

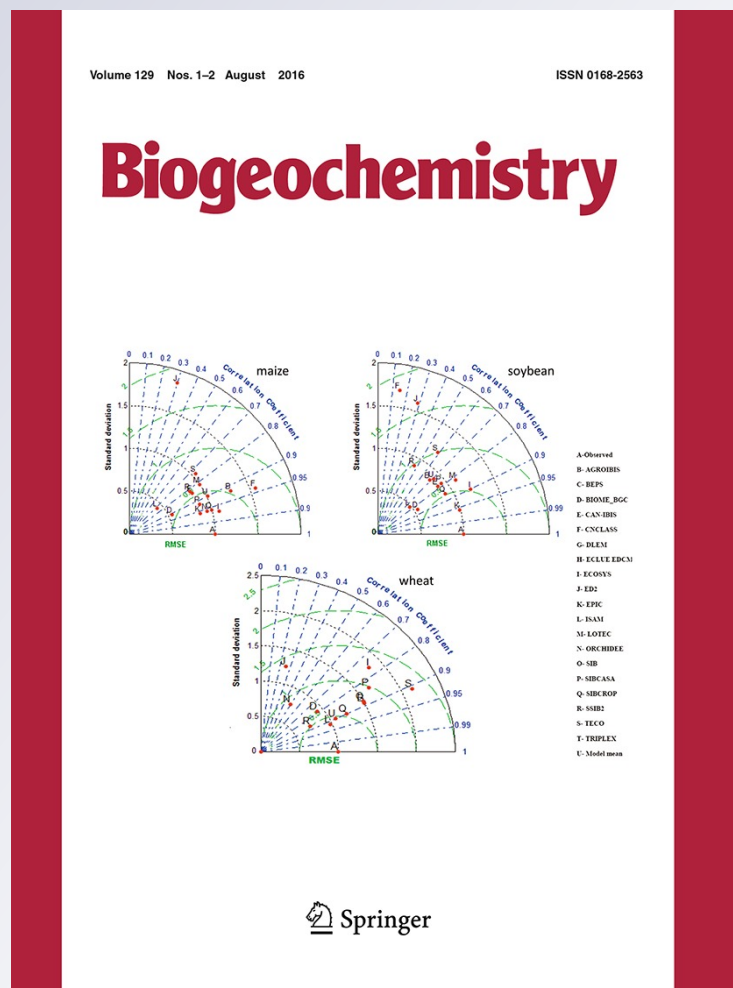
# Carbon and energy fluxes in cropland ecosystems: a model-data comparison

**E. Lokupitiya, A. S. Denning,  
K. Schaefer, D. Ricciuto, R. Anderson,  
M. A. Arain, I. Baker, A. G. Barr,  
G. Chen, J. M. Chen, et al.**

**Biogeochemistry**  
An International Journal

ISSN 0168-2563  
Volume 129  
Combined 1-2

Biogeochemistry (2016) 129:53-76  
DOI 10.1007/s10533-016-0219-3



**Your article is protected by copyright and all rights are held exclusively by Springer International Publishing Switzerland. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at [link.springer.com](http://link.springer.com)".**

# Carbon and energy fluxes in cropland ecosystems: a model-data comparison

E. Lokupitiya · A. S. Denning · K. Schaefer · D. Ricciuto · R. Anderson · M. A. Arain · I. Baker · A. G. Barr · G. Chen · J. M. Chen · P. Ciais · D. R. Cook · M. Dietze · M. El Maayar · M. Fischer · R. Grant · D. Hollinger · C. Izaurralde · A. Jain · C. Kucharik · Z. Li · S. Liu · L. Li · R. Matamala · P. Peylin · D. Price · S. W. Running · A. Sahoo · M. Sprintsin · A. E. Suyker · H. Tian · C. Tonitto · M. Torn · Hans Verbeeck · S. B. Verma · Y. Xue

Received: 4 January 2016 / Accepted: 13 May 2016 / Published online: 3 June 2016  
© Springer International Publishing Switzerland 2016

**Abstract** Croplands are highly productive ecosystems that contribute to land–atmosphere exchange of carbon, energy, and water during their short growing seasons. We evaluated and compared net ecosystem exchange (NEE), latent heat flux (LE), and sensible heat flux (H) simulated by a suite of ecosystem models at five agricultural eddy covariance flux tower sites in the central United States as part of the North American

Carbon Program Site Synthesis project. Most of the models overestimated H and underestimated LE during the growing season, leading to overall higher Bowen ratios compared to the observations. Most models systematically under predicted NEE, especially at rain-fed sites. Certain crop-specific models that were developed considering the high productivity and associated physiological changes in specific crops better predicted the NEE and LE at both rain-fed and irrigated sites. Models with specific parameterization

Responsible Editor: E. Matzner

E. Lokupitiya (✉)  
Department of Zoology and Environment Sciences,  
Faculty of Science, University of Colombo, Colombo 03,  
Sri Lanka  
e-mail: erandi@sci.cmb.ac.lk

A. S. Denning · I. Baker  
Department of Atmospheric Science, Colorado State  
University, 1371 Campus Delivery, Fort Collins,  
CO 80523, USA

K. Schaefer  
National Snow and Ice Data Center (NSIDC), University  
of Colorado, Boulder, CO 80309, USA

D. Ricciuto  
Environmental Sciences Division, Oak Ridge National  
Laboratory, Oak Ridge, TN 37831, USA

R. Anderson · S. W. Running  
Numerical Terradynamic Simulation Group, University of  
Montana, Missoula, MT 59812, USA

M. A. Arain  
School of Geography and Earth Sciences and McMaster  
Centre for Climate Change, McMaster University,  
Hamilton, ON L8S 4K1, Canada

A. G. Barr  
Science and Technology Branch, Environment Canada,  
National Hydrology Research Centre, Innovation  
Boulevard, Saskatoon, SK S7N 3H5, Canada

G. Chen · H. Tian  
Ecosystem Dynamics and Global Ecology Laboratory,  
School of Forestry and Wildlife Science, Auburn  
University, Auburn, AL 36849, USA

J. M. Chen  
Department of Geography, University of Toronto,  
Toronto, ON M5S 3G3, Canada

for different crops better simulated the inter-annual variability of NEE for maize-soybean rotation compared to those models with a single generic crop type. Stratification according to basic model formulation and phenological methodology did not explain significant variation in model performance across these sites and crops. The under prediction of NEE and LE and over prediction of H by most of the models suggests that models developed and parameterized for natural ecosystems cannot accurately predict the more robust physiology of highly bred and intensively managed crop ecosystems. When coupled in Earth System Models, it is likely that the excessive physiological stress simulated in many land surface component models leads to overestimation of temperature and atmospheric boundary layer depth, and underestimation of humidity and CO<sub>2</sub> seasonal uptake over agricultural regions.

**Keywords** Carbon and energy fluxes · Cropland ecosystems · Land–atmosphere exchange · Model–data comparison · Cropland carbon and energy exchange

### Abbreviations

CO<sub>2</sub> Carbon dioxide  
GPP Gross primary productivity

P. Ciais · P. Peylin  
Laboratoire des Sciences du Climat et de l'Environnement (LSCE), CEA CNRS UVSQ, 91191 Gif-sur-Yvette, France

D. R. Cook · R. Matamala  
Environmental Science Division, Argonne National Laboratory, Lemont, IL 60439, USA

M. Dietze  
Department of Earth and Environment, Boston University, Boston, MA 02215, USA

M. El Maayar  
Energy, Environment and Water Research Center, The Cyprus Institute, Nicosia, Cyprus

M. Fischer · M. Torn  
Lawrence Berkeley National Laboratory, Berkeley, CA 94718, USA

R. Grant  
Department of Renewable Resources, University of Alberta, Edmonton, AB T6G 2E3, Canada

H Sensible heat flux  
LAI Leaf area index  
LE Latent heat flux  
MARE Mean absolute relative error  
NACP North American Carbon Program  
NCDC National Climate Data Center  
NEE Net ecosystem exchange  
R Ecosystem respiration  
RMSE Root mean square error  
STD Standard deviation

### Introduction

Croplands are highly productive ecosystems that take up a significant amount of carbon dioxide (CO<sub>2</sub>) from the atmosphere during their short but intense growing seasons. In the US, cropland ecosystems occupy about one fifth of the total land area (Boryan et al. 2011). Maize, soybean, and wheat are the main crops, each occupying over 20 % of the total harvested cropland area (Lokupitiya et al. 2012) of the country. Maize has much higher growing season CO<sub>2</sub> uptake compared to the other two crops. The US has 40 % of global maize production (FAO 2010), and 20 % of the country's production is exported (ERS USDA 2010). With

D. Hollinger  
Northern Research Station, USDA Forest Service, Durham, NH 03824, USA

C. Izaurrealde  
Pacific Northwest National Laboratory and University of Maryland, College Park, MD 20740, USA

A. Jain  
Department of Atmospheric Sciences, University of Illinois, Urbana, IL 61801, USA

C. Kucharik  
Department of Agronomy & Nelson Institute Center for Sustainability and the Global Environment, University of Wisconsin - Madison, Madison, WI 53706, USA

Z. Li  
Teleobservation Research LLC, Columbia, MD 21044, USA

increasing CO<sub>2</sub> and other greenhouse gases, the Earth's climate is changing, creating a challenge in predicting the impacts of future climate change on food production and overall CO<sub>2</sub> uptake and energy exchanges (i.e. latent and sensible heat fluxes) by agricultural ecosystems. Thus having more realistic models of crop performance under any given climatic and environmental conditions is needed, and the reliability of the existing models to predict land-atmosphere carbon and energy exchanges by croplands is important in evaluating the contribution of these ecosystems towards global food production, carbon cycling, surface energy budgets, land-surface climate, and atmospheric circulation.

Models of land-atmosphere interaction and carbon cycling are primarily based on natural ecosystems, which have developed complex strategies for conserving scarce resources like water and nutrients, and can become stressed by heat and aridity during the growing season. Crop plants, by contrast, are bred for maximum productivity without the need to conserve resources at the end of the growing season. They are intensively managed to avoid resource constraints, for example by tilling, irrigation, fertilization, and weeding. Eddy covariance measurements over growing maize show net ecosystem exchange (NEE, expressed as ecosystem respiration minus photosynthesis) of

CO<sub>2</sub> as high as 75 μMol m<sup>-2</sup> s<sup>-1</sup> at midday (Lokupitiya et al. 2009), more than three times the typical rate over the Amazon rainforest (Saleska et al. 2003). As a result, seasonal drawdown of about 35 ppm of CO<sub>2</sub> in the atmospheric boundary layer was observed by a network of instrumented towers in the central US Corn Belt; it was the strongest seasonal CO<sub>2</sub> cycle ever observed (Miles et al. 2012). Atmospheric inverse models using these data found that uptake rates had to be substantially increased relative to Bayesian prior estimates from ecosystem models (Schuh et al. 2013; Ogle et al. 2015). The physiological resilience of crop ecosystems has likely contributed to the observed amplification of the seasonal cycle of CO<sub>2</sub> at background stations since 1960 (Graven et al. 2013), especially due to enhanced technology, irrigation and high yielding varieties (Zeng et al. 2014). Production of maize, wheat, rice, and soybean within the Northern Hemisphere grew by 240 % between 1961 and 2008, leading to a significant increase in the net carbon uptake by croplands during their short growing season by 0.33 petagrams, with maize alone accounting for two-thirds of this change (Gray et al. 2014). Models developed to represent natural ecosystems may struggle to achieve the very high rates of photosynthesis and transpiration in managed crop ecosystems. A recent study using space-based monitoring of sun-

---

S. Liu

U.S. Geological Survey (USGS) Earth Resources  
Observation and Science (EROS) Center, Sioux Falls,  
SD 57198, USA

L. Li

School of Life Sciences, University of Technology  
Sydney, PO Box 123, Broadway, NSW 2007, Australia

D. Price

Natural Resources Canada, Northern Forestry Centre,  
5320-120 Street, Edmonton, AB T6H3S5, Canada

A. Sahoo

Department of Civil and Environmental Engineering,  
Princeton University, E324 Engineering Quad, Princeton,  
NJ 08544, USA

M. Sprintsin

Forest Management and GIS Department, Jewish National  
Fund-Keren Kayemet LeIsrael, Jerusalem, Israel

A. E. Suyker · S. B. Verma

School of Natural Resources, University of Nebraska, 807  
Hardin Hall, Lincoln, NE 68583-0968, USA

C. Tonitto

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

H. Verbeeck

CAVElab – Computational and Applied Vegetation  
Ecology, Faculty of Bioscience Engineering, Ghent  
University, 9000 Ghent, Belgium

Y. Xue

Department of Geography, University of California, Los  
Angeles, Los Angeles, CA 90095-1524, USA

induced chlorophyll fluorescence (SIF) from terrestrial vegetation found that SIF-based crop gross primary productivity (GPP) estimates over regions such as the US Corn Belt are 50–75 % higher than those from state-of-the-art carbon cycle models (Guanter et al. 2014). Gross primary production and transpiration are physiologically linked via stomatal conductance, so models that underestimate GPP are also likely to simulate warmer temperatures and a drier atmosphere over agricultural regions when linked to climate models.

Many model-data and inter-model comparison studies have been carried out for different ecosystems or vegetation types concerning different spatial domains. Previous model comparison studies mostly focused on forested (e.g. Ryan et al. 1996; Amthor et al. 2001; Grant et al. 2005), and agricultural (e.g. Semenov et al. 1996; Frohling et al. 1998; Ciais et al. 2010; Asseng et al. 2013, 2015; Bassu et al. 2014; Martre et al. 2015) ecosystems. Semenov et al. (1996) compared the performance of five wheat models at two sites in Europe: Rothamsted, United Kingdom, and Seville, Spain. The aim of the study was to compare the model predictions under several climate change scenarios and investigate the effects of changes in climatic variability on model predictions; Semenov et al. (1996) compared grain yields simulated by the five models, but the study did not address any model limitations or errors. Frohling et al. (1998) compared nitrous oxide (N<sub>2</sub>O) fluxes simulated by four models against field measurements from five agricultural sites in three countries (United States, Scotland, and Germany). The field sites included sites with different fertilizer treatments. Model performance varied depending on site characteristics and the model output had a daily resolution for majority of the parameters simulated. It was concluded that for better model performance, the models should have algorithms for special characteristics corresponding to different cropping systems such as cover-crops, underseeds, mulches, reduced tillage, different organic fertilizers, etc. Ciais et al. (2010) evaluated the performance of crop productivity by several models on croplands in Europe. They identified certain limitations within the models including the lack of detail on management practices and structural limitations.

The North American Carbon Program (NACP) covers the United States, Canada, and Mexico. The objectives of NACP are to measure, understand,

predict and explain the sources and sinks of CO<sub>2</sub>, methane, and carbon monoxide in North America and in adjacent ocean regions (Denning et al. 2005). The NACP has organized several synthesis activities to evaluate and compare ecosystem models and observations at local and regional scales. The NACP Site-Level Interim Synthesis (Site Synthesis) compared the simulated and observed fluxes at 47 eddy covariance flux towers across North America that included conifer and deciduous forests, grasslands, peatlands, tundra, and both irrigated and rain-fed crop lands, to determine how well the models agreed with observations. Out of the models that participated in the Site Synthesis 20 ecosystem models simulated the five agricultural sites considered.

Previous model-data comparisons conducted under NACP Site Synthesis focused mostly on carbon dynamics. For instance, Schwalm et al. (2010) analyzed the CO<sub>2</sub> exchange simulated by 22 terrestrial biosphere models at 44 eddy covariance flux towers in North America and found that few models simulating different biomes and sites, the mean model ensemble, and a model that used data assimilation for parameter optimization showed high consistency with observations. Dietz et al. (2012) used spectral analyses to determine the performance of 21 ecosystem models at multiple time scales considering 9 eddy covariance flux tower sites; this study found that the model performance was related to model time step, soil hydrology, and the representation of photosynthesis and phenology in the models. Stoy et al. (2013) used wavelet coherence to analyze the model performance of 20 ecosystem models at 10 different eddy covariance research sites in simulating NEE at different time scales and identified the need for better parameterization and mechanistic improvement of models for more accurate predictions. Schaefer et al. (2008) compared the daily average GPP simulated by 26 models against estimated GPP at 39 eddy covariance flux tower sites across the United States and Canada, and highlighted seasonal differences in GPP based on ecosystem types and moisture availability in soil.

Recent studies by Asseng et al. (2013, 2015) and Martre et al. (2015) compared the performance of multiple wheat models (27–30 models) in simulating crop yields and growth variables and highlighted the better performance shown by the ensemble mean or median compared to individual models. Similarly, Bassu et al. (2014) highlighted the better accuracy of

the ensemble mean in evaluating the performance of 23 models simulating maize yields at four sites.

In this study, we consider model simulations of carbon and energy fluxes at four selected agricultural sites with maize, soybean, and wheat crops in central North America against eddy covariance flux tower measurements as observations. The specific variables evaluated in the comparison were net ecosystem exchange of CO<sub>2</sub> (NEE, the difference between ecosystem respiration and photosynthesis) and turbulent fluxes of latent and sensible heat (LE and H, respectively). Most of the models in the intercomparison are intended for general application in simulating ecosystem-atmosphere interactions, and were not developed specifically for agroecosystems, although some of them had the parameterization for certain crops. Since only a few models incorporate parameterizations to simulate crop yields, we did not evaluate yields. The specific objectives of this study were to (a) evaluate the overall model performance against the observed carbon and energy fluxes at the eddy covariance flux tower sites considered under the NACP Site Synthesis, (b) assess the model performance for individual crops and sites based on the current parameterization and capabilities of the models, and (c) identify potential changes needed to improve model performance.

## Materials and methods

### Site description and observed fluxes

Table 1 shows the five eddy covariance flux tower sites that were identified as agricultural sites under the NACP Site Synthesis. Four sites had maize and soybean crops and the remaining site (i.e. Southern Great Plains eddy covariance flux tower site (Fischer et al. 2007) under the Atmospheric Radiation Measurement (ARM) program in Oklahoma; US-ARM) had winter wheat and occasional summer crops (maize and soybean). All these sites were located in the US Midwest agricultural plains and varied by location (Fig. 1), soils and weather, presence of a single crop or crops in rotation, management (e.g. tillage and fertilizer levels, use of irrigation), and other site specific parameters (Table 1).

We used measured hourly or half-hourly carbon and energy fluxes for the above sites in evaluating

model performance. Although the gaps in observed NEE data were filled using the modified Fluxnet-Canada Method (Barr et al. 2004, 2013), the gap-filling methodology occasionally eliminated certain years that had sparse NEE data. Therefore in the current analyses, we used only non-gap filled, observed flux data.

### Models and simulation details

Output from ecosystem models that participated in the model simulations of CO<sub>2</sub> and energy fluxes at the agricultural sites under the NACP Site Synthesis were analyzed and compared against the observed fluxes. Of the 20 models, five were agricultural models (Table 2), simulating only agroecosystems, while the rest of the models simulated other ecosystems as well. The five agricultural models and a few other ecosystem models had parameterization corresponding to the specific crops considered in this study (maize, soybean, and wheat), while the rest of the models had a generic crop (i.e. universal crop parameterization). Most of the models calculated changes in leaf area index (LAI) using a prognostic method using a combination of climate, photosynthesis, and carbon allocation. Five models used diagnostic phenology, in which LAI was prescribed from remotely sensed monthly Global Inventory Modeling and Mapping Studies (GIMMS) Normalized Difference Vegetation Index data (Tucker et al. 2005).

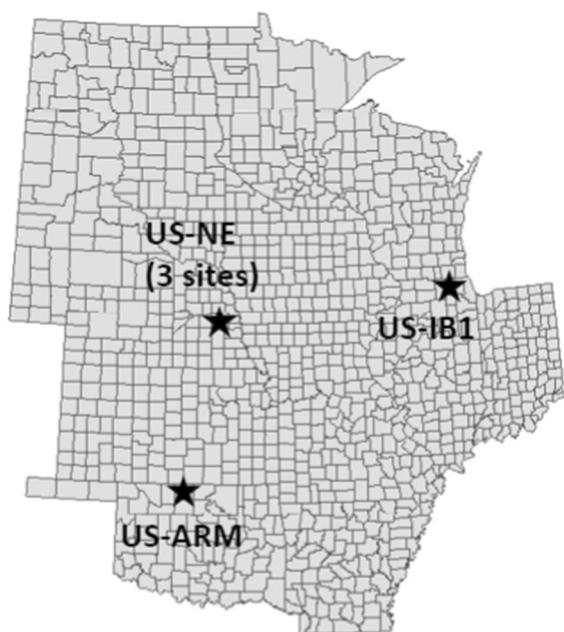
Table 2 briefly describes each model and outlines the inputs required for estimating the variables analyzed in this study. The models varied in temporal resolution: 12 models had hourly or sub-hourly temporal resolution and 6 had daily resolution (Table 3); one model (ECLUE EDCM) had monthly resolution and one model (ISAM) had weekly resolution. The models were spun up to steady-state initial conditions using site-specific, gap-filled observed weather data as input (Ricciuto et al. 2009, 2013).

The meteorological variables used in the model simulations included air temperature, specific humidity, wind speed, precipitation, atmospheric pressure, surface incident shortwave and longwave radiation, and CO<sub>2</sub> concentration. The three sites at Mead, Nebraska (US-NE), had hourly weather data and the Fermi agricultural site, Illinois (US-IB1), and the US-ARM sites had half-hourly weather data. Gaps in weather data were initially filled using the available

**Table 1** Site characteristics

Site	Code	Location	Latitude	Longitude	Rain-fed/ irrigated	Main crop/s	Reference
Mead irrigated	US-NE1	Mead, Nebraska	41.1651	−96.4766	Irrigated	Maize	Suyker and Verma (2008), Verma et al. (2005)
Mead irrigated rotation	US-NE2	Mead, Nebraska	41.1649	−96.4701	irrigated	Maize and soybean in rotation	Suyker et al. (2004), Verma et al. (2005)
Mead rain-fed	US-NE3	Mead, Nebraska	41.1797	−96.4396	Rain-fed	Maize and soybean in rotation	Suyker et al. (2004), Verma et al. (2005)
Fermi agricultural site	US-IB1	Illinois	41.8593	−88.2227	Rain-fed	Maize and soybean in rotation	Xiao et al. (2008)
Atmospheric Radiation Measurement (ARM) program Central Facility site	US-ARM <sup>a</sup>	Oklahoma	36.6058	−97.4888	Rain-fed	Wheat	Fischer et al. (2007)

<sup>a</sup> This site was the only site with wheat. Thus the main crop (i.e. wheat) was considered for the analyses



**Fig. 1** Site locations within the US Midwest region (US-NE: three Mead sites at Saunders County, Nebraska; US-IB1: Fermi agricultural site in DuPage and Kane Counties in Illinois; US-ARM: Atmospheric Radiation Measurement (ARM) program Central Facility site in Grant County, Oklahoma)

weather data from a nearby flux tower site within 30 km distance and 150 m elevation and any remaining gaps were filled using National Climate Data

Center (NCDC) climate station data, using any hourly data from the nearest station first and then using the daily NCDC data or DAYMET (<http://www.daymet.org/>) daily data. When a choice of daily datasets was available, the daily dataset that had better correlation with observations was used for gap-filling, and was temporally downsampled to sub-daily values using an average diurnal cycle (Ricciuto et al. 2009).

For the model simulations, site-specific ancillary data provided by the principal investigators at the eddy covariance flux tower sites were used. These included (1) site location (latitude and longitude), (2) soil properties (soil texture, %age of silt, clay, and sand, soil depth and nutrient (C and N) content, water holding capacity, etc.), (3) crop rotation and management (planting and harvesting times, disturbance and land use history information), and (4) measured LAI and biomass in different plant pools. Descriptions of basic model characteristics and simulation output for the variables NEE, LE, and H were provided by participating modelers.

#### Model comparison

Although there is considerable variation in the model parameterization and characteristics among the models, analyses relating model formulation to



**Table 2** Overview of ecosystem models based on the information provided by the modeling teams

Model	Photosynthesis (GPP and/or NPP)	Phenology and LAI	Presence of generic/specific crops
Agro-IBIS <sup>a</sup>	Enzyme Kinetic Model	Prognostic; LAI based on leaf carbon and SLA	Specific crops
BEPS	Enzyme Kinetic Model	Diagnostic; depends on remote sensing	No crop specificity; distinct C3 versus C4 vegetation
Biome-BGC	Stomatal Conductance Model	Prognostic; LAI based on leaf carbon and SLA	No crop specificity; distinct C3 versus C4 vegetation
Can-IBIS	Enzyme Kinetic Model	Prognostic	No crop specificity; distinct C3 versus C4 (grass)
CN-CLASS	Enzyme Kinetic Model	Prognostic	Crop specific parameters can be applied
DLEM	Stomatal Conductance Model	Prognostic	Specific crops
DNDC <sup>a</sup>	Light Use Efficiency Model	Prognostic	Specific crops
ECLUE EDCM	Statistical	Prognostic	Specific crops
Ecosys	Enzyme Kinetic Model	Prognostic; LAI based on leaf carbon growth, senescence	Specific crops
ED2	Enzyme Kinetic Model	Prognostic; LAI based on leaf carbon and SLA	C3 (soybean) and C4 (maize) crops
EPIC <sup>a</sup>	Light Use Efficiency model	Prognostic	Specific crops
ISAM	Stomatal Conductance Model	Diagnostic; LAI based on remote sensing	No crop specificity
LOTEC	Enzyme Kinetic Model	prognostic	No crop specificity; C3 versus C4 parameters tuned to match observations
ORCHIDEE – STICS <sup>a</sup>	Enzyme Kinetic Model	Prognostic; LAI has a specific phenology calculated by STICS crop model with parameters of two French Maize (DK604) and Wheat (Soisson) varieties	Specific crops (maize and wheat)
SiB3	Enzyme Kinetic and Stomatal Conductance Models	Diagnostic; depends on remote sensing	C3 versus C4 grasses and generic crops
SiBCASA	Enzyme Kinetic and Stomatal Conductance Models	Diagnostic; depends on remote sensing	C3 versus C4 grasses and generic crops
SiBcrop <sup>a</sup>	Enzyme Kinetic and Stomatal conductance Models	Prognostic; LAI based on leaf C and SLA	Specific crops
SSiB2	Stomatal Conductance Model	Diagnostic; depends on remote sensing	
TECO	Stomatal Conductance Model	Prognostic	No crop specificity
TRIPLEX-Flux	Stomatal Conductance Model	LAI –measured or calculated as a function of leaf biomass and SLA	No crop specificity

Model summaries emphasize controlling factors and mechanism relevant to NEE and energy fluxes. The information provided in this table and the model output used in the analyses correspond to November 2009, and do not include any subsequent model modifications. Description of the abbreviations:

*LAI* leaf area index, *SLA* specific leaf area

<sup>a</sup> Agricultural models

performance could be performed only to the extent feasible based on the limited basic information provided by the modelers under a predetermined

common format designed for the Site Synthesis participants. Similar to some previous analyses done under the NACP Site Synthesis (Schwalm et al. 2010;

Dietz et al. 2012; Schaefer et al. 2008; Stoy et al. 2013), the output from those models that simulated agricultural sites was analyzed at a monthly time scale, to account for the variation in the temporal resolution among the models (hourly, daily, weekly, and monthly). Taylor diagrams based on modeled flux standard deviation (STD), root mean square error (RMSE), and correlations with the observed data were used to evaluate the model output against the observed, non-gap filled fluxes of carbon (NEE), and energy (LE, and H).

#### Analyses of model performance by site and by crop

We evaluated model skill *by site* considering each crop separately, and *by crop* with all sites having a given crop combined, using Taylor skill (Taylor 2001):

$$S = \frac{4(1+R)^4}{\left(\hat{\sigma}_f + \frac{1}{\hat{\sigma}_f}\right)^2 (1+R_0)^4} \quad (1)$$

where  $S$  is Taylor skill,  $R$  is the correlation coefficient,  $\hat{\sigma}_f$  is the normalized standard deviation of the model output (i.e. model standard deviation/observed standard deviation), and  $R_0$  is the maximum, potentially realizable correlation. Taylor skill indicates the model's accuracy in reproducing both the magnitude and the phase of observed variability. Taylor skill scores could vary between zero (least skillful) and one (most skillful) (Taylor 2001). For the analyses *by site*, we compared monthly averaged model output for NEE, LE, and H fluxes corresponding to each crop at each site against observed values. For the analyses *by crop*, we compared the monthly averaged model output for NEE, LE, and H fluxes from each crop with all the sites and years combined against observed values. In our analyses, we considered those models with a skill score greater than 0.8 to have high performance, and those with a skill score of  $< 0.5$  to have poor performance.

We also calculated mean absolute relative error (MARE):

$$MARE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - x_i}{x_i} \right|, \quad (2)$$

where  $N$  is total number of observations,  $y_i$  is  $i$ th model estimate, and  $x_i$  is the  $i$ th observation (de La Casinière et al. 1997).

We calculated MARE for the peak crop growth period (July and August for maize and soybean, and April and May for wheat) to evaluate model performance when each crop is in its highest productivity. We combined the output from all the sites with the relevant crop. MARE increases as the absolute difference between the observations and model output increases. A MARE of  $\sim 1$  indicates that the difference between the observations and the predictions is about the same magnitude as the observed flux. Thus the model with the lowest MARE has the best performance.

#### Annual and diurnal cycles of the carbon and energy fluxes

Out of the three fluxes (i.e. NEE, LE, and H), the one that was consistently simulated by the models was NEE or the carbon flux, and NEE at US-NE3 site that had maize and soybean in rotation was simulated by the majority of the models. Therefore, the carbon fluxes at this site was analyzed to evaluate the ability of the models to simulate the pattern of variation and magnitude of crop annual cycles and overall inter-annual variability that depend on the physiology and type of crop present in the field; for studying the variability in annual cycles, monthly mean NEE at US-NE3 site was used.

Although the main analyses of the study were done at a monthly scale, diurnal cycles were also evaluated for those models that had sub-daily (i.e. hourly and half-hourly) resolution, to have a better understanding on their performance in simulating carbon and energy fluxes in croplands. The diurnal cycles of NEE, LE, and H were analyzed for the crop sites at Mead, Nebraska, for model performance under rain-fed and irrigated management practices; the output from 12 models with hourly/half-hourly temporal resolution were used (Table 2). The amplitude and phase of diurnal variation simulated by each model (and the overall model mean) during the peak growing season were studied and compared against the observed mean and standard deviation for the same period. In the analyses, NEE was defined as (R–GPP), and thus negative values in the diurnal cycle would indicate a net uptake of CO<sub>2</sub> from the atmosphere. The same sign convention was used for the diurnal fluxes of energy (i.e. flux to the atmosphere indicated as ‘+’ and the reverse indicated as ‘–’).

**Table 3** Model details and simulation specifics

Model	Reference <sup>a</sup>	Temporal Resolution	No. of crop sites simulated <sup>b</sup>	Model applicability	Simulation specifics
Agro-IBIS	Kucharik (2003), Kucharik and Twine (2007)	Half-hourly	5 (All)	Crop sites only	
BEPS	Chen et al. (1999), Liu et al. (1999)	Daily	1 (US-NE3)		Soybean only
Biome-BGC	Running and Hunt (1993), Thornton et al. (2002)	Daily	5		NEE and LE only
Can-IBIS	Foley et al. (1996), El Maayar et al. (2002)	Half-hourly	1(US-NE3)		NEE and LE only
CN-CLASS	Arain et al. 2006; Kothavala et al. 2005	Half-hourly	4 (sites excluding US-ARM)		
DLEM	Ren et al. 2007; Tian et al. 2008	Daily	4 (sites excluding US-ARM)		NEE only
DNDC	Li et al. 1992	Daily	5	Crop sites only	
ECLUE EDCM	Liu et al. 2003	Monthly	1(US-NE3)		NEE only
Ecosys	Grant et al. (2007a, b)	Hourly	5	all terrestrial ecosystems	
ED2	Moorcroft et al. (2001), Medvigy et al. (1995)	Half-hourly	5		NEE and LE only
EPIC	Williams (1995), Izaurrealde et al. (2006)	Daily	1(US-NE3)	Crop sites only	NEE and LE only
LOTEC	Hanson et al. (2004)	Hourly	1(US-NE3)		NEE and LE only
ISAM	Jain et al. (2005)	Weekly	2 (US-NE3 and US-ARM)		Energy fluxes only
ORCHIDEE-STICS	De Noblet-Ducoudré et al. 2004, Krinner et al. (2005)	Half-hourly	5	Crop sites only	maize and wheat only
SiB3	Baker et al. (2008)	Half-hourly	1(US-ARM)		
SiBCASA	Schaefer et al. (2008)	Half-hourly	5		
SiBCROP	Lokupitiya et al. (2009)	Half-hourly	5	Crop sites only	
SSiB2	Xue et al. 1991, Zhan et al. (2003)	Half-hourly	5		
TECO	Weng and Luo (2008)	Hourly	5		NEE and LE only
TRIPLEX	Sun et al. (2008), Zhou et al. (2008)	Daily	1(US-NE3)		NEE only

<sup>a</sup> Only up to two references are given

<sup>b</sup> Models either simulated all the sites or only one (i.e. Mead, NE rain-fed site for all the other models excepting SiB3 and ISAM, which simulated US-ARM site only)

The monthly NEE simulated by the models were grouped and statistically compared to relate model performance to the basic model formulation, considering 1) method of GPP calculation and 2) prognostic versus diagnostic phenology and LAI (Table 2). For each group, a mean bias (mean of observed minus predicted NEE) was calculated first and then either one-way ANOVA (for the multiple groups based on GPP calculation) or two sample *t* test with 95 percent confidence level (for the two groups based on

phenology) was used on those biases to test the statistical significance of differences among model groups. These analyses were performed separately for C4 (maize) and C3 (soybean and wheat) crops.

## Results and discussion

According to our findings, there was no single model that could perform equally well with regard to all three

fluxes (i.e. NEE, LE, and H) analyzed in the current study, and the models showed significant variability in simulating both seasonal and diurnal cycles of the different fluxes. However, certain general trends could be identified, as discussed below.

### Model performance for individual fluxes

#### *Net ecosystem exchange*

**Performance by site** At all three **US-NE** sites, the majority of the models underestimated NEE during the growing season for both maize and soybean crops, which led to lower Taylor skills. Depending on the site, only 20–30 % of the models that simulated maize and 10–20 % of the models that simulated soybean at US-NE sites showed high performance with skills of greater than 0.8 (Fig. 2). The site that was simulated by the largest number of models (i.e. 17 out of 20 models) was US-NE3 rain-fed site, while the irrigated US-NE1 and US-NE2 sites were simulated by a lower number (i.e. 14) of models. With regard to maize, the majority of the models and the overall model mean had higher skills at the irrigated sites, compared to the rain-fed site (Fig. 2), showing poorer model performance at the rain-fed site. However, the model performance for soybean was not much different between the rain-fed (US-NE3) and irrigated (US-NE2) sites. Only a few models (less than 20 %) consistently simulated both crops well at the rain-fed and irrigated sites (Fig. 2).

The above findings could be related to the way the models address irrigation and the fact that some models do not have adequate parameterization to simulate soil moisture or drought stress under rain-fed conditions. Only a few models had parameterizations to represent irrigation. For instance, in ORCHIDEE–STICS, the water requirements for optimal crop growth on irrigated land is calculated as the difference between maximum transpiration and available water (defined as the difference between precipitation and total runoff); irrigation in Ecosys is handled explicitly through the addition of selected amounts of water (in mm) during selected hours on selected dates for irrigation events. These values are typically those reported from the field site, and may follow an automated protocol if field values are not available; in SiBcrop, the irrigation is represented by restricting the moisture availability between the field capacity and

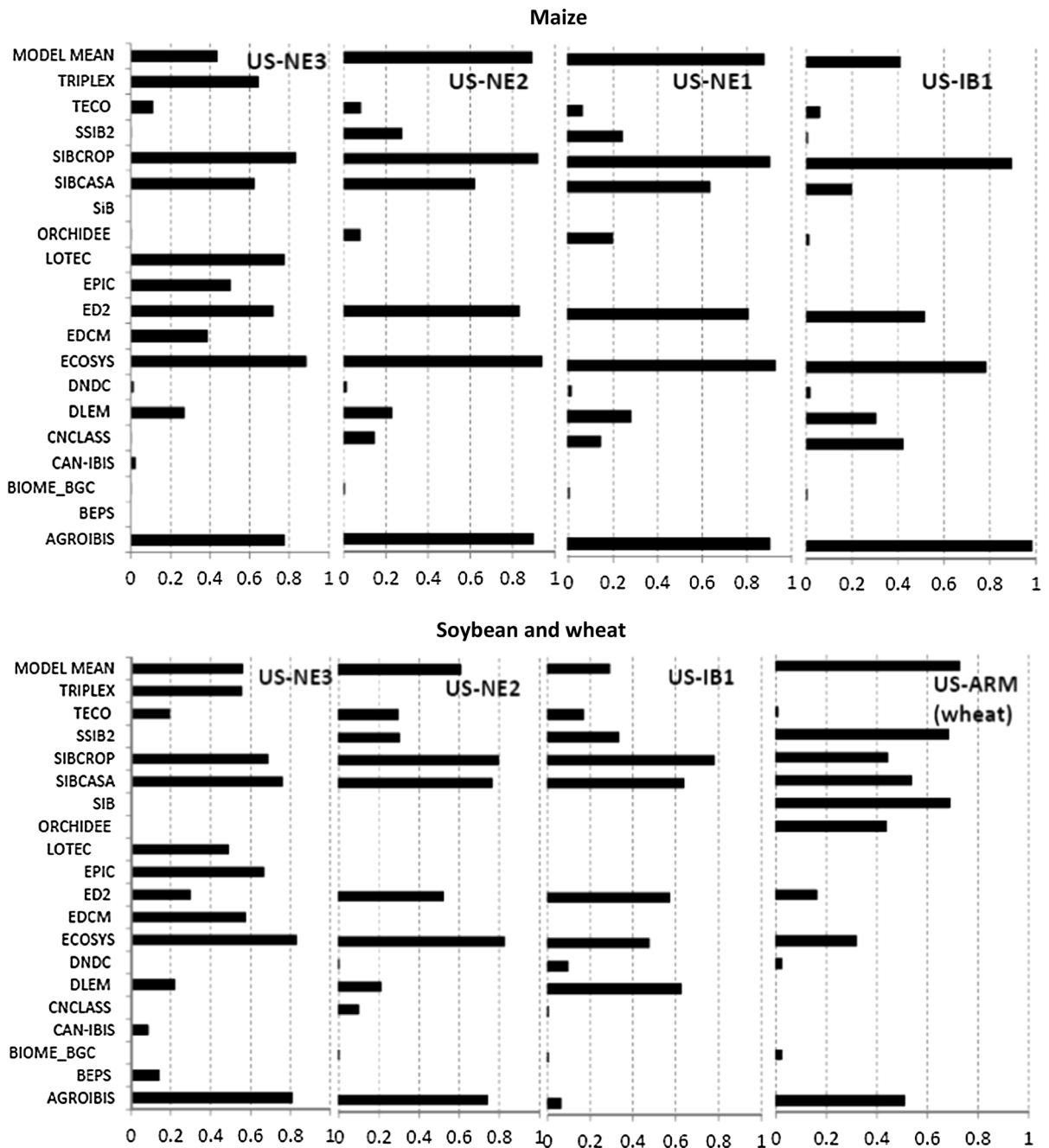
saturation in response to irrigation, thus minimizing drought stress. The problems in simulating drought stress were mostly obvious in rain-fed sites, especially for maize, and the majority of the models showed a bias for drought stress at rain-fed sites. The tendency of most models to overestimate drought stress may reflect their heritage as simulators for natural ecosystems rather than crops.

The overall model performance for NEE was the lowest at US-IB1 site, as depicted by the skills of individual models and overall model mean (Fig. 2). About 60 % of the models and the model mean had skills of less than 0.4, and this was true for both maize and soybean crops.

At the **US-ARM** site, the majority of the models had a skill of less than 0.6, and the model mean had a skill of about 0.7. Thus the model mean showed greater skill at US-ARM compared to most of the individual models, indicating that the mean skill has been elevated by the very high skills of a few models when the ensemble mean is taken for the comparison against the observations (Fig. 2).

**Performance by crop** The Taylor diagrams in Fig. 3 illustrate the model performance by each crop when all the sites were combined. Since the normalized standard deviation in the Taylor diagrams was calculated without removing the seasonal cycle, the distance from the origin to each radial curve was essentially the ratio of simulated to observed NEE amplitude. Therefore those models that had a normalized standard deviation close to 1.0 produced a seasonal amplitude very close to the observed. The overall best model performance was shown by the models that also had the highest correlation and lowest RMSE (Fig. 3).

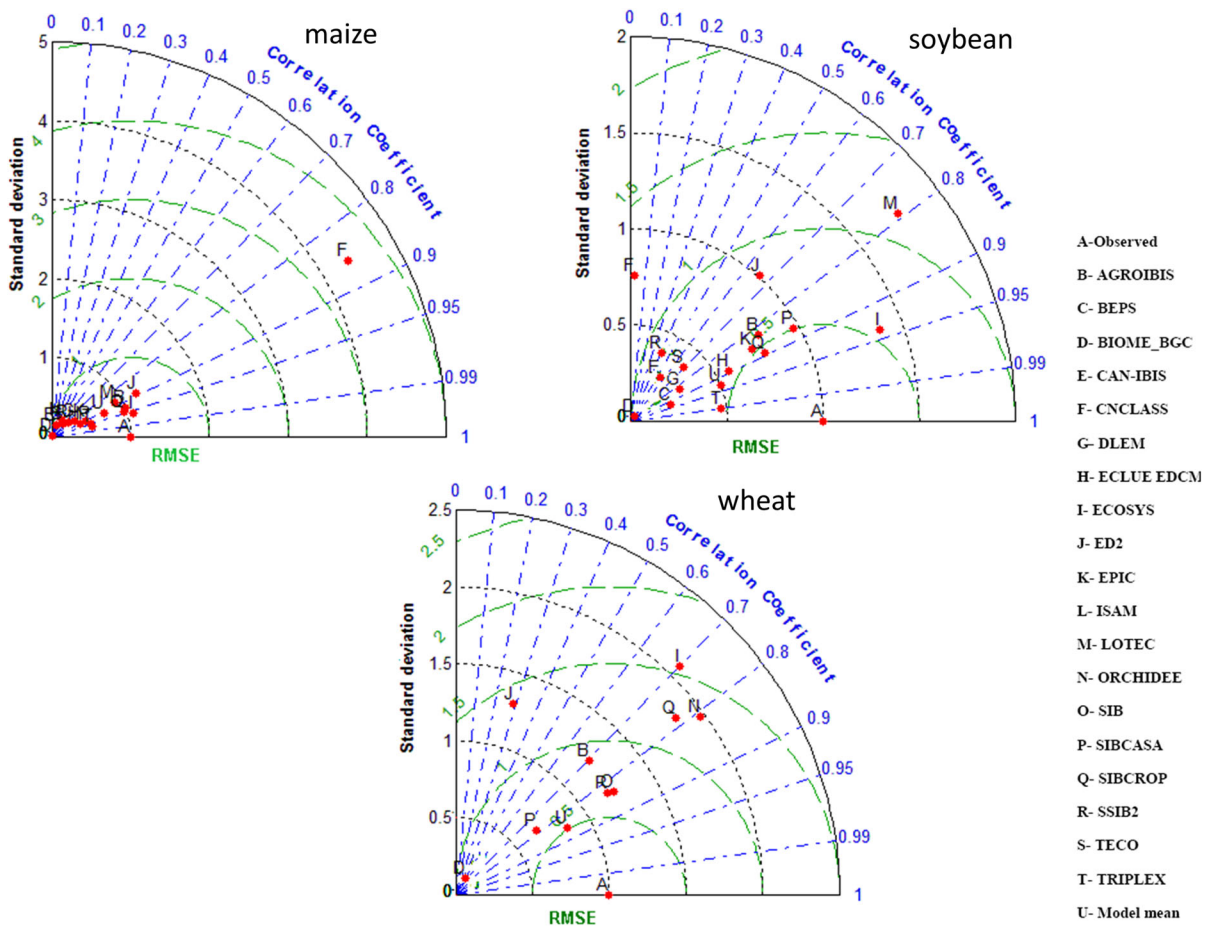
For **maize**, there was a significant difference in the phasing and the amplitude of seasonal carbon uptake among different models, as shown by the correlation and normalized standard deviation (and RMSE) in the Taylor diagram (Fig. 3). The amplitude and timing of the seasonal cycle of the observed maize NEE was more closely simulated by the models corresponding to the points B, I, and Q in the Taylor diagram, which had higher correlations and RMSE values closer to the reference point that corresponded to the observed data (indicated by the letter A in Fig. 3). The above three models had a crop specific parameterization for maize with model skills of  $\sim 0.9$  and the lowest MARE ( $<0.2$ ) during the peak



**Fig. 2** Comparison of the model skill values for net ecosystem exchange (monthly means) simulated by different models for the different sites against the observed data

growing season. The majority (60 %) of the models had a skill of more than 0.5 in predicting the maize NEE (when all the irrigated and rain-fed sites combined). The model mean had a skill of 0.8 and a MARE of 0.3 during the peak growing season.

For **soybeans**, the magnitude of NEE and standard deviation were quite different among the models, leading to significant differences between model skills. Model skills of ~0.8 were found for the models corresponding to I, P, and Q in the Taylor diagram



**Fig. 3** Taylor diagrams for crop wise model performance in simulating monthly mean NEE against the observations

(Fig. 3), showing the best model performance compared to the observations (Fig. 3). Models represented by points B and K had a skill of 0.7, while the model mean and rest of the models had lower skills. In general the models' ability to simulate both the magnitude and timing of NEE seasonal cycle was poor for soybean, as shown by the pattern of distribution of the points corresponding to different models in Fig. 3. MARE during the peak growing season for most of the models was higher for soybean than that for maize, indicating more deviation of the estimated NEE from the observed NEE for soybean. For instance, B, I, and Q had a MARE of less than 0.2 for maize, but had a MARE of 0.4, 0.5, and 0.3, respectively, for soybean. Soybean was simulated by altogether 17 models, and 70 % of the models had MARE greater than 0.5, showing poorer performance compared to maize.

As mentioned above, **wheat** at the US-ARM site was best simulated by the ensemble mean followed by the individual models represented by O, P, and R (Fig. 3). The lowest MARE during peak growth was obtained for O and P (both having MARE = 0.2), followed by model mean (MARE = 0.3). Wheat crop was simulated by 11 models, and 8 of them (70 %) had MARE greater than 0.5, again showing poorer performance compared to maize. Therefore, when we consider the skill scores and MARE among the crops, the overall highest model performance by crop was found for maize, followed by the C3 crops, soybean and wheat, respectively. However, when we consider the performance by the model mean, the best performance was shown by the ensemble mean for wheat, implying a random distribution of the models about the true value.

*Inter-annual variability of NEE* There was considerable variation in the model performance with regard to the inter-annual variability of net carbon uptake. Inter-annual variability in NEE is mostly driven by crop rotation. During the years with maize crop (2001 and 2003) there was much higher net uptake of CO<sub>2</sub> (maximum uptake  $24 \pm 0.11 \mu\text{mol m}^{-2} \text{s}^{-1}$ ) compared to the years with soybean (2002 and 2004; maximum uptake  $14 \pm 2.7 \mu\text{mol m}^{-2} \text{s}^{-1}$ ) in the observations at US-NE3 site. Soybean has LAI and gross primary productivity compared to maize, and during the years with soybean, there is lower net carbon uptake due to lower GPP and higher respiration from decomposition of the crop residues from the previous year's maize crop. By contrast, during the years with maize crop there is higher GPP and lower respiration from decomposition of the lower amounts of residue inputs from the previous year's soybean crop. Thus higher NEE is found in the years with maize crop compared to the years with soybean in the field. The lower carbon uptake during soybean years was predicted to varying degrees by the models and the model mean.

Models that used crop specific parameterizations for maize and soybean better predicted the pattern of inter-annual variability in maize-soybean rotation with maize crop showing larger CO<sub>2</sub> uptake despite the significant variation in the magnitude in the predicted uptake by the different models. Although some of the models were not specifically crop models, using crop-specific parameterization for soybean and maize in these models helped in getting the expected inter-annual variability. Those models with a generic crop (Table 2) could not capture the difference in the amplitude of annual carbon uptake between maize and soybean well, leading to approximately the same amplitude of NEE for both crops in rotation.

*Diurnal cycles of NEE for maize and soybean* The observed average diurnal cycle during the growing season (from the beginning of July to beginning of September, as considered in this analysis) for US-NE2 and US-NE3 sites is illustrated in Fig. 4. At both sites, the models showed significant variation in the overall magnitude of the diurnal cycle of carbon uptake by maize and soybean.

At rain-fed US-NE3, the average observed diurnal cycle for carbon uptake peaked around

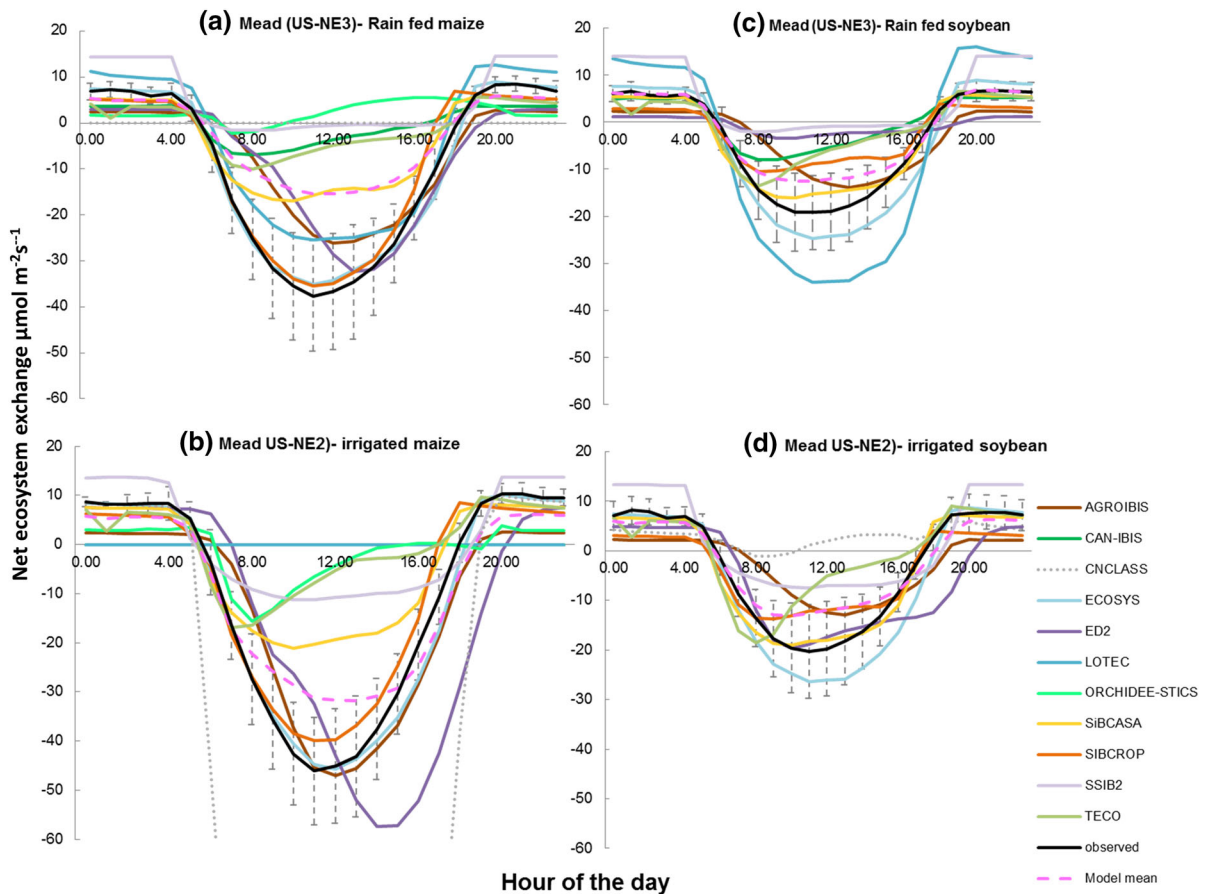
$40 \mu\text{mol m}^{-2} \text{s}^{-1}$  for maize. The amplitude of uptake was closely simulated by only two models, while the remaining models simulated 65 % or less of the observed carbon uptake (Fig. 4a). The models differed in the algorithms used in the calculation of the components of NEE (i.e. GPP and R); the influence of the differences in model structure and parameterization were also evident. For instance, certain models had much higher ecosystem respiration leading to greater positive NEE during night time and less carbon uptake during daytime. Similarly, the daytime carbon uptake by maize was weaker than the observations in models with a generic crop with C3 physiology. The diurnal cycle of NEE for soybean at US-NE3 was poorly simulated by the models. Only two models and the model mean had a performance falling within one standard deviation of the observations (Fig. 4b).

The irrigated US-NE2 site had much larger diurnal carbon uptake by maize in both the observations and the model estimates, compared to the rain-fed site; three models had an amplitude of carbon uptake very close to the observations, while most of the remaining models simulated less than half the uptake compared to the observed diurnal cycle (Fig. 4c). For soybean at US-NE2, the majority of the models had lower amplitudes in the diurnal cycle, compared to the observations (Fig. 4d). Because so many of the models substantially underestimated NEE, the few models actually parameterized for crop physiology outperformed the ensemble mean.

#### *Latent heat flux*

*Performance by site* At the US-NE sites, overall model performance for LE was better for maize and soybean at irrigated sites compared to the rain-fed US-NE3 site, as indicated by the skills of individual models and the overall model mean (Fig. 5). At US-NE1, US-NE2, and US-NE3 sites, the %age of the models with a skill of over 0.8 for maize was 60, 70, and 50 %, respectively, with a larger %age at the two irrigated sites. The %age of models showing a skill score of over 0.8 for soybean crop at US-NE2 and US-NE3 was 20 and 30 %, respectively (Fig. 5).

At the Fermi agricultural site, US-IB1, the models had far better skills for maize LE compared to soybean. Skills greater than 0.8 were observed for the models B, I, N, Q and the model mean, in simulating maize. Soybean crop was poorly simulated



**Fig. 4** Diurnal cycles of NEE: model performance against the site level observations at Mead rain-fed and irrigated sites during maize (*left*) and soybean (*right*) years, considering the average NEE over the growing season. The vertical error bars

correspond to one standard deviation from the observed mean across the years that had observed data (i.e. 2001 and 2003 for maize and 2002 and 2004 for soybean)

with a very low skill for the model mean and the individual models (i.e. skills <0.5). Of all the sites, models and model mean showed the poorest performance for US-IB1 site during soybean years.

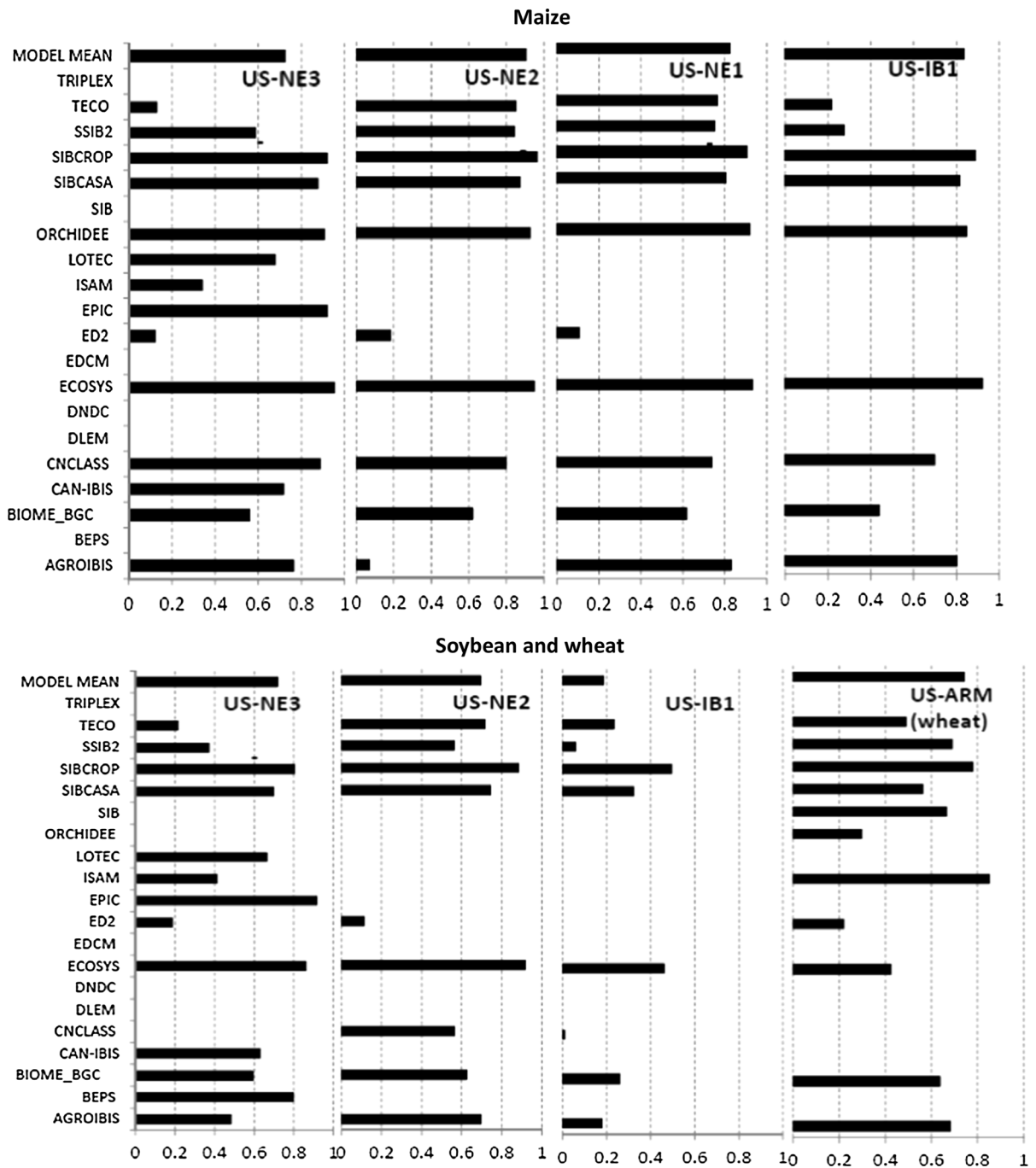
At the **US-ARM** wheat site, the model mean and a majority of the models had a skill greater than 0.6, with only two models having a skill of 0.8 (Fig. 5). Thus no consistent pattern could be observed across the sites for model performance in simulating LE, except the better performance at irrigated sites compared to the rain-fed site at Mead, NE.

*Performance by crop* Although the majority of models had a correlation of over 0.8 for each crop, there was significant variation in the magnitude of the LE fluxes, as illustrated by RMSE in the Taylor diagrams in

Fig. 6. The closest performance to the observed LE for maize was found in I, K, N, and Q, all of which had model skills of over 0.9. They also had minimum MARE (<0.2) during the peak growing season for maize, indicating a close performance to the observations. I, K, and Q had the lowest MARE (<0.2) during the peak growing season of soybean as well. In general, MARE values were larger for wheat LE, indicating poorer model performance compared to the other two crops, with the majority of the models having MARE greater than 0.3.

The pattern shown by models within the Taylor plots revealed more about model performance (Fig. 6). For maize, although the correlations were high, the standard deviations varied widely for majority of the models, indicating that the models got the timing of the seasonal cycle right, but the

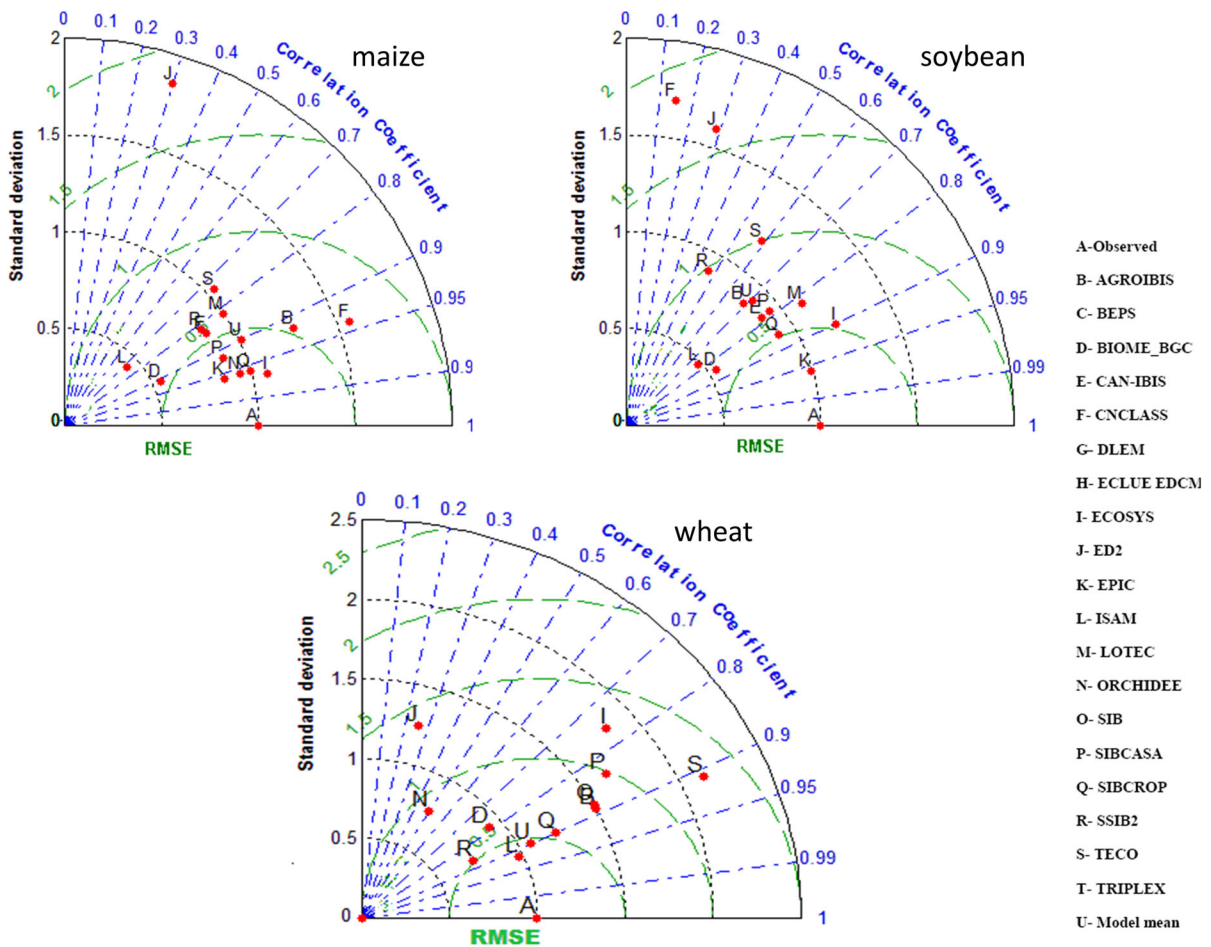




**Fig. 5** Comparison of the model skill values for latent heat (monthly means) simulated by different models for the different sites against the observed data

seasonal amplitude was wrong. For soybean, all the models had standard deviations close to one, with a wide range of correlations, indicating that the models

got the magnitude of the seasonal cycle correct, but not the timing. For wheat, the models showed a large spread in both standard deviation and correlation,



**Fig. 6** Taylor diagrams showing the model performance against observations with regard to crop wise latent heat flux simulations

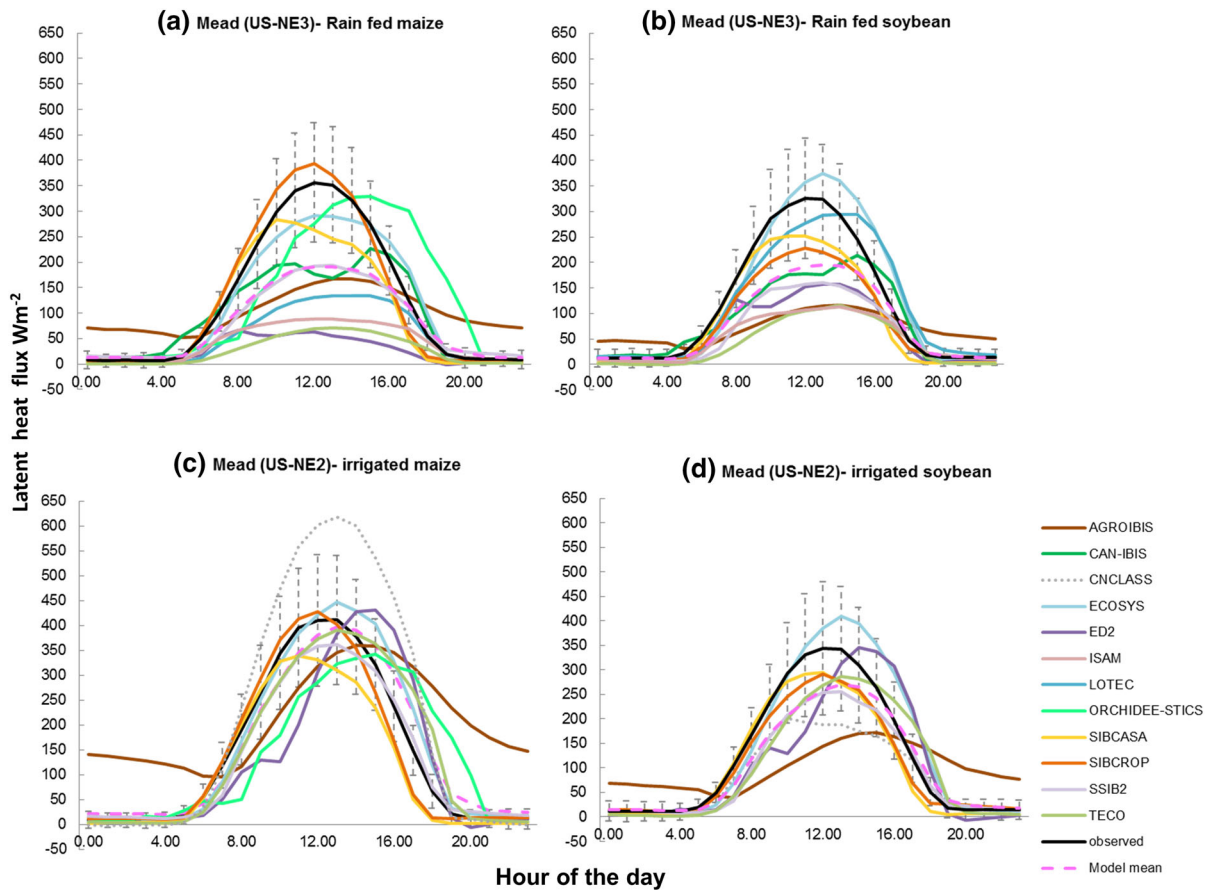
indicating that they got neither the seasonal amplitude nor the timing correct.

**Diurnal cycles of LE** The model estimates of LE had significant differences in amplitude and pattern of variation compared to the observed diurnal cycle of LE. Overall, the majority of models had lower LE values compared to the observations at US-NE3 rain-fed site. Models had better performance at the irrigated US-NE2 site; most of the models fell within one standard deviation from the observed curve for most of their diurnal cycles; this was true for both maize and soybean (Fig. 7). Thus similar to NEE, most models substantially underestimated LE at the rain-fed site. Because both NEE and LE are linked to stomatal conductance, this result is consistent with our interpretation that most models overestimate drought stress in crop ecosystems.

*Sensible heat flux*

**Performance by site** Compared to NEE or LE, simulated monthly H values at the US-NE sites (US-NE1, US-NE2, US-NE3) had smaller skill levels (Fig. 8). The majority of the models had a skill of less than 0.3; the highest skill observed at individual sites fell within the range 0.6–0.7. The skill of the overall model mean was also relatively low (i.e. 0.2–0.5) for all three sites, showing poor overall model performance for H (Fig. 8). There was not much difference in the model skills observed for irrigated versus rain-fed sites either.

Sensible heat flux was overestimated by most of the models during the growing season. Productive crop ecosystems typically do not experience significant drought stress and they exhibit high rates of transpiration and relatively low sensible heat flux, making the



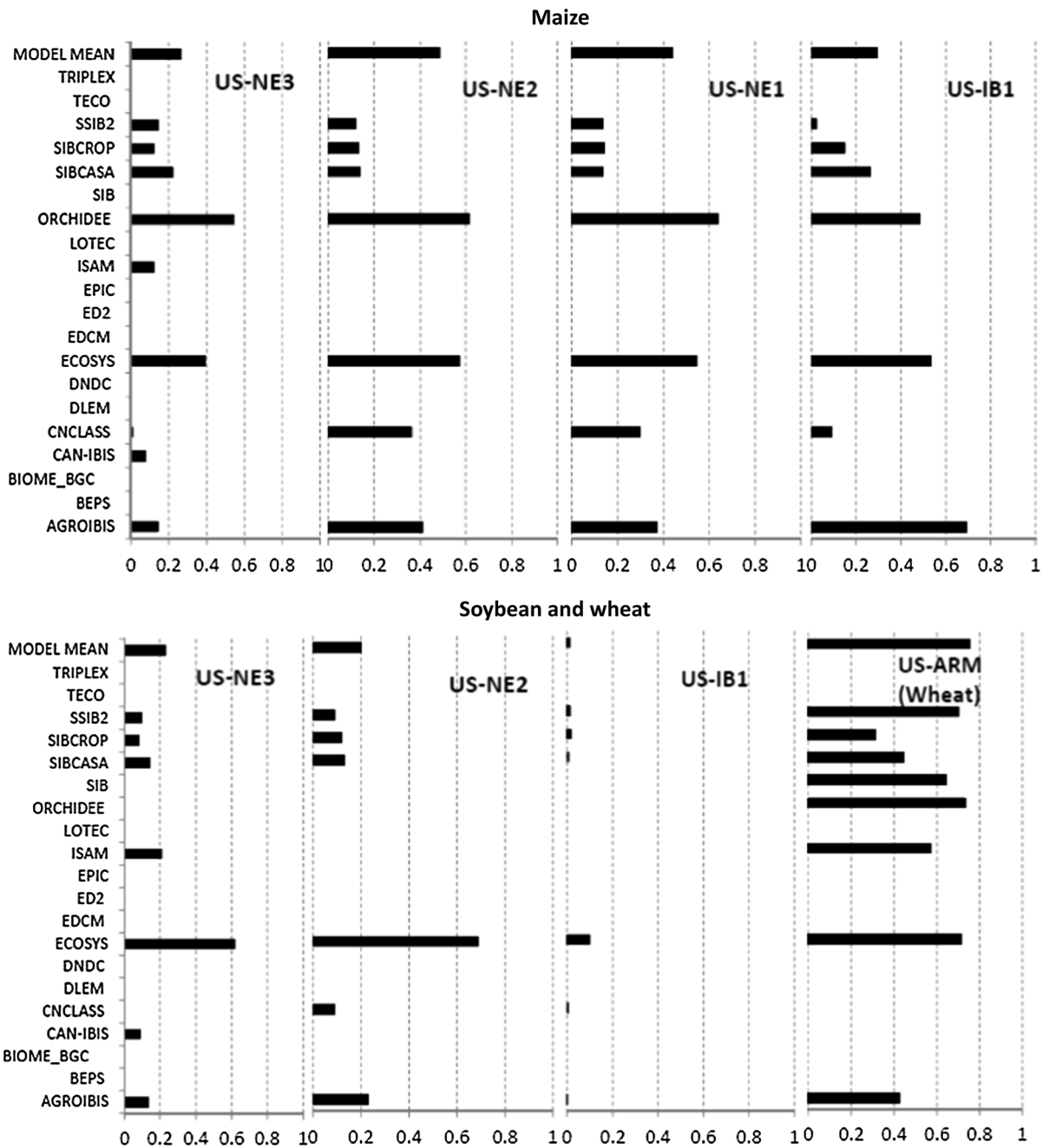
**Fig. 7** Latent heat flux average diurnal cycle across the time series for Mead rain-fed (US-NE3; *top*) and irrigated (US-NE2; *bottom*) sites

denominator in Eq. 2 smaller compared to the situation for NEE or LE fluxes and therefore giving rise to larger MARE.

At Fermi agricultural site, **US-IB1**, overall model performance was poor, but similar to LE, it was better for maize compared to soybean; half of the models had a skill of greater than 0.3 for maize; the skill for soybean was less than 0.1 for the majority of the models (Fig. 8). Overall, the site-level model performance was lower for H, compared to NEE or LE, and the model skills did not exceed 0.8 at any given site. The maximum model skills observed ranged between 0.7 and 0.75, as shown by three of the models and the model mean at US-ARM wheat site (Fig. 8). Thus a higher model performance was found at US-ARM site compared to the rest of the sites (Fig. 8), with ensemble mean having the highest skill.

**Performance by crop** For maize, the estimated MARE during peak growth was over 1.0 for all the models and the overall model mean, indicating that the difference between the observed and predicted H was more than twice the magnitude of observed H during the peak growth; the same was true for soybean. However, for wheat there were comparatively lower relative errors during the peak growth; the majority (5 out of 8) had MARE of less than 0.4 for H, while the rest of the models that simulated wheat H had MARE of over 0.8.

The RMSE and overall model performance for different models in simulating H of different crops is given in Fig. 9. For maize and soybean crops, both the correlation and standard deviation significantly varied among the models, indicating that timing and amplitude of the seasonal cycle of the H flux from these

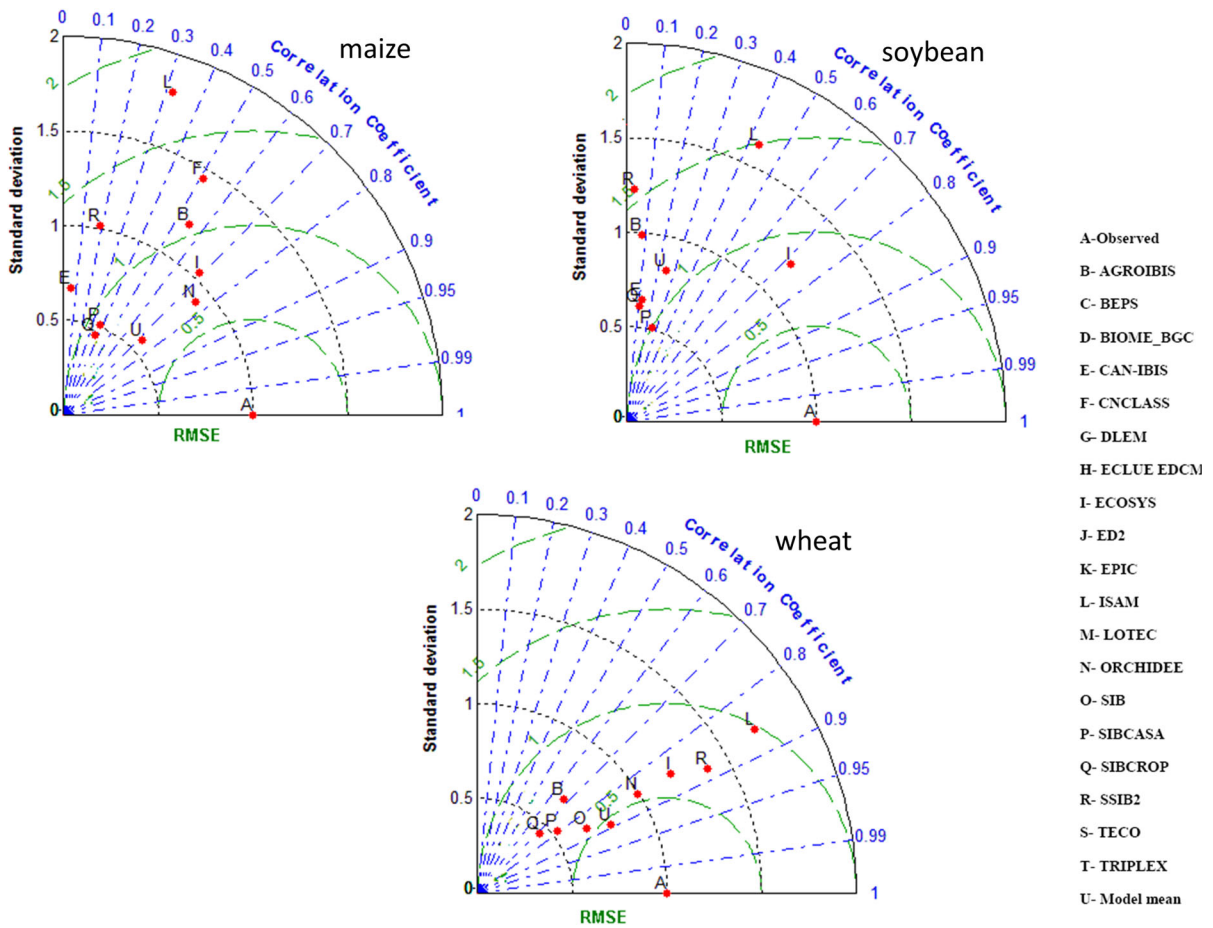


**Fig. 8** Comparison of the model skill values for sensible heat (monthly means) simulated by different models for the different sites against the observed data. DNDC (not shown here due to

estimated near-zero skill), a model with daily temporal resolution had also simulated sensible heat flux

crops were poorly simulated by the models. The most poorly simulated crop H across all the sites was found for soybean, with low skills, very low correlations and a wide range of standard deviations in the Taylor

diagrams. However, for the wheat crop, better correlations and a range of standard deviations were shown by a majority of the models, indicating better timing (but not the amplitude) of the seasonal cycle,



**Fig. 9** Taylor diagrams showing model performance by crop (i.e. combined sites) of sensible heat flux against the observations

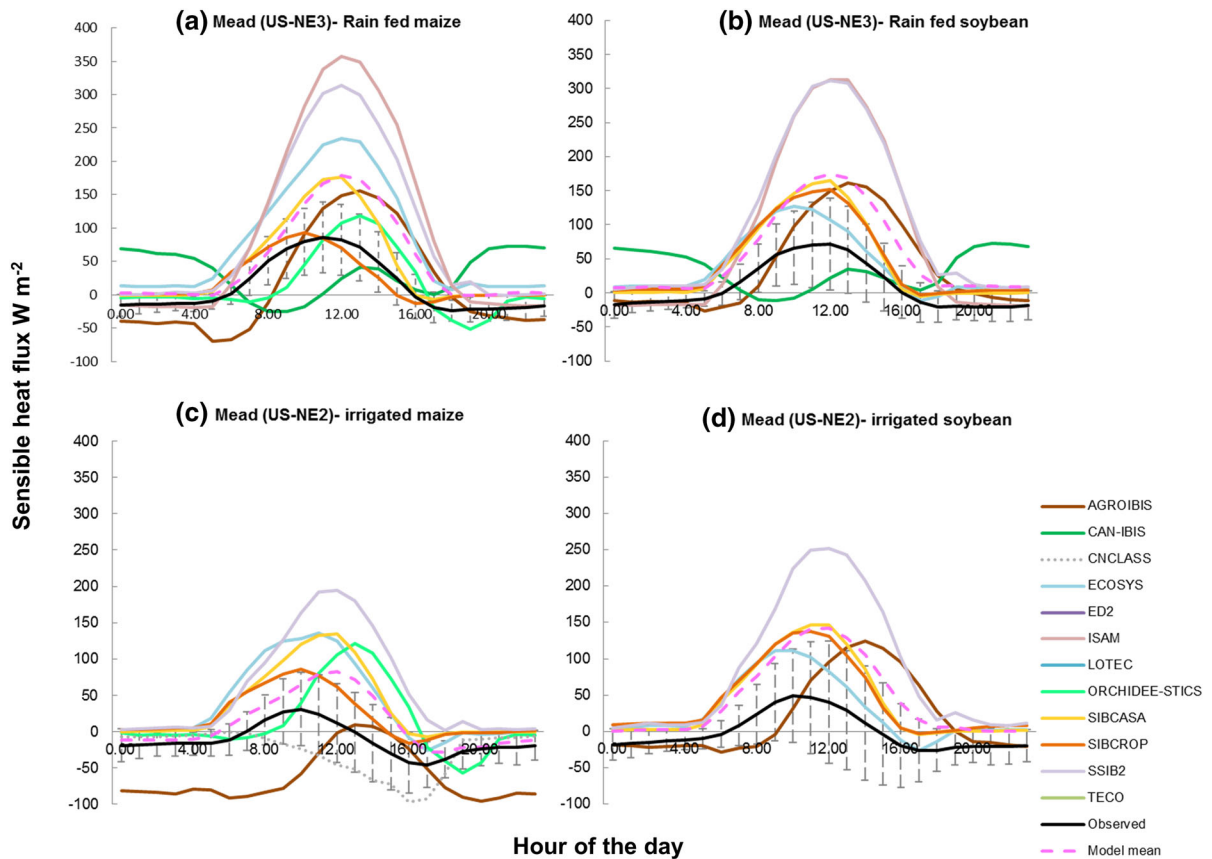
compared to the H flux predicted for the other two crops. This might be due to model performance based on noticeable weather changes at the site during the growth of wheat crop, as the wheat crop starts greening up when soils warm up after the winter dormancy period. The models seem to capture the timing of this warming in soils better compared to the two summer crops involved in this study.

Overall, most of the models showed a strong positive bias in simulating H fluxes, with larger RMSE, lower correlation and overall lower skills, compared to the other two variables. This is consistent with the results for NEE and LE, and with our interpretation that most ecosystem models in this study systematically overestimate drought stress for crop ecosystems. At the site level, better model performance could be found for US-ARM site, yielding higher performance by the model mean for

wheat crop, compared to the other two crops; several models showed better performance for wheat, showing better skills and correlations in the Taylor diagrams.

*Diurnal cycles of H* There was a substantial variation in the diurnal cycle of H simulated by different models during the peak growing season at both rain-fed and irrigated sites. Sensible heat flux at the Mead irrigated site was lower compared to the rain-fed site in observations and the majority of the model estimates (Fig. 10). For both crops, the number of models that fell within one standard deviation of the observed mean was quite low (Fig. 10).

The majority of the models had much higher H and lower LE compared to the observed values for the diurnal cycles, showing higher Bowen ratios. There are several possible reasons for the poor model



**Fig. 10** Sensible heat flux average diurnal cycle across the time series for maize and soybean at Mead rain-fed (US-NE3; *top*) and irrigated (US-NE2; *bottom*) sites

performance with regard to the energy fluxes at diurnal scale. Observed surface energy fluxes at eddy covariance flux tower sites do not conserve energy (the energy budget does not “close” or sum to zero) (Twine et al. 2000; Wilson et al. 2002; Foken 2008), producing a bias in the fluxes mostly due to under measured latent and sensible heat fluxes. Alternatively, there could be a mismatch in energy partitioning in the models, compared to that of the observations (a Bowen Ratio bias). For instance, Grant et al. (2005) found larger partitioning to H in the models used for boreal forest simulations. However, systematic overestimation of H, and underestimation of both LE and NEE across most models in our study strongly suggest a problem in simulating excessive drought stress relative to the observations. Thus, improving the simulation of drought stress would lead to more accurate and better predictions of carbon and energy fluxes from croplands.

#### Performance based on basic model formulation

There was no statistical difference in model performance for NEE based on model grouping according to the GPP calculation (enzyme kinetic, stomatal conductance, light use efficiency, or statistical model). Based on one-way ANOVA, all the groups had similar performance ( $p > 0.05$ ). Overall, there were no statistically significant patterns among model performance when stratifying models according to phenology (prognostic versus diagnostic) or model structure and formulation.

We have identified errors in crop seasonal and diurnal cycles of the fluxes during the peak growth of crops: the models were able to capture seasonality better compared to the diurnal variation. When there was poor performance in the diurnal cycle by the majority of the models, the model mean was more deviant (beyond one standard deviation) from the

observations. This was especially evident for NEE and LE for rain-fed maize at US-NE3, rain-fed soybean LE, and overall diurnal cycles of H.

Unlike in some other previous studies (e.g. Asseng et al. 2013, 2015; Bassu et al. 2014; Martre et al. 2015) the ensemble mean did not generally outperform the individual models, which was especially true for the sites with maize and soybean crops, which had few models with crop specificity that had better skills compared to the ensemble mean.

## Conclusion

Overall higher H, lower LE, and lower NEE compared to the observations were produced by the majority of the models that simulated the agricultural sites considered in this study. Our analyses revealed that having crop-specific parameterization within the models could help better simulation of the inter-annual variability of fluxes under crop rotation. However, only a few models that had crop specificity could accurately simulate the magnitude of the carbon flux. Overall, there was no difference in model performance based on the basic model formulation for photosynthetic calculation or on diagnostic versus prognostic phenology.

The greater productivity and evapotranspiration by croplands as dynamic artificial ecosystems could not be simulated accurately by the majority of the models, the parametrization of which are mostly based on natural ecosystems. The main reason for this seems to be that the models overestimate drought and heat stress which could be handled effectively by the improved crop varieties in the field, yielding higher productivity in reality. Thus to get the magnitude of the fluxes (especially the energy fluxes) right, the models should improve the simulation of physiological stress. The improved technology, management practices, and high yielding and drought-tolerant crop varieties found in reality, makes it a difficult challenge for the existing models to accurately simulate the correct magnitude of net carbon uptake and energy exchange found in these man-made ecosystems. The systematic tendency toward excessive drought stress exhibited by these models is also likely to lead to overestimation of air temperature and atmospheric boundary layer depth, and to underestimation of atmospheric humidity and clouds over agricultural

regions when such models are coupled in Earth System Models.

**Acknowledgments** 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. We acknowledge the comments given by Dr. Andrew Richardson during the initial stages of this manuscript. This research was partly funded by the U.S. Department of Energy (DoE; under contract Nos DE-FG02-06ER64317 and DE-AC02-05CH11231) and National Oceanic and Atmospheric Administration Award NA07OAR4310115. Data from the US-ARM site was supported by the Office of Biological and Environmental Research of the U.S. Department of Energy (under grant or contract DE-AC02-05CH11231) as part of the Atmospheric Radiation Measurement Program. We also acknowledge the support from the Center for Multiscale Modeling of Atmospheric Processes (CMMAP; NSF-ATM-0425247). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

## References

- Amthor JS, Chen JM, Clein JS, Frohling SE, Goulden ML, Grant RF, Kimball JS, King AW, McGuire AD, Nikolov NT, Potter CS, Wang S, Wofsy SC (2001) Boreal forest CO<sub>2</sub> exchange and evapotranspiration predicted by nine ecosystem process models: intermodel comparisons and relationships to field measurements. *J Geophys Res* 106(D24):33623–33648. doi:[10.1029/2000JD900850](https://doi.org/10.1029/2000JD900850)
- Araïn MA, Yaun F, Black TA (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.2006.03.02](https://doi.org/10.1016/j.agrformet.2006.03.02)
- Asseng S et al (2013) Uncertainty in simulating wheat yields under climate change. *Nat Clim Change* 3:827–832
- Asseng S et al (2015) Rising temperatures reduce global wheat production. *Nat Clim Change* 5:143–147
- 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](https://doi.org/10.1029/2007JG000644)
- Barr AG, Ricciuto DM, Schaefer K, Richardson A, Agarwal D, Thornton PE, Davis K, Jackson B, Cook RB, Hollinger DT, van Ingen C, Amiro B, Andrews A, Araïn MA, Baldocchi D, Black TA, Bolstad P, Curtis P, Desai A, Dragoni D, Flanagan L, Gu L, Katul G, Law BE, Lafleur P, Margolis H, Matamala R, Meyers T, McCaughey H, Monson R, Munger JW, Oechel W, Oren R, Roulet N, Torn M, Verma S (2013) NACP Site: Tower Meteorology, Flux Observations with Uncertainty, and Ancillary Data, Data set, Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA. doi:[10.3334/ORNLDAC/1178](https://doi.org/10.3334/ORNLDAC/1178)

- Barr AG et al (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
- Bassu S et al (2014) How do various maize crop models vary in their responses to climate change factors? *Glob Change Biol* 20:2301–2320
- Boryan C, Yang Z, Mueller R, Craig M (2011) Monitoring US agriculture: the US department of agriculture, national agricultural statistics service, cropland data layer program. *Geocarto Int* 26(5):341–358
- Chen JM, Liu J, Cihlar J, Guolden ML (1999) Daily canopy photosynthesis model through temporal and spatial scaling for remote sensing applications. *Ecol Model* 124:99–119
- Ciais P et al (2010) The European carbon balance. Part 2: croplands. *Glob Change Biol* 16:1409–1428
- de La Casinie're A, Bokoye AI, Cabot T (1997) Direct solar spectral irradiance measurements and updated simple transmittance models. *J Appl Meteorol* 36:509–520
- de Noblet-Ducoudré N, Gervois S, Ciais P, Biovy N, Brissson N, Seguin B, Perrier A (2004) Coupling the soil-vegetation atmosphere transfer scheme ORCHIDEE to the agronomy model STICS to study the influence of croplands on the European carbon and water budgets. *Agronomie* 24:397–407
- Denning, A. S., et al. (2005) Science Implementation strategy for the North American Carbon Program. Available online at <http://www.nacarbon.org>
- Dietz MC et al (2012) Characterizing the performance of ecosystem models across time scales: a spectral analysis of the North American Carbon Program site-level synthesis. *Journal of Geophysical Research: Biogeosciences* 116: G04029. doi:10.1029/2011JG001661
- El Maayar M, Price DT, Black TA, Humphreys ER, Jork EM (2002) Sensitivity tests of the integrated biosphere simulator to soil and vegetation characteristics in a pacific coastal coniferous forest. *Atmos Ocean* 40:313–332
- ERS USDA (2010) Corn, briefing rooms of economic research service of the United States Department of Agriculture. Available at <http://www.ers.usda.gov/Briefing/Corn/>
- FAO (2010) FAOSTAT, Food and Agriculture Organization of the United Nations. Available at <http://faostat.fao.org/site/567/DesktopDefault.aspx?PageID=567#ancor>
- Fischer ML, Billesbach DP, Riley WJ, Berry JA, Torn MS (2007) Spatiotemporal variations in growing season exchanges of CO<sub>2</sub>, H<sub>2</sub>O, and sensible heat in agricultural fields of the southern Great Plains. *Earth Interact.* 11:1–21
- Foken T (2008) The energy budget closure: an overview. *Ecol Appl* 18:1351–1367
- Foley JA, Prentice IC, Ramankutty N, Levis S, Pollard D, Sitch S, Haxeltine A (1996) An integrated biosphere model of land surface processes, terrestrial carbon balance, and vegetation dynamics. *Glob Biogeochem Cycles* 10:603–623
- Frolking SE et al (1998) Comparison of N<sub>2</sub>O emissions from soils at three temperate agricultural sites: simulations of year-round measurements by four models. *Nutr Cycl Agroecosys* 52:77–105
- Grant RF, Arain A, Arora V, Barr A, Black TA, Chen J, Wang S, Yuan F, Zhang Y (2005) Intercomparison of techniques to model high temperature effects on CO<sub>2</sub> and energy exchange in temperate and boreal coniferous forests. *Ecol Model* 188:217–252
- Grant RF, Arkebauer TJ, Dobermann A, Hubbard KG, Schimelfenig TT, Suyker AE, Verma SB, Walters DT (2007a) Net biome productivity of irrigated and rain-fed maize—soybean rotations: modelling vs. measurements. *Agron. J.* 99:1404–1423
- Grant RF, Barr AG, Black TA, Iwashita H, Kidson J, McCaughey H, Morgenstern K, Murayama S, Nesic Z, Saigusa N, Shashkov A, Zha T (2007b) Net ecosystem productivity of boreal jack pine stands regenerating from clearcutting under current and future climates. *Glob Change Biol* 13:1423–1440
- Graven HD et al (2013) Enhanced seasonal exchange of CO<sub>2</sub> by northern ecosystems since 1960. *Science* 341(6150):1085–1089
- Gray JM, Frolking S, Kort EA, Ray DK, Kucharik CJ, Ramankutty N, Friedl MA (2014) Direct human influence on atmospheric CO<sub>2</sub> seasonality from increased cropland productivity. *Nature* 515(7527):398–401
- Guanter L et al (2014) Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence. *Proc Natl Acad Sci* 111(14):E1327–E1333
- Hanson PJ, Amthor JS, Wullschlegel SD 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
- Izaurrealde RC, Williams JR, McGill WB, Rosenberg NJ, Quiroga Jakas MC (2006) Simulating soil C dynamics with EPIC: model description and testing against long-term data. *Ecol Model* 192:362–384
- Jain AK, West TO, Yang X, Post WM (2005) Assessing the impact of changes in climate and CO<sub>2</sub> on potential carbon sequestration in agricultural soils. *Geophys Res Lett* 32. doi:10.1029/2005GL023922
- Kothavala Z, Arain MA, Black TA, Verseghy D (2005) Evaluating fluxes of energy, water vapour and carbon dioxide over common crops. *Agric For Meteorol* 133:89–108
- Krinner G, Biovy N, de Noblet-Ducoudre N, Ogee 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.* 19:GB1015
- Kucharik CJ (2003) Evaluation of a process-based agroecosystem model (Agro-IBIS) across the U.S. cornbelt: simulations of the inter-annual variability in maize yield. *Earth Interact.* 7:1–33
- Kucharik CJ, Twine TE (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
- Li C, Frolking S, Frolking TA (1992) A model of nitrous oxide evolution from soil driven by rainfall events: 1. Model structure and sensitivity. *J Geophys Res* 97:9759–9776
- Liu J, Chen JM, Cihlar J, Chen W (1999) Net primary productivity distribution in the BOREAS study region from a process model driven by satellite and surface data. *J Geophys Res* 104(D22):27735–27754
- Liu S, Bliss N, Sundquist E, Huntington T (2003) Modelling carbon dynamics in vegetation and soil under the impact of soil erosion and deposition. *Glob Biogeochem* 17:1074. 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
- Lokupitiya E, Paustian K, Easter M, Williams S, Andr n O, K tterer T (2012) Carbon balance in US croplands during the last two decades of the 20th century. *Biogeochemistry* 107:207–225
- Martre P et al (2015) Multimodel ensembles of wheat growth: many models are better than one. *Glob Change Biol* 21:911–925
- Medvigy D, Wofsy SC, Munger JW, Hollinger DY, Poulton PR, Melillo JM, Borchers J, Chaney J et al (1995) Vegetation ecosystem modeling and analysis project—comparing biogeography and biogeochemistry models in a continental-scale study of terrestrial ecosystem responses to climate-change and CO<sub>2</sub> doubling. *Global Biogeochem* 9:407–437
- Miles NL, Richardson SJ, Davis KJ, Lauvaux T, Andrews AE, West TO, Bandaru V, Crosson ER (2012) Large amplitude spatial and temporal gradients in atmospheric boundary layer CO<sub>2</sub> mole fractions detected with a tower-based network in the U.S. upper Midwest. *J Geophys Res* 117:G01019. doi:10.1029/2011JG001781
- Moorcroft PR, Hurtt GC, Pacala SW (2001) A method for scaling vegetation dynamics: the ecosystem demography model (ED). *Ecol Monogr* 71:557–586
- Ogle SM, Davis K, Lauvaux T, Schuh A, Cooley D, West TO, Heath LS, Miles NL, Richardson S, Jay Breidt F, Smith JE, McCart JL, Gurney KR, Tans P, Denning AS (2015) An approach for verifying biogenic greenhouse gas emissions inventories with atmospheric CO<sub>2</sub> concentration data. *Environ Res Lett* 10(2015):034012. doi:10.1088/1748-9326/10/3/034012
- Ren W, Tian H, Liu M, Zhang C, Chen G, Pan S, Felzer B, Xu X (2007) Effects of tropospheric ozone pollution on net primary productivity and carbon storage in terrestrial ecosystems of China. *J. Geophys Res* 112:D22S09. doi:10.1029/2007JD008521
- Ricciuto DM, Thornton PE, Schaefer K, Cook RB, Davis KJ (2009) How uncertainty in gap-filled meteorological input forcing at eddy covariance sites impacts modeled carbon and energy flu., *Eos Trans. AGU*, 90(52) Fall Meet. Suppl., Abstract B54A-03
- Ricciuto DM, Schaefer K, Thornton PE, Davis K, Cook RB, Liu S, Anderson R, Arain MA, Baker I, Chen JM, Dietze M, Grant R, Izaurrealde C, Jain AK, King AW, Kucharik C, Liu S, Lokupitiya E, Luo Y, Peng C, Poulter B, Price D, Riley W, Sahoo A, Tian H, Tonitto C, Verbeeck H (2013) NACP Site: terrestrial biosphere model and aggregated flux data in standard format, data set. Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA. doi:10.3334/ORNLDAAAC/1183
- Running SW, Hunt ER Jr (1993) Generalization of a forest ecosystem process model for other biomes, BIOME-BGC, and an application for global-scale models. In: Ehleringer JR, Field C (eds) *Scaling physiological processes: leaf to globe*. Academic Press, San Diego, pp 141–158
- Ryan MG, McMurtrie RE,  gren GI, Hunt ER Jr, Aber JD, Friend AD, Rastetter EB, Pulliam WJ (1996) Comparing models of ecosystem function for coniferous forests. II. Predictions of response to changes in atmospheric CO<sub>2</sub> and climate. In: Breymer AI, Hall DO,  gren GI, Melillo JM (eds) *Global change: effects on coniferous forests and grasslands (SCOPE)*. Wiley, New York, pp 363–387
- Saleska SR et al (2003) Carbon fluxes in old-growth Amazonian rainforest: seasonality and disturbance-induced net carbon loss. *Science* 302:1554–1557
- Schaefer K, Collatz GJ, Tans P, Denning AS, Baker I, Berry J, Prihodko L, Suits N, Philpott A (2008) Combined simple biosphere/Carnegie-Ames-Stanford Approach terrestrial carbon cycle model. *J Geophys Res* 113:G03034. doi:10.1029/2007JG000603
- Schuh A, Lauvaux T, Denning A, West T, Davis K, Miles N, Richardson S, Uliasz M, Lokupitiya E, Cooley D, Andrews A, Ogle SM (2013) Evaluating atmospheric CO<sub>2</sub> inversions at multiple scales over a highly-inventoried agricultural landscape. *Glob Change Biol* 19:1424–1439
- Schwalm CR et al (2010) A model-data intercomparison of CO<sub>2</sub> exchange across North America: results from the North American carbon program site synthesis. *J Geophys Res* 115:G00H05. doi:10.1029/2009JG001229
- Semenov MA, Wolf J, Evans LG, Eckersten H, Iglesias A (1996) Comparison of wheat simulation models under climate change. 2. Application of climate change scenarios. *Clim Res* 7:271–281
- Stoy PC et al (2013) Evaluating the agreement between measurements and models of net ecosystem exchange at different times and timescales using wavelet coherence: an example using data from the North American Carbon Program Site-Level Interim Synthesis. *Biogeosciences* 10:6893–6909
- Sun J, Peng C, McCaughey H, Zhou X, Thomas V, Berninger F, St-Onge B, Hua D (2008) Simulating carbon exchange of Canadian boreal forests: II. Comparing the carbon budgets of a boreal mixedwood stand to a black spruce forest stand. *Ecol Model* 219:276–286
- Suyker AE, Verma SB (2008) Interannual water vapor and energy exchange in an irrigated maize-based agroecosystem. *Agric For Meteorol* 148(3):417–427
- Suyker AE, Verma SB, Burba GG, Arkebauer TJ, Walters DT, Hubbard KG (2004) Growing season carbon dioxide exchange in irrigated and rainfed maize. *Agric For Meteorol* 124(1–2):1–13
- Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res* 106:7183–7192
- Thornton P, Law BE, Gholz HL, Clark KL, Falge E, Ellsworth DS, Goldstein AH, Monson RK, Hollinger D, Falk M, Chen J, Sparks JP (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
- Tian HQ, Xu X, Zhang C, Ren W, Chen G, Liu M, Lu D, Pan S (2008) Forecasting and assessing the large-scale and long-term impacts of global environmental change on terrestrial ecosystems in the United States and China. In: Miao S, Carstenn S, Nungesser M (eds) *Real world ecology: large-scale and long-term case studies and methods*. Springer-Verlag, New York
- Tucker CJ, Pinzon JE, Brown ME, Slayback DA, Pak EW et al (2005) An extended AVHRR 8-km NDVI dataset

- compatible with MODIS and SPOT vegetation NDVI data. *Int J Remote Sens* 26:4485–4498
- Twine TE, Kustas WP, Norman JM, Cook DR, Houser PR, Meyers TP, Prueger JH, Starks PJ, Wesely ML (2000) Correcting eddy-covariance flux underestimates over a grassland. *Agric For Meteorol* 103:279–300
- Verma SB, Dobermann A, Cassman KG, Walters DT, Knops JM, Arkebauer TJ, Suyker AE, Burba GG, Amos B, Yang H, Ginting D, Hubbard KG, Gitelson AA, Walter-Shea EA (2005) Annual carbon dioxide exchange in irrigated and rain-fed maize-based agroecosystems. *Agric For Meteorol* 131:77–96
- Weng E, Luo Y (2008) Soil hydrological properties regulate grassland ecosystem responses to multifactor global change: a modeling analysis. *J Geophys Res* 113. doi:10.1029/2007JG000539
- Williams JR (1995) The EPIC model. In: Singh VP (ed) *Computer Models of Watershed Hydrology*. Water Resources Publications, Highlands Ranch, pp 909–1000
- Wilson K, Goldstein A, Falge E, Aubinet M, Baldocchi D, Berbigier P, Bernhofer C, Ceulemans R, Dolman H, Field C, Grelle A, Ibrom A, Law BE, Kowalski A, Meyers T, Moncrieff J, Monson R, Oechel W, Tenhunen J, Valentini R, Verma S (2002) Energy balance closure at FLUXNET sites. *Agric For Meteorol* 113:223–243
- Xiao JF, Zhuang QL, Baldocchi DD, Law BE, Richardson AD, Chen JQ, Oren R, Starr G, Noormets A, Ma SY, Verma SB, Wharton S, Bolstad PV, Burns SP, Cook DR, Curtis PS, Drake BG, Falk M, Foster DR, Gu LH, Hollinger DY, Katul GG, Matamala R, Monson RK, Munger JW, Sun KTPUG, Tom MS (2008) Estimation of net ecosystem carbon exchange for the conterminous United States by combining MODIS and AmeriFlux data. *Agric For Meteorol* 148:1827–1847
- Xue Y, Sellers PJ, Kinter JL III, Shukla J (1991) A simplified biosphere model for global climate studies. *J Climate* 4:345–364
- Zeng N, Zhao F, Collatz GJ, Kalnay E, Salawitch RJ, West TO, Guanter L (2014) Agricultural Green Revolution as a driver of increasing atmospheric CO<sub>2</sub> seasonal amplitude. *Nature* 515(7527):394–397
- Zhan X, Xue Y, Collatz GJ (2003) An analytical approach for estimating CO<sub>2</sub> and heat fluxes over the Amazonian region. *Ecol Model* 162:97–117
- Zhou X, Peng C, Dang Q-L, Sun J, Wu H, Hua D (2008) Simulating carbon exchange in Canadian Boreal forests: I. Model structure, validation, and sensitivity analysis. *Ecol Model* 219:287–299