Assessment of model estimates of land-atmosphere CO$_2$ exchange across Northern Eurasia


To cite this version:
Assessment of model estimates of land-atmosphere CO$_2$ exchange across Northern Eurasia

M. A. Rawlins$^1$, A. D. McGuire$^2$, J. S. Kimball$^3$, P. Dass$^1$, D. Lawrence$^4$, E. Burke$^5$, X. Chen$^6$, C. Delire$^7$, C. Koven$^8$, A. MacDougall$^9$, S. Peng$^{10,16}$, A. Rinke$^{11,12}$, K. Saito$^{13}$, W. Zhang$^{14}$, R. Alkama$^7$, T. J. Bohn$^{15}$, P. Ciais$^{10}$, B. Decharme$^7$, I. Gouttevin$^{16,17}$, T. Hajima$^{13}$, D. Ji$^{11}$, G. Krinner$^{16}$, D. P. Lettenmaier$^{18}$, P. Miller$^{14}$, J. C. Moore$^{11}$, B. Smith$^{14}$, and T. Sueyoshi$^{19,13}$

$^1$Climate System Research Center, Department of Geosciences, University of Massachusetts, Amherst, MA, USA
$^2$US Geological Survey, Alaska Cooperative Fish and Wildlife Research Unit, University of Alaska, Fairbanks, Alaska 99775, USA
$^3$NTSG, University of Montana, Missoula, MT, USA
$^4$National Center for Atmospheric Research, Boulder, CO, USA
$^5$Met Office Hadley Centre, FitzRoy Road, Exeter, EX1 3PB, UK
$^6$Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, USA
$^7$CRNM-GAME, Unité mixte de recherche CNRS/Meteo-France (UMR 3589), 42 av Coriolis, 31057 Toulouse, CEDEX, France
$^8$Lawrence Berkeley National Laboratory, Berkeley, CA, USA
$^9$School of Earth and Ocean Sciences, University of Victoria, Victoria, BC, Canada
$^{10}$Laboratoire des Sciences du Climat et de l’Environnement, CEA-CNRS-UVSQ, UMR8212, 91191 Gif-sur-Yvette, France
$^{11}$State Key Laboratory of Earth Surface Processes and Resource Ecology, College of Global Change and Earth System Science, Beijing Normal University, Beijing, China
$^{12}$Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Potsdam, Germany
$^{13}$Department of Integrated Climate Change Projection Research, Japan Agency for Marine-Earth Science and Technology, Yokohama, Kanagawa, Japan
$^{14}$Department of Physical Geography and Ecosystem Science, Lund University, Sölvegatan 12, SE 223 62 Lund, Sweden
$^{15}$School of Earth and Space Exploration, Arizona State University, Tempe, AZ, USA
$^{16}$CNRS and Université Grenoble Alpes, LGGE, 38041, Grenoble, France
$^{17}$Irstea, UR HHLY, 5 rue de la Doua, CS 70077, 69626 Villeurbanne, CEDEX, France
$^{18}$Department of Geography, University of California, Los Angeles, CA, USA
$^{19}$National Institute of Polar Research, Tachikawa, Tokyo, Japan

Correspondence to: M. A. Rawlins (rawlins@geo.umass.edu)

Received: 08 January 2015 – Published in Biogeosciences Discuss.: 03 February 2015
Revised: 18 May 2015 – Accepted: 01 July 2015 – Published: 28 July 2015

Abstract. A warming climate is altering land-atmosphere exchanges of carbon, with a potential for increased vegetation productivity as well as the mobilization of permafrost soil carbon stores. Here we investigate land-atmosphere carbon dioxide (CO$_2$) cycling through analysis of net ecosystem productivity (NEP) and its component fluxes of gross primary productivity (GPP) and ecosystem respiration (ER) and soil carbon residence time, simulated by a set of land surface models (LSMs) over a region spanning the drainage basin of Northern Eurasia. The retrospective simulations cover the period 1960–2009 at 0.5° resolution, which is a scale common among many global carbon and climate model simulations. Model performance benchmarks were drawn from comparisons against both observed CO$_2$ fluxes derived from site-based eddy covariance measurements as well as regional-scale GPP estimates based on satellite remote-sensing data.

Published by Copernicus Publications on behalf of the European Geosciences Union.
The site-based comparisons depict a tendency for overestimates in GPP and ER for several of the models, particularly at the two sites to the south. For several models the spatial pattern in GPP explains less than half the variance in the MODIS MOD17 GPP product. Across the models NEP increases by as little as 0.01 to as much as 0.79 g C m$^{-2}$ yr$^{-2}$, equivalent to 3 to 340 % of the respective model means, over the analysis period. For the multimodel average the increase is 135 % of the mean from the first to last 10 years of record (1960–1969 vs. 2000–2009), with a weakening CO$_2$ sink over the latter decades. Vegetation net primary productivity increased by 8 to 30 % from the first to last 10 years, contributing to soil carbon storage gains. The range in regional mean NEP among the group is twice the multimodel mean, indicative of the uncertainty in CO$_2$ sink strength. The models simulate that inputs to the soil carbon pool exceeded losses, resulting in a net soil carbon gain amid a decrease in residence time. Our analysis points to improvements in model elements controlling vegetation productivity and soil respiration as being needed for reducing uncertainty in land-atmosphere CO$_2$ exchange. These advances will require collection of new field data on vegetation and soil dynamics, the development of benchmarking data sets from measurements and remote-sensing observations, and investments in future model development and intercomparison studies.

1 Introduction

Northern boreal regions are known to play a major role in the land-atmosphere exchange of CO$_2$ at high latitudes (Graven et al., 2013). During the Holocene the Arctic is believed to have been a net sink of carbon (Pries et al., 2012). During modern times, often referred to as the anthropocene (Crutzen, 2006), warming across the high northern latitudes has occurred at a faster rate than the rest of the globe (Serreze et al., 2006). The enhanced warming is attributable to feedbacks involving biogeochemical and biogeophysical processes (Chapin III et al., 2005; Serreze and Barry, 2011; Schuur et al., 2015). Warming may increase soil microbial decomposition, placing the large permafrost carbon pool at greater risk for being mobilized and transferred to the atmosphere as greenhouse gases (GHGs), thus providing a positive feedback to global climate (Dutta et al., 2006; Vogel et al., 2009; Schuur et al., 2009). Warming may also lead to longer growing seasons, contributing to increased plant productivity and ecosystem carbon sequestration (Mellilo et al., 1993; Euskirchen et al., 2006). At the same time, warming may also lead to respiration increases through enhanced microbial activity and/or increased input of plant photosynthates into the soil (Högberg et al., 2001), offsetting some productivity increases and resulting in relatively low net carbon uptake (Parmentier et al., 2011). Satellite observations show broad greening trends in tundra regions (Myneni et al., 1997; Goetz et al., 2005; Zhang et al., 2008), suggesting a potential increase in the land sink of atmospheric CO$_2$. Some areas, however, are browning (Goetz et al., 2006).

Research studies point to uncertainty in the sign, magnitude and temporal trends in contemporary land-atmosphere exchanges of CO$_2$. A recent synthesis of observations and models by McGuire et al. (2012) suggests that tundra regions across the pan-Arctic were a sink for atmospheric CO$_2$ and a source of CH$_4$ from 1990–2009. However, a meta-analysis of 40 years of CO$_2$ flux observations from 54 studies spanning 32 sites across northern high latitudes found that tundra was an annual CO$_2$ source from the mid-1980s until the 2000s, with the data suggesting an increase in winter respiration rates, particularly over the last decade (Belshe et al., 2013).

In an analysis of outputs from several models from recent terrestrial biosphere model intercomparison projects, Fisher et al. (2014) found that spatial patterns in carbon stocks and fluxes over Alaska in 2003 varied widely, with some models showing a strong carbon sink, others a strong carbon source, and some showing the region as carbon neutral. It is critical to understand the net carbon sink as recent studies suggest that with continued warming the Arctic may transition from a net sink of atmospheric CO$_2$ to a net source over the coming decades (Hayes et al., 2011; Koven et al., 2011; Schaefer et al., 2011; MacDougall et al., 2013; Oechel et al., 2014). In a study using a process model which included disturbances, Hayes et al. (2011) estimated a 73 % reduction in the strength of the pan-Arctic land-based CO$_2$ sink over 1997–2006 vs. previous decades in the late 20th century.

Recent studies have provided new insights into model uncertainties relevant to our understanding of the land-based CO$_2$ sink across Northern Eurasia. Examining several independent estimates of the carbon balance of Russia including two dynamic global vegetation models (DGVMs), two atmospheric inversion methods, and a landscape-ecosystem approach (LEA) incorporating observed data, Quegan et al. (2011) concluded that estimates of heterotrophic respiration were biased high in the two DGVMs, and that the LEA appeared to give the most credible estimates of the fluxes. In an analysis of the terrestrial carbon budget of Russia using inventory-based, eddy covariance, and inversion methods, and a landscape-ecosystem approach (LEA) incorporating observed data, Quegan et al. (2011) concluded that estimates of heterotrophic respiration were biased high in the two DGVMs, and that the LEA appeared to give the most credible estimates of the fluxes. In an analysis of the terrestrial carbon budget of Russia using inventory-based, eddy covariance, and inversion methods, Dolman et al. (2012) noted good agreement in net ecosystem exchange among these bottom-up and top-down methods, estimating an average CO$_2$ sink across the three methods of 613.5 Tg C yr$^{-1}$. Their examination of outputs from a set of DGVMs, however, showed a much lower sink of 91 Tg C yr$^{-1}$. Graven et al. (2013) point to specification of vegetation dynamics and nitrogen cycling in a subset of CMIP5 models as a potential cause for their underestimation of changes in net productivity over the past 50 years. These analyses highlight the need for comprehensive assessments of numerical model estimates of spatial and temporal variations in land-atmosphere CO$_2$ exchange against independent benchmarking data. A lack of direct flux measurements...
across northern land areas presents considerable challenges for model validation efforts (Fisher et al., 2014).

In this study we examine model estimates of net ecosystem productivity (NEP) and component fluxes gross primary productivity (GPP) and ecosystem respiration (ER) across the arctic basin of Northern Eurasia from a series of retrospective simulations for the period 1960–2009. Our analysis for the region is unique in its synthesis of a large suite of land-surface models, available site-level data, and a remote-sensing product. Study goals are two-fold. First, using the available in situ data derived from tower-based measurements and the remote-sensing GPP product we seek to assess model efficacy in simulating spatial and temporal variations in GPP, ER, and NEP across the region. In doing so we elucidate issues complicating evaluations of model carbon cycle estimates across Northern Eurasia and, by extension, other areas of the northern high latitudes. Second, we estimate time changes in NEP and soil organic carbon (SOC) residence time and its controls as an indicator of climate sensitivity and potential vulnerability of soil carbon stocks. We focus the analysis and discussion on assessing how well the models capture the seasonal cycle and spatial patterns in GPP and ER flux rates, evaluating uncertainties in the net CO$_2$ exchange given reported biases in respiration rates, and in advancing understanding of the land–atmosphere cycling of CO$_2$ over recent decades.

2 Methods

2.1 Study Region

The spatial domain is the arctic drainage basin of Northern Eurasia which comprises all land areas draining to the Arctic Ocean, a region of some 13.5 million km$^2$ (Fig. 1). The basin covers roughly half of the Northern Eurasian Earth Science Partnership Initiative (NEESPI) study area, generally defined as the region between 15° E in the west, the Pacific Coast in the east, 40° N in the south, and the Arctic Ocean coastal zone in the north (Groisman et al., 2009). Warming and associated environmental changes to this region are among the most pronounced globally (Groisman and Bartalev, 2007; Groisman and Soja, 2009). Tundra vegetation is common across northern areas, with boreal forest and taiga comprising much of the remainder of the region. Steppes and grasslands are found across a relative small area in the extreme southwest. Continuous permafrost underlies over half of the region. Sporadic and relic permafrost comprise the southwest portion of the domain. West to east, the Ob, Yenisei, Lena, and Kolyma rivers drain a large fraction of the total river discharge from the Northern Eurasian basin.

2.2 Modeled data

We used outputs from retrospective simulations of nine models participating in the model integration group of the Permafrost Carbon Network. All simulation outputs available at the time of writing were included in the analysis (http://www.permafrostcarbon.org). The simulation protocol allowed for the choice of a model’s driving data sets for atmospheric CO$_2$, N deposition, climate, disturbance, and other forcings (Tables 1 and 2). Simulations were run at daily or sub-daily time steps in some models and at 0.5° resolution over all land areas north of 45° N latitude. The present study focuses on analysis of spatial patterns and temporal changes in land-atmosphere CO$_2$ fluxes over the period 1960–2009. Quantities analyzed are GPP, ER, and NEP, defined here as NEP = GPP − ER, where a positive value represents a net sink of CO$_2$ into the ecosystem. ER is the sum of
heterotrophic respiration and autotrophic respiration as estimated by the models. In this study we follow the conceptual framework for NEP and related terms as described in Chapin III et al. (2005). For this Permafrost Carbon Network activity modeling groups are providing gridded data for permafrost regions of the northern hemisphere. The nine models examined here (full model names in Table 1) are the (1) CLM version 4.5 (hereafter CLM4.5, Oleson et al., 2013); (2) CoLM (Ji et al., 2014); (3) ISBA (Decharme et al., 2011); (4) JULES (Best et al., 2011; Clark et al., 2011); (5) LPJ Guess WHyMe (hereafter LPJG, Smith et al., 2001; Wania et al., 2009b, a, 2010; Miller and Smith, 2012); (6) MIROC-ESM (Watanabe et al., 2011); (7) ORCHIDEE-IPSL (Koven et al., 2009, 2011; Gouttevin et al., 2012); (8) UVic (Avis et al., 2011; MacDougall et al., 2013); and (9) UW-VIC (Bohn et al., 2013). Table 2 lists the model elements most closely related to CO2 source and sink dynamics. These include model land cover initialization, time series forcings, light use efficiency, and CO2 and nitrogen fertilization. Among the models there is a wide range of accounting for processes related to disturbances such as fire and land use change (Table 2). All but two of the nine models (ISBA and UW-VIC) are considered to be dynamic global vegetation models (DGVMs), possessing the ability for vegetation to change over the model simulation. For ORCHIDEE, dynamic vegetation was not enabled in the simulation examined in this study. While studies that examine the overall ecosystem carbon balance (i.e. the net ecosystem carbon balance, NECB) are elemental to our understanding of the carbon cycle of Northern Eurasia, the present study focuses on the patterns in NEP and component fluxes GPP and ER, common in all of the models, in order to avoid the uncertainties given the range of model formulations related to the full carbon balance. Outputs from several of the nine models have been examined in other recent studies. The LPJG and ORCHIDEE were used in the synthesis of data and models presented by McGuire et al. (2012). JULES, LPJG, ORCHIDEE, and CLM4.5 participated in the TRENDY MIP (Piao et al., 2013). CLM4.5, ORCHIDEE, and LPJG were three of the eight models examined in the study of Dolman et al. (2012).

2.3 Observational data

2.3.1 Flux tower eddy covariance data

Model estimates for GPP, ER, and NEP are evaluated against data from six eddy covariance flux towers in four research areas located across Russia. The data are contained in the La Thuile global FLUXNET data set (Baldocchi, 2008). FLUXNET represents a global network of tower eddy covariance measurement sites for monitoring land-atmosphere exchanges of carbon dioxide and water vapor (http://daac.ornl.gov/FLUXNET/fluxnet.shtml). For these sites, GPP and ER data records overlap in the years 2002–2005. Observations during colder months are few. Tower sites are identified...
Figure 3. As in Fig. 2, for ER.

Figure 4. As in Fig. 2, for NEP. NEP = GPP−ER.
Table 1. Models participating in the Vulnerability of Permafrost Carbon Research Coordination Network (RCN) retrospective simulations. Modeling groups provided outputs for year 1960–2009, with the exception of CLM (–2005); JULES (–1999); UW-VIC (–2006).

<table>
<thead>
<tr>
<th>Model</th>
<th>Institution</th>
<th>Climate Data Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Community Land Model (CLM4.5)</td>
<td>National Center for Atmospheric Research, USA</td>
<td>CRUNCEP4</td>
</tr>
<tr>
<td>Common Land Model (CoLM) Interaction Sol-Biosphère-Atmosphere (ISBA)</td>
<td>Beijing Normal University, China</td>
<td>Princeton</td>
</tr>
<tr>
<td>Joint UK Land Environment Simulator (JULES)</td>
<td>Met Office, United Kingdom</td>
<td>WATCH3</td>
</tr>
<tr>
<td>Lund-Potsdam-Jenna General Ecosystem Simulator (LPJG)</td>
<td>Lund University, Sweden</td>
<td>CRU TS 3.1</td>
</tr>
<tr>
<td>Model for Interdisciplinary Research on Climate, Earth System Model (MIROC)</td>
<td>Japan Agency for Marine-Earth Science and Technology, Japan</td>
<td>CMIP5</td>
</tr>
<tr>
<td>Organising Carbon and Hydrology In Dynamic Ecosystems (ORCHIDEE)</td>
<td>Institute Pierre Simon Laplace (IPSL), France</td>
<td>WATCH3</td>
</tr>
<tr>
<td>University of Victoria (UVic)</td>
<td>University of Victoria, Canada</td>
<td>CRUNCEP4</td>
</tr>
<tr>
<td>Variable Infiltration Capacity (UW-VIC)</td>
<td>University of Washington, USA</td>
<td>CRU7, UDel8, NCEP-NCAR</td>
</tr>
</tbody>
</table>


3 Results

3.1 Model evaluation and benchmarking

3.1.1 Site-level evaluations

Confident assessment of uncertainties in land-atmosphere CO₂ fluxes is dependent on robust comparisons of model estimates against consistent benchmarking data. We begin by assessing the seven models which provided estimates through 2005, along with MOD17 GPP product. Monthly GPP from the models and MOD17 are compared with the cumulative monthly tower values by extracting the model values for the grid cell encompassing each tower site. Error measures that are based on absolute values of differences, like the...
Age than the GPP errors for comparisons where (i) ER errors
determine whether model errors in ER exceed the errors in
for Hakasija and Zotino during late summer and autumn are
overestimated by a considerable degree. Overestimates in ER
nine models overestimate GPP and ER at Zotino, with ER
is a tendency for overestimation in GPP and ER. All mod-
are also noted (Table 4). For the two sites in the south there
is a tendency for overestimation during the growing season peak at Hakasija and Zotino, respectively. The MBE for ER are 8, 35, 43, and
20 g C m²month⁻¹ for Chersky and Chokurdakh. Aver-
CO₂ in GPP (Fig. 2) and ER (Fig. 3), including the timing of peak
errors in GPP (Fig. 2) and ER (Fig. 3), including the timing of peak
errors in GPP (Fig. 2) and ER (Fig. 3), including the timing of peak
errors in GPP (Fig. 2) and ER (Fig. 3), including the timing of peak
errors in GPP (Fig. 2) and ER (Fig. 3), including the timing of peak
errors in GPP (Fig. 2) and ER (Fig. 3), including the timing of peak
errors in GPP (Fig. 2) and ER (Fig. 3), including the timing of peak
errors in GPP (Fig. 2) and ER (Fig. 3), including the timing of peak
A wide range of model performance is evident from Table 5. As with the mean errors shown in Table 4, agreements with observations are generally better at Chersky and Chokurdakh than Hakasija and Zotino. ER errors are also greater than GPP errors. Nash-Sutcliffe Es are negative for all models for both GPP and ER at Hakasija, and for most of the comparisons at Chokurdakh. Models CLM4.5, ISBA and UW-VIC exhibit the largest disagreements among the seven models for which estimates are available over the 2002–2005 period.

3.1.2 Regional-level evaluation of model GPP

Estimates from the MOD17 product provide a temporally and spatially continuous benchmark to assess model simulated GPP over the study domain. Average annual-total GPP from MOD17 over the period 2000–2009 is shown in Fig. 5. The MOD17 product clearly captures three distinct land cover zones over the region, representing: (i) grasslands across the south; (ii) boreal forests in the center of the region; and (iii) tundra to the north. Highest production occurs in the western forests where mean annual temperatures are higher. Both the steppe and tundra areas show annual GPP of less than 300 g C m$^{-2}$ yr$^{-1}$. Areas of low productivity in high elevation areas to the north are well delineated. The spatially averaged mean across the region is approximately 470 g C m$^{-2}$ yr$^{-1}$. In most of the models the patterns in GPP broadly represent the major biome areas captured in the MODIS land cover product (Fig. 1a). The east to west gradient is broadly captured in most of the models. However, grid-based correlations with the MOD17 GPP estimates (upper left of map panels in Fig. 5) show a wide range of agreement across the models. Spatial averages of the correlations across the domain range from $r = 0.92$ (ISBA) to $r = 0.48$ (ORCHIDEE). Four of the nine (LPJG, MIROC, ORCHIDEE, UVic) simulate a GPP field that explains less than 44% of the variability in GPP found within the MOD17 product. Annual GPP in the LPJG is notably low across the eastern half of the region. The CLM4.5 tends to predict lower GPP than MOD17 over tundra areas and higher productivity in the boreal zone. As estimated by the coefficient of variation (CV, upper right panel of Fig. 5), agreement in GPP is best across the higher productivity taiga biome. Figure 6

Table 3. Flux tower sites from the LaThuile data set (Baldocchi, 2008) used in this study. Site Hakasija consists of records from 3 sub-sites which all fall within the same RCN model grid. Each sub-site is represented with a different symbol in Figs. 2c, 3c, 4c. GPP and ER in the LaThuile data set are calculated using methodologies described in Reichstein et al. (2005).

Table 4. Average model error in g C m$^{-2}$ month$^{-1}$ for site-level comparisons over the years 2002–2005 shown in Figs. 2–4. Errors are calculated as the average ($\epsilon_j$) over all years and months for which a model estimate and site estimate are available at a given site. Thus, for each site and month, the mean bias error (MBE) is calculated as the average difference between the model and observed values: $\epsilon_j = C_j - C_{obs}$, where $C_j$ is GPP, ER or NEP for model $j$ and $C_{obs}$ is the observed value from the LaThuile FLUXNET observations (Baldocchi, 2008). The last column lists mean NEP error (NEP) across all sites. Model estimates for years 2002–2005 are not available for CoLM and JULES. Differences were evaluated using a 2-way repeated measures ANOVA test. Test design was a comparison of GPP vs ER $r$ tests for (i) each area separately; (ii) GPP and ER pooled for the two tundra sites and across the two forest sites; and (iii) GPP errors pooled across the four sites vs. ER errors pooled across the four sites.

Table 5. GPP, ER and NEP (g C m$^{-2}$ yr$^{-1}$) for the seven models compared with observations at the four sites (LaThuile, Hakasija, Chersky and Zotino). Models are ordered by mean error at the site where GPP is the closest estimate to MOD17. The last column lists the mean NEP error at the site (NEP). Model CLM4.5 is excluded from the analysis due to the low number of available records. Nash-Sutcliffe Es are negative for all models at both Hakasija and Zotino. The mean error at Hakasija is higher than at Zotino. The westernmost sub-region (Zotino) has the highest mean error at Hakasija and Zotino. ER errors are also greater than GPP errors at Hakasija and Zotino. Nash-Sutcliffe Es are negative for all models for both GPP and ER at Hakasija, and for most of the comparisons at Chokurdakh. Models CLM4.5, ISBA and UW-VIC exhibit the largest disagreements among the seven models for which estimates are available over the 2002–2005 period.
Table 5. Nash-Sutcliffe coefficient of efficiency \((E)\) (Nash and Sutcliffe, 1970) and Willmott’s refined index of agreement \((d_r)\) (Willmott et al., 2011) for comparison of GPP and ER errors derived from comparisons at sites shown in Table 4.

<table>
<thead>
<tr>
<th>Model</th>
<th>CHE GPP</th>
<th>CHE ER</th>
<th>COK GPP</th>
<th>COK ER</th>
<th>HAK GPP</th>
<th>HAK ER</th>
<th>ZOT GPP</th>
<th>ZOT ER</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLM4.5</td>
<td>0.15,0.67</td>
<td>−0.09,0.50</td>
<td>−0.74,0.44</td>
<td>−1.52,0.15</td>
<td>−1.20,0.39</td>
<td>−2.77,−0.03</td>
<td>−1.19,0.66</td>
<td>−5.34,−0.19</td>
</tr>
<tr>
<td>ISBA</td>
<td>0.43,0.67</td>
<td>−0.79,0.34</td>
<td>−0.04,0.54</td>
<td>−5.64,−0.26</td>
<td>−10.25,−0.24</td>
<td>−19.44,−0.55</td>
<td>−0.82,0.62</td>
<td>−10.56,−0.34</td>
</tr>
<tr>
<td>LPJG</td>
<td>0.64,0.77</td>
<td>0.68,0.76</td>
<td>0.86,0.83</td>
<td>0.62,0.71</td>
<td>−5.37,−0.09</td>
<td>−26.99,−0.64</td>
<td>0.76,0.85</td>
<td>0.64,0.76</td>
</tr>
<tr>
<td>MIROC</td>
<td>0.49,0.76</td>
<td>−0.38,0.48</td>
<td>−1.23,0.33</td>
<td>−8.02,−0.29</td>
<td>−2.69,0.24</td>
<td>−2.85,−0.01</td>
<td>0.95,0.94</td>
<td>0.35,0.60</td>
</tr>
<tr>
<td>ORCHIDEE</td>
<td>0.44,0.69</td>
<td>0.45,0.66</td>
<td>−1.08,0.32</td>
<td>−3.37,−0.04</td>
<td>−2.39,0.33</td>
<td>−1.29,0.21</td>
<td>0.80,0.87</td>
<td>0.74,0.83</td>
</tr>
<tr>
<td>UVic</td>
<td>0.35,0.68</td>
<td>0.69,0.76</td>
<td>0.59,0.74</td>
<td>−3.98,−0.14</td>
<td>−1.93,−0.44</td>
<td>−9.50,−0.41</td>
<td>0.91,0.87</td>
<td>−0.17,0.50</td>
</tr>
<tr>
<td>VIC</td>
<td>0.14,0.67</td>
<td>−3.41,0.10</td>
<td>−14.88,−0.45</td>
<td>−60.73,−0.74</td>
<td>−2.04,0.30</td>
<td>−0.32,0.61</td>
<td>0.83,0.87</td>
<td>−0.27,0.56</td>
</tr>
</tbody>
</table>

Figure 5. Mean annual gross primary productivity (GPP) from the permafrost RCN models and from the MOD17 product. The averaging period is 2000–2009 for GPP from the MOD17 product and all models with the exception of CLM4.5 (1995–2004); CoLM (1991–2000); and JULES (1991–2000). Spatial correlations between MOD17 GPP and each model GPP for all grids is shown at upper left in each map panel. Map panel at upper right is coefficient of variation (CV) for GPP. At each grid the CV is estimated from the mean and standard deviation across the nine models (MOD17 not included).

Figure 6. Distributions for mean annual GPP from the models and the MOD17 product over the averaging period listed in Fig. 5. The rectangles bracket the 25th and 75th percentiles. Whiskers extend to the 5th and 95th percentiles. Thick and thin horizontal lines mark the mean and median respectively.

In general, the models bracket the MOD17 estimates, with several models showing a larger spread and several showing a reduced spread. Regional averages from each model fall within ±20 % of the MOD17 average of 468 g C m⁻² yr⁻¹, with the exception of the LPJG model for which annual GPP is 40 % lower than MOD17.

For each model the spatial pattern in ER (not shown) closely matches the pattern in GPP, consistent with the strong dependence of autotrophic respiration and litterfall on vegetation productivity (Waring et al., 1998; Bond-Lamberty et al., 2004). Area-averaged GPP and ER are highly correlated \((r = 0.99, \text{Fig. 7})\). That is, models which simulate low (high) GPP also simulate low (high) ER.

3.1.3 Spatial patterns and area averages

In this study net ecosystem productivity (NEP) represents the net exchange of CO₂ between the land surface and the atmosphere. NEP is defined as the difference between GPP and ER. We do not examine other emission components of land-atmosphere CO₂ exchange (Hayes and Turner, 2012), as several of the models possess limited representation of disturbance processes important for carbon cycling in boreal forest...
regions (e.g. fire and forest harvest). The multimodel mean NEP is highest over the south-central part of the region and lowest in the tundra to the north (Fig. 8a). Only 0.3 % of the region is a net annual source of CO₂, notably two small areas in Scandinavia. Tundra areas are a net sink of approximately 15 g C m⁻² yr⁻¹ based on the multimodel mean NEP. As measured by the coefficient of variation (CV), the agreement in NEP among the models is highest across the boreal region and lowest in the tundra to the north and grasslands to the south (Fig. 8b). The multimodel mean NEP is approximately 20 g C m⁻² yr⁻¹ or 270 Tg C yr⁻¹ over the simulation period (Fig. 9). Among the models, NEP varies from 4 (UVic) to 48 (JULES) g C m⁻² yr⁻¹, a range that is double the multimodel mean. The UVic simulates a negative NEP (CO₂ source) for nearly half of the region, and the CoLM and MIROC for nearly 25 % of the region.

3.2 Temporal changes over the period 1960–2009

Figure 10 shows the time series of regionally averaged annual NEP each year over the period 1960–2009 for each model. Across the model group annual NEP is positive in most but not all years. Several models show a net source of CO₂ in some years, primarily during the earlier decades of the period. Among the models NEP increases by 0.01 to 0.79 g C m⁻² yr⁻², (3 to 340 % of the respective model means) based on a linear least squares (LLS) regression (Table 6). Seven of the models (CLM4.5, CoLM, ISBA, JULES, LPJG, MIROC, ORCHIDEE) show statistically significant trends at the $p < 0.01$. Taking averages over the first decade

Figure 7. Spatially averaged ER vs. GPP over the period 1960–2009. Horizontal and vertical lines span the range across the 5th and 75th percentiles for GPP and ER, respectively. The GPP 5th and 75th percentiles are shown in Fig. 6. NEP is equal to the difference GPP minus ER.

Figure 8. (a) Annual NEP (1960–2009) averaged across the nine models. Areas in blue are a net annual source of CO₂. (b) Coefficient of variation as estimated from the across model mean and standard deviation for each grid.

Figure 9. Distributions for mean annual NEP from the models over the averaging period listed in Fig. 5. Boxplot quartiles are as described in caption for Fig. 6.
weaken over the latter decades (Fig. 12). The uncertainty range for the multimodel mean shows that the region has been a net sink for CO\textsubscript{2} over the simulation period. Interestingly the uncertainty range reflects relatively better model agreement in annual NEP (lower variance) during the years 1960–1965 and in the low NEP years 1978 and 1996. Amid this increase there is evidence of a deceleration in NEP. The deceleration is apparent when examining trend magnitude and significance across all time intervals (minimum 20-year interval) over the simulation period (Fig. 13). Here several models (ISBA, LPJG, ORCHIDEE) exhibit weaker linear trends over time and all models show a lack of significant positive trends for time intervals spanning the latter decades (e.g. 1980–1999 or 1982–2009). While temporal trends in NEP are highly variable across the models, it is clear that the greatest increases in NEP occurred during the earliest decades of the simulation period. The LLS trend is significant for 20 of 42 (48 %) possible time periods beginning in 1975 or later, whereas 72 of 107 (67 %) are significant for periods starting in 1960–1962.

### 3.3 Residence Time

Annual estimates of residence time (RT) are calculated for each model and at each grid cell over the period 1960–2009 using model soil carbon storage and the rate of heterotrophic respiration ($R_\text{h}$). Among the models RT (long-term climatological mean) varies from 40 (CoLM) to 400 years (CLM4.5), and largely by model soil carbon amount, which
Table 6. Trend in GPP, ER, and NEP over simulation period for each model. Trend slopes (g C m\(^{-2}\) yr\(^{-2}\)) are estimated using an auto-regressive AR[1] model to account for temporal autocorrelation. Standard error for the regression is indicated in ( ). Standard deviation of the model means is shown in [ ]. Significant trends (p < 0.01) are denoted with an asterisk (*).

<table>
<thead>
<tr>
<th>Model</th>
<th>GPP</th>
<th>ER</th>
<th>NEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLM4.5</td>
<td>1.3*(0.18)</td>
<td>1.0*(0.15)</td>
<td>0.27*(0.06)</td>
</tr>
<tr>
<td>CoLM</td>
<td>1.3*(0.19)</td>
<td>0.9*(0.18)</td>
<td>0.31*(0.07)</td>
</tr>
<tr>
<td>ISBA</td>
<td>3.9*(0.29)</td>
<td>3.1*(0.23)</td>
<td>0.78*(0.11)</td>
</tr>
<tr>
<td>JULES</td>
<td>1.7(0.27)</td>
<td>1.3(0.19)</td>
<td>0.33(0.11)</td>
</tr>
<tr>
<td>LPJG</td>
<td>1.2*(0.11)</td>
<td>1.0*(0.11)</td>
<td>0.17*(0.06)</td>
</tr>
<tr>
<td>MIROC</td>
<td>1.9*(0.16)</td>
<td>1.7*(0.15)</td>
<td>0.24*(0.12)</td>
</tr>
<tr>
<td>ORCHIDEE</td>
<td>1.6*(0.15)</td>
<td>1.1*(0.13)</td>
<td>0.43*(0.08)</td>
</tr>
<tr>
<td>UVic</td>
<td>1.7*(0.18)</td>
<td>1.6*(0.18)</td>
<td>0.11(0.06)</td>
</tr>
<tr>
<td>UW-VIC</td>
<td>1.4*(0.12)</td>
<td>1.4*(0.13)</td>
<td>0.02(0.05)</td>
</tr>
<tr>
<td>mean</td>
<td>1.8[0.78]</td>
<td>1.5[0.64]</td>
<td>0.29[0.18]</td>
</tr>
</tbody>
</table>

The potential for alterations to the terrestrial sink of atmospheric CO\(_2\) across the high northern latitudes motivates our examination of model estimates of land-atmosphere exchanges of CO\(_2\) across the arctic drainage basin of Northern Eurasia. Validation of model estimates through comparisons to measured flux tower data is hindered by several factors. The limited extent of available measurements from a sparse regional tower network clearly challenges the validation of model estimates and, in turn, identification of model processes which require refinement. There are also inherent uncertainties in GPP and ER data derived from net ecosystem exchange (NEE) measurements at the eddy covariance tower sites. ER is generally assumed to equal NEE during nighttime hours (Lasslop et al., 2010). An empirical relationship is derived to estimate ER during that time and it is extrapolated into the daylight hours. GPP is then generally calculated as the difference between NEE and ER (accounting for appropriate signs). Since there is generally daylight for photosynthesis during the middle of the summer, ER could potentially be underestimated if primary production had occurred during the hours used for ER model calibration. Direct validation of the partitioning of measured NEE flux to GPP and ER is not possible. However, in a recent sensitivity study Lasslop et al. (2010) compared two independent methods for partitioning and found general agreement in the results. This agreement across methods increases our confidence in the partitioned GPP and ER estimates in the LaThuile FLUXNET data set. When measurements come from nearly-ideal sites the error bound on the net annual exchange of CO\(_2\) has been esti-
Figure 13. Magnitude of linear trend in NEP over given time interval for all trends significant at $p < 0.05$. For each model, linear trends are calculated for all time intervals of 20 years or more. For example, 1960–1979, 1960–1980, ..., 1990–2009. Intervals for which the trend is significant are marked with a line from the start to end year of the interval and shaded by the trend magnitude. As an example, one time interval is identified with a significant NEP trend for UW-VIC, from 1964–1993.

4.2 Model uncertainties contributing to errors in net CO$_2$ sink/source activity

Regionally averaged GPP is within 20% of the MOD17 average (470 g C m$^{-2}$ yr$^{-1}$) for 8 of the 9 models. While the models broadly capture the three major biomes across the region, a wide range in spatial GPP estimates is evident. This result may reflect differences in model forcings, initial conditions, parameterization and the dynamic vs static nature of vegetation and LAI (Table 2). While these differences make it difficult to unambiguously determine the underlying causes for many of the mismatches, the evaluations, in the context of prior studies, point to particular biases. The timing of peak summer GPP is generally well captured in most of the models (Fig. 4). Despite the agreement in peak GPP (and ER) timing, several models overestimate the small source of CO$_2$ before, and to some degree after, winter dormancy at the Hakasija sites and Zotino. Overestimates in GPP and ER are more common than underestimates (Table 4). Indeed, all errors are positive for site Hakasija and five of the seven models show relatively large overestimates in ER at Zotino. The tendency to overestimate GPP suggests that parametrizations and process specifications controlling primary production (e.g. # 1, 2, 3, 4, 6, 8 in Table 2) may require refinement. It should be noted that large seasonal flux errors (e.g. Keenan et al., 2012; Richardson et al., 2012; Schaefer et al., 2012) will appear as more modest monthly errors such as those noted in our analysis. While it is not possible to evaluate sources of error separately for $R_h$ and autotrophic respiration ($R_a$), our results and those from prior studies implicating $R_h$ in the model uncertainties (Dolman et al., 2012; Quegan et al., 2011) suggest...
a need for further investigation of model processes controlling respiration. Only one of the nine models, the CLM4.5, simulated limits on productivity due to nitrogen availability. None account for competition for nitrogen. Lack of accounting for nitrogen limits on photosynthesis may be leading to overestimates in simulated GPP, since nitrogen availability limits terrestrial carbon sequestration in boreal regions (Zäehle, 2013). While accounting for fire is important for estimates of impacts on recently disturbed areas, and may be contributing to the wide range in GPP exhibited by CLM4.5, CoLM, and LPJG (Fig. 6), climate variability is a more dominant influence on regional fluxes (Yi et al., 2013). Regarding errors in respiration rates, models with the highest soil carbon amounts (CLM4.5 and UW-VIC) exhibit relatively high ER rates when compared to the observations at several sites (Fig. 3). This tendency is consistent with results described by Exbrayat et al. (2013), who suggest that initial carbon pool size is the main driver of the response to warming, with the magnitude of the carbon pool strongly controlling the sensitivity of \( R_h \) to changes in temperature and moisture. While all of the models incorporate temperature and moisture in their formulations for \( R_h \), only three of the nine account for the effect of vegetation type on soil thermal dynamics. A wide range in process specifications for soil thermal dynamics is present across the models.

In a study of nine models from the TRENDY project, Peng et al. (2015) found that the models overestimate both GPP and ER, and underestimate NEE at most of the flux sites examined, and for the Northern Hemisphere based on upscaled measurements. A low NEE, or NEP, may be attributable to model biases in respiration exceeding those in productivity. Averaged across the nine models and the region of the present study, NEP of approximately 20 g C m\(^{-2}\) yr\(^{-1}\) (270 Tg C yr\(^{-1}\)) is broadly consistent with inventory assessments for Eurasian forests, which range between 93 and 347 Tg C yr\(^{-1}\) (Hayes et al., 2011). Quegan et al. (2011) concluded that NPP simulated by two DGVMs examined was nearly balanced by the models’ estimate of \( R_h \). Dolman et al. (2012) found that GPP increased during the years 1920 to 2008, with the GPP increase in the DGVMs balanced equally by increases in respiration. They reported NEP over the Russian territory as an average of three methods at nearly 30 g C m\(^{-2}\) yr\(^{-1}\). The DGVM average, however, was only 4.4 g C m\(^{-2}\) yr\(^{-1}\) and so low that the authors chose to remove it from their final carbon budget. This under-estimate was attributed to an excess in \( R_h \). While the mean NEP of 20 g C m\(^{-2}\) yr\(^{-1}\) in the present study is more consistent with the three-method average of Dolman et al. (2012) than their lower DGVM estimates, our comparisons against tower-based data and results of other studies suggest the sink strength is underestimated. Of the three models common to that study and the present one, the CLM4.5 and ORCHIDEE rank on the low end of model NEP magnitudes (Fig. 9).

Recent research points to phenology as one of the principle sources of error in model simulations of land-atmosphere exchanges of \( \text{CO}_2 \). Graven et al. (2013) found that the change in NEP simulated by a set of CMIP5 models could not account for the observed increase in the seasonal cycle amplitude in atmospheric \( \text{CO}_2 \) concentrations. They point to data showing that boreal regions have experienced greening and shifting age composition which strongly influence NEP and suggest that process models under-represent the observed changes. Model inability to capture canopy phenology has been identified as a major source of model uncertainty leading to large seasonal errors in carbon fluxes such as GPP (Keenan et al., 2012; Richardson et al., 2012; Schaefer et al., 2012). Indeed, evaluated against flux tower data across the eastern USA, current state-of-the-art terrestrial biosphere models have been found to mis-characterize the temperature sensitivity of phenology, which contributes to poor model
4.3 Uncertainties in temporal trend estimates

Uncertainties exist as to whether tundra areas are presently a net sink or source of CO\(_2\). Across tundra regions, process models indicate a stronger sink in the 2000s compared with the 1990s, attributable to a greater increase in vegetation net primary production than heterotrophic respiration in response to warming (McGuire et al., 2000). However, a greater increase in winter CO\(_2\) emissions in tundra may help resolve uncertainties in processes within land surface models and provide a means to connect a warming climate with vegetation changes, permafrost thaw and CO\(_2\) dynamics.
the soil were greater than the enhancement in decomposition. In a recent study involving CMIP5 models, Carvalhais et al. (2014) found that while the coupled climate/carbon-cycle models reproduce the latitudinal patterns of carbon turnover times, differences between the models of more than one order of magnitude were also noted. The authors suggest that more accurate descriptions of hydrological processes and water–carbon interactions are needed to improve the model estimates of ecosystem carbon turnover times. The reduction in soil carbon residence time may at least partially be a direct response to increasing NEP, rather than through warming effects on respiration. A recent study (Koven et al., 2015) using a set of simulations from five CMIP5 models found that, because heterotrophic respiration equilibrates faster to the increasing NPP than the soil carbon stocks, increased productivity leads to reductions in inferred residence times even when there are no changes to the environmental controls on decomposition rates, a process they refer to as false priming. Because the experimental protocol analyzed here does not include a fixed-climate simulation, it is not possible to unambiguously separate the contribution from the false priming effect from that due to warming-related respiration increases, but the fact that soil C stocks increase over the period of simulation suggests that it is the dominant effect. Apart from climatological factors, vegetation growth is also dependent on biological nitrogen availability. Failure to account for nitrogen limitation may thus impart a bias in the modeled carbon flux estimates. However, more process models are incorporating linkages between carbon and nitrogen dynamics (Thornton et al., 2009). Given the broad range in spatial patterns in GPP across the models, a closer examination of processes related to nitrogen limitations and primary productivity is needed. The lower rate of NEP increase over the latter decades of the simulation period suggests a weakening of the land CO$_2$ sink, driven by increased $R_h$ from warming, associated permafrost thaw, and an upward trend in fire emissions (Hayes et al., 2011).

As the climate warms, the amount of carbon emitted as CH$_4$ and CO$_2$ will depend on whether soils become wetter or drier. A synthesis of observations and models points to intensification of the pan-Arctic hydrological cycle over recent decades (Rawlins et al., 2010), manifested prominently by increasing river discharge from Northern Eurasia (Peterson et al., 2002). In addition to hydrological cycle intensification and deepening soil active layer (Romanovsky et al., 2010), rapid thaw and ground collapse will also likely alter the landscape and impact land-atmosphere carbon exchanges. Land surface models are now beginning to implement new process formulations to account for these fine scale perturbations. Several of the models examined in this study incorporate the effect of soil freeze-thaw state on decomposition of organic carbon (Table 2). Only four of the nine models, however, account for methane emissions. Six simulate talik formation, and among these a variety of approaches are employed to compute snow insulation type.

5 Conclusions

Outputs from a suite of land surface models were evaluated against independent data sets and used to investigate elements of the land-atmosphere exchange of CO$_2$ across Northern Eurasia over the period 1960–2009. The models exhibit a wide range in spatial patterns and regional mean magnitudes. Compared to tower-based data, overestimates in both GPP and ER are noted in several of the models, with larger errors in ER relative to GPP, particularly for the comparisons at the southern higher productivity sites. Regarding agreement in the spatial pattern in GPP, less than half of the variance in GPP expressed in the MOD17 product is explained by the GPP pattern from four of the nine models. The NEP increases range from 3 to 340 % of the model means, further illustrating uncertainties in sink strength. The models exhibit a decrease in residence time of the soil carbon pool that is driven by an increase in $R_h$, simultaneous with an increase in soil carbon storage. This result suggests that net primary productivity (NPP) inputs to the pool increased more than $R_h$ fluxes out. Among the quantities examined, uncertainties are lowest for GPP across the forest/taiga biome and highest for residence time over tundra and steppe areas. Amid the uncertainty in NEP magnitude, the results of this study and others suggest that the CO$_2$ sink of the region is underestimated.

Several recommendations are made as a result of this analysis. The range in area and climatological mean NEP across the models, more than double the mean value, illustrates the considerable uncertainty in the magnitude of the contemporary CO$_2$ sink. The results of the site-level comparison point to a need to better understand the connections between model-simulated productivity rates, soil dynamics controlling heterotrophic respiration rates, and associated uncertainties in total ER. Given the strong connections between soil thermal and hydrological variations and soil respiration, we recommend that model improvements are targeted at processes and parameterizations controlling respiration with depth in the soil profile. These validation efforts are especially important given the likelihood of net carbon transfer from ecosystems to the atmosphere from permafrost thaw (Schuur and Abbott, 2012; Schuur et al., 2015). Model responses to CO$_2$ fertilization and nitrogen limitation, processes largely underrepresented in the models, should be evaluated in the context of ecosystem productivity. While insights have been gained by examining the model estimates of GPP, ER, and NEP, an improved understanding of net CO$_2$ sink/source dynamics will require the continued development and application of model formulations for carbon emissions from fire and other disturbances. The limited number of measured site data across this important region clearly hampers model assessments, highlighting the critical need for new field, tower, and aircraft data for model validation and parametrization. Specifically, new observations in the boreal zone are required to better evaluate model bi-
cases documented in this and in other recent studies. Moreover, our finding of biases in CO$_2$ source activity during the shoulder seasons points to a critical need for observations during autumn, winter, and spring. Given our results, conclusions drawn from studies which use a single model should be viewed cautiously in the absence of rigorous validation against observations across the region of interest.

New observations from current and upcoming field campaigns such as Carbon in Arctic Reservoirs Vulnerability Experiment (CARVE) and the Arctic Boreal Vulnerability Experiment (ABoVE) should be used to confirm the results of this study. Future model evaluations will benefit from continued development of consistent benchmarking data sets from field measurements and remote sensing. Regarding tower data, any new measurements must be supported by refinements in the models used to partition the measured NEE flux into GPP and ER components. Regarding these and similar model intercomparisons, investments must be made which will minimize or eliminate differences in a priori climate forcings used in the simulations. At a programmatic level support for these activities should lead to well-designed model intercomparisons which minimize, to the extent possible, differences in model spinup, forcings and other elements which confound model intercomparisons.


Acknowledgements. This research was supported by the US National Aeronautics and Space Administration NASA grant NNX11AR16G and the Permafrost Carbon Network (http://www.permafrostcarbon.org/) funded by the National Science Foundation. The MODIS Land Cover Type product data was obtained through the online Data Pool at the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota (https://lpdaac.usgs.gov/data_access). We thank Hans Dolman and a second reviewer for their insightful comments which helped improve the manuscript. We thank the researchers working at FLUXNET sites for making available their CO$_2$ flux data. We also thank Eugenie Euskirchen and Dan Hayes for comments on an earlier version of the manuscript, and Yonghong Yi for assistance with the FLUXNET data. Charles Koven was supported by the Director of the Office of Biological and Environmental Research, Office of Science, US Department of Energy, under Contract DE-AC02-05CH11231 as part of the Regional and Global Climate Modeling Program (RGCM). Eleanor J. Burke was supported by the Joint UK DECC/Defra Met Office Hadley Centre Climate Programme (GA01101) and the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 282700. Bertrand Decharme and Christine Delire were supported by the French Agence Nationale de la Recherche under agreement ANR-10-CEPL-012-03. Several of the authors were funded by the European Union 7th Framework Programme under project Page21 (grant 282700). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the US Government.

Edited by: U. Seibt

References


Hobie, S. E., Schimel, J. P., Trombore, S. E., and Randerson, J. R.: Controls over carbon storage and turnover in high-latitude


