Manuscript prepared for Biogeosciences
with version 5.0 of the LaTeX class copernicus.cls.
Date: 13 May 2015

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[13]

[1]Climate System Research Center, Department of Geosciences, University of Massachusetts, Amherst, MA USA
[2]U.S. Geological Survey, Alaska Cooperative Fish and Wildlife Research Unit, University of Alaska, Fairbanks, Alaska 99775 USA
[3]FLBS/NTSG, Univ. 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, F-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

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

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 degree 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. 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 ten 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 ten 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 datasets 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 E. A. G. 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 (Melillo 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 any 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) find 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 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, 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 (Figure 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, Yenesey, 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 this writing were included in the analysis (http://www.permafrostcarbon.org, accessed May 10, 2014). The simulation protocol allowed for the choice of a model's driving datasets 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 degree 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 CO_2 source and sink dynamics. These include model land cover initialization, time series forcings, light use efficiency, and CO_2 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 dataset (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 here by their locations: Chersky (CHE), Chokurdakh (COK), Hakasija (HAK), and Zotino (ZOT). Data from three towers are available for Hakasija; HAK1 is in an area of grassland-steppe; HAK2 is grassland; HAK3 an abandoned agricultural fields. Chersky and Chokurdakh are in northeast Russia in the general zone of tundra vegetation. Hakasija and Zotino are in an area of generally higher productivity in southern Siberia (Figure 1). Data are available for years 2002–2004 at Chersky, Hakasija and Zotino, and 2003–2005 at Chokurdakh. General characteristics of these sites are summarized in Table 3. In this dataset GPP and ER are derived from an empirical model driven by field-based eddy covariance measurements of net ecosystem CO_2 exchange (NEE) using methodologies described in Reichstein et al. (2005).

2.3.2 Satellite-based estimates of GPP

Satellite data driven estimates of annual total GPP are also obtained from the MODIS (Moderate Resolution Imaging Spectroradiometer) MOD17 operational product (Running et al., 2004; Zhao et al., 2005). The MOD17 product has been derived operationally from the NASA EOS MODIS sensors since 2000 and provides a globally consistent and continuous estimation of vegetation productivity at 1-km resolution and 8-day intervals. MOD17 uses a light use efficiency algorithm driven by global land cover classification and canopy fractional photosynthetically active radiation (FPAR) inputs from MODIS. The product also uses daily surface meteorology inputs from global reanalysis data (Zhao and Running, 2010), and land cover class specific biophysical response functions to estimate the conversion efficiency of canopy absorbed photosynthetically active radiation to vegetation biomass (g C MJ^{-1}) and GPP (Running et al., 2004). The MOD17 algorithms and productivity estimates have been extensively evaluated for a range of regional and global applications, including northern, boreal and Arctic domains (Heinsch et al., 2006; Turner et al., 2006; Zhang et al., 2008; Zhao and Running, 2010). We use the MOD17 Collection 5 product, which has undergone five major reprocessing improvements since 2000. The MOD17 data are used in this study as a consistent satellite-derived baseline for evaluating GPP simulations from the detailed carbon process models.

3 Results

3.1 Model evaluation and benchmarking

3.1.1 Site-level evaluations

Confident assessment of uncertainties in land-atmosphere CO_2 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 mean-absolute error (MAE) and mean bias error (MBE) are preferable to those based on squared differences (Willmott and Matsuura, 2005; Willmott et al., 2011). Model performance is evaluated here using the mean bias error (MBE), defined as the 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 tower value.

As shown in (Figure 2), MOD17 GPP agrees well with the tower estimates for Chersky and Chokurdakh, with MBE over the three years of -2 and -11 g C m^{-2} month^{-1}, respectively (Table 4). MOD17 GPP broadly agrees with the observations at Hakasija and Zotino. Average MBEs are 13 and 10 g C m^{-2} month^{-1}, respectively, for these sites with higher productivity than Chersky and Chokurdakh. Averaged across all models the error in GPP is 7, 34, 34 and 13 g C m^{-2} month^{-1} for Chersky, Chokurdakh, Hakasija and Zotino, respectively. The MBE for ER are 8, 35, 43 and 33 g C m^{-2} month^{-1}, respectively.

Overall the models simulate fairly well the seasonal cycle in GPP (Figure 2) and ER (Figure 3), including the timing of peak CO_2 drawdown. Modest overestimates are noted near growing season peak at Hakasija and Zotino. However, for all four sites significant over- and under- estimates in GPP and ER are also noted (Table 4). For the two sites in the south there is a tendency for overestimation in GPP and ER. All models overestimate both GPP and ER at Hakasija. Seven of the nine models overestimate GPP and ER at Zotino, with ER overestimated by a considerable degree. Overestimates in ER for Hakasija and Zotino during late summer and autumn are particularly noteworthy. An ANOVA test was carried out to determine whether model errors in ER exceed the errors in GPP. The tests confirm that that ER errors are greater on average than the GPP errors for comparisons where (i) ER errors for all sites are pooled together and compared against GPP pooled across all sites and (ii) ER and GPP errors for the two northern sites are pooled and compared against ER and GPP errors from the two southern sites.

The tendency to overestimate ER leads to discrepancies in net CO_2 source (negative NEP) at Hakasija and Zotino, particularly in autumn (Figure 4). Average NEP errors are -11 and -20 g C m^{-2} month^{-1} for Hakasija and Zotino, respectively (Table 4). Errors in the magnitude and timing

of NEP prior to and following the dormant season are much smaller at Chersky, and to some extent Chokurdakh. However, a lack of available tower-based data during the colder months limits the robustness of our assessments during that time of year.

We further evaluate model performance through two additional error metrics, the refined index of agreement (d_r) (Willmott et al., 2011) and the Nash-Sutcliffe coefficient of efficiency (E) (Nash and Sutcliffe, 1970). As described by Willmott et al. (2011), the refined index of agreement (d_r) involves the sum of the magnitudes of the differences between the model-predicted and observed deviations about the observed mean, relative to the sum of the magnitudes of the perfect-model (model predicted = observed) and observed deviations about the observed mean. It is bounded between -1 and $+1$. When d_r equals 0.0, it signifies that the sum of the magnitudes of the errors and the sum of the perfect-model-deviation and observed-deviation magnitudes are equivalent. Like d_r, the Nash-Sutcliffe E considers observed deviations within the basis of comparison. For both metrics, values closer to 1 indicate higher model accuracy. Nash-Sutcliffe's E is also positively correlated with d_r. Values of E less than zero occur when the residual model variance is larger than the data variance.

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. As well, 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 Figure 5. The MOD17 product clearly captures three distinct landcover 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 landcover product (Figure 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 Figure 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 Figure 5), agreement in GPP is best across the higher productivity tiaga biome. Figure 6 shows the distribution of GPP for all grids of each model. 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 $\pm$ 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, Figure 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 CO_2 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 CO_2 exchange (Hayes and Turner, 2012), as several of the models posses 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 (Figure 8a). Only 0.3% of the region is a net annual source of CO_2, 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 (Figure 8b). The multimodel mean NEP is approximately 20 g C m^{-2} yr^{-1} or 270 Tg C yr^{-1} over the simulation period (Figure 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 (CO_2 source) for nearly half of the region, and the CoLM and MIROC for nearly 25% of the region.

3.2 Temporal changes over 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 CO_2 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}, (5 to 40 g C m^{-2} total over the period) 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 (1960–1969) and last decade (2000–

2009) we estimate that the NEP change ranges from 10% to 400% of the first decade mean, with a nine model average of 135%. For each model the GPP trend magnitude exceeds the ER trend magnitude (Table 6), hence the increase in NEP over time. The increases from the first to last decade of the simulations range from 9–35% of the early decade average for GPP and 8–30% for ER. Total cumulative NEP over the 50 year period and averaged across all models is approximately 12 (range 3–20) Pg C (Figure 11). Averaged across the models, NEP exhibits an increase during mainly the earliest decades that tends to weaken over the latter decades (Figure 12). The uncertainty range for the multimodel mean suggests that the region has been a net sink for CO_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 (Figure 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 (eg. 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 heteorotrophic respiration (R_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 varies by an order of magnitude across the models. Over the period examined all of the models simulate a statistically significant ($p < 0.01$) decrease in the regionally-averaged RT. Across the models the decrease from first to last decade of the study period ranges from -5% to -16% of each model's mean. The decline occurs amid an increase in SOC storage over time. All models with the exception of CoLM simulate a statistically significant increase in soil carbon and all exhibit an increase in R_h. The increases in carbon storage range from 0.2% to 3.6% while the increases in R_h range from 7% to 22%. Likewise the models simulate an increase in the the rate of net primary production (NPP) of 8% to 30%. Across the model group the change in RT is highly correlated (r = 0.99) with change in R_h. In essence, higher rates in R_h and NPP led to a decrease in soil carbon RT, with increased soil carbon storage resulting from enhanced vegetation productivity and litterfall inputs.

The spatial pattern in RT changes suggests that controlling influences are leading to both decreases and increases over different parts of the region. The largest decreases are found across north-central

Russia and the eastern third of the domain (Figure 14a). The decline in RT is statistically significant ($p < 0.01$) for just over 46% of the region. exceeding -20% for approximately 16% of the region. An increase in RT is noted for less than 5% of the region, including a small area in the far north and across extreme southern parts of the region. The change, however, is not significant in those areas. The CV map (Figure 14b) lends further confidence to the RT decreases across much of the center of the region. High uncertainties (CVs > 10) are noted in the areas where the multimodel average suggests an increase in RT.

4 Discussion

4.1 Uncertainties in tower-based measurements

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 challenged 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 dataset. When measurements come from nearly ideal sites the error bound on the net annual exchange of CO_2 has been estimated to be less than ± 50 g C m^{-2} yr^{-1} (Baldocchi, 2003). Systematic errors in eddy covariance fluxes due to non-ideal observation conditions are uncertain at this time. Total error is likely below the value of 200 g C m^{-2} yr^{-1} that has been conservatively estimated (Reichstein et al., 2007). The model errors estimated in this present study often exceed that level for site Hakasija and, for a few models, Zotino as well. Lastly, any conclusions about the CO_2 sink strength drawn from such a limited number of eddy covariance sites should be viewed with caution.

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 (Figure 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 (eg. # 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) suggests 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 (Zaehle, 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 (Figure 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 (Figure 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} (Figure 9) (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 from 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 underestimate 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 in 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 (Figure 9).

Recent research points to phenology as one of the principle sources of error in model simulations of land-atmosphere exchanges of 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 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 US, current state-of-the-art terrestrial biosphere models have been found to mis-characterize the temperature sensitivity of phenology, which contributes to poor model performance (Keenan et al., 2014). Two recent studies using eight land surface models from the TRENDY comparison (Murray-Tortarolo et al., 2013) (several examined in the present study) and 11 coupled carbon-climate models (Anav et al., 2013) have found that models consistently overestimate leaf area index (LAI) and have a longer growing season, mostly due to a later autumn dormancy, compared to satellite data. However, when estimated using model GPP, dormancy was much earlier than previously predicted using LAI. The authors conclude that the models are keeping inactive leaves for longer than they should, but with little impact on carbon cycle fluxes. Anav et al. (2013) further suggested that it was unlikely that differences in climate in the coupled models were solely responsible for the positive bias. (Fisher et al., 2014) also concluded that variability in land model model fluxes was driven primarily by differences in model physics rather than differences in forcing data.

Simulated R_h estimates among the DGVMs analyzed by Dolman et al. (2012) vary in the range between 200 to 225 g C m^{-2} yr^{-1}. In the present study the nine model average is 190 g C m^{-2} yr^{-1}. Dolman et al. (2012) point to lower estimates from Kurganova and Nilsson (2003) of 139 g C

m^{-2} yr^{-1} and Schepaschenko et al. (2013) of 174 g C m^{-2} yr^{-1} as being more representative for the region. Our benchmark comparisons of ER against tower-based data are consistent with these recent studies and suggest that several models are overestimating R_h, particularly over the boreal forest zone. Among the model examined in this study a wide range in soil carbon parameterizations is noted (Table 2). Not surprisingly the effects of active layer depth on the availability of soil organic carbon for decomposition and combustion has been recognized as a key sensitivity in process models (Hayes et al., 2014). Regarding below-ground processes, model parameterizations and processes controlling carbon storage and turnover such as litter decomposition rates and biological activity in frozen soils (Hobbie et al., 2000) require close examination as well. Model simulations of R_h during the nongrowing season are sensitive to the presence or absence of snow (McGuire et al., 2000), suggesting that future studies of mechanisms controlling 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.

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., 2012; Belshe et al., 2013). The spatial pattern in multimodel mean NEP in this study points to small areas in Scandinavia ($<$ 1% of the domain) as sources of CO_2. Broadly, areas classified as tundra are a modest CO_2 sink of approximately 15 g C m^{-2} yr^{-1}. Across-model standard deviations in areas of small positive and negative NEP are a factor of ten or more greater than the multimodel mean in some areas, and are generally high across the tundra (Figure 8b). Estimates of NEP sink magnitudes must be interpreted with caution given that the models in general possess inadequate representation of disturbances which are an important component of the overall carbon balance (Hayes et al., 2011). Among this model group, four of the nine account for fire. The nature of model initialization and spinup is also a strong influence on simulated NEP changes. For example, spin-up procedures can explain some of the discrepancies. ISBA, for instance, was equilibrated using the 10 coldest year of the WATCH forcing repeatedly to emulate preindustrial climate. As a result, soil and vegetation carbon were fairly low at the beginning of the 20th century run, much lower than the equilibrium that would result from the 1960s climate. Due to the large characteristic time scale of soil carbon, part of ISBA's large trend during the 1960–2009 period (Figure 11) can be traced to the climate used for the model spinup procedure.

Previous studies have pointed to changes in the seasonal drawdown and release of CO_2 across the northern high latitudes (Graven et al., 2013). A change in the seasonal cycle of GPP and ER is also noted (figure not shown), with the models analyzed in this study simulating a relatively higher productivity rate from late spring to mid-summer. Indeed, increased productivity did not occur uni-

formly across the growing season, as most of the models show little change in August or September NEP over time. The models also simulate little change in NEP over the cold season. Greater productivity in spring and early summer may be due in part to earlier spring thawing and temporal advance in growing season initiation (McDonald et al., 2004), whereas GPP and NEP are more strongly constrained by moisture limitations later in the growing season (Yi et al., 2014). Extension of the growing season is therefore attributed more to a regional warming driven advance in spring thaw than a delay in autumn freeze-up (Kimball et al., 2006; Euskirchen et al., 2006; Kim et al., 2012) which correlates with regional annual evapotranspiration for the region above 40°N (Zhang et al., 2011). There are, however, signs of a delay in the timing of the fall freeze (-5.4 days decade^{-1}) across Eurasia over the period 1988–2002 (Smith et al., 2004) consistent with fall satellite snow cover (SCE) increases, and attributed to greater fall/winter snowfall and regional cooling (Cohen et al., 2012). Consistent with the advance in spring thaw, the models examined here show a greater NEP increase in spring compared to autumn.

Soil carbon storage across the region increased significantly over the study period in eight of the nine models. A relatively larger increase in R_h is correlated strongly with the associated decline in soil carbon residence time. This suggests that amid recent warming, vegetation carbon inputs to 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 production 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. Over the simulation period NEP increases between 10% and 400% of the respective model mean. 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/tiaga biome and highest for residence time over tundra and steppe areas. Amid the uncertainty in NEP magnitude, the results of this study and others suggests 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 errors 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 E. A. G. 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 biases 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 the study. Future model evaluations will benefit from continued development of consistent benchmarking datasets 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 forcings and other elements which confound model intercomparisons.

Author Contributions

M. A. Rawlins conceived the study with input from A. D. McGuire, J. K. Kimball and P. Dass. Co-authors D. Lawrence, E. Burke, X. Chen, C. Delire, C. Koven, A. MacDougall, S. Peng, A. Rinke, K. Saito, W. Zhang, R. Alkama, T. J. Bohn, P. Ciais, B. Decharme, I. Gouttevin, T. Hajima, D. Ji, G. Krinner, D. P. Lettenmaier, P. Miller, J. C. Moore, B. Smith, and T. Sueyoshi provided model simulation outputs. M. A. Rawlins analyzed the outputs and other data. M. A. Rawlins prepared the manuscript with contributions from all co-authors.

Acknowledgements. This research was supported by the U.S. 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, U.S. 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 n°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 7^{th} 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 U.S. Government.

References

Adam, J. C. and Lettenmaier, D. P.: Adjustment of global gridded precipitation for systematic bias, Journal of Geophysical Research: Atmospheres (1984–2012), 108, 2003. 27

Adam, J. C., Clark, E. A., Lettenmaier, D. P., and Wood, E. F.: Correction of global precipitation products for orographic effects, Journal of Climate, 19, 15–38, 2006. 27

Anav, A., Murray-Tortarolo, G., Friedlingstein, P., Sitch, S., Piao, S., and Zhu, Z.: Evaluation of land surface models in reproducing satellite Derived leaf area index over the high-latitude northern hemisphere. Part II: Earth system models, Remote Sensing, 5, 3637–3661, 2013. 13

Avis, C. A., Weaver, A. J., and Meissner, K. J.: Reduction in areal extent of high-latitude wetlands in response to permafrost thaw, Nature Geoscience, 4, 444–448, 2011. 5

Baldocchi, D.: TURNER REVIEW No. 15.'Breathing'of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems, Australian Journal of Botany, 56, 1–26, 2008. 6, 29, 30

Baldocchi, D. D.: Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future, Global Change Biology, 9, 479–492, 2003. 11

Belshe, E., Schuur, E., and Bolker, B.: Tundra ecosystems observed to be CO2 sources due to differential amplification of the carbon cycle, Ecology letters, 16, 1307–1315, 2013. 3, 14

Best, M., Pryor, M., Clark, D., Rooney, G., Essery, R., Ménard, C., Edwards, J., Hendry, M., Porson, A., Gedney, N., et al.: The Joint UK Land Environment Simulator (JULES), model description–Part 1: energy and water fluxes, Geoscientific Model Development, 4, 677–699, 2011. 5

Bohn, T., Podest, E., Schroeder, R., Pinto, N., McDonald, K., Glagolev, M., Filippov, I., Maksyutov, S., Heimann, M., Chen, X., et al.: Modeling the large-scale effects of surface moisture heterogeneity on wetland carbon fluxes in the West Siberian Lowland., Biogeosciences, 10, 2013. 5

Bond-Lamberty, B., Wang, C., and Gower, S. T.: A global relationship between the heterotrophic and autotrophic components of soil respiration?, Global Change Biology, 10, 1756–1766, 2004. 9

Carvalhais, N., Forkel, M., Khomik, M., Bellarby, J., Jung, M., Migliavacca, M., Mu, M., Saatchi, S., Santoro, M., Thurner, M., Weber, U., Ahrens, B., Beer, C., Cescatti, A., Randerson, J. T., and Reichstein, M.: Global covariation of carbon turnover times with climate in terrestrial ecosystems, Nature, 514, 213–217, doi:http://dx.doi.org/10.1038/nature13731 10.1038/nature13731, http://www.nature.com/nature/journal/v514/n7521/abs/nature13731.html#supplementary-information, 2014. 15

Chapin III, F. S., Sturm, M., Serreze, M. C., McFadden, J. P., Key, J. R., Lloyd, A. H., McGuire, A. D., Rupp, T. S., Lynch, A. H., Schimel, J. P., Beringer, J., Chapman, W. L., Epstein, H. E., Euskirchen, E. S., Hinzman, L. D., Jia, G., Ping, C.-L., Tape, K. D., Thompson, C. D. C., Walker, D. A., and Welker, J. M.: Role of Land-Surface Changes in Arctic Summer Warming, Science, 310, 657–660, dOI: 10.1126/science.1117368, 2005. 2, 5

Clark, D., Mercado, L., Sitch, S., Jones, C., Gedney, N., Best, M., Pryor, M., Rooney, G., Essery, R., Blyth, E., et al.: The Joint UK Land Environment Simulator (JULES), model description–Part 2: carbon fluxes and vegetation dynamics, Geoscientific Model Development, 4, 701–722, 2011. 5

Cohen, J. L., Furtado, J. C., Barlow, M. A., Alexeev, V. A., and Cherry, J. E.: Arctic warming, increasing snow

cover and widespread boreal winter cooling, Environmental Research Letters, 7, 014 007, 2012. 15

Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A., and Totterdell, I. J.: Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model, Nature, 408, 184–187, doi:10.1038/35041539, http://www.nature.com/nature/journal/v408/n6809/abs/408184a0.html, 2000.

Crutzen, P. J.: The "Anthropocene", in: Earth System Science in the Anthropocene, edited by Ehlers, P. D. E. and Krafft, D. T., pp. 13–18, Springer Berlin Heidelberg, http://link.springer.com/chapter/10.1007/3-540-26590-2_3, 2006. 2

Decharme, B., Boone, A., Delire, C., and Noilhan, J.: Local evaluation of the Interaction between Soil Biosphere Atmosphere soil multilayer diffusion scheme using four pedotransfer functions, Journal of Geophysical Research: Atmospheres (1984–2012), 116, 2011. 5

Dolman, A. J., Shvidenko, A., Schepaschenko, D., Ciais, P., Tchebakova, N., Chen, T., van der Molen, M. K., Belelli Marchesini, L., Maximov, T. C., Maksyutov, S., and Schulze, E.-D.: An estimate of the terrestrial carbon budget of Russia using inventory-based, eddy covariance and inversion methods, Biogeosciences, 9, 5323–5340, doi:10.5194/bg-9-5323-2012, http://www.biogeosciences.net/9/5323/2012, 2012. 3, 5, 12, 13

Dutta, K., Schuur, E. a. G., Neff, J. C., and Zimov, S. A.: Potential carbon release from permafrost soils of Northeastern Siberia, Global Change Biology, 12, 2336–2351, doi:10.1111/j.1365-2486.2006.01259.x, http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2486.2006.01259.x/abstract, 2006. 3

Euskirchen, E., McGuire, A. D., Kicklighter, D. W., Zhuang, Q., Clein, J. S., Dargaville, R., Dye, D., Kimball, J. S., McDonald, K. C., Melillo, J. M., et al.: Importance of recent shifts in soil thermal dynamics on growing season length, productivity, and carbon sequestration in terrestrial high-latitude ecosystems, Global Change Biology, 12, 731–750, 2006. 3, 15

Exbrayat, J.-F., Pitman, A. J., Zhang, Q., Abramowitz, G., and Wang, Y.-P.: Examining soil carbon uncertainty in a global model: response of microbial decomposition to temperature, moisture and nutrient limitation, Biogeosciences, 10, 7095–7108, doi:10.5194/bg-10-7095-2013, http://www.biogeosciences.net/10/7095/2013/, 2013. 12

Fisher, J., Sikka, M., Oechel, W., Huntzinger, D., Melton, J., Koven, C., Ahlström, A., Arain, A., Baker, I., Chen, J., et al.: Carbon cycle uncertainty in the Alaskan Arctic, Biogeosciences Discussions, 11, 2887–2932, 2014. 3, 4, 13

Goetz, S. J., Bunn, A. G., Fiske, G. J., and Houghton, R.: Satellite-observed photosynthetic trends across boreal North America associated with climate and fire disturbance, Proceedings of the National Academy of Sciences of the United States of America, 102, 13 521–13 525, 2005. 3

Goetz, S. J., Fiske, G. J., and Bunn, A. G.: Using satellite time-series data sets to analyze fire disturbance and forest recovery across Canada, Remote Sensing of Environment, 101, 352–365, 2006. 3

Gouttevin, I., Menegoz, M., Dominé, F., Krinner, G., Koven, C., Ciais, P., Tarnocai, C., and Boike, J.: How the insulating properties of snow affect soil carbon distribution in the continental pan-Arctic area, Journal of Geophysical Research: Biogeosciences (2005–2012), 117, 2012. 5

Graven, H., Keeling, R., Piper, S., Patra, P., Stephens, B., Wofsy, S., Welp, L., Sweeney, C., Tans, P., Kelley, J., et al.: Enhanced Seasonal Exchange of CO_2 by Northern Ecosystems Since 1960, Science, 341, 1085–1089, 2013. 2, 3, 13, 14

Groisman, P. and Soja, A. J.: Ongoing climatic change in Northern Eurasia: justification for expedient research,

Environmental Research Letters, 4, 045 002, 2009. 4

Groisman, P. Y. and Bartalev, S. A.: Northern Eurasia Earth Science Partnership Initiative (NEESPI), science plan overview, Global and Planetary Change, 56, 215–234, 2007. 4

Groisman, P. Y., Clark, E. A., Lettenmaier, D. P., Kattsov, V. M., Sokolik, I. N., Aizen, V. B., Cartus, O., Chen, J., Schmullius, C. C., Conard, S., et al.: The Northern Eurasia earth science partnership: an example of science applied to societal needs, Bulletin of the American Meteorological Society, 90, 671–688, 2009. 4

Harris, I., Jones, P., Osborn, T., and Lister, D.: Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset, International Journal of Climatology, 34, 623–642, 2014. 27

Hayes, D. and Turner, D.: The need for "apples-to-apples" comparisons of carbon dioxide source and sink estimates, Eos, Transactions American Geophysical Union, 93, 404–405, 2012. 9

Hayes, D. J., McGuire, A. D., Kicklighter, D. W., Gurney, K., Burnside, T., and Melillo, J. M.: Is the northern high-latitude land-based CO_2 sink weakening?, Global Biogeochemical Cycles, 25, GB3018, doi:10.1029/2010GB003813, 2011. 3, 13, 14, 16

Hayes, D. J., Kicklighter, D. W., McGuire, A. D., Chen, M., Zhuang, Q., Yuan, F., Melillo, J. M., and Wullschleger, S. D.: The impacts of recent permafrost thaw on land–atmosphere greenhouse gas exchange, Environmental Research Letters, 9, 045 005, 2014. 14

Heinsch, F. A., Zhao, M., Running, S. W., Kimball, J. S., Nemani, R. R., Davis, K. J., Bolstad, P. V., Cook, B. D., Desai, A. R., Ricciuto, D. M., et al.: Evaluation of remote sensing based terrestrial productivity from MODIS using regional tower eddy flux network observations, Geoscience and Remote Sensing, IEEE Transactions on, 44, 1908–1925, 2006. 6

Hobbie, S. E., Schimel, J. P., Trumbore, S. E., and Randerson, J. R.: Controls over carbon storage and turnover in high-latitude soils, Global Change Biology, 6, 196–210, doi:10.1046/j.1365-2486.2000.06021.x, http://onlinelibrary.wiley.com/doi/10.1046/j.1365-2486.2000.06021.x/abstract, 2000. 14

Högberg, P., Nordgren, A., Buchmann, N., Taylor, A. F. S., Ekblad, A., Högberg, M. N., Nyberg, G., Ottosson-Löfvenius, M., and Read, D. J.: Large-scale forest girdling shows that current photosynthesis drives soil respiration, Nature, 411, 789–792, doi:10.1038/35081058, http://www.nature.com/nature/journal/v411/n6839/full/411789a0.html, 2001. 3

International Permafrost Association Standing Committee on Data Information and Communication (comp.). : Circumpolar Active-Layer Permafrost System (CAPS), Tech. rep., National Snow and Ice Data Center, http://dx.doi.org/10.7265/N5SF2T3B., 2003.

Ji, D., Wang, L., Feng, J., Wu, Q., Cheng, H., Zhang, Q., Yang, J., Dong, W., Dai, Y., Gong, D., et al.: Description and basic evaluation of Beijing Normal University Earth System Model (BNU-ESM) version 1, Geoscientific Model Development, 7, 2039–2064, 2014. 5

Kalnay, E., Cai, M., Li, H., and Tobin, J.: Estimation of the impact of land-surface forcings on temperature trends in eastern United States, Journal of Geophysical Research: Atmospheres (1984–2012), 111, 2006. 27

Keenan, T., Baker, I., Barr, A., Ciais, P., Davis, K., Dietze, M., Dragoni, D., Gough, C. M., Grant, R., Hollinger, D., et al.: Terrestrial biosphere model performance for inter-annual variability of land-atmosphere CO2 exchange, Global Change Biology, 18, 1971–1987, 2012. 12, 13

Keenan, T. F., Gray, J., Friedl, M. A., Toomey, M., Bohrer, G., Hollinger, D. Y., Munger, J. W., O'Keefe, J., Schmid, H. P., Wing, I. S., et al.: Net carbon uptake has increased through warming-induced changes in

temperate forest phenology, Nature Climate Change, 2014. 13

Kim, Y., Kimball, J., Zhang, K., and McDonald, K.: Satellite detection of increasing Northern Hemisphere non-frozen seasons from 1979 to 2008: Implications for regional vegetation growth, Remote Sensing of Environment, 121, 472–487, 2012. 15

Kimball, J., McDonald, K., and Zhao, M.: Spring thaw and its effect on terrestrial vegetation productivity in the western Arctic observed from satellite microwave and optical remote sensing, Earth Interactions, 10, 1–22, 2006. 15

Koven, C., Friedlingstein, P., Ciais, P., Khvorostyanov, D., Krinner, G., and Tarnocai, C.: On the formation of high-latitude soil carbon stocks: Effects of cryoturbation and insulation by organic matter in a land surface model, Geophysical Research Letters, 36, 2009. 5

Koven, C. D., Ringeval, B., Friedlingstein, P., Ciais, P., Cadule, P., Khvorostyanov, D., Krinner, G., and Tarnocai, C.: Permafrost carbon-climate feedbacks accelerate global warming, Proceedings of the National Academy of Sciences, 108, 14 769–14 774, 2011. 3, 5

Koven, C. D., Chambers, J. Q., Georgiou, K., Knox, R., Negron-Juarez, R., Riley, W. J., Arora, V. K., Brovkin, V., Friedlingstein, P., and Jones, C. D.: Controls on terrestrial carbon feedbacks by productivity vs. turnover in the CMIP5 Earth System Models, Biogeosciences Discussions, 12, 5757–5801, doi:10.5194/bgd-12-5757-2015, http://www.biogeosciences-discuss.net/12/5757/2015/, 2015. 15

Kurganova, I. and Nilsson, S.: Carbon dioxide emission from soils of Russian terrestrial ecosystems, Interim Report, IR-02, 70, 2003. 13

Lasslop, G., Reichstein, M., Papale, D., Richardson, A. D., Arneth, A., Barr, A., Stoy, P., and Wohlfahrt, G.: Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation, Global Change Biology, 16, 187–208, 2010. 11

MacDougall, A. H., Eby, M., and Weaver, A. J.: If Anthropogenic CO2 Emissions Cease, Will Atmospheric CO2 Concentration Continue to Increase?, Journal of Climate, 26, 9563–9576, 2013. 3, 5

McDonald, K. C., Kimball, J. S., Njoku, E., Zimmermann, R., and Zhao, M.: Variability in Springtime Thaw in the Terrestrial High Latitudes: Monitoring a Major Control on Biospheric Assimilation of Atmospheric CO_2 with Spaceborne Microwave Remote Sensing, Earth Interactions, 8, 1–23, 2004. 15

McGuire, A., Melillo, J., Randerson, J., Parton, W., Heimann, M., Meier, R., Clein, J., Kicklighter, D., and Sauf, W.: Modeling the effects of snowpack on heterotrophic respiration across northern temperate and high latitude regions: Comparison with measurements of atmospheric carbon dioxide in high latitudes, Biogeochemistry, 48, 91–114, doi:10.1023/A:1006286804351, http://dx.doi.org/10.1023/A:1006286804351, 2000. 14

McGuire, A., Christensen, T., Hayes, D., Heroult, A., Euskirchen, E., Yi, Y., Kimball, J., Koven, C., Lafleur, P., Miller, P., et al.: An assessment of the carbon balance of Arctic tundra: comparisons among observations, process models, and atmospheric inversions, Biogeosciences Discussions, 9, 4543–4594, 2012. 3, 5, 14

Melillo, J. M., McGuire, A. D., Kicklighter, D. W., Moore, B., Vorosmarty, C. J., and Schloss, A. L.: Global climate change and terrestrial net primary production, Nature, 363, 234–240, 1993. 3

Miller, P. A. and Smith, B.: Modelling tundra vegetation response to recent arctic warming, Ambio, 41, 281–291, 2012. 5

Mitchell, T. D. and Jones, P. D.: An improved method of constructing a database of monthly climate observa-

tions and associated high-resolution grids, International journal of climatology, 25, 693–712, 2005. 27

Murray-Tortarolo, G., Anav, A., Friedlingstein, P., Sitch, S., Piao, S., Zhu, Z., Poulter, B., Zaehle, S., Ahlström, A., Lomas, M., et al.: Evaluation of Land Surface Models in Reproducing Satellite-Derived LAI over the High-Latitude Northern Hemisphere. Part I: Uncoupled DGVMs, Remote Sensing, 5, 4819–4838, 2013. 13

Myneni, R. B., Keeling, C. D., Tucker, C. J., Asrar, G., and Nemani, R. R.: Increased plant growth in the northern high latitudes from 1981 to 1991, Nature, 386, 698–702, doi:10.1038/386698a0, 1997. 3

Nash, J. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I—A discussion of principles, Journal of hydrology, 10, 282–290, 1970. 8, 31

Oak Ridge National Laboratory: MODIS (MCD12Q1) Land Cover Product, https://lpdaac.usgs.gov, available online from Land Processes Distributed Active Archive Center (LP DAAC), Sioux Falls, South Dakota, U.S.A.. Retrieved 2014-07-01, 2014.

Oechel, W. C., Laskowski, C. A., Burba, G., Gioli, B., and Kalhori, A. A. M.: Annual patterns and budget of CO2 flux in an Arctic tussock tundra ecosystem, Journal of Geophysical Research: Biogeosciences, 119, 323–339, doi:10.1002/2013JG002431, http://dx.doi.org/10.1002/2013JG002431, 2014. 3

Oleson, K., Lawrence, D., Bonan, G., Drewniak, B., Huang, M., Koven, C., Levis, S., Li, F., Riley, W., Subin, Z., Swenson, S., Thornton, P., Bozbiyik, A., Fisher, R., Kluzek, E., Lamarque, J., Lawrence, P., Leung, L., Lipscomb, W., Muszala, S., Ricciuto, D., Sacks, W., Sun, Y., Tang, J., and Z.L., Y.: Technical Description of version 4.5 of the Community Land Model (CLM). Ncar Technical Note NCAR/TN-503+STR, Tech. rep., National Center for Atmospheric Research, boulder, CO, 422 pp, DOI: 10.5065/D6RR1W7M, 2013. 5

Parmentier, F., Van Der Molen, M., Van Huissteden, J., Karsanaev, S., Kononov, A., Suzdalov, D., Maximov, T., and Dolman, A.: Longer growing seasons do not increase net carbon uptake in the northeastern Siberian tundra, Journal of Geophysical Research: Biogeosciences (2005–2012), 116, 2011. 3

Peng, S., Ciais, P., Chevallier, F., Peylin, P., Cadule, P., Sitch, S., Piao, S., Ahlström, A., Huntingford, C., Levy, P., et al.: Benchmarking the seasonal cycle of CO2 fluxes simulated by terrestrial ecosystem models, Global Biogeochemical Cycles, 29, 46–64, doi:10.1002/2014GB004931, 2015. 12

Piao, S., Sitch, S., Ciais, P., Friedlingstein, P., Peylin, P., Wang, X., Ahlström, A., Anav, A., Canadell, J. G., Cong, N., et al.: Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO2 trends, Global change biology, 19, 2117–2132, 2013. 5

Pries, C. E. H., Schuur, E. A. G., and Crummer, K. G.: Holocene Carbon Stocks and Carbon Accumulation Rates Altered in Soils Undergoing Permafrost Thaw, Ecosystems, 15, 162–173, doi:10.1007/s10021-011-9500-4, http://link.springer.com/article/10.1007/s10021-011-9500-4, 2012. 2

Quegan, S., Beer, C., Shvidenko, A., McCallum, I., Handoh, I. C., Peylin, P., Roedenbeck, C., Lucht, W., Nilsson, S., and Schmullius, C.: Estimating the carbon balance of central Siberia using a landscape-ecosystem approach, atmospheric inversion and Dynamic Global Vegetation Models, Global change biology, 17, 351–365, 2011. 3, 12, 13

Peterson, B. J., Holmes, R. M., McClelland, J. W., Vörösmarty, C. J., Lammers, R. B., Shiklomanov, A. I., Shiklomanov, I. A., and Rahmstorf, S.: Increasing river discharge to the Arctic Ocean, Science, 298, 2171–2173, doi:10.1126/science.1077445, http://www.sciencemag.org/content/298/5601/2171.short, 2002. 16

Rawlins, M. A., et al. (2010), Analysis of the Arctic System for Freshwater Cycle Intensification: Observations and Expectations, J. Clim., 23, 5715–5737. 16

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., et al.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, Global Change Biology, 11, 1424–1439, 2005. 6, 29

Reichstein, M., Papale, D., Valentini, R., Aubinet, M., Bernhofer, C., Knohl, A., Laurila, T., Lindroth, A., Moors, E., Pilegaard, K., et al.: Determinants of terrestrial ecosystem carbon balance inferred from European eddy covariance flux sites, Geophysical Research Letters, 34, 2007. 11

Richardson, A. D., Anderson, R. S., Arain, M. A., Barr, A. G., Bohrer, G., Chen, G., Chen, J. M., Ciais, P., Davis, K. J., Desai, A. R., et al.: Terrestrial biosphere models need better representation of vegetation phenology: results from the North American Carbon Program Site Synthesis, Global Change Biology, 18, 566–584, 2012. 12, 13

Romanovsky, V. E., Smith, S. L., and Christiansen, H. H.: Permafrost thermal state in the polar Northern Hemisphere during the international polar year 2007–2009: a synthesis, Permafrost Periglacial Proc., 21, 106–116, doi:10.1002/ppp.689, http://dx.doi.org/10.1002/ppp.689, 2010. 16

Running, S., Nemani, R., Heinsch, F., Zhao, M., Reeves, M., and Hashimoto, H.: A continuous satellite-derived measure of global terrestrial primary production, Bioscience, 54, 547–560, 2004. 6

Schaefer, K., Zhang, T., Bruhwiler, L., and Barrett, A.: Amount and timing of permafrost carbon release in response to climate warming, Tellus B, 2011. 3

Schaefer, K., Schwalm, C. R., Williams, C., Arain, M. A., Barr, A., Chen, J. M., Davis, K. J., Dimitrov, D., Hilton, T. W., Hollinger, D. Y., et al.: A model-data comparison of gross primary productivity: Results from the North American Carbon Program site synthesis, Journal of Geophysical Research: Biogeosciences (2005–2012), 117, 2012. 12, 13

Schepaschenko, D., Mukhortova, L., Shvidenko, A., and Vedrova, E.: The pool of organic carbon in the soils of Russia, Eurasian Soil Science, 46, 107–116, 2013. 14

Schuur, E., Vogel, J. G., Crummer, K. G., Lee, H., Sickman, J. O., and Osterkamp, T. E.: The effect of permafrost thaw on old carbon release and net carbon exchange from tundra, Nature, 459, 2009. 3

Schuur, E. A. G. and Abbott, B.: Climate change: High risk of permafrost thaw, Nature, 480, 32–33, doi:http://dx.doi.org/10.1038/480032a, 10.1038/480032a, 2011. 17

Schuur E. A. G., McGuire A. D., Schadel C., Grosse G., Harden J. W., Hayes D. J., Hugelius G., Koven C. D., Kuhry P., Lawrence D. M., Natali S. M., Olefeldt D., Romanovsky V. E., Schaefer K., Turetsky M. R., Treat C. C., and Vonk J. E.: Climate change and the permafrost carbon feedback, Nature, 520, 171–179, doi:http://dx.doi.org/10.1038/nature14338 10.1038/nature14338, 2015. 2, 17

Serreze, M. C. and Barry, R. G.: Processes and impacts of Arctic amplification: A research synthesis, Global and Planetary Change, 77, 85–96, 2011. 2

Serreze, M. C., Barrett, A. P., Slater, A. G., Woodgate, R. A., Aagaard, K., Lammers, R. B., Steele, M., Moritz, R., Meredith, M., and Lee, C. M.: The large-scale freshwater cycle of the Arctic, J. Geophys. Res., 111, doi:10.1029/2005JC003424, http://onlinelibrary.wiley.com/doi/10.1029/2005JC003424/full, 2006. 2

Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling, J. Climate, 19, 3088–3111, 2006. 27

Smith, B., Prentice, I. C., and Sykes, M. T.: Representation of vegetation dynamics in the modelling of terrestrial ecosystems: comparing two contrasting approaches within European climate space, Global Ecology

and Biogeography, 10, 621–637, 2001. 5

Smith, N. V., Saatchi, S. S., and Randerson, J. T.: Trends in high northern latitude soil freeze and thaw cycles from 1988 to 2002, Journal of Geophysical Research: Atmospheres (1984–2012), 109, 2004. 15

Thornton, P. E., Doney, S. C., Lindsay, K., Moore, J. K., Mahowald, N., Randerson, J. T., Fung, I., Lamarque, J.-F., Feddema, J. J., and Lee, Y.-H.: Carbon-nitrogen interactions regulate climate-carbon cycle feedbacks: results from an atmosphere-ocean general circulation model, Biogeosciences, 6, 2099–2120, 2009. 15

Turner, D. P., Ritts, W. D., Cohen, W. B., Gower, S. T., Running, S. W., Zhao, M., Costa, M. H., Kirschbaum, A. A., Ham, J. M., Saleska, S. R., et al.: Evaluation of MODIS NPP and GPP products across multiple biomes, Remote Sensing of Environment, 102, 282–292, 2006. 6

Viovy, N. and Ciais, P.: CRUNCEP data set for 1901–2008, Tech. Rep. Version 4, Tech. rep., Labora- toire des Sciences du Climat et de l'Environnement, see http://dods.extra.cea.fr/data/p529viov/cruncep/readme.htm, 2011. 27

Vogel, J., Schuur, E. A. G., Trucco, C., and Lee, H.: Response of CO2 exchange in a tussock tundra ecosystem to permafrost thaw and thermokarst development, Journal of Geophysical Research: Biogeosciences, 114, G04 018, doi:10.1029/2008JG000901, http://onlinelibrary.wiley.com/doi/10.1029/2008JG000901/abstract, 2009. 3

Wania, R., Ross, I., and Prentice, I.: Integrating peatlands and permafrost into a dynamic global vegetation model: 2. Evaluation and sensitivity of vegetation and carbon cycle processes, Global Biogeochemical Cycles, 23, 2009a. 5

Wania, R., Ross, I., and Prentice, I.: Integrating peatlands and permafrost into a dynamic global vegetation model: 1. Evaluation and sensitivity of physical land surface processes, Global Biogeochemical Cycles, 23, 2009b. 5

Wania, R., Ross, I., and Prentice, I.: Implementation and evaluation of a new methane model within a dynamic global vegetation model: LPJ-WHyMe v1. 3.1, Geoscientific Model Development, 3, 565–584, 2010. 5

Waring, R., Landsberg, J., and Williams, M.: Net primary production of forests: a constant fraction of gross primary production?, Tree Physiology, 18, 129–134, 1998. 9

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takcmura, T., Okajima, H., Nozawa, T., Kawase, H., Abe, M., Yokohata, T., et al.: MIROC-ESM 2010: model description and basic results of CMIP 5-20 c 3 m experiments, Geoscientific Model Development, 4, 845–872, 2011. 5, 27

Weedon, G., Gomes, S., Viterbo, P., Shuttleworth, W., Blyth, E., Österle, H., Adam, J., Bellouin, N., Boucher, O., and Best, M.: Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century, Journal of Hydrometeorology, 12, 823–848, 2011. 27

Willmott, C. J. and Matsuura, K.: Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, Climate research, 30, 79, 2005. 7

Willmott, C. J., Robeson, S. M., and Matsuura, K.: A refined index of model performance, International Journal of Climatology, 32, 2088–2094, 2011. 7, 8, 31

Yi, Y., Kimball, J. S., Jones, L. A., Reichle, R. H., Nemani, R., and Margolis, H. A.: Recent climate and fire disturbance impacts on boreal and arctic ecosystem productivity estimated using a satellite-based terrestrial carbon flux model, Journal of Geophysical Research: Biogeosciences, pp. 1–17, 2013. 12

Yi, Y., Kimball, J. S., and Reichle, R. H.: Spring hydrology determines summer net carbon uptake in northern ecosystems, Environmental Research Letters, 9, 064 003, 2014. 15

Zaehle, S.: Terrestrial nitrogen–carbon cycle interactions at the global scale, Philosophical Transactions of the Royal Society B: Biological Sciences, 368, 20130 125, doi:http://dx.doi.org/10.1098/rstb.2013.0125, 2013. 12

Zhang, K., Kimball, J. S., Hogg, E. H., Zhao, M., Oechel, W. C., Cassano, J. J., and Running, S. W.: Satellite-based model detection of recent climate-driven changes in northern high-latitude vegetation productivity, J. Geophys. Res., 113, g03033, doi:10.1029/2007JG000621, 2008. 3, 6

Zhang, X., He, J., Hinzman, L. D., Polaykov, I., Walsh, J. E., Inoue, J., and Zhao, P.: Enhanced poleward moisture transport: Predominant driver of northern high latitude wetting, in preparation, 2011. 15

Zhao, M. and Running, S. W.: Drought-induced reduction in global terrestrial net primary production from 2000 through 2009, science, 329, 940–943, 2010. 6

Zhao, M., Heinsch, F. A., Nemani, R. R., and Running, S. W.: Improvements of the MODIS terrestrial gross and net primary production global data set, Remote sensing of Environment, 95, 164–176, 2005. 6

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).

Model	Institution	Climate Data Set
Community Land Model (CLM4.5)	National Center for Atmospheric Research, USA	CRUNCEP4[1]
Common Land Model (CoLM)	Beijing Normal University, China	Princeton[2]
Interaction Sol-Biosphère-Atmosphere (ISBA)	National Centre for Meteorological Research, France	WATCH[3] WFDEI[6,*]
Joint UK Land Environment Simulator (JULES)	Met Office, United Kingdom	WATCH[3]
Lund-Potsdam-Jenna General Ecosystem Simulator (LPJG)	Lund University, Sweden	CRU TS 3.1[4]
Model for Interdisciplinary Research on Climate, Earth System Model (MIROC)	Japan Agency for Marine-Earth Science and Technology, Japan	CMIP5[5]
Organising Carbon and Hydrology In Dynamic Ecosystems (ORCHIDEE)	Institute Pierre Simon Laplace (IPSL), France	WATCH[3] WFDEI[6,*]
University of Victoria (UVic)	University of Victoria, Canada	CRUNCEP4[1]
Variable Infiltration Capacity (UW-VIC)	University of Washington, USA	CRU[7], UDel[8], NCEP-NCAR[9]

[1] Viovy and Ciais (2011) (http://dods.extra.cea.fr/data/p529viov/cruncep/readme.htm)

[2] Sheffield et al. (2006) (http://hydrology.princeton.edu/data.pgf.php)

[3] Weedon et al. (2011) (http://www.waterandclimatechange.eu/about/watch-forcing-data-20th-century)

[4] Harris et al. (2014)

[5] Watanabe et al. (2011)

[6] http://www.eu-watch.org/gfx_content/documents/README-WFDEI.pdf

[7] Mitchell and Jones (2005) for temperature

[8] Willmott and Matsura (2001) for precipitation; Adam and Lettenmaier (2003) and Adam et al. (2006) for precipitation adjustments

[9] Kalnay et al. (2006) for wind speed

[*] WATCH used for 1901–1978; WFDEI used for 1978–2009

Table 2. Properties in each model relevant to simulation of land-atmosphere CO_2 dynamics, particularly for the northern high latitude terrestrial biosphere. Properties are indicated as present (✓), absent (✗) or otherwise (see footnote for details).

	CLM4.5	CoLM	ISBA	JULES	LPJG	MIROC	ORCHIDEE	UVic	UW-VIC
Tree mortality/senescence included?	✓/✓	✓/✓	✗/✗	✓/✗	✓/✓	✓/✓	✓/✓	✗/✗	✗/✗
Light limits phtosynthesis?	✓	✓	✓	✓	✓	✓	✓	✓	✓
N limits photosynthesis?	✗	✗	✗	✗	✓	✗	✗	✗	✓
Vegetation competes for light/water/nitrogen?	✗/✓/✗	✓/✓/✗	✗/✗/✗	✓/✗/✗	✓/✓/✗	✓/✓/✗	✓/✓/✗	✓/✓/✗	✗/✗/✗
No. of PFTs	16	14	9	5	15	13	12	5	20
CO_2 fertilization?	✗	✗	✓	✓	✓	✗	✓	✗	✗
Turnover time of carbon in heartwood (yr)	50	process dependent	30–50	PFT dependent	PFT dependent	20	20–80	PFT dependent	33.3
Turnover time of carbon in sapwood (yr)	50	29	30–50	PFT dependent	PFT dependent	20	1	PFT dependent	33.3
Turnover time of carbon in leaves (yr)	1	0.5–2	0.4–1	PFT dependent	PFT dependent	0.15–4.5	80days	PFT dependent	2.86
Turnover time of carbon in coarse/fine roots	50yrs	1–2yrs	150–365days	PFT dependent	PFT dependent	20/1.1–6.25yrs	80days	PFT dependent	33.3
Time step of carbon cycle	0.5hr	1hr	30min–1day	0.5hr	1month	1day	0.5hr–1day	1hr	3hr
Disturbance (F/L/I)[c]?	F+L	F	✗	✗	F	F+L	✗	L	✗
Vegetation dynamic?	✓	✓	✗	✓	✓	✓	✗	✓	✗
Vegetation dynamics time step	NA	1yr	NA	10days	1month	1yr	1yr	5days	NA
LAI[d] dynamic?	✓	✓	✓	✓	✓	✗	✓	✓	✗
LAI_{max} prescribed?	✗	✗	✗	✓	✗	✓	✓	✗	✗
LAI time step	0.5 hr	1 day	1 day	1 day	1 month	1 day	1 day	5 days	30 days
Max veg height prescribed?	✗	✓	✓	✓	✗*	✓	✗	✗	✓
Max rooting depth	variable	3.4m	2m	3m	2m	1m	variable	3.35m	1m
C_{soil}[b] layered? (Depth)	✓(4m)	✗(3.4m)	✗(1m)	implicit	implicit	implicit	✓(2-47m)	✓(3.35m)	✗
Soil layers for hydrology	10	10	14	30	2	6	11	8	25
Biogenic CH_4 fluxes	✓	✗	✗	✗	✓	✗	✗	✗	✓
Depth of water extraction (m)	PFT dependent	3.4	PFT dependent	PFT dependent	2	2	Soil depth limited	3.35	1
Approach to soil thermal dynamics	heat diffusion	heat diffusion	multi-layer (Fourier law)	multi-layer finite difference model	multi-layer finite difference model	heat conduction	1D Fourier	Avis (2011)	Finite difference
Effect of vegetation on soil thermal dynamics?	✓	✗	✓(only at surface)	✓	✓	✗	✗	✓(water+albedo)	✗
Snow insulation type	multi-layer	multi-layer	multi-layer	multi-layer	implicit	multi-layer	implicit	-	bulk
Capable of talik formation and dynamics?	✗	✗	✗	✓	✗	✓	✗	✓	✓

[a] Heterotrophic respiration

[b] Soil carbon

[c] **F**ire; **L**and-use change; **I**nsects

[d] **L**eaf **A**rea **I**ndex

[e] **M**oisture; **T**emperature; **C**arbon/**N**itrogen ratio; **O**xygen

[*] max height prescribed for shrubs

Table 3. Flux tower sites from the LaThuile dataset (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 Figures 2c, 3c, 4c. GPP and ER in the La Thuile dataset are calculated using methodologies described in Reichstein et al. (2005).

site	coordinates	IGBP class	start/end years
Chersky (CHE)	68.61N, 161.34E	mixed forest	2002–
Chokurdakh (COK)	70.62N, 147.88E	open shrubland	2003–2005
Hakasija* (HAK)	54.77N, 89.95E	grassland	2002–2004
Zotino (ZOT)	60.80N, 89.35E	evergreen needleleaf forest	2002–

* Data used from three research sites (HAK1, HA2, HAK3)

Table 4. Average model error in g C m^{-2} $month^{-1}$ for site-level comparisons over the years 2002–2005 shown in Figures 2–4. Errors are calculated as the average ($\hat{\epsilon_j}$) 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 La Thuile FLUXNET observations (Baldocchi, 2008). The last column lists mean NEP error ($\overline{\text{NEP}}$) 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 t-tests for (i) each area separately; (ii) GPP and ER pooled for the 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.

	CHE			COK			HAK			ZOT			
Model	GPP	ER	NEP	GPP	ER	NEP	GPP	ER	NEP	GPP	ER	NEP	$\overline{\text{NEP}}$
MOD17	−2	-	-	−11	-	-	13	-	-	10	-	-	
CLM4.5	−25	−19	−6	−42	−23	−19	8	22	−15	78	81	−3	−11
ISBA	27	25	2	34	41	−7	82	78	3	82	98	−16	−5
LPJG	−10	−5	−5	−5	−1	−4	53	74	−22	−34	−13	−20	−13
MIROC	20	18	2	49	43	6	28	37	−10	−4	21	−25	−7
ORCHIDEE	23	12	11	49	32	17	16	21	−6	−30	−6	−24	−1
UVic	−14	−7	−7	16	36	−20	30	38	−9	−7	31	−38	−19
UW-VIC	27	34	−6	140	119	19	18	33	−16	2	20	−18	−5
mean	7	8	−1	34	35	−1	34	43	−11	13	33	−20	−8

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.

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

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 (*).

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

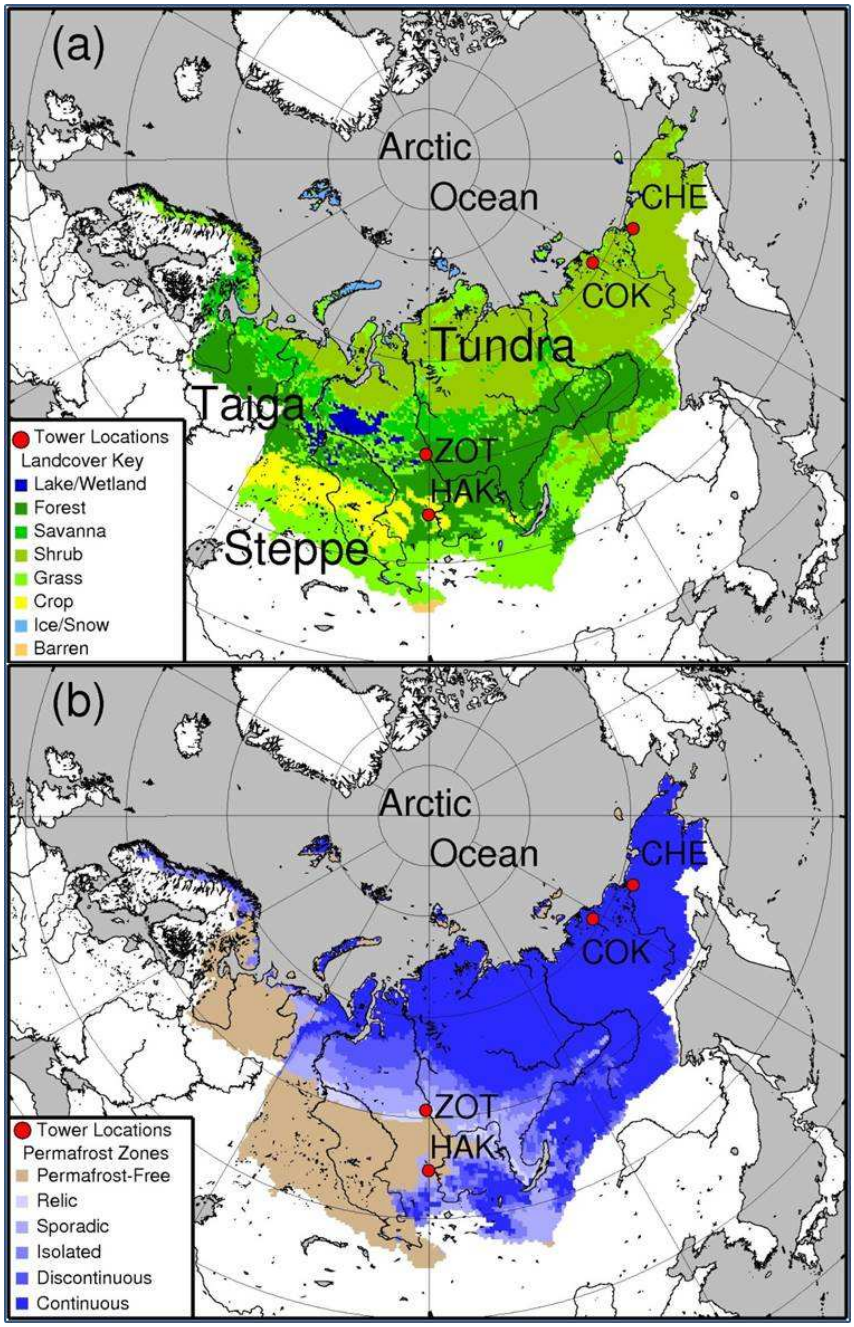

Fig. 1. Study domain spanning the arctic drainage basin in northern Eurasia. Map panels show a) plant functional types (PFTs) and b) permafrost classification along with tower sites used in the study: a) Chersky, b) Chokurdakh, c) Hakasija, and d) Zotino locations (Table 3). Gridded PFTs are from the MODIS MOD12 product (Oak Ridge National Laboratory, 2014). Permafrost classes for each grid are drawn from the CAPS dataset (International Permafrost Association Standing Committee on Data Information and Communication, 2003) .

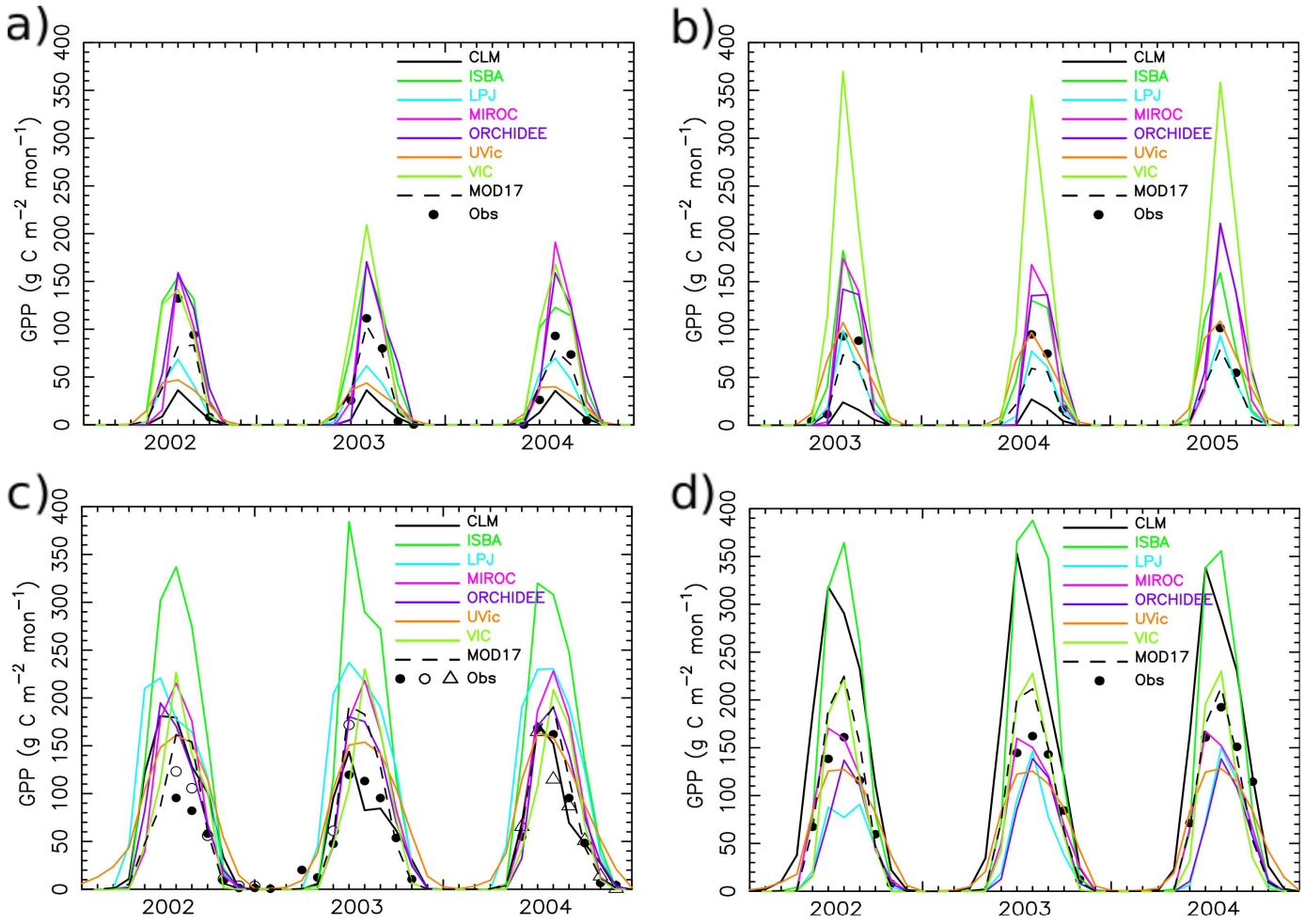

Fig. 2. Monthly GPP at sites a) Chersky, b) Chokurdakh, c) Hakasija, and d) Zotino (Obs, Table 3). Colored lines trace monthly GPP for each model grid that encompassing the tower location. Site Hakasija includes research areas Ha1 (filled circle), Ha2 (open circle), and Ha3 (triangle)

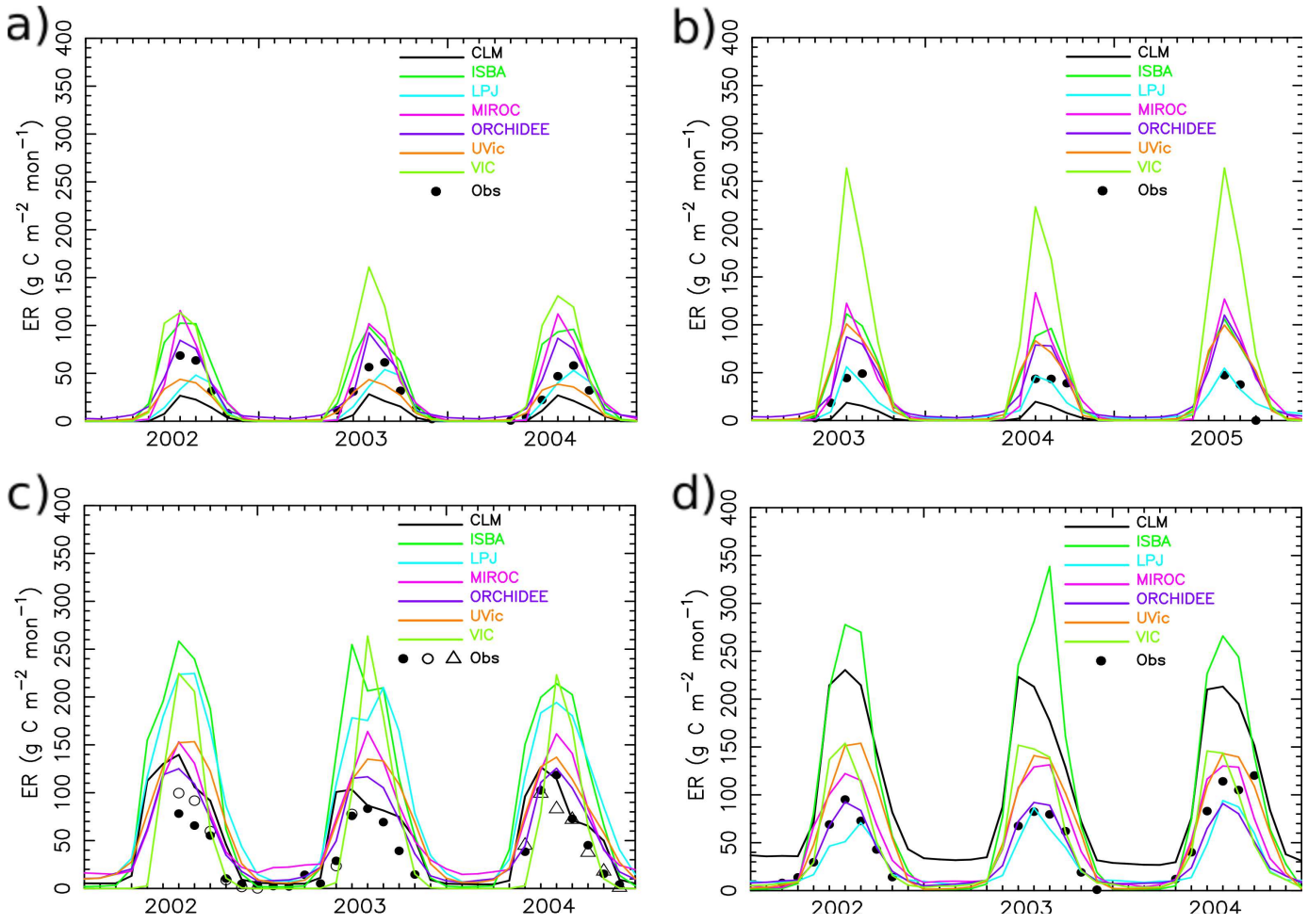

Fig. 3. As in figure 2, for ER.

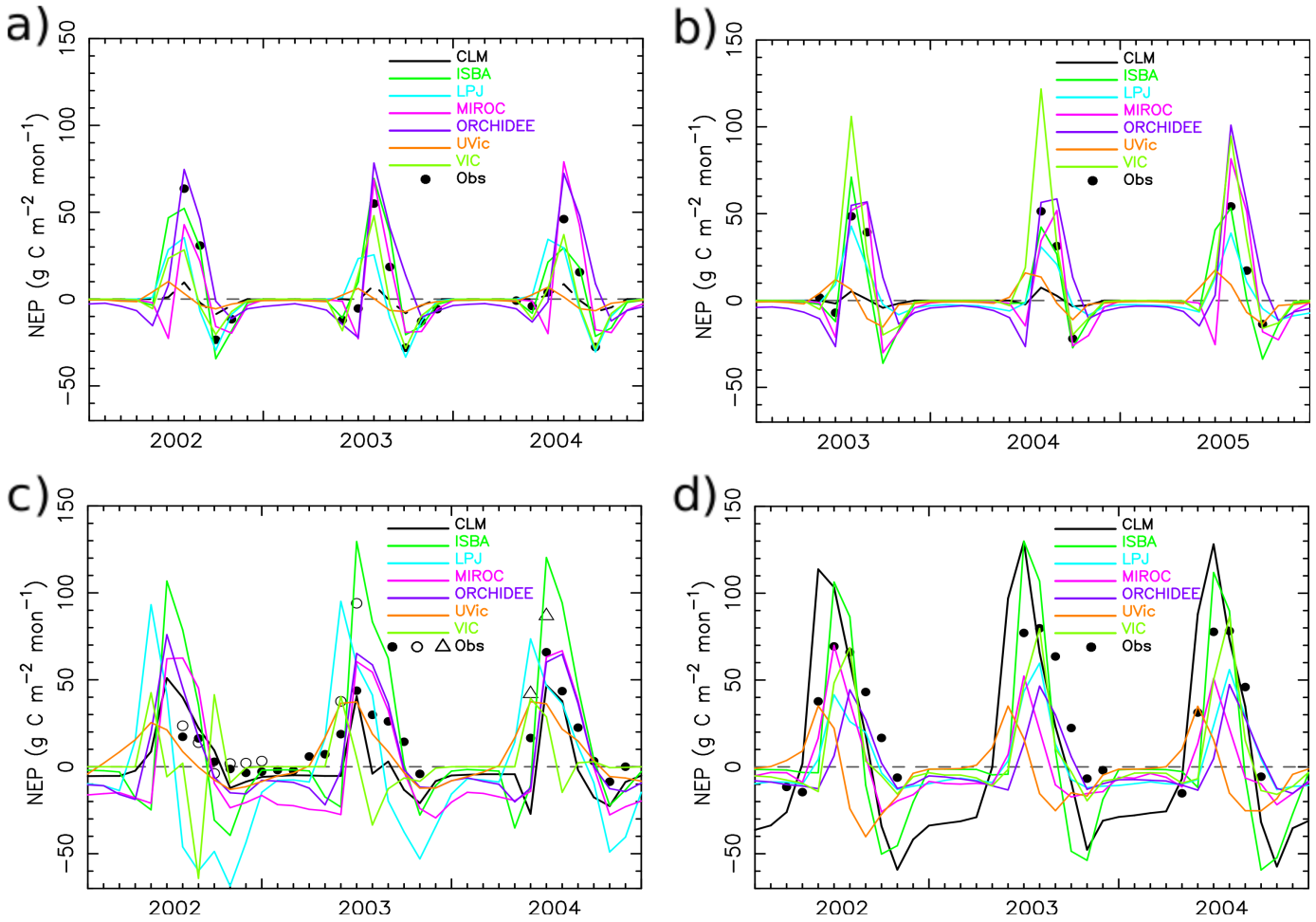

Fig. 4. As in figure 2, for NEP. NEP = GPP−ER.

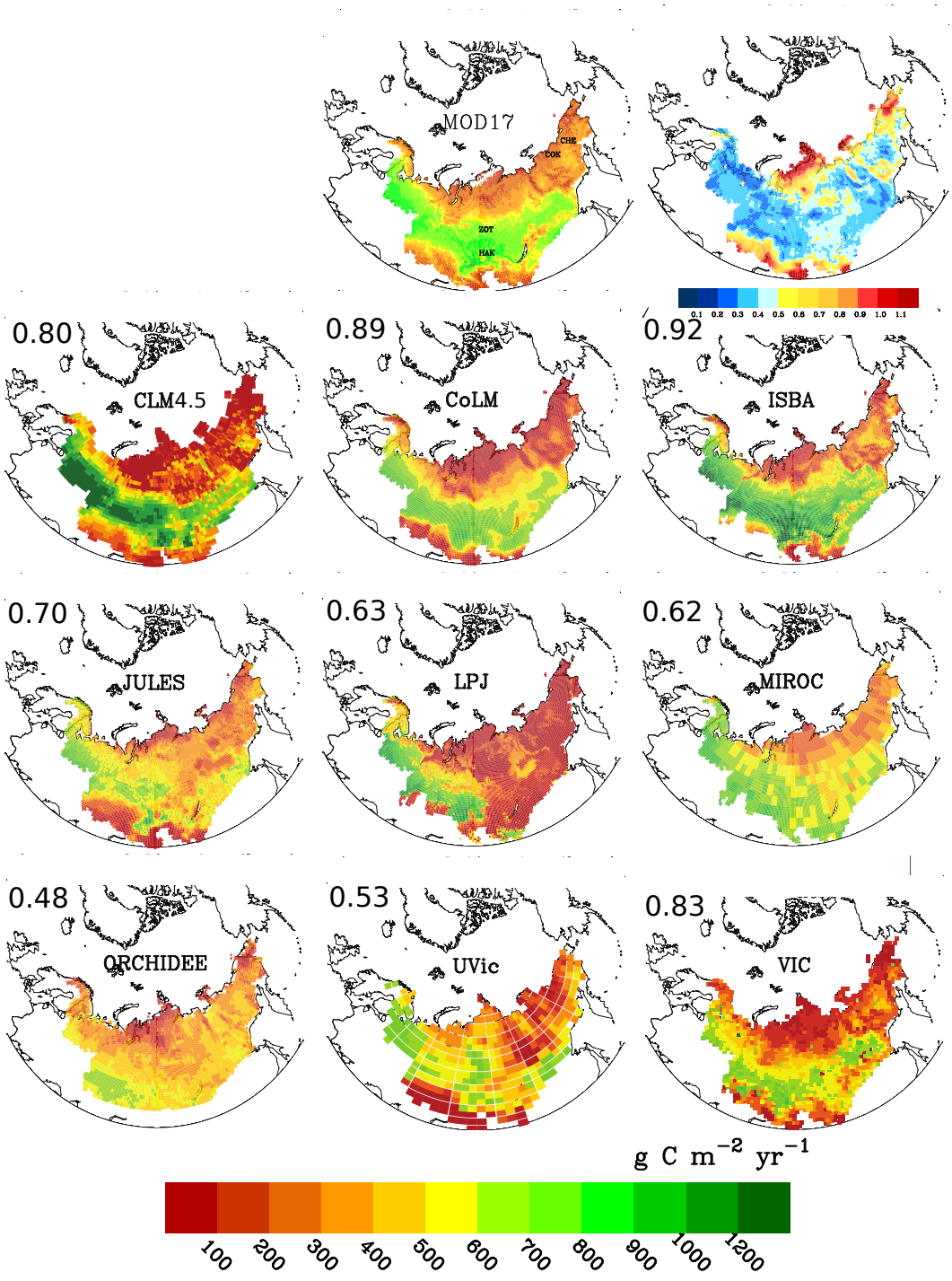

Fig. 5. Mean annual Gross Primary Productivity (GPP) from the permafrost RCN models and from the MOD17 product. The averaging period is 2000–2009 for GPP from the MOD17 product and all models with the exception of CLM4.5 (1995–2004); CoLM (1991–2000); and JULES (1991–2000). Spatial correlations between 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).

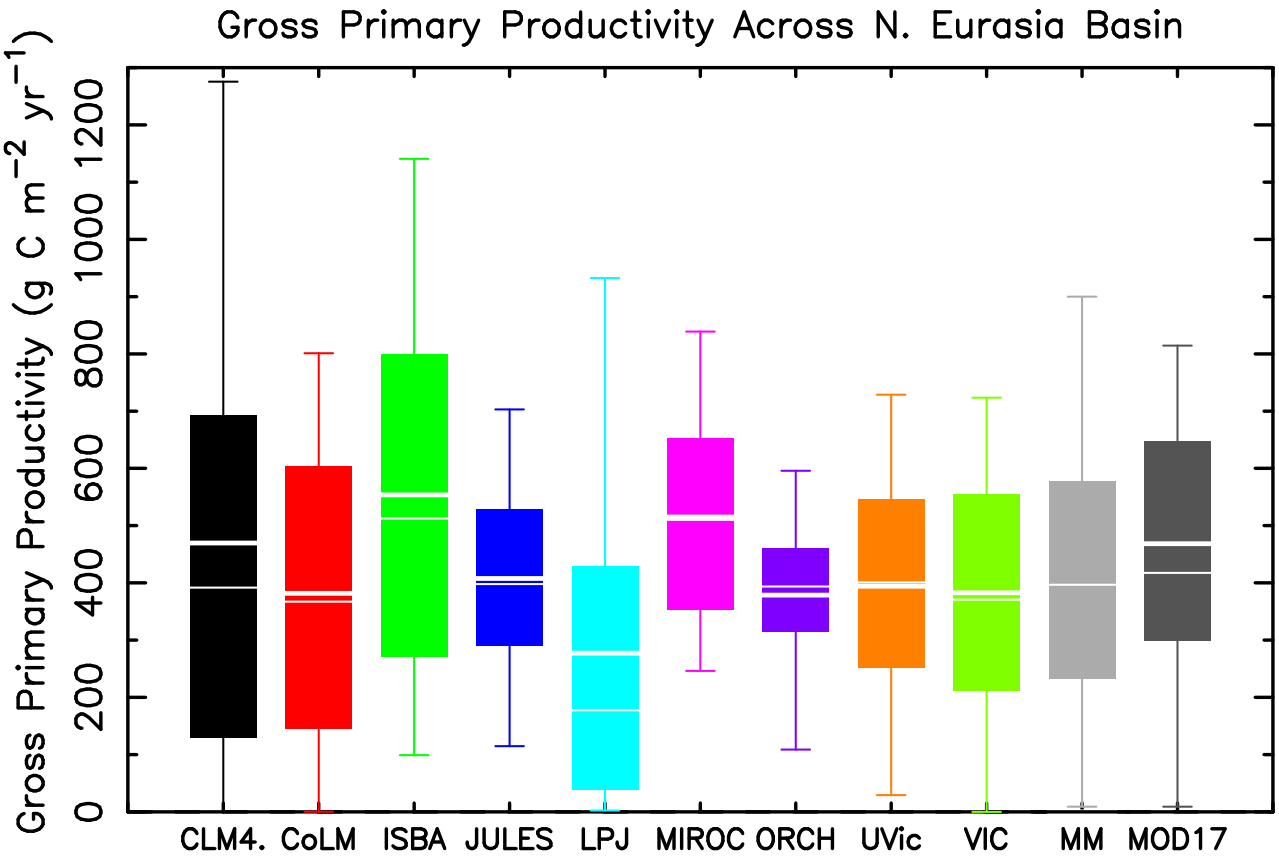

Fig. 6. Distributions for mean annual GPP from the models and the MOD17 product over the averaging period listed in Figure 5. The rectangles bracket the 25th and 75th percentiles. Whiskers extend to the 5th and 95th percentiles. Thick and thin horizontal lines mark the mean and median respectively.

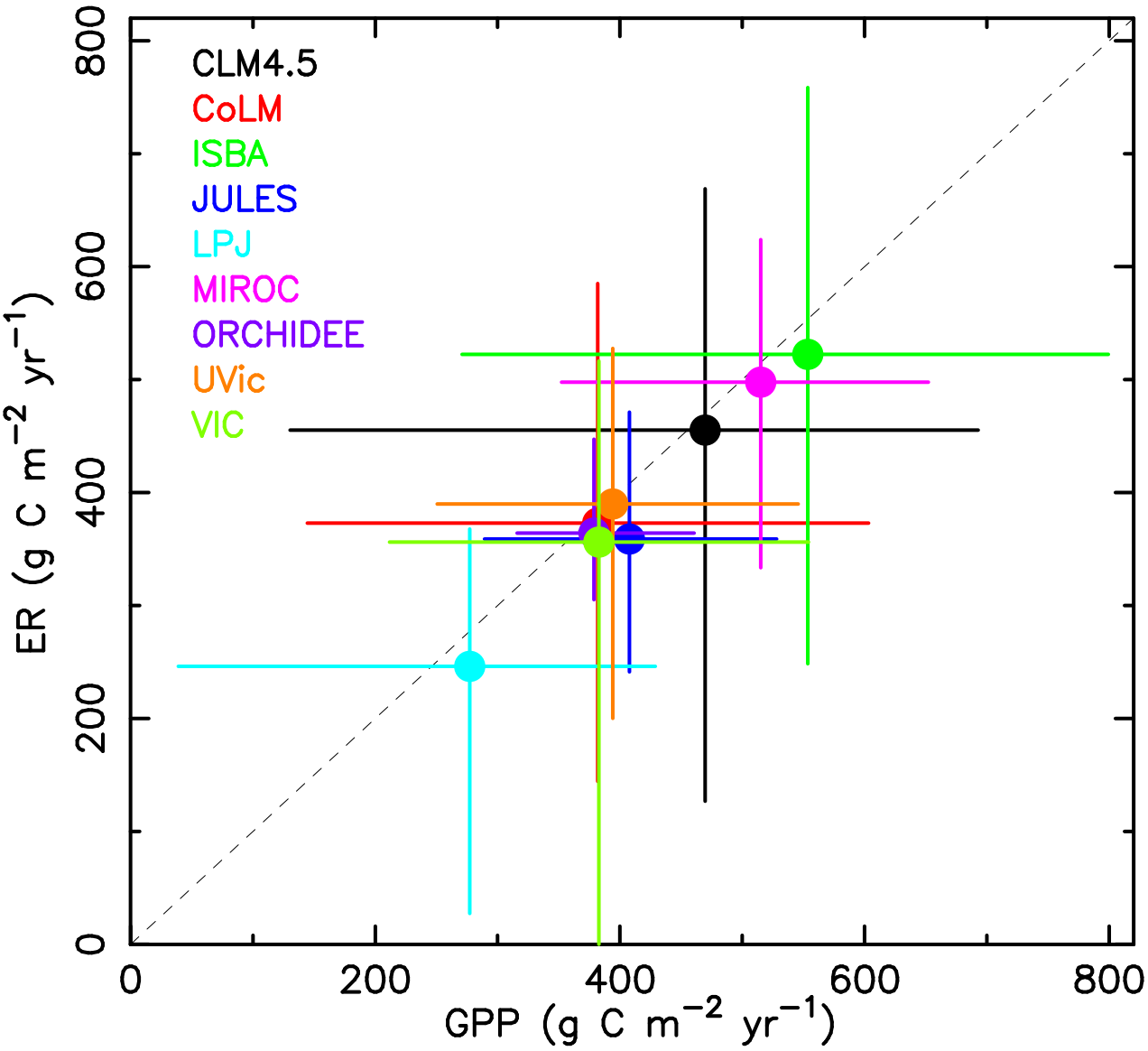

Fig. 7. Spatially averaged ER vs. GPP over the period 1960–2009. Horizontal and vertical lines span the range across the 5th and 75th percentiles for GPP and ER, respectively. The GPP 5th and 75th percentiles are shown in Figure 6. NEP is equal to the difference GPP minus ER.

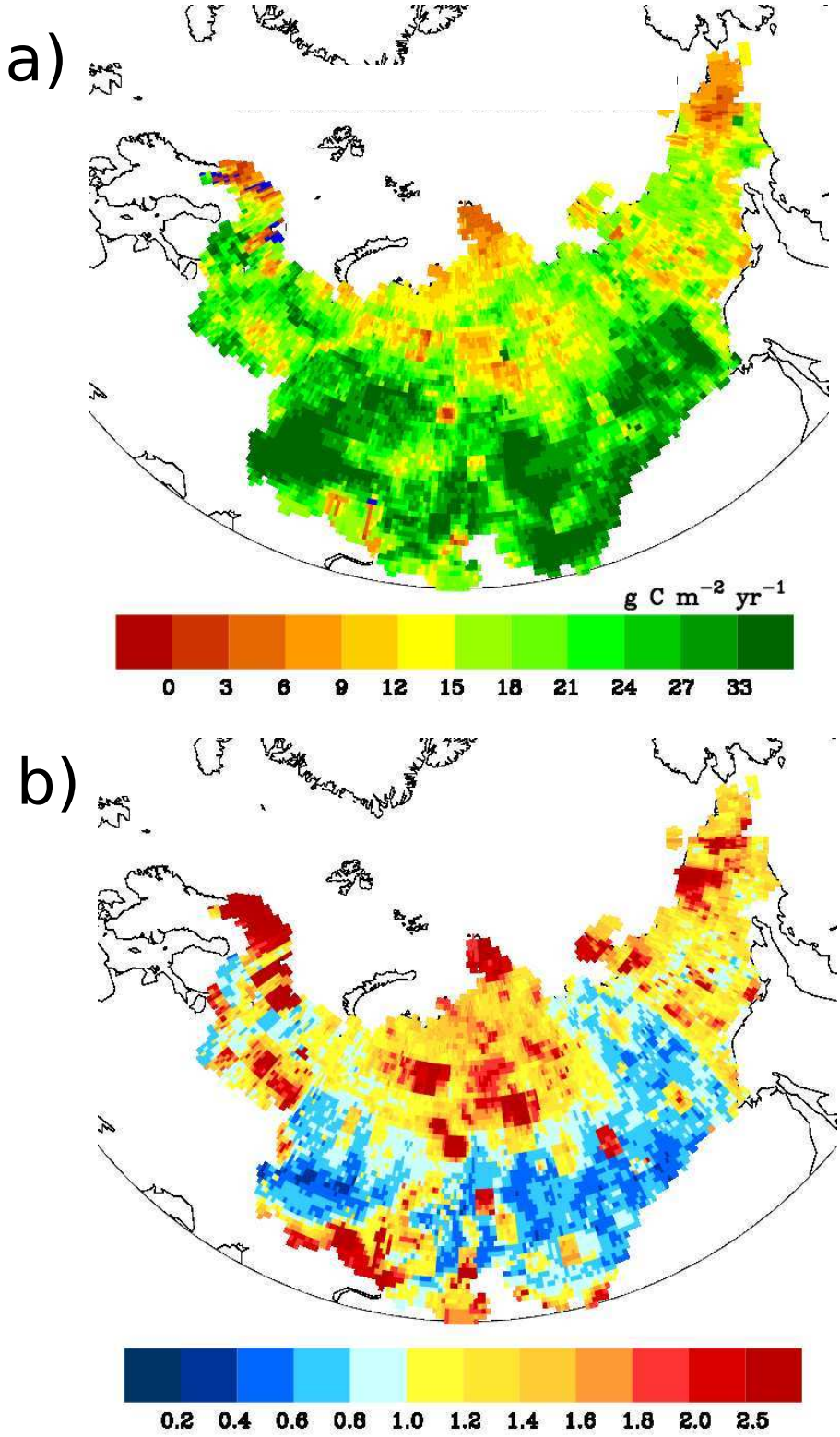

Fig. 8. a) Annual NEP (1960–2009) averaged across the nine models. Areas in blue are a net annual source of CO_2. b) Coefficient of variation as estimated from the across model mean and standard deviation for each grid.

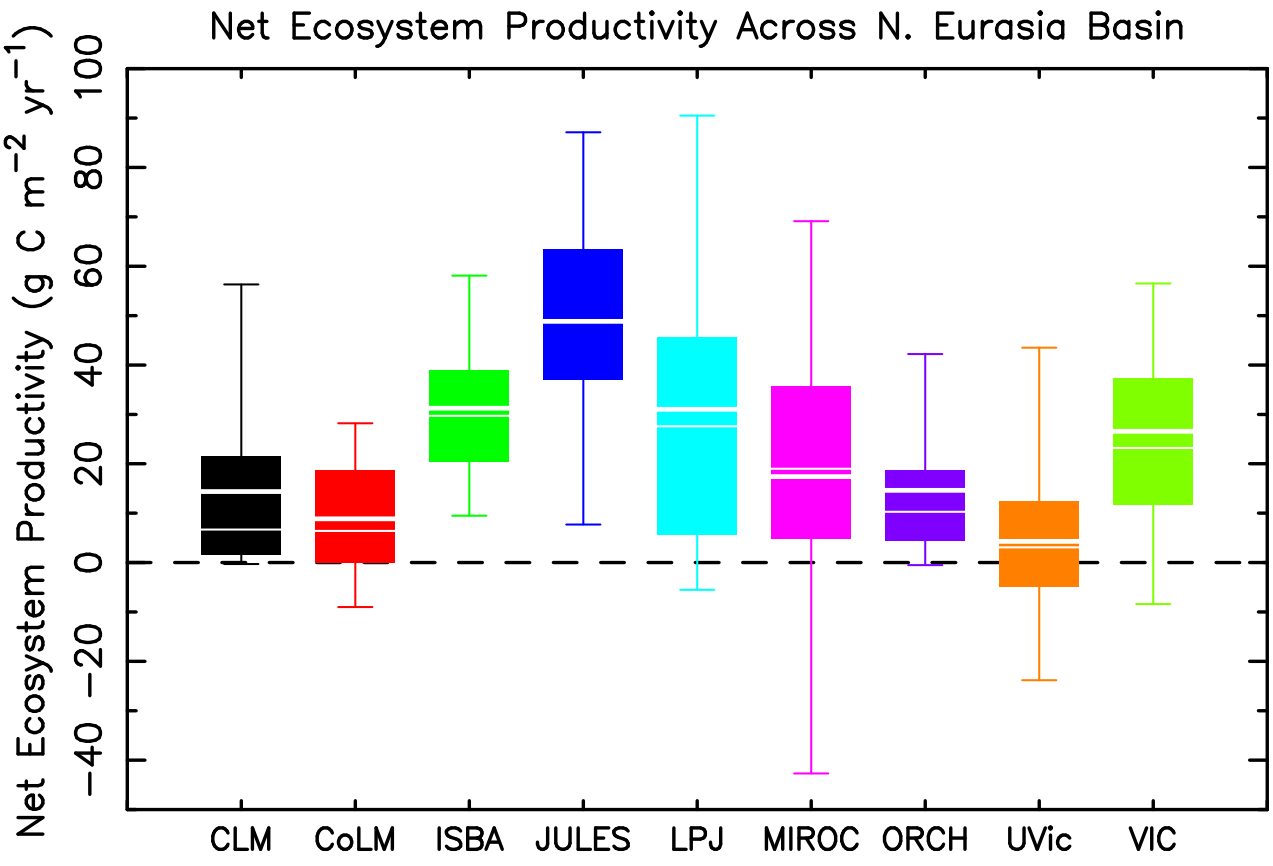

Fig. 9. Distributions for mean annual NEP from the models over the averaging period listed in Figure 5. Boxplot quartiles are as described in caption for Figure 6.

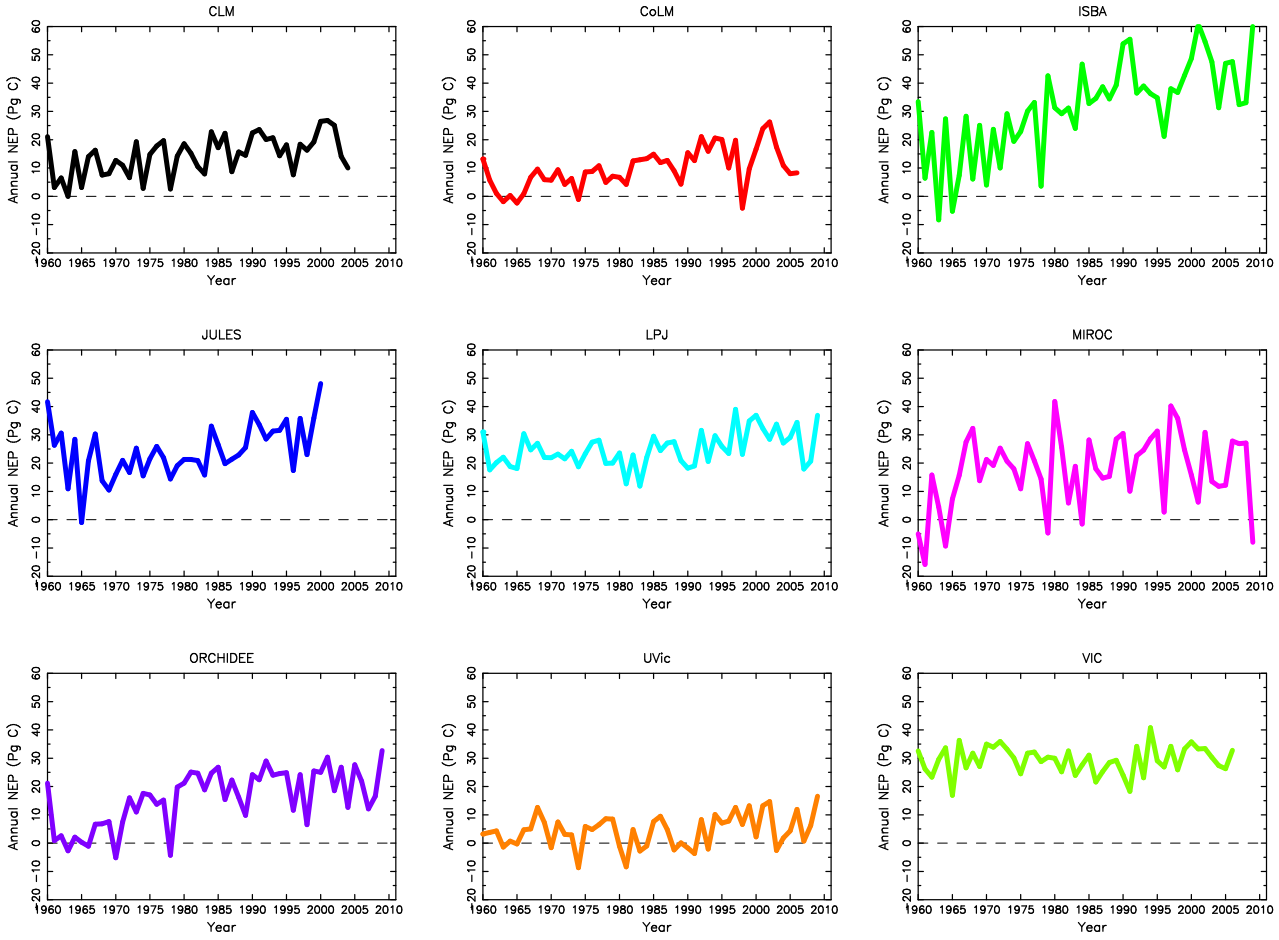

Fig. 10. Annual NEP as a spatial average across the region for each year 1960–2009.

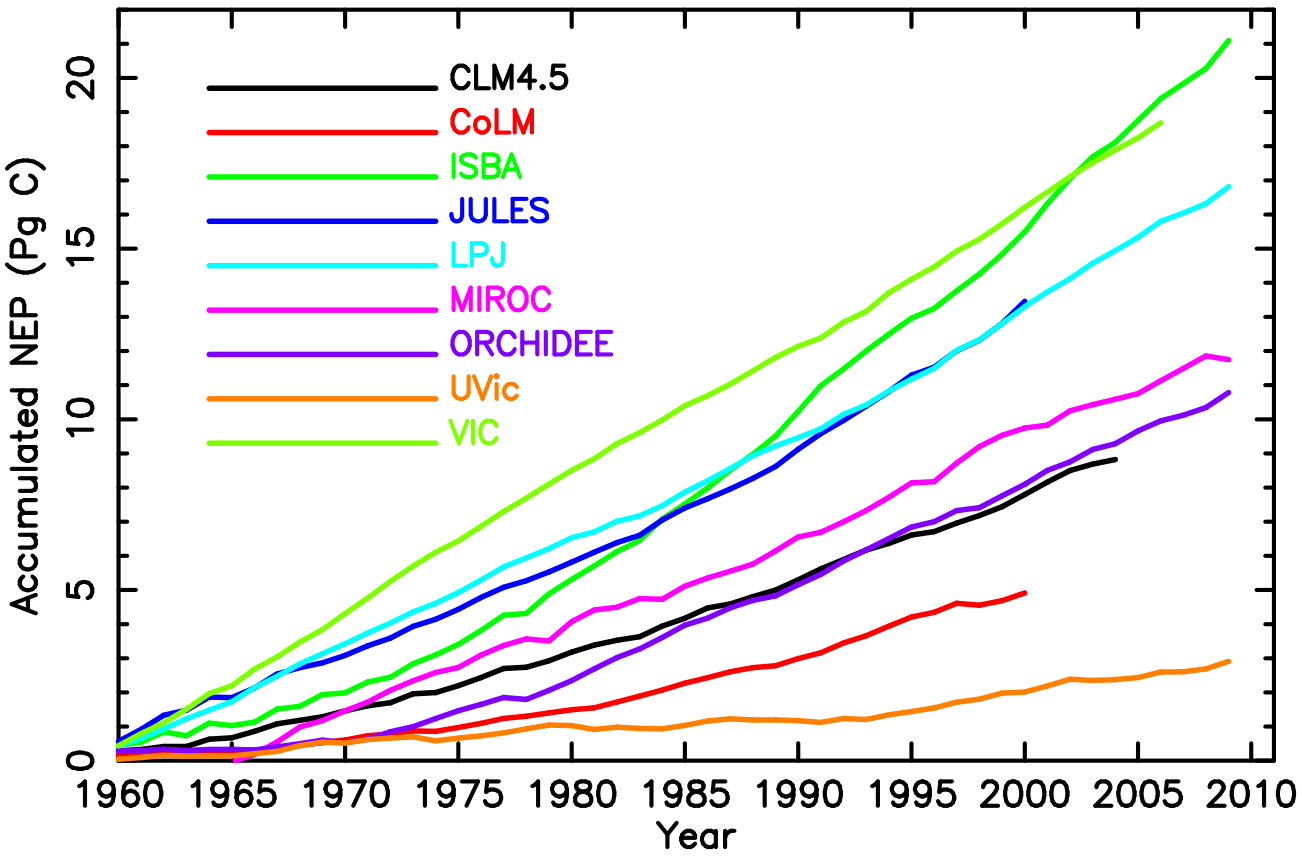

Fig. 11. Cumulative NEP in Pg C over the simulation period for each model.

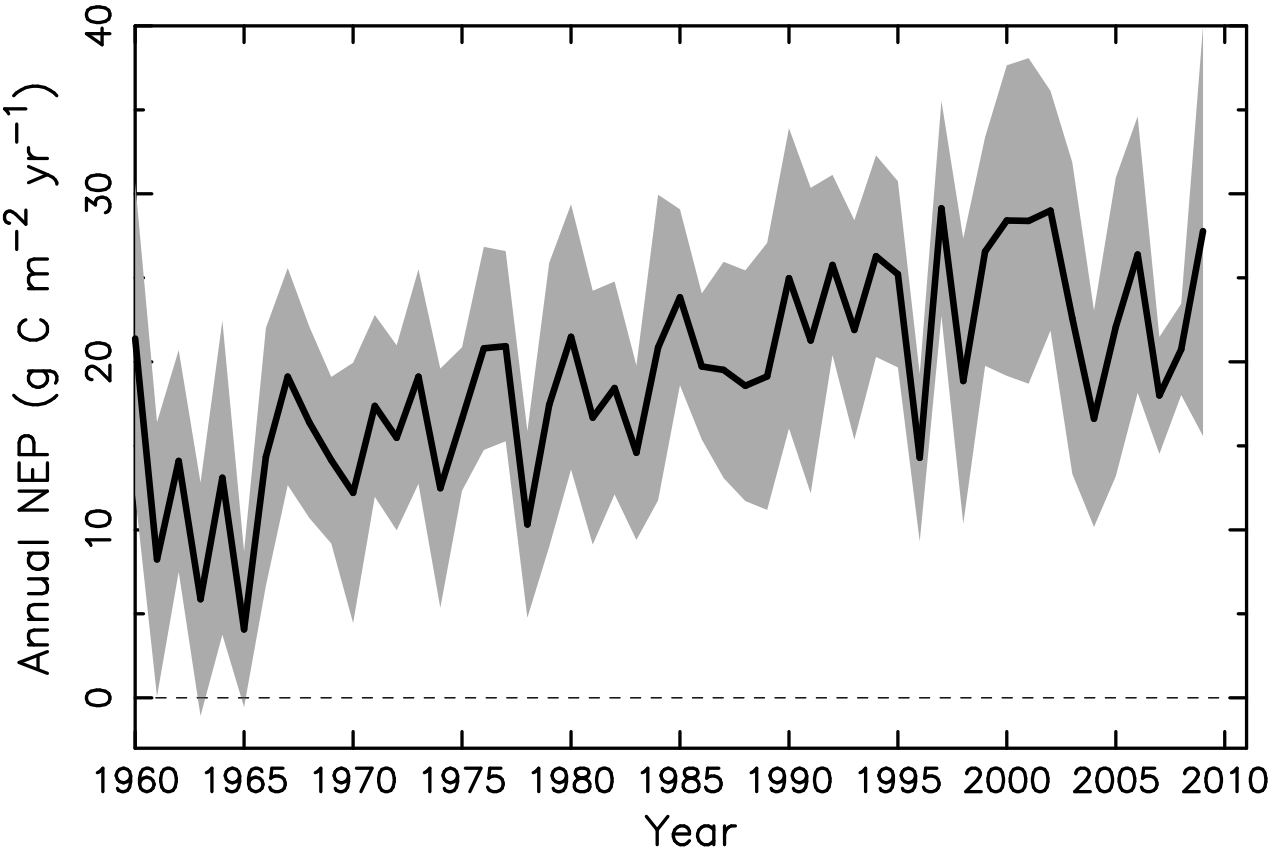

Fig. 12. Spatially averaged annual NEP as an average across the nine models. Gray region marks the 95th confidence interval, where CI = $\mu \pm$ (SE $\times$ 1.96), where μ is the nine model average and SE is the standard error. Standard deviation (σ) used to estimate SE is obtained each year from the set of nine model NEP values used to obtain the yearly average.

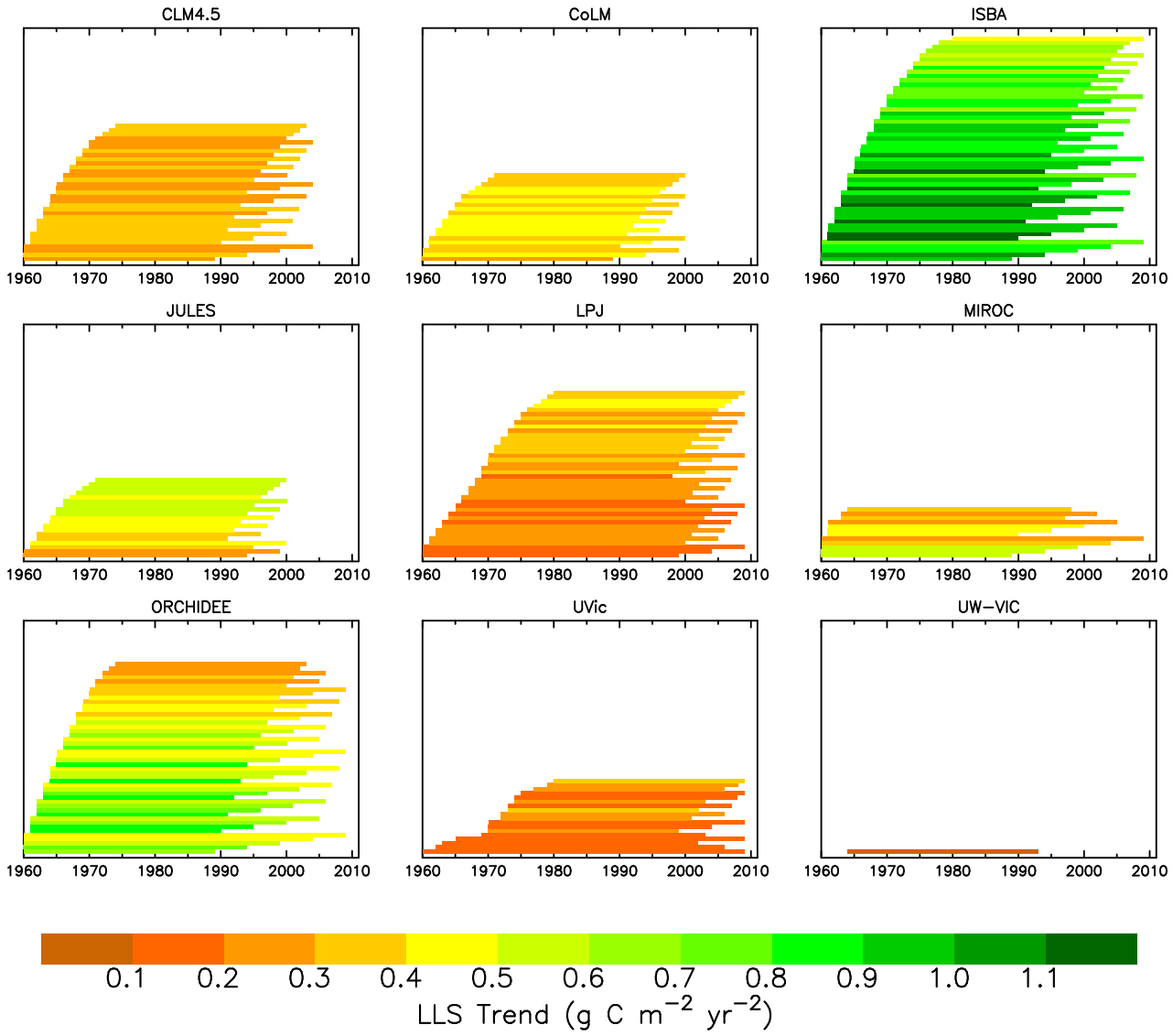

Fig. 13. Magnitude of linear trend in NEP over given time interval for all trends significant at $p < 0.05$. For each model, linear trends are calculated for all time intervals of 20 years or more. For example, 1960–1979, 1960–1980, ..., 1990–2009. Intervals for which the trend is significant are marked 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 indentified with a significant NEP trend for UW-VIC, from 1964–1993.

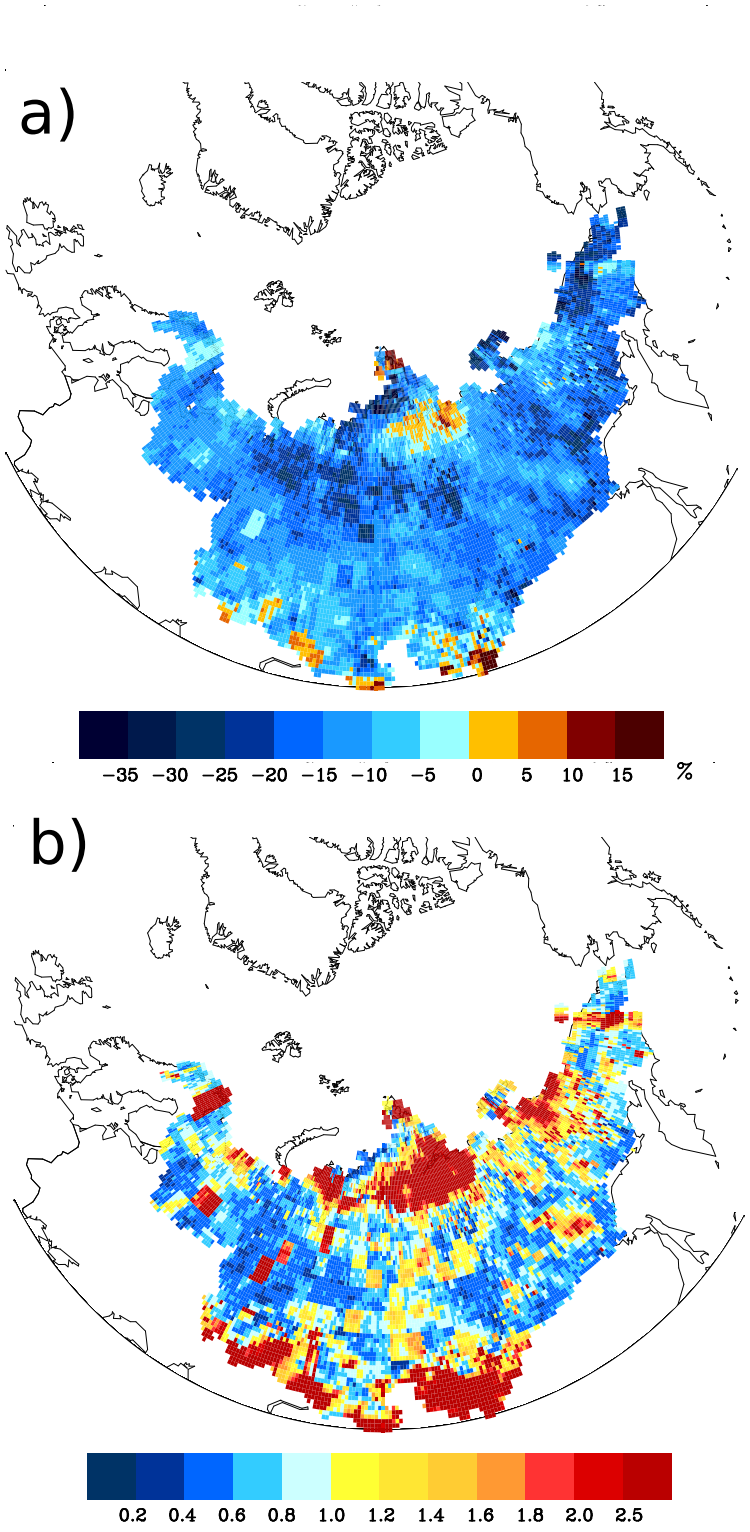

Fig. 14. a) Change in soil organic carbon (SOC) residence time (RT) averaged across all nine models. Change is significant for 46% of the region, predominantly negative changes (decreases). b) CV for RT as estimated from the across-model mean and standard deviation at each grid.