{4\delta_{NEE}}\right)^2, \qquad (2)$$

where δ_{NEE} is uncertainty of monthly *NEE* (see below), 4 normalizes the uncertainty in observed *NEE* to correspond to a 95% confidence interval, the summation is across any arbitrary data group (denoted by subscripts on the summation operator). χ^2 values are linked to error variance with < 1 indicating an overestimation and > 1 an underestimation.

A value of unity indicates that model and data are in agreement relative to data uncertainty.

A final characterization of model performance used Taylor diagrams (Taylor, 2001); visual displays based on pattern matching, i.e., the degree to which simulations matched the temporal evolution of monthly *NEE*. Taylor plots are polar coordinate displays of the linear correlation coefficient (ρ), centered *RMSE* (pattern error without considering bias), and the standard deviation of *NEE* (σ) where all three quantities were calculated using the full data record for each combination of site and model (ranging from 7 to 178 months). Taylor diagrams were constructed for the mean model ensemble and across-site mean model performance. More generally, each coordinate point for any arbitrary data group can be scored:

$$S = \frac{2(1+\rho)}{\left(\sigma_{norm} + 1/\sigma_{norm}\right)^2}, \qquad (3)$$

where *S* is the model skill metric bound by zero and unity where unity indicates perfect agreement, and σ_{norm} is the ratio of simulated to observed normalized standard deviation (Taylor, 2001).

To scale any model skill metric across gradients of site, biome, model, seasonality, and dryness level we aggregated across data groups weighting each by sample size. For example, χ^2 for model *I* is given by:

$$\chi^2_{i=I} = \sum_{jkl} \frac{n_{jkl}\chi^2_{jkl}}{n_{i=I}}, \quad (4)$$

where the summation is across all groups where model *I* was used, and $\chi^2_{i=I}$ references model *I* across all sites, seasons, and levels of dryness as denoted by subscripts on the summation operator. We did not evaluate model performance for any data group with $n <$ 3. In sum, Taylor displays and skill examined models' ability to mimic the monthly trajectory of observed *NEE*, the calculation of *NMAE* quantified bias in units of mean observed *NEE*, and χ^2 values quantified how well model-data mismatch scales with flux uncertainty.

Observational flux uncertainty

We calculated the standard error of monthly *NEE* (δ_{NEE}) by combining random uncertainty and uncertainty associated with the friction velocity threshold (u_*^{Th}), a value use to identify and reject spurious nighttime *NEE* measurements. Random uncertainty was estimated following Richardson & Hollinger (2007): (i) generate synthetic *NEE* data using the gap-filling model (Barr et al., 2004) for a given site-year, (ii) introduce gaps as in the observed data with u_*^{Th} filtering, (iii) add noise, (iv) infill gaps using gap-filling model, (v) repeat the process 1000 times for each site. The random uncertainty component of δ_{NEE} was then the standard deviation across all 1000 realizations aggregated to months.

The u_*^{Th} uncertainty component of δ_{NEE} was also estimated using the Monte Carlo methods. Here 1000 realizations of *NEE* were generated using 1000 draws from a

distribution of u_*^{Th}. The distribution of u_*^{Th} was based on binning the raw flux data with respect to climatic season, temperature, and site-year and estimating u_*^{Th} in each bin (Papale et al., 2005). The standard deviation across all realizations gave the u_*^{Th} uncertainty component of δ_{NEE}. Both components were combined in quadrature to one standard error of monthly *NEE* (= δ_{NEE}) (ftp://nacp.ornl.gov/synthesis/2008/firenze/site/observations/uncertainty/).

Relating model skill to model structure and site history

The models evaluated here range widely in their emphasis and structure (Table 1). Some focus on biophysical calculations (SiB3, BEPS), where as others emphasize biogeochemistry (ORCHIDEE) or ecosystem dynamics (ED2). However as terrestrial biosphere models simulate carbon cycling with hydrological variables most models contain both biophysics and biogeochemistry. This motivated characterizing model structure with definite attributes, e.g., prognostic vs. prescribed canopy, number of soil pools, and type of *NEE* algorithm (Table 3). To resolve how such characteristics and site history impacted model skill we calculated *S* for all observed combinations of site, model, seasonality, and drought level and cross-referenced these with 13 site history and 15 model attributes (Table 3). Only 20 models were available for the RTA exercise, the MEAN model and the optimized LoTEC were excluded. We used *S* as it is bound by zero (no agreement) and unity (perfect agreement) in contrast to *NMAE* and χ^2 which are unbound. The Taylor skill metric (*S*) was discretized into three classes based on terciles. These classes, representing three tiers of model-data agreement, were then related to biome, climatic season, drought level, site history, and model structure using regression

tree analysis (RTA) as a supervised classification algorithm. RTA is a form of binary recursive partitioning (Breiman et al., 1984) that successively splits the data (Taylor skill classes as the response; all other attributes as predictors) into subsets (nodes) by minimizing within-subset variation. The result is a pruned tree-like topology whereby predicted values (Taylor skill metric class) are derived by a top-to-bottom traversal following the rules (branches) that govern subset membership until a predicted value is reached (terminal node). The splitting rules at each node as well as its position allow for a calculation of relative variable importance (Breiman et al., 1984) with the most important variable given a score of 100. Variables of high importance were further analyzed using conditional means, i.e., comparing mean values for each predictor value, with statistical differences determined using Bonferroni methods.

Results

Model-data agreement relative to climatic season, dryness, and biome

Overall agreement across $n = 31025$ months was better in forested than non-forest biomes; both *NMAE* (Table 4) and χ^2 values (Table 5) were closer to zero. At the biome level model skill was loosely ranked in four tiers: evergreen needleleaf forests in the temperate zone > all other forested biomes > grasslands, woody savannahs > croplands, shrublands, wetlands > tundra. These rankings were robust across models used in the majority of biomes although some divergence was apparent for croplands and shrublands (Figure 1). Relative to seasonality and drought level models were most consistent with observations during periods of peak biological activity (climatic summer) under dry conditions (Figure 2). However, across the three levels of dryness changes in model-data

agreement were negligible for *NMAE* (~4% change, Table 4) but more pronounced for χ^2 (8.10 to 12.72, Table 5). Averaged over the warm season (excluding climatic winter) dry/wet conditions were coincident with worse model-data agreement, e.g., *NMAE* was -1.32, -1.26, and -1.74 for dry, normal, and wet respectively. In biomes with a clear seasonal cycle in leaf area index (LAI) a loss of model skill occurred during climatic spring and fall (Table 4 & 5).

Skill metrics by model

Model skill was, regardless of metric, highly variable. Of the three, *NMAE* was related to both Taylor skill and χ^2 (ρ = -0.65; $p < 0.0001$). Jointly, high Taylor skill co-occurred with *NMAE* and χ^2 values closer to zero (Figure 3). Across models *NMAE* ranged from -0.42 of the overall mean observed flux to -2.18 for LoTEC and DNDC respectively. Values of χ^2 varied from 2.17 to 29.87 for LoTEC and CN-CLASS. Alternatively, the degree of model-data mismatch (the distance between observations and simulations) was at least 2.17 times the observational flux uncertainty. Similarly, Taylor skill showed a high degree of scatter across sites and within models (Figure 4), although two crop only models (SiBcrop and AgroIBIS), LoTEC, and ISOLSM were more conservative and showed a high degree of consistency with observations ($S > 0.8$).

Among crop models, SiBCrop and AgroIBIS performed well, especially in climatic spring and during wet conditions. In contrast, the crop only DNDC model exhibited poor model-data agreement with $\chi^2 > 15$ in climate spring and summer as well as across all drought levels. Although four crop only simulators were analyzed, the best agreement in

croplands (*NMAE* and χ^2 closer to zero) was achieved by SiB3 and Ecosys, models used in multiple biomes. With the highest Taylor skill and *NMAE* and χ^2 closest to zero, the LoTEC model (*NMAE* = -0.42, χ^2 = 2.17, *S* = 0.95) was most consistent with observations across all sites, dryness levels, and climatic seasons. This platform was optimized using a data assimilation technique, unique among model runs evaluated here, and was applied at 10 sites. In addition, the mean model ensemble (MEAN) also performed well (*NMAE* = -0.74, χ^2 = 3.35, *S* = 0.80). For individual models (*n* = 12) used at a wider range of sites (at least 24 sites) model consistency with observations was highest for Ecosys (*NMAE* = -0.69, χ^2 = 7.71, *S* = 0.94) and lowest for CN-CLASS (*NMAE* = -1.50, χ^2 = 29.87, *S* = 0.48).

Site-level model-data agreement also showed a high degree of variability (Figure 5). At three croplands sites (US-Ne1, US-Ne2, and US-Ne3) Taylor skill across all models ranged from zero to unity. Both *NMAE* and χ^2 exhibited similar scatter by site (not shown). Even for the best predicted site on average (US-Syv) *S* ranged from 0.19 to 0.95. Similarly, two forested sites (CA-Qfo and CA-TP4) were predicted well (*S* > 0.5) by all models; whereas only one tundra site (US-Atq) was consistently poorly predicted (S < 0.5). Despite the wide range in model performance, model skill (*NMAE*, χ^2, and *S*) was not related to the number of sites ($|\rho| < 0.2$; $p > 0.5$) or biomes ($|\rho| < 0.2$; $p > 0.3$) simulated, i.e., using a more general rather than a specialized model did not result in a loss in model performance. Also, model-data agreement was not better at sites with longer data records ($|\rho| < 0.1$; $p > 0.1$). Lastly, model skill was not related to how close model spinup and initial conditions approximated steady state. Although long-term simulated *NEE* varied from near zero (steady state) to ~-500 g C m^{-2} yr^{-1}, there was no

relationship ($|\rho| < 0.1$; $p > 0.2$) between these values and any of the three model skill metrics.

Model and site-specific consistency with observations using Taylor diagrams

Average model performance (both across-site and across-model) was evaluated using Taylor diagrams based on all simulated and observed *NEE* monthly integrals. Better model performance was indicated by proximity to the benchmark, representing the observed state. The benchmark was normalized by observed standard deviation such that the distance of σ and *RMSE* from the benchmark was in observed σ units. Similar to model skill metrics, forested sites were better predicted than non-forested ones. The MEAN model, i.e., average model performance across all models, at all sites (excluding CA-SJ2 and US-Atq) showed $\rho \geq 0.2$ but generally (33 of 44 sites) underpredicted the variability associated with monthly *NEE* at forested (Figure 6) and non-forested (Figure 7) sites. Similarly, 40 of 44 sites were predicted with *RMSE* < σ. Also 8 (6 forested and two croplands sites: CA-Obs, CA-Qfo, CA-TP4, US-Ho1, US-IB1, US-MMS, US-Ne3, US-UMB) of the 44 sites were predicted with $\rho \geq 0.95$ and *RMSE* < 1. The worst predicted site was CA-SJ2 with $\rho = -0.67$, $\sigma = 4.3$, and *RMSE* = 5.1.

Overall model performance, aggregated across sites, was similar (Figure 8). Most models underpredicted variability and showed *RMSE* < σ. Of all 22 models only DNDC exhibited $\rho < 0.2$. Based on proximity to the benchmark, i.e., a high *S* value (Figure 3), the best models were: EPIC (crop only model used on one site), ISOLSM (used on 9 sites), LoTEC (data assimilation model), SiBcrop and AgroIBIS (crop only models),

EDCM (used on 10 sites), Ecosys and SiBCASA (more generalist models used on 39 and 35 sites respectively), and MEAN (mean model ensemble for all 44 sites). All of these "best" models had $\rho > 0.8$, $RMSE < 0.8$ and slightly underpredicted variability; except the crop only models and Ecosys where variability was overpredicted. Models whose average behavior was furthest away from the benchmark were DNDC followed by BEPS.

Links between model skill, model structure, and site history

Across all sampled combinations of site, model, climatic season and drought ($n = 3132$ groups), biome was the most important factor in the distribution of model skill (Figure 9). Biome was followed by climatic season and stand age, the highest scored site-specific attribute. None of the 12 evaluated site disturbances (Table 3) achieved an importance score of at least 25. Apart from drought, the remaining variables were model-specific structural attributes: the number of soil layers, vegetation pools, and soil pools. Three carbon flux calculations also had a variable score > 25, with *NEE* being the highest scored algorithm.

With 8 of the 14 evaluated model-specific factors having a variable importance score of at least 25, model attributes were linked to model skill. Comparing mean S across all levels of a single model attribute (Figure 10) revealed three instances where model-specific factors were statistically related to model skill: Models using a prescribed canopy achieved higher S (= 0.54) than either prognostic or semi-prognostic models ($S = 0.42$; $p < 0.05$). Using a daily time step was coincident ($S = 0.42$) with a loss in model skill relative to non-daily time steps ($S = 0.53$; $p < 0.05$). Finally, calculating *NEE* as the

difference between *GPP* and autotrophic respiration was superior ($S = 0.53$) to other calculation methods ($S = 0.42$; $p < 0.05$). Both vegetation pools and soil layers exhibited a weak pattern whereby models with coarse or very detailed vegetation and soil structure were more consistent with observed data (Figure 10).

Despite these effects model attributes were of secondary importance. The change in *S* relative to biome varied from 0.30 to 0.55; a much larger range than seen for model attributes. Similarly, the high variable importance scores for biome and climatic season, as well as the lower score for drought level, corroborated the relationships between these factors and model skill as seen with *NMAE* and χ^2. While the regression tree algorithm achieved an accuracy of 66% for predicting Taylor skill class, the site history and model characteristics considered here did not explain the underlying cause of biome and seasonal differences in model skill.

Discussion

Effect of parameter sets on model performance

Model parameter sets are a large source of variability in terms of model performance (Jung et al., 2007b). They influence output and accuracy (Grant et al., 2005) and are more important for simulating CO_2 exchange than interannual climatic variability (Amthor et al., 2001). This is related to the use of biome-specific parameters relative to within-biome variability (Purves & Pacala, 2008). A corollary occurs in the context of EC observations as tower footprints can exhibit heterogeneity, particularly in soils, that is not reproduced in model site-specific parameters (Amthor et al., 2001).

The importance of model parameter sets was visible in this intercomparison in two ways. Firstly, biome had the highest variable importance score. Insomuch as models rely on biome-specific parameter values this finding indicates model parameter sets as a key factor in the distribution of model skill. Furthermore, the variability in model skill across sets of biome-specific constants (Figure 5) underscores that biomes may be too heterogeneous in time (Stoy et al., 2005; 2009) and space to assume constant parameter values relative to within-biome variability linked to climate (Hargrove et al. 2003). Secondly, the general high degree of within-model variation in model skill (Figure 4) suggested that model parameter sets may need to be refined to capture local, site-specific realities.

Effect of model structure on model performance

The smaller variable importance scores and changes in model skill relative to biome and climatic season (Figure 9 & 10) revealed only a secondary link between model skill and the model structural attributes evaluated here. In general, model-data agreement was highest when models did not use a daily time step, used a prescribed canopy, and calculated *NEE* as the difference between *GPP* and autotrophic respiration. Similarly, for soil layers, vegetation pools, and, to a lesser extent, soil pools, consistency with observations was highest in those models with either the simplest structure (e.g., one soil carbon pool in ISOLSM) or the most complex (e.g., Ecosys and SiBCASA with 15 soil layers). Models that exhibited all of these structural characteristics (SiBCASA, SiB3, and ISOLSM) showed high degrees of model-data agreement across all three skill metrics.

Similarly, Ecosys, which used a prognostic canopy but otherwise had similar structural characteristics as SiBCASA, also performed well. It is noteworthy that models with a prognostic canopy required more detailed carbon pool and soil layer characterizations such as Ecosys. There was no model with a low number of carbon pools and soil layers and a prognostic canopy that was in the top tercile of model skill based on any skill metric except SiBcrop and AgroIBIS for Taylor skill in croplands only.

The success of those models with desirable, determined from predicting model skill classes using model attributes, model structure was not confined to single biomes. While non-forested sites were not as well simulated as forested sites, model-data consistency in these biomes, especially croplands, sites was highest with Ecosys and SiB3; not the crop only simulators. This suggests that a single model with requisite model structural attributes can be used in all types of ecosystems. As multiple models with both good performance and desirable structural attributes were available, model choice is recast as a tradeoff between ease of parameterization and spin up. Finally, while the choice of model structure is an important consideration for model-data agreement using multiple models (MEAN) or data assimilation to optimize model parameter sets (LoTEC) resulted in better consistency with observations.

Links between model performance and environmental factors

Dryness level was related to model skill. During the warm season (all climatic seasons excluding winter) model performance declined under dry/wet conditions.

While this points to process uncertainty, especially during dry conditions (Sitch et al., 2008), ecosystem response to longer-term drought can exhibit lags and positive feedbacks (Arnone et al., 2008; Granier et al., 2007; Thomas et al., 2009; Williams et al., 2009) that were not explicitly included in the drought metric used in this study.

In spring and fall, especially for biomes with a significant deciduous component, models showed a decline in model skill relative to periods of peak biological activity (climatic summer) (see also Morales et al., 2005). Phenological cues are known to influence the annual carbon balance at multiple scales (Barr et al., 2007; Delpierre et al., 2009; Keeling et al., 1996). The loss of model skill seen in this study during spring and fall was likely linked to uncertainty (process- and parameter-based) in leaf initiation and senescence (Hanson et al., 2004) as well as the initiation and cessation of canopy photosynthesis relative to soil moisture and soil temperature (Waring & Running, 2007). In this study seasonality was second only to biome in driving model skill (Figure 9). This and the lack of link between model skill and site history strongly implicate phenology as a needed refinement of terrestrial biosphere simulators.

The evergreen needleleaf forest biome diverged in performance based on whether the sites were located in the temperate or boreal zones. A similar divergence was reported using Biome-BGC, LPJ and ORCHIDEE to simulate gross CO_2 uptake across a temperature gradient in Europe (Jung et al., 2007a); average relative *RMSE* was higher for evergreen needleleaf forests in the boreal zone. This was linked to an overestimation of LAI at the boreal sites and relationships between resource availability and leaf area

(Friedlingstein et al., 2006; Jung et al., 2007a; Stich et al., 2008). Additionally, recent observations in the circumboreal region, where all boreal evergreen needleleaf forested sites are located, suggest that transient effects of climate change, e.g., increased severity and intensity of natural disturbances (fire, pest outbreaks) and divergence from climate normals in temperature, have already occurred (Soja et al., 2007). We speculate the loss of model skill in boreal relative to temperate evergreen needleleaf forests was linked to insufficient characterization of cold temperature sensitivity of metabolic processes and water flow in plants as well as freeze-thaw dynamics (Schaefer et al., 2007; 2009) and that this was exacerbated by the effects of transient climate change.

Effects of site history and protocol on model evaluation

Disturbance regime and how a model treats disturbance are known to impact model performance (Ito, 2008). In this study stand age impacted model skill whereas site history was of marginal importance (Figure 9). Comparing sites with and without fire or harvest activity showed no significant difference in mean model skill ($p > 0.25$), i.e., disturbance and age were confounded. However, CA-SJ2, the worst predicted site (Figure 6), was harvested in 2000 and scarified in 2002, and US-SO2, a second poorly predicted shrubland site (Figure 7), suffered catastrophic wildfire during the analyzed data record. The poor model performance for recently disturbed sites follows from the steady state assumption used in simulation and the absence of modeling logic to accommodate disturbance. However, the distribution of site history metrics was skewed; only few sites were burned, harvested, or in the early stages of recovery from disturbance when *NEE* is more nonlinear relative to established stands. Furthermore, age class was biased toward

older stands; of the 17 forested sites only one was classified as a young stand. Other site characteristics were also unbalanced; all non-forested biomes occurred on five or less sites; with only site each for shrublands and woody savannahs. While regression trees are inherently robust additional observed and simulated fluxes in rapidly growing young forested stands and undersampled biomes are desirable to better characterize model performance.

Other aspects of the NACP site synthesis protocol also influenced the interpretation of our results. Firstly, this analysis focused solely on non-gap-filled data to allow the model-data intercomparison to inform model development. However, the low turbulence (friction velocity) filtering removes more data at night than during the day, so our analysis may be skewed towards daytime conditions. Secondly, each model that used remotely sensed inputs (such as LAI) repeated an average seasonal cycle calculated from site-specific time series based on all pixels within 1 km of the tower site. This likely deflated relevant variable importance scores (Figure 9) and precluded a full comparison of prescribed vs. prognostic LAI. While only few models used such inputs (Table 1), including one of the best performing generalist models SiBCASA, removing the inherent bias of an invariant seasonal cycle over multiple years may improve model performance. Incorporating disturbance information to recreate historical land use and disturbance, especially for recent site entries, could also improve model performance. Lastly, despite the model simulation protocol's emphasis on steady state, few, if any, of the sites are actually at steady state. This resulted in an inherent bias between simulated and observed *NEE* for all sites regardless of site history. Relaxing the steady state assumption

(Carvalhais et al., 2008) or initializing using observed wood biomass and the quasi-steady state assumption (Schaefer et al., 2008) could improve model performance.

Conclusion

We used observed CO_2 exchange from 44 eddy covariance towers in North America with simulations from 22 terrestrial biosphere models to examine model skill across gradients in dryness, seasonality, and biome. Models' ability to match observed monthly net ecosystem exchange across levels of dryness and seasonality was generally poor; the mean squared distance (χ^2) between observations and simulations was ~10-times observational error. Overall, forested sites were better predicted than non-forested sites. Weaknesses in model performance concerned model parameter sets, phenology, especially for biomes with a clear seasonal cycle in leaf area index, and abnormally dry or wet conditions during the growing season. Sites with disturbances during the analyzed data record and undersampled biomes (grasslands, shrublands, wetlands, woody savannah, and tundra) also showed a large divergence between observations and simulations. In contrast, performance was higher when evaluating month-to-month trajectories using Taylor diagrams. This indicated that the temporal evolution of *NEE* is better modeled than responses to finer scaled changes in driving variables, albeit at the cost of under and overpredictions that cancel out over longer-term simulations. The highest degree of model-data agreement occurred in temperate evergreen forests across all climatic seasons and during summer across all biomes. Terrestrial biosphere simulators that calculated *NEE* as the difference between *GPP* and autotrophic respiration and with either simplistic (e.g., single soil carbon pool) or complex (e.g., 15

soil layers) characterizations of carbon pools and soil layers showed higher model skill relative to other model structural attributes. Models most consistent with observations included generalist models applied over a wide range of sites and biomes as well as an optimized model (parameters tuned using data assimilation), and a model ensemble (mean simulated value across several models). As generalist models (e.g., SiB3, Ecosys) with preferred model structural attributes performed well across all biomes, different model architectures were not needed for different types of ecosystems and model choice is recast as a function of ease of parameterization and initialization.

Acknowledgements

CRS, CAW, and KS were supported by the U.S. National Science Foundation grant ATM-0910766. We would like to thank the North American Carbon Program Site-Level Interim Synthesis team, the Modeling and Synthesis Thematic Data Center, 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. This study was in part supported by the U.S. National Aeronautics and Space Administration (NASA) grant NNX06AE65G, the U.S. National Oceanic and Atmospheric Administration (NOAA) grant NA07OAR4310115, and the U.S. National Science Foundation (NSF) grant OPP-0352957 to the University of Colorado at Boulder. This work was also partially supported by the U.S. Department of Energy and LBNL under Contract No. DE-AC02-05CH11231.

References

Amthor, JS et al. (2001) Boreal forest CO_2 exchange and evapotranspiration predicted by nine ecosystem process models: Intermodel comparisons and relationships to field measurements. J. Geophys. Res., 106(D24), 33,623-33,648.

Angert A, Biraud S, Bonfils C et al. (2005) Drier summers cancel out the CO_2 uptake enhancement induced by warmer springs. Proceedings of the National Academy of Sciences of the United States of America, 102, 10823-10827.

Arain MA, Yaun F, Black TA (2006) Soil-plant nitrogen cycling modulated carbon exchanges in a western temperate conifer forest in Canada. Agricultural and Forest Meteorology, 140, 171-192.

Arnone, JA, Verburg PSJ, Johnson DW et al. (2008) Prolonged suppression of ecosystem carbon dioxide uptake after an anomalously warm year. Nature 455:383-386.

Baldocchi D, Falge E, Gu LH 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. Bulletin of the American Meteorological Society, 82, 2415-2434.

Baldocchi, D (2008) Breathing of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. Australian Journal of Botany, 56, 1-26.

Baker IT, Prihodko L, Denning AS, Goulden M, Miller S, da Rocha HR (2008) Seasonal drought stress in the Amazon: Reconciling models and observations. J. Geophys. Res., 113, G00B01, doi:10.1029/2007JG000644.

Barr AG, Black TA, Hogg EH et al. 2004. Inter-annual variability in the leaf area index of a boreal aspen-hazelnut forest in relation to net ecosystem production. Agricultural Forest Meteorology, 126: 237-255.

Barr AG, Black TA, Hogg EH et al. (2007) Climatic controls on the carbon and water balances of a boreal aspen forest, 1994–2003. Global Change Biology, 13, 561-576.

Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Wadsworth, Belmont, CA, 358 pp.

Carvalhais N et al. (2008) Implications of the carbon cycle steady state assumption for biogeochemical modeling performance and inverse parameter retrieval. Global Biogeochem. Cycles, 22, GB2007, doi:10.1029/2007GB003033.

Causarano HJ, Shaw JN, Franzluebbers AJ, Reeves DW, Raper RL, Balkcom KS, Norfleet ML, Izaurralde RC (2007) Simulating field-scale soil organic carbon dynamics using EPIC. Soil Science Society of America Journal, 71, 1174-1185.

Chen JM, Govind A, Sonnentag O, Zhang Z, Barr A, Amiro B (2006) Leaf area index measurements at Fluxnet-Canada forest sites. Agric. For. Meteorol., 140, 257-268.

Ciais P et al (2005) Europe-wide reduction in primary productivity caused by the heat and drought in 2003. Nature 437:529-533

Cook ER, Woodhouse CA, Eakin CM, Meko DM, Stahle DW (2004) Long-term aridity changes in the western United States. Science, 306, 1015-1018.

Cook ER, Seager R, Cane MA, Stahle DW (2007) North American droughts: reconstructions, causes and consequences. Earth-Sci. Rev., 81, 93-134.

Dai A, Trenberth KE, Qian T (2004) A global data set of Palmer Drought Severity Index for 1870-2002: Relationship with soil moisture and effects of surface warming. J. Hydrometeorol., 5, 1117-1130.

Delpierre N, Soudani K, Francois C et al. (2009) Exceptional carbon uptake in European forests during the warm spring of 2007: a data–model analysis. Glob. Change Biol., 15, 1455-1474.

Friedlingstein P, Cox PM, Betts R et al. (2006) Climate-carbon cycle feedback analysis, results from the C^4MIP model intercomparison. Journal of Climate, 19, 3337-3353.

Granier A, Reichstein M, Breda N et al. (2007) Evidence for soil water control on carbon and water dynamics in European forests during the extremely dry year: 2003. Agricultural and Forest Meteorology, 143, 123–145.

Grant RF et al. (2005) Intercomparison of techniques to model high temperature effects on CO_2 and energy exchange in temperate and boreal coniferous forests. Ecol. Model., 188, 217–252.

Hanson PJ, Amthor JS, Wullschleger SD et al. (2004) Oak forest carbon and water simulations: model intercomparisons and evaluations against independent data. Ecological Monographs, 74, 443-489.

Hargrove WW, Hoffman FM, Law BE (2003) New Analysis Reveals Representativeness of AmeriFlux Network. Earth Observing System Transactions, American Geophysical Union 84(48):529.

Higuchi K, Shashkov A, Chan D, Saigusa N, Murayama S, Yamamoto S, Kondo H, Chen JM, Liu J, Chen B (2005) Simulations of seasonal and inter-annual variability of net CO_2 flux at Takayama with BEPS Ecosystem Model. Agricultural and Forest Meteorology, 134, 143-150.

Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Xia D, Maskell K, Johnson CA (eds) (2001) Climate Change 2001: The Scientific Basis: Contributions of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, New York, 881 pp.

Huntington TG (2006) Evidence for intensification of the global water cycle: Review and synthesis. J. Hydrol., 319, 83-95.

Hwang Y, Carbone GJ (2009) Ensemble forecasts of drought indices using a conditional residual resampling technique. Journal of Applied Meteorology and Climatology, 48, 1289-1301.

Ichii K, Suzuki T, Kato T et al (2009) Multi-model analysis of terrestrial carbon cycles in Japan: reducing uncertainties in model outputs among different terrestrial biosphere models using flux observations, Biogeosciences Discuss., 6, 8455-8502.

Ito A (2008) The regional carbon budget of East Asia simulated with a terrestrial ecosystem model and validated using AsiaFlux data. Agr. Forest Meteorol., 148, 738-747.

Jung M, Le Maire G, Zaehle S et al. (2007a) Assessing the ability of three land ecosystem models to simulate gross carbon uptake of forests from boreal to Mediterranean climate in Europe. Biogeosciences, 4, 647-656.

Jung M, Vetter M, Herold M et al. (2007b) Uncertainties of modeling gross primary productivity over Europe: A systematic study on the effects of using different drivers and terrestrial biosphere models. Global Biogeochem. Cycles, 21, GB4021, doi:10.1029/2006GB002915.

Keeling CD, Chin JF, Whorf TP (1996) Increased activity of northern vegetation inferred from atmospheric CO_2 measurements. Nature, 382, 146-149.

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

Kucharik CJ, Twine TE (2007) Residue, respiration, and residuals: Evaluation of a dynamic agroecosystem model using eddy flux measurements and biometric data. Agricultural and Forest Meteorology, 146, 134-158.

Larcher W (1995) Physiological Plant Ecology. Springer Verlag, Berlin, 506 pp.

Li et al. (2009) Modeling impacts of alternative farming management practices on greenhouse gas emissions from a winter wheat-maize rotation system in China. Agriculture, Ecosystems and Environment, 135, 24-33.

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

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

Loveland TR, Reed BC, Brown JF, Ohlen DO, Zhu J, Yang L, Merchant JW (2001) Development of a global land cover characteristics database and IGBP DISCover from 1-km AVHRR data. International Journal of Remote Sensing, 21, 1,303-1,330.

McKee TB, Doeskin NJ, Kleist J (1993) The relationship of drought frequency and duration to time scales. Proc. 8th Conf. on Applied Climatology, January 17-22, 1993, American Meteorological Society, Boston, Massachusetts, 179-184.

Medlyn BE, Robinson AP, Clement R, McMurtrie RE (2005) On the validation of models of forest CO_2 exchange using eddy covariance data: some perils and pitfalls. Tree Physiology, 25, 839-857.

Medvigy D, Wofsy SC, Munger JW, Hollinger DY, Moorcroft PR (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.

Meehl GA, Tebaldi C (2004) More intense, more frequent, and longer lasting Heat waves in the 21st Century. Science, 305, 994-997.

Mishra VR, Desai AK (2006) Drought forecasting using feed-forward recursive neural network. Ecol. Modell., 198, 127-138.

Moffat A, Papale D, Reichstein M et al. (2007) Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. Agricultural and Forest Meteorology, 147, 209-232.

Morales P, Sykes MT, Prentice IC et al (2005) Comparing and evaluating process-based ecosystem model predictions of carbon and water fluxes in major European forest biomes. Global Change Biol., 11, 2211-2233.

Papale D, Reichstein M, Aubinet M et al. (2006) Towards a standardized processing of net ecosystem exchange measured with eddy covariance technique: algorithms and uncertainty estimation. Biogeosciences, 3, 571-583.

Peel MC, Finlayson BL, McMahon TA (2007) Updated world map of the Köppen–Geiger climate classification. Hydrol. Earth Syst. Sci., 11, 1633-1644.

Piao SL, Friedlingstein P, Peylin P, Reichstein M, Luyssaert S, Margolis H, Fang JY, Barr AG, Chen A, Grelle A, Hollinger DY, Laurila T, Lindroth A, Richardson A, Vesala T (2008) Net carbon dioxide losses of northern ecosystems in response to autumn warming. Nature 451, 49-53.

Purves DW, Pacala S (2008) Predictive models of forest dynamics. Science, 320, 1452-1453.

Ricciuto DM, King AW, Gu L, Post WM (2008) Estimates of terrestrial carbon cycle model parameters by assimilation of FLUXNET data: Do parameter variations cause bias in regional flux estimates? Eos Trans. AGU, 89(53), Fall Meet. Suppl., Abstract B54A-03.

Richardson AD and Hollinger DY. 2007. A method to estimate the additional uncertainty in gap-filled NEE resulting from long gaps in the CO2 flux record. Agricultural and Forest Meteorology, 147, 199-208.

Riley WJ, Still CJ, Torn MS, Berry JA (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, 1095, doi:10.1029/2002GB001878.

Schaefer K, Zhang T, Tans P, Stöckli R (2007) Temperature anomaly reemergence in seasonally frozen soils. J. Geophys. Res., 112, D20102, doi:10.1029/2007JD008630.

Schaefer K, Collatz GJ, Tans P et al. (2008) Combined Simple Biosphere/Carnegie-Ames-Stanford Approach terrestrial carbon cycle model, J. Geophys. Res., 113, G03034, doi:10.1029/2007JG000603.

Schaefer K, Zhang T, Slater AG, Lu L, Etringer A, Baker I (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.

Schwalm CR, Black TA, Amiro BD et al. (2006) Photosynthetic light use efficiency of three biomes across an east–west continental-scale transect in Canada. Agric. Forest Meteorol., 140, 269-286.

Schwalm CR, Williams CA, Schaefer KS et al. (2009) Assimilation exceeds respiration sensitivity to drought: A FLUXNET synthesis. Global Change Biology, doi: 10.1111/j.1365-2486.2009.01991.x

Seager R (2007a) The turn of the century drought across North America: global context, dynamics and past analogues. Journal of Climate, 20, 5527-5552.

Seager R, Ting MF, Held IM (2007b) Model projections of an imminent transition to a more arid climate in Southwestern North America, Science, 316, 1181-1184.

Seager R, Tzanova A, Nakamura J (2009) Drought in the Southeastern United States: Causes, variability over the last millennium and the potential for future hydroclimate change, Journal of Climate, 22, 5021-5045.

Sheffield J, Wood EF (2008) Projected changes in drought occurrence under future global warming from multi-model, multi-scenario, IPCC AR4 simulations. Climate Dynamics, 13, 79-105.

Siqueira MB, Katul GG, Sampson DA, Stoy PC, Juang J-Y, McCarthy HR, Oren R (2006) Multiscale model intercomparisons of CO_2 and H_2O exchange rates in a maturing southeastern US pine forest. Global Change Biology, 12, 1189-1207.

Sitch S, Smith B, Prentice IC, Arneth A, Bondeau A, Cramer W, Kaplan JO, Levis S, Lucht W, Sykes MT, Thonicke K, Venevsky S (2003) Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. Global Change Biology, 9, 161-185.

Sitch S, Huntingford C, Gedney N et al. (2008) Evaluation of the terrestrial carbon cycle future plant geography and climate-carbon cycle feedbacks using five Dynamic Global Vegetation Models (DGVMs). Glob. Change Biol, 14, 2015-2039.

Soja AJ, Tchebakova NM, French NHF, Flannigan MD, Shugart HH, Stocks BJ, Sukhinin AI, Varfenova EI, Chapin FS, Stackhouse PW (2007). Climate induced boreal forest change: Predictions versus current observations. Global and Planetary Change, 56, 274-296.

St-Hilaire F, Wu J, Roulet NT, Frolking S, Lafleur PM, Humphreys ER, Arora V (2008) McGill Wetland Model: evaluation of a peatland carbon simulator developed for global assessments. Biogeosciences Discussions, 5, 1689-1725.

Stoy P, Katul G, Siqueira M, Juang J, McCarthy H, Kim H, Oishi A, Oren R (2005) Variability in net ecosystem exchange from hourly to inter-annual time scales at adjacent pine and hardwood forests: a wavelet analysis. Tree Physiology, 25, 887-902.

Stoy P, Richardson A, Baldocchi D, Katul G, Stanovick J, Mahecha M, Reichstein M, Detto M, Law B, Wohlfahrt G, Arriga N, Campos J, McCaughey J, Montagnani L, Paw U, KT, Sevanto S, Williams M (2009) Biosphere-atmosphere exchange of CO_2 in relation to climate: a cross-biome analysis across multiple time scales. Biogeosciences, 6, 2297-2312.

Taylor JR (1996) An introduction to error analysis. University Science Books, Mill Valley, California, USA, 327 pp.

Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res., 106, 7183-7192.

Thomas CK, Law BE, Irvine J, Martin JG, Pettijohn JC, Davis KJ (2009) Seasonal hydrology explains interannual and seasonal variation in carbon and water exchange in a semi-arid mature ponderosa pine forest in Central Oregon. JGR Biogeosciences. In press.

Thornton, PE, Running SW, White MA (1997) Generating surfaces of daily meteorological variables over large regions of complex terrain. J. of Hydrology, 3-4, 214-251

Thornton PE, Running SW, Hunt ER (2005) Biome-BGC: Terrestrial Ecosystem Process Model, Version 4.1.1. Model product. Available on-line [http://daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. doi:10.3334/ORNLDAAC/805.

Tian, HQ, Chen G, Liu M, Zhang C, Sun G, Lu C, Xu X, Ren W, Pan P, Chappelka A (2009) Model Estimates of Ecosystem Net Primary Productivity, Evapotranspiration, and Water Use Efficiency in the Southern United States during 1895-2007. Forest Ecology and Management (in press).

Trenberth KE, Jones PD, Ambenje P et al. (2007) Observations: Surface and atmospheric climate change. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Solomon SD et al.). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Waring RH, Running SW (2007) Forest ecosystems - analysis at multiple scales (third edition). Academic Press, San Diego, 440 pp.

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

Williams CA, Hanan NP, Scholes RJ, Kutsch WL (2009) Complexity in water and carbon dioxide fluxes following rain pulses in an African savanna. Oecologia, 161, 469-480.

Williamson TB, Price DT, Beverly JL, Bothwell PM, Frenkel B, Park J, Patriquin MN (2008) Assessing potential biophysical and socioeconomic impacts of climate change on forest-based communities: a methodological case study. Nat. Resour. Can., Can. For. Serv., North. For. Cent., Edmonton, AB. Inf. Rep. NOR-X-415E.

Zhan XW, Xue YK, Collatz GJ (2003) An analytical approach for estimating CO_2 and heat fluxes over the Amazonian region. Ecol. Model. 162, 97-117.

Zhou XL, Peng CH, Dang QL, Sun JF, Wu HB, Hua D (2008) Simulating carbon exchange in Canadian Boreal forests I: model structure, validation, and sensitivity analysis. Ecological Modelling, 219, 287-299.

Table 1. Summary of Model Characteristics

Model Attribute	Model									
	AgroIBIS	BEPS	Biome-BGC	Can-IBIS	CN-CLASS	DLEM	DNDC	Ecosys	ED2	EDCM
Temporal Resolution	Half-hourly	Daily	Daily	Half-hourly	Half-hourly	Daily	Daily	Hourly	Half-hourly	Monthly
Vegetation Pools	4	4	7	3	4	6	3	9	9	8
Soil Pools	7	9	4	7	3	3	9	9	4	5
Soil Layers	11	3	1	7	3	2	10	15	9	10
Canopy Phenology	Prognostic	Semiprognostic	Prognostic	Prognostic	Prognostic	Semiprognostic	Prognostic	Prognostic	Prognostic	Prognostic
Nitrogen Cycle	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gross Primary Productivity (GPP)	Enzyme Kinetic Model	Enzyme Kinetic Model	Stomatal Conductance Model	Enzyme Kinetic Model	Enzyme Kinetic Model	Stomatal Conductance Model	Light Use Efficiency Model	Enzyme Kinetic Model	Enzyme Kinetic Model	Light Use Efficiency Model
Heterotrophic Respiration (HR)	First or Greater Order Model	Air Temperature Soil Temperature Precipitation Soil Moisture Evaporation Soil Carbon Soil Nitrogen	Soil Temperature Soil Moisture Soil Carbon	First or Greater Order Model	First or Greater Order Model	Decay Methane Air Temperature Soil Temperature Litter and Soil Carbon Soil Nitrogen Soil Moisture	Decay Methane Soil Temperature Precipitation Soil Moisture Soil Carbon Vegetation Carbon Soil Nitrogen	Decay Methane CO_2 Diffusion Dissolved Carbon Loss Soil Temperature Soil Moisture Surface Incident Shortwave Radiation Surface Incident Longwave Radiation Soil Carbon Vegetation Carbon Soil Nitrogen Leaf Nitrogen	Soil Temperature Soil Moisture Soil Carbon Soil Nitrogen	Soil Temperature Soil Moisture Soil Carbon Dissolved Carbon Loss Vegetation Carbon Soil Nitrogen
Autotrophic Respiration (AR)	Air Temperature Soil Temperature Precipitation Soil Moisture Surface Incident Shortwave Radiation Surface Incident Longwave Radiation Vegetation Carbon	Air Temperature GPP	Air Temperature Vegetation Carbon Leaf Nitrogen	Air Temperature Soil Temperature Precipitation Soil Moisture Surface Incident Shortwave Radiation Surface Incident Longwave Radiation Vegetation Carbon	Fraction of Instantaneous GPP	Air Temperature Vegetation Carbon Leaf Nitrogen GPP	Soil Temperature	Air Temperature Soil Temperature Vegetation Carbon Leaf Nitrogen	Air Temperature Soil Temperature Vegetation Carbon Leaf Nitrogen GPP	Proportional to Growth
Ecosystem Respiration (R)	AR + HR	AR + HR	Air Temperature Soil Temperature Soil Moisture Soil Carbon Vegetation Carbon LAI	AR + HR	AR + HR	AR + HR	AR + HR	AR + HR	AR + HR	AR + HR
Net Primary Production (NPP)	GPP - AR	GPP - AR	Surface Incident Shortwave Radiation Vapor Pressure Deficit CO_2 Vegetation Carbon Leaf Nitrogen LAI	GPP - AR	Fraction of Instantaneous GPP	GPP - AR	Air Temperature Precipitation Soil Moisture Potential Evaporation Vegetation Carbon Soil Nitrogen Leaf Nitrogen fPAR	GPP - AR	GPP - AR	Air Temperature Precipitation Soil Carbon Soil Nitrogen Soil Moisture Vegetation Carbon Leaf Nitrogen LAI
Net Ecosystem Exchange (NEE)	NPP - HR	NPP - HR	Soil Temperature Soil Moisture Surface Incident Shortwave Radiation Vapor Pressure Deficit	NPP - HR	GPP - R	NPP - HR	NPP - HR	GPP - R	NPP - HR	NPP - HR

Table 1. (continued)

Model Attribute	Model									
	AgroIBIS	BEPS	Biome-BGC	Can-IBIS	CN-CLASS	DLEM	DNDC	Ecosys	ED2	EDCM
Biomes Simulated	Croplands	6	8	10	9	10	Croplands	10	6	6
Sites Simulated	5	10	36	27	31	33	5	39	25	10
Months Simulated	192	945	2001	1978	2082	2246	192	2450	1684	658
Source	*Kucharik and Twine* [2007]	*Liu et al.* [1999]	*Thornton et al.* [2005]	*Williamson et al.* [2008]	*Arain et al.* [2006]	*Tian et al.* [2010]	*Li et al.* [2010]	*Grant et al.* [2005]	*Medvigy et al.* [2009]	*Liu et al.* [2003]

Model Attribute	Model										
	EPIC	ISOLSM	LoTEC	LPJ	ORCHIDEE	SiB3	SiBCASA	SiBcrop	SSiB2	TECO	Triplex-FLUX
Temporal Resolution	Daily	Half-hourly	Half-hourly	Daily	Half-hourly	Half-hourly	10 min	Half-hourly	Half-hourly	Hourly	Half-hourly
Vegetation Pools	3	0	4	3	8	0	8	4	0	3	0
Soil Pools	0	1	5	2	8	0	5	1	0	5	0
Soil Layers	15	0	14	2	0	10	15	10	3	10	0
Canopy Phenology	Prognostic	Prescribed	Prognostic	Prognostic	Prognostic	Prescribed	Prescribed	Prognostic	Prescribed	Prognostic	Prescribed
Nitrogen Cycle	Yes	No	No	No	No	Yes	No	Yes	No	No	No
Gross Primary Productivity (GPP)	Nil	Stomatal Conductance Model	Enzyme Kinetic Model	Stomatal Conductance Model	Enzyme Kinetic Model	Enzyme Kinetic Model	Enzyme Kinetic Model	Enzyme Kinetic Model	Stomatal Conductance Model	Stomatal Conductance Model	Stomatal Conductance Model
Heterotrophic Respiration (HR)	CO_2 Diffusion Dissolved Carbon Loss Air Temperature Soil Temperature Precipitation Soil Moisture	First or Greater Order Model	Soil Temperature Soil Moisture Soil Carbon	Soil Temperature Soil Moisture Soil Carbon	Soil Temperature Soil Moisture Soil Carbon	Zero-order Model	Soil Temperature Soil Moisture Soil Carbon	Soil Temperature Soil Carbon	Zero-order Model	First or Greater Order Model	First or Greater Order Model
Autotrophic Respiration (AR)	Nil	Fraction of Instantaneous GPP	Air Temperature Soil Temperature Soil Moisture Vegetation Carbon GPP	Air Temperature Soil Moisture Vegetation Carbon	Air Temperature Vegetation Carbon	Fraction of Instantaneous GPP	Air Temperature Soil Moisture Vegetation Carbon	Air Temperature Vegetation Carbon GPP	Air Temperature Soil Moisture Surface Incident Shortwave Radiation Relative Humidity LAI fPAR CO_2	Air Temperature Vegetation Carbon	Fraction of Annual GPP
Ecosystem Respiration (R)	AR + HR	AR + HR	AR + HR	AR + HR	AR + HR	Forced Annual Balance	AR + HR	Forced Annual Balance	Forced Annual Balance	AR + HR	AR + HR
Net Primary Production (NPP)	Light Use Efficiency Model	Nil	GPP - AR	GPP - AR	GPP - AR	GPP - AR	Air Temperature Soil Moisture CO_2 Relative Humidity	GPP - AR	GPP - AR	GPP - AR	Fraction of Instantaneous GPP
Net Ecosystem Exchange (NEE)	NPP - HR	GPP - R	NPP - HR	NPP - HR	GPP - R	GPP - R	GPP - R	GPP - R	GPP - R	GPP - R	GPP - R
Biomes Simulated	Croplands	5	6	9	10	10	10	Croplands	10	10	3
Sites Simulated	U.S.-Ne3	9	10	29	35	31	35	5	44	35	7
Months Simulated	48	909	825	2126	2332	2258	2402	192	2800	2414	291
Source	*Causarano et al.* [2007]	*Riley et al.* [2002]	*Hanson et al.* [2004]	*Sitch et al.* [2003]	*Krinner et al.* [2005]	*Baker et al.* [2008]	*Schaefer et al.* [2009]	*Lokupitiya et al.* [2009]	*Zhan et al.* [2003]	*Weng and Luo* [2008]	*Zhou et al.* [2008]

Table 2. Summary of Site Characteristics[a]

Site ID	Name	Priority	Country	Latitude	Longitude	Elevation (m a.s.l.)	IGBP Class	Köppen-Geiger Climate Classification
CA-Ca1	British Columbia - Campbell River - Mature Forest Site	1	Canada	49.87	−125.33	300	ENF	Maritime temperate
CA-Ca2	British Columbia - Campbell River - Clearcut Site	2	Canada	49.87	−125.29	180	ENF	Maritime temperate
CA-Ca3	British Columbia - Campbell River - Young Plantation Site	2	Canada	49.53	−124.90	165	ENF	Maritime temperate
CA-Gro	Ontario - Groundhog River - Mature Boreal Mixed Wood	1	Canada	48.22	−82.16	300	MF	Warm summer continental
CA-Let	Lethbridge	1	Canada	49.71	−112.94	960	GRA	Warm summer continental
CA-Mer	Eastern Peatland - Mer Bleue	1	Canada	45.41	−75.52	70	WET	Warm summer continental
CA-Oas	Sask. - SSA Old Aspen	1	Canada	53.63	−106.20	530	DBF	Continental subarctic
CA-Obs	Sask. - SSA Old Black Spruce	1	Canada	53.99	−105.12	629	ENF	Continental subarctic
CA-Ojp	Sask. - SSA Old Jack Pine	1	Canada	53.92	−104.69	579	ENF	Continental subarctic
CA-Qfo	Quebec Mature Boreal Forest Site	1	Canada	49.69	−74.34	382	ENF	Continental subarctic
CA-SJ1	Sask. - 1994 Harvested Jack Pine	2	Canada	53.91	−104.66	580	ENF	Continental subarctic
CA-SJ2	Sask. - 2002 Harvested Jack Pine	2	Canada	53.94	−104.65	518	ENF	Continental subarctic
CA-SJ3	Sask. - SSA 1975 Harvested Young Jack Pine	2	Canada	53.88	−104.64	511	ENF	Continental subarctic
CA-TP3	Ontario - Turkey Point Middle-aged White Pine	2	Canada	42.71	−80.35	219	ENF	Warm summer continental
CA-TP4	Ontario - Turkey Point Mature White Pine	1	Canada	42.71	−80.36	219	ENF	Warm summer continental
CA-WP1	Western Peatland - LaBiche-Black Spruce/Larch Fen	1	Canada	54.95	−112.47	540	MF	Continental subarctic
U.S.-ARM	OK - ARM Southern Great Plains Site - Lamont	1	USA	36.61	−97.49	310	CRO	Humid subtropical
U.S.-Atq	AK - Atqasuk	1	USA	70.47	−157.41	16	WET	Tundra
U.S.-Brw	AK - Barrow	1	USA	71.32	−156.63	1	WET	Tundra
U.S.-Dk2	NC - Duke Forest - Hardwoods	1	USA	35.97	−79.10	160	DBF	Humid subtropical
U.S.-Dk3	NC - Duke Forest - Loblolly Pine	1	USA	35.98	−79.09	163	ENF	Humid subtropical
U.S.-Ha1	MA - Harvard Forest EMS Tower (HFR1)	1	USA	42.54	−72.17	303	DBF	Warm summer continental
U.S.-Ho1	ME - Howland Forest (Main Tower)	1	USA	45.20	−68.74	60	ENF	Warm summer continental
U.S.-IB1	IL - Fermi National Accelerator Laboratory - Batavia (Agricultural Site)	1	USA	41.86	−88.22	227	CRO	Hot summer continental
U.S.-IB2	IL - Fermi National Accelerator Laboratory - Batavia (Prairie Site)	1	USA	41.84	−88.24	227	GRA	Hot summer continental
U.S.-Los	WI - Lost Creek	1	USA	46.08	−89.98	480	CSH	Warm summer continental
U.S.-MMS	IN - Morgan Monroe State Forest	1	USA	39.32	−86.41	275	DBF	Humid subtropical
U.S.-MOz	MO - Missouri Ozark Site	1	USA	38.74	−92.20	219	DBF	Humid subtropical
U.S.-Me2	OR - Metolius - Intermediate Aged Ponderosa Pine	1	USA	44.45	−121.56	1253	ENF	Dry-summer subtropical
U.S.-Me3		2	USA	44.32	−121.61	1005	ENF	

Table 2. (continued)

Site ID	Name	Priority	Country	Latitude	Longitude	Elevation (m a.s.l.)	IGBP Class	Köppen-Geiger Climate Classification
	OR - Metolius - Second Young Aged Pine							Dry-summer subtropical
U.S.-Me4	OR - Metolius - Old Aged Ponderosa Pine	2	USA	44.50	−121.62	915	ENF	Dry-summer subtropical
U.S.-Me5	OR - Metolius - First Young Aged Pine	2	USA	44.44	−121.57	1183	ENF	Dry-summer subtropical
U.S.-NR1	CO - Niwot Ridge Forest (LTER NWT1)	1	USA	40.03	−105.55	3050	ENF	Continental subarctic
U.S.-Ne1	NE - Mead - Irrigated Continuous Maize Site	1	USA	41.17	−96.48	361	CRO	Hot summer continental
U.S.-Ne2	NE - Mead - Irrigated Maize - Soybean Rotation Site	1	USA	41.16	−96.47	361	CRO	Hot summer continental
U.S.-Ne3	NE - Mead - Rainfed Maize - Soybean Rotation Site	1	USA	41.18	−96.44	361	CRO	Hot summer continental
U.S.-PFa	WI - Park Falls/WLEF	1	USA	45.95	−90.27	485	MF	Warm summer continental
U.S.-SO2	CA - Sky Oaks - Old Stand	1	USA	33.37	−116.62	1392	CSH	Dry-summer subtropical
U.S.-Shd	OK - Shidler- Oklahoma	1	USA	36.93	−96.68	350	GRA	Humid subtropical
U.S.-Syv	MI - Sylvania Wilderness Area	1	USA	46.24	−89.35	540	MF	Warm summer continental
U.S.-Ton	CA - Tonzi Ranch	1	USA	38.43	−120.97	177	WSA	Dry-summer subtropical
U.S.-UMB	MI - University of Michigan Biological Station	1	USA	45.56	−84.71	234	DBF	Warm summer continental
U.S.-Var	CA - Vaira Ranch - Ione	1	USA	38.41	−120.95	129	GRA	Dry-summer subtropical
U.S.-WCr	WI - Willow Creek	1	USA	45.81	−90.08	520	DBF	Warm summer continental

Site ID	Annual NEE (g C m^2−)	Annual NEE Error (g C m^2−)	Daytime Data Coverage (%)	Nighttime Data Coverage (%)	LAI	Annual Precipitation (mm)	Mean Annual Air Temperature (°C)	Measurement Period	Biome	Source
CA-Ca1	−244.3	61.1	99	26	6.1	1256	8.7	1998–2006	ENFT	*Schwalm et al.* [2007]
CA-Ca2	571.7	31.5	96	23	4.4	1222	8.8	2001–2006	ENFT	*Schwalm et al.* [2007]
CA-Ca3	91.2	37.9	91	27	2.2	1554	9.5	2002–2006	ENFT	*Schwalm et al.* [2007]
CA-Gro	−36.5	33.5	93	34	4.1	427	3.3	2004–2006	MF	*McCaughey et al.* [2006]
CA-Let	−132.9	14.3	96	46	0.7	335	6.5	1997–2006	GRA	*Flanagan et al.* [2002]
CA-Mer	−68.5	21.6	79	56	1.3	935	6.2	1999–2006	WET	*Lafleur et al.* [2003]
CA-Oas	−158.0	28.5	94	56	3.8	460	2.3	1997–2006	DBF	*Barr et al.* [2004]
CA-Obs	−56.3	16.1	89	45	5.6	470	1.6	2000–2006	ENFB	*Griffis et al.* [2003]
CA-Ojp	−29.9	16.6	91	50	3.4	461	1.5	2000–2006	ENFB	*Griffis et al.* [2003]
CA-Qfo	−13.7	21.0	93	40	4	819	2.7	2004–2006	ENFB	*Bergeron et al.* [2007]
CA-SJ1	28.0	15.3	87	31	0.8	344	0.6	2002–2005	ENFB	*Zha et al.* [2009]
CA-SJ2	117.0	6.1	89	47	1.3	537	0.1	2003–2006	ENFB	*Zha et al.* [2009]
CA-SJ3	−82.0	17.7	92	34	4.3	694	0.8	2004–2005	ENFB	*Zha et al.* [2009]
CA-TP4	−133.2	29.5	95	43	3.5	959	8.6	2002–2007	ENFT	*Peichl and Arain* [2007]
CA-WP1	−195.8	16.4	96	50	2.7	481	1.7	2003–2007	WET	*Syed et al.* [2006]

Table 2. (continued)

Site ID	Annual NEE (g C m^2−)	Annual NEE Error (g C m^2−)	Daytime Data Coverage (%)	Nighttime Data Coverage (%)	LAI	Annual Precipitation (mm)	Mean Annual Air Temperature (°C)	Measurement Period	Biome	Source
U.S.-ARM	−128.4	74.4	89	36	3.1	629	15.6	2000–2006	CRO	*Fischer et al.* [2007]
CA-TP4	−133.2	29.5	95	43	3.5	959	8.6	2002–2007	ENFT	*Peichl and Arain* [2007]
U.S.-Atq	−12.8	-	50	22	1.1	118	−10.6	1999–2006	TUN	*Oberbauer et al.* [2007]
U.S.-Brw	−72.0	-	49	29	1.5	108	−10.9	1999–2002	TUN	*Harazono et al.* [2003]
U.S.-Dk2	−718.1	-	48	1	7	1091	15.1	2003–2005	DBF	*Siqueira et al.* [2006]
U.S.-Dk3	−350.0	139.0	75	37	5.6	1126	14.7	1998–2005	ENFT	*Siqueira et al.* [2006]
U.S.-Ha1	−217.4	65.9	78	34	3.38	1122	7.9	1991–2006	DBF	*Urbanski et al.* [2007]
U.S.-Ho1	−223.0	33.4	70	47	5.2	818	6.6	1996–2004	ENFT	*Richardson et al.* [2009]
U.S.-IB1	−269.0	31.3	92	46	1.29	718	10.1	2005–2007	CRO	*Post et al.* [2004]
U.S.-IB2	−86.0	42.0	80	49	5.38	818	10.4	2004–2007	GRA	*Post et al.* [2004]
U.S.-Los	−78.0	19.2	82	54	4.24	666	3.8	2000–2006	WET	*Sulman et al.* [2009]
U.S.-MMS	−346.1	66.3	97	46	4.9	1109	12.4	1999–2006	DBF	*Schmid et al.* [2000]
U.S.-MOz	−305.7	48.9	94	33	3.91	730	13.3	2004–2007	DBF	*Gu et al.* [2006]
U.S.-Me2	−536.0	65.8	63	46	3.62	434	7.6	2002–2007	ENFT	*Thomas et al.* [2009]
U.S.-Me3	−198.0	32.7	83	28	0.52	423	8.5	2004–2005	ENFT	*Vickers et al.* [2009]
U.S.-Me4	−612.3	-	55	41	2.1	641	8.3	1996–2000	ENFT	*Irvine et al.* [2004]
U.S.-Me5	−206.0	10.6	97	48	1.1	350	7.6	1999–2002	ENFT	*Irvine et al.* [2004]
U.S.-NR1	−37.2	27.0	89	44	4.2	663	2.5	1998–2007	ENFT	*Bradford et al.* [2008]
U.S.-Ne1	−424.0	41.8	93	42	6.5	832	11.1	2001–2006	CRO	*Verma et al.* [2005]
U.S.-Ne2	−382.0	41.8	96	51	6.5	823	10.8	2001–2006	CRO	*Verma et al.* [2005]
U.S.-Ne3	−258.0	43.3	94	55	6.2	627	10.9	2001–2006	CRO	*Verma et al.* [2005]
U.S.-PFa	45.0	41.1	85	30	4.05	736	5.1	1997–2005	MF	*Davis et al.* [2003]
U.S.-SO2	22.4	25.6	87	30	3	695	13.8	1998–2006	SHR	*Luo et al.* [2007]
U.S.-Shd	−75.5	22.0	96	49	5.9	1179	14.8	1997–2001	GRA	*Suyker et al.* [2003]
U.S.-Syv	48.5	34.7	53	51	4.1	700	4.4	2001–2006	MF	*Desai et al.* [2005]
U.S.-Ton	−67.8	52.0	77	25	0.6	549	16.4	2001–2007	WSA	*Ma et al.* [2007]
U.S.-UMB	−132.0	42.4	86	39	4.23	629	7.4	1998–2006	DBF	*Schmid et al.* [2003]
U.S.-Var	7.3	110.6	80	22	2.4	563	15.9	2001–2007	GRA	*Ma et al.* [2007]
U.S.-WCr	−222.6	54.1	48	55	5.36	712	5.3	1998–2006	DBF	*Cook et al.* [2004a]

[a]Sources: IGBP classification, *Loveland et al.* [2001]; Köppen-Geiger, *Peel et al.* [2007]; LAI for USA sites, http://public.ornl.gov/ameriflux/; LAI for Canadian sites, *Chen et al.* [2006] and *Schwalm et al.* [2006]. Annual precipitation and mean annual air temperature are measurement period averages of meteorological inputs used to drive model simulations. NEE values show yearly integrals and associated error: one standard deviation based on uncertainty due to random noise and the friction velocity threshold aggregated to yearly values and summed in quadrature [*Barr et al.*, 2009]. Data coverages are percentages of half-hourly NEE measurements that satisfied quality control standards (friction velocity threshold) for day- and nighttime separately. Priority: (1) Primary sites with complete (includes ancillary and biological data templates) records; (2) Secondary chronosequence sites. Standardized Precipitation Index available only for Priority 1 sites excluding US-Atq, US-Brw, US-Dk2, US-IB1, and US-Shd. CA-TP3, US-Atq, US-Brw, US-Dk2, and US-Me4 sites used postprocessing protocol from the La Thuile and Asilomar FLUXNET Synthesis data set (http://www.fluxdata.org/) [*Moffat et al.*, 2007; *Papale et al.*, 2006] and lack NEE uncertainties. Biome is combination of IGBP class and Köppen-Geiger climate. US-Atq and US-Brw-Arctic wetlands classified as tundra biome CA-WP1, treed fen (IGBP mixed forest) grouped with wetlands biome US-Los, shrub wetland site (IGBP closed shrublands) grouped with wetlands biome US-SO2, closed shrublands grouped with shrublands (open or closed) biome. IGBP class and biome codes: CRO, croplands; CSH, closed shrublands; GRA, grasslands, ENF, evergreen needleleaf forest; ENFB, evergreen needleleaf forest-boreal zone; ENFT, evergreen needleleaf forest-temperate zone; DBF, deciduous broadleaf forest; MF, mixed (deciduous/evergreen) forest; WSA, woody savanna; SHR, shrublands; TUN, tundra; WET, wetlands.

Table 3. Model structural and site history predictors used to classify Taylor skill with regression tree analysis. Taylor skill (*S*; Eq 3) was divided into three classes using terciles. Model structural predictants are from the Metadata for Forward (Ecosystem) Model Intercomparison survey collated by the NACP Site Synthesis (http://daac.ornl.gov/SURVEY8/survey_results.shtml). Site history data are from http://public.ornl.gov/ameriflux/, www.fluxnet.org, and Schwalm et al. (2006).

Predictor	Value
Model temporal resolution	Daily, half-hourly or less, hourly, monthly
Canopy	Prognostic, semi-prognostic, prescribed. Prescribed canopy from remote sensing, semi-prognostic has some prescribed input into canopy leaf biomass but calculates phenology with other prognostic variables.
Number of vegetation pools	Number of pools, both dynamic and static
Number of soil pools	Number of pools, both dynamic and static
Number of soil layers	Number of layers
Nitrogen	True if the model has a nitrogen cycle; otherwise false.
Steady state	True if the simulated long-term *NEE* integral approaches zero; otherwise false.
Autotrophic respiration (*AR*)	Fraction of annual *GPP*, fraction of instantaneous *GPP*, explicitly calculated, nil, proportional to growth
Ecosystem respiration (*R*)	*AR* + *HR*, explicitly calculated, forced annual balance
Gross primary productivity (*GPP*)	Enzyme kinetic model, light use efficiency model, nil, stomatal conductance model
Heterotrophic respiration (*HR*)	Explicitly calculated, first or greater order model, zero-order model
Net ecosystem exchange (*NEE*)	Explicitly calculated, *GPP* - *R*, *NPP* - *HR*
Net primary productivity (*NPP*)	Explicitly calculated, fraction of instantaneous *GPP*, *GPP* - *AR*, light use efficiency model
Overall model complexity	Low, average, high. Values correspond to terciles of the total amount of first-order functional arguments for the following model-generated variables/outputs: *AR*, canopy leaf biomass, *R*, evapotranspiration, *GPP*, *HR*, *NEE*, *NPP*, soil moisture.
Site history	True if the below listed management activity or disturbance or event occurred on site; otherwise false. Grazed, fertilized, fire, harvest, herbicide, insects and pathogens, irrigation, natural regeneration, pesticide, planted, residue management, thinning
Stand age class	Young, intermediate, nil, mature, multi-cohort. Values based on stand age in forested sites; stands without a clear dominant stratum are treated as multi-cohort; non-forest types have nil.

Table 4. Normalized mean average error (*NMAE*) by climatic season, drought level, and biome. Drought level was based on monthly values of 3-month Standard Precipitation Index (SPI): dry value of < -0.8; wet > +0.8. Otherwise normal conditions existed.

	Climatic season				Drought level			
Biome[#]	Winter	Spring	Summer	Fall	Dry	Normal	Wet	Overall
CRO	1.90	4.64	-0.79	12.73	-1.43	-1.54	-1.59	-1.55
DBF	0.81	93.7	-0.52	-2.14	-1.01	-1.00	-0.95	-1.00
ENFB	1.52	-1.12	-0.69	-1.92	-0.87	-1.15	-3.43	-1.12
ENFT	-6.34	-0.66	-0.50	-0.76	-0.63	-0.72	-0.63	-0.68
GRA	-25.46	-0.84	-1.11	5.19	-1.52	-1.32	-3.07	-1.51
MF	1.10	-7.48	-0.47	57.70	-1.42	-1.04	-1.15	-1.12
SHR	-87.37	-1.37	-3.03	-140.17	-1.82	-2.18	-41.13	-2.88
TUN	-1.43	-11.07	-20.63	6.38	19.22	-24.06	-1.81	-20.15
WET	1.80	-5.07	-0.59	-4.72	-1.21	-1.20	-2.38	-1.27
WSA	-2.73	-0.75	-1.47	10.56	-1.39	-1.32	-1.51	-1.37
Overall	2.42	-1.35	-0.61	-1.94	-0.97	-1.01	-1.00	-1.00

[#] Biome codes: CRO = cropland, GRA = grassland, ENFB = evergreen needleleaf forest – boreal zone, ENFT = evergreen needleleaf forest – temperate zone, DBF = deciduous broadleaf forest, MF = mixed (deciduous/evergreen) forest, WSA = woody savanna, SHR = shrubland, TUN = tundra, WET = wetland.

Table 5. Reduced χ^2 statistic by climatic season, drought level, and biome. Drought level was based on monthly values of 3-month Standard Precipitation Index (SPI): dry value of < -0.8; wet > +0.8. Otherwise normal conditions existed.

	Climatic season				Drought level			
Biome[#]	Winter	Spring	Summer	Fall	Dry	Normal	Wet	Overall
CRO	3.22	10.66	39.75	49.71	14.43	23.54	32.75	25.8
DBF	5.29	10.74	8.77	4.55	5.58	7.86	8.67	7.34
ENFB	21.25	17.75	4.98	6.61	11.64	12.02	18.51	12.61
ENFT	4.39	7.90	3.27	2.26	4.71	4.29	4.60	4.45
GRA	10.89	11.38	25.01	17.22	13.97	10.99	26.01	16.07
MF	3.74	4.67	2.05	2.02	2.92	3.24	2.98	3.08
SHR	13.34	27.98	12.52	11.2	9.26	21.31	10.31	16.26
WET	23.65	27.27	11.74	7.54	21.51	17.36	12.91	17.47
WSA	0.61	5.81	11.88	3.39	6.73	4.64	6.35	5.37
Overall	8.18	11.95	11.27	9.45	8.10	9.98	12.72	10.26

[#] Biome codes: CRO = cropland, GRA = grassland, ENFB = evergreen needleleaf forest – boreal zone, ENFT = evergreen needleleaf forest – temperate zone, DBF = deciduous broadleaf forest, MF = mixed (deciduous/evergreen) forest, WSA = woody savanna, SHR = shrubland, WET = wetland.

Figure 1.

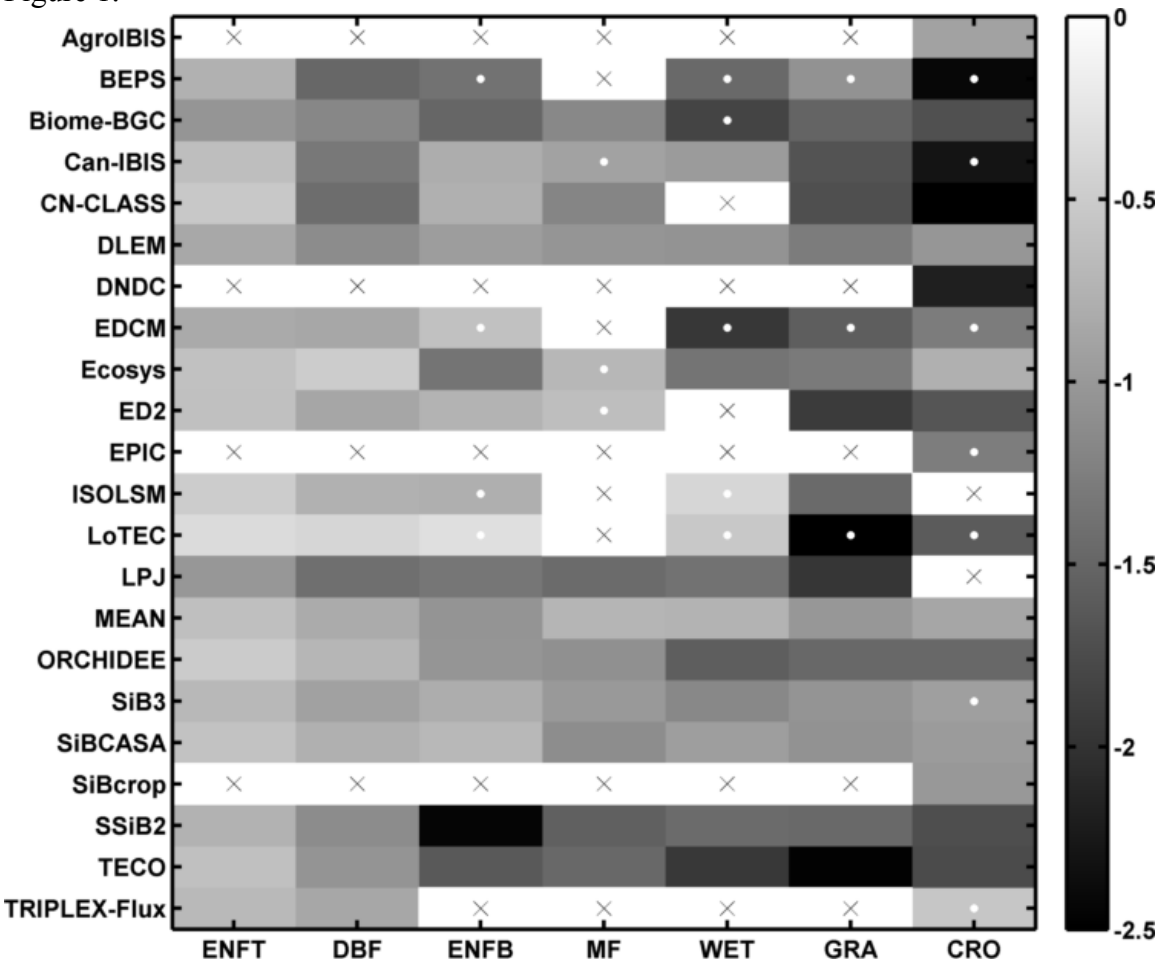

Figure 1. Normalized mean average error (*NMAE*) by biome for each model. Biomes in ascending order based on model-specific *NMAE*; biomes on the left show better average agreement with observations. *NMAE* is normalized by mean observed flux. Across all sites, seasons, and drought levels within a given biome this value is negative ($NEE < 0$), indicating a sink. *NMAE* values closer to zero coincide with a higher degree of model-data agreement. Woody savannahs and shrublands not shown: only one site each. Tundra ($n = 2$ sites) has $NMAE < -10$ for all models. CN-CLASS croplands value is off-scale (= -8.98). Black cross: no observations; white circle: undersampled ($n < 100$ months).

Figure 2.

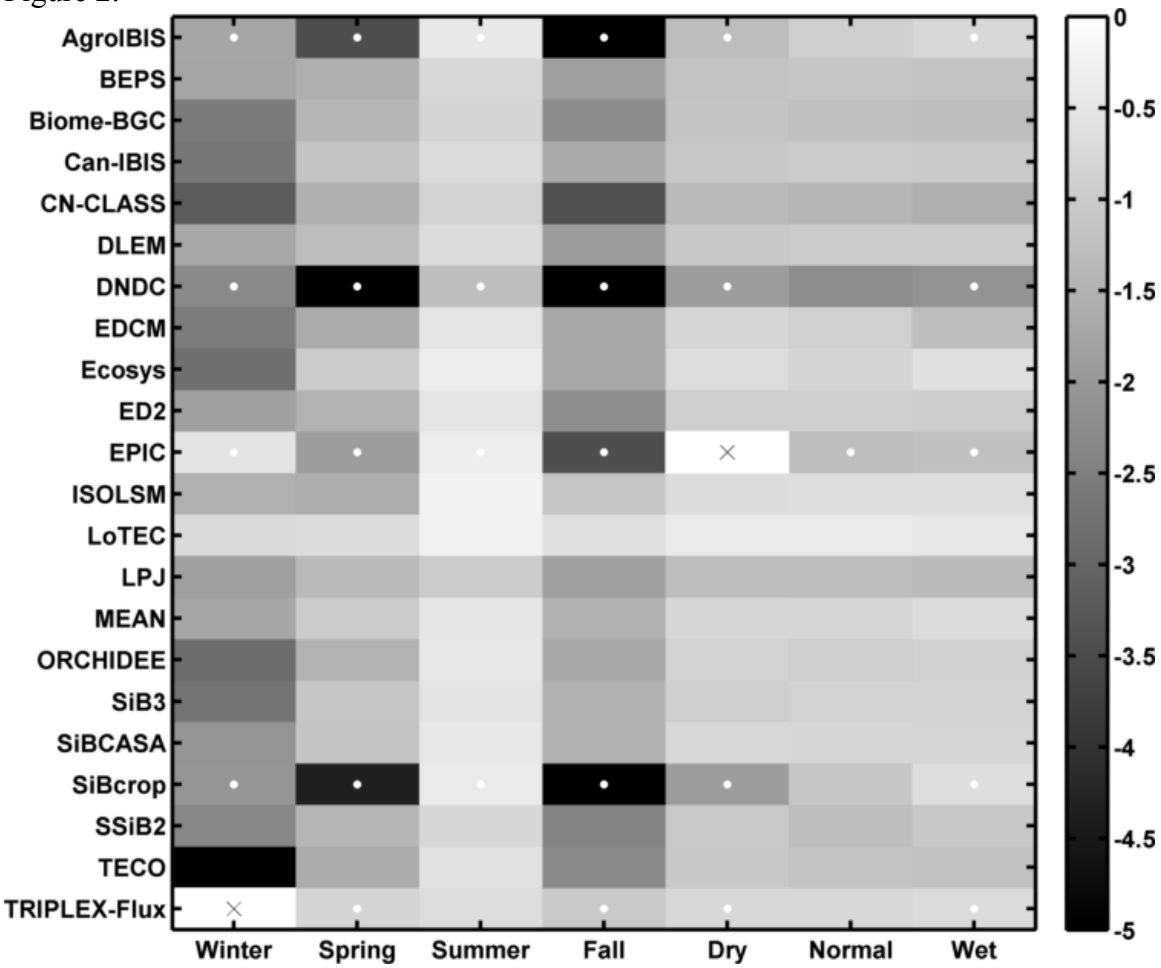

Figure 2. Normalized mean average error (*NMAE*) by climatic season and drought level. *NMAE* is normalized by mean observed flux such that most values are negative (*NEE* < 0), indicating a sink. Positive values, displayed on same color bar with opposite sign, indicate a source (*NEE* > 0). These occur in winter (all models) as well as spring and fall (all crop only models: AgroIBIS, DNDC, EPIC, SiBcrop). Off-scale values: AgroIBIS and SiBcrop in fall are -7.1 and -11.1 respectively. DNDC in fall and spring is -11.4 and -8.7 respectively. Black cross: no observations; white circle: undersampled (n < 100 months).

Figure 3.

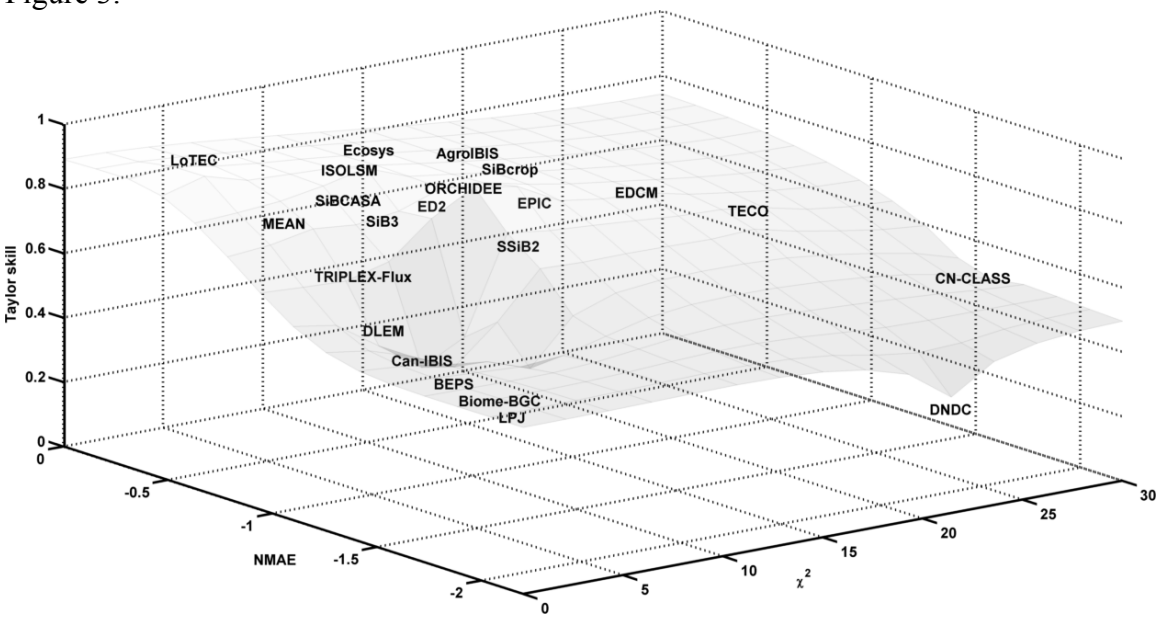

Figure 3. Model skill metrics for all 22 models. Skill metrics are Taylor skill (*S*; Eq. 3), normalized mean average error (*NMAE*), and reduced χ^2 statistic (χ^2). Better model-data agreement corresponds to the upper left with Taylor skill close to unity; *NMAE* and χ^2 closer to zero. Gray interpolated surface added and model names jittered to improve readability.

Figure 4.

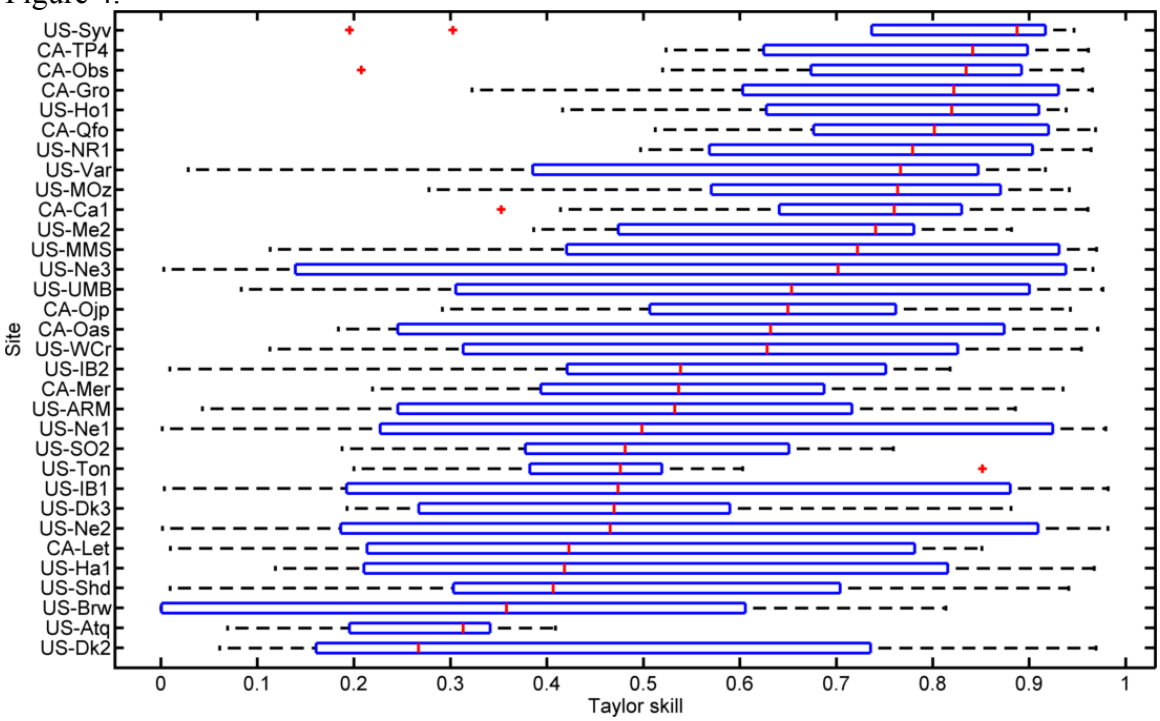

Figure 4. Boxplots of Taylor skill by site. Taylor skill (*S*; Eq. 3) is a single value summary of a Taylor diagram where unity indicates perfect agreement with observations. Panels show interquartile range (blue box), median (solid red line), range (whiskers), and outliers (red cross; values more than 1.5 x interquartile range from the median). Only sites (n = 32) simulated with at least 10 unique models, excluding the mean model ensemble and the assimilated LoTEC, shown. Sites sorted by median Taylor skill.

Figure 5

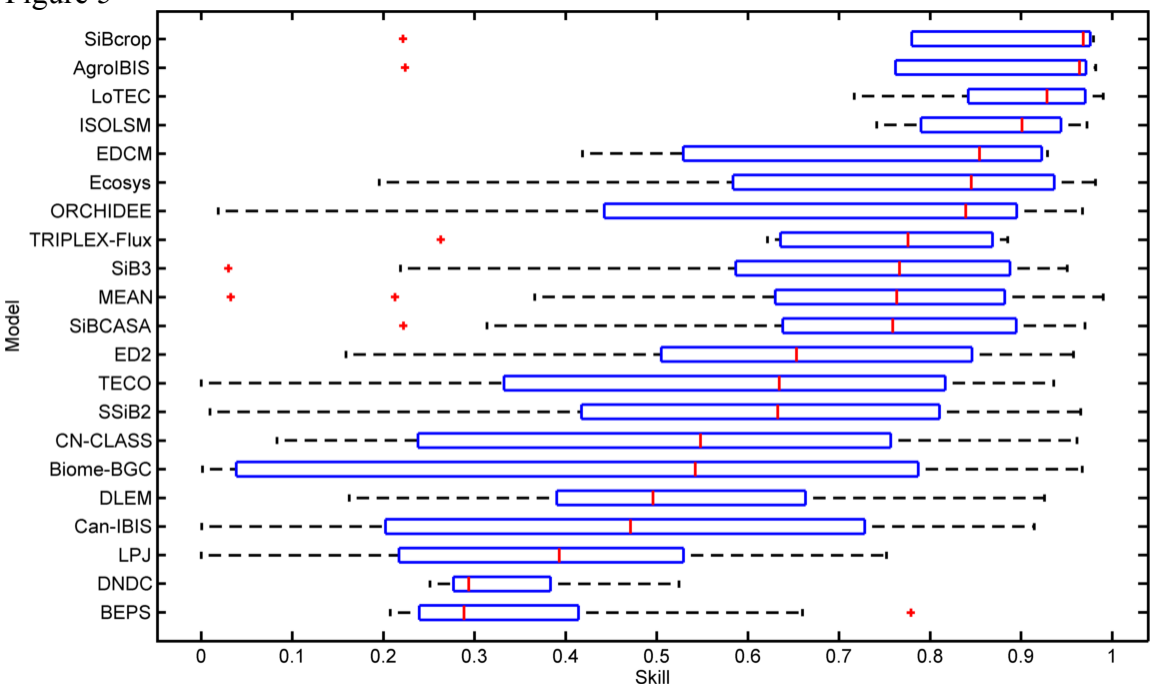

Figure 5. Boxplots of Taylor skill by model. Taylor skill (*S*; Eq. 3) is a single value summary of a Taylor diagram where unity indicates perfect agreement with observations. Panels show interquartile range (blue box), median (solid red line), range (whiskers), and outliers (red cross; values more than 1.5 x interquartile range from the median). Only models (n = 21) used on at least two sites shown. Sites sorted by median Taylor skill.

Figure 6.

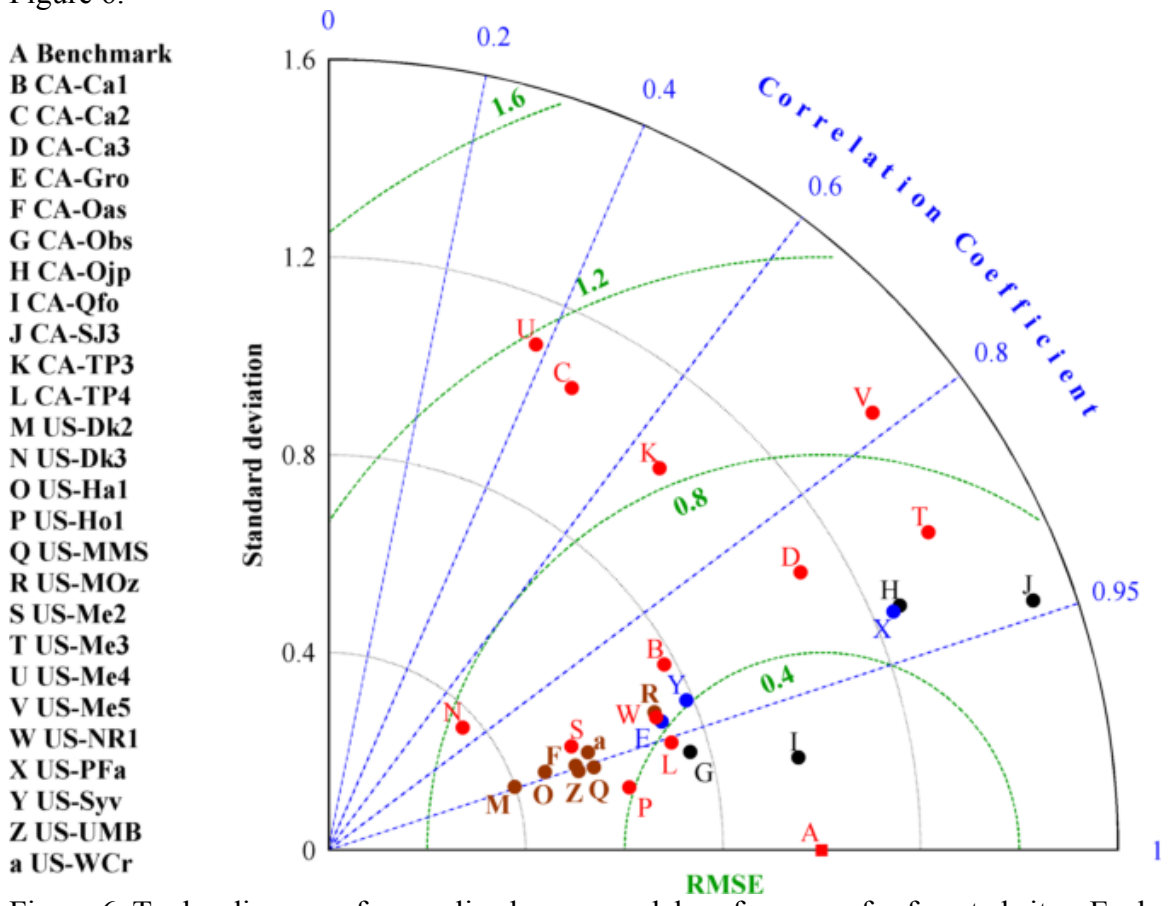

Figure 6. Taylor diagram of normalized mean model performance for forested sites. Each circle (n = 26 sites) is the site-specific mean model ensemble (MEAN). Benchmark (red square) corresponds to observed normalized monthly *NEE*; units of σ and *RMSE* are multiples of observed σ. Color coding of site letter and circles indicates biome: evergreen needleleaf forest – temperate zone (red), deciduous broadleaf forest (brown), mixed (deciduous/evergreen) forest (blue), evergreen needleleaf forest – boreal zone (black). Outlying sites (evergreen needleleaf forest – boreal zone) not shown: CA-SJ1 (ρ = 0.81, σ = 3.9, *RMSE* = 3.1) and CA-SJ2 (ρ = -0.67, σ = 4.3, *RMSE* = 5.1).

Figure 7.

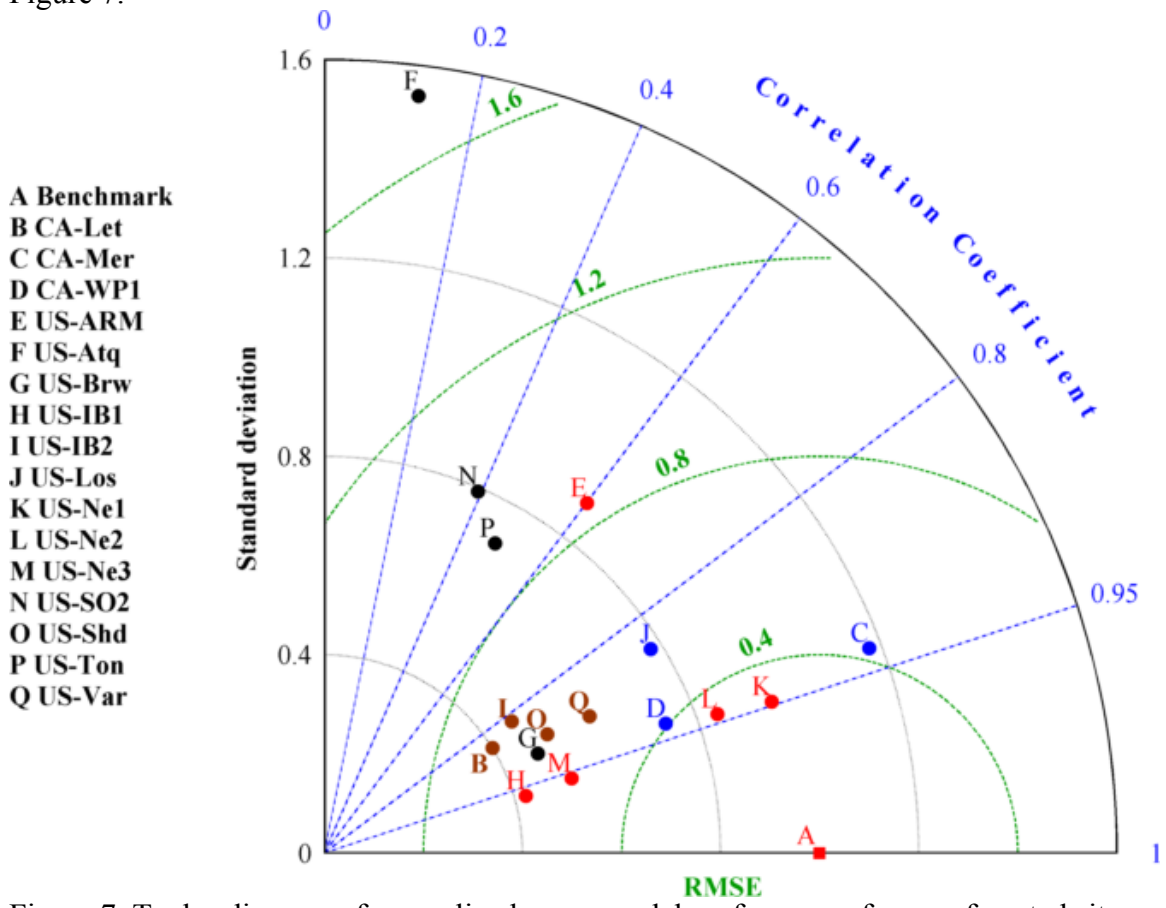

Figure 7. Taylor diagram of normalized mean model performance for non-forested sites. Each circle (n = 16 sites) is the site-specific mean model ensemble (MEAN). Benchmark (red square) corresponds to observed normalized monthly *NEE*; units of σ and *RMSE* are multiples of observed σ. Color coding of site letter and circles indicates biome: croplands (red), grasslands (brown), wetlands (blue), all other biomes (black).

Figure 8.

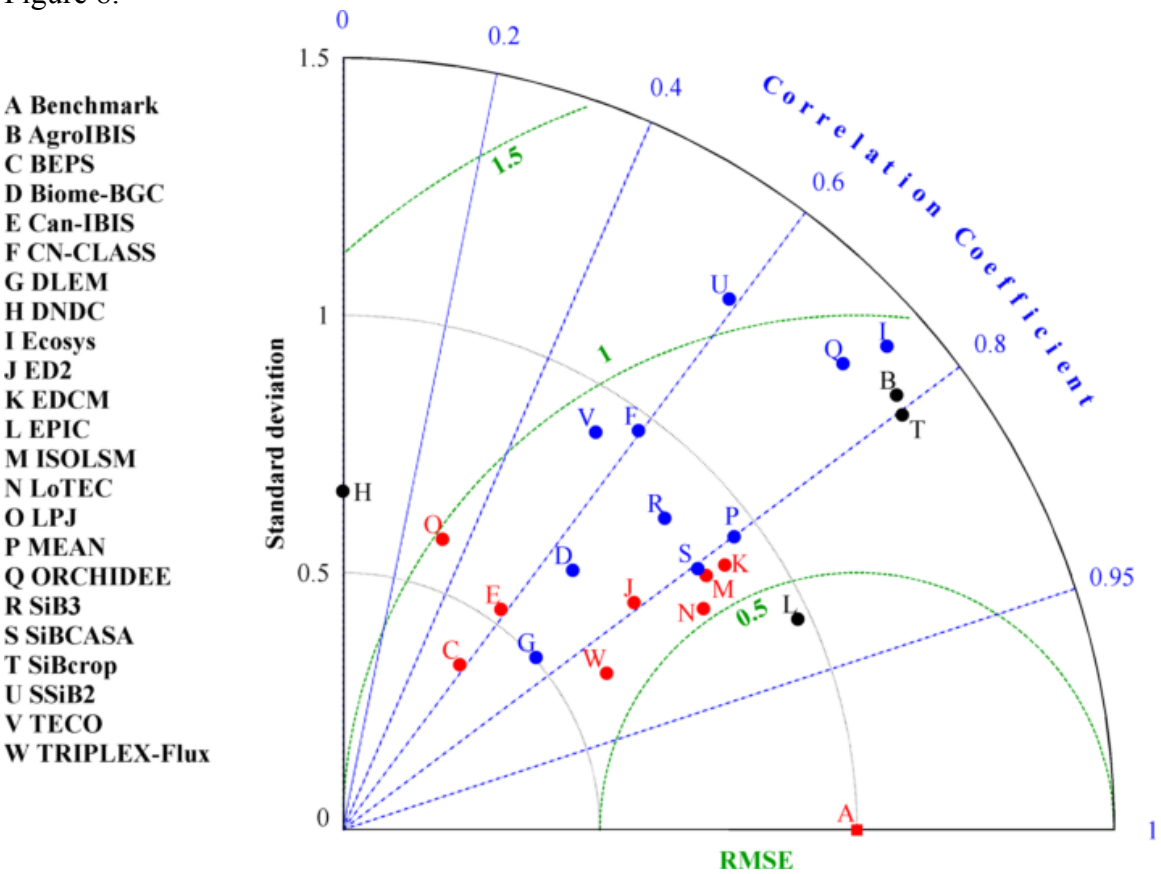

Figure 8. Taylor diagram of normalized across-site average model performance. Model σ and *RMSE* were normalized by observed σ. Each circle (n = 22 models) corresponds to the mean across all sites. Benchmark (red square) corresponds to observed normalized monthly *NEE*; units of σ and *RMSE* are multiples of observed σ. Color coding of model letter and circles indicates generality of model performance: specialist models used only in croplands ($n \leq 5$ sites; black), generalist models used across range of biomes and sites ($n \geq 30$ sites, blue), all other models (red). The correlation for DNDC (ρ = -0.13) is displayed as zero for readability.

Figure 9.

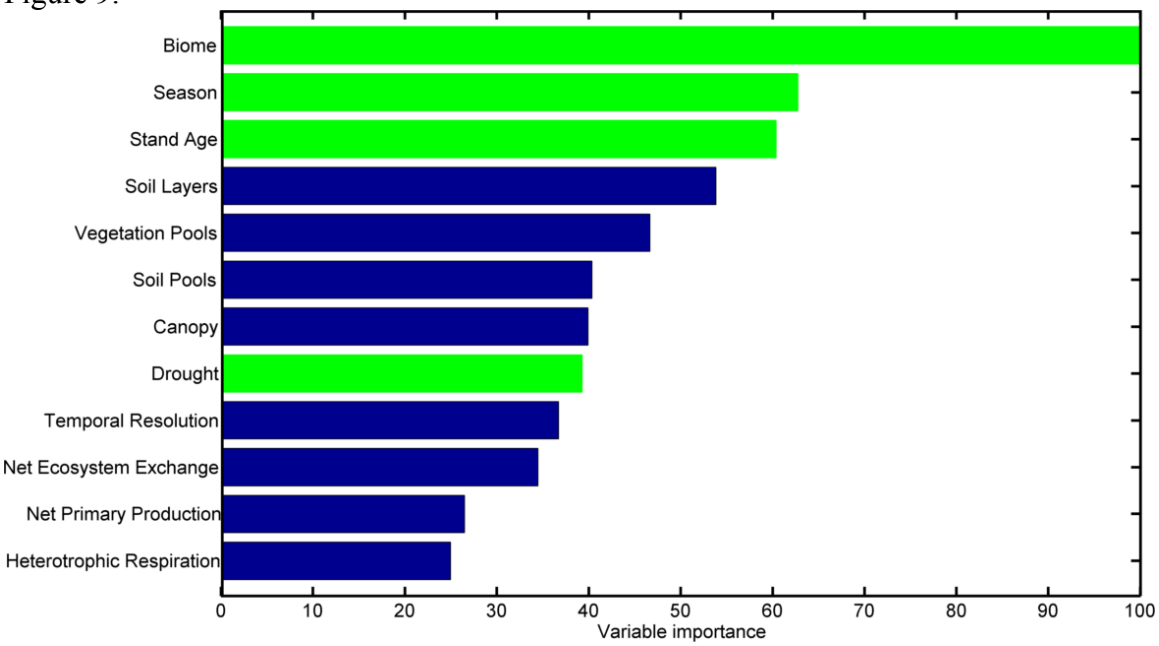

Figure 9. Variable importance scores for model-specific (blue) and site-specific (green) predictors. Scores were generated from a regression tree with the Taylor skill classes based on terciles (n = 3132) as the response. Only the 12 of 28 predictants with score > 25 shown; see Table 3 for complete listing of evaluated model structural and site attributes.

Figure 10.

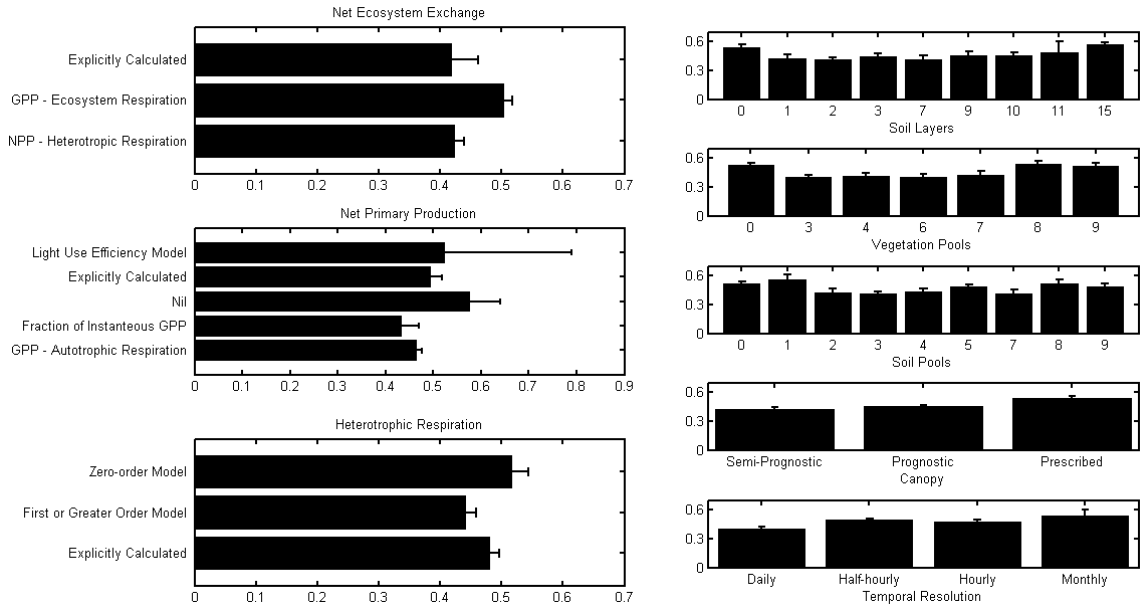

Figure 10. Bar graphs of mean Taylor skill by model attribute. Whiskers represent one standard error of the mean. Only model-specific attributes with variable important scores > 25 shown.

DISCLAIMER