Statistical upscaling of ecosystem CO_2 fluxes across the terrestrial tundra and boreal domain: regional patterns and uncertainties

Authors: Anna-Maria Virkkala[1,2], Juha Aalto[1,3], Brendan M. Rogers[2], Torbern Tagesson[4,5], Claire C. Treat[6], Susan M. Natali[2], Jennifer D. Watts[2], Stefano Potter[2], Aleksi Lehtonen[7], Marguerite Mauritz[8], Edward A. G. Schuur[9], John Kochendorfer[10], Donatella Zona[11,12], Walter Oechel[11,13], Hideki Kobayashi[14], Masahito Ueyama[14], Elyn Humphreys[15], Mathias Goeckede[16], Hiroki Iwata[17], Peter Lafleur[18], Eugenie S. Euskirchen[19], Stef Bokhorst[20], Maija Marushchak[21,22], Bo Elberling[23], Carolina Voigt[22,24], Oliver Sonnentag[24], Frans-Jan W. Parmentier[25,26], Gerardo Celis[27], Vincent L. St.Loius[28], Craig A. Emmerton[28], Matthias Peichl[29], Jinshu Chi[29], Järvi Järveoja[29], Steven F. Oberbauer[30], Margaret S. Torn[31], Sang-Jong Park[32], Han Dolman[33], Ivan Mammarella[34], Namyi Chae[35], Rafael Poyatos[36], Mats. B. Nilsson[29], Christina Biasi[22,24], Pertti Martikainen[22], Efrén Lopéz-Blanco[37,38], Torben Røjle Christensen[38], Min Jung Kwon[39], M. Luoto[1]

[1]Department of Geosciences and Geography, University of Helsinki, Finland; [2]Woodwell Climate Research Center, USA; [3]Finnish Meteorological Institute, Finland; [4]Lund University, Sweden; [5]Copenhagen University, Denmark; [6]Alfred-Wegener-Institut, Germany; [7]Natural Resources Institute Finland, Finland; [8] University of Texas, at El Paso, USA; [9]Northern Arizona University, USA; [10]Atmosperic Turbulence and Diffiusion Division of NOAA's Air Resources Laboratory, Oak Ridge, TN, USA; [11]SDSU, USA; [12]University of Sheffiled, UK; [13]University of Exeter, UK; [14]Research Institute for Global Change, Japan Agency for Marine-Earth Science and Technology, Japan; [15]Carleton University, Canada; [16]Max Planck Institute for Biogeochemistry, Germany; [17]Shinshu University, Japan; [18]Trent University, Canada; [19]University of Alaska Fairbanks, Institute of Arctic Biology, USA; [20]Vrije Universiteit Amsterdam, Netherlands; [21]University of Jyväskylä, Finland; [22]University of Eastern Finland, Finland; [23]Center for Permafrost, Department of Geoscience and Natural Resource Management, University of Copenhagen, Denmark; [24]University of Montreal, Canada; [25]Deparment of Geosciences, University of Oslo, Norway; [26]Department of Physical Geography and Ecosystem Science, Lund University, Sweden; [27]Agronomy Department, University of Florida, USA; [28]Department of Biological Sciences, University of Alberta, Canada; [29]Department of Forest Ecology and Management, Swedish University of Agricultural Sciences, Swedem; [30]Department of Biologial Sciences, Florida International University, USA; [31]Berkeley Lab and UC Berkeley, USA; [32]Korea Polar Research Institute, South Korea; [33]Department of Earth Sciences, Free University Amsterdam, Netherlands; [34]Institute for Atmospheric and Earth System Research/Physics, Faculty of Science, University of Helsinki, Finland; [35]Institutes of Life Sciences and Natural Resources, Korea University, South Korea; [36]CREAF, Cerdanyola del Vallès, Spain; [37]Department of Environment and Minerals, Greenland Institute of Natural Resources, Denmark; [38]Department of Bioscience, Arctic Research Center, Aarhus University, Denmark; [39]Laboratoire des Sciences du Climat et de l'Environnement, France

Corresponding author: Anna-Maria Virkkala, avirkkala@woodwellclimate.org

Citation: This paper has been accepted for publication by Global Change Biology.

Virkkala, A.-M., Aalto, J., Rogers, B.M., Tagesson, T., Treat, C.C., Natali, S.M., Watts, J.D., Potter, S., Lehtonen, A., Mauritz, M., Schuur, E.A.G., Kochendorfer, J., Zona, D., Oechel, W., Kobayashi, H., Humphreys, E., Goeckede, M., Iwata, H., Lafleur, P.M., Euskirchen, E.S., Bokhorst, S., Marushchak, M., Martikainen, P.J., Elberling, B., Voigt, C., Biasi, C., Sonnentag, O., Parmentier, F.-J.W., Ueyama, M., Celis, G., St.Louis, V.L., Emmerton, C.A., Peichl, M., Chi, J., Järveoja, J., Nilsson, M.B., Oberbauer, S.F., Torn, M.S., Park, S.-J., Dolman, H., Mammarella, I., Chae, N., Poyatos, R., López-Blanco, E., Christensen, T.R., Kwon, M.J., Sachs, T., Holl, D. and Luoto, M. (2021), Statistical upscaling of ecosystem CO_2 fluxes across the terrestrial tundra and boreal domain: Regional patterns and uncertainties. Glob. Change Biol., 27: 4040-4059.

https://doi.org/10.1111/gcb.15659

Abstract

The regional variability in tundra and boreal carbon dioxide (CO_2) fluxes can be high, complicating efforts to quantify sink-source patterns across the entire region. Statistical models are increasingly used to predict (i.e., upscale) CO_2 fluxes across large spatial domains, but the reliability of different modeling techniques, each with different specifications and assumptions, has not been assessed in detail. Here, we compile eddy covariance and chamber measurements of annual and growing season CO_2 fluxes of gross primary productivity (GPP), ecosystem respiration (ER), and net ecosystem exchange (NEE) during 1990–2015 from 148 terrestrial high-latitude (i.e., tundra and boreal) sites to analyze the spatial patterns and drivers of CO_2 fluxes and test the accuracy and uncertainty of different statistical models. CO_2 fluxes were upscaled at relatively high spatial resolution (1 km^2) across the high-latitude region using five commonly-used statistical models and their ensemble, i.e., the median of all five models, using climatic, vegetation, and soil predictors. We found the performance of machine learning and ensemble predictions to outperform traditional regression methods. We also found the predictive performance of NEE-focused models to be low, relative to models predicting GPP and ER. Our data compilation and ensemble predictions showed that CO_2 sink strength was larger in boreal biome (observed and predicted average annual NEE –46 and –29 g C m^{-2} yr^{-1}, respectively) compared to tundra (average annual NEE +10 and –2 g C m^{-2} yr^{-1}). This pattern was associated with large spatial variability, reflecting local heterogeneity in soil organic carbon stocks, climate, and vegetation productivity. The terrestrial ecosystem CO_2 budget, estimated using the annual NEE ensemble prediction, suggests the high-latitude region was on average an annual CO_2 sink during 1990–2015, although uncertainty remains high.

Keywords: land, empirical, Arctic, permafrost, greenhouse gas, CO2 balance, remote sensing

1. Introduction

The terrestrial ecosystem carbon dioxide (CO_2) balance is one of the largest uncertainties in the global carbon budget (Friedlingstein et al., 2020), with high-latitudes (i.e., tundra and boreal biomes) representing one of the least constrained budgets (López-Blanco et al., 2019; Schuur et al., 2015; Zscheischler et al., 2017). Moreover, due to polar amplification and large carbon stocks, the high latitudes have the potential for substantial positive feedbacks to climate warming (Abbott et al., 2016; Gasser et al., 2018; Schuur et al., 2008; Turetsky et al., 2020). Currently, in the absence of major disturbances (e.g., fire), boreal forests are generally CO_2 sinks (Bradshaw & Warkentin, 2015; Pan et al., 2011), while regional estimates of tundra vary from sinks (McGuire et al., 2009, 2012, 2016) to sources (Belshe et al., 2013). Both the winter and growing seasons are important for these annual budget estimates. A recent synthesis by Natali et al., (2019) found that winter soil CO_2 emissions from the northern permafrost region are larger than previously estimated, however CO_2 uptake by plants over the growing season can be substantial and is often the dominant component of the annual CO_2 budget (Alekseychik et al., 2017; Kolari et al., 2009; Lafleur et al., 2012). The current state of the annual terrestrial high-latitude CO_2 budget (net sink or source) remains highly uncertain. A key research priority is to develop and compare methods used to estimate CO_2 budgets so that best practices can be identified and regional boreal and tundra budgets constrained at annual and seasonal time scales.

Estimating high-latitude CO_2 fluxes across large areas and over long timescales is challenging due to their high spatiotemporal variability (Ai et al., 2018; Wilkman et al., 2018) that is controlled by a range of environmental variables (Camps-Valls et al., 2015; Lund et al., 2010). The ecosystem CO_2 balance (net ecosystem CO_2 exchange; NEE) is a relatively small difference between the two large CO_2 fluxes of photosynthesis (gross primary production; GPP) and ecosystem respiration (ER; comprising autotrophic and heterotrophic respiration). Although NEE can be measured with the eddy covariance (EC) and chamber techniques (Baldocchi et al., 1988; Lundegårdh, 1927), GPP and ER are estimated indirectly using environmental light and temperature measurements for EC sites (Lasslop et al., 2010; Reichstein et al., 2005) and light manipulations for chamber sites (Shaver et al., 2007). Field studies have shown that GPP, ER, and NEE depend on climatic conditions (e.g., temperature, precipitation, and radiation) (López-Blanco et al., 2017; Nobrega and Grogan, 2008; Zhang et al., 2018), vegetation (Cahoon et al., 2012; Fox et al., 2008; Järveoja et al., 2018), and soil properties (e.g., soil nutrients and moisture) (Arens et al., 2008; Dagg and Lafleur, 2011; Lund et al., 2009). However, our understanding of the influence of these drivers on GPP and ER, and particularly on NEE, across the entire boreal and tundra domain remains limited (see e.g., Belshe et al., 2013; Lund et al., 2010).

Knowledge of the contemporary high-latitude terrestrial CO_2 budget is further limited by an increasing, but still relatively sparse, flux measurement network (Alton, 2020; Chu et al., 2017; Virkkala et al., 2018). The majority of flux sites are concentrated within a few intensively studied regions, particularly Alaska and Fennoscandia (Metcalfe et al., 2018; Pastorello et al., 2020; Virkkala et al., 2019), with just a few sites in other large regions such as Siberia and northern Canada. Consequently, several methodological issues related to the temporal, geographical and environmental representativeness of the measurements need to be addressed to accurately estimate high-latitude carbon budgets. Previous studies have used a variety of synthesis approaches (Belshe et al., 2013; McGuire et al., 2012), and statistical (Natali et al., 2019), process-based (Lopéz-Blanco et al., 2019; McGuire et al., 2018; Rawlins et al., 2015; Wania et al., 2009) and atmospheric inversion models (McGuire et al., 2012), yielding highly different sink-source patterns. Most of these modeling studies have been conducted at coarse spatial resolutions (25 – 100 km km; Natali et al., 2019; Rawlins et al., 2015; López-Blanco et al., 2019) that do not fully capture the local heterogeneity in high-latitude environments despite their importance for the regional CO_2 budgets (Treat et al., 2018). New efforts synthesizing the current distribution of flux data and developing models at high spatial resolution are required to improve our understanding on the spatial patterns and magnitudes of CO_2 fluxes.

Models that rely on the statistical relationships between CO_2 flux and predictor variables have been increasingly employed (e.g., Jung et al., 2020; Natali et al., 2019; Warner et al., 2019). These statistical models are useful for predicting fluxes across larger areas (i.e., upscaling) because they directly draw upon relationships between fluxes and environmental variables, can account for environmental variability across space and time at high resolutions, and are able to handle biases in the geographic representation of the data (Jung et al., 2020; Natali et al., 2019; Warner et al., 2019). A broad range of statistical models and data sources are available for upscaling, but not all of these have been fully utilized. For example, many past studies have upscaled high-latitude fluxes using a single model (Natali et al., 2019; Peltola et al., 2019; Ueyama, Ichii, et al., 2013), but how different models compare with each other is not well known (with exception of Jung et al., 2017 and Tramontana et al., 2016). Further, most of the studies have primarily used machine learning models due to their ability to capture non-linear relationships in data and lack of required assumptions (Elith et al., 2008). However, traditional regression methods can be a powerful tool in upscaling high-latitude ground conditions due to their ability to extrapolate beyond the range of data used for training, and due to their generalizability and ease of interpretation (Aalto et al., 2018). Finally, many of the recent upscaling studies have relied on EC flux measurements, neglecting chamber measurements despite their importance as additional data sources, especially in remote, sparsely-measured treeless tundra where chambers can capture the entire ecosystem CO_2 balance and directly measure NEE and ER (Natali et al., 2019). Thus, a compilation of both EC and chamber flux measurements and the comparison of available modeling techniques is clearly required to ensure accurate CO_2 flux estimates from existing data and models.

Here, we synthesize annual and growing season CO_2 fluxes from EC and chamber measurements across the high-latitude terrestrial tundra and boreal biomes. We then use this new database to upscale annual average ecosystem CO_2 fluxes at relatively high spatial resolution (1 km^2) across the high-latitude domain using several statistical models. We compare our new database of *in situ* CO_2 fluxes to past tundra syntheses (Belshe et al., 2013; McGuire et al., 2012), provide a detailed assessment of model performance, analyze the spatial patterns and drivers of CO_2 fluxes, and discuss the resulting CO_2 budget estimates and recommendations for future work. We focus on understanding the spatial variability in average CO_2 fluxes instead of a temporal analysis of CO_2 flux change; however, our modeling framework also considers the interannual variability in fluxes.

2. Material and Methods

2.1 Data Collection

2.1.1 Collection of CO_2 flux data

Our study area was defined by the high-latitude tundra and boreal biomes (>45 °N) based on global ecoregions (20.6 x 10^6 km^2; Fig. 1; Dinerstein et al., 2017). We first conducted a literature survey to identify existing EC and chamber-based terrestrial CO_2 flux observations of GPP, ER, and NEE over annual and growing season periods across the domain. Potential sites were identified from previous studies (Ichii et al., 2017; Marushchak et al., 2013; McCallum et al., 2013; Watts et al., 2014) and prior synthesis efforts (Belshe et al., 2013; McGuire et al., 2012; Virkkala et al., 2018). We augmented the resulting site list using a Web Of Science search with key words ("tundra" or "boreal" or "arctic") and ("CO_2 flux" or "CO_2 exchange" or "CO_2 budget"). Additionally, a community call was solicited through a CO_2 flux synthesis workshop (Parmentier et al., 2019), whereby investigators contributed their most current unpublished data. Additional EC data were downloaded from FLUXNET2015 (Pastorello et al., 2020).

The compiled data set represents all natural vegetation types (categorized by needle- or broadleaf forest, shrubland, grassland, wetland, and sparse vegetation) present in the study domain. We included flux measurements from managed forests and wetlands but excluded croplands. While the EC observations represent all vegetation types, chamber data from forest sites were not included since they do not represent whole ecosystem fluxes. EC measures NEE directly, whereas GPP and ER are indirect estimates acquired from various partitioning methods (Lasslop et al., 2010; Reichstein et al., 2005). NEE is also often gap filled with the indirect GPP and ER estimates. Chambers measure NEE and ER directly, out of which GPP can be estimated. If a given site reported both EC and chamber fluxes for the same year and period, EC fluxes were selected over chambers as EC footprints are larger and correspond better with the scale of our gridded predictor variables. In experimental manipulation studies, only the fluxes from the control plot were included. We aggregated spatial replicates of chamber fluxes within a given site and year by calculating the median flux.

We included studies and sites with NEE, GPP, and ER estimates over a full growing season or year (i.e., cumulative flux). Growing season flux measurements are provided by EC and chambers. Winter flux measurements include a variety of methods in addition to EC and chambers (e.g., a gas diffusion method by Björkman et al., 2010, soda lime by Welker et al., 2004, or an empirical model by Vogel et al., 2009). Growing season length and measurement period were defined in multiple ways at individual sites. To allow inter-site comparison, we filtered out measurements that did not represent the entire growing season and standardized the remaining measurements (see Supplementary Text Section 1.1 and a similar approach in Belshe et al., 2013). From this filtered data set, we calculated average growing season daily flux rates based on the reported measurement length and standardized the fluxes based on a common growing season length. The final list of sites having representative annual or growing season measurements is provided in Supplementary Table 1, sites that were dropped are in Supplementary Table 2.

The resulting dataset included 148 sites with CO_2 fluxes from 1990 to 2015 from variable measurement periods (Fig. 1). We compiled 1390 cumulative annual and growing season flux values (when chamber measurements were aggregated per site; Fig. 1); 78 % of the aggregated observations are from EC and 22 % are from chambers. Annual and growing season NEE were the most widely reported fluxes in the dataset (Fig. 1). Unlike McGuire et al., (2012) and Belshe et al., (2013) we also included data from the boreal biome, additional tundra sites, and wetlands (not synthesized in Belshe et al., 2013; Supplementary Fig. 1). Similar to McGuire et al., (2012) and Belshe et al., (2013), our database primarily represents undisturbed environments. However, it also includes measurements from ca. 10 sites that have experienced high natural, anthropogenic or anthropogenically-induced disturbances, such as rapid permafrost thaw (Bäckstrand et al., 2010; Cassidy et al., 2016; Trucco et al., 2012), fires (Iwata et al., 2011; Ueyama et al., 2019), insect outbreaks (Heliasz et al., 2011; López-Blanco et al., 2017; Lund et al., 2017), or extensive harvesting practices (Coursolle et al., 2012; Machimura et al., 2005). Throughout the text, positive numbers for NEE indicate net CO_2 loss to the atmosphere (i.e., CO_2 source) and negative numbers indicate net CO_2 gain (i.e., CO_2 sink). GPP and ER are always given as positive numbers.

2.1.2 Gridded predictors and reference flux data

We acquired 10 eco-physiologically relevant predictors at 1 km^2 resolution (0.0083°) representing topographic, soil, climate, and vegetation conditions: topographic wetness index (TWI), potential incoming direct annual radiation (RAD; MJ cm^{-2} yr^{-1}), soil organic carbon stocks in the upper 2 m (SOC; tons per ha), topsoil (0-5 cm) pH, topsoil clay content (CLAY; %), growing degree days (GDD3; °C), freezing degree days (FDD; °C), water balance (WAB; mm), normalized difference index (NDVI) and land cover (LC; classes were mixed or broadleaved forest, needleleaved forest, grassland and shrubland, wetland, sparse vegetation; see Supplementary Text Section 1.2 and Supplementary Fig. 2 for more information about the predictors). These predictors characterize previously identified key relationships between CO_2 fluxes and summer and winter temperatures, radiation, precipitation, local hydrology and soil conditions, soil carbon stocks, and vegetation properties (i.e., see Beer et al., 2010; Belshe et al., 2013; Lund et al., 2010; Natali et al., 2019; Ueyama, Iwata, et al., 2013). We recognize that GPP and ER partitioning and gap filling rely on some environmental data (e.g., temperature and radiation), and consequently these fluxes already include some information about variables that we also used as predictors in our statistical models. We used annual (1990–2015) data for GDD3, FDD, WAB, and maximum summer NDVI; the remaining predictors were considered to be static. In cases where an annual flux value extended over multiple years (i.e., measurement period from October to September of the following year, or where a study reported an average flux from multiple years), a median climate or NDVI value for those years was used. All predictor data sets were masked to only include tundra and boreal biomes (Dinerstein et al., 2017), and to exclude permanent water bodies, urban areas, and croplands based on a land cover dataset developed by ESA, (2017).

We compared our annual ecosystem NEE predictions and budgets (see Section 2.2.1) with FLUXCOM, a global product derived from FLUXNET EC towers and machine learning at 0.5 ° resolution (Baldocchi et al., 2001; Jung et al., 2017; Tramontana et al., 2016) and an ensemble of global Earth system models from the Coupled Model Intercomparison Project Phase 5 (CMIP5) at 1.92 x 1.5 ° resolution (Taylor et al., 2012) (Supplementary Text Section 1.2).

2.2 Data Analysis

2.2.1 Statistical Modeling

Our main response variables were annual and growing season cumulative GPP, ER, and NEE, but we also modeled daily average GPP, ER, and NEE during the growing season. Annual and growing season CO_2 fluxes were linked to the environmental predictors using a range of different statistical modeling methods (Supplementary Fig. 3). We used five statistical models; two were extensions of linear regression models, and three were based on machine-learning. All of these models have been widely used in empirical CO_2 flux upscaling studies (Bond-Lamberty and Thomson, 2010; Hursh et al., 2016; Tramontana et al., 2016; Ueyama, Ichii, et al., 2013). Specifically, we examined generalized linear models (GLMs); generalized additive models (GAMs); generalized boosted regression trees (GBMs); random forest (RF models); and support vector machines (SVMs). GLM is an extension of linear regression models where the response variable can have a non-normal distribution, and the regression is generalized by linking the linear model to the response variable via a link function (Nelder and Wdderburn, 1972). GAM is a more flexible method than generalized linear modeling, as it can use local spline smoothing functions constrained by the user to fit non-linear relationships between the response variable and the predictor (Hastie and Tibshirani, 1987). GBM and RF are tree-based machine learning methods, where modeling is based on splitting the data into multiple trees (Breiman, 2001; Elith et al., 2008). SVM is a powerful machine learning method based on projecting vectors into a high-dimension space with a kernel function and then fitting an optimal hyperplane (Smola and Schölkopf, 2004).

We used several model approaches because individual models have inherent strengths and weaknesses (Supplementary Text Section 2). For example, machine learning methods might suffer from overfitting, whereas regression methods might result in unrealistic values when extrapolated outside the model data range. Further, individual models may detect different patterns in the data, and the best performing models are not always the same for different response variables (Segurado and Araújo, 2004). We also produced an ensemble prediction by calculating a median prediction over the five predictions from the individual modeling methods (see also Tramontana et al., 2016). We used the median instead of the mean to avoid extreme predicted values inflating the ensemble prediction. In this procedure, the uncertainty of the ensemble is expected to be lower than the uncertainty of a single model (Aalto et al., 2018). Consequently, we produced six model predictions for each of our response variables.

To determine the main drivers of the spatial patterns of response variables, the relative contribution of predictors in the models was assessed using a prediction re-shuffling approach (Niittynen and Luoto, 2018). We first fit the model and developed predictions using the original data, and then repeated this procedure with the values for one predictor randomly permuted. The contribution of a variable was calculated as a correlation between these two predictions (i.e., original model and the model with a shuffled predictor) subtracted from one:

$$\text{Relative contribution} = 1 - \text{correlation}\,(\text{Prediction}_{\text{original data}}, \text{Prediction}_{\text{Randomly permuted data}})$$

Values close to 1 indicate that the two predictions were different, indicating high variable importance of the predictor variable. Each predictor was randomly permuted 100 times for each flux with each of the modelling methods, and an ensemble contribution was derived by taking a mean of the values. To visualize a predictor's effect on a response variable after controlling for the effects of other predictors, partial dependence plots were derived from the random forest model. For both variable importance and partial dependence plot analyses, we used daily average growing season fluxes because the growing season length estimates that were used to calculate growing season fluxes are not independent from GDD3. We found that the daily average fluxes correlated strongly with the growing season fluxes (Pearson's correlation 0.93-0.94), so they can be assumed to reflect the same relationships with the predictors.

To extrapolate across the domain, we fit the models using the entire data set to produce annual flux predictions and their ensembles that were subsequently averaged to 1990-2015 mean values. Because the ensemble predictions were among the most accurate and least uncertain predictions across all response variables, and because their use is generally recommended in predictive efforts (Araújo and New, 2007), our final flux maps were based on the flux ensemble. Because growing season length has been estimated in several different ways in previous studies, we aggregated growing season budgets for two additional periods to compare the tundra and northern permafrost region growing season budgets to previous studies: Belshe et al., (2013) and Natali et al., (2019). Belshe et al., (2013) estimated the growing season to be 100 days at each site, and Natali et al., (2019) used the May-September period (153 days) for the growing season. For this comparison, we calculated a growing season NEE budget by multiplying the growing season daily NEE predictions by 100 and 153 days. However, we suggest our time-varying growing season estimate more reliably represents true growing season length as it captures the variability in growing season length across the high-latitude region. Regional budgets of annual NEE and the time-varying 100- or 153-day growing season NEE were calculated for the entire study domain (i.e., tundra and boreal biomes; Dinerstein et al., 2017), the northern permafrost region (Brown et al., 2002; excluding permafrost south of the boreal biome; includes regions both in tundra and boreal biomes), the non-permafrost region located within our study domain (includes boreal regions in Fennoscandia and some parts of Russia and Canada), and the boreal and tundra wetland and upland regions (based on the biomes and wetland and non-wetland classes in LC; ESA, 2017) by averaging the budgets estimated from annual ensemble predictions over the 26-year period. In addition to annual and growing season budgets, we also calculated a non-growing season budget (see Supplementary Table 3). We had different numbers of observations and sites available for each flux and model, and consequently observed and predicted ER and GPP fluxes and budgets do not sum up to NEE.

2.2.2 Model fit, predictive performance and uncertainty

To evaluate model fit, we predicted fluxes over the entire model training data. To assess the predictive performance of the models, we used a leave-one-site-out cross validation scheme in which each site was iteratively left out from the data set, and the remaining data were used to predict fluxes for the excluded site (Bodesheim et al., 2018). For both model fit and predictive performance, we calculated bias an average of the absolute error between prediction and actual observation, Pearson correlation (r) to determine the extent of linear relationship between the observed and predicted fluxes, and root mean squared error (RMSE) to estimate the model error. We use the terms "observed" and "predicted" to distinguish between field measurements and model predictions but acknowledge that some of these observed values represent indirect estimates of fluxes.

We evaluated the prediction uncertainty of all flux models and the budget uncertainty of annual and growing season NEE models using a repeated random resampling procedure (Aalto et al., 2018). Prediction uncertainty was calculated to characterize the spatial variability in flux predictions across the high-latitude region, whereas budget uncertainty quantified the range of potential NEE budget values. We used bootstrapping (fractional resampling with replacement based on LC classes) to subset the model training data into 200 different data sets, all of which had the same number of observations as the original flux data itself. These 200 data sets were then used to produce 200 individual predictions with all five statistical models and their ensemble for each flux and for each year from 1990 to 2015 to assess prediction uncertainty which was summarized using the prediction interval (PI; 95th percentile – 5th percentile). Uncertainty for annual and growing season NEE budgets was estimated by calculating the range of budgets from the 50 first ensemble predictions out of the 200 predictions for each year from 1990 to 2015, due to computational constraints. The prediction uncertainty of annual NEE was also assessed by comparing the average annual NEE budgets with the annual NEE derived from annual ER and GPP predictions, by examining alternative estimates from other studies (i.e., FLUXCOM and CMIP5) and by calculating a standard deviation across these products to evaluate where the regional differences occur. For more details, see Supplementary Text section 2.3 and Supplementary Fig. 4.

3. Results

3.1 Observed flux variation

Flux measurements showed considerable variation in magnitudes and signs (sink vs source) across the high-latitude environments (Fig. 1 and Table 1). Observed annual NEE (no upscaling) was on average a small source of CO_2 in the most northern parts of the study domain (tundra: +10 g C m^{-2} yr^{-1}, 42 sites, northern permafrost region: +6 g C m^{-2} yr^{-1} based on 63 sites) and in drier environments (tundra upland: +16 g C m^{-2} yr^{-1}, 34 sites), whereas the boreal biome (–46 g C m^{-2} yr^{-1}, 41 sites), and in particular boreal uplands (–47 C m^{-2} yr^{-1}, 36 sites), and non-permafrost-boreal regions (–90 g C m^{-2} yr^{-1}, 20 sites) were net ecosystem CO_2 sinks. All environmental categories were, on average, net CO_2 sinks during the growing season, with the average NEE ranging from –37 to –115 g C m^{-2} $period^{-1}$ (Table 1). Tundra upland and non-permafrost regions had the lowest average growing season sink strength. The non-permafrost region sink was greatly reduced by one disturbed site that had large source values up to +600 g C m^{-2} $period^{-1}$ (Petrone et al., 2014), but this was not apparent in the annual averages because the same site did not report annual fluxes. Although the distribution of environmental conditions at the sites were fairly representative

(Supplementary Fig. 5), colder environments with low NDVI and GDD3 as well as high FDD were less well represented (e.g., large areas of Siberia; Fig. 1).

3.2 Predictive performance of the models

The model fit and predictive performance analyses indicated that the GBM, RF and SVM (machine learning) methods outperformed the GLM and GAM (regression model) approaches across most of the response variables (in particular with NEE, but also with GPP and ER; model fit of annual machine learning models: r = 0.69–0.99 vs. regression models: r = 0.6–0.92; predictive performance of annual machine learning methods: r = 0.2–0.73 vs. regression models: r = 0.12–0.72; Fig. 2). We found that the machine learning-based methods were less uncertain (Supplementary Fig. 6) and always predicted values within the range of the observed fluxes as opposed to regression models. However, the machine learning method that performed best and had the least uncertainties varied depending on the flux response variable.

Ensemble predictions were among the best performing models (model fit of annual and growing season ensemble models: r = 0.68–0.94; predictive performance of annual and growing season ensemble models: r = 0.21–0.73; Fig. 2 and Supplementary Fig. 7). However, similar to the individual models, model fit and predictive performance was lower for annual and growing season NEE compared to GPP and ER (model fit for GPP and ER: r = 0.89–0.94 vs. NEE: r = 0.68–0.77; predictive performance for GPP and ER: r = 0.53–0.71 vs. NEE: r = 0.21–0.27; Fig. 2 and Supplementary Fig. 7). Annual models for ER and NEE exhibited a better fit and predictive performance than the growing season models, whereas the opposite was true for GPP (Fig. 2 and Supplementary Fig. 7). The growing season GPP model fit and predictive performance was higher than that of the ER models, but annual GPP and ER models performed equally well. In most predictive performance analyses the lowest and highest observed fluxes were over- and underestimated, respectively, indicating overall poor predictive performance at the extremes (Supplementary Fig. 8–9).

Average predicted and observed fluxes were of similar magnitude (Table 1). However, there was a tendency for the average predicted values to have slightly larger GPP and ER values (e.g., observed and predicted annual GPP in the tundra: 250 g C m^{-2} and 378 g C m^{-2}, respectively) and stronger net CO_2 sink values than what was observed (e.g. observed and predicted annual NEE in the tundra: +10 g C m^{-2} and –2 g C m^{-2}, respectively). Our cross-comparison of annual and growing season flux ensemble predictions showed there was a mismatch between annual and growing season component fluxes in approximately 2 % of the pixels (growing season GPP/ER larger than annual GPP/ER) and that unrealistic flux values (negative GPP or ER) were found in less than 0.01 % of the pixels in the ensemble predictions.

3.3 Predicted flux variation

Predicted fluxes showed pronounced spatial variability across the region with a general trend towards increasing fluxes and sink strength with decreasing latitude for GPP, ER, and NEE (Fig. 3 and Supplementary

Fig. 10). The variability was related to differences in climate (GDD3 and FDD), solar radiation (RAD) and vegetation greenness (NDVI), which had the strongest influence on most of the fluxes (Fig. 4). Moreover, SOC, CLAY, and LC were important variables for annual NEE; CLAY and SOC both had a positive yet saturating relationship. The relationship between LC and annual NEE suggested that the annual and growing season net sink strength was largest in wetlands and smallest in sparse vegetation (Supplementary Fig. 11–12). Some variables had a very low variable importance for most of the fluxes (e.g. TWI, soil pH).

Our predictions revealed regional hot spots in annual and growing season NEE, GPP, and ER. Strong annual and growing season CO_2 sinks, having low ER and high GPP, were found in forested regions with high GDD3, NDVI, RAD, and low FDD across Fennoscandia and European Russia, southern Canada, and southern Siberia (Fig. 3 and Supplementary Fig. 10). Annual CO_2 sources were identified within northern and central Siberia, Greenland, northern and central Alaska, as well as northern Canada. These regions were located mainly in the tundra, characterized by high FDD, and low GDD3 and NDVI. Growing season CO_2 sources were located in southeastern Siberia, northern Siberia and some parts of southern and northern Canada. Largest uncertainties in flux predictions were found in areas with relatively strong CO_2 sinks in the boreal biome, such as in Fennoscandia and eastern Canada, but also in the tundra (e.g., Canadian Arctic Archipelago; Fig. 3 and Supplementary Fig. 10). The largest differences across our annual NEE, and CMIP5 and FLUXCOM predictions were found in European Russia, Fennoscandia, and southeastern Canada (Fig. 5a-d).

3.4 Terrestrial ecosystem NEE budget for the high-latitude region

Our ensemble predictions showed that the annual terrestrial ecosystem CO_2 sink was considerable for the high-latitude tundra and boreal region over the 26-year (1990–2015) study period (Table 2). The annual NEE budget (based on upscaled NEE data) averaged –419 Tg C yr^{-1} (90 % uncertainty range: –559 to –189 Tg C yr^{-1}; range of budgets across the 26-year time period: –449 to –366 Tg C yr^{-1}). When estimating annual NEE according to the separately modeled annual GPP (11,344 Tg C yr^{-1}) and ER (10,397 Tg C yr^{-1}) budgets, we obtain a NEE budget of –948 Tg C yr^{-1}. The average high-latitude growing season NEE budget over the period of 1990–2015 was –1,018 Tg C yr^{-1} (–1,332 to –455 Tg C yr^{-1}, 90 % uncertainty range), which was supported by the difference between the average growing season ER (5,800 Tg C yr^{-1}) and GPP (7,016 Tg C yr^{-1}) budgets. For the regional budgets, see Table 2.

The average annual NEE budgets over the study period from CMIP5 and FLUXCOM were –488 and –1056 Tg C yr^{-1}, respectively (Supplementary Table 4). In the boreal biome, average annual GPP in our study was 8,850 compared to 8,561 Tg C yr^{-1} in FLUXCOM. In the tundra biome, the average annual GPP in this study was twice as high as in FLUXCOM (2,495 and 1,267 Tg C yr^{-1}, respectively). Differences were larger for annual ER. Our annual ER budget for the boreal and tundra biomes was 8,241 and 2,156 Tg C yr^{-1}, respectively, but the same budgets were only 6,363 and 1,200 Tg C yr^{-1} in FLUXCOM. For the regional NEE budgets estimated with CMIP5 and FLUXCOM, see Supplementary Table 4.

4. Discussion

This study provides a conceptual and methodological framework to bridge the gap between local, regional, and high-latitude scales in statistical flux upscaling. Our framework is unique in that it 1) compiles a new data synthesis of growing season and annual fluxes using EC and chamber data and investigates the drivers of these fluxes; 2) quantifies the performance of different statistical models; and 3) provides the first spatially continuous high-latitude maps of CO_2 fluxes and their uncertainties at high spatial resolution, capturing the inherent spatial heterogeneity in predictors and fluxes and minimizing biases in upscaling compared to coarser scale models (Fig. 5e). The better geographical and environmental coverage of the flux measurements compared to past efforts improves our understanding of the spatial patterns and regional budgets of terrestrial ecosystem CO_2 fluxes, however uncertainties in our direct model estimates of NEE remained rather high.

4.1. Drivers and spatial patterns of GPP, ER, and NEE

Our results suggest that climatic, vegetation, and soil variables were all important predictors for terrestrial ecosystem CO_2 fluxes. However, almost all CO_2 fluxes were strongly driven by the broad climatic gradients and spatiotemporal variability in radiation, growing and winter season climatic conditions, water balance, and the resulting vegetation greenness patterns, supporting the findings of previous syntheses (Belshe et al., 2013; Lund et al., 2010; Natali et al., 2019). Even though these climatic variables are not independent of our GPP and ER estimates (see section 4.2.), confidence in these results can be drawn from the underlying mechanistic relationships between the climate drivers and fluxes. For example, GPP across large scales is dependent on growing season temperatures, length of season, and radiation, which regulate and provide resources for plant growth (López-Blanco et al., 2017; Lund et al., 2010), and ER is largely driven by enzymatic processes, which are tightly linked with temperatures (Davidson et al., 2006) as well as plant growth (La Puma et al., 2007). In general, we found that warmer, moderately wet, and greener conditions (i.e., environments of higher biomass as indicated by NDVI) increased the magnitude of annual GPP and ER. However, our results also indicate that the overall net sink strength increases with larger greenness, warmer and shorter winters, and wetter climate. These results suggest that GPP and ER respond rather similarly to changes in climate and vegetation conditions across the high-latitude region, although GPP might increase even more due to increases in vegetation greenness (Berner et al., 2020) and changing climate (Lund et al., 2010). However, differences in these relationships might occur in different regions (Belshe et al., 2013) and land cover types (Baldocchi et al., 2018; Lafleur et al., 2012).

In addition to the climate and greenness variables operating mostly at large scales, other more local-scale variables such as soil organic carbon stock and land cover helped explain CO_2 fluxes. Soil organic carbon stock was the most important predictor for annual NEE, and it had a positive relationship with it, demonstrating that areas with high carbon stocks might lose more CO_2. However, this result was not supported by the annual ER models, which would represent the main process behind this relationship (i.e., larger carbon stocks have more potential for increased CO_2 emissions, particularly in dry conditions (Voigt et al., 2019)). The lack of this relationship might be due to annual ER models not covering the full range of conditions represented by the annual NEE models, or spurious causal relationships being identified by the relatively poorly-performing NEE models. The importance of land cover was expected as it summarizes many key processes related to carbon cycling (e.g. the carbon uptake capacity, temperature sensitivity, as well as quantity and quality of carbon inputs into the soil; Sørensen et al., 2019) and distinguishes other environmental characteristics across the land cover types (e.g., soil moisture is likely higher in wetlands than in sparse vegetation).

Our ensemble prediction suggested that most of the southern high-latitude terrestrial region was an annual net ecosystem CO_2 sink while the central and northern regions were neutral or small net CO_2 sources. Observed and predicted spatial patterns in fluxes were similar to those described by most previous studies. For example, our compiled field observations and predictions are consistent with the majority of Alaskan tundra being an annual ecosystem CO_2 source on average, similar to the average observed fluxes in McGuire et al., (2012) or the prediction in Ueyama et al., (2013). The strongest annual ecosystem CO_2 sinks in our study were located in southern European Russia, Fennoscandia, and southern Canada, as also observed in the FLUXCOM products (Jung et al., 2017; Tramontana et al., 2016).

For some regions, our ensemble prediction differed from the predictions of previous studies. The distribution of annual CO_2 sources across the tundra biome was larger in our prediction compared to FLUXCOM, particularly in Siberia and Canada. This was likely explained by our models being based on a larger number of tundra sites from Canada, Greenland, European Russia, and Siberia, which were not covered by the FLUXCOM model training data. Some of the sites in these regions were annual CO_2 sources on some years (Emmerton et al., 2016; Karelin et al., 2013). A similar disagreement was found between an Asia-focused statistical upscaling analysis by Ichii et al., (2017) which suggested stronger sink strength across large parts of Siberia, likely due to a limited number of northern eddy covariance sites used to train their models. The largest regional differences between our predictions, CMIP5, and FLUXCOM occurred in central Siberia, Fennoscandia, European Russia, and Canada, and these differences were primarily driven by the fact that CMIP5 showed these regions to be primarily sources whereas they were sinks in FLUXCOM and our analysis (Fig. 5). These regional differences demonstrate that these particular areas should be studied further to understand the sink-source patterns more accurately in the future.

Our uncertainty estimation suggests that CO_2 flux predictions should be interpreted carefully in areas that lack sampling locations or have large variability in fluxes that cannot be captured by the predictor variables.

Such areas are particularly concentrated in European Russia, eastern Canada, and the Canadian Arctic Archipelago. As the accuracy of the prediction can usually be improved with increases in the quantity and quality of data, new measurements in these regions would likely improve the accuracy of high-latitude CO_2 flux models.

4.2 Key sources of uncertainty in our modeling approach

No single best model could be identified across the five modeling methods. However, the three machine learning methods outperformed the two regression models, particularly for NEE, as demonstrated by the improved model performance, lower uncertainty and the lack of unrealistically high or low flux values in predictions. The better performance of the machine learning methods was likely related to their flexibility and capability to find complex structures in the flux data (Elith et al., 2008). Our results demonstrate that several machine learning methods should be tested to produce the most accurate high-latitude flux predictions and that ensemble methods provide robust predictions (Araújo and New, 2007). Our results also indicate that an ensemble prediction based on machine learning methods alone would likely lead to higher model accuracy and transferability (see also Tramontana et al., 2016).

Our models performed well when predicting to the same data that the models were trained with, but the models had challenges when predicting to new data. The predictive performance of our ensemble predictions was comparable to (annual GPP and ER) or less than (growing season GPP, ER, NEE, and annual NEE) that of in other global and regional upscaling studies (Ichii et al., 2017; Natali et al., 2019; Peltola et al., 2019; Tramontana et al., 2016; Ueyama, Ichii, et al., 2013). However, comparisons of cross-validation results are hampered by different cross-validation techniques used in studies, with some of the studies including observations from the same site both in the model training and validation data, therefore providing overly optimistic accuracy estimates based on non-independent data. Moreover, these other studies primarily focused on a smaller area and/or shorter time period (with the exception of Tramontana et al., 2016), and used a different set of predictors, further complicating this comparison. In these other studies, the correlation (r) between observed and predicted fluxes (derived with cross validation), measured mostly throughout the year as daily-to-monthly fluxes, was roughly 0.65–0.7 for NEE and 0.7–0.8 for GPP and ER. There are several reasons for why some of our models performed more poorly than these previous studies, which we explain below.

The lower quantity of measurements and weaker comparability of fluxes derived with EC and chamber techniques and with variable measurement lengths might explain the lower predictive performance in our study compared to the other upscaling studies. As we used aggregated fluxes over the growing season and annual time scales, the sample size in our models was smaller than in other efforts which all used daily or monthly fluxes (a few hundred observations versus thousands of observations). A larger sample size usually increases the predictive performance of the models, particularly when these measurements cover variable environmental conditions that can be captured by the predictors. For example, FLUXCOM models (Jung et al., 2017, 2020; Tramontana et al., 2016) might have had a higher predictive performance than our models because they use a global FLUXNET database (Pastorello et al., 2020), which covers broad environmental gradients. However, FLUXNET data originates mostly from lower latitudes (e.g., only five sites from the Arctic and 34 from the boreal out of 224 global sites in total used in Tramontana et al., 2016). This could explain the larger net sink strength in FLUXCOM compared to our predictions. The higher predictive performance of FLUXCOM compared to our prediction might also be explained by the fact that FLUXNET is based on a single flux measurement technique (EC) with standardized filtering, gap-filling, and partitioning procedures. Although the inclusion of chambers in this study was crucial for adequate environmental coverage, using both chamber and EC measurements, and different partitioning methods for EC, increased the number of different flux measurement techniques and study designs, and may have made the comparison of fluxes across sites more uncertain (Fox et al., 2008; Tramontana et al., 2016). Further, the lower predictive performance of growing season models compared to annual models was potentially related to the variable growing season measurement periods used across the studies. We accepted this variability because our goal was to use as many published fluxes as possible to improve the geographical and environmental coverage of sites.

The accuracy of our ensemble predictions varied depending on the flux, with the predictive performance being lowest for NEE models (r=0.21–0.27). The predictive performance of our GPP and ER models was much higher (r=0.53–0.73) and is comparable to past efforts (Ichii et al., 2017; Natali et al., 2019; Tramontana et al., 2016; Ueyama, Ichii, et al., 2013) because these fluxes better represent the biophysical processes describing carbon uptake and loss. GPP and ER also already included some information about variables that we used as predictors in our statistical models, which may introduce some circularity and artificially inflate the model performance. Our NEE models over- and underestimated low and high (i.e., large negative and positive) values, respectively, by approximately 100–200 g C m^{-2} yr^{-1}, which has also been demonstrated with NEE and other fluxes in previous upscaling studies (Ichii et al., 2017; Tramontana et al., 2016; Warner et al., 2019). These extreme values were often from disturbed sites experiencing for example, permafrost thaw or extreme forest management practices, or an observation that was notably different from the site mean. Based on the cross validation results of the individually-modeled annual NEE, a substantial fraction (54 %) of annual source observations were predicted to be sinks (similar to the pattern observed in Ichii et al., (2017) Fig. 3b), but some sink observations (24 %) were also predicted as sources. We also discovered that the observed average annual NEE was often larger (more positive) than the individually-predicted average NEE, which was either a result of the model not being able to predict sources accurately, or of the distribution of flux sites being biased towards environments with larger CO_2 source observations than the entire region on average (see the large number of sites with source observations originating primarily only from Alaska in Fig. 1). These results demonstrate that the predictors included in our analyses did not fully represent the spatial gradients and dynamic temporal variability in environmental conditions that influence carbon cycle processes, and particularly the high and low NEE conditions. Further research should explore improvements offered by other current and potential future predictors related to the disturbance and permafrost conditions, snow cover duration and snow depth, soil moisture and nutrient availability, and phenology, root properties, and microbial communities (Illeris et al., 2003; Järveoja et al., 2018; Nobrega and Grogan, 2007).

Even though the geographical and environmental coverage of the flux sites was improved in our study compared to previous efforts, our models included only ca. 10 sites from heavily disturbed conditions (see Section 2.1.1). Consequently, our sites did not cover the full range of disturbance and post-disturbance conditions and the associated impacts on CO_2 fluxes. For example, rapidly thawing permafrost and burned landscapes remained largely under-sampled across Siberia. These disturbances have a substantial impact on carbon cycling in high-latitude ecosystems (Abbott et al., 2016; Walker et al., 2019), including direct emissions from the disturbance (not estimated with our models) and typically increased net CO_2 emissions for several years to decades after the disturbance (Coursolle et al., 2012; Lund, Raundrup, et al., 2017; Turetsky et al., 2020) which should ideally be captured by our models. The lack of flux data representing disturbed conditions likely leads to underestimations in net ecosystem CO_2 emissions, and is generally thought as one of the key limitations in statistical upscaling efforts (Jung et al., 2020; Zscheischler et al., 2017).

4.3 Terrestrial ecosystem CO_2 budget and its uncertainty

Although our models may be biased towards sinks, our results suggested that high-latitude terrestrial ecosystems were on average an annual net CO_2 sink during 1990–2015. The uncertainty of this budget was high, as demonstrated by the low predictive performance of the annual NEE model, and the fact that budgets derived from different predictions (individual NEE predictions and ER-GPP predictions) differed by ca. 500 Tg C yr^{-1} – the latter most likely being linked to the different numbers of observations and sites available for each flux and model (Fig. 1). Nevertheless, the annual NEE budget was of similar magnitude to the one estimated by CMIP5 models and larger (less negative) than the one estimated by FLUXCOM (Supplementary Table 4). The boreal biome was responsible for most of this sink strength (–406 Tg C yr^{-1}, from–499 to –239 Tg C yr^{-1}; 13.9 x 10^6 km^2), whereas the tundra biome was on average a small sink (–13 Tg C yr^{-1}, from –81 to +62 Tg C yr^{-1}; 6.7 x 10^6 km^2) or a small source (+10 g C m^{-2} yr^{-1}), based on our average predictions and observations. This suggests that the tundra biome was on average close to CO_2 neutral, suggesting that the strong CO_2 sink strength, indicated by the large soil organic carbon stocks of this region (Hugelius et al., 2014), might be declining, demonstrating the sensitivity of the tundra carbon cycle to climate change (IPCC, 2019). Our tundra budget is within the range (though on average more positive, indicating stronger source) of the one comprising process and inversion models, and field-based estimates by McGuire et al., (2012) (−103 Tg C yr^{-1}, from −297 to +89 Tg C yr^{-1}). However, it differs from the source budget (+462 Tg C yr^{-1}, from +94 to +840 Tg C yr^{-1}; 10.5 x 10^6 km^2; wetlands not included) estimated by Belshe et al., (2013). The divergence of average annual NEE across our and Belshe et al. (2013) study is likely explained by our inclusion of fluxes from wetlands, which were on average annual net ecosystem CO_2 sinks (Table 1). The discrepancy between our and the McGuire et al., (2012) study can be explained by a 50 % increase in new annual tundra source observations in our data set (see e.g., Celis et al., 2017; Euskirchen et al., 2014), which were not included in the McGuire et al. (2012) analysis. Further, there are some differences in the study domain boundaries (e.g., Belshe et al., 2013 included alpine tundra across the globe to their aerial estimate of 10.5 x 10^6 km^2) which might explain some of the discrepancies between these studies, although the general patterns of these boundaries were rather similar (see e.g. Fig 1. in McGuire et al., 2012 vs. our tundra domain in Fig. 1).

Our findings suggest that both the boreal and tundra biomes were strong CO_2 sinks during the growing season. Our growing season CO_2 budgets estimated for the same seasons as in previous studies (see Section 2.2.1), derived both by predicting NEE as well as subtracting GPP from ER suggest that the growing season net uptake is stronger than or similar to the estimates in Belshe et al., (2013) and Natali et al., (2019). The growing season NEE budget calculated for 100 days in the tundra was –296 Tg C yr^{-1} in this study, compared to –137 +- 80 Tg C yr^{-1} in Belshe et al., (2013). The growing season NEE budget estimated for 153 days in the northern permafrost region in this study was –1,122 Tg C yr^{-1}, whereas the process model estimates varied between –687 and –1,647 Tg C yr^{-1} in Natali et al., (2019). Further, the observed daily average growing season NEE in tundra demonstrated a stronger sink strength than the average growing season NEE reported in McGuire et al., (2012) and Belshe et al., (2013) (–0.6, –0.2, and –0.3 g C m^{-2}, respectively). Even though we acknowledge that some plant uptake and CO_2 emissions occur outside of our defined growing season (i.e., our growing season estimates did not capture the spring and autumn seasons), our results demonstrate that growing season CO_2 uptake might be larger than previously thought.

4.4. Summary and next steps in high-latitude CO_2 flux upscaling

Overall, our findings suggest that statistical predictions aimed at describing high-latitude CO_2 flux patterns provide new insights into the understanding of broad GPP and ER patterns but require caution when attempting to directly estimate NEE. Furthermore, this study demonstrates that machine learning models are a robust and accurate empirical approach to predicting high-latitude terrestrial CO_2 fluxes, and that, at least in our case, no individual machine learning model definitively outperformed the others. This therefore supports the use of ensemble predictions to reduce uncertainties associated with a single method and to produce more robust predictions. Nevertheless, the building of better models with improved data remains the highest research priority. Our results suggest that the next steps for future high-latitude upscaling efforts are to 1) measure fluxes over the entire year in as many sites as possible, 2) establish new sites in data-poor regions and regions where CO_2 predictions were most uncertain, such as in European Russia, Siberia, eastern Canada, and Canadian Arctic Archipelago, and specifically in disturbed and high-Arctic conditions, 3) develop better geospatial predictors (e.g., describing soil moisture and nutrients or permafrost thaw) to explain fluxes, 4) conduct detailed sensitivity tests of the importance of the flux measurement method, data distribution, and different predictor data sets influencing the budgets, and 5) build models at a finer temporal resolution than annual and growing season, to capture rapidly changing transition periods and bypass issues associated with temporal aggregation and varying definitions of seasons. High-latitude specific models are needed to more accurately monitor current emissions and improve understanding of the role of high-latitude regions in the global carbon cycle, as large changes in carbon cycling are likely in the near future.

Acknowledgements

AMV was supported by Nordenskiöld-samfundet, The Finnish Cultural Foundation, Alfred Kordelin Foundation, Väisälä fund, and Jenny and Antti Wihuri Foundation. AMV and ML were also funded by the Academy of Finland (project 286950). JA acknowledges the funding by Academy of Finland (project 33761), while AL acknowledges strategic research funding by the Academy of Finland for SOMPA project (grant 312912). TT was funded by the Swedish National Space Board (SNSB Dnr 95/16). BR was supported by the NASA Carbon Cycle Science and Arctic-Boreal Vulnerability Experiment programs (ABoVE grant NNX17AE13G), SMN by NASA ABoVE (grant NNX15AT81A) and JDW by NNX15AT81A and NASA NIP grant NNH17ZDA001N. AMV, BR, SN, and JDW were also funded by the Gordon and Betty Moore foundation #8414. JK acknowledges NSF grant 1203583, DZ NSF 1204263 and 1702797 and WO NSF 1204263, and 1702798. HK, MU, and HI were funded by Arctic Challenge for Sustainability II grant JPMXD1420318865, and EH and PL by Natural Sciences and Engineering Research Council. MG acknowledges European Commission (INTAROS project, H2020-BG-09-2016, project 727890) and ESE NSF grants DEB-1636476, AON 856864, 1304271, 0632264, and 1107892, and the US Geological Survey. MM was funded by Academy of Finland (project: 317054) and the EU 6th Framework Programme project CARBO-North (project 036993) and CB and CV by the EU FP7-ENV project PAGE21 (project 282700) and the Nordic Center of Excellence project DEFROST. CB was further funded by the Academy of Finland (project: 314630), and CV by Academy of Finland (project 332196). BE acknowledges Danish National Research Foundation (CENPERM DNRF100) and FJWP Research Council of Norway (Winterproof, project 274711) and Swedish Research Council (WinterGap, project 2017-05268). VLSL and CE were funded by the Natural Sciences and Engineering Research Council and MP by FORMAS #2016-01289. JJ was funded by the Swedish Forest Society Foundation (2018-485-Steg 2 2017) and SFO by NSF grants PLR1504381 and PLR1836898. MST acknowledges Office of Biological and Environmental Research, DOE Office of Science, SJP Korean government (NRF-2016M1A5A1901769, KOPRI-PN20081) and NC Korean government (MSIP) (NRF-2018R1D1A1B07047778 and NRF-2016-M1A5A1901790). HS was funded by NESSC, Netherlands Earth System Science Centre and IM by Academy of Finland Flagship funding (project 337549) and ICOS-Finland by University of Helsinki funding. RP was funded by Humboldt Fellowship for Experienced Researchers, MBN by Swedish Research Council, and ELB by Greenland Research Council.

Author contributions

AMV and ML designed the study. AMV extracted the flux data from the literature and the data from the community call was designed and gathered by MM, TS et al. AMV, JA, and SP prepared the gridded data sets. ML, JA, and AMV developed the modeling framework. TT, CT, BR, JDW, and SMN commented on the analysis and AMV, with the help of JA and ML, conducted the analysis. Other authors contributed data and all authors were involved in the writing.

Data availability

Data are archived and freely available at Zenodo. The synthesis dataset is available at [link added after next week]. Averaged flux predictions and their uncertainties are available at [link added after next week]. The codes to run the statistical models and predictions together with the uncertainty estimation can be found in an R Markdown file as a supplement (Virkkalaetal_CO2flux_upscaling.pdf) [final edits to the document after next week].

References

Aalto J, Karjalainen O, Hjort J, et al. (2018) Statistical Forecasting of Current and Future Circum-Arctic Ground Temperatures and Active Layer Thickness. *Geophysical Research Letters* 45: 4889–4898.

Abbott BW, Jones JB, G Schuur EA, et al. (2016) Biomass offsets little or none of permafrost carbon release from soils, streams, and wildfire: an expert assessment. *Environmental Research Letters* 11(3): 1–13.

Ai J, Jia G, Epstein HE, et al. (2018) MODIS-Based Estimates of Global Terrestrial Ecosystem Respiration. *Journal of Geophysical Research: Biogeosciences* 123(2): 326–352.

Alekseychik P, Mammarella I, Karpov D, et al. (2017) Net ecosystem exchange and energy fluxes in a West Siberian bog. *Atmospheric Chemistry and Physics Discussions* (February): 1–20.

Alton PB (2020) Representativeness of global climate and vegetation by carbon-monitoring networks;

implications for estimates of gross and net primary productivity at biome and global levels. *Agricultural and Forest Meteorology*, Elsevier 290(April): 108017.

Araújo MB and New M (2007) Ensemble forecasting of species distributions. *Trends in Ecology and Evolution* 22(1): 42–47.

Arens SJT, Sullivan PF and Welker JM (2008) Nonlinear responses to nitrogen and strong interactions with nitrogen and phosphorus additions drastically alter the structure and function of a high arctic ecosystem. *Journal of Geophysical Research-Biogeosciences* 113(G3): 10.

Bäckstrand K, Crill PM, Jackowicz-Korczyński M, et al. (2010) Annual carbon gas budget for a subarctic peatland. *Biogeosciences* 7(1): 95–108.

Baldocchi D, Falge E, Gu L, et al. (2001) FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem-Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities. *Bulletin of the American Meteorological Society* 82(11): 2415–2434.

Baldocchi D, Chu H and Reichstein M (2018) Inter-annual variability of net and gross ecosystem carbon fluxes: A review. *Agricultural and Forest Meteorology* 249: 520–533.

Baldocchi DD, Hicks BB and Meyers TP (1988) Measuring biosphere-atmosphere exchanges of biologically related gases with micrometeorological methods. *Ecology* 69(5): 1331–1340.

Beer C, Reichstein M, Tomelleri E, et al. (2010) Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate. *Science* 329(5993): 834–838.

Belshe EF, Schuur EAG and Bolker BM (2013) Tundra ecosystems observed to be CO2 sources due to differential amplification of the carbon cycle. *Ecology Letters* 16(10): 1307–1315.

Berner LT, Massey R, Jantz P, et al. (2020) Summer warming explains widespread but not uniform greening in the Arctic tundra biome. *Nature Communications*, Springer US 11(1): 1–12.

Björkman MP, Morgner E, Björk RG, et al. (2010) A comparison of annual and seasonal carbon dioxide effluxes between sub-Arctic Sweden and High-Arctic Svalbard. *Polar Research* 29(1): 75–84.

Bodesheim P, Jung M, Gans F, et al. (2018) Upscaled diurnal cycles of land-Atmosphere fluxes: A new global half-hourly data product. *Earth System Science Data* 10(3): 1327–1365.

Bond-Lamberty B and Thomson A (2010) Temperature-associated increases in the global soil respiration record. *Nature* 464: 579–582.

Bradshaw CJA and Warkentin IG (2015) Global estimates of boreal forest carbon stocks and flux. *Global and*

Planetary Change 128: 24–30.

Breiman L (2001) Random Forest. *Machine Learning* 45: 5–32.

Brown J, Ferrians O, Heginbottom JA, et al. (2002) Circum-Arctic Map of Permafrost and Ground-Ice Conditions, Version 2. *Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center.*

Cahoon SMP, Sullivan PF, Shaver GR, et al. (2012) Interactions among shrub cover and the soil microclimate may determine future Arctic carbon budgets. *Ecology Letters* 15: 1415–1422.

Camps-Valls G, Jung M, Ichii K, et al. (2015) Ranking drivers of global carbon and energy fluxes over land. *International Geoscience and Remote Sensing Symposium*: 4416–4419.

Cassidy AE, Christen A and Henry GHR (2016) The effect of a permafrost disturbance on growing-season carbon-dioxide fluxes in a high Arctic tundra ecosystem. *Biogeosciences* 13(8): 2291–2303.

Celis G, Mauritz M, Bracho R, et al. (2017) Tundra is a consistent source of CO2 at a site with progressive permafrost thaw during 6 years of chamber and eddy covariance measurements. *Journal of Geophysical Research: Biogeosciences* 122(6): 1471–1485.

Chu H, Baldocchi DD, John R, et al. (2017) Fluxes All of the Time? A Primer on the Temporal Representativeness of FLUXNET. *Journal of Geophysical Research: Biogeosciences* 122(2): 289–307.

Coursolle C, Giasson M-A, Margolis HA, et al. (2012) Moving towards carbon neutrality: CO_2 exchange of a black spruce forest ecosystem during the first 10 years of recovery after harvest. *Canadian Journal of Forest Research* 42(11): 1908–1918.

Dagg J and Lafleur P (2011) Vegetation Community, Foliar Nitrogen, and Temperature Effects on Tundra CO2 Exchange across a Soil Moisture Gradient. *Arctic Antarctic and Alpine Research* 43(2): 189–197.

Davidson EAJEA, Janssens IAIA, Marks D, et al. (2006) Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature* 440(7081): 165–73.

Dinerstein E, Olson D, Joshi A, et al. (2017) An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm. *BioScience* 67(6): 534–545.

Elith J, Leathwick JR and Hastie T (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*.

Emmerton CA, St. Louis VL, Humphreys ER, et al. (2016) Net ecosystem exchange of CO2 with rapidly changing high Arctic landscapes. *Global Change Biology* 22(3): 1185–1200.

ESA (2017) *Land Cover CCI Product User Guide Version 2. Tech. Rep.*

Euskirchen ES, Edgar CW, Turetsky MR, et al. (2014) Differential response of carbon fluxes to climate in three peatland ecosystems that vary in the presence and stability of permafrost. *Journal of Geophysical Research G: Biogeosciences* 119(8): 1576–1595.

Fox AM, Huntley B, Lloyd CR, et al. (2008) Net ecosystem exchange over heterogeneous Arctic tundra: Scaling between chamber and eddy covariance measurements. *Global Biogeochemical Cycles* 22(2): 1–15.

Friedlingstein P, Sullivan MO, Jones MW, et al. (2020) Global Carbon Budget 2020. *Earth System Science Data* 12: 3269–3340.

Gasser T, Kechiar M, Ciais P, et al. (2018) Path-dependent reductions in CO2 emission budgets caused by permafrost carbon release. *Nature Geoscience*, Springer US 11(11): 830–835.

Hastie T and Tibshirani R (1987) Generalized Additive Models : Some Applications Generalized Additive Models : Some Applications. *Journal of American Statistical Association* 82(398): 371–386.

Heliasz M, Johansson T, Lindroth A, et al. (2011) Quantification of C uptake in subarctic birch forest after setback by an extreme insect outbreak. *Geophysical Research Letters* 38(1): 1–5.

Hugelius G, Strauss J, Zubrzycki S, et al. (2014) Estimated stocks of circumpolar permafrost carbon with quantified uncertainty ranges and identified data gaps. *Biogeosciences* 11(23): 6573–6593.

Hursh A, Ballantyne A, Cooper L, et al. (2016) The sensitivity of soil respiration to soil temperature, moisture, and carbon supply at the global scale. *Global Change Biology*: 2090–2103.

Ichii K, Ueyama M, Kondo M, et al. (2017) New data-driven estimation of terrestrial CO_2 fluxes in Asia using a standardized database of eddy covariance measurements, remote sensing data, and support vector regression. *Journal of Geophysical Research: Biogeosciences* 122(4): 767–795.

Illeris L, Michelsen A and Jonasson S (2003) Soil plus root respiration and microbial biomass following water, nitrogen, and phosphorus application at a high arctic semi desert. *Biogeochemistry* 65(1): 15–29.

IPCC (2019) *Special Report on the Ocean and Cryosphere in a Changing Climate*. Pörtner H-O, Roberts DC, Masson-Delmotte V, et al. (eds).

Iwata H, Ueyama M, Harazono Y, et al. (2011) Quick recovery of carbon dioxide exchanges in a burned black spruce forest in Interior Alaska. *Scientific Online Letters on the Atmosphere* 7(1): 105–108.

Järveoja J, Nilsson MB, Gažovič M, et al. (2018) Partitioning of the net CO2exchange using an automated chamber system reveals plant phenology as key control of production and respiration fluxes in a boreal peatland. *Global Change Biology* 24(8): 3436–3451.

Jung M, Reichstein M, Schwalm CR, et al. (2017) Compensatory water effects link yearly global land CO_2 sink changes to temperature. *Nature* 541(7638): 516–520.

Jung M, Schwalm C, Migliavacca M, et al. (2020) Scaling carbon fluxes from eddy covariance sites to globe: Synthesis and evaluation of the FLUXCOM approach. *Biogeosciences* 17: 1343–1365.

Karelin D V., Zamolodchikov DG, Zukert N V., et al. (2013) Interannual changes in PAR and soil moisture during the warm season may be more important for directing of annual carbon balance in tundra than temperature fluctuations. *Zhurnal Obshchei Biologii* 74(1): 3–22.

Kolari P, Kulmala L, Pumpanen J, et al. (2009) CO_2 exchange and component CO_2 fl uxes of a boreal Scots pine forest. *Boreal environment research* 14: 761–783.

La Puma IP, Philippi TE and Oberbauer SF (2007) Relating NDVI to ecosystem CO_2 exchange patterns in response to season length and soil warming manipulations in arctic Alaska. *Remote Sensing of Environment* 109(2): 225–236.

Lafleur PM, Humphreys ER, St. Louis VL, et al. (2012) Variation in peak growing season net ecosystem production across the Canadian arctic. *Environmental Science and Technology* 46(15): 7971–7977.

Lasslop G, Reichstein M, Papale D, et al. (2010) Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: Critical issues and global evaluation. *Global Change Biology* 16(1): 187–208.

López-Blanco E, Lund M, Williams M, et al. (2017) Exchange of CO_2in Arctic tundra: Impacts of meteorological variations and biological disturbance. *Biogeosciences* 14(19): 4467–4483.

López-Blanco E, Exbrayat JF, Lund M, et al. (2019) Evaluation of terrestrial pan-Arctic carbon cycling using a data-assimilation system. *Earth System Dynamics* 10(2): 233–255.

Lopez Blanco E, Exbrayat JF, Lund M, et al. (2019) Evaluation of terrestrial pan-Arctic carbon cycling using a data-assimilation system. *Earth System Dynamics* 10(2): 233–255.

Lund M, Christensen TR, Mastepanov M, et al. (2009) Effects of N and P fertilization on the greenhouse gas exchange in two northern peatlands with contrasting N deposition rates. *Biogeosciences* 6(10): 2135–2144.

Lund M, Lafleur PM, Roulet NT, et al. (2010) Variability in exchange of CO_2 across 12 northern peatland and tundra sites. *Global Change Biology* 16(9): 2436–2448.

Lund M, Raundrup K, Westergaard-Nielsen A, et al. (2017) Larval outbreaks in West Greenland: Instant and subsequent effects on tundra ecosystem productivity and CO2 exchange. *Ambio* 46: 26–38.

Lundegårdh H (1927) Carbon Dioxide Evolution of Soil and Crop Growth. *Soil Science* 23(6): 417–453.

Machimura T, Kobayashi Y, Iwahana G, et al. (2005) Change of Carbon Dioxide Budget during Three Years after Deforestation in Eastern Siberian Larch Forest. *J. Agric. Meteorol.* 60(5): 653–656.

Marushchak ME, Kiepe I, Biasi C, et al. (2013) Carbon dioxide balance of subarctic tundra from plot to regional scales. *Biogeosciences* 10(1): 437–452.

McCallum I, Franklin O, Moltchanova E, et al. (2013) Improved light and temperature responses for light-use-efficiency-based GPP models. *Biogeosciences* 10(10): 6577–6590.

McGuire AD, Anderson LG, Christensen TR, et al. (2009) Sensitivity of the carbon cycle in the Arctic to climate change. *Ecological Monographs* 79(4): 523–555.

McGuire AD, Christensen TR, Hayes D, et al. (2012) An assessment of the carbon balance of Arctic tundra: comparisons among observations, process models, and atmospheric inversions. *Biogeosciences* 9(8): 3185–3204.

McGuire AD, Koven C, Lawrence DM, et al. (2016) Variability in the sensitivity among model simulations of permafrost and carbon dynamics in the permafrost region between 1960 and 2009. *Global Biogeochemical Cycles* 30: 1015–1037.

McGuire AD, Lawrence DM, Koven C, et al. (2018) Dependence of the evolution of carbon dynamics in the northern permafrost region on the trajectory of climate change. *Proceedings of the National Academy of Sciences of the United States of America* 115(15): 3882–3887.

Metcalfe DB, Hermans TDG, Ahlstrand J, et al. (2018) Patchy field sampling biases understanding of climate change impacts across the Arctic. *Nature Ecology and Evolution*: 1–6.

Natali SM, Watts JD, Rogers BM, et al. (2019) Large loss of CO 2 in winter observed across the northern permafrost region. *Nature Climate Change* 9: 852–857.

Nelder JA and Wdderburn RWM (1972) Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)* 135(3): 370.

Niittynen P and Luoto M (2018) The importance of snow in species distribution models of arctic vegetation. *Ecography* 41(6): 1024–1037.

Nobrega S and Grogan P (2007) Deeper snow enhances winter respiration from both plant-associated and bulk soil carbon pools in birch hummock tundra. *Ecosystems* 10(3): 419–431.

Nobrega S and Grogan P (2008) Landscape and ecosystem-level controls on net carbon dioxide exchange along a natural moisture gradient in Canadian low arctic tundra. *Ecosystems* 11(3): 377–396.

Pan Y, Birdsey RA, Fang J, et al. (2011) A Large and Persistent Carbon Sink in the World's Forests. *Science* 333: 1239–1243.

Parmentier F-JW, Sonnentag O, Mauritz M, et al. (2019) Is the northern permafrost zone a source or a sink for carbon? *EOS* 100.

Pastorello G, Trotta C, Canfora E, et al. (2020) The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Scientific Data* 7(1): 225.

Peltola O, Vesala T, Gao Y, et al. (2019) Monthly gridded data product of northern wetland methane emissions based on upscaling eddy covariance observations. *Earth Syst. Sci. Data* 11: 1263–1289.

Petrone RM., Chasmer L., Hopkinson C., et al. (2014) Effects of harvesting and drought on $CO_2$and H_2o fluxes in an aspen-dominated western boreal plain forest: Early chronosequence recovery. *Canadian Journal of Forest Research* 45(1): 87–100.

Rawlins MA, McGuire AD, Kimball JS, et al. (2015) Assessment of model estimates of land-atmosphere CO_2 exchange across Northern Eurasia. *Biogeosciences* 12(14): 4385–4405.

Reichstein M, Falge E, Baldocchi D, et al. (2005) On the separation of net ecosystem exchange into assimilation and ecosystem respiration: Review and improved algorithm. *Global Change Biology* 11(9): 1424–1439.

Schuur EAG, Bockheim J, Canadell JG, et al. (2008) Vulnerability of Permafrost Carbon to Climate Change: Implications for the Global Carbon Cycle. *BioScience* 58(8): 701.

Schuur EAG, McGuire AD, Schädel C, et al. (2015) Climate change and the permafrost carbon feedback. *Nature* 520(7546): 171–179.

Segurado P and Araújo MB (2004) An evaluation of methods for modelling species distributions. *Journal of Biogeography* 31(10): 1555–1568.

Shaver GR, Street LE, Rastetter EB, et al. (2007) Functional convergence in regulation of net CO_2 flux in heterogeneous tundra landscapes in Alaska and Sweden. *Journal of Ecology* 95(4): 802–817.

Smola AJ and Schölkopf B (2004) A tutorial on support vector regression. *Statistics and Computing* 14: 199–222.

Sørensen MV, Graae BJ, Classen A, et al. (2019) Drivers of C cycling in three arctic-alpine plant communities. *Arctic, Antarctic, and Alpine Research* 51(1): 128–147.

Taylor KE, Stouffer RJ and Meehl GA (2012) An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society* 93(4): 485–498.

Tramontana G, Jung M, Schwalm CR, et al. (2016) Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. *Biogeosciences* 13(14): 4291–4313.

Treat CC, Marushchak ME, Voigt C, et al. (2018) Tundra landscape heterogeneity, not interannual variability, controls the decadal regional carbon balance in the Western Russian Arctic. *Global Change Biology* 24(11): 5188–5204.

Trucco C, Schuur EAGG, Natali SM, et al. (2012) Seven-year trends of CO<inf>2</inf> exchange in a tundra ecosystem affected by long-term permafrost thaw. *Journal of Geophysical Research: Biogeosciences* 117(2): 1–12.

Turetsky MR, Abbott BW, Jones MC, et al. (2020) Carbon release through abrupt permafrost thaw. *Nature Geoscience* 13(2): 138–143.

Ueyama M, Iwata H, Harazono Y, et al. (2013) Growing season and spatial variations of carbon fluxes of Arctic and boreal ecosystems in Alaska (USA). *Ecological Applications*, Ecological Society of America 23(8): 1798–1816.

Ueyama M, Ichii K, Iwata H, et al. (2013) Upscaling terrestrial carbon dioxide fluxes in Alaska with satellite remote sensing and support vector regression. *Journal of Geophysical Research: Biogeosciences* 118(3): 1266–1281.

Ueyama M, Iwata H, Nagano H, et al. (2019) Carbon dioxide balance in early-successional forests after forest fires in interior Alaska. *Agricultural and Forest Meteorology* 275(November 2018): 196–207.

Virkkala A-M, Virtanen T, Lehtonen A, et al. (2018) The current state of CO2 flux chamber studies in the Arctic tundra : A review. *Progress in Physical Geography* 42(2): 162–184.

Virkkala A-M, Abdi AM, Luoto M, et al. (2019) Identifying multidisciplinary research gaps across Arctic terrestrial gradients. *Environmental Research Letters* 14.

Vogel J, Schuur EAGG, Trucco C, et al. (2009) Response of CO2 exchange in a tussock tundra ecosystem to permafrost thaw and thermokarst development. *Journal of Geophysical Research-Biogeosciences* 114(G4): 1–14.

Voigt C, Marushchak ME, Mastepanov M, et al. (2019) Ecosystem carbon response of an Arctic peatland to simulated permafrost thaw. *Global Change Biology* 25(5): 1746–1764.

Walker XJ, Baltzer JL, Cumming SG, et al. (2019) Increasing wildfires threaten historic carbon sink of boreal forest soils. *Nature*, Springer US 572(7770): 520–523.

Wania R, Ross I and Prentice IC (2009) Integrating peatlands and permafrost into a dynamic global vegetation model: 2. Evaluation and sensitivity of vegetation and carbon cycle processes. *Global Biogeochemical Cycles* 23(GB3015).

Warner DL, Bond-Lamberty B, Jian J, et al. (2019) Spatial predictions and associated uncertainty of annual soil respiration at the global scale. *Global Biogeochemical Cycles*.

Watts JD, Kimball JS, Parmentier FJW, et al. (2014) A satellite data driven biophysical modeling approach for estimating northern peatland and tundra CO_2 and CH_4 fluxes. *Biogeosciences* 11(7): 1961–1980.

Welker JM, Fahnestock JT, Henry GHR, et al. (2004) CO_2 exchange in three Canadian High Arctic ecosystems: response to long-term experimental warming. *Global Change Biology* 10(12): 1981–1995.

Wilkman E, Zona D, Tang Y, et al. (2018) Temperature Response of Respiration Across the Heterogeneous Landscape of the Alaskan Arctic Tundra. *Journal of Geophysical Research: Biogeosciences RESEARCH* 123: 2287–2302.

Zhang W, Jansson PE, Schurgers G, et al. (2018) Process-Oriented Modeling of a High Arctic Tundra Ecosystem: Long-Term Carbon Budget and Ecosystem Responses to Interannual Variations of Climate. *Journal of Geophysical Research: Biogeosciences* 123(4): 1178–1196.

Zscheischler J, Mahecha MD, Avitabile V, et al. (2017) Reviews and syntheses : An empirical spatiotemporal description of the global surface – atmosphere carbon fluxes : opportunities and data limitations. *Biogeosciences* 14: 3685–3703.

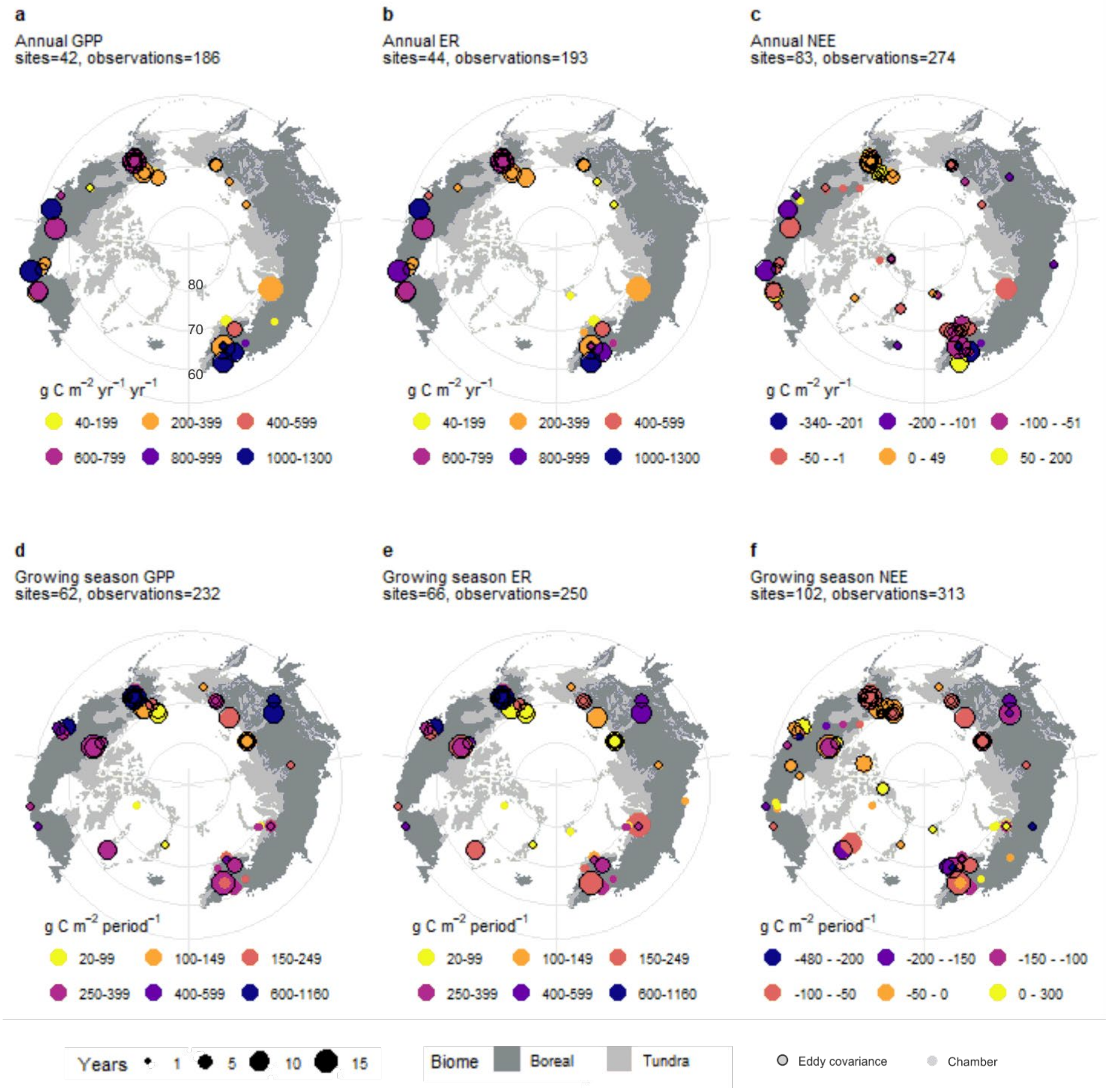

Figure 1. Measured median annual **(a-c)** and growing season **(d-f)** fluxes of GPP (gross primary production), ER (ecosystem respiration), and NEE (net ecosystem exchange) in the study domain (>45 °N). The color of the point defines the median flux of the site (i.e., a sampling location), and the size of the point the number of observations that was measured (i.e., number of years). The background map represents the high-latitude region (dark gray = boreal biome, light gray = tundra biome). In all panels, sites that had only eddy covariance measurements are shown with black outline color around the point, and chamber measurements are without outline. One site had both eddy covariance and chamber measurements, but this is shown with black outline color. Positive numbers for NEE indicate net ecosystem CO_2 loss to the atmosphere (i.e. CO_2 source) and negative numbers indicate net ecosystem CO_2 gain (i.e. CO_2 sink).

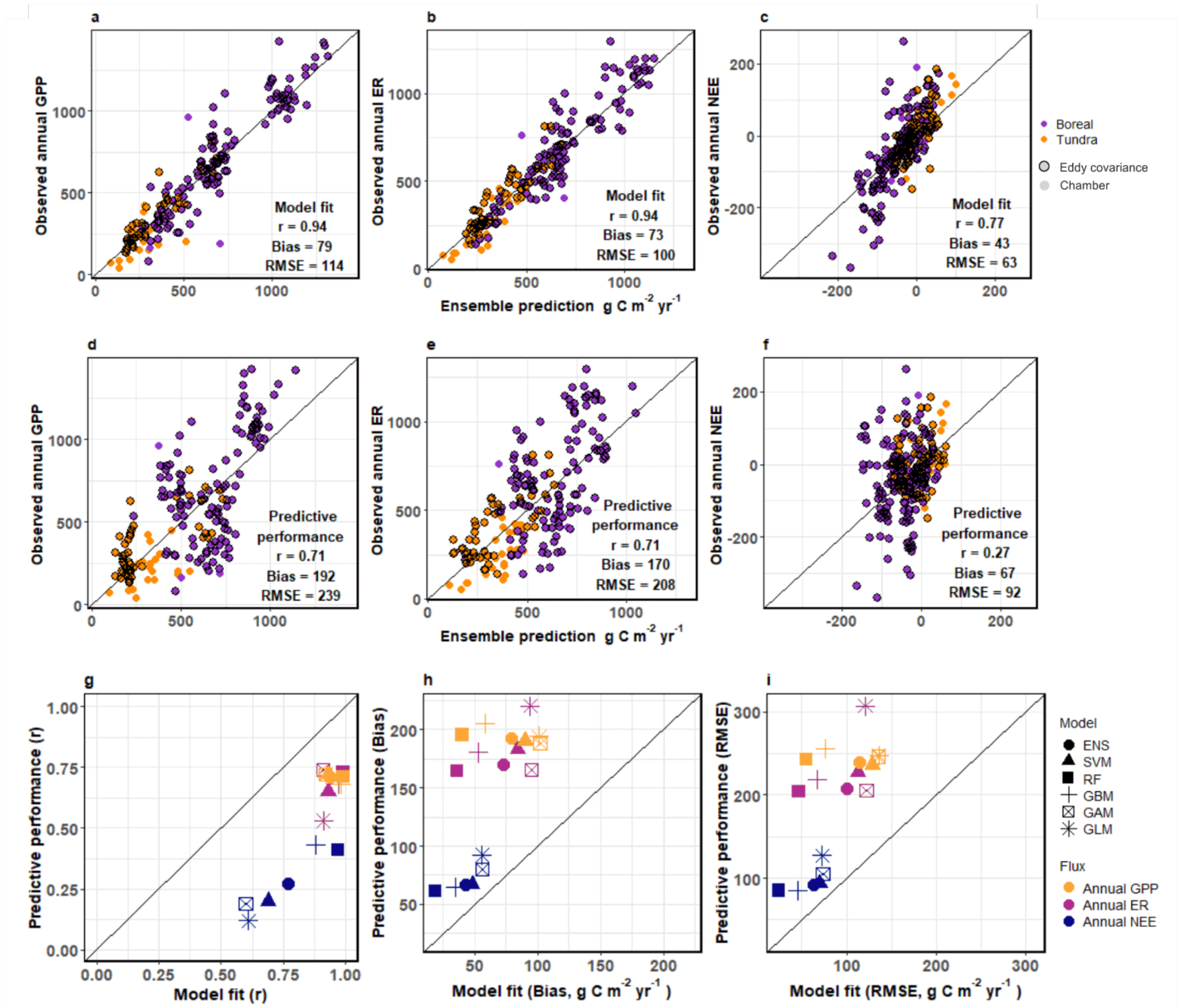

Figure 2. Observed and predicted annual fluxes of GPP (gross primary production; **a & d**), ER (ecosystem respiration; **b & e**), and NEE (net ecosystem exchange; **c & f**) based on model fit (a-c) and predictive performance (d-e). Model fit was evaluated by predicting fluxes over the entire model training data, while predictive performance was assessed using a leave-one-site-out cross validation scheme in which each site was iteratively left out from the data set, and the remaining data were used to predict fluxes for the excluded site. Model fit and predictive performance statistics (r = Pearson's correlation between observed and predicted fluxes, **g**; Bias = mean absolute bias, **h**; RMSE = root mean square error, **i**) across annual fluxes and five modeling methods (GLM = generalized linear model, GAM = generalized additive model, GBM = generalized boosted regression tree, RF = random forest, SVM = support vector machine) and their median ensemble (ENS) are shown in subfigures g-i. The black line indicates a 1:1 relationship.

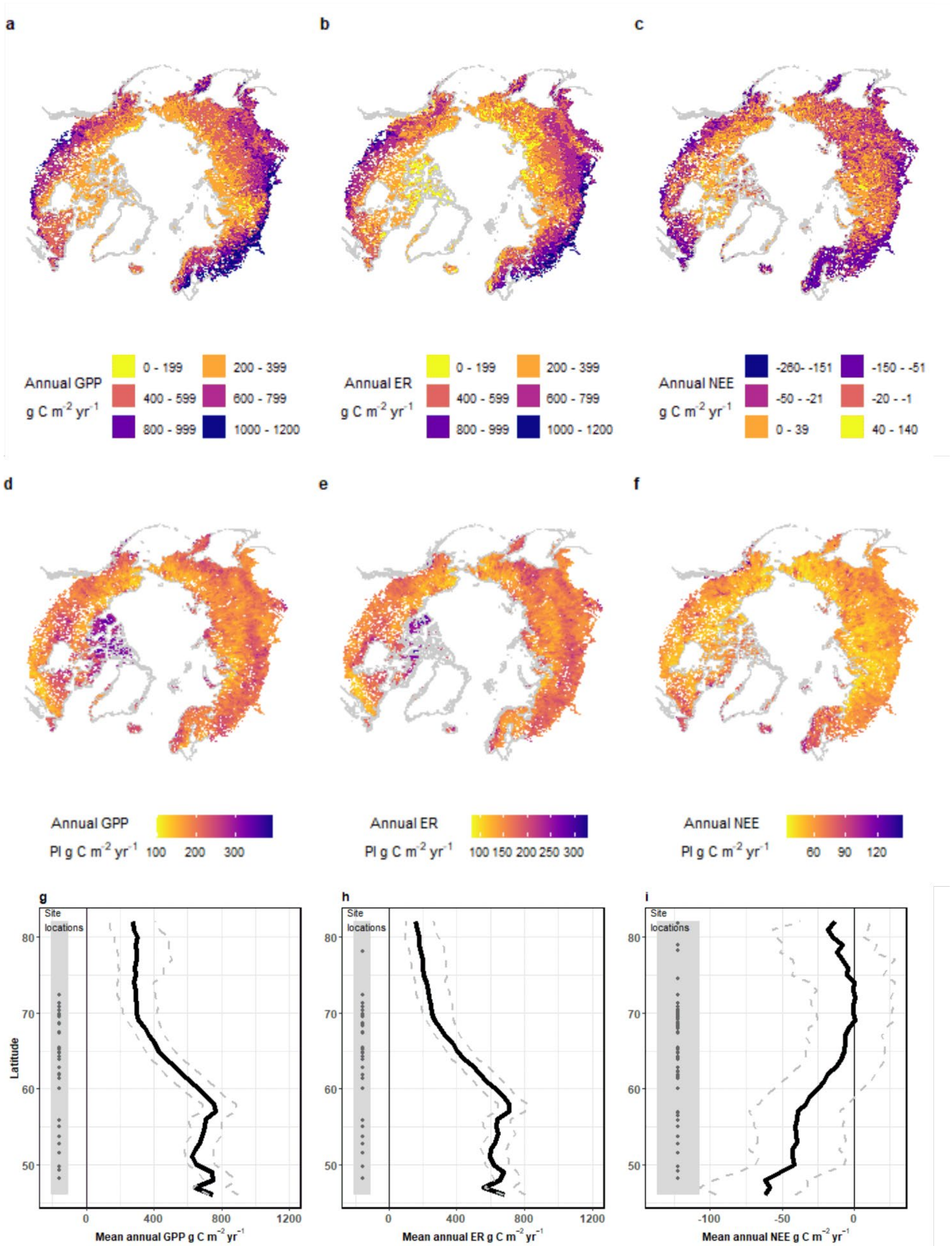

Figure 3. Average predictions of annual CO_2 fluxes at 1 km^2 resolution over 1990–2015. Annual predictions **(a-c)**, associated uncertainties **(d-f)** and mean fluxes and uncertainties along latitudes **(g-i)** of GPP (gross primary production), ER (ecosystem respiration), and NEE (net ecosystem exchange) of the statistical model

ensembles over 1990–2015. The uncertainty (prediction interval, PI; 90 % uncertainty range) is quantified as the variability of predictions over a random subset of pixels (n = 10 000) interpolated across the study domain based on a repeated (n = 200) bootstrap sampling procedure. It demonstrates how robust the relationships in the models are and how differences in model training data influence the predictions. The gray lines in **a-f** represent the borders of northern countries and points in **g-i** site locations.

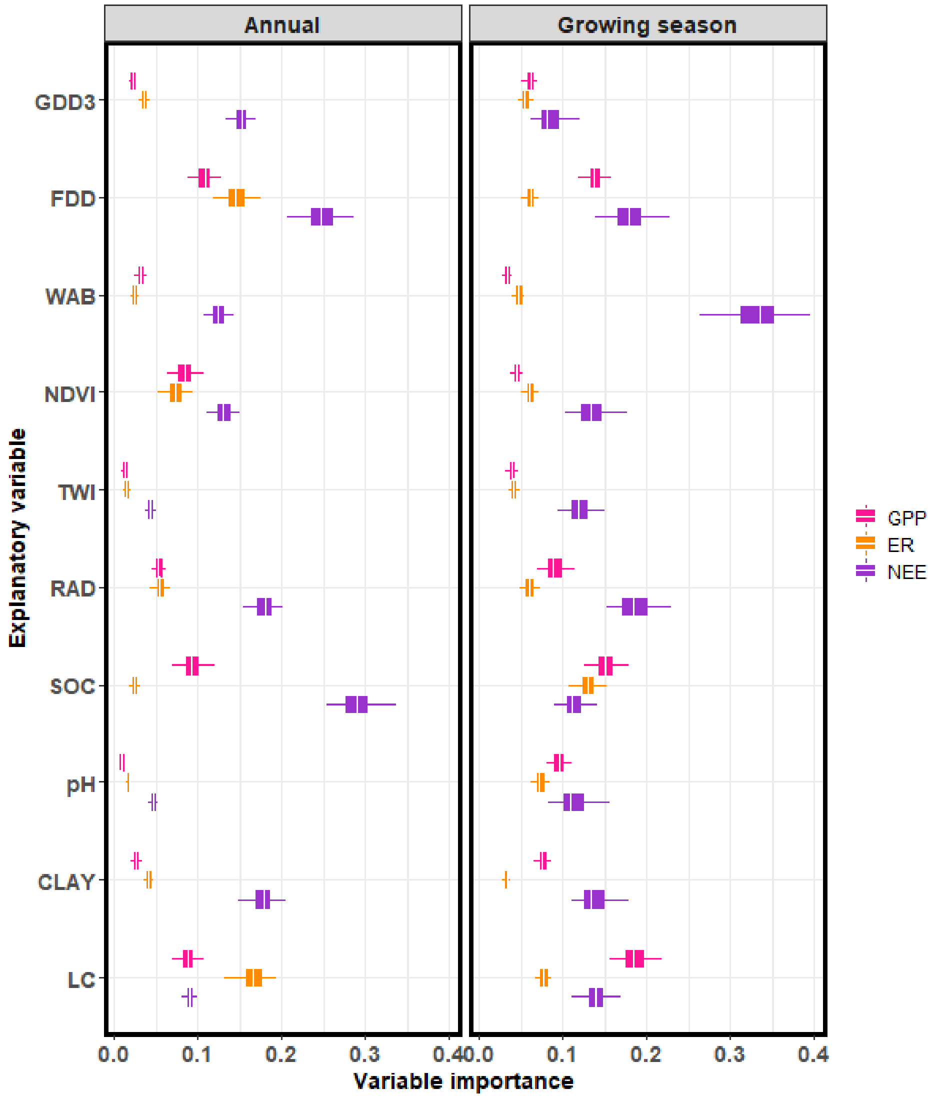

Figure 4. Variable importance for annual and growing season fluxes of GPP (gross primary production), ER (ecosystem respiration), and NEE (net ecosystem exchange). Explanatory variables are GDD3 (growing degree days), FDD (freezing degree days), WAB (water balance), NDVI (normalized difference vegetation index), TWI (topographic wetness index), RAD (potential incoming direct annual radiation), SOC (soil organic carbon stocks up to 2 m), pH (topsoil pH), CLAY (topsoil clay content), and LC (land cover). Variable importance was calculated by assessing how a randomly permuted predictor influences the predictions across all five statistical models. Values close to 0 and 1 indicate low and high importance of the predictor variable, respectively. The box corresponds to the 25th and 75th percentiles. The lines denote the 1.5 IQR of the lower and higher quartile, where IQR is the inter-quartile range, or distance between the first and third quartiles.

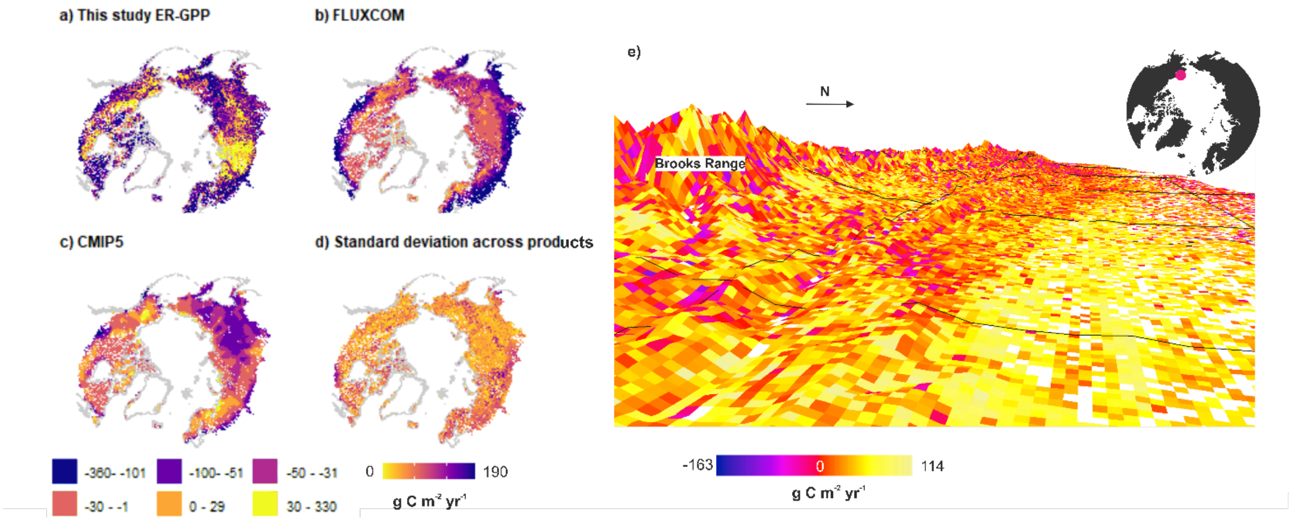

Figure 5. Complementing annual NEE predictions averaged over 1990–2015. Mean annual NEE derived by subtracting annual ER (ecosystem respiration) from GPP (gross primary production) in this study (**a**), from a global upscaling product FLUXCOM (**b**), and from a process model ensemble CMIP5 (Coupled Model Intercomparison Project Phase 5; **c**), and the standard deviation of these and the annual NEE developed in in this study (visualized in Fig. 3c) (**d**). A regional-scale example of the spatial variation of annual NEE in our prediction in northern Alaska, with black outlines depicting the size of the pixel in one of the highest resolution (smallest pixel size) models in the CMIP5 ensemble (1.92 x 1.5 °; **e**).

Table 1. Summary statistics of observed and predicted (using the average ensemble prediction) annual and growing season GPP (gross primary productivity), ER (ecosystem respiration), and NEE (net ecosystem exchange) fluxes (g C m^{-2} yr^{-1} for annual and g C m^{-2} $period^{-1}$ for growing season fluxes) in different environments across the high-latitude region over 1990–2015. Positive numbers for NEE indicate net CO_2 loss to the atmosphere (i.e., CO_2 source) and negative numbers indicate net CO_2 gain (i.e., CO_2 sink). The time-series of the sites were averaged prior calculating the observed mean flux (i.e., one flux value from one site was used when the regional averages were calculated). Note that ER and GPP do not sum up to NEE as different numbers of observations and sites were available for each flux and model. Moreover, some plant uptake occurs outside of our defined growing season, and consequently growing season GPP and annual GPP do not equal to each other. The average fluxes were calculated based on the extent of the high-latitude tundra and boreal biomes (Dinerstein et al., 2017), permafrost zones (Brown et al., 2002), and land cover (i.e. wetlands, and everything else is upland; ESA, 2017). The confidence intervals for the observed fluxes and the uncertainty ranges for the predicted fluxes can be found in the Supplementary Table 5.

Category	Annual GPP	Annual ER	Annual NEE	Growing season GPP	Growing season ER	Growing season NEE
Observed mean flux						
High-latitude	482	456	-17	317	262	-63
Boreal	624	605	-46	420	347	-87
Tundra	250	259	10	232	192	-44
Boreal upland	676	647	-47	432	350	-84
Boreal wetland	406	381	-38	347	330	-102
Tundra upland	250	259	16	232	192	-37
Tundra wetland			-24			-115
No permafrost	831	773	-90	405	370	-37
Permafrost	342	350	6	302	241	-67
Predicted mean flux						
High-latitude	554	508	-20	343	283	-50
Boreal	638	594	-29	396	327	-52
Tundra	378	326	-2	230	192	-46
Boreal upland	653	604	-30	399	328	-51
Boreal wetland	437	458	-18	358	303	-64
Tundra upland	378	326	-1	229	191	-45
Tundra wetland	367	347	-29	281	242	-71
No permafrost	805	736	-56	447	375	-53
Permafrost	489	448	-11	315	259	-49

Table 2. Annual and growing season average GPP, ER, and NEE budgets (Tg C yr^{-1}) over 1990–2015 across the environments and the spatial extent of each environmental category when permanent water bodies, urban areas, and croplands were masked away. The NEE budgets are based on upscaled NEE data and include an uncertainty range derived by bootstrapping. The budgets were calculated based on the extent of the high-latitude tundra and boreal biomes (Dinerstein et al., 2017), permafrost zones (Brown et al., 2002), and land cover (i.e. wetlands, and everything non-wetland is upland; ESA, 2017). Our area estimate of the permafrost region lacks a small permafrost region in the southeastern Asia, which did not belong to the tundra and boreal biomes. For the non-growing season CO_2 budgets estimated based on annual and growing season budgets, see Supplementary Table 4.

Category	**Annual GPP**	**Annual ER**	**Annual NEE**	**Growing season GPP**	**Growing season ER**	**Growing season NEE**	**Area x 10^6 km^2**
High-latitude	11,344	10,397	-419 (-559 - -189)	7,016	5,800	-1,018 (-1,332 – -455)	20.6
Boreal	8,850	8,241	-406 (-499 - -239)	5,496	4,531	-715 (-1,037 – -224)	13.9
Tundra	2,495	2,156	-13 (-81 - 62)	1,520	1,269	-303 (-338 – -224)	6.7
Boreal upland	8,437	7,808	-389 (-475 - -226)	5,158	4,245	-655 (-973 - -196)	12.9
Boreal wetland	412	433	-17 (-28 - -10)	338	287	-60 (-70 - -29)	0.9
Tundra upland	2,451	2,115	-9 (-78 – 66)	1,486	1,240	-294 (-330 - -218)	6.6
Tundra wetland	44	41	-4 (-3 - -1)	34	29	-8 (-9 - -6)	0.1
No permafrost	3,407	3,116	-238 (-288 - -185)	1,895	1,587	-223 (-353 - -45)	4.2
Permafrost	7,924	7,269	-181 (-305 - 32)	5,114	4,207	-793 (-1000 – -414)	16.3