Benchmarking the seasonal cycle of CO 2 fluxes simulated by terrestrial ecosystem models

Shushi Peng, Philippe Ciais, Frédéric Chevallier, Philippe Peylin, Patricia Cadule, Stephen Sitch, Shilong Piao, Anders Ahlström, Chris Huntingford, Peter Levy, et al.

AGU PUBLICATIONS

Global Biogeochemical Cycles

RESEARCH ARTICLE
10.1002/2014GB004931

Key Points:
- Seasonality of CO_2 fluxes from terrestrial ecosystem models is evaluated
- Most models overestimate the seasonal amplitude of GPP
- Growing season net uptake is 7.9 ± 2.0 Pg C yr^{-1} for the northern Hemisphere

Supporting Information:
- Readme
- Figures S1–S12

Correspondence to:
S. Peng,
Shushi.Peng@lsce.ipsl.fr

Citation:
Peng, S., et al. (2015), Benchmarking the seasonal cycle of CO_2 fluxes simulated by terrestrial ecosystem models, *Global Biogeochem. Cycles*, *29*, 46–64, doi:10.1002/2014GB004931.

Received 4 JUL 2014
Accepted 26 NOV 2014
Accepted article online 28 NOV 2014
Published online 12 JAN 2015

Benchmarking the seasonal cycle of CO_2 fluxes simulated by terrestrial ecosystem models

Shushi Peng[1,2], Philippe Ciais[1,3], Frédéric Chevallier[1], Philippe Peylin[1], Patricia Cadule[1], Stephen Sitch[4], Shilong Piao[3], Anders Ahlström[5], Chris Huntingford[6], Peter Levy[7], Xiran Li[3], Yongwen Liu[3], Mark Lomas[8], Benjamin Poulter[1,9], Nicolas Viovy[1], Tao Wang[1], Xuhui Wang[3], Sönke Zaehle[10], Ning Zeng[11], Fang Zhao[11], and Hongfang Zhao[3]

[1]Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ, Gif-sur-Yvette, France, [2]Laboratoire de Glaciologie et Géophysique de l'Environnement, UMR5183, CNRS and UJF Grenoble 1, Grenoble, France, [3]Sino-French Institute for Earth System Science, College of Urban and Environmental Sciences, Peking University, Beijing, China, [4]School of Geography, University of Exeter, Exeter, UK, [5]Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden, [6]Centre for Ecology and Hydrology, Wallingford, UK, [7]Centre for Ecology and Hydrology, Penicuik, UK, [8]Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK, [9]Institute on Ecosystems and Department of Ecology, Montana State University, Bozeman, USA, [10]Max Planck Institute for Biogeochemistry, Jena, Germany, [11]Department of Atmospheric and Oceanic Science, University of Maryland, College Park, Maryland, USA

Abstract We evaluated the seasonality of CO_2 fluxes simulated by nine terrestrial ecosystem models of the TRENDY project against (1) the seasonal cycle of gross primary production (GPP) and net ecosystem exchange (NEE) measured at flux tower sites over different biomes, (2) gridded monthly Model Tree Ensembles-estimated GPP (MTE-GPP) and MTE-NEE obtained by interpolating many flux tower measurements with a machine-learning algorithm, (3) atmospheric CO_2 mole fraction measurements at surface sites, and (4) CO_2 total columns (X_{CO2}) measurements from the Total Carbon Column Observing Network (TCCON). For comparison with atmospheric CO_2 measurements, the LMDZ4 transport model was run with time-varying CO_2 fluxes of each model as surface boundary conditions. Seven out of the nine models overestimate the seasonal amplitude of GPP and produce a too early start in spring at most flux sites. Despite their positive bias for GPP, the nine models underestimate NEE at most flux sites and in the Northern Hemisphere compared with MTE-NEE. Comparison with surface atmospheric CO_2 measurements confirms that most models underestimate the seasonal amplitude of NEE in the Northern Hemisphere (except CLM4C and SDGVM). Comparison with TCCON data also shows that the seasonal amplitude of X_{CO2} is underestimated by more than 10% for seven out of the nine models (except for CLM4C and SDGVM) and that the MTE-NEE product is closer to the TCCON data using LMDZ4. From CO_2 columns measured routinely at 10 TCCON sites, the constrained amplitude of NEE over the Northern Hemisphere is of 1.6 ± 0.4 gC m^{-2} d^{-1}, which translates into a net CO_2 uptake during the carbon uptake period in the Northern Hemisphere of 7.9 ± 2.0 PgC yr^{-1}.

1. Introduction

Terrestrial ecosystem models are used to assess the impacts of climate change on ecosystems [e.g., *Ciais et al.*, 2005; *Friedlingstein et al.*, 2006; *Sitch et al.*, 2008; *Piao et al.*, 2013]. However, model results are still plagued by large uncertainties, as evidenced by their spread in intercomparison exercises [*Friedlingstein et al.*, 2006; *Sitch et al.*, 2008; *Le Quere et al.*, 2009; *Todd-Brown et al.*, 2013; *Graven et al.*, 2013]. Comparison of terrestrial ecosystem models against observation helps to document their systematic errors for current conditions [e.g., *Cadule et al.*, 2010; *Randerson et al.*, 2009; *Maignan et al.*, 2011; *Zhao et al.*, 2012; *Luo et al.*, 2012; *Kelley et al.*, 2013] and can also give heuristic constraints about their projections for the future [*Cox et al.*, 2013; *Stegehuis et al.*, 2013].

Terrestrial CO_2 fluxes vary temporally from diurnal, seasonal, interannual, and longer time scales and spatially across climate and vegetation zones [*Keeling et al.*, 1996; *Heimann et al.*, 1998]. Eddy covariance measurements of continuous ecosystem-atmosphere CO_2 exchange are available for a large number of sites globally. These local data are a valuable tool for model development and validation. For example, *Stoeckli et al.* [2008]] used 15 flux sites to identify terrestrial water storage and carbon-nitrogen and deficiencies in CLM3.0 and showed improvement of those processes in CLM3.5. *Randerson et al.* [2009] used 74 flux sites across different ecosystems (e.g., tundra, needleleaf/broadleaf evergreen/deciduous forests) from Ameriflux sites to evaluate

the performance of two models (CASA and CLM-CN). *Richardson et al*. [2012] showed a systematic bias toward too early photosynthesis for an ensemble of 14 terrestrial ecosystem models evaluated at 10 North American flux tower sites. Systematic errors in modeling the onset of GPP were shown to propagate into errors in the phase of NEE by *Wang et al*. [2012]. It is interesting to test whether a bias of the phase of GPP attenuates (i.e., compensated by a bias of respiration of opposite sign) or amplifies bias in the phase of NEE, especially for models that couple GPP with respiration on subannual time scales though short-lived carbon pools.

At large spatial scale, the measurements of atmospheric CO_2 mole fractions are valuable data to test NEE from ecosystem models (which allow propagating these fluxes through atmospheric inversion). *Nemry et al*. [1999] used an atmospheric transport model and 25 atmospheric CO_2 monitoring stations to evaluate NEE across eight terrestrial biosphere models and showed that the simulated seasonality of atmospheric CO_2 was poor in the Southern Hemisphere. *Cadule et al*. [2010] evaluated three coupled climate carbon cycle models based on a series of metrics qualifying skills of model for the long-term trend, the seasonal cycle, the interannual variability of atmospheric CO_2, and the sensitivity of atmospheric CO_2 growth rate to climate variability. The seasonal cycle of CO_2, dominated by terrestrial exchange, provides an integrated constraint about the phase and amplitude of NEE in the Northern Hemisphere. *Cadule et al*. [2010] found that the seasonal CO_2 amplitude at Mauna Loa (MLO) was underestimated in two out of the three coupled models. *Kelley et al*. [2013] found an earlier atmospheric CO_2 drawdown and timing of minimum for three terrestrial ecosystem models at Northern Hemisphere stations.

Besides CO_2 measurements made at the ground-level, ground-based remote sensing of column-averaged dry air mole fraction of CO_2 (X_{CO2}) from the Total Carbon Column Observing Network (TCCON) [*Wunch et al*., 2011] were also used to evaluate the large-scale seasonality of NEE from terrestrial ecosystem models by *Yang et al*. [2007] and *Messerschmidt et al*. [2013]. TCCON data have also been used to optimize the NEE from the ORCHIDEE model using an atmospheric inversion [*Chevallier et al*., 2011]. The seasonality of TCCON X_{CO2} is sensitive to the seasonality of NEE at continental to hemispheric scales, while the seasonal cycle of CO_2 in the boundary layer CO_2 is sensitive to both regional NEE fluxes and continental to hemispheric fluxes [*Kaminski et al*., 1996; *Keppel-Aleks et al*., 2012].

In this study, we used (1) local eddy covariance CO_2 fluxes from selected sites as in *Randerson et al*. [2009], (2) large-scale gridded products MTE-GPP and MTE-NEE obtained from a set of eddy covariance measurements (253 sites) by a machine-learning algorithm [*Jung et al*., 2011], (3) atmospheric CO_2 mole fraction in the boundary layer; and (4) X_{CO2} from TCCON [*Wunch et al*., 2011] to test the seasonal cycle of GPP and NEE from nine terrestrial ecosystem models run in the TRENDY project (dgvm.ceh.ac.uk) [*Sitch et al*., 2013]. The testing of TRENDY models NEE with (3) and (4) involves the use of a 3-D atmospheric tracer transport model to transform NEE into 4-D CO_2 concentration fields. To this aim we used the LMDZ4 transport model from the Laboratoire de Météorologie Dynamique [*Hourdin et al*., 2006]. We combined the various sets of model-data misfits by correlating the seasonal biases of the terrestrial models for GPP and NEE from (1) and (2) with the biases of atmospheric CO_2 simulated by each terrestrial model coupled to the LMDZ4 atmospheric model in (3) and (4). We investigated the correlation of model biases across the various evaluation data sets and scales.

In the TRENDY ensemble, three models calculate the seasonality of NEE (using hereafter the definition of *Hayes* et al. [2012]) from the imbalance between GPP and ecosystem respiration, five models additionally simulate explicitly fire emissions, and two models have a simple parameterization of biomass harvest and its subsequent respiration flux. In the real world, the seasonal cycle of NEE (at large scale) is due to these processes but also accounts for the seasonality of CO_2 outgassing by rivers, lakes and estuaries [e.g., *Raymond et al*., 2000], and of CO_2 emitted by products decaying in landfills and waste, and other processes such as carbonate formation and dissolution [e.g., *Roland et al*., 2013] and emission of reduced carbon species that oxidize to CO_2 in the atmosphere through chemical reactions [*Folberth et al*., 2005]. The seasonal amplitude of these NEE components is not known, some of these fluxes are small, and they were not accounted for in our study. Further, when comparing with the seasonality of atmospheric CO_2, uncertainty on the seasonality of air-sea fluxes and fossil fuel CO_2 emissions also adds to uncertain atmospheric transport as a source of bias in testing NEE [e.g., *Peylin et al*., 2011].

In section 2, the models, data sets, and evaluation metrics are described. The results of the comparison for GPP and NEE between models and flux measurements are shown in section 3. In section 4, modeled GPP and

Table 1. Nine TRENDY Models (http://www-lscedods.cea.fr/invsat/RECCAP/V2)

Model Name	Abbreviation	Spatial Resolution	Period	Reference
Community Land Model 4C	CLM4C	2.5° × 1.875°	1901–2009	*Oleson et al.* [2010] and *Lawrence et al.* [2011]
Community Land Model 4C	CLM4CN	2.5° × 1.875°	1901–2009	*Oleson et al.* [2010] and *Lawrence et al.* [2011]
Lund-Postdam-Jena	LPJ	0.5° × 0.5°	1901–2009	*Sitch et al.* [2003]
LPJ_GUESS	LPJ_GUESS	0.5° × 0.5°	1901–2009	*Smith et al.* [2001]
ORCHIDEE	ORCHIDEE	0.5° × 0.5°	1901–2009	*Krinner et al.* [2005]
ORCHIDEE-CN	OCN	3.75° × 2.5°	1901–2009	*Zaehle et al.* [2010] and *Zaehle et al.* [2010]
Sheffield-DGVM	SDGVM	3.75° × 2.5°	1901–2009	*Woodward et al.* [1995]
TRIFFID	TRI	3.75° × 2.5°	1901–2009	*Cox et al.* [2000]
VEGAS	VEGAS	0.5° × 0.5°	1901–2009	*Zeng et al.* [2005]

NEE in the Northern Hemisphere are compared with gridded MTE-GPP and MTE-NEE data products [*Jung et al.*, 2011]. Comparison between modeled and observed atmospheric CO_2 concentration is shown in section 5. Correlation of systematic errors attached to the seasonality of GPP and NEE across scales is discussed in section 6. Section 7 concludes this study.

2. Methods

2.1. Terrestrial Ecosystem Models

We evaluated monthly GPP and NEE (negative value indicates carbon uptake) simulated by the nine TRENDY models of Table 1 (TRENDY, http://dgvm.ceh.ac.uk/, accessed 11 July 2013). We used the configuration S2 of TRENDY, in which models are driven by rising atmospheric CO_2 concentration and climate change from 1901 to 2010, without land use change CO_2 emissions.

2.2. Observations

We selected 16 sites from the FLUXNET database (www.fluxdata.org) that span diverse vegetation types in temperate, boreal, and arctic regions over the Northern Hemisphere. The representativeness of a flux tower site with respect to the model processes to be evaluated is difficult to define. Since the TRENDY models do not account for local disturbance and site history, we have chosen FLUXNET data from ecosystems far from their last disturbance. The selected sites also have long records, in order to calculate a robust mean seasonal cycle of NEE and GPP. The location, vegetation type, climate type, and data available period of the 16 sites are listed in Figure 1 and Table 2. For each flux site, continuous half-hourly GPP and NEE (negative value indicate carbon uptake) during the measurement period have been quality controlled and gap filled using the method of *Papale et al.* [2006] and *Moffat et al.* [2007]. GPP from each flux site was separated from NEE with the algorithm developed by *Lasslop et al.* [2010], which combines temperature sensitivity of respiration and vapor pressure deficit limitation and light response of photosynthesis. The uncertainty of monthly GPP and NEE at flux sites are ~0.3 gC m^{-2} d^{-1} and ~0.4 gC m^{-2} d^{-1}, respectively [e.g., *Papale et al.*, 2006; *Lasslop et al.*, 2010].

We also use an observation-driven global monthly gridded GPP, NEE statistical model (hereafter referred to as MTE-GPP and MTE-NEE) derived from FLUXNET measurements (not restricted to the 16 sites of the previous paragraph but global 253 flux sites) by statistical upscaling [*Jung et al.*, 2011]. The statistical MTE model integrates remote-sensing indices, climate and meteorological data, and information on land cover. The resulting product has a spatial resolution of 0.5° × 0.5° and covers the period from 1982 to 2011. The typical uncertainty of MTE-GPP and MTE-NEE are ~46 gC m^{-2} yr^{-1} (5%) and ~37 gC m^{-2} yr^{-1} (28%), respectively, which are defined as the standard deviation of the 25 model tree ensembles (from the MTE algorithm) [*Jung et al.*, 2011].

It should be noted that the uncertainty of MTE-NEE is larger than that of MTE-GPP and that disturbance related CO_2 emissions in MTE-NEE calculation are not included. Further, MTE-NEE shows a large positive bias over the Northern Hemisphere or globe (NEE is 4.9 ± 0.4 PgC yr^{-1} north of 25°N) compared to the terrestrial CO_2 sink [*Peylin et al.*, 2013], and we do not know if this bias is constant or seasonally variable. We also used atmospheric CO_2 observations to evaluate NEE at large scale. We selected atmospheric CO_2 records from 15 Northern Hemisphere boundary layer sites from the NOAA global cooperative air sampling network (http://www.esrl.noaa.gov/gmd/ccgg/globalview/). The amplitude and phase of the seasonal cycle of atmospheric CO_2 relates to

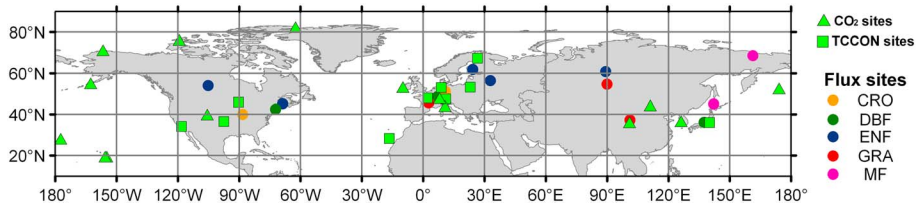

Figure 1. The location of 16 FLUXNET sites (filled circles with different colors indicate vegetation type), 15 flask and continuous atmospheric CO_2 measurement sites (green triangle), and 10 TCCON sites (green square).

NEE influencing each station through atmospheric transport, and the "footprint" range of each atmospheric station depends on the location of the measurement station and transport fields [*Cadule et al.*, 2010]. The locations and measurement periods of the atmospheric CO_2 mole fraction measurement sites are listed in Figure 1 and Table 3. The model fluxes are compared to CO_2 mole fraction measurements through the simulation of atmospheric transport by the LMDZ4 model. Total column CO_2 concentration (X_{CO2}) observations are obtained from 10 TCCON stations (Figure 1 and Table 4) [*Wunch et al.*, 2011]. X_{CO2} measurements are taken during clear-sky daytime, from the direct solar spectra [*Wunch et al.*, 2011]. The uncertainty of observed X_{CO2} from TCCON sites is less than 0.25% (defined as 1 standard deviation, less than 1 ppm) and 0.1% under clear-sky conditions [*Wunch et al.*, 2011].

2.3. Analysis

We aggregated half-hourly GPP and NEE measurements at each site into monthly mean values and computed the local mean seasonal cycle of GPP and NEE over the measurements (Table 2). GPP and NEE from all the plant functional types (PFT) present in a grid cell have been summed together in the TRENDY model output (i.e., PFT specific fluxes were not archived separately by modelers). To compare this PFT-weighted climatological-simulated GPP and NEE with each flux site, we first searched for the closest model grid cell dominated by the observed PFT of each site and then sampled the monthly modeled GPP and NEE during the same period of flux site measurements for this "closest grid cell" in each ecosystem model. The comparison between grid scale and observation from flux site suggests a systematic error other than vegetation cover difference or resolution [*Raczka et al.*, 2013]. For the Northern Hemisphere scale (>25°N), a comparison between TRENDY models and the MTE products was analyzed, using area-weighted GPP and NEE.

We used the 3-D atmospheric transport model LMDZ4 of *Hourdin et al.* [2006], nudged to horizontal winds analyzed by the European Centre for Medium-Range Weather Forecasts, to transport modeled NEE (if fire and harvest were included in models, then carbon flux from fire and harvest were also accounted for in NEE). Ocean flux and fossil fuel emission were included in the transport simulations to produce monthly CO_2 mole fractions at each station. Because NEE is monthly, covariance between NEE and transport at diurnal [*Denning et al.*, 1996] and synoptic time scales causing monthly CO_2 gradients, is not accounted for. The air-sea CO_2 flux prescribed to LMDZ4 is from the climatology *Takahashi et al.* [2009], and fossil fuel emission are from EDGARv4 (http://edgar.jrc.ec.europa.eu/index.php#) adjusted by Carbon Dioxide Information Analysis Center annual country totals [*Boden et al.*, 2012]. These emissions are assumed to have no seasonality and are all injected in the first model layer (thickness 70 m). The version of LMDZ4 that was used has a horizontal resolution of 3.75° longitude × 2.5° latitude and 19 vertical levels. After simulating 30 min time series of CO_2 with LMDZ4 applied to monthly NEE, we used equation (1) to generate a detrended monthly smoothed seasonal cycle of CO_2 composed of eight harmonics at each site (to capture the nonsymmetrical seasonal shape of CO_2 which has a long winter maximum and a shorter summer minimum).

$$Y = \sum_{i=0}^{2} p_i \times t^i + \sum_{i=3}^{6} p_i \times \sin\left((i-2) \times 2 \times \pi \times t\right) + \sum_{i=7}^{10} p_i \times \cos\left((i-6) \times 2 \times \pi \times t\right) \quad (1)$$

where Y is a smoothed CO_2 mole fraction including a trend, t is the time unit, and $p_0 - p_{10}$ are the regressed parameters. The seasonal cycle is defined by $p_3 - p_{10}$.

Table 2. Vegetation Type (Veg), Location (Latitude (Lat), Longitude (Lon)), Climatic Type, and Studied Periods of Flux Sites in This Study[a]

Site Name	Veg	Lat (°N)	Lon (°E)	Climate	MAT (°C)	MAP (mm)	Studied Period
CA-Obs	ENF	53.99	−105.12	Boreal	0.8	406	1999–2005
US-Ho1	ENF	45.20	−68.74	Temperate	5.3	1070	1996–2004
US-Ha1	DBF	42.54	−72.17	Temperate	6.6	1071	1991–2006
US-Bo1	CRO	40.01	−88.29	Temperate	11.0	991	1996–2007
US-Brw	WET	71.32	−156.63	Arctic	−12.8	86	1998–2002
FI-Hyy	ENF	61.85	24.29	Boreal	2.2	620	1996–2006
FR-Hes	DBF	48.67	7.06	Temperate	9.2	793	1997–2006
FR-Lq1	GRA	45.64	2.74	Temperate	8.3	730	2004–2006
DE-Geb	CRO	51.10	10.91	Temperate	8.7	444	2004–2006
RU-Fyo	ENF	56.46	32.92	Temperate	4.4	671	1998–2006
RU-Zot	ENF	60.80	89.35	Boreal	−3.3	943	2002–2004
RU-Ha1	GRA	54.73	90.00	Boreal	−0.1	592	2002–2004
RU-Che	MF	68.61	161.34	Boreal	−12.1	609	2002–2005
JP-Tak	DBF	36.15	137.42	Temperate	6.5	1024	1999–2004
JP-Tef	MF	45.06	142.11	Temperate	5.7	1000	2001–2005
CN-HaM	GRA	37.37	101.18	Arctic	−1.7	567	2002–2004

[a]ENF, DBF, MF, GRA, CRO, and WET stand for evergreen needleleaf forest, deciduous broadleaf forest, mix forest, grassland, cropland, and wetland, respectively.

2.4. Metrics

We compared the amplitude and phase of the seasonal cycle between observations and simulations. The climatological amplitude of the seasonal cycle in GPP, NEE, and atmospheric CO_2 (see equation (1) above) is defined as the difference between the maximum and minimum monthly values between January and December. The phase of the seasonality of the fluxes and of CO_2 concentration is computed by a fit to the cosine function.

$$Y = b_0 + \sum_{i=1}^{3} b_i \left(\frac{2^i \times \pi \times D - ph_i \times 2 \times \pi}{365} \right) \tag{2}$$

where Y is GPP, NEE, or CO_2 mole fraction, D is Julian date of the year, ph_1, ph_2 and ph_3 are the phases of the seasonality, and b_0 – b_3 are regressed parameters.

The amplitude bias of modeled GPP, NEE, or CO_2 mole fraction is defined as the difference between the amplitude of modeled and observed GPP, NEE, or CO_2 mole fraction. For the phase bias, we consider only the phase of the first harmonics (ph_1 in equation (2)) to show the date of seasonal peak in GPP or NEE. The phase bias of modeled GPP, NEE, or CO_2 mole fraction is defined as the difference between modeled and observed ph_1. We also used the mean bias (MB), root mean square error (RMSE), systematic RMSE ($RMSE_{sys}$), and unbiased RMSE ($RMSE_{unbias}$) to quantify differences between observations and model simulations.

$$\mathrm{MB} = \frac{\sum_{i=1}^{N} (m_i - o_i)}{N} \tag{3}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (m_i - o_i)^2}{N}} \tag{4}$$

$$\mathrm{RMSE_{sys}} = \sqrt{\frac{\sum_{i=1}^{N} (\hat{m}_i - o_i)^2}{N}} \tag{5}$$

$$\mathrm{RMSE_{unbias}} = \sqrt{\frac{\sum_{i=1}^{N} (\hat{m}_i - m_i)^2}{N}} \tag{6}$$

Table 3. The 15 Atmospheric CO_2 Mole Fraction Record Stations in This Study

Station name	Abbreviation	Lat (°N)	Lon (°E)	Altitude (m)	Studied period
Alert	ALT	82.45	−62.52	210	1985–2010
Barrow	BRW	71.32	−156.61	11	1971–2010
Cold Bay	CBA	55.21	−162.72	57	1978–2010
Mt. Cimone	CMN	44.18	10.70	2165	1979–2006
Cape Kumukahi	KUM	19.52	−154.82	8	1976–2010
Mould Bay	MBC	76.25	−119.35	58	1980–1997
Mace Head	MHD	53.33	−9.90	25	1992–2010
Sand Island, Midway	MID	28.21	−177.38	8	1985–2010
Mauna Loa	MLO	19.54	−155.58	3402	1969–2010
Niwot Ridge	NWR	40.05	−105.58	3526	1968–2010
Schauinsland	SCH	48.00	8.00	1205	1972–2004
Shemya Island	SHM	52.72	174.10	40	1985–2010
Tae-ahn Peninsula	TAP	36.73	126.13	20	1990–2010
Ulaan Uul	UUM	44.45	111.10	914	1992–2010
Mt. Waliguan	WLG	36.29	100.90	3810	1990–2010

where m_i and o_i are respectively the modeled and observed GPP, NEE, and atmospheric CO_2 in the ith month; $\hat{m}_i$ is the predicted value in the ith month by linear regression between modeled and observed GPP, NEE, and atmospheric CO_2 ($m_i = a^* o_i + b + \varepsilon$, where a and b are the regressed parameters, and ε is the regression error); N is set to 12 months.

3. Local Scale Results

3.1. Monthly GPP

3.1.1. Mean Bias of GPP

Figure 2 shows the model-observation comparison for GPP. MTE-GPP is also reported in the figure to show how this product compares with local measurements and models. CLM4C, CLM4CN, LPJ, LPJ_GUESS, and ORCHIDEE overestimate GPP for each month of the growing season at most sites, while SDGVM and VEGAS underestimate GPP. GPP of the five evergreen needleleaf forest sites are better simulated than the grass and cropland sites [*Vuichard et al.*, 2007].

Figure 3 summarizes the mean bias between modeled and observed site level GPP. The mean bias of modeled GPP in CLM4C, LPJ, and ORCHIDEE is positive at more than 12 flux sites out of 16. A positive mean bias of GPP is also found at 9–10 flux sites in CLM4CN, LPJ_GUESS, OCN, and TRI (Figure 3). In contrast, a negative mean bias of GPP for SDGVM and VEGAS is found at 13 and 12 out of the 16 flux sites, with the largest negative bias being found at forest sites (Figure 3).

Although the range of mean bias of GPP across models is large (−2.1–5.0 gC m^{-2} d^{-1}, larger than uncertainty of measurement), the mean bias of the model ensemble is smaller than 1 gC m^{-2} d^{-1} except at a deciduous broadleaf forest site (JP-Tak) and at two grassland sites (FR-Lq1 and RU-Ha1), showing a large diversity of model results without any systematic pattern among the models for GPP. By contrast, the mean bias of MTE-GPP at the 16 flux sites ranges from −0.8 gC m^{-2} d^{-1} to 0.9 gC m^{-2} d^{-1}, which is closer to measured GPP than the models (−2.6–5.0 gC m^{-2} d^{-1}). MTE-GPP still has a small average negative bias compared with the 16 flux sites (except at RU-Zot, JP-Tak, and US-Ha1 sites where the bias is positive). The median of the mean bias of GPP across the 16 flux sites is positive in seven out of the nine models in Table 1 (except SDGVM and VEGAS), with a value of −0.6 ± 0.9 gC m^{-2} d^{-1} (± standard deviation across the 16 flux sites) in VEGAS to 1.2 ± 1.3 gC m^{-2} d^{-1} in LPJ (Figure 3). GPP of the model average is closer to observed GPP than that of any of the nine models (median of the model average bias of 0.1 ± 0.8 gC m^{-2} d^{-1} across sites). In comparison, the median of the mean bias of MTE-GPP at the 16 flux sites is −0.2 ± 0.6 gC m^{-2} d^{-1}.

3.1.2. RMSE of Monthly GPP

The RMSE between modeled and measured GPP is generally smaller at evergreen needleleaf forests (ENF) (the first to fifth rows in Figure 3) than at other flux sites (Figure 3). The median RMSE of monthly GPP across the 16 flux sites is ~1.5 ± 1.0 gC m^{-2} d^{-1} in eight out of the nine models (except 2.9 ± 1.0 gC m^{-2} d^{-1} in LPJ). The $RMSE_{sys}$ of GPP has similar magnitude than RMSE of GPP. $RMSE_{unbias}$ of GPP is smaller than $RMSE_{sys}$ of GPP at most sites, except at the cropland site US-Bo1 (Figure S1 in the supporting information). This suggests

Table 4. The 10 TCCON Stations in This Study

Name	Country	Start Time	Lat	Lon	Altitude (km)
Bialystok	Poland	3/2009	53.23	23.03	0.18
Bremen	Germany	7/2004	53.10	8.85	0.027
Garmisch	Germany	7/2007	47.48	11.06	0.74
Izaña	Tenerife	5/2007	28.30	−16.50	2.37
JPL	USA	7/2007	34.202	−118.175	0.21
Lamont	USA	7/2008	36.60	−97.49	0.32
Orleans	France	9/2009	47.97	2.11	0.13
Park	USA	5/2004	45.95	−90.27	0.44
Sodankylä	Finland	1/2009	67.37	26.63	0.18
Tsukuba	Japan	12/2008	36.05	140.12	0.03

that the nine models capture the shape of the GPP seasonal variation well but with either a high or low bias. For the cropland sites or managed grassland sites, modeled GPP generally has a larger $RMSE_{unbias}$ (Figure S1), likely because the effects of management and other missing processes, such as grazing, specific crop and pasture phenology, irrigation, and fertilization are not represented in some TRENDY models [e.g., *Chang et al.*, 2013; *Valade et al.*, 2014; *Guanter et al.*, 2014].

3.1.3. Amplitude Bias of GPP

The amplitude bias of modeled GPP is positively correlated with the mean bias of GPP across the nine models at all sites (Figure S2). The median of the amplitude bias of GPP across sites ranges from -3.9 ± 2.8 gC m^{-2} d^{-1} in SDGVM to 1.2 ± 4.2 gC m^{-2} d^{-1} in LPJ (Figure S2). The median of the amplitude bias of GPP across models is of -1.6 ± 2.5 gC m^{-2} d^{-1}. MTE-GPP also underestimates the amplitude of GPP at 12 out of the 16 flux sites, with a median amplitude bias of -0.8 ± 1.7 gC m^{-2} d^{-1}.

3.1.4. Phase Bias of GPP

The median phase bias of GPP across the flux sites for the nine models is small, from -10 ± 15 days to 2 ± 15 days. Small phase biases (less than 10 days) are found at evergreen needleleaf forest sites for all nine models, except TRI (Figure S2). CLM4C, CLM4CN, LPJ_GUESS, ORCHIDEE, and SDGVM have a negative phase bias of GPP at most sites, related to their earlier start of photosynthesis and earlier-than-observed timing of peak GPP (Figure 2). The positive GPP phase bias of LPJ and VEGAS is related to a later-than-observed peak GPP, i.e., a GPP bias during the recession of the growing season, despite an earlier start of photosynthesis in these two models (Figure 2). The phase bias of MTE-GPP is less than 10 days at 14 out of 16 flux sites (except US-Brw and RU-Ha1), thus comparable to that of the process models. Models that overestimate GPP (CLM4C, CLM4CN, LPJ_GUESS, LPJ, and ORCHIDEE) generally have an earlier start and a later end of growing season at flux sites, consistent with the results of *Richardson et al.* [2012] using 14 models in North America, among which two models (LPJ and ORCHIDEE) are participating to TRENDY.

3.2. Monthly NEE

3.2.1. Mean Bias of NEE

Figure 4 shows the comparison between modeled and observed NEE at the 16 flux sites. Across all models and flux sites, the mean bias of NEE ranges from -1.0 gC m^{-2} d^{-1} to 1.3 gC m^{-2} d^{-1} (Figure 5). The mean bias of NEE is smaller than that of GPP. All models simulate larger (less carbon uptake) than observed NEE, i.e., a positive mean bias at 13 out of the 16 sites (except at RU-Fyo, ~ -0.5 gC m^{-2} d^{-1}; RU-Che, ~ -0.1 gC m^{-2} d^{-1}, and JP-Tef, ~ -0.9 gC m^{-2} d^{-1} where the NEE bias is negative) (Figure 5). The median of mean bias of NEE across the flux sites ranges from 0.3 ± 0.5 to 0.6 ± 0.5 gC m^{-2} d^{-1}. The median of mean bias of NEE across all models is of 0.5 ± 0.5 gC m^{-2} d^{-1}. In comparison, the mean bias of NEE for MTE-NEE is less than 0.3 gC m^{-2} d^{-1} at 10 out of the 16 flux sites, and the median of mean bias of NEE is 0.0 ± 0.5 gC m^{-2} d^{-1}. The positive mean bias of both GPP and NEE of seven models (except OCN and VEGAS) suggests a systematic overestimation of respiration, at least in the way the models were run and sampled to match local FLUXNET measurements. Note that equilibrium ecosystem carbon balance in TRENDY models (long-term mean NEE plus harvest and fire emissions for those models simulating these fluxes, in each grid point being zero) prevails in the year 1901 from the spin-up of each model. In contrast, forests where NEE is measured at the FLUXNET sites usually are in a growing phase and are net annual carbon sinks (unlike modeled mature forests subject only to CO_2 and climate forcing). For this reason, the TRENDY models likely underestimate carbon uptake (i.e., have a too small NEE) at

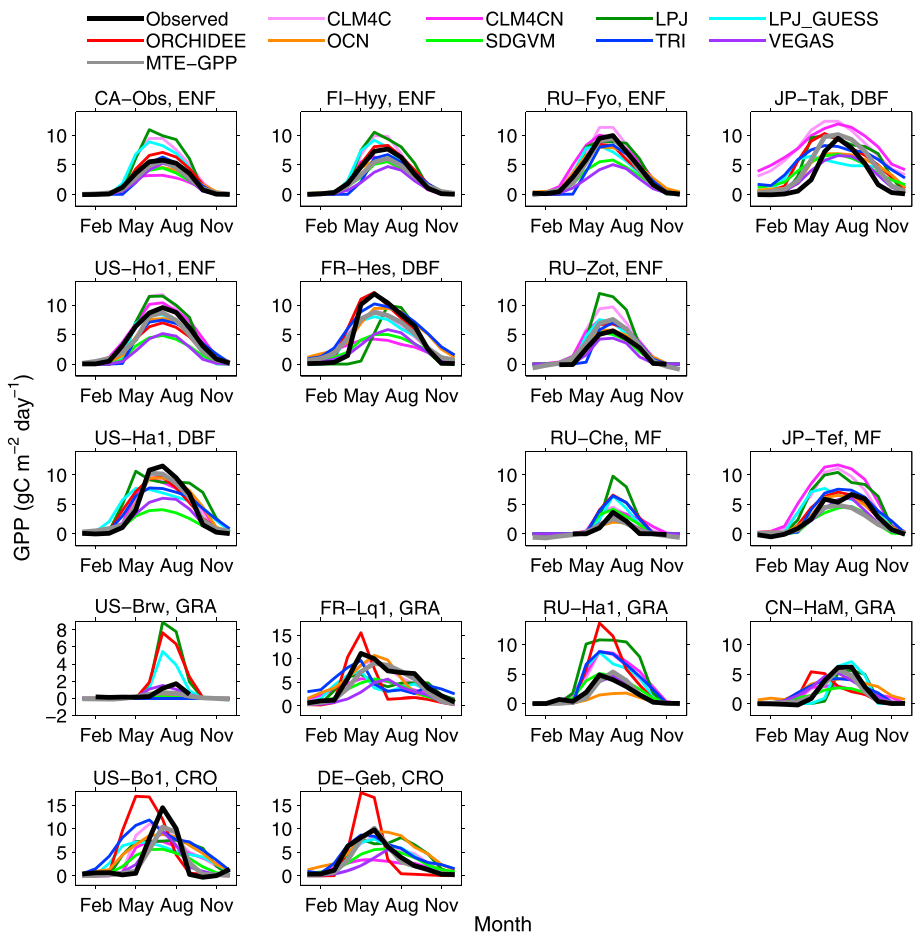

Figure 2. Observed (black), modeled and MTE-GPP at the 16 flux sites listed in Table 2. (from left to right) The first to fourth columns gather flux sites from North America, Europe, Russia, and East Asia, respectively. (from top to bottom) The first to third rows show forest flux sites, the fourth row shows grassland flux sites, and the fifth row shows cropland flux sites.

forest sites. There may also be a bias from the mix of different PFT present in a model grid cell when compared to mono-PFT site data, an error that could be reduced in future studies by archiving fluxes of each PFT in future comparison projects.

3.2.2. RMSE of Monthly NEE

The median RMSE of monthly NEE across sites is 0.9 ± 0.6–1.1 ± 0.6 gC m^{-2} d^{-1} for all the models and is close to the median RMSE of the average of the models (0.9 ± 0.7 gC m^{-2} d^{-1}). It is more than twice as good as the RMSE of daily NEE (see *Chevallier et al.* [2012], in the case of ORCHIDEE). The RMSE of monthly MTE-NEE is of similar magnitude, i.e., larger than 1.0 gC m^{-2} d^{-1} at seven out of the 16 flux sites. The median RMSE of MTE-NEE across sites is 0.9 ± 0.5 gC m^{-2} d^{-1} and is also close to the models' performance. Across models and flux sites, the $RMSE_{sys}$ of NEE ranges from 0.2 to 3.9 gC m^{-2} d^{-1} and the $RMSE_{unbias}$ of NEE ranges from 0.2 to 3.4 gC m^{-2} d^{-1}. The $RMSE_{sys}$ of NEE is thus almost equal to the $RMSE_{unbias}$ of NEE for most models (Figure S3, except OCN, TRI, and VEGAS). Overall, all models have similar magnitudes of NEE RMSE and do not perform well at simulating NEE at most flux sites.

Generally, RMSE of monthly NEE has a similar magnitude than RMSE of monthly GPP for all models. There is a significant positive correlation between median RMSE of GPP and median RMSE of NEE (Figures 6 and S5). Comparing the RMSE of GPP and of NEE for different PFTs, the models (from their median RMSE) have a lower RMSE of GPP and NEE for evergreen needleleaf forest than for deciduous broadleaf forest and cropland sites (Figure 6). There is also a positive correlation of biases between GPP and NEE, even though an overestimated GPP can be partially compensated by an overestimated respiration in some models like CLM4C, LPJ, LPJ_GUESS, and ORCHIDEE.

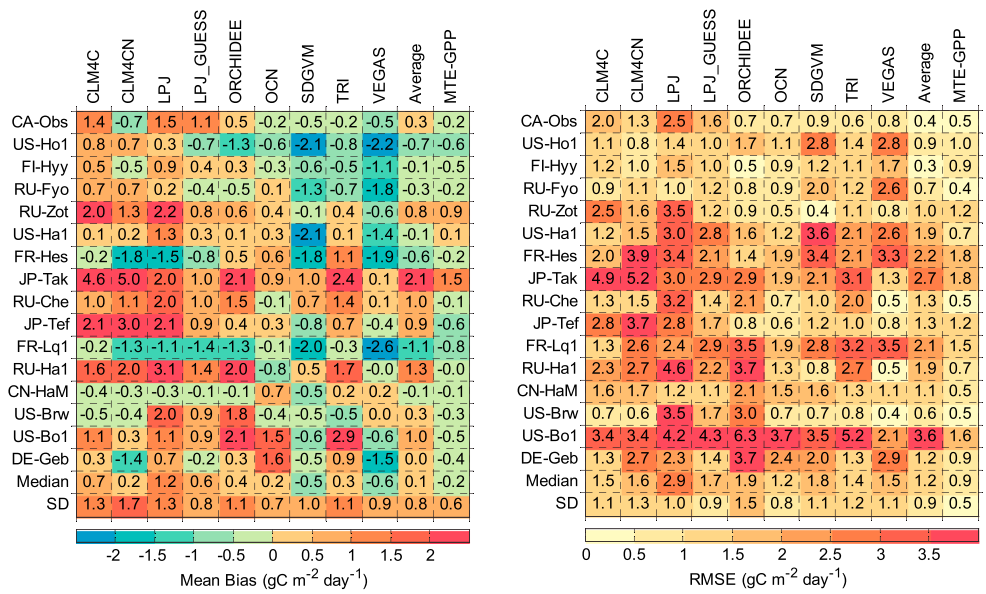

Figure 3. The mean bias and RMSE of monthly GPP at the 16 flux sites listed in Table 2. (left) The mean GPP bias and (right) the RMSE of monthly GPP. The last two rows in the two panels indicate the median and standard deviation (SD) of mean bias or RMSE of GPP across the 16 flux sites.

3.2.3. Amplitude Bias of NEE

The amplitude bias of NEE shows different patterns across sites and models (Figure S4). Across sites, the amplitude bias of NEE varies from -7.0 to 5.3 gC m^{-2} d^{-1}. The median amplitude bias of NEE across sites ranges from -1.9 ± 2.0 gC m^{-2} d^{-1} in TRI to 1.4 ± 2.6 gC m^{-2} d^{-1} in CLM4C. CLM4C, LPJ, ORCHIDEE, and SDGVM have a positive amplitude bias at more than eight sites (Figure S3), i.e., these models overestimate NEE in summer (too much uptake) and underestimate it in winter (too much respiration). The high bias of GPP in CLM4C, LPJ, and ORCHIDEE (Figure S2) may propagate to a bias in simulated NEE, causing an overestimated summer NEE (too little uptake). In contrast, SDGVM underestimates summer GPP and overestimates NEE (too much uptake), which indicates underestimated summer respiration together and overestimated winter respiration resulting in overestimated amplitude of NEE at most flux sites in SDGVM (Figure 4).

3.2.4. Phase Bias of NEE

The magnitude of phase bias in NEE at most sites is larger than 10 days. The median phase bias of NEE across sites is less than 10 days in five out of the nine models (Figure S4). LPJ_GUESS and SGDVM have a negative phase bias at most sites. The phase bias of NEE averaged across models is negative at most sites, and it is smaller at forest sites than that at cropland and grassland sites (Figure S4). The median phase bias of average modeled NEE across models is of -6 ± 40 days. By comparison, the median phase bias of MTE-NEE is of 5 ± 40 days.

4. Hemisphere-Scale Results

4.1. Average GPP

Figure 7 shows the mean seasonal cycle of GPP north of 25°N compared with MTE-GPP as a reference. Seven out of the nine models simulate GPP larger than MTE-GPP during spring and autumn (except TRI and VEGAS), which is consistent with the comparison with FLUXNET data at the site level. Nine out of nine models (except TRI and VEGAS) simulate a larger mean monthly GPP than MTE-GPP (Figure 7). The mean bias of modeled GPP compared with MTE-GPP varies from -0.4 gC m^{-2} d^{-1} in VEGAS to 0.6 gC m^{-2} d^{-1} in LPJ. Most models underestimate the amplitude of GPP compared with MTE-GPP except LPJ (Figure 7).

4.2. Average NEE

All models simulate a higher-average NEE compared with MTE-NEE by about 0.2/0.1 gC m^{-2} d^{-1} (models have less uptake than MTE, consistent with the finding that the mean value of NEE is of 0.09 ± 0.02 gC m^{-2} d^{-1} north

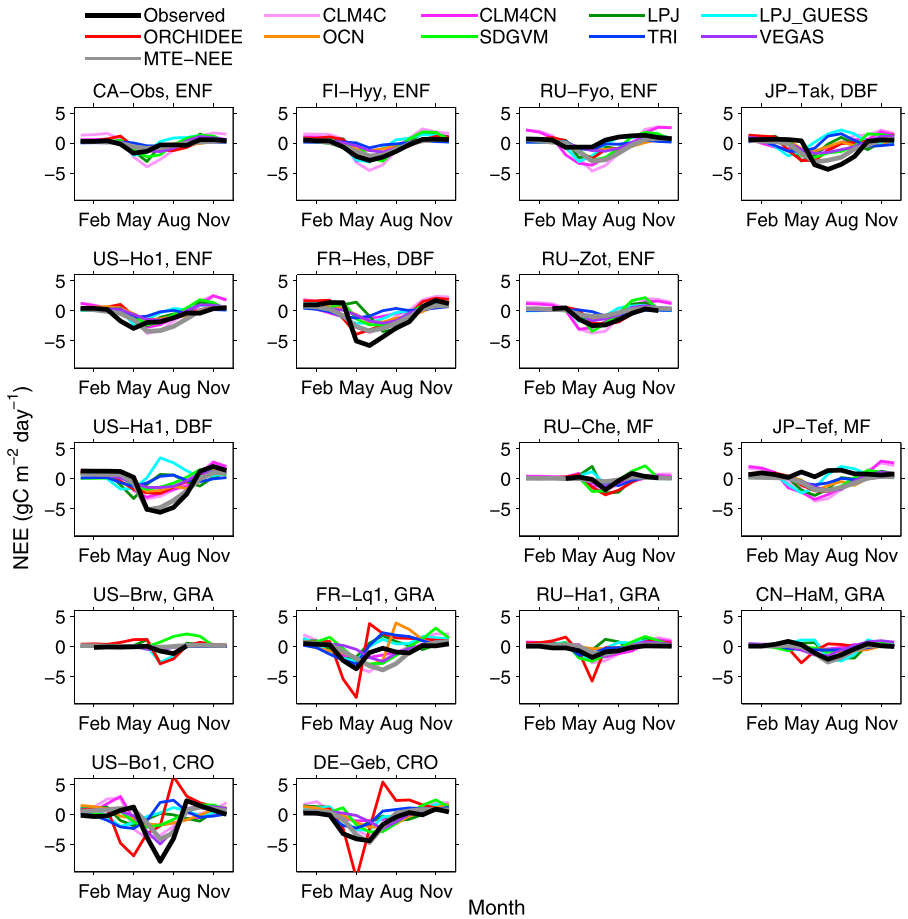

Figure 4. Observed (black), modeled, and MTE-NEE at the 16 flux sites listed in Table 2. Negative NEE indicates carbon uptake. (from left to right) The first to fourth columns gather flux sites from North America, Europe, Russia, and East Asia, respectively. (from top to bottom) The first to third rows show forest flux sites, the fourth row shows grassland flux sites, and the fifth row shows cropland flux sites.

of 25°N [*Peylin et al.*, 2013]). Most models simulate a higher NEE than MTE-NEE both in summer (too little uptake) and winter (overestimated winter respiration). The difference in minimum monthly NEE between the models and MTE-NEE is positive in seven out of nine models (except CLM4C and SDGVM), with a range from $-0.4\,\mathrm{gC\,m^{-2}\,d^{-1}}$ in SDGVM to $0.7\,\mathrm{gC\,m^{-2}\,d^{-1}}$ in LPJ and TRI. CLM4C and SDGCM simulate larger amplitude of NEE than MTE-NEE, by $0.7\,\mathrm{gC\,m^{-2}\,d^{-1}}$ and $1.1\,\mathrm{gC\,m^{-2}\,d^{-1}}$ respectively (Figure 7). In contrast, the other seven models simulate smaller amplitude than MTE-NEE over the entire northern domain (-0.9 to $-0.1\,\mathrm{gC\,m^{-2}\,d^{-1}}$).

4.3. Spatial Patterns of GPP

The spatial patterns of GPP in LPJ, LPJ_GUESS, and ORCHIDEE show higher GPP than MTE-GPP in most areas of the Northern Hemisphere (Figure S6). The other six models (CLM4C, CLM4CN, OCN, SDGVM, TRI, and VEGAS) have mixed positive and negative regional GPP biases compared with MTE-GPP (Figure S6). For GPP amplitude, CLM4C, CLM4CN, OCN, SDGVM, and VEGAS simulate a lower GPP amplitude than MTE-GPP over Siberia. LPJ, LPJ_GUESS, and ORCHIDEE simulate a higher GPP amplitude than MTE-GPP over North America and southwest of Asia. In addition, most models have a negative phase bias of GPP over south of Asia and a positive phase bias of GPP over Siberia (Figure S7).

4.4. Spatial Patterns of NEE

All models have very similar spatial distributions of the NEE difference with MTE-NEE. Models produce larger NEE than MTE-NEE over Europe, east of Siberia, south of China, and southeast of North America and smaller values over north of Canada and southwest of Asia than JU11 (Figure S8). For the amplitude of NEE,

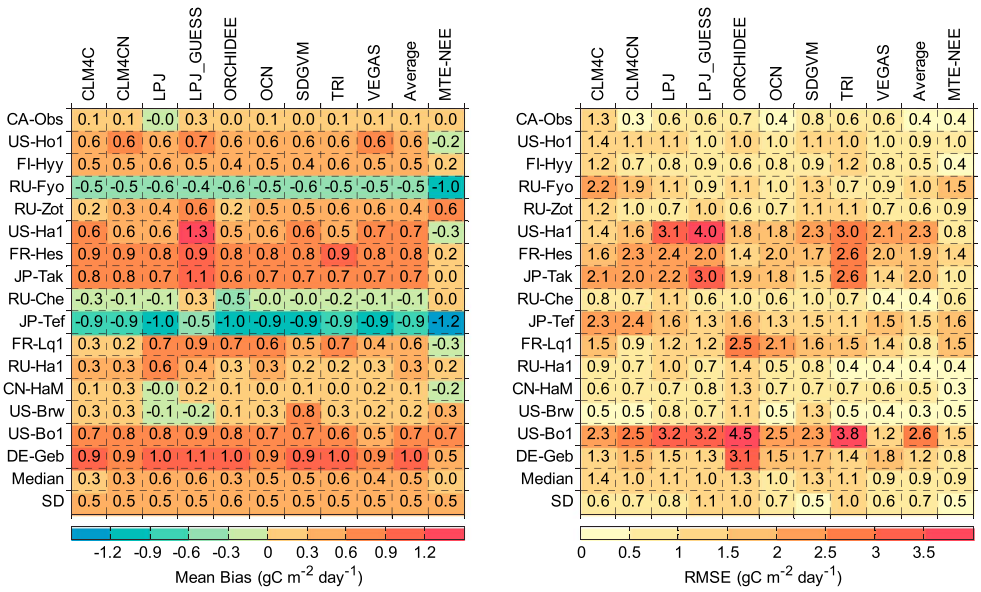

Figure 5. The mean bias and RMSE of monthly NEE at the 16 flux sites listed in Table 2. (left) The mean NEE bias and (right) the RMSE of monthly NEE. The last two rows in the two panels indicate the median and standard deviation (SD) of mean bias or RMSE of NEE across the 16 flux sites.

CLM4, CLM4CN, LPJ, LPJ_GUESS, ORCHIDEE, and SDGVM simulate larger amplitude than MTE-NEE, while OCN and TRI simulate smaller amplitude in most areas of the Northern Hemisphere (Figure S9). In addition, LPJ_GUESS and SDGVM simulate an earlier NEE phase (< -20 days) than MTE-NEE, and other models have patchy patterns of earlier and a later phase than MTE-NEE (Figure S9) but with a consistent earlier phase than MTE-NEE over western Russia and boreal North America.

5. Comparison With Atmospheric CO_2 Seasonal Cycle

5.1. Surface Stations CO_2

Figure 8 shows the comparison between the observed seasonal cycle of atmospheric CO_2 and NEE of each model including MTE-NEE after transport by LMDZ4 at 15 stations. Most simulations capture the shape of the seasonality of atmospheric CO_2 mole fractions, except for TRI that shows an opposite seasonality than observed. Most simulations show a too early CO_2 uptake in spring at high-latitude sites ALT, BRW, and MBC. Across stations, the median amplitude bias of CO_2 for all models ranges from -8.8 ± 2.3 ppm to 2.7 ± 2.8 ppm, with most models underestimating the amplitude of CO_2 (except CLM4 and SDGVM) (Figure 10). Interestingly, the LMDZ4 simulation with MTE-NEE also underestimates the amplitude of CO_2 at most stations (-1.8 ± 2.0 ppm). CLM4C overestimates the amplitude of CO_2 at all stations, and SDGVM overestimates it at 11 out of the 15 sites (except MID, MLO, NWR, and SHM).

The amplitude bias of CO_2 for ORCHIDEE significantly and positively correlates with the latitude of the measurement sites (Figure S10; ORCHIDEE overestimates the seasonal amplitude of atmospheric CO_2 by 2–5 ppm at high-latitude sites ALT, BRW, CBA, and MBC but underestimates the amplitude at the low-latitude site MLO), which suggests that ORCHIDEE overestimates the amplitude of NEE in high-latitude regions and underestimates it in midlatitude regions. Among the nine models, a positive correlation between simulated CO_2 amplitude bias and latitude is found in four models (Figure S11; CLM4C, LPJ, ORCHIDEE, and SDGVM), and this correlation is significant for ORCHIDEE and SDGVM, which is consistent with the spatial distribution of the difference between NEE of models and MTE-NEE (Figure S8), even though the amplitude of MTE-NEE is more underestimated at high-latitude stations than at midlatitude ones. Note that the negative amplitude bias of CO_2 for MTE-NEE at all stations suggests that this product underestimates the NEE amplitude (Figure 10), which raises a caution flag when using it for evaluation of models.

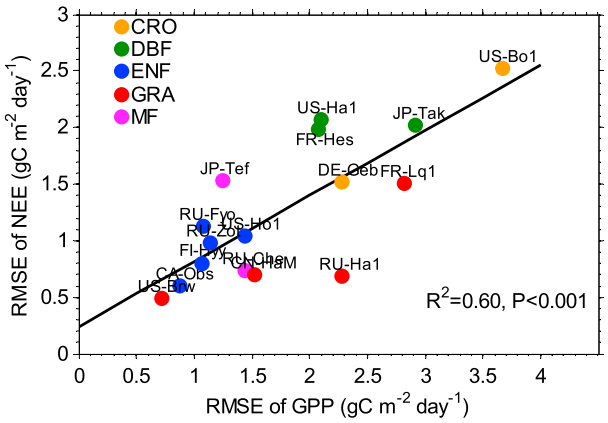

Figure 6. The relationship between median RMSE of monthly GPP and NEE for the nine terrestrial ecosystem models across the 16 flux sites.

5.2. X_{CO2}

The comparison of the LMDZ4 simulations with TCCON column measurements also shows earlier CO_2 uptake in spring for most models (Figure 9), like at surface stations. Across the 10 TCCON sites, the X_{CO2} amplitude is overestimated by CLM4C and SDGVM and underestimated by the other models (Figure 10). CLM4C simulates higher carbon release from the biosphere by winter respiration and higher CO_2 uptake in summer. SDGVM simulates higher CO_2 uptake in summer (Figure 9). Across the 10 TCCON sites, the X_{CO2} amplitude is overestimated by 3.0 ± 3.7 ppm in CLM4C and 2.1 ± 2.3 ppm in SDGVM. The X_{CO2} amplitude is moderately underestimated by -1.2 ± 2.0 ppm in CLM4CN, -1.8 ± 1.9 ppm in ORCHIDEE, and -1.2 ± 1.4 ppm in VEGAS. In LPJ, LPJ_GUESS, OCN, and TRI, the X_{CO2} amplitude is underestimated by -4.2 ± 1.8 ppm, -2.4 ± 1.9 ppm, -4.0 ± 1.3 ppm, and -6.0 ± 1.6 ppm, respectively. MTE-NEE produces a larger than observed X_{CO2} amplitude at four out of the 10 TCCON sites, and the median X_{CO2} amplitude bias for this product is -0.3 ± 1.3 ppm.

The seasonal CO_2 amplitude during the period 2009–2011 was underestimated in five CMIP5 models and overestimated in three Coupled Model Intercomparison Project Phase 5 (CMIP5) models [*Graven et al.*, 2013]. Two models (CLM4CN and ORCHIDEE) in this study were used as land surface models in CMIP5 (part of the CCSM4 and IPSL-CM5A-MR Earth System Models, respectively) [*Graven et al.*, 2013]. In our study, CLM4CN underestimates the seasonal CO_2 amplitude at the boreal surface CO_2 sites, but ORCHIDEE overestimates the seasonal CO_2 amplitude at these sites. These results are consistent with underestimation and overestimation in seasonal CO_2 amplitude over 45°N–90°N in CCSM4 and IPSL-CM5A-MR, respectively, as found by *Graven et al.* [2013], despite the fact that in our study models are forced by observed climate, whereas in *Graven et al.* [2013] they were forced by (biased) general circulation model climate. This bias of seasonal CO_2 amplitude and its changes in the models show the limited ability of predicting future land carbon sinks [e.g., *McGuire et al.*, 2001; *Graven et al.*, 2013], suggesting that improved seasonal cycle of NEE and its changes are needed in the future.

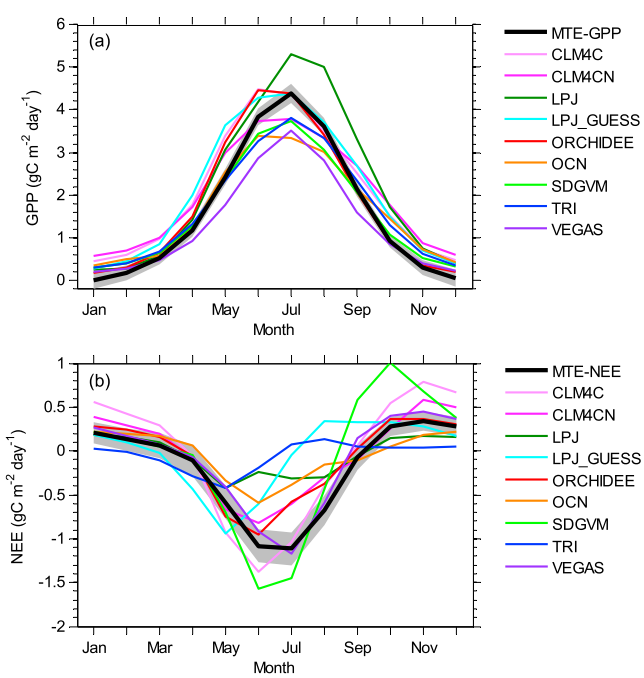

Figure 7. The seasonal cycle of (a) GPP and (b) NEE over land north of 25°N.

The phase bias of X_{CO2} is negative for CLM4C, CLM4CN, LPJ_GUESS, ORCHIDEE, SDGVM, TRI, and VEGAS but positive for LPJ and OCN at most sites (Figure 10). Across surface sites and TCCON sites, the median phase bias of CO_2 seasonality is negative in seven out of nine models (except LPJ and OCN; Figure 10), which results from an earlier-than-observed drawdown of atmospheric CO_2 in spring (Figures 8 and 9). The median phase bias of X_{CO2} across all the sites is less than 11 days in most models (except LPJ_GUESS,

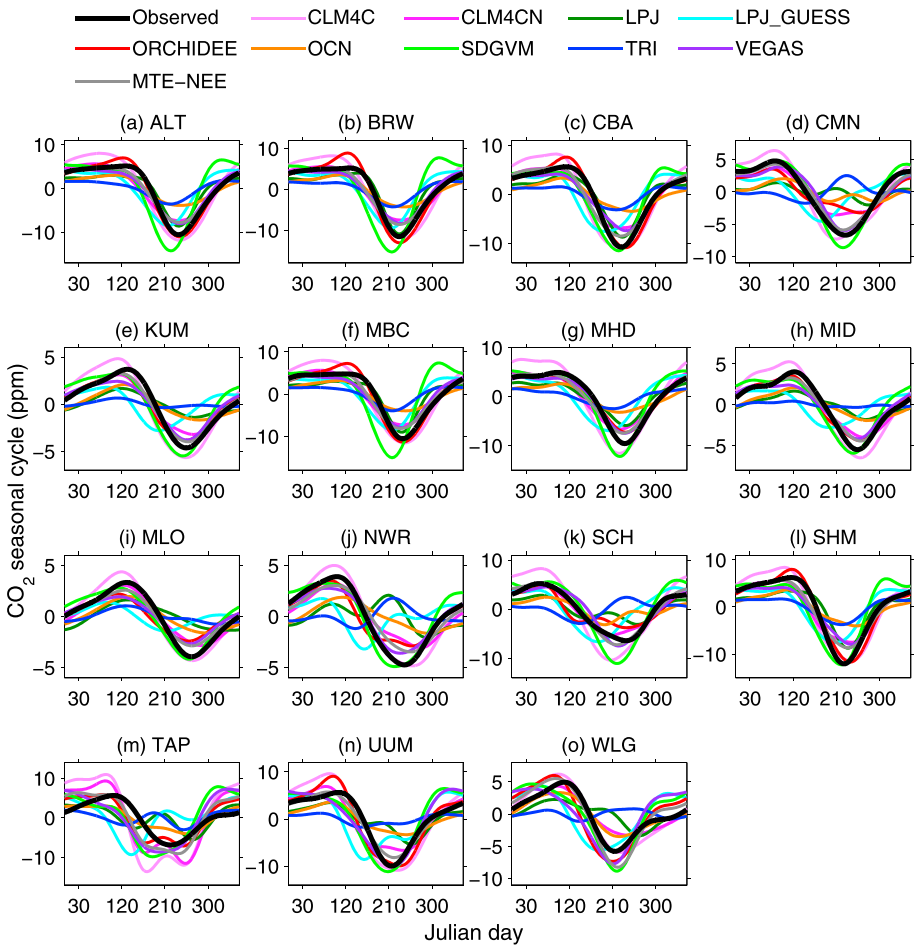

Figure 8. Observed and modeled averaged CO_2 seasonal cycle at the 15 atmospheric CO_2 concentration measurement sites listed in Table 3.

SDGVM, and TRI). The phase bias of CO_2 for JU11 is negative at most sites and less than 10 days, the median phase bias of CO_2 for JU11 is -7 ± 9 days.

6. Synthesis of Systematic GPP and NEE Errors Across Scales

Generally, the median of amplitude bias of GPP (NEE) at the 16 flux sites significantly and positively correlated with the amplitude bias of the MTE-GPP (NEE) across the models (Figure S12). The median of amplitude bias of NEE at the 16 flux sites also significantly correlated with the median of the amplitude bias of the CO_2 seasonal cycle at the 25 surface stations and TCCON CO_2 sites. This suggests that evaluation of modeled GPP and NEE at the 16 flux sites in this study is broadly representative of model performance at large scale. However, there are still some discrepancies between the comparison at flux sites versus the MTE products for some models. This suggests that both flux data and data products representing large scale (MTE products and CO_2 concentration) are needed to evaluate the performance of models because they cover different scales.

Over the Northern Hemisphere, the amplitude of modeled GPP spans a large range (1.6–5.1 gC m^{-2} d^{-1}), while the amplitude of MTE-GPP is of 4.4 ± 0.3 gC m^{-2} d^{-1}. The amplitude of modeled NEE over land north of 25°N spans a range from 0.7 gC m^{-2} d^{-1} to 2.5 gC m^{-2} d^{-1}. At large scale, the amplitude of NEE significantly correlates with the amplitude of simulated X_{CO2} across models for each TCCON site. Figure 11 shows the correlations between the amplitude of NEE north of 25°N and X_{CO2} across models at two TCCON sites (Bremen and Park Falls; see Table 4) to illustrate this. The amplitude of modeled NEE explains 92% and 88% of the cross-model variation in the amplitude of simulated X_{CO2} at Bremen and Park Falls, respectively. If we use

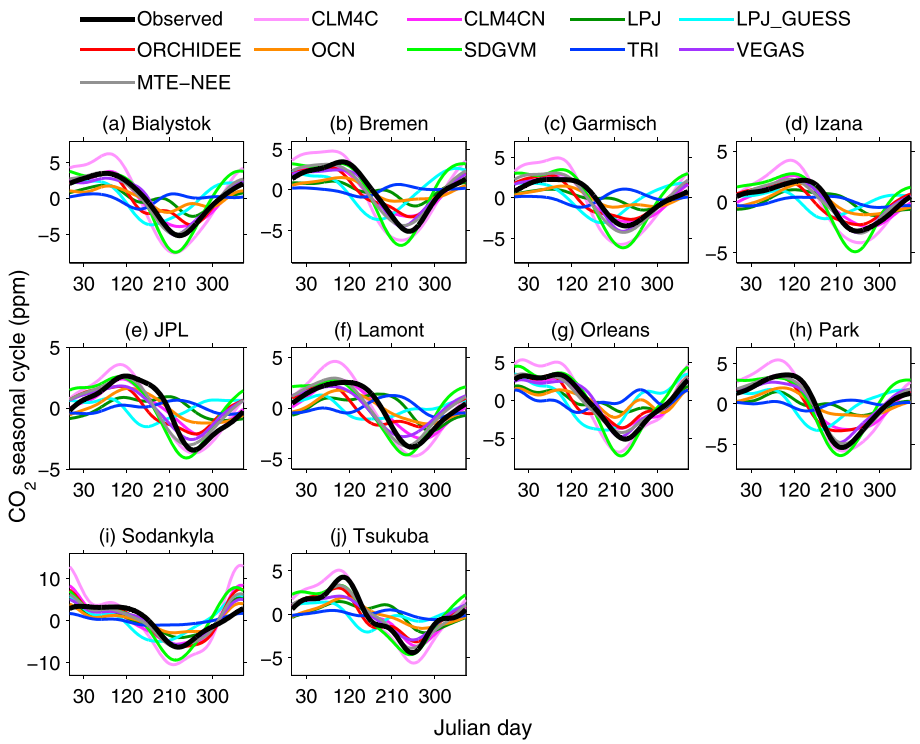

Figure 9. Observed and modeled averaged CO_2 seasonal cycle at the 10 TCCON sites listed in Table 4.

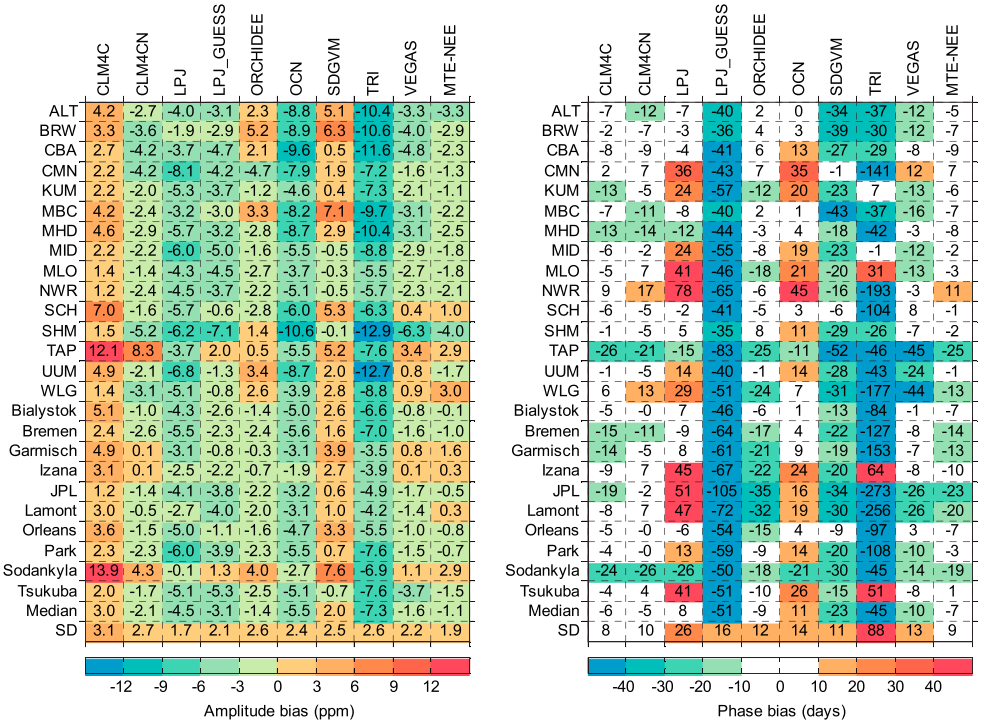

Figure 10. The amplitude and phase bias of modeled averaged CO_2 seasonal cycle in comparison to observed averaged CO_2 seasonal cycle at the 15 atmospheric CO_2 concentration measurement sites and 10 TCCON sites listed in Tables 3 and 4, respectively. (left) The amplitude bias and (right) the phase bias. The last two rows in the two panels indicate the median and standard deviation (SD) of amplitude bias or phase bias among the 25 sites.

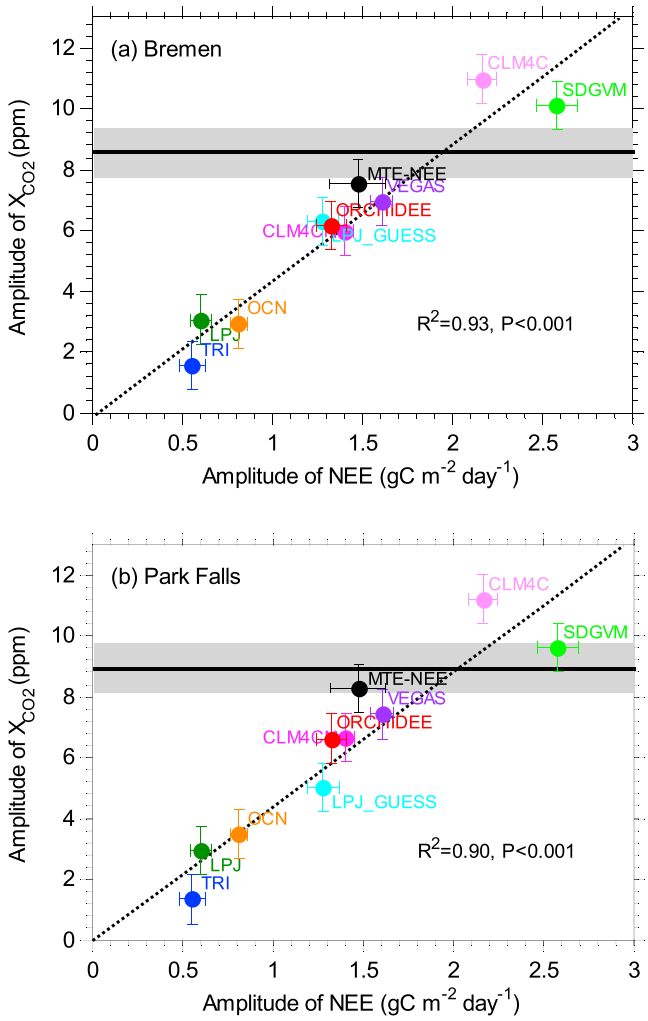

Figure 11. Amplitude of modeled NEE over the land north of 25°N versus amplitude of X_{CO2} measured at (a) Bremen and (b) Park Falls (TCCON sites). The dot line shows the linear regressed line across the nine models. The black point indicates MTE-NEE. The black solid line is the amplitude of observed X_{CO2}, with 1 standard deviation as the gray shade.

the observed amplitude of X_{CO2} at Bremen (8.6 ± 0.8 ppm) or Park Falls (8.9 ± 0.8 ppm) to estimate the amplitude of NEE (the intersection point by solid black line and dot line in Figure 11), we get a constrained amplitude of NEE of 1.8 ± 0.2 gC m^{-2} d^{-1} and 1.9 ± 0.2 gC m^{-2} d^{-1} for Bremen and Park Falls, respectively. Because the footprint of each TCCON site relative to continental-scale NEE may be different from each other and smaller than the whole Northern Hemisphere, the average of the TCCON-constrained amplitude of NEE is of 1.6 ± 0.4 gC m^{-2} d^{-1} across the 10 TCCON sites, which is a little larger than the amplitude of MTE-NEE (1.5 ± 0.2 gC m^{-2} d^{-1}). The uncertainty of amplitude of NEE across the nine models (±0.7 gC m^{-2} d^{-1}) is reduced by ~40% by TCCON-constrained method.

The mean carbon uptake during the growing season over the Northern Hemisphere (April to September) estimated from MTE-NEE is 7.6 ± 2.0 Pg C yr^{-1}. Considering that MTE-NEE underestimates the average amplitude of X_{CO2} across the 10 TCCON sites by ~4 ± 26% with LMDZ4, the total carbon uptake during the growing season over Northern Hemisphere adjusted by scaling MTE-NEE by a factor 1.04 ± 0.26 is of 7.9 ± 2.0 Pg C yr^{-1}. By comparison, the carbon uptake during the growing season was estimated to be 7.9 Pg C yr^{-1} using the CASA modeled NEE transported with 12 TransCom 3 experiment models [*Gurney et al.*, 2003] and Park Falls X_{CO2} by *Yang et al.* [2007], which is exactly the same number than our estimation from the LMDZ4 model but our estimation covers five more latitude degrees.

For the phase of X_{CO2} at the TCCON sites, most models simulate an earlier drawdown of X_{CO2} than observed (except for LPJ and OCN) (Figure 10). The phase of modeled NEE over the Northern Hemisphere significantly correlates with the phase of X_{CO2} simulated by LMDZ4 across models at all the TCCON sites ($R = 0.77$–0.96, $P < 0.01$). Figure 12 shows an example of the significantly positive correlation between the phase of NEE and phase of X_{CO2} at Bremen ($R^2 = 0.92$, $P < 0.001$) and Park Falls ($R^2 = 0.96$, $P < 0.001$) sites. Constraining the phase of NEE using the phase of X_{CO2} observed at the 10 TCCON sites, the constrained phase of NEE is 356 ± 12 days, i.e., the peak of NEE over Northern Hemisphere occurs at 174 ± 12 days.

7. Discussion and Conclusions

Using GPP and NEE measured at 16 FLUXNET sites distributed in the Northern Hemisphere across different PFTs, gridded GPP, and NEE products upscaled from flux sites by MTE, and atmospheric CO_2 measurements at surface sites and averaged over the atmospheric column at TCCON sites, we evaluated the seasonality of GPP and NEE of

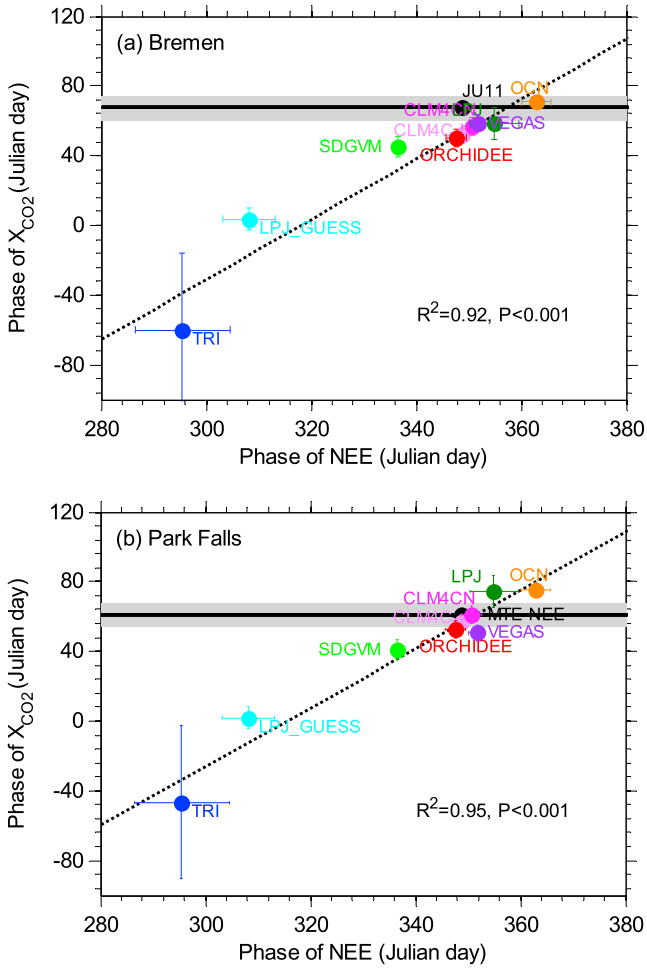

Figure 12. Phase of modeled NEE over the land north of 25°N versus phase of X_{CO2} measured at (a) Bremen and (b) Park Falls (TCCON sites). The dot line shows the linear regressed line across the nine models. The black point indicates MTE-NEE. The black solid line is the phase of observed X_{CO2}, with 1 standard deviation as the gray shade.

the nine terrestrial ecosystem models from the TRENDY intercomparison project. Seven models out of the nine have a bias to overestimate GPP and have an earlier start of GPP in spring at most flux sites. However, all models estimate higher NEE in most flux sites. Generally, the models perform better at evergreen needleleaf forest flux sites than at deciduous broadleaf forest, grassland, and cropland sites. The errors in simulated GPP propagate to errors in simulated NEE. Over the Northern Hemisphere, five models out of the nine models simulate higher-average monthly GPP than MTE, especially in spring and autumn. This overestimated GPP in some models may reflect the lack of nutrient limitations on GPP in some models [*Bonan et al.*, 2012; *Guenet et al.*, 2013]. Although some models overestimate GPP, seven out of the nine models underestimate the seasonal amplitude of NEE, which points out to an overcompensation of the GPP positive bias by a larger positive bias in ecosystem respiration and disturbance-induced CO_2 emissions. All models simulate less update than MTE over the Northern Hemisphere by 0.2/0.1 gC m^{-2} d^{-1}.

The comparison with atmospheric CO_2, with the help of a transport model, also shows the underestimated amplitude of CO_2 at most sites for seven out of the nine models (except CLM4C and SDGVM), which confirms that most models underestimate the Northern Hemisphere carbon uptake, and even more so during the carbon uptake period. The too small NEE seasonal amplitude in models can result from missing processes. On the one hand, models may underestimate NEE due to missing disturbance in models such as insects, windthrow, or small fires. If these disturbances emit CO_2 in the growing season, including them in models should further decrease the amplitude of simulated NEE. On the other hand, harvest of crop and wood biomass and the oxidation of harvested products causing CO_2 emissions with a nonseasonal profile are not systematically incorporated in models [e.g., *Zeng et al.*, 2014]. Including these processes [*Ciais et al.*, 2007] would tend to increase the seasonal amplitude of NEE.

The overestimated ecosystem respiration during the growing season could also be one important reason for the lower amplitude of NEE. A high bias of heterotrophic respiration can be related to the parameterization of the temperature sensitivity and asymmetric response to daytime and nighttime temperature [*Peng et al.*, 2013], as well as the turnover rate and storage of labile carbon pools that determine the seasonal cycle of heterotrophic respiration. For instance, in CLM4, *Bonan et al.* [2012] compared the decomposition rate of litter with observations at many sites and concluded too high decomposition of litter pools, which is consistent with too high respiration. If this bias of CLM4 is also common to other models, then it could explain why the NEE amplitude tends to be underestimated while the GPP amplitude is overestimated. Evaluation of models against soil C incubation data [*Moyano et al.*, 2012] and field studies of litter decomposition should bring clues to this problem.

The bias in biomass of models could also introduce a positive bias in autotrophic respiration, because models that do not describe forest harvest and disturbance tend to overestimate biomass and thus autotrophic respiration. Thus, comparison with storage and turnover rate of soil organic carbon and biomass [e.g., *Todd-Brown et al*., 2013] should be included in future systematic benchmarks alongside with evaluation of heterotrophic respiration against in situ and field data. These improvements can be valuable to get better seasonal NEE amplitude and could enhance the prediction in coupled models [*Graven et al*., 2013].

We also note that a common bias of most models is that they underestimate harvest by humans and do not include the lateral export of carbon from soils by erosion and the fraction of soil C transferred to rivers. *Regnier et al*. [2013] estimated that an amount of 1.9 PgC yr^{-1} was transferred from soil to rivers. *Quinton et al*. [2010] estimated an additional 0.5 PgC yr^{-1} erosion of agricultural soil carbon, which could be not available for seasonal soil heterotrophic respiration. In addition, crop harvest [*Ciais et al*., 2007] and wood harvest also decrease available litter for heterotrophic respiration. Models ignore all or part of these processes, and it is thus logical that despite their GPP being high biased, their respiration has an even larger positive bias.

Northern Hemisphere TCCON sites further suggest that the amplitude of X_{CO2} is underestimated by 4% for MTE and by more than 10% for seven out of the nine models. Using observed amplitude of X_{CO2} at the 10 TCCON sites, the amplitude of NEE over the Northern Hemisphere was estimated at 1.6 ± 0.4 gC m^{-2} d^{-1}, which corresponds to an uptake during the growing season of 7.9 ± 2.0 Pg C yr^{-1}.

Acknowledgments
TCCON data were obtained from the TCCON Data Archive, operated by the California Institute of Technology from the website at http://tccon.ipac.caltech.edu/. Support for TCCON is provided by many national research organizations that are listed on the TCCON website. We thank NOAA/ESRL, CarboEurope, ICOS-France, and IMS Italy for making the surface CO_2 measurements and contributing them to data achieve. We thank Martin Jung for producing and making available the MTE-GPP and MTE-NEE products. We thank Samuel Levis for the NCAR-CLM4 simulations. We also thank Nicholas Deutscher's valuable comments.

References

Boden, T. A., G. Marland, and R. J. Andres (2012), Global, Regional, and National Fossil-Fuel CO_2 Emissions. Carbon Dioxide Information Analysis Center, Oak Ridge Natl. Lab., U.S. Dep. of Energy, Oak Ridge, Tenn., doi:10.3334/CDIAC/00001_V2012.

Bonan, G. B., K. W. Oleson, R. A. Fisher, G. Lasslop, and M. Reichstein (2012), Reconciling leaf physiological traits and canopy flux data: Use of the TRY and FLUXNET databases in the Community Land Model version 4, *J. Geophys. Res.*, *117*, G02026, doi:10.1029/2011JG001913.

Cadule, P., P. Friedlingstein, L. Bopp, S. Sitch, C. D. Jones, P. Ciais, S. L. Piao, and P. Peylin (2010), Benchmarking coupled climate-carbon models against long-term atmospheric CO_2 measurements, *Global Biogeochem. Cycles*, *24*, Gb2016, doi:10.1029/2009GB003556.

Chang, J. F., et al. (2013), Incorporating grassland management in ORCHIDEE: Model description and evaluation at 11 eddy-covariance sites in Europe, *Geosci. Model Dev.*, *6*, 2165–2181, doi:10.5194/gmd-6-2165-2013.

Chevallier, F., et al. (2011), Global CO_2 fluxes inferred from surface air-sample measurements and from TCCON retrievals of the CO_2 total column, *Geophys. Res. Lett.*, *38*, L24810, doi:10.1029/2011GL049899.

Chevallier, F., et al. (2012), What eddy-covariance measurements tell us about prior land flux errors in CO_2-flux inversion schemes, *Global Biogeochem. Cycles*, *26*, GB1021, doi:10.1029/2010GB003974.

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

Ciais, P., P. Bousquet, A. Freibauer, and T. Naegler (2007), Horizontal displacement of carbon associated with agriculture and its impacts on atmospheric CO_2, *Global Biogeochem. Cycles*, *21*, GB2014, doi:10.1029/2006GB002741.

Cox, P. M., R. A. Betts, C. D. Jones, S. A. Spall, and I. J. Totterdell (2000), Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model, *Nature*, *408*, 184–187.

Cox, P. M., D. Pearson, B. B. Booth, P. Friedlingstein, C. Huntingford, C. D. Jones, and C. M. Luke (2013), Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability, *Nature*, *494*, 341–344.

Denning, A. S., D. A. Randall, G. J. Collatz, and P. J. Sellers (1996), Simulations of terrestrial carbon metabolism and atmospheric CO_2 in a general circulation model. Part 2: Simulated CO_2 concentrations, *Tellus, Ser. B*, *48*, 543–567.

Folberth, G., D. A. Hauglustaine, P. Ciais, and J. Lathière (2005), On the role of atmospheric chemistry in the global CO_2 budget, *Geophys. Res. Lett.*, *32*, L08801, doi:10.1029/2004GL021812.

Friedlingstein, P., et al. (2006), Climate-carbon cycle feedback analysis: Results from the (CMIP)-M-4 model intercomparison, *J. Clim.*, *19*, 3337–3353.

Graven, H. D., et al. (2013), Enhanced seasonal exchange of CO_2 by northern ecosystems since 1960, *Science*, *341*, 1085–1089.

Guanter, L., et al. (2014), Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence, *Proc. Natl. Acad. Sci. U.S.A.*, *111*, E1327–E1333.

Guenet, B., P. Cadule, S. Zaehle, S. L. Piao, P. Peylin, F. Maignan, P. Ciais, and P. Friedlingstein (2013), Does the integration of the dynamic nitrogen cycle in a terrestrial biosphere model improve the long-term trend of the leaf area index?, *Clim. Dyn.*, *40*, 2535–2548.

Gurney, K. R., et al. (2003), TransCom 3 CO_2 inversion intercomparison: 1. Annual mean control results and sensitivity to transport and prior flux information, *Tellus B*, *55*, 555–579.

Heimann, M., et al. (1998), Evaluation of terrestrial carbon cycle models through simulations of the seasonal cycle of atmospheric CO(2): First results of a model intercomparison study, *Global Biogeochem. Cycles*, *12*, 1–24, doi:10.1029/97GB01936.

Hourdin, F., et al. (2006), The LMDZ4 general circulation model: Climate performance and sensitivity to parametrized physics with emphasis on tropical convection, *Clim. Dyn.*, *27*, 787–813.

Jung, M., et al. (2011), Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, *J. Geophys. Res.*, *116*, G00J07, doi:10.1029/2010JG001566.

Kaminski, T., R. Giering, and M. Heimann (1996), Sensitivity of the seasonal cycle of CO_2 at remote monitoring stations with respect to seasonal surface exchange fluxes determined with the adjoint of an atmospheric transport model, *Phys. Chem. Earth*, *21*, 457–462.

Keeling, C. D., J. F. S. Chin, and T. P. Whorf (1996), Increased activity of northern vegetation inferred from atmospheric CO_2 measurements, *Nature*, *382*, 146–149.

Kelley, D. I., I. C. Prentice, S. P. Harrison, H. Wang, M. Simard, J. B. Fisher, and K. O. Willis (2013), A comprehensive benchmarking system for evaluating global vegetation models, *Biogeosciences*, *10*, 3313–3340.

Keppel-Aleks, G., et al. (2012), The imprint of surface fluxes and transport on variations in total column carbon dioxide, *Biogeosciences*, *9*, 875–891.

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

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

Lawrence, D. M., et al. (2011), Parameterization improvements and functional and structural advances in Version 4 of the Community Land Model, *J. Adv. Model. Earth Syst.*, *3*, M03001.

Le Quere, C., et al. (2009), Trends in the sources and sinks of carbon dioxide, *Nat. Geosci.*, *2*, 831–836.

Luo, Y. Q., et al. (2012), A framework for benchmarking land models, *Biogeosciences*, *9*, 3857–3874.

Maignan, F., F. M. Breon, F. Chevallier, N. Viovy, P. Ciais, C. Garrec, J. Trules, and M. Mancip (2011), Evaluation of a global vegetation model using time series of satellite vegetation indices, *Geosci. Model Dev.*, *4*, 1103–1114.

McGuire, A. D., et al. (2001), Carbon balance of the terrestrial biosphere in the Twentieth Century: Analyses of CO_2, climate and land use effects with four process-based ecosystem models, *Global Biogeochem. Cycles*, *15*, 183–206, doi:10.1029/2000GB001298.

Messerschmidt, J., N. Parazoo, D. Wunch, N. M. Deutscher, C. Roehl, T. Warneke, and P. O. Wennberg (2013), Evaluation of seasonal atmosphere–biosphere exchange estimations with TCCON measurements, *Atmos. Chem. Phys.*, *13*, 5103–5115.

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

Moyano, F. E., et al. (2012), The moisture response of soil heterotrophic respiration: Interaction with soil properties, *Biogeosciences*, *9*, 1173–1182.

Nemry, B., L. François, J.-C. Gérard, A. Bondeau, M. Heimann, and The Participants of the Potsdam NPP Model Intercomparison (1999), Comparing global models of terrestrial net primary productivity (NPP): Analysis of the seasonal atmospheric CO_2 signal, *Global Change Biol.*, *5*, 65–76.

Oleson, K. W., et al. (2010), Technical description of version 4.0 of the Community Land Model (CLM), *NCAR Tech. Note NCAR/TN-478+STR*, 257 pp.

Papale, D., et al. (2006), Towards a standardized processing of net ecosystem exchange measured with eddy covariance technique: Algorithms and uncertainty estimation, *Biogeosciences*, *3*, 571–583.

Peng, S., S. Piao, P. Ciais, R. B. Myneni, A. Chen, F. Chevallier, A. J. Dolman, I. A. Janssens, J. Penuelas, and G. Zhang (2013), Asymmetric effects of daytime and night-time warming on Northern Hemisphere vegetation, *Nature*, *501*, 88–92.

Peylin, P., et al. (2011), Importance of fossil fuel emission uncertainties over Europe for CO_2 modeling: Model intercomparison, *Atmos. Chem. Phys.*, *11*, 6607–6622.

Peylin, P., et al. (2013), Global atmospheric carbon budget: Results from an ensemble of atmospheric CO_2 inversions, *Biogeosciences*, *10*, 6699–6720.

Piao, S., et al. (2013), Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO_2 trends, *Global Change Biol.*, *19*, 2117–2132.

Quinton, J. N., G. Govers, K. Van Oost, and R. D. Bardgett (2010), The impact of agricultural soil erosion on biogeochemical cycling, *Nature Geosci.*, *3*, 311–314.

Raczka, B. M., et al. (2013), Evaluation of continental carbon cycle simulations with North American flux tower observations, *Ecol. Monogr.*, *83*, 531–556.

Randerson, J. T., et al. (2009), Systematic assessment of terrestrial biogeochemistry in coupled climate-carbon models, *Global Change Biol.*, *15*, 2462–2484.

Raymond, P. A., J. E. Bauer, and J. J. Cole (2000), Atmospheric CO_2 evasion, dissolved inorganic carbon production, and net heterotrophy in the York River estuary, *Limnol. Oceanogr.*, *45*, 1707–1717.

Regnier, P., et al. (2013), Anthropogenic perturbation of the carbon fluxes from land to ocean, *Nat. Geosci.*, *6*, 597–607.

Richardson, A. D., et al. (2012), Terrestrial biosphere models need better representation of vegetation phenology: Results from the North American Carbon Program Site Synthesis, *Global Change Biol.*, *18*, 566–584.

Roland, M., et al. (2013), Atmospheric turbulence triggers pronounced diel pattern in karst carbonate geochemistry, *Biogeosci. Discuss.*, *10*, 1207–1227.

Sitch, S., et al. (2003), Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model, *Global Change Biol.*, *9*, 161–185.

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

Sitch, S., et al. (2013), Trends and drivers of regional sources and sinks of carbon dioxide over the past two decades, *Biogeosci. Discuss.*, *10*, 20,113–20,177.

Smith, B., I. C. Prentice, and M. T. Sykes (2001), Representation of vegetation dynamics in the modelling of terrestrial ecosystems: Comparing two contrasting approaches within European climate space, *Global Ecol. Biogeogr.*, *10*, 621–637.

Stegehuis, A. I., A. J. Teuling, P. Ciais, R. Vautard, and M. Jung (2013), Future European temperature change uncertainties reduced by using land heat flux observations, *Geophys. Res. Lett.*, *40*, 2242–2245, doi:10.1002/grl.50404.

Stoeckli, R., D. M. Lawrence, G. Y. Niu, K. W. Oleson, P. E. Thornton, Z. L. Yang, G. B. Bonan, A. S. Denning, and S. W. Running (2008), Use of FLUXNET in the community land model development, *J. Geophys. Res.*, *113*, G01025, doi:10.1029/2007JG000562.

Takahashi, T., et al. (2009), Climatological mean and decadal change in surface ocean pCO_2, and net sea–air CO_2 flux over the global oceans, *Deep Sea Res., Part II*, *56*, 554–577.

Todd-Brown, K. E. O., J. T. Randerson, W. M. Post, F. M. Hoffman, C. Tarnocai, E. A. G. Schuur, and S. D. Allison (2013), Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations, *Biogeosciences*, *10*, 1717–1736.

Valade, A., N. Vuichard, P. Ciais, F. Ruget, N. Viovy, B. Gabrielle, N. Huth, and J.-F. Martiné (2014), ORCHIDEE-STICS, a process-based model of sugarcane biomass production: Calibration of model parameters governing phenology, *GCB Bioenergy*, *6*, 606–620.

Vuichard, N., J.-F. Soussana, P. Ciais, N. Viovy, C. Ammann, P. Calanca, J. Clifton-Brown, J. Fuhrer, M. Jones, and C. Martin (2007), Estimating the greenhouse gas fluxes of European grasslands with a process-based model: 1. Model evaluation from in situ measurements, *Global Biogeochem. Cycles*, *21*, GB1004, doi:10.1029/2005GB002611.

Wang, T., et al. (2012), State-dependent errors in a land surface model across biomes inferred from eddy covariance observations on multiple timescales, *Ecol. Modell.*, *246*, 11–25.

Woodward, F. I., T. M. Smith, and W. R. Emanuel (1995), A global land primary productivity and phytogeography model, *Global Biogeochem. Cycles*, *9*, 471–490, doi:10.1029/95GB02432.

Wunch, D., G. C. Toon, J.-F. L. Blavier, R. A. Washenfelder, J. Notholt, B. J. Connor, D. W. T. Griffith, V. Sherlock, and P. O. Wennberg (2011), The total carbon column observing network, *Philos. Trans. R. Soc. London, Ser. A*, *369*, 2087–2112.

Yang, Z., R. A. Washenfelder, G. Keppel-Aleks, N. Y. Krakauer, J. T. Randerson, P. P. Tans, C. Sweeney, and P. O. Wennberg (2007), New constraints on Northern Hemisphere growing season net flux, *Geophys. Res. Lett.*, *34*, L12807, doi:10.1029/2007GL029742.

Zaehle, S., A. D. Friend, P. Friedlingstein, F. Dentener, P. Peylin, and M. Schulz (2010), Carbon and nitrogen cycle dynamics in the O-CN land surface model: 2. Role of the nitrogen cycle in the historical terrestrial carbon balance, *Global Biogeochem. Cycles*, *24*, GB1006, doi:10.1029/2009GB003522.

Zeng, N., A. Mariotti, and P. Wetzel (2005), Terrestrial mechanisms of interannual CO_2 variability, *Global Biogeochem. Cycles*, *19*, GB1016, doi:10.1029/2004GB0022763.

Zeng, N., F. Zhao, G. J. Collatz, E. Kalnay, R. J. Salawitch, T. O. West, and L. Guanter (2014), Agricultural Green Revolution as a driver of increasing atmospheric CO_2 seasonal amplitude, *Nature*, *515*, 394–397.

Zhao, Y., et al. (2012), How errors on meteorological variables impact simulated ecosystem fluxes: A case study for six French sites, *Biogeosciences*, *9*, 2537–2564.