Monthly gridded data product of northern wetland methane emissions based on upscaling eddy covariance observations

Olli Peltola\textsuperscript{1}, Timo Vesala\textsuperscript{2,3}, Yao Gao\textsuperscript{1}, Olle Räty\textsuperscript{4}, Pavel Alekseychik\textsuperscript{5}, Mika Aurela\textsuperscript{1}, Bogdan Chojnicki\textsuperscript{6}, Ankr R. Desai\textsuperscript{7}, Albertus J. Dolman\textsuperscript{8}, Eugenie S. Euskirchen\textsuperscript{9}, Thomas Friborg\textsuperscript{10}, Mathias Göckede\textsuperscript{11}, Manuel Helbig\textsuperscript{12,13}, Elyn Humphreys\textsuperscript{14}, Robert B. Jackson\textsuperscript{15}, Georg Jocher\textsuperscript{16,c}, Fortunat Joos\textsuperscript{17}, Janina Klatt\textsuperscript{18}, Sara H. Knox\textsuperscript{19}, Natalia Kowalska\textsuperscript{a}, Lars Kutzbach\textsuperscript{20}, Sebastian Lienert\textsuperscript{17}, Annela Lohila\textsuperscript{1,2}, Ivan Mammarella\textsuperscript{2}, Daniel F. Nadeau\textsuperscript{21}, Mats B. Nilsson\textsuperscript{16}, Walter C. Oechel\textsuperscript{22,23}, Matthias Peichi\textsuperscript{16}, Thomas Pykker\textsuperscript{24}, William Quinton\textsuperscript{25}, Janne Rinne\textsuperscript{26}, Torsten Sachs\textsuperscript{27}, Mateusz Samson\textsuperscript{6}, Hans Peter Schmid\textsuperscript{18}, Oliver Sonnentag\textsuperscript{13}, Christian Wille\textsuperscript{27}, Donatella Zona\textsuperscript{22,28}, and Tuula Aalto\textsuperscript{1}

\textsuperscript{1}Climate Research Programme, Finnish Meteorological Institute, P.O. Box 503, 00101 Helsinki, Finland
\textsuperscript{2}Institute for Atmosphere and Earth System Research/Physics, Faculty of Science, University of Helsinki, P.O. Box 68, 00014 Helsinki, Finland
\textsuperscript{3}Institute for Atmospheric and Earth System Research/Forest Sciences, Faculty of Agriculture and Forestry, University of Helsinki, P.O. Box 27, 00014 Helsinki, Finland
\textsuperscript{4}Meteorological Research, Finnish Meteorological Institute, P.O. Box 503, 00101 Helsinki, Finland
\textsuperscript{5}Natural Resources Institute Finland (LUKE), 00790 Helsinki, Finland
\textsuperscript{6}Department of Meteorology, Faculty of Environmental Engineering and Spatial Management, Poznań University of Life Sciences, 60-649 Poznań, Poland
\textsuperscript{7}Department of Atmospheric and Oceanic Sciences, University of Wisconsin-Madison, 1225 W Dayton St, Madison, WI 53706, USA
\textsuperscript{8}Department of Earth Sciences, Faculty of Sciences, Vrije Universiteit Amsterdam, Boelelaan 1085, 1081 HV Amsterdam, the Netherlands
\textsuperscript{9}University of Alaska Fairbanks, Institute of Arctic Biology, 2140 Koyukuk Dr., Fairbanks, AK 99775, USA
\textsuperscript{10}Department of Geosciences and Natural Resource Management, University of Copenhagen, Copenhagen, Denmark
\textsuperscript{11}Max Planck Institute for Biogeochemistry, Hans-Knöll-Strasse 10, 07745 Jena, Germany
\textsuperscript{12}School of Geography and Earth Sciences, McMaster University, Hamilton, ON L8S 4K1, Canada
\textsuperscript{13}Département de géographie, Université de Montréal, Montréal, QC H2V 3W8, Canada
\textsuperscript{14}Department of Geography & Environmental Studies, Carleton University, Ottawa, ON K1S 5B6, Canada
\textsuperscript{15}Department of Earth System Science, Woods Institute for the Environment, and Precourt Institute for Energy, Stanford University, Stanford, CA 94305, USA
\textsuperscript{16}Department of Forest Ecology and Management, Swedish University of Agricultural Sciences, Umeå, Sweden
\textsuperscript{17}Climate and Environmental Physics, Physics Institute and Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland
\textsuperscript{18}Institute of Meteorology and Climatology – Atmospheric Environmental Research (IMK-IFU), Karlsruhe Institute of Technology (KIT), Kreuzbeckbahnstrasse 19, 82467 Garmisch-Partenkirchen, Germany
\textsuperscript{19}Department of Geography, The University of British Columbia, Vancouver, BC V6T 1Z2, Canada
\textsuperscript{20}Institute of Soil Science, Center for Earth System Research and Sustainability, Universität Hamburg, Allende-Platz 2, 20146 Hamburg, Germany
\textsuperscript{21}Department of Civil and Water Engineering, Université Laval, Québec, QC G1V 0A6, Canada
\textsuperscript{22}Global Change Research Group, Dept. Biology, San Diego State University, San Diego, CA 92182, USA

Published by Copernicus Publications.
Natural wetlands constitute the largest and most uncertain source of methane (CH$_4$) to the atmosphere and a large fraction of them are found in the northern latitudes. These emissions are typically estimated using process (“bottom-up”) or inversion (“top-down”) models. However, estimates from these two types of models are not independent of each other since the top-down estimates usually rely on the a priori estimation of these emissions obtained with process models. Hence, independent spatially explicit validation data are needed. Here we utilize a random forest (RF) machine-learning technique to upscale CH$_4$ eddy covariance flux measurements from 25 sites to estimate CH$_4$ wetland emissions from the northern latitudes (north of 45°N). Eddy covariance data from 2005 to 2016 are used for model development. The model is then used to predict emissions during 2013 and 2014. The predictive performance of the RF model is evaluated using a leave-one-site-out cross-validation scheme. The performance (Nash–Sutcliffe model efficiency = 0.47) is comparable to previous studies upscaling net ecosystem exchange of carbon dioxide and studies comparing process model output against site-level CH$_4$ emission data. The global distribution of wetlands is one major source of uncertainty for upscaling CH$_4$. Thus, three wetland distribution maps are utilized in the upscaling. Depending on the wetland distribution map, the annual emissions for the northern wetlands yield 32 (22.3–41.2, 95 % confidence interval calculated from a RF model ensemble), 31 (21.4–39.9) or 38 (25.9–49.5) Tg(CH$_4$) yr$^{-1}$. To further evaluate the uncertainties of the upscaled CH$_4$ flux data products we also compared them against output from two process models (LPX-Bern and WetCHARTs), and methodological issues related to CH$_4$ flux upscaling are discussed. The monthly upcaled CH$_4$ flux data products are available at https://doi.org/10.5281/zenodo.2560163 (Peltola et al., 2019).
cess model improvements will also directly affect the uncertainty of inversion results since they provide important a priori information to the inversion models (Bergamaschi et al., 2013).

Boreal and arctic wetlands comprise up to 50% of the total global wetland area (e.g. Lehner and Döll, 2004) and the wetlands in these northern latitudes substantially contribute to total terrestrial wetland CH4 emissions (ca. 27%, based on the sum of regional budgets for boreal North America, Europe and Russia in Saunois et al., 2016). In wetlands, CH4 is produced by methanogenic Archaea under anaerobic conditions, and hence the production takes place predominantly under water-saturated conditions (e.g. Whalen, 2005). The microbial activity and the resulting CH4 production is thus controlled by the quality and quantity of the available substrates, competing electron acceptors, and temperature (Le Mer and Roger, 2001). Once produced, the CH4 can be emitted to the atmosphere via three pathways: ebullition, molecular diffusion through soil matrix and water column, or plant transport. If plants capable of transporting CH4 are present, plant transport is generally the dominating emission pathway (Knoblauch et al., 2015; Kwon et al., 2017; Waddington et al., 1996; Whiting and Chanton, 1992). A large fraction of CH4 transported via molecular diffusion is oxidized into CO2 by methanotrophic bacteria in the aerobic layers of wetland soils and hence never reaches the atmosphere (Sundh et al., 1995), whereas CH4 transported via ebullition and plant transport can largely bypass oxidation (Le Mer and Roger, 2001; McEwing et al., 2015). Furthermore, processes related to permafrost (e.g. active layer, thermokarst) and snow cover dynamics (e.g. snow melt, insulation) have an impact on CH4 flux seasonality and variability (Friberg et al., 1997; Helbig et al., 2017; Mastepanov et al., 2008; Zona et al., 2016; Zhao et al., 2016). Hence, wetland CH4 emissions to the atmosphere largely depend on interplay between various controls, including water table position, temperature, vegetation composition, methane consumption, availability of substrates and competing electron acceptors.

During the past 2 decades, eddy covariance (EC) measurements of wetland CH4 emissions have become more common, due to rapid development in sensor technology (e.g. Detto et al., 2011; Peltola et al., 2013, 2014). The latest generation of low-power and low-maintenance instruments are rugged enough for long-term field deployment (Nemitz et al., 2018; McDermitt et al., 2010); thus, the number of sites where CH4 flux measurements have been made is increasing. Due to this progress, EC CH4 flux synthesis studies have been emerging (Petrescu et al., 2015; Knox et al., 2019). Similar progress was made with CO2 and energy flux measurements in the 1990s and now these measurements form the backbone of the global EC measurement network FLUXNET (https://fluxnet.fluxdata.org/, last access: 6 August 2019), whose data have provided invaluable insights into terrestrial carbon and water cycles. Some of the most important results have been obtained by upscaling FLUXNET observations using machine-learning algorithms to evaluate terrestrial carbon balance components and evapotranspiration (Beer et al., 2010; Bodesheim et al., 2018; Jung et al., 2010, 2011, 2017; Mahecha et al., 2010). These results are now widely used by the modelling community to evaluate process model performance (e.g. Wu et al., 2017) and to validate satellite-derived carbon cycle data products (e.g. Sun et al., 2017; Y. Zhang et al., 2017).

In this study, we synthesized EC CH4 flux data from 25 EC CH4 flux sites and developed an observation-based monthly gridded data product of northern wetland CH4 emissions. We focus on northern wetlands (north of 45°N) due to their significance in the global CH4 budget and relatively good data coverage and process understanding, at least compared to tropical systems (Knox et al., 2019). High-latitude regions are projected to warm during the next century at a faster rate than any other region, which will likely significantly impact the carbon cycling of wetland ecosystems (Tarnocai, 2009; Z. Zhang et al., 2017) and permafrost areas of the Arctic boreal region (Schuur et al., 2015). To date, CH4 emission estimates for northern wetlands are typically based on process models (Bohn et al., 2015; Bloom et al., 2017a; Chen et al., 2015; Melton et al., 2013; Stocker et al., 2013; Wania et al., 2010; Watts et al., 2014; Zhang et al., 2016) or inversion modelling (Bohn et al., 2015; Bruhwiler et al., 2014; Spahni et al., 2011; Thompson et al., 2017; Thonat et al., 2017; Warwick et al., 2016), yet scaling of existing chamber measurements to the northern wetland area has also been published (Zhu et al., 2013). However, CH4 emission estimates obtained with the former two approaches are not independent since the attribution of CH4 emissions derived using inversion models to different emission sources (e.g. wetlands) depends largely on a priori estimates of these emissions (i.e. process models for wetland emissions), highlighting the tight coupling between these two approaches (Bergamaschi et al., 2013; Spahni et al., 2011). Hence, the main objective of this study is to produce an independent data-driven estimate of northern wetland CH4 emissions. This product could be used as an additional constraint for the wetland emissions and hence aid in process model refinement and development. Additionally, the drivers causing CH4 flux variability at the ecosystem scale are also evaluated and methodological issues are discussed, which will support future CH4 wetland flux upscaling studies.

2 Materials and methods

Data from flux measurement sites (Fig. 1) were acquired and used together with forcing data to estimate CH4 emissions from northern wetlands with a monthly time resolution using a random forest (RF) modelling approach. Both in situ measurements and remote sensing are utilized in this study. In this section, the RF approach is briefly introduced (Sect. 2.1)
Random forest (RF) is a machine-learning algorithm that can be used for classification or regression analyses (Breiman, 2001). In this study the RF models consist of a large ensemble of regression trees. Each individual regression tree is built by training it with a random subset of training data and the trees are trained independently of each other. The RF model output is then the average of all the predictions made by individual regression trees in the forest. Hence, the RF algorithm applies the bootstrap aggregation (bagging) algorithm and takes full advantage of the fact that ensemble averaging decreases the noise of the prediction. In addition to random selection of training data, the predictor variables used in split nodes are also selected from a random sample of all predictors, which minimizes the possible correlation between trees in the forest (Breiman, 2001) and decreases the possibility of overfitting. The predictor variables can be either categorical or continuous. The variables are then used in the split nodes to divide the data into two (e.g. categorical variable true or false or a continuous variable, such as temperature above or below 5 °C).

Performance of RF algorithms to predict CO$_2$ and energy fluxes across FLUXNET sites have been compared against other machine-learning algorithms, such as artificial neural networks and multivariate regression splines, by Tramontana et al. (2016), who showed that differences between methods were negligible. We anticipate a similarly negligible effect of machine-learning algorithm choice for CH$_4$ fluxes. For a thorough description of the RF algorithm for flux upscaling purposes, the reader is referred to Bodesheim et al. (2018) (and references therein).

In this study, the RF models were developed using the MATLAB 9.4.0 (R2018a) TreeBagger function with default values similar to those of Bodesheim et al. (2018). These settings included a minimum of five samples in a leaf node and used mean squared error (MSE) as a metric for deciding the split criterion in split nodes. Each trained forest consisted of 300 randomized regression trees.

### 2.1 Random forest algorithm

Random forest (RF) is a machine-learning algorithm that can be used for classification or regression analyses (Breiman, 2001). In this study the RF models consist of a large ensemble of regression trees. Each individual regression tree is built by training it with a random subset of training data and the

---

**Figure 1.** Map showing the locations of the EC measurements. The distribution of wetlands shown in the figure is based on Xu et al. (2018). Hudson Bay lowlands (50–60° N, 75–96° W) and western Siberian lowlands (52–74° N, 60–94.5° E) are highlighted with dashed red lines.
2.1.2 RF model development for CH₄ flux upscaling

For upscaling purposes, one RF model was developed using all the available data in order to maximize the information content for the global (> 45° N) CH₄ flux map. The model performance or uncertainty, however, was evaluated using two approaches. (1) The predictive performance of the model was assessed using the widely used “leave-one-site-out” cross-validation scheme (e.g., Jung et al., 2011). In order to avoid correlation between training data and validation data, sites located nearby (closer than 100 km) were excluded from the training data (Roberts et al., 2016). (2) The uncertainty of the upscaled fluxes was estimated by bootstrapping. The 200 independent RF models were trained using a bootstrap sample of the available data. This yielded 200 predictions for each grid cell and time step in the upscaled CH₄ flux map. The variability over this prediction ensemble was used as an uncertainty measure following, e.g., Aalto et al. (2018) and Zhu et al. (2013). This uncertainty estimate reflects the ability of the RF model to capture the dependence of CH₄ flux on the used predictors in the available data. However, it does not have any reference to actual in situ CH₄ fluxes, unlike the model predictive performance estimated with cross-validation.

Predictors for the RF model used in the upscaling were determined following Moffat et al. (2010). First, the RF models were trained for each site using one predictor at a time (see all the predictors in Table 1). The single predictor which yielded the best match against validation data (leave-one-site-out scheme) was selected as the primary driver. Then, the RF models were trained again with the primary driver, plus each of the other predictors in turn as secondary drivers. Then the RF model performance was again evaluated and the best predictor pair was selected for the next round. This procedure was continued until all the predictors were included in the RF model. The smallest set of predictors capable of producing optimal RF model performance was used for flux upscaling.

2.2 Metrics for model performance evaluation

The RF model performance was evaluated against independent validation data using a set of statistical metrics, which were related to different aspects of model performance. During the RF model training MSE was optimized as follows:

\[ \text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (p_i - o_i)^2, \]  \hspace{1cm} (1)

where \( o \) and \( p \) are vectors containing the observed and predicted values, respectively, and the overbar denotes averaging.

The Nash–Sutcliffe model efficiency (NSE; Nash and Sutcliffe, 1970) was used to evaluate how well the model was able to predict validation data when compared against a reference (typically the mean of the validation data):

\[ \text{NSE} = 1 - \frac{\sum_{i=1}^{n} (o_i - p_i)^2}{\sum_{i=1}^{n} (o_i - \bar{o})^2}, \]  \hspace{1cm} (2)

where \( i \) is index running over all the \( n \) values in the \( o \) and \( p \) vectors. When NSE is equal to 1, there is a perfect match between prediction and observations. Values above 0 imply that the model predicts the observations better than the mean of observations and values below 0 indicate that the predictive capacity of the model is worse than the mean of validation data. Note that NSE calculated with Eq. (2) above is equivalent to the coefficient of determination calculated using residual sum of squares and total sum of squares. However, following the approach used in previous upscaling studies (e.g., Bodesheim et al., 2018; Tramontana et al., 2016), we opted to call this metric NSE. Instead, the coefficient of determination \((R^2)\) was estimated as the squared Pearson correlation coefficient. Note that \( R^2 \) and NSE are equal when there is no bias between \( o \) and \( p \) and the residuals follow Gaussian distribution. Pearson correlation coefficients obtained with different model runs are compared using Fisher’s \( r \) to \( z \) transformation.

The standard deviation (\( \sigma \)) of the model residuals was used to evaluate the spread of model residual values (RE):

\[ \text{RE} = \sigma (o - p), \]  \hspace{1cm} (3)

whereas biases between model predictions and validation data were used to estimate the systematic uncertainty in the upscaled fluxes (BE):

\[ \text{BE} = o - p. \]  \hspace{1cm} (4)

Note that RE equals RMSE when there is no systematic difference between the model predictions and observations (i.e. when BE equals zero).

2.3 Data

2.3.1 Data from eddy covariance flux measurement sites

Data were acquired from 25 sites that (1) measure CH₄ fluxes with the EC technique, (2) are located north of 45° N and (3) are wetlands as defined above and without substantial human influence on ecosystem functioning (see the site locations in Fig. 1 and the site list in Appendix A). The sites were evenly distributed among wetland types, including fens \((n = 9)\), bogs \((7)\), and wet tundra \((9)\), as well as among different biomes, including tundra \((11)\), boreal \((8)\), and temperate \((6)\) biomes, as defined in Olson et al. (2001). At 15 of the 25 sites, sedges (e.g., *Rhynchospora alba*, *Eriophorum vaginatum*, *Carex limosa*) were the dominant vascular plant functional type in the flux measurement source area. Most of the sites \((18 \text{ out of } 25)\) were located north of 60° N and the highest densities of sites were in Fennoscandia and Alaska.
The magnitude of monthly CH$_4$ flux data varied between sites and the median time series length was 14.5 months of CH$_4$ flux data per site. Overall, the dataset spanned between 2005 and 2016. The sites represent northern wetlands sufficiently well to create an upscaled CH$_4$ flux product based on EC data. Sites are referred to with their FLUXNET IDs and if these were not available then new temporary site IDs were generated for this study (see Appendix A).

Site principal investigators (PIs) provided CH$_4$ fluxes and their potential drivers (air temperature and pressure, precipitation, wind speed and direction, friction velocity, net ecosystem exchange of CO$_2$ and its components – i.e. canopy photosynthesis and ecosystem respiration – photosynthetically active radiation, water table depth, and soil temperature) . However, out of the in situ measurements only air temperature and precipitation were used for developing the RF model for flux upscaling since gridded data products of the other potentially important drivers were not readily available and/or the data for the other drivers were missing from several sites.

The 30 min averaged flux data were acquired from 21 sites and daily data were provided for 4 sites. The flux time series were quality filtered by removing fluxes with the worst quality flag (based on 0,1,2 flagging scheme, Mauder et al., 2013) and with friction velocity below a site-specific threshold (if friction velocity and threshold were available for the site). After filtering, daily medians were calculated if the daily data coverage was above 60% and days with lower data coverage, no daily values calculated. Shorter gaps (< 2 h) were filled with linear interpolation, whereas longer gaps (between 2 to 14.5 h) were replaced with mean diurnal variation within a 30 d moving window. However, for precipitation, daily sums were calculated without any gap filling. Besides the measurements at the sites, potential solar radiation ($R_{pot}$) and its time derivative (der($R_{pot}$)) were calculated based on latitude and time of measurement. In order to remove the $R_{pot}$ latitudinal dependence it was normalized to be between 0 and 1 before usage.

CH$_4$ flux drivers measured in situ, in addition to the remote-sensing data (Sect. 2.3.2), were used for the gap filling of CH$_4$ time series with the RF algorithm (Sect. 2.1.1). For each site the gap-filling models generally agreed well with the independent validation data (mean NSE $= 0.74$ and mean RMSE $= 9$ nmol m$^{-2}$ s$^{-1}$). After gap filling, the CH$_4$ flux time series were aggregated to monthly values if the monthly data coverage prior to gap filling was at least 20%.

The daily time series of air temperature and precipitation were used for developing the RF model for flux upscaling of CH$_4$ flux data at each site. However, at sites with Phragmites cover a relatively strong diurnal cycle can be observed (e.g. Kim et al., 1999; Kowalska et al., 2013).

Unlike the CH$_4$ flux data, the other in situ data from the sites were gap filled prior to the calculation of daily values. The gap filling was done only if the daily data coverage was above 60% and days with lower data coverage, no daily values calculated. Shorter gaps (< 2 h) were filled with linear interpolation, whereas longer gaps (between 2 to 14.5 h) were replaced with mean diurnal variation within a 30 d moving window. However, for precipitation, daily sums were calculated without any gap filling. Besides the measurements at the sites, potential solar radiation ($R_{pot}$) and its time derivative (der($R_{pot}$)) were calculated based on latitude and time of measurement. In order to remove the $R_{pot}$ latitudinal dependence it was normalized to be between 0 and 1 before usage.
(e.g. Räisänen and Räty, 2013; Räty et al., 2014). For precipitation, the mean of WFDEI data were simply adjusted to match site mean precipitation. For air temperature, the bias correction was done for each month separately using quantile mapping with smoothing within a moving 7-month window. Quantile mapping compares the cumulative distribution functions (CDFs) of WFDEI and site measurements against each other and adjusts the WFDEI data so that after adjustment its CDF matches with the CDF of the site measurements (e.g. Räisänen and Räty, 2013). After gap filling the daily time series with WFDEI data, monthly and annual precipitation were calculated, in addition to monthly mean air temperature.

2.3.2 Remote-sensing data

Several data products from the Moderate Resolution Imaging Spectrometer (MODIS) were used in this study to derive various driving variables. For RF model development the following data products at 500 or 1000 m spatial resolution were used: MOD10A1 snow cover (Hall and Rigs, 2016), MOD11A2 daytime and night-time land surface temperature (LSTd and LSTn, Wan et al., 2015a), MOD13A3 enhanced vegetation index (EV1, Didan, 2015), and MOD09A1 surface reflectance (Vermote, 2015). More elaborate data products estimating ecosystem gross primary productivity (GPP) and net primary productivity (NPP; MOD17) were not included here for two reasons: (1) many of the sites included here were misclassified in the land cover map used in MOD17 (e.g. as woody savanna), hence severely influencing the estimated GPP and NPP (Zhao et al., 2005), and (2) sites that were correctly classified as permanent wetlands were in fact assigned a fill value and removed from the product since the product is not strictly valid for these areas (Lees et al., 2018). All the remote-sensing data products were quality filtered using the quality flags provided along with the data.

The MODIS snow cover ranged from 0 (no snow) to 100 (full snow cover) and was converted to a simple snow cover flag (SC) consisting of 0 and 1 depending whether the snow cover data were below or above 50, respectively. A vector containing days since snow melt (DSSM) was calculated using the snow cover flag and normalized to 0 (beginning) and 1 (end) for each growing season (Mastepanov et al., 2013). The MOD09A1 surface reflectance at bands 2 (841–876 nm) and 5 (1230–1250 nm) were used to calculate the simple ratio water index (SRWI = band 2/band 5), following Zarco-Tejada and Ustin (2001). SRWI showed spurious values when there was snow cover, and hence these points were replaced with the mean SRWI observed at each site when there was no snow. Meingast et al. (2014) showed that SRWI can be used as a proxy for wetland water table depth, although their results were affected by changes in vegetation cover, which might hinder across-site comparability in this study. Additionally, following the temperature and greenness modelling approach (Sims et al., 2008), a product of EVI and LSTd was included in the analysis as a proxy for GPP, following a previous peatland study (Schubert et al., 2010). The remote-sensing data were provided with daily (MOD10A1), 8 d (MOD09A1, MOD11A2) or monthly (MOD13A3) time resolution and the data were aggregated to monthly means prior to usage.

2.3.3 Additional categorical variables

The sites were also classified based on the presence of permafrost in the source area (present or absent) and according to biome type. Biome types (temperate, boreal and tundra) were determined from Olson et al. (2001) and the information about the permafrost was provided by the site PIs. Furthermore, the data were categorized based on wetland type and sedge cover as in Treat et al. (2018) and Turetsky et al. (2014). However, such information is not available in the gridded format needed for upscaling; nevertheless, inclusion of these variables can be used to assess how much they increase the predictive performance of the model.

2.4 Gridded datasets used in flux upscaling

For upscaling CH4 fluxes using the developed RF model, the LST data were acquired from the aggregated product MOD11C3 (Wan et al., 2015b) and snow cover data from MOD10CM (Hall and Riggs, 2018). Distribution of permafrost in the northern latitudes were estimated using the circum-Arctic map of permafrost derived by National Snow and Ice Data Center (Brown et al., 2002). The resolution of the gridded data was adjusted to match the resolution of the wetland maps using bilinear interpolation if needed. Additionally, land and ocean masks (Jet Propulsion Laboratory, 2013) were utilized when processing the gridded datasets.

2.5 Wetland maps

Upscaled fluxes were initially estimated in flux densities per wetland area, i.e. amount of CH4 per area of wetland per unit of time. To create a gridded product of CH4 emissions from the northern wetlands, these upscaled flux densities were converted into (amount of CH4 per (grid cell area) per (unit of time) using different wetland maps. Wetland mapping is an ongoing field of research and the usage of different wetland maps contributes to the uncertainty of global wetland CH4 emission estimates (e.g. Bloom et al., 2017a; Z. Zhang et al., 2017). Hence, three different wetland maps (PEATMAP, DYPTOP and GLWD) were used in this study to evaluate how much they affect the overall estimates of northern high-latitude wetland CH4 emissions.

The recently developed static wetland map PEATMAP (Xu et al., 2018) combines detailed geospatial information from various sources to produce a global map of wetland extent. Here, the polygons in PEATMAP were converted to fractions of wetland in 0.5° grid cells. While PEATMAP is
focused on mapping peatlands, marshes and swamps (typically on mineral soil) are included in the product for certain areas in the northern latitudes. However, most of the wetlands in the northern latitudes are peatlands, and thus PEATMAP is suitable for our upscaling purposes. The dynamic wetland map estimated by the DYPTOP model (Stocker et al., 2014) was used by aggregating peat and inundated areas to form one dynamic wetland map with 1° resolution. The widely used Global Lakes and Wetlands Database (GLWD, Lehner and Döll, 2004) is a static wetland map with 30 arcsec resolution and since it has been widely used here it provided a point of reference for the other two maps. The map was aggregated to 0.5° resolution and lakes, reservoirs and rivers were excluded from the aggregated map.

2.6 Process models

The upscaled CH\textsubscript{4} fluxes were compared against the output from two process models: LPX-Bern (Spahni et al., 2013; Stocker et al., 2013; Zürcher et al., 2013) and the model ensemble WetCHARTs version 1.0 (Bloom et al., 2017a, 2017b). LPX-Bern is a dynamic global vegetation model that models carbon and nitrogen cycling in terrestrial ecosystems. The model has a separate peatland module with peatland-specific plant functional types (for more details, see Spahni et al., 2013). The wetland extent in LPX-Bern was dynamically estimated using the DYPTOP approach with 1° resolution (Stocker et al., 2014). WetCHARTs combines several prescribed wetland maps with different gridded products for heterotrophic respiration and temperature sensitivity (\(Q_{10}\)) parameterizations for CH\textsubscript{4} production to form a model ensemble of wetland CH\textsubscript{4} emissions (Bloom et al., 2017b). Here we used the extended ensemble of WetCHARTs.

3 Results

3.1 Selecting the predictors for the RF model

The predictors in Table 1 were selected one by one using the procedure described in Sect. 2.1.2. The order in which the predictors were selected is shown in Fig. 2. LST\textsubscript{n} alone gave NSE = 0.29. After including the category permafrost presence and absence, \(R_{\text{pot}}\), SC and biome class increased NSE to 0.47. However, the influence of SC and biome class on the model performance was marginal based on the small increase in NSE. Additional predictors did not increase the model performance further because (1) they were strongly correlated with a predictor already included in the model (e.g. \(T_{\text{air}}\) is correlated with LST\textsubscript{n}) or (2) the predictors did not contain any information about CH\textsubscript{4} flux variability. The model response to predictors other than biome category was physically reasonable (e.g. permafrost and snow cover decrease fluxes, close to exponential dependence on LST\textsubscript{n}), whereas the response to biome category was contrary to expectations. The RF model estimated the CH\textsubscript{4} flux magnitude from the different biomes to be in the following order: tundra < temperate < boreal. However, in prior studies it has been shown to be in the following order: tundra < boreal < temperate (Knox et al., 2019; Treat et al., 2018; Turetsky et al., 2014). This discrepancy may be due to the limited number of measurement sites and related sampling bias problems. Hence, in order not to upscale an incorrect pattern of decreasing CH\textsubscript{4} emissions when moving from boreal to temperate regions, the biome class was omitted from upscaling. In the subsequent analysis and flux upscaling only the four first predictors (LST\textsubscript{n}, permafrost category, \(R_{\text{pot}}\) and SC) are utilized.

We further tested whether information about wetland type or sedge cover would improve the model performance even though these categorical variables were not available in gridded format and hence were not usable for upscaling. Including the sedge flag increased the NSE to 0.53, although the increase in Pearson correlation was not statistically significant (\(p > 0.05\), comparison of correlation coefficients using Fisher’s \(r\) to \(z\) transformation). Also, wetland type did not have a statistically significant influence on the model performance (\(p > 0.05\) and NSE = 0.49 if type included). Using
too many categorical variables in a RF model may be problematic because each site may end up with a unique combination of categorical variables.

The most important predictor for the model was temperature, similar to numerous studies showing that wetland CH\textsubscript{4} emissions are strongly correlated to soil temperature (Christensen et al., 2003; Helbig et al., 2017; Jackowicz-Korczyński et al., 2010; Rinne et al., 2018; Yvon-Durocher et al., 2003; Helbig et al., 2017; Jackowicz-Korczyński et al., 2010; Rinne et al., 2018; Yvon-Durocher et al., 2014; Knox et al., 2019). Selection of LSTn as the primary driver instead of the other temperature variables was likely an outcome of the available data and the algorithm used to select the drivers. With slightly different dataset (more sites) other temperature variables (e.g. $T_{air}$) might have been more important drivers for the CH\textsubscript{4} flux variability. Estimating apparent $Q_{10}$ from the RF model LSTn dependence yielded a value of 1.90 ± 0.03 and for validation data it was slightly higher (1.97 ± 0.06) (Fig. 3). These values are comparable to the ones reported in Turetsky et al. (2014) for CH\textsubscript{4} chamber measurements at bog and fen sites. The temperature dependence of CH\textsubscript{4} production is modelled in many process models with the parameter $Q_{10}$ value close to 2 (Xu et al., 2016b), which agrees with the CH\textsubscript{4} emission temperature dependence shown here. However, one should note that CH\textsubscript{4} oxidation also depends on temperature and the derived apparent $Q_{10}$ value describes the temperature dependence of surface CH\textsubscript{4} emission, which is always a combination of CH\textsubscript{4} production and oxidation.

3.2 Model agreement with validation data

The overall systematic bias (BE) between the RF predictions and validation data was negligible (Fig. 4), whereas the spread of the data (RE) was more pronounced (Fig. 4). Following Moffat et al. (2010), RE was analysed further by binning the data based on CH\textsubscript{4} flux magnitude and calculating RE for each bin. RE was clearly correlated with flux magnitude (RE = $(0.52 ± 0.06)F_{CH4} + (3.3 ± 2.0)$ nmol m$^{-2}$ s$^{-1}$, where $F_{CH4}$ denotes CH\textsubscript{4} flux), indicating that the random error of the RF model prediction was nearly constant and approximately 50 % for high fluxes. The systematic error BE did not show a clear dependence on flux magnitude. The RF model performance was worse on a site mean level than with monthly data. When comparing site means, NSE and $R^2$ were both 0.25 and RE and BE were 27.0 and 1.5 nmol m$^{-2}$ s$^{-1}$, respectively. Possible drivers causing the remaining CH\textsubscript{4} flux variability not captured by the RF model (i.e. the scatter in Fig. 4) are discussed in Sect. 4.2.1.

When considering the model performance for each site separately, the agreement shows different characteristics (see Fig. 5 for four examples). For individual sites the magnitude of BE is typically somewhat higher (median of absolute value of BE approximately 11 nmol m$^{-2}$ s$^{-1}$), whereas RE is lower than for the overall agreement (median RE approximately 10 nmol m$^{-2}$ s$^{-1}$). These results indicate that the upscaled CH\textsubscript{4} fluxes have, in general, relatively low bias and high random error, whereas individual pixels in the upscaled CH\textsubscript{4} map may have higher bias but lower random error.

The mean annual cycle of CH\textsubscript{4} emission predicted by the RF model agrees well with the mean annual cycle cal-
culated from the validation data (not shown). During the non-growing season the RF model slightly overestimates the fluxes (15 % overestimation) but such differences were negligible during the rest of the year (< 1 %). However, for individual sites CH$_4$ emission seasonality agrees less. For instance, at US-Los the modelled CH$_4$ emissions start to increase 1 month earlier in the spring (Fig. 5b). The nongrowing season fluxes are overestimated at four example sites (FI-Sii, US-Los, US-Atq and RU-Ch2; Fig. 5). The mean flux magnitude is modelled well at FI-Sii (Fig. 5a), whereas at US-Los (Fig. 5b) and US-Atq (Fig. 5c) the RF model overestimates and at RU-Ch2 (Fig. 5d) underestimates the CH$_4$ emissions. The flux bias had a relatively large impact on site-specific NSE. For example, for US-Atq NSE was −1.85, meaning that the observation mean would be a better predictor for this site than the RF model (see the NSE definition in Sect. 2.2). The RF model is not able to replicate the between-year differences in CH$_4$ emissions at the example sites. Capturing interannual variability has also been difficult in previous upscaling studies of CO$_2$ and energy fluxes (e.g. Tramontana et al., 2016).

In general, the RF model performance was better for permafrost-free sites than for sites with permafrost ($r = 0.66$ and $r = 0.51$, respectively; $p < 0.05$), which is likely related to the fact that at sites with permafrost the MODIS LSTn is not as directly related to the soil temperature than at sites without permafrost. Hence, LSTn is not as good proxy for the temperature that is controlling both CH$_4$ production and consumption and this results in a worse performance than at sites without permafrost.

### 3.3 Upscaled CH$_4$ fluxes

The RF model developed in this study was used together with the gridded input datasets (Sect. 2.4) and wetland distribution maps (Sect. 2.5) to estimate CH$_4$ emissions from northern wetlands in 2013 and 2014. The mean CH$_4$ emissions of the 2 years from the RF model are plotted in Fig. 6 together with CH$_4$ wetland emission maps from the process model LPX-Bern and model ensemble WetCHARTs. Differences between the process model estimations and upscaled fluxes are shown in Fig. 7. In general, the spatial patterns are similar among emission maps, which is not surprising given that the spatial variability is largely controlled by the underlying wetland distributions. One noteworthy difference is that WetCHARTs, RF-PEATMAP (i.e. RF modelling with
PEATMAP) and RF-GLWD show higher emissions from western Canada than LPX-Bern or the upscaled fluxes using the wetland map from that process model (RF-DYPTOP). The other difference is that RF-GLWD show negligible emissions from Fennoscandia (Fig. 6c). These differences are related to differences in the underlying wetland maps. While the wetland maps differ, there is no consensus on which is more accurate, so comparisons indicate the uncertainty in upscaling emanating from uncertainties in wetland distribution.

Three statistical metrics (NSE, $R^2$ and RE) were calculated between RF-DYPTOP and LPX-Bern for each grid cell (Fig. 8). The figure illustrates how well the temporal variability of CH$_4$ emissions estimated by RF-DYPTOP and LPX-Bern agree in each grid cell. NSE values are low in areas where the systematic difference between RF-DYPTOP and LPX-Bern was high (compare Figs. 8a and 7a) since the bias strongly penalizes NSE. The $R^2$ values are high throughout the study domain, likely due to the fact that the seasonal cycle of CH$_4$ emissions dominated the temporal variability in most of the grid cells and the seasonal cycles were in phase between RF-DYPTOP and LPX-Bern. RE values calculated between RF-DYPTOP and LPX-Bern were high in areas where the emissions estimated by RF-DYPTOP were also high (compare Figs. 8c and 6a). This is likely due to the fact that, even though the seasonal cycles were in phase, their amplitudes were different which increased the variability between LPX-Bern and RF-DYPTOP (i.e. increase in RE).

The uncertainties of the upscaled fluxes were estimated from the spread of predictions made with the ensemble of 200 RF models (Fig. 9). The uncertainty mostly scales with the flux magnitude (compare Fig. 6a–c with Fig. 9a–c), meaning that grid cells with high fluxes tend to also have high uncertainties. However, the relative flux uncertainty does have some geographical variation (Fig. 9d–f). The highest relative uncertainties are typically at the highest and lowest latitudes of the study domain. In these locations the dependencies between the predictors and the CH$_4$ flux are not as well defined as in the locations with lower uncertainties leading to larger spread in the ensemble of RF model predictions. For instance, at low latitudes LSTn may go beyond the range of LSTn values in the training data (see the range in Fig. 3), and hence the RF model predictions are not well constrained in these situations. On the other hand, lower relative uncertainties are typically obtained for locations close to the measurement sites incorporated in this study (compare Figs. 1 and 9), since the dependencies between the predictors and the CH$_4$ flux are better defined.

The seasonalities of the upscaled fluxes and CH$_4$ fluxes from process models are similar with the highest CH$_4$ emissions in July–August and the lowest in February. This seasonal pattern is consistent throughout the study domain (Fig. 10). Warwick et al. (2016) and Thonat et al. (2017) showed that the northern wetland CH$_4$ emissions should peak in August–September in order to correctly explain the seasonality of atmospheric CH$_4$ mixing ratios and isotopes measured across the Arctic. Hence, the wetland CH$_4$ emissions presented here are peaking approximately 1 month too early to perfectly match with their findings. CH$_4$ flux magnitude agrees well between WetCHARTs and the upscaled flux during spring and midsummer (April–July), whereas LPX-Bern estimates lower fluxes (0 % and 26 % difference, respectively). During late summer and autumn (August–October) both process models estimate slightly lower fluxes than the upscaled estimate (17 % and 19 % difference, respectively). The upscaled fluxes also show somewhat higher emissions during the non-growing season (November–March) than the two process models (27 % and 35 % difference; see Table 2), and the upscaled estimates of non-growing season emissions are relatively close to a recent model estimate (Treat et al., 2018). This result promotes the recent notion that process models might be underestimating non-growing season fluxes at high latitudes (e.g. Treat et al., 2018; Xu et al., 2016a; Zona et al., 2016).

Treat et al. (2018) adjusted WetCHARTs model output so that it matches with their estimates of non-growing season CH$_4$ emissions and then estimated annual wetland CH$_4$ emissions north of 40° N to be 37 ± 7 Tg(CH$_4$) yr$^{-1}$ using this adjusted model output. The estimates derived here for the annual emissions using the three wetland maps are similar (see Table 2), especially when considering our slightly smaller study domain (above 45° N). The two process models included in this study estimated slightly lower mean annual emissions than the upscaled fluxes (11 % and 26 % difference between the mean upscaled estimate and WetCHARTs and LPX-Bern, respectively; see also Table 2). However, given the uncertainties in upscaling as well as in process models, this can be regarded as relatively good agreement. Different process models may be driven with different climate forcing data and they may have discrepancies in the underlying wetland distributions, in addition to the different parameterizations and descriptions of the processes behind the CH$_4$ emissions. These sources of uncertainty should be recognized when models are compared against each other or against upscaling products.

In order to further evaluate the agreement between the upscaled fluxes and process models we focused on two specific regions: Hudson Bay lowlands (HBL) and western Siberian lowlands (WSL) (see locations in Fig. 1). The upscaled fluxes indicate higher annual emissions for both subdomains compared to the two process models or previously published estimate (Table 2). For WSL the upscaled estimates are within the range of variability observed between process models and inversion modelling in WETCHIMP-WSL (Bohn et al., 2015) and close to Thompson et al. (2017). The upscaled estimates by Glagolev et al. (2011) might underestimate CH$_4$ emissions from the WSL area (Bohn et al., 2015). Furthermore, the process models in Bohn et al. (2015) are likely underestimating the non-growing season CH$_4$ emissions which might partly explain the discrepancy to the upscaled estimates in this study. Hence, the upscaled CH$_4$ emission esti-
Figure 6. Mean annual CH$_4$ wetland emissions during years 2013–2014 estimated by upscaling EC data using the RF model and three wetland maps (a, b, c) and process models (d, e). Grid cells with low CH$_4$ wetland emissions (below 0.1 g(CH$_4$) m$^{-2}$ yr$^{-1}$) are shown in grey. The flux rates refer to total unit area in a grid cell.

Figure 7. Difference in mean annual CH$_4$ wetland emissions during the years 2013–2014 estimated by upscaling EC data using the RF model with different wetland maps and process models. All the CH$_4$ emission maps were aggregated to 1° resolution before comparison. The flux rates refer to total unit area in a grid cell.

mates for the WSL area, while large, are still in a reasonable range.

For HBL, the discrepancy between upscaled emission estimates and the estimates based on process models or previous studies is larger (Table 2). The upscaling results agree with Zhang et al. (2016) and Melton et al. (2013) but show emissions that are twice as large for the HBL than the other estimates (Table 2). This cannot be explained by wetland mapping since the difference also holds when the DYP-TOP wetland map is used in upscaling. There are only few
long-term EC flux studies conducted in the HBL area and the only one found (Hanis et al., 2013) showed on average 6.9 g(CH₄) m⁻² annual emissions at a subarctic fen located in the HBL. If the upscaled CH₄ emissions are downscaled back to ecosystem level in the HBL area with wetland maps, we get on average 11.0 g(CH₄) m⁻² annual CH₄ emission for the HBL area based on the RF model output, which is 1.6 times larger than the estimate by Hanis et al. (2013). While Hanis et al. (2013) studied only one wetland during different years than those used here (years 2008–2011 in Hanis et al., 2013, here 2013–2014), it is still noteworthy that the relative difference between Hanis et al. (2013) and this study is similar to the discrepancy between this study and the inversion estimates (Pickett-Heaps et al., 2011; Thompson et al., 2017) at the whole HBL scale. Pickett-Heaps et al. (2011) and Thompson et al. (2017) show near zero CH₄ emissions during October–April and onset of CH₄ emissions in mid-May or even June, largely dependent on when the ground was free of snow and unfrozen. This is somewhat surprising given that only 32 % of wetlands in the area are underlain by permafrost (based on amalgam of PEATMAP and the permafrost map), and hence the soils are likely not completely frozen and some non-growing season CH₄ emissions are likely to occur in such conditions (e.g. Treat et al., 2018). The upscaled non-growing season CH₄ emissions show on average 1.1 Tg(CH₄) yr⁻¹ emissions for the HBL area. This partly, but not completely, explains the discrepancy between the CH₄ emission estimates for the HBL area. All these results suggest that the upscaled product likely overestimates CH₄ emissions from the HBL area.

4 Discussion

4.1 Comparing the RF model predictive performance to previous studies

The RF model performance was worse when compared against independent validation data than what has been achieved in previous upscaling studies for GPP and energy fluxes ($R^2 > 0.7$) and ecosystem respiration ($R_{eco}$; $R^2 > 0.6$) (e.g. Jung et al., 2010; Tramontana et al., 2016). However, the RF model performance for monthly CH₄ emissions was comparable to net ecosystem exchange of CO₂ (NEE) ($R^2 < 0.5$) (e.g. Jung et al., 2010; Tramontana et al., 2016). Likely reasons for this finding include, for instance, that for other fluxes there is simply more data available from several sites spanning the globe. For example, the La Thuile synthesis dataset used by Jung et al. (2010) and Tramontana et al. (2016) consists of 965 site years of data from over 252 EC stations. Here we have data from 25 sites with CH₄ fluxes. Furthermore, the drivers (or proxies for the drivers) of, for example, GPP and $R_{eco}$ are never directly measured with the EC technique, they are always at least partly derived products (Lasslop et al., 2009; Reichstein et al., 2005). Hence, direct functional relationships between GPP and $R_{eco}$ and their environmental drivers are inherently included in these flux estimates, whereas NEE and CH₄ emissions are directly measured without additional modelling. Also, both NEE and CH₄ emissions are differences between component fluxes (NEE:
Figure 9. Absolute (a–c) and relative (d–f) uncertainties of the upscaled CH$_4$ fluxes using different wetland maps. Uncertainty is estimated as 1σ variability of the predictions by 200 RF models developed by bootstrapping the training data (Sect. 2.1.2). Grid cells with low CH$_4$ wetland emissions (below 0.1 g(CH$_4$) m$^{-2}$ yr$^{-1}$) are shown in grey. The absolute uncertainties refer to total unit area in a grid cell.

Figure 10. Monthly time series of zonal mean CH$_4$ fluxes. The upscaled fluxes with different wetland maps are shown in (a, b, c) and wetland CH$_4$ emissions estimated with the two process models are given in (d, e).
nia et al. (2010) found on average RMSE instance, McNorton et al. (2016) validated the land surface emissions from northern wetlands (November–March) for several process models for CH4 flux data from seven sites, respectively. Zürcher et al. (2013) found the time-integrated CH4 flux to be well represented by LPX-Bern model across different sites. A tight correlation ($R^2 = 0.92$) is found between simulated and measured cumulative site emissions after calibrating the model against the measurements. While Xu et al. (2016a) did not explicitly show any statistical metrics, their model (CLM4.5) comparison against site level CH4 flux data seemed to be somewhat better than in Wanina et al. (2010) or McNorton et al. (2016). Xu et al. (2016a) emphasize the importance of non-growing season emissions and the fact that their model was clearly underestimating these emissions. Zhu et al. (2014) calibrated their model (CLM4.5) comparison against site level CH4 emissions from eight models ±1σ of interannual variation in the model estimates for the period 1993–2004. 5 Process model tuned to match atmospheric observations. 6 North of 45$^\circ$ N. 7 Mean ±1σ over the LPJ-wsl model results using different wetland extents for the period 1980–2000.

### Table 2. Annual CH4 wetland emissions in different subdomains (Hudson Bay lowlands and western Siberian lowlands; see Fig. 1) and time periods. The values are given in Tg(CH4) yr$^{-1}$. Note that estimates from some reference studies are not for the same period as the one studied here (2013–2014). For WetCHARTs the mean of the model ensemble together with the range (in parentheses) is given, whereas for the upscaling results the 95% confidence intervals for the estimated emissions are given.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Hudson Bay lowlands</th>
<th>Western Siberian lowlands</th>
<th>Non-growing season fluxes from northern wetlands (November–March)</th>
<th>Annual emissions north of 45$^\circ$ N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inversion models</td>
<td>Bohn et al. (2015), WETCHIMP-WSL</td>
<td>6.06 ± 1.22</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Brauhiler et al. (2014)$^a$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kim et al. (2011)</td>
<td>2.9 ± 1.7 and 3.0 ± 1.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Miller et al. (2014)</td>
<td>2.4 ± 0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spahni et al. (2011)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thompson et al. (2017)</td>
<td>6.9 ± 3.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Process models</td>
<td>Bohn et al. (2015), WETCHIMP-WSL</td>
<td>5.34 ± 0.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chen et al. (2015)$^b$</td>
<td>3.11 ± 0.45</td>
<td>35.0 ± 6.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Melton et al. (2013), WETCHIMP$^c$</td>
<td>5.4 ± 3.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pickett-Heaps et al. (2011)$^d$</td>
<td>2.3 ± 0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Treat et al. (2018)$^e$</td>
<td>6.1 ± 1.5</td>
<td>37 ± 7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Watts et al. (2014)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zhang et al. (2016)$^f$</td>
<td>4.6 ± 0.6</td>
<td>30.3 ± 5.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>This study, LPX-Bern</td>
<td>4.4</td>
<td>24.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>This study, WetCHARTs</td>
<td>5.1 (0.6–17.0)</td>
<td>29.7 (8.7–74.0)</td>
<td></td>
</tr>
<tr>
<td>Flux measurement upscaling</td>
<td>Glagolev et al. (2011)</td>
<td>3.9 ± 1.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zhu et al. (2013)</td>
<td></td>
<td>44.0–53.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>This study, RF-PEATMAP</td>
<td>4.6 (3.1–6.0)</td>
<td>7.0 (5.2–8.8)</td>
<td>6.2 (4.6–7.8)</td>
</tr>
<tr>
<td></td>
<td>This study, RF-DYPTOP</td>
<td>4.9 (3.4–6.5)</td>
<td>6.8 (5.0–8.5)</td>
<td>8.0 (5.8–10.2)</td>
</tr>
<tr>
<td></td>
<td>This study, RF-GLWD</td>
<td>4.6 (3.1–6.0)</td>
<td>7.0 (5.2–8.8)</td>
<td>6.2 (4.6–7.8)</td>
</tr>
</tbody>
</table>

$^a$ Approximately north of 47$^\circ$ N.  $^b$ Approximately north of 45$^\circ$ N.  $^c$ Mean annual CH4 emissions from eight models ±1σ of interannual variation in the model estimates for the period 1993–2004.  $^d$ Process model tuned to match atmospheric observations.  $^e$ North of 46$^\circ$ N.  $^f$ Mean ±1σ over the LPJ-wsl model results using different wetland extents for the period 1980–2000.

GPP and $R_{ec}$; CH4 flux: production and oxidation). Therefore, GPP and $R_{ec}$ upscaling algorithms show better correspondence with validation data than for NEE or CH4 emissions and the results for NEE would be the correct point of reference for the RF model performance presented here.

While the RF model performance in this study was inferior to previous upscaling studies for other fluxes when evaluated using different statistical metrics, it was still comparable to what has been shown before for several process models for CH4 emissions (McNorton et al., 2016; Wanina et al., 2010; Zürcher et al., 2013; Zhu et al., 2014; Xu et al., 2016a). For instance, McNorton et al. (2016) validated the land surface model JULES against CH4 flux data from 13 sites and found $R^2 = 0.10$ between the validation data and the model. Wanina et al. (2010) found on average RMSE = 29 nmol m$^{-2}$ s$^{-1}$ and RMSE = 42 nmol m$^{-2}$ s$^{-1}$ with and without tuning their model LPJ-WHyME against CH4 flux data from seven sites, respectively. Zürcher et al. (2013) found the time-integrated CH4 flux to be well represented by LPX-Bern model across different sites. A tight correlation ($R^2 = 0.92$) is found between simulated and measured cumulative site emissions after calibrating the model against the measurements. While Xu et al. (2016a) did not explicitly show any statistical metrics, their model (CLM4.5) comparison against site level CH4 flux data seemed to be somewhat better than in Wanina et al. (2010) or McNorton et al. (2016). Xu et al. (2016a) emphasize the importance of non-growing season emissions and the fact that their model was clearly underestimating these emissions. Zhu et al. (2014) calibrated their model (TRIPLEX-GHG) for each measurement site by changing,
e.g. the $Q_{10}$ for CH4 production and CH4 to CO2 release ratio to be site-specific and found on average $R^2 = 0.64$ when comparing the calibrated model against measurements at 17 CH4 flux measurement sites. However, their findings are not directly comparable to the RF model agreement with validation data shown here due to their model calibration against data before comparison. Nevertheless, their results show that, even after calibration, the process models are not fully able to capture the CH4 flux variability in measurements. Miller et al. (2014) argued that the structure of some of the process models is so complex that the required forcing variables may not be reliable at larger spatial scales. All of these five models (JULES, LPJ-WHYME, LPX-Bern, CLM4.5 and TRIPLEX-GH) are contributing to the global CH4 budget estimation within the Global Methane Project (Saunois et al., 2016), highlighting that these results summarize the agreement between state-of-the-art process models and field measurements.

4.2 Methods to improve RF model predictive performance

4.2.1 Missing predictors

In this study a statistical model was developed using the RF algorithm, and the model was able to yield $R^2 = 0.47$ against monthly CH4 flux validation data. Our upscaling using the RF model focused on 2013–2014, as these were the years with the largest overlap of collected data. However, all data from all the years (2005–2016) were used to develop and validate the model. The incomplete match between the RF model and validation data is likely caused by the fact that not all the possible drivers causing inter- and intra-site variability in CH4 emissions were included in the analysis, and hence all the variability could not be explained by the model.

Christensen et al. (2003) were able to explain practically all the variability ($R^2 = 0.92$) in annual CH4 emissions in their multi-site chamber study with only two predictors: temperature and the availability of substrates for CH4 production. Also, Yvon-Durocher et al. (2014) speculate that the amount of substrates for microbial CH4 production explains across-site variability of CH4 fluxes in their data. However, gridded data on spatially explicit substrate information are currently nonexistent. Hence, proxies for the substrates available for methanogenesis are needed. The current paradigm on wetland CH4 emissions is that most of the emitted CH4 is produced from recently fixed carbon being used as precursors for the CH4-producing Archaea (e.g. Chanton et al., 1995; Whiting and Chanton, 1993). Most process models are based on the premise that a certain fraction of ecosystem net primary productivity (NPP) is available and used for CH4 production or alternatively a fraction of heterotrophic respiration is allocated to CH4 emissions (e.g. Xu et al., 2016b). Thus, NPP (or GPP) could potentially be included as a predictor for the RF model and used as a proxy for the amount of substrates available for CH4 production. However, the RF model performance in this study was not enhanced if variables closely related to NPP (EVI and the product of EVI and LSTd) were included as predictors. Also, Knox et al. (2019) did not find GPP as an important predictor of CH4 emission variability in their multi-site synthesis study.

Using NPP (or proxies for it) for the RF model development might be an oversimplification, since it has been shown that the deep-rooted sedges and their NPP are especially important for CH4 production (Joabsson and Christensen, 2002; Ström et al., 2003, 2012; Waddington et al., 1996). Hence, information about plant functional types (PFTs) would be needed to better explain the CH4 flux variability (Davidson et al., 2017; Gray et al., 2013). Furthermore, the fraction of the fixed carbon allocated to the roots and released as root exudates (hence, available for CH4 production) varies between species and root age (Proctor and He, 2017; Ström et al., 2003), further complicating the connection between NPP and CH4 emissions. The sedges also act as conduits for CH4, allowing the CH4 produced below water level to rapidly escape to the atmosphere and bypass the oxic zone in which the CH4 might have otherwise been oxidized (Waddington et al., 1996; Whiting and Chanton, 1992). Besides sedges, Spagnum mosses are also important because methanotrophic bacteria that live in symbiosis with these mosses significantly decrease the CH4 emissions to the atmosphere when they are present (Larmola et al., 2010; Liebner et al., 2011; Parmentier et al., 2011b; Raghoebarsing et al., 2005; Sundh et al., 1995). In a modelling study, Li et al. (2016) showed that it was essential to consider the vegetation differences between sites when modelling CH4 emissions from two northern peatlands. Hence, ideally one should have gridded information on wetland species composition and associated NPP across the high latitudes to significantly improve the upscaling results. Unfortunately, such information is not yet available and therefore modelled estimates could be used (e.g. LPX-Bern, which includes several peatland-specific PFTs allowed to freely evolve during the model run) (Spahni et al., 2013). However, in such cases the upscaled CH4 emission estimates would no longer be independent of the model and therefore would be less suitable for model validation. We also note that many process models have only one PFT per wetland.

Different variables related to water input to the ecosystem (i.e. $P$, $P_{an}$) or surface moisture (SRWI) did not enhance the RF model predictive performance, not only reflecting that water table depth (WTD) is not solely controlled by input of water via precipitation but also that evapotranspiration and lateral flows affect wetland WTD, data that were missing from our study. These findings are consistent with previous studies (e.g. Christensen et al., 2003; Rinne et al., 2018; Pugh et al., 2018 and Knox et al., 2019), who showed only a modest CH4 flux dependence on WTD in wetlands and peatlands. In contrast, several chamber-based studies have shown a positive relationship between WTD and CH4 fluxes (Granberg...
et al., 1997; Olefeldt et al., 2012; Treat et al., 2018; Turetsky et al., 2014). In general, chamber-based studies often show spatial dependency of CH$_4$ flux on WTD, whereas studies done at an ecosystem scale with EC generally do not show temporal WTD dependency, albeit there are exceptions (e.g. Zona et al., 2009). This might indicate that WTD controls metre-scale spatial heterogeneity of CH$_4$ flux between microtopographical features (e.g. Granberg et al., 1997) but not temporal variability on the ecosystem scale, provided that WTD stays relative close to the surface. Also, the chamber studies tend to observe spatial variation, which can be indirectly influenced by WTD via its influence on plant communities, whereas EC studies observe typically temporal variation in sub-annual timescales. However, the effect of WTD might be masked by a confounding effect caused by plant phenology, since vegetation biomass often peaks at the same time as the WTD is at its lowest. While the variables related to WTD did not increase the RF model performance, WTD might still play a role in controlling ecosystem-scale CH$_4$ variability when it is exceptionally high or low. For instance, the year 2006 was exceptionally dry at the Siikaneva fen, and hence CH$_4$ emissions during that year were lower than on average (see Fig. 5a). However, in order to accurately capture such dependencies with machine-learning techniques (such as RF), they should be frequent enough so that the model can learn these dependencies.

RF model performance was better at permafrost-free than at sites with permafrost, which might indicate that the LSTn might not be an appropriate proxy for the temperature controlling the CH$_4$ production and oxidation rates at sites with permafrost. Also, no information on the development of the seasonally unfrozen, hydrologically and biogeochemically active layer was included in the RF model. Furthermore, Zona et al. (2016) showed strong hysteresis between soil temperatures and CH$_4$ emissions at their permafrost sites in Alaska, whereas Rinne et al. (2018) show a synchronous exponential dependence between soil temperature and CH$_4$ emissions at a boreal fen without permafrost. The hysteresis observed in Zona et al. (2016) could be explained by the fact that part of the produced CH$_4$ at these permafrost sites is stored below ground for several months before it is being emitted to the atmosphere, causing a temporal lag between soil temperature and observed surface flux. In any case, more knowledge of soil processes (soil thawing and freezing, CH$_4$ production and storage) is needed before the CH$_4$ emissions from these permafrost ecosystems can be extrapolated to other areas with greater confidence.

It should be emphasized that the drivers causing across-site variability in ecosystem-scale CH$_4$ emissions are, in general, unknown since studies comparing EC CH$_4$ fluxes from multiple wetland sites have only recently been published (Baldocchi, 2014; Knox et al., 2019; Petrescu et al., 2015). Most previous CH$_4$ synthesis studies were based on plot-scale measurements (Bartlett and Harriss, 1993; Olefeldt et al., 2012; Treat et al., 2018; Turetsky et al., 2014). However, the CH$_4$ flux responses to environmental drivers and their relative importance might be different at an ecosystem scale since CH$_4$ fluxes typically show significant spatial variability at sub-metre scale (e.g. Sachs et al., 2010). Furthermore, the temporal coverage of plot-scale measurements with chambers is usually relatively poor, whereas EC measurements provide continuous data on ecosystem scale. This study and Knox et al. (2019) show that temperature is important when predicting CH$_4$ flux variability in a multi-site CH$_4$ flux dataset, but a significant fraction of CH$_4$ flux variability is still left unexplained. It remains a challenge for future EC CH$_4$ flux synthesis studies to discover the drivers explaining the rest of the variability.

4.2.2 Quality and representativeness of CH$_4$ flux data

The RF model performance may improve if instrumentation, measurement setup and the data processing are harmonized across sites, since these discrepancies between flux sites might have caused spurious differences in CH$_4$ fluxes. These differences would have created additional variability in the synthesis dataset, which would in turn (1) influence the training of RF model and (2) decrease, for example, NSE values obtained against validation data, since there would be artificial variability in the validation data, which is not related to the predictors. In this study, the site PIs processed the data themselves using different processing codes, albeit the gap filling was done centrally in a standardized way.

While these issues mentioned above could impact the upscaling results shown here, prior studies have shown that the usage of different instruments or processing codes does not significantly impact CH$_4$ flux estimates. For instance, Mammarella et al. (2016) showed that the usage of different processing codes (EddyPro and EddyUH) resulted, in general, in a 1% difference in long-term CH$_4$ emissions. On the other hand, CH$_4$ instrument cross comparisons have shown small differences (typically less than 7%) between the long-term CH$_4$ emission estimates derived using different instruments (Goodrich et al., 2016; Peltola et al., 2013, 2014). While these studies show consistent CH$_4$ emissions, they also stress that the data should be carefully processed to achieve such good agreement across processing codes and instruments. In addition, many issues related to, for example, friction velocity filtering and gap filling of CH$_4$ fluxes are still unresolved, and the role of short-term emission bursts, which are common in methane flux time series, needs to be further investigated (e.g. Schaller et al., 2017). However, recently Nemitz et al. (2018) advanced these issues by proposing a methodological protocol for EC measurements of CH$_4$ fluxes used to standardize CH$_4$ flux measurements within the ICOS measurement network (Franz et al., 2018).

A total of 25 flux measurement sites were included in this study and they were distributed across the Arctic boreal region (see Fig. 1). The measurements were largely concentrated in Fennoscandia and Alaska, whereas data from, for
example, the HBL and WSL areas, were missing. Long-term EC CH$_4$ flux measurements are largely missing from these vast wetland areas, casting uncertainty on wetland CH$_4$ emissions from these areas. The location of a flux site is typically restricted by practical limitations related to, for example, ease of access and availability of grid power. Hence, open-path instruments with low power requirements potentially open up new areas for flux measurements (McDermitt et al., 2010), yet they need continuous maintenance, which is not necessarily easy in remote locations. However, one could argue that the geographical location of flux sites is not vital for upscaling, more important is that the available data represents well the full range of CH$_4$ fluxes across the northern latitudes and more importantly the CH$_4$ flux responses to the environmental drivers. Also, sites should ideally cover all different wetlands with varying plant species composition, whereas geographical representation is not necessarily as important. CH$_4$ flux site representativeness could be potentially assessed in the same vein as in previous studies for other measurement networks (Hargrove et al., 2003; Hoffmann et al., 2013; Papale et al., 2015; Sulkava et al., 2011). However, before such analysis can be done, the main drivers causing across-site variability in ecosystem-scale CH$_4$ fluxes should be better identified.

Most of the CH$_4$ flux data here and in the literature have been recorded during the growing season when the CH$_4$ fluxes are at a maximum, whereas year-round continuous CH$_4$ flux measurements are not as common. This is likely due to the harsh conditions in the Arctic during winter that make continuous high-quality flux measurements very demanding (e.g. Goodrich et al., 2016; Kittler et al., 2017a) but also in part since the large-scale importance of non-growing season emissions has just recently been recognized (Kittler et al., 2017b; Treat et al., 2018; Xu et al., 2016a; Zona et al., 2016). For upscaling year-round CH$_4$ emissions, continuous measurements are vital to accurately constrain also the non-growing season emissions and their drivers.

5 Data availability

The presented upscaled CH$_4$ flux maps (RF-DYPTOP, RF-PEATMAP and RF-GLWD), their uncertainties and the underlying CH$_4$ flux densities are accessible via an open-data repository Zenodo (Peltola et al., 2019). The datasets are saved in netCDF files and they are accompanied by a readme file. The dataset can be downloaded from https://doi.org/10.5281/zenodo.2560163.

6 Conclusions

Methane (CH$_4$) emission data comprising over 40 site years from 25 eddy covariance flux measurement sites across the Arctic boreal region were assembled and upscaled to estimate CH$_4$ emissions from northern (> 45° N) wetlands. The upscaling was done using the random forest (RF) algorithm. The performance of the RF model was evaluated against independent validation data utilizing the leave-one-site-out scheme, which yielded value of 0.47 for both the Nash–Sutcliffe model efficiency and $R^2$. These results are similar to previous upscaling studies for the net ecosystem exchange of carbon dioxide (NEE) but worse for the individual components of NEE or energy fluxes (e.g. Jung et al., 2010; Tramontana et al., 2016). The performance is also comparable to studies where process models are compared against site CH$_4$ flux measurements (McNorton et al., 2016; Wania et al., 2010; Zürcher et al., 2013; Zhu et al., 2014; Xu et al., 2016a). Hence, despite the relatively high fraction of unexplained variability in the CH$_4$ flux data, the upscaling results are useful for comparing against models and could be used to evaluate model results. The three gridded CH$_4$ wetland flux estimates and their uncertainties are openly available for further usage (Peltola et al., 2019).

The upscaling to the regions > 45° N resulted in mean annual CH$_4$ emissions comparable to prior studies on wetland CH$_4$ emissions from these areas (Bruhwiler et al., 2014; Chen et al., 2015; Spahni et al., 2011; Treat et al., 2018; Watts et al., 2014; Zhang et al., 2016; Zhu et al., 2013) and hence, in general, support the prior modelling results for the northern wetland CH$_4$ emissions. When compared to two validation areas, the upscaling likely overestimated CH$_4$ emissions from the Hudson Bay lowlands, whereas emission estimates for the western Siberian lowlands were in a reasonable range. Future CH$_4$ flux upscaling studies would benefit from long-term continuous CH$_4$ flux measurements, centralized data processing and better incorporation of CH$_4$ flux drivers (e.g. wetland vegetation composition and carbon cycle) from remote-sensing data needed for scaling the fluxes from the site level to the whole Arctic boreal region.
### Table A1. Description of eddy covariance sites included in this study.

<table>
<thead>
<tr>
<th>Name</th>
<th>Site ID</th>
<th>PI</th>
<th>Latitude, longitude</th>
<th>Amount of monthly CH$_4$ flux data available</th>
<th>Reference</th>
<th>Permafrost present (true/false)</th>
<th>Sedges as dominant vegetation type</th>
<th>Biome based on Olson et al. (2011)</th>
<th>Wetland type</th>
<th>Time resolution of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schechenfliz Nord</td>
<td>DE-SfN</td>
<td>Janina Klatt, Hans Peter Schmid</td>
<td>47.8064, 11.3275</td>
<td>2 Hommeltenberg et al. (2014)</td>
<td>false</td>
<td>false</td>
<td>temperate bog</td>
<td>true</td>
<td>wet tundra</td>
<td>30 min</td>
</tr>
<tr>
<td>Chokurdakh</td>
<td>RU-Cok</td>
<td>Albertus J. Dolman</td>
<td>70.8291, 147.4943</td>
<td>5 Parmentier et al. (2011a)</td>
<td>true</td>
<td>true</td>
<td>tundra</td>
<td>false</td>
<td>wet tundra</td>
<td>30 min</td>
</tr>
<tr>
<td>Yorkuta</td>
<td>RU-Vor</td>
<td>Thomas Friborg</td>
<td>67.0547, 62.9405</td>
<td>5 Marushchak et al. (2016)</td>
<td>true</td>
<td>false</td>
<td>tundra</td>
<td>false</td>
<td>wet tundra</td>
<td>30 min</td>
</tr>
<tr>
<td>Stordalen</td>
<td>SE-St1</td>
<td>Thomas Friborg</td>
<td>68.3542, 19.0503</td>
<td>6 Jammet et al. (2017)</td>
<td>false</td>
<td>true</td>
<td>tundra</td>
<td>false</td>
<td>fen</td>
<td>30 min</td>
</tr>
<tr>
<td>Stordalen (ICOS)*</td>
<td>SE-Sto</td>
<td>Janne Rinne</td>
<td>68.3560, 19.0452</td>
<td>55 true and false</td>
<td>true</td>
<td>false</td>
<td>fen</td>
<td>true</td>
<td>bog</td>
<td>30 min</td>
</tr>
<tr>
<td>Siikaneva 1</td>
<td>FI-Si1</td>
<td>Timo Vesala, Ivan Mammarella</td>
<td>61.8327, 24.1928</td>
<td>104 Rinne et al. (2018)</td>
<td>true</td>
<td>true</td>
<td>boreal fen</td>
<td>false</td>
<td>daily</td>
<td>30 min</td>
</tr>
<tr>
<td>Siikaneva 2</td>
<td>FI-Si2</td>
<td>Timo Vesala, Ivan Mammarella</td>
<td>61.8375, 24.1699</td>
<td>26 Korensoalo et al. (2018)</td>
<td>true</td>
<td>false</td>
<td>boreal bog</td>
<td>true</td>
<td>daily</td>
<td>30 min</td>
</tr>
<tr>
<td>Lompolojänkkä</td>
<td>FI-Lom</td>
<td>Anna-Lotta Kohila</td>
<td>67.9972, 24.2092</td>
<td>59 Aurela et al. (2009)</td>
<td>false</td>
<td>true</td>
<td>boreal fen</td>
<td>false</td>
<td>daily</td>
<td>30 min</td>
</tr>
<tr>
<td>James Bay lowlands</td>
<td>CA-JBL</td>
<td>Daniel F. Nadeau</td>
<td>53.674, −78.1706</td>
<td>3 Nadeau et al. (2013)</td>
<td>true</td>
<td>false</td>
<td>boreal bog</td>
<td>false</td>
<td>daily</td>
<td>30 min</td>
</tr>
<tr>
<td>Lost Creek</td>
<td>US-Los</td>
<td>Ankur R. Desai</td>
<td>46.0827, −89.9792</td>
<td>30 Pugh et al. (2018)</td>
<td>false</td>
<td>true</td>
<td>temperate fen</td>
<td>true</td>
<td>fen</td>
<td>30 min</td>
</tr>
<tr>
<td>Anquask</td>
<td>US-Aq</td>
<td>Donatella Zona</td>
<td>70.4696, −157.4089</td>
<td>11 Zona et al. (2016)</td>
<td>true</td>
<td>true</td>
<td>tundra</td>
<td>true</td>
<td>wet tundra</td>
<td>30 min</td>
</tr>
<tr>
<td>Barrow Environmental Observatory</td>
<td>US-Bes</td>
<td>Donatella Zona</td>
<td>71.2810, −156.6123</td>
<td>16 Zona et al. (2016)</td>
<td>true</td>
<td>true</td>
<td>tundra</td>
<td>true</td>
<td>wet tundra</td>
<td>30 min</td>
</tr>
<tr>
<td>Biocomplexity</td>
<td>US-Bes</td>
<td>Donatella Zona</td>
<td>71.2809, −156.5965</td>
<td>16 Zona et al. (2016)</td>
<td>true</td>
<td>true</td>
<td>tundra</td>
<td>true</td>
<td>wet tundra</td>
<td>30 min</td>
</tr>
<tr>
<td>Ivonuk</td>
<td>US-Ivo</td>
<td>Donatella Zona</td>
<td>68.4865, 155.7502</td>
<td>15 Zona et al. (2016)</td>
<td>true</td>
<td>true</td>
<td>wet tundra</td>
<td>true</td>
<td>wet tundra</td>
<td>30 min</td>
</tr>
<tr>
<td>Western peatland 1</td>
<td>CA-WP1</td>
<td>Lawrence B. Flanagan</td>
<td>54.9538, −112.4670</td>
<td>5 Long et al. (2010)</td>
<td>false</td>
<td>false</td>
<td>temperate fen</td>
<td>true</td>
<td>fen</td>
<td>30 min</td>
</tr>
<tr>
<td>Mer Blesie</td>
<td>CA-Mer</td>
<td>Elyn Humphreys</td>
<td>45.4094, −75.5186</td>
<td>16 Brown et al. (2014)</td>
<td>false</td>
<td>false</td>
<td>temperate bog</td>
<td>true</td>
<td>daily</td>
<td>30 min</td>
</tr>
<tr>
<td>Chersky reference</td>
<td>RU-Ch2</td>
<td>Mathias Göckede</td>
<td>68.6169, 161.3509</td>
<td>21 Kitter et al. (2017)</td>
<td>true</td>
<td>true</td>
<td>boreal</td>
<td>true</td>
<td>daily</td>
<td>30 min</td>
</tr>
<tr>
<td>Raczin</td>
<td>PL-wet</td>
<td>Janusz Olejnik</td>
<td>52.7622, 16.3094</td>
<td>4 Kowalski et al. (2013)</td>
<td>true</td>
<td>false</td>
<td>temperate fen</td>
<td>true</td>
<td>fen</td>
<td>30 min</td>
</tr>
<tr>
<td>Degerø Stormyr</td>
<td>SE-Dreg</td>
<td>Mats B. Nilsson, Matthias Peichl</td>
<td>64.1820, 19.5567</td>
<td>22 Nilsson et al. (2008)</td>
<td>true</td>
<td>true</td>
<td>boreal</td>
<td>true</td>
<td>fen</td>
<td>30 min</td>
</tr>
<tr>
<td>Seney</td>
<td>US-Sen</td>
<td>Thomas Pyper</td>
<td>46.3167, −86.0500</td>
<td>5 Pyper et al. (2013)</td>
<td>true</td>
<td>false</td>
<td>temperate fen</td>
<td>true</td>
<td>fen</td>
<td>30 min</td>
</tr>
<tr>
<td>Scotty Creek</td>
<td>CA-SCC</td>
<td>Oliver Sonnentag</td>
<td>61.3000, −121.300</td>
<td>14 Helbig et al. (2016)</td>
<td>false</td>
<td>false</td>
<td>boreal</td>
<td>false</td>
<td>daily</td>
<td>30 min</td>
</tr>
<tr>
<td>Samoylov</td>
<td>RU-Sam</td>
<td>Torsten Sachs</td>
<td>72.3667, 126.5000</td>
<td>11 Sach et al. (2008)</td>
<td>true</td>
<td>true</td>
<td>tundra</td>
<td>true</td>
<td>wet tundra</td>
<td>30 min</td>
</tr>
<tr>
<td>Immavait Creek</td>
<td>US-ICh</td>
<td>Eugenie S. Euskirchen</td>
<td>68.6000, 149.3110</td>
<td>7 true</td>
<td>true</td>
<td>false</td>
<td>tundra</td>
<td>true</td>
<td>wet tundra</td>
<td>30 min</td>
</tr>
<tr>
<td>Bonanza Creek, fen</td>
<td>US-BCF</td>
<td>Eugenie S. Euskirchen</td>
<td>64.7040, 148.3130</td>
<td>16 Euskirchen et al. (2014)</td>
<td>true</td>
<td>false</td>
<td>boreal</td>
<td>true</td>
<td>fen</td>
<td>30 min</td>
</tr>
<tr>
<td>Bonanza Creek, bog</td>
<td>US-BCB</td>
<td>Eugenie S. Euskirchen</td>
<td>64.7000, 148.3200</td>
<td>14 Euskirchen et al. (2014)</td>
<td>false</td>
<td>false</td>
<td>boreal</td>
<td>false</td>
<td>daily</td>
<td>30 min</td>
</tr>
</tbody>
</table>

* Data from this site is divided into two since data from two wind directions differ from each other (with and without permafrost).
Author contributions. OP, TA and TV designed the study and YG contributed further ideas for the study. OP did the data processing and analysis. OR prepared the PEATMAP map for the study. PA, MA, BC, ARD, AJD, ESE, TF, MG, MH, EH, GJ, JK, NK, LK, AL, IM, DFN, MBN, WCO, MP, TP, WQ, JR, TS, MS, HPS, OS, CW and DZ provided CH₄ fluxes and other in situ data for the study. FJ and SL did the LPX-Bern model runs. OP wrote the first version of the manuscript and all authors provided input.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. Lawrence B. Flanagan is acknowledged for providing data from CA-WP1 site. Lawrence B. Flanagan acknowledges support from the Natural Sciences and Engineering Council of Canada and Canadian Foundation for Climate and Atmospheric Sciences. Olli Peltola is supported by the postdoctoral researcher project (decision 315424) funded by the Academy of Finland. Olle Räty is supported by the Academy of Finland IDA-MARI project (decision 313828). Financial support from the Academy of Finland Centre of Excellence (grant nos. 312571 and 282842), ICOS-Finland (grant no. 281255) and CARB-ARC project (grant no. 285630) is acknowledged. Sara H. Knox and Robert B. Jackson acknowledge support from the Gordon and Betty Moore Foundation (grant no. 281255, 285630, 272041 and 307331), the Gordon and Betty Moore Academy of Finland (grant nos. 315424, 313828, 312571, 282842, 312571 and 282842, 313828, 312571, 282842, 312571 and 282842, 313828, 312571, 282842, 312571 and 282842), the Vermont Center of Excellence (grant nos. 272041 and 307331), the Gordon and Betty Moore Foundation Professor projects (grant nos. 312571 and 282842), ICOS-Finland (grant no. 281255) and CARB-ARC project (grant no. 285630) is acknowledged. Sara H. Knox and Robert B. Jackson acknowledge support from the Gordon and Betty Moore Foundation through grant GBMF5439 “Advancing Understanding of the Global Methane Cycle”. Ankur R. Desai acknowledges support of the DOE Ameriflux Network Management Project. Albertus J. Dolman acknowledges support from the Netherlands Earth System Science Centre, NESSC. Torsten Sachs was supported by the Helmholtz Association of German Research Centres (grant no. VH-NG-821). Ivan Mammarella and Timo Vesala thank the EU for supporting the RINGO project funded by the Horizon 2020 Research and Innovation Programme (grant no. 730944). The EU-H2020 CRESCENDO project (grant no. 641816) is also acknowledged. Fortunat Joos and Matthias Peichl acknowledge support from the National Research Programme (grant no. VH-NG-821). The Gordon and Betty Moore Foundation (grant no. GBMF5439), the Helmholtz Association (grant no. VH-NG-821), and Horizon 2020 (RINGO (grant no. 730944) and CRESCENDO (grant no. 641816)).

Financial support. This research has been supported by the Academy of Finland (grant nos. 315424, 313828, 312571, 282842, 281255, 285630, 272041 and 307331), the Gordon and Betty Moore Foundation (grant no. GBMF5439), the Helmholtz Association (grant no. VH-NG-821), and Horizon 2020 (RINGO (grant no. 730944) and CRESCENDO (grant no. 641816)).

Review statement. This paper was edited by David Carlson and reviewed by two anonymous referees.

References


Hall, D. K. and Riggs, G. A.: MODIS/Terra Snow Cover Monthly L3 Global 0.05 Deg CMG, Version 6, NASA National Snow and Ice Data Center Distributed Active Archive Center Boulder, Col-


