Biogeosciences

Open Access

EGU

Supplement of

High-resolution spatial patterns and drivers of terrestrial ecosystem carbon dioxide, methane, and nitrous oxide fluxes in the tundra

Anna-Maria Virkkala et al.

Correspondence to: Anna-Maria Virkkala (avirkkala@woodwellclimate.org)

S1. Chamber measurement details

We installed the collars 48 hours prior to the flux measurements to minimise the potential disturbance caused by the installation. We sealed the edges of the collars using moist inert quartz sand to secure an air-tight closure. We installed two steel collars at each GHG flux plot: one for the CO_2 flux measurements, and another one for the CH_4 and N_2O measurements. These two were placed in similar conditions ca. 0.4 to 2 meters away from each other. We used separate collars for the two measurements due to destructive sampling conducted in CO_2 flux collars after the measurements.

We hypothesized that most of the N_2O fluxes and CH_4 uptake fluxes to be small in this landscape dominated by upland tundra, and therefore used a 50-min chamber enclosure time to detect small changes in these concentrations (for a similar closure time, see e.g. Marushchak et al., 2021; Voigt et al., 2017). We used an opaque chamber, covered by space tape that reflects the sun, and did not thus observe any clear signs of heating of the chamber. The chamber headspace temperature difference during the start and end of the measurement ranged from -2.3 to 0.5 ° C (25th and 75th quantiles). Despite the long chamber enclosure time, the relationship between CH_4 concentrations and measurement time at sites with high CH_4 emissions (wetlands) was linear, indicating no issues with the chamber closure time (see Fig. S9).

Measurements were made under clear weather conditions but it also rained during the study period. Rainfall can impact the soil moisture conditions and thus GHG fluxes. It rained on 8 days during July 2018, and three of the days had heavier rain (>8 mm per day, FMI 2018). We made flux measurements during two of these days because during the measurement time, the conditions were sunny. Nevertheless, the three heavier rainfall days had clear but small impacts on soil moisture (volumetric water content (%) increased by 0.01-0.08 units) and it took approximately 0-3 days for soil moisture to return to the preceding soil moisture level after the rain (Fig. S10). Our data show that the range and mean of CH_4 flux is similar both in the plots measured during or 1-3 days after the rain and during other days, suggesting that rainfall events did not have a major influence on our results. The mean CH_4 flux during or 1-3 days after the rain was -1.8, and range -4.7 and 0.2 mg C m^{-2} d^{-1} (n=14), and during other days -1.5 (mean), and -4.9 and 0.1 (range) mg C m^{-2} d^{-1} (n=72); note that wetlands were not considered in this comparison because of their uneven distribution during these time periods which also impacts the summary statistics due to their high VWC% levels. In our upscaling framework, we control for the rainfall events as the GHG flux predictions are based on bi-hourly soil moisture and temperature maps that should reflect changes in soil moisture conditions after rain.

Mean air temperatures in the chamber were 22 °C during the CH_4 and N_2O measurements, and 25 °C during the CO_2 measurements, and the mean July temperature between 10 am and 5 pm from the temperature loggers (15 cm above the soil surface) in the study domain was 20 °C. Mean soil temperatures were 11 °C during CH_4 and N_2O measurements, and 13 °C during the CO_2 measurements (range between 3 and 19 °C). Mean July soil temperature between 10 am and 5 pm from the temperature loggers in the study domain was 14 °C (range between 5 and 27 °C). Mean soil moisture was 27 % during the CH_4 and N_2O flux measurements and 24 % during the CO_2 flux measurements. Mean July soil moisture between 10 am and 5 pm was 30 %. Note that not

all flux plots had continuous temperature and moisture loggers; the differences in the distribution of microclimate loggers might thus explain some of the differences between the means.

S2. Flux calculation statistics

The median explanatory power (R^2) of linear regression used to estimate CO_2 fluxes was 0.91 and root mean square error (RMSE) 0.91 ppm CO_2. Small fluxes from plots with little biomass had in general lower R^2 estimates but their RMSE was relatively similar to larger fluxes. For example, median R^2 for CO_2 fluxes between -0.5 and 0.5 g C m-2 d^{-1} was 0.11 and RMSE 0.89 ppm CO_2, whereas the same values for fluxes larger than those (i.e. more negative and more positive) were 0.94 and 0.91 ppm CO_2. We decided to keep all these fluxes in our data to guarantee that our measurements cover a broad range of fluxes and environmental conditions. The mean R^2 of linear regression for CH_4 was 0.87, and lower, 0.30 for N_2O due to the high proportion of near-zero fluxes.

S3. Microclimate data processing

Due to logger failures and loggers that were lifted from the soil during the measurement period, some data are missing in the temperature time series. This was often caused by reindeer picking the loggers up, but could have also been related to frost heave or other disturbance. To fill these gaps we utilized all logger data from the broader site (a total of 250 loggers) from the whole measurement period from July 2016 to September 2019 aggregated to daily mean values and combined these data with corresponding temperature records from nearby weather stations. This dataset was used to impute the missing values with the functions of Amelia R library (Honaker et al., 2011). Amelia provides multiple imputations for multivariate time series with a bootstrapped version of expectation maximisation algorithm to replace the missing values with draws from posterior distribution.

S4. Remote sensing variables

The remotely sensed data were extracted to our GHG flux and environmental data plots in their original spatial resolution, but they were resampled to a 2-meter resolution for the statistical upscaling.

S4.1 Vegetation classification and NDVI

For upscaling, we conducted a supervised classification of multitemporal satellite imagery (three PlanetScope scenes from dates 2018-07-02, 2018-07-29 and 2018-09-09; four spectral bands at 3-m resolution; georectified surface reflectance products). We stacked the images and calculated normalized difference spectral indices between all possible combinations of two spectral bands in the image stack resulting in 78 raster layers. As training data, we used the locations and vegetation classes of the GHG flux plots (n=101) and an additional set of 6034 plots within the study area classified in the field in summer 2016 matched with the five vegetation classes used in this study. After merging the plot data, the distribution of observations across the five classes was highly unbalanced. Thus, we collected 389 additional points covering mostly the less well covered vegetation classes (mainly barren and wetlands) based on the authors' expertise from the study area and by

interpreting high-resolution aerial imagery provided by the National Land Survey of Finland. We fitted a random forests classifier model (Breiman, 2001) with the vegetation classes treated as the response variable and used the spectral and index values from the satellite image stack of the corresponding locations as explanatory variables. The resulting model was then used to predict a spatially continuous vegetation classification map to the satellite imagery covering the entire study area. The internal out-of-bag cross-validation of the Random Forest model showed that 65.5% of the observations were correctly classified. The classification accuracy was highest in the deciduous shrub and wetland classes and lowest in barren. We used the same three PlanetScope images to calculate a Normalized Difference Vegetation Index (NDVI), a commonly used proxy for the amount of photosynthetic biomass.

S4.2 Snow cover duration

We used 94 PlanetScope images to construct a snow melting date map for the study area. The imagery is from April to September 2017-2018. We manually digitized a cloud mask for each image separately and created a training dataset from the masked images by selecting clear pixels manually and labeling them as land, water or snow. Then we fitted a Random Forests model (Breiman, 2001) and used the model to classify all the pixels in the 94 images and then reclassified the pixels either as snow or no-snow. We also predicted the class probabilities for each pixel. Then we stacked the binary snow images and created a pixel-wise binomial generalized linear model to determine the average melting date for each pixel separately (for more details on the methodology, see Niittynen and Luoto, 2018). The predicted class probabilities were used as weights in the models to give more weight for observations of high classification certainty.

S4.3 Topography

The topography variables were calculated from a digital terrain model (DTM; 2 m resolution) based on light detection and ranging (LiDAR) data from the open file service of the National Land Survey of Finland (2019). Following Kemppinen et al. (2021) we calculated elevation (m), topographic wetness index (twi), topographic position index (tpi), potential incoming solar radiation (kWh/m2), slope (radians), and aspect.

TWI was calculated using Equation TWI = ln (SCA / local slope), in which the SCA refers to the specific catchment area. We calculated the total catchment area (TCA) from a filled DTM using the multiple flow–direction algorithm (Freeman, 1991; Wang and Liu, 2006). Then we calculated the SCA assuming that flow width equals the grid resolution (2 m2) (i.e., SCA = TCA/2). We calculated the local slope and aspect following (Zevenbergen and Thorne, 1987). Aspect was transformed from a circular variable to a one-directional variable by taking a cosine of the aspect values. The final aspect variable represents northness, and values range from -1 (at due south) to 1 (at due north).

TPI describes the position of the plot on a topographic gradient at a given radius (Ågren et al., 2014). Positive values indicate that the plot is located on a ridge top, negative values that the plot is in a depression, and values close to zero that the plot is on a slope or flat ground. We used an unfilled DTM to calculate TPI using 5 m and 30 m radii (Kemppinen et al. 2021).

Radiation was calculated for June, July, and August, which constitute the growing season. We used the "Potential incoming solar radiation" tool in SAGA GIS v. 2.3.2. using the sky view–factor option (Böhner and Antonić, 2009). The position of the Sun was calculated for every fifth day (4-h interval). The atmospheric transmittance was calculated using the lumped atmosphere option.

S4.4 Surficial deposits

Surficial deposits were classified following (Kemppinen et al., 2018) and were based on field observations and aerial images (0.5 m resolution) provided by the National Land Survey of Finland. The classes represent peat and fluvial deposits, glacial till, and boulders and rock outcrops. Finally, the data was rasterized to a 2-meter resolution.

S4.5 ESA CCI land cover

ESA CCI land cover product at 300 m was used in Fig. 1 to visualize the cover of key vegetation types across the Arctic. To produce the map, we grouped some classes to the four key vegetation types in the tundra. We used the following classes to the shrub class: mosaic tree, shrub, or herbaceous cover (100 and 110), shrubland (120-122). Bare vegetation was based on sparse vegetation (150-153) and bare area (200-202) classes. Grasslands (class 130) were used to represent meadows. Wetlands were defined based on the flooded vegetation classes (160-180).

S5. Models

S5.1 Light response model details

All models were run in R with the package brms (Bürkner, 2017), which is an interface to the bayesian modelling platform Stan (Carpenter et al., 2017). We used weakly informative priors for the plot-specific intercept terms, informed by those reported in Williams et al. 2006 and Happonen et al 2022. The means and standard deviations for the priors were 1 and 2 for the logarithm of the ER (posterior mean: 0.65 and posterior standard deviation 0.13), 0 and 1 for the temperature effect on logarithm of the ER (posteriors: 0.02 and 0.01), 10 and 10 for maximum GPP (posteriors: 8.54 and 0.52), and 6.2 and 0.3 for the logarithm of the half-saturation parameter (posteriors: 5.92 and 0.06). We used logarithms for some parameters to normalize their error distribution. The model was fit with 4 MCMC chains, which were run for 2000 iterations each. The first 1000 iterations were discarded as warmup, leaving a total of 4000 samples of each parameter. We refer to these standardized variables as NEE, GPP, and ER. For more details on the model structure, see Happonen et al. 2022 (section 2.4.1) and the code light_response_model.R in Virkkala et al. 2023.

The Bayesian R_2 of the model was 0.96, all family- and population-specific mean parameters had an Rhat less than 1.03, and posterior predictive draws matched well with the observations, indicating model convergence and good predictive performance.

We used several models because they treat the data differently and thus, may individually detect different patterns and relationships in the data, and because the best performing models are not always the same for different response variables. These models were based on regression trees (generalized boosted regression models, GBM; random forest, RF) and kernel methods (support vector machine regression, SVM). GBM splits the data internally several times into training and evaluation data, and builds trees recursively using the information from the previous ones to improve the accuracy of the current tree in a process called boosting (Elith et al., 2008). RF bootstraps the data several times and samples the predictor variables as candidates at each split during the tree building, after which it builds an ensemble prediction (Breiman, 2001). SVM projects vectors into a high-dimensional space with a kernel function and fits an optimal hyperplane for the final model (Drake et al., 2006). The user can control these models by defining different parameters related to the trees, splits and variables, or the kernel properties.

For all the models, we assumed Gaussian error distribution. Parameters for machine learning models were tuned separately for each response variable with caret package (Kuhn, 2012) using the leave-one-out cross validation. We allowed the GBM to estimate the best model by testing different numbers of trees (50 to 1000), number of splits the model has to perform on a tree (5 to 13), learning rate (0.01 to 0.1), and the minimum number of observations in the trees' terminal nodes (2 to 5). For RF, we defined the number of variables randomly sampled as candidates at each split from four options (number of predictors divided by three to number of predictors divided by three plus six). And for SVM, we tested different values between 2^{-15} to 2^0 for the kernel width and values between 1 to 32 for the cost of constraints violation. The best model with the final set of parameters was chosen based on the lowest root mean square error (RMSE) values.

We used partial dependence plots and estimated variable importance of the predictors from each of the models (Section Machine learning models in the main text). The values on the y axis of each partial dependence plot can be interpreted as followed: yhat is conditional on other predictors in the model and their relationships with the predictor in the plot in question. Therefore, yhat values should not be directly compared with observed or predicted values, rather the patterns in yhat should be explored more generally. The x-axis represents the actual predictor values and can be used to infer, for example, conditions that lead to drastic changes in yhat (tipping points). For the variable importance analyses, we ran the permutation-based importance 100 times and averaged the results to reduce the error introduced by the randomness in the permutation procedure. A highly influential variable is represented by high variable importance scores.

We did not consider a cross-validation scheme that would have considered the potential spatial dependence of near-by GHG flux patterns as we did not find any spatial autocorrelation in the model residuals (see Fig. S8).

Uncertainty in GHG flux predictions was derived by bootstrapping (fractional resampling with replacement based on vegetation type classes) to subset the model training data into 30 different data sets, all of which had the same number of observations as the original data itself. These 30 data sets were then used to produce 30 individual predictions for a subset of the times with all three machine learning models and their ensemble for each response variable. We made these predictions to a subset of the time steps during the study period. We

randomly selected six time periods during each of the 5 weeks; three samples were selected from morning-noon (8-14) and three from the afternoon-evening (16-20). Thus, we used 30 out of 231 time periods in our uncertainty estimation, producing in total 18,000 files. We calculated an average GHG flux across the time periods for each of the bootstrapped model training data sets, and then took the 2.5 and 97.5 quantiles across these average GHG flux estimates to quantify prediction uncertainty. We also produced an uncertainty map by estimating the prediction interval for each pixel by subtracting the quantiles from each other. This method does not consider the uncertainty in predictors, but gives an estimate of how the distribution of the flux data influences our conclusions about the GHG budgets.

S5.3 Diagnosing model overfitting

Our machine learning models showed some signs of overfitting as demonstrated by the high overall model fit statistics that were higher compared to predictive performance statistics (Table S3). Ideally, the metrics derived from both validation strategies would be similar - i.e., the model would predict both the training data and new test data in a similar way, however, in our case the cross validation led to lower performance compared to training data-based estimates which were high (e.g., cross-validated predictive performance R^2 0.23 to 0.53 versus model-fit based R^2 0.33 to 0.94; excluding N_2O flux models that all had poor performance). Our models were potentially overfitting because we had a relatively small dataset (n=101), and machine learning models can learn the patterns in the data extremely well. This is a common problem and has been described in many other studies, and can be an indication of high uncertainties when extrapolating outside the data domain; however in our case we were not predicting to new areas outside our study domain. Moreover, the relationships we observed were logical and comparable to those observed in earlier studies.

Nevertheless, to decrease model overfitting, we tested models with less predictors (only three; biomass, soil moisture, and annual soil temperature), however this resulted in a similar mismatch with weaker performance, suggesting that decreasing the number of predictors does not help with model overfitting in our study (Table S3). We also tested generalized additive models (GAMs), a simpler and more traditional regression method (Hastie and Tibshirani 1987), where overfitting was clearly less pronounced (Table S3). We used gaussian distribution for all the fluxes and smooth responses for the continuous variables, and qq plots showed no issues with the residuals. However, when predicting to our full study domain, GAMs extrapolated significantly outside the observed flux data range (e.g., showing unrealistically low upscaled CH_4 fluxes with -33 mg C m^{-2} d^{-1} whereas the minimum in-situ fluxes were -5 g C m^{-2} d^{-1}), leading to highly uncertain upscaled fluxes (data not shown). Thus, we kept our original set of models and predictors but note that our models might suffer from overfitting and should thus be treated with caution when extrapolating to areas outside our study domain or different time periods.

Supplementary Tables and Figures

Table S1. The number of plots with field measurements for each variable used to explain GHG fluxes varied due to practical or analytical reasons. We used the spatial predictions of biomass, soil temperature, soil moisture, soil organic carbon stocks, and soil C/N to fill the gaps in the model training data. Specifically, we had the coordination information for all the plots, and extracted the missing information from the maps using the coordination information.

Variable	Number of plots
Chamber CO_2, CH_4, and N_2O flux measurements	101
Coordinates of the chamber plot	101
Vegetation type	101 (5413 for the environmental predictions)
Biomass	74 (110 for the environmental predictions)
Soil temperature	101 (139 for the environmental predictions)
Annual soil temperature	80 (139 for the environmental predictions)
Soil moisture	101 for CO_2 flux measurements, 100 for CH_4 and N_2O flux measurements (50 for the environmental predictions)
Soil organic carbon stock	99 (168 for the environmental predictions)
Soil C/N	98 (147 for the environmental predictions)

Table S2. In-situ GHG flux and environmental summaries and areal extents of the vegetation types. Negative values in NEE, CH_4 flux and N_2O flux indicate net sink of CO_2 from the atmosphere to the ecosystems. GPP and ER are given as positive values.

Vegetation type		Barren	Deciduous shrub	Evergreen shrub	Graminoid	Wetland
Number of plots		8	14	38	26	15
The proportional area of the vegetation type in the landscape		0.12	0.36	0.48	0.08	0.03
NEE g C m^{-2} d^{-1}	Mean	-1.18	-1.61	-0.91	-3.06	-2.44
	Median	-1.03	-1.21	-0.75	-3.06	-2.33
	25th q	-1.37	-2.15	-1.94	-3.99	-2.82
	75th q	-0.06	-0.75	0.06	-1.51	-1.79
GPP g C m^{-2} d^{-1}	Mean	3.67	5.47	4.32	7.57	5.41
	Median	2.81	4.72	3.82	7.13	5.42
	25th q	1.84	4.26	2.83	4.73	4.23
	75th q	3.94	5.27	5.37	9.74	7.22
ER g C m^{-2} d^{-1}	Mean	2.49	3.86	3.41	4.51	2.97
	Median	1.9	3.35	2.85	4.26	2.92
	25th q	1.63	3.03	2.31	3.62	2.28
	75th q	2.58	4.07	4.54	5.85	3.65
CH_4 flux mg C m^{-2} d^{-1}	Mean	-1.27	-2.41	-1.57	-1.09	46.55

	Median	-1.32	-2.53	-1.35	-0.59	23.16
	25th q	-1.63	-2.88	-1.83	-1.54	3.56
	75th q	-0.45	-2	-0.66	-0.13	43.68
N_2O flux μg N m^{-2} d^{-1}	Mean	-0.24	-26.96	-4.25	-2.5	12.77
	Median	5.73	-8.27	-3.18	-3.5	7.64
	25th q	-7	-41.05	-19.41	-38.82	-3.82
	75th q	10.98	2.7	28.8	31.5	25.46
Soil moisture VWC %	Mean	11.6	12.16	14.24	36.45	67.65
	Median	14.45	10.95	10.75	36.5	65.2
	25th q	6.15	8.25	7.7	26.9	63.85
	75th q	16.1	16.12	18.65	44.9	70.25
Soil temperature °C	Mean	11.81	10.14	10.59	11.37	13.77
	Median	11.6	10.8	11	11.9	14.2
	25th q	10.1	8.25	8.65	9.85	13.6
	75th q	13.88	11.8	12.43	13.28	14.75
Annual soil temperature °C	Mean	0.32	1.82	0.9	1.72	1.78
	Median	-0.16	1.81	1.33	1.82	1.78
	25th q	-0.87	1.42	-0.1	1.24	1.7
	75th q	1.99	2.21	1.84	2.26	1.86

Soil organic carbon stock kg C m^{-2}	Mean	3.91	5.76	5.42	10	21.69
	Median	3.43	5.8	3.92	6.41	20.06
	25th q	3.2	3.74	2.92	4.13	19.01
	75th q	4.17	7.09	7.44	13.39	24.56
Dissolved organic carbon mg C g^{-1}	Mean	1.73	2.01	1.98	1.61	2.84
	Median	1.95	2	1.86	1.59	2.88
	25th q	1.07	1.71	1.45	0.99	1.91
	75th q	2.32	2.72	2.52	1.95	3.45
Soil C/N	Mean	25.3	22.44	24.32	14.49	14.68
	Median	25.26	21.85	24.66	14.04	14.19
	25th q	22.4	19.42	20.47	13.39	13.22
	75th q	29.67	24.67	28.31	15.67	15.23
Soil pH	Mean	4.69	4.54	4.54	5.2	5.55
	Median	4.77	4.53	4.57	5.11	5.46
	25th q	4.44	4.38	4.32	4.85	5.3
	75th q	4.93	4.72	4.8	5.42	5.72
Soil nitrogen stock kg N m^{-2}	Mean	0.32	0.41	0.37	0.81	1.56
	Median	0.31	0.43	0.29	0.57	1.52
	25th q	0.27	0.25	0.19	0.32	1.15

	75th q	0.32	0.49	0.55	1.06	1.8
Biomass g m^{-2}	Mean	231.71	328.73	332.59	105.16	142.36
	Median	172.59	299.61	242.82	94.79	139.77
	25th q	86.21	187.82	171.37	69.91	120.31
	75th q	259.84	421.92	313.58	138.74	150.75

Table S3. Model fit and predictive performance statistics of the final individual models (generalized boosted regression models, GBM; random forest, RF; and support vector machine regression, SVM) and models trained with three predictors, including the generalized additive model (GAM) estimates as a comparison against machine learning-based estimates.

			Model fit statistics				Predictive performance statistics (cross validation-based)		
Model structure	Machine learning model	Response variable	RMSE	R^2	MAE		RMSE	R^2	MAE
Final full model									
	GBM	NEE g C m^{-2} d^{-1}	1.1	0.81	0.85		1.78	0.34	1.35
	GBM	GPP g C m^{-2} d^{-1}	1.26	0.89	0.94		2.32	0.5	1.76
	GBM	ER g C m^{-2} d^{-1}	0.7	0.87	0.56		1.28	0.38	1
	GBM	CH$_4$ flux mg C m^{-2} d^{-1}	12.88	0.85	4.4		21.49	0.44	7.71

11

	GBM	N$_2$O flux mg N m^{-2} d^{-1}	0.07	0.54	0.04		0.08	0	0.05
	RF	NEE g C m^{-2} d^{-1}	0.84	0.92	0.61		1.78	0.34	1.32
	RF	GPP g C m^{-2} d^{-1}	1.04	0.94	0.77		2.36	0.48	1.81
	RF	ER g C m^{-2} d^{-1}	0.61	0.93	0.48		1.33	0.35	1.06
	RF	CH$_4$ flux mg C m^{-2} d^{-1}	12.06	0.85	4		22.03	0.42	7.89
	RF	N$_2$O flux mg N m^{-2} d^{-1}	0.05	0.86	0.03		0.08	0	0.05
	SVM	NEE g C m^{-2} d^{-1}	1.51	0.57	1.05		1.77	0.35	1.37
	SVM	GPP g C m^{-2} d^{-1}	1.91	0.68	1.32		2.27	0.53	1.73
	SVM	ER g C m^{-2} d^{-1}	0.94	0.69	0.63		1.32	0.38	1.02
	SVM	CH$_4$ flux mg C m^{-2} d^{-1}	24.69	0.33	6.52		25.59	0.23	7.67
	SVM	N$_2$O flux mg N m^{-2} d^{-1}	0.08	0.03	0.05		0.08	0.09	0.05
	GAM	NEE g C m^{-2} d^{-1}	1.52	0.52	1.15		1.97	0.26	1.53
	GAM	GPP g C m^{-2} d^{-1}	1.93	0.65	1.4		2.46	0.45	1.81

	GAM	ER g C m^{-2} d^{-1}	1.03	0.61	0.78		1.3	0.39	1.01
	GAM	CH$_4$ flux mg C m^{-2} d^{-1}	14.6	0.74	5.3		29.41	0.22	11.2
	GAM	N$_2$O flux mg N m^{-2} d^{-1}	0.07	0.19	0.04		0.09	0	0.05
Model with 3 predictors	GBM	NEE g C m^{-2} d^{-1}	1.24	0.74	0.92		1.8	0.32	1.32
	GBM	GPP g C m^{-2} d^{-1}	1.6	0.81	1.16		2.47	0.42	1.86
	GBM	ER g C m^{-2} d^{-1}	0.72	0.83	0.58		1.35	0.33	1.09
	GBM	CH$_4$ flux mg C m^{-2} d^{-1}	6.18	0.96	2.45		19.39	0.55	8.48
	GBM	N$_2$O flux mg N m^{-2} d^{-1}	0.07	0.68	0.04		0.08	0	0.05
	RF	NEE g C m^{-2} d^{-1}	0.84	0.89	0.61		1.81	0.32	1.34
	RF	GPP g C m^{-2} d^{-1}	1.16	0.9	0.85		2.51	0.41	1.89
	RF	ER g C m^{-2} d^{-1}	0.62	0.91	0.49		1.4	0.27	1.12
	RF	CH$_4$ flux mg C m^{-2} d^{-1}	11.23	0.86	3.72		20.08	0.51	7.17

	RF	N_2O flux mg N m^{-2} d^{-1}	0.04	0.87	0.02		0.08	0.01	0.05
	SVM	NEE g C m^{-2} d^{-1}	1.78	0.38	1.25		1.82	0.34	1.3
	SVM	GPP g C m^{-2} d^{-1}	2.44	0.47	1.66		2.61	0.38	1.85
	SVM	ER g C m^{-2} d^{-1}	1.23	0.44	0.9		1.31	0.37	1
	SVM	CH_4 flux mg C m^{-2} d^{-1}	22.1	0.52	6.74		23.55	0.39	7.65
	SVM	N_2O flux mg N m^{-2} d^{-1}	0.08	0.04	0.04		0.08	0.01	0.05
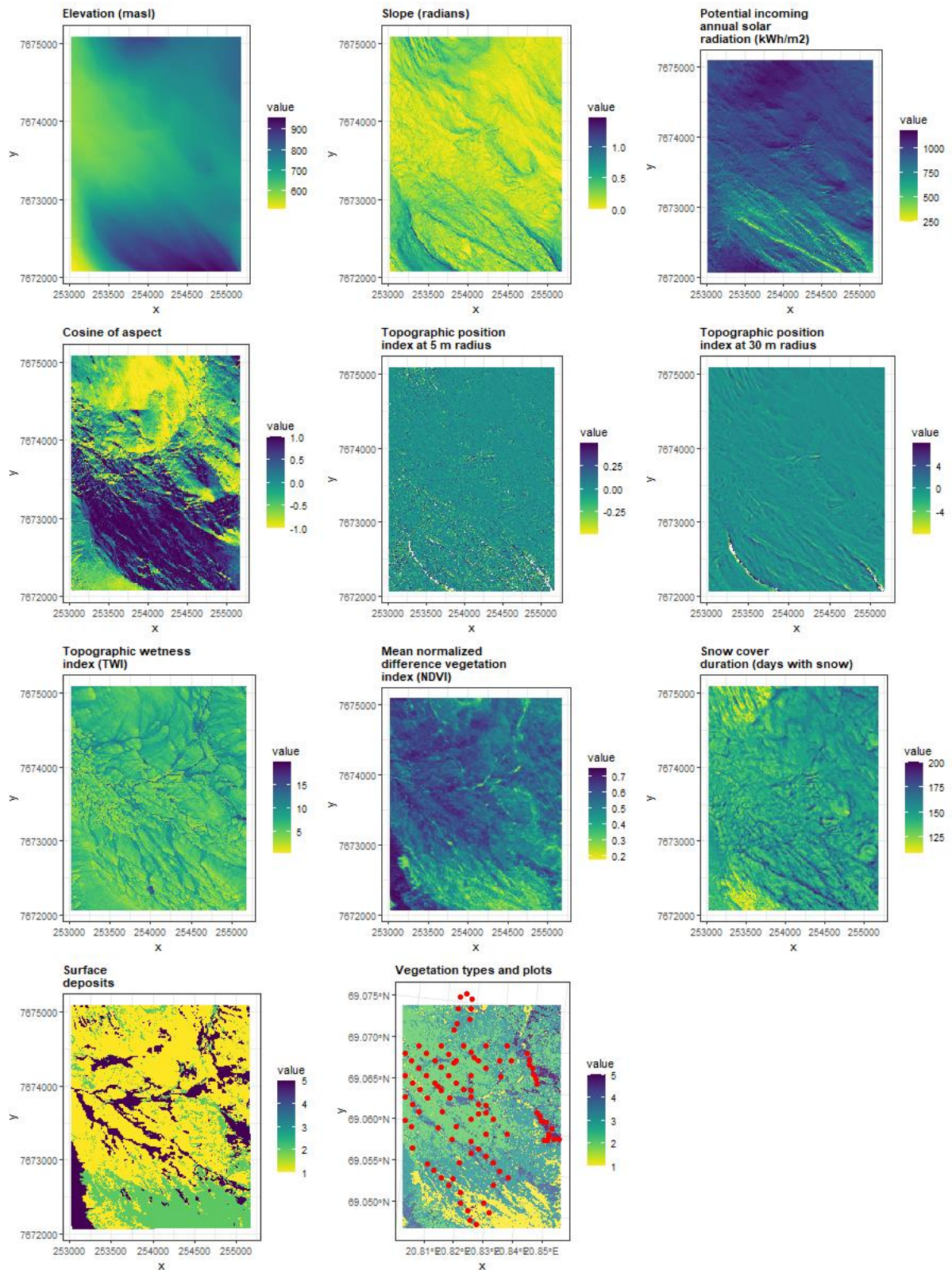
Fig. S1. Remotely sensed data used in machine learning models predicting environmental conditions. The first seven variables are based on light detection and ranging (LiDAR) data, whereas ndvi and snow are based on Planet Imagery and surficial deposits on field observations and high-resolution (0.5 m) aerial images. LiDAR data and aerial images are provided by the National Land Survey of Finland (2019). Surface deposits

representing soil types are 1=glacial till, 2=boulders, rock outcrops, bare soils, 5=peat or fluvial material. Vegetation types are 1=barren, 2=deciduous shrub, 3=evergreen shrub, 4=meadow, 5=wetland.
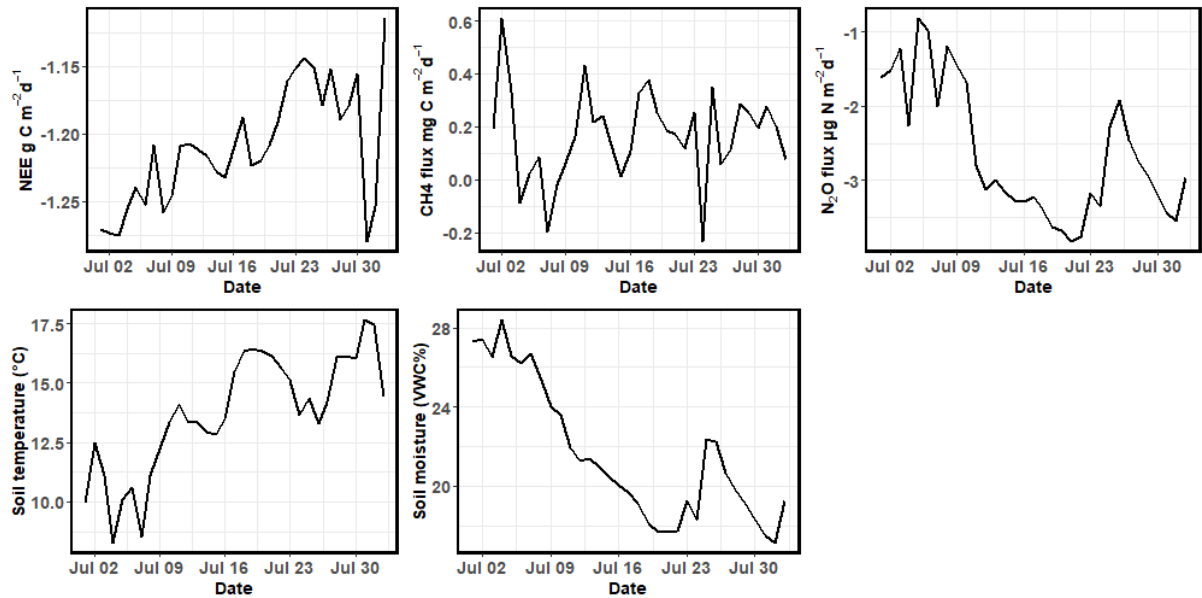
Fig. S2. Time series of day-time (8 am to 8 pm) average daily GHG fluxes and soil temperature and moisture conditions based on the upscaled maps.
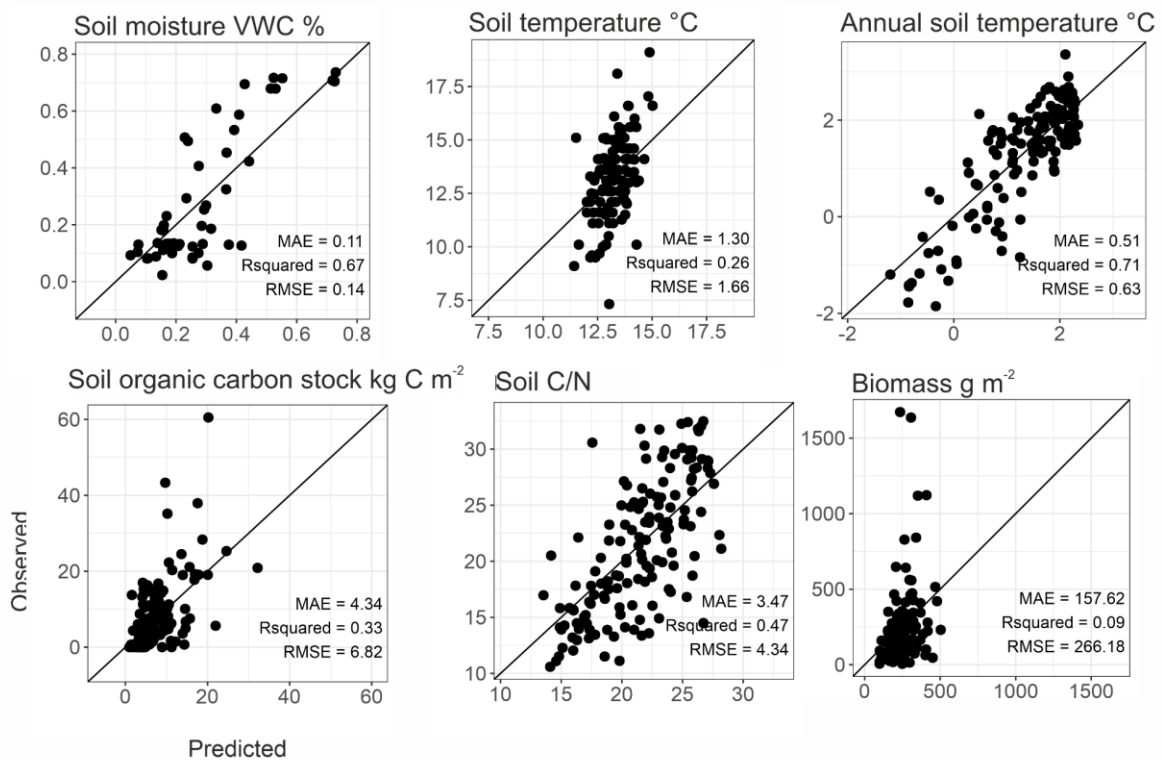
Fig. S3. Predictive performance of models used to predict environmental conditions. The soil moisture and

temperature models represent the performance of the model from July 16th, 2018 at 12 pm, but the performance remains somewhat similar throughout the growing season.
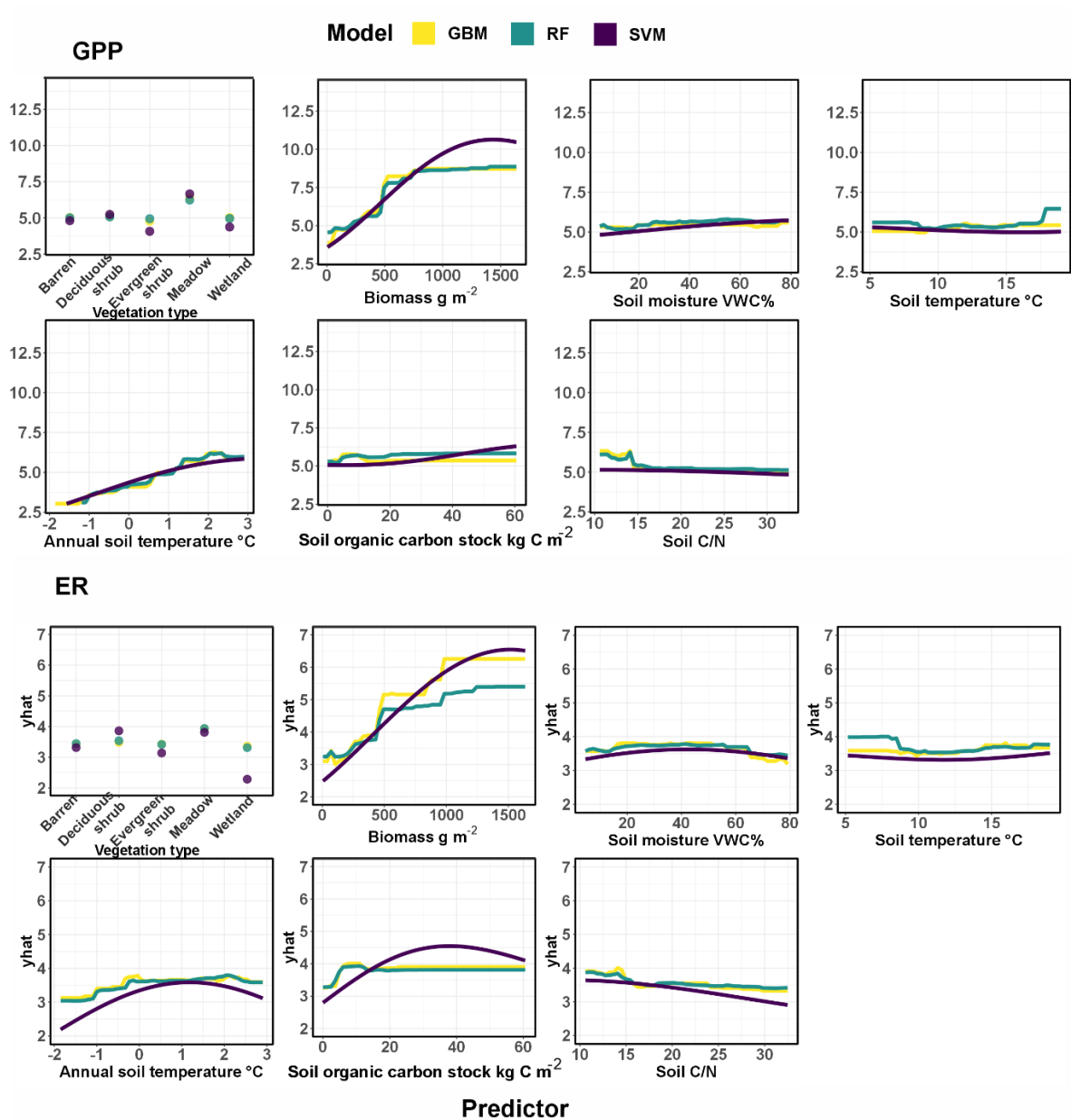
Fig. S4. The partial dependence plots for GPP and ER.
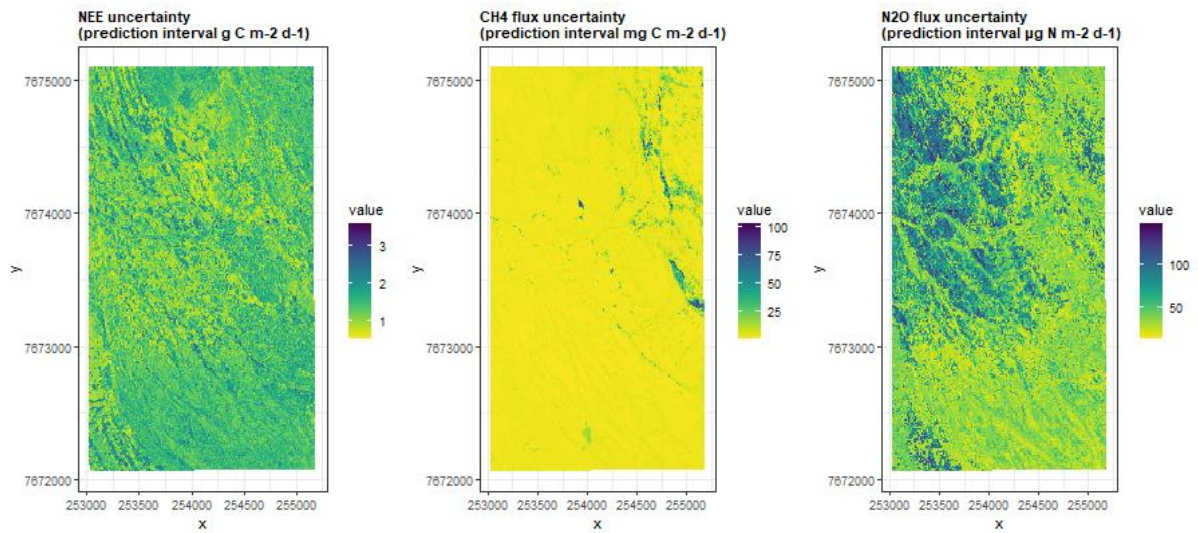
Fig. S5. Uncertainty in GHG flux predictions, estimated by bootstrapping the model training data and creating 2,700 different predictions for each flux. High prediction interval estimates reflect large differences in model predictions across the model ensembles.
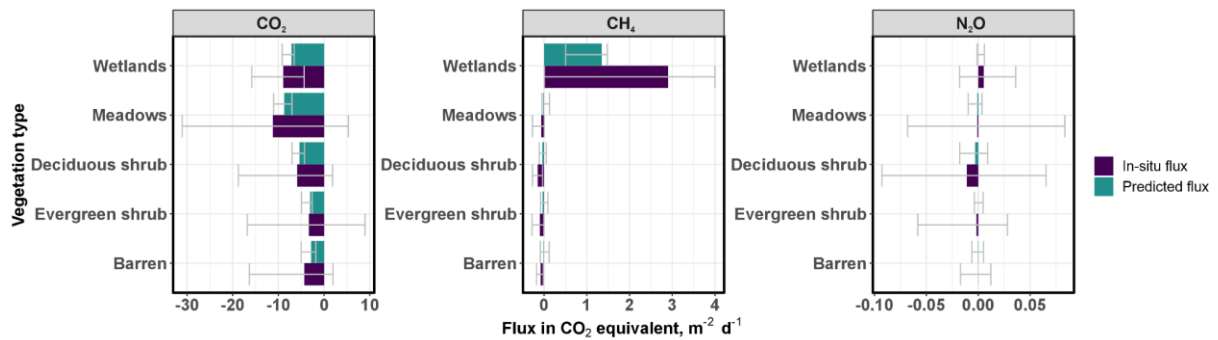
Fig. S6. Growing season mean and percentile (0.025 and 0.975) GHG flux contributions based on in-situ data and upscaled flux predictions, averaged across the entire study period (only daytime variability between 8 am and 8 pm considered) and across vegetation types. Note that the scale for the x axis is different for each gas species, and that the uncertainties in in-situ versus predicted mean fluxes cannot be directly compared with each other. The uncertainty in wetland CH_4 continues up to 11.2 but was cropped for visualization purposes. CO_2 equivalents were calculated using the sustained GWP approach.
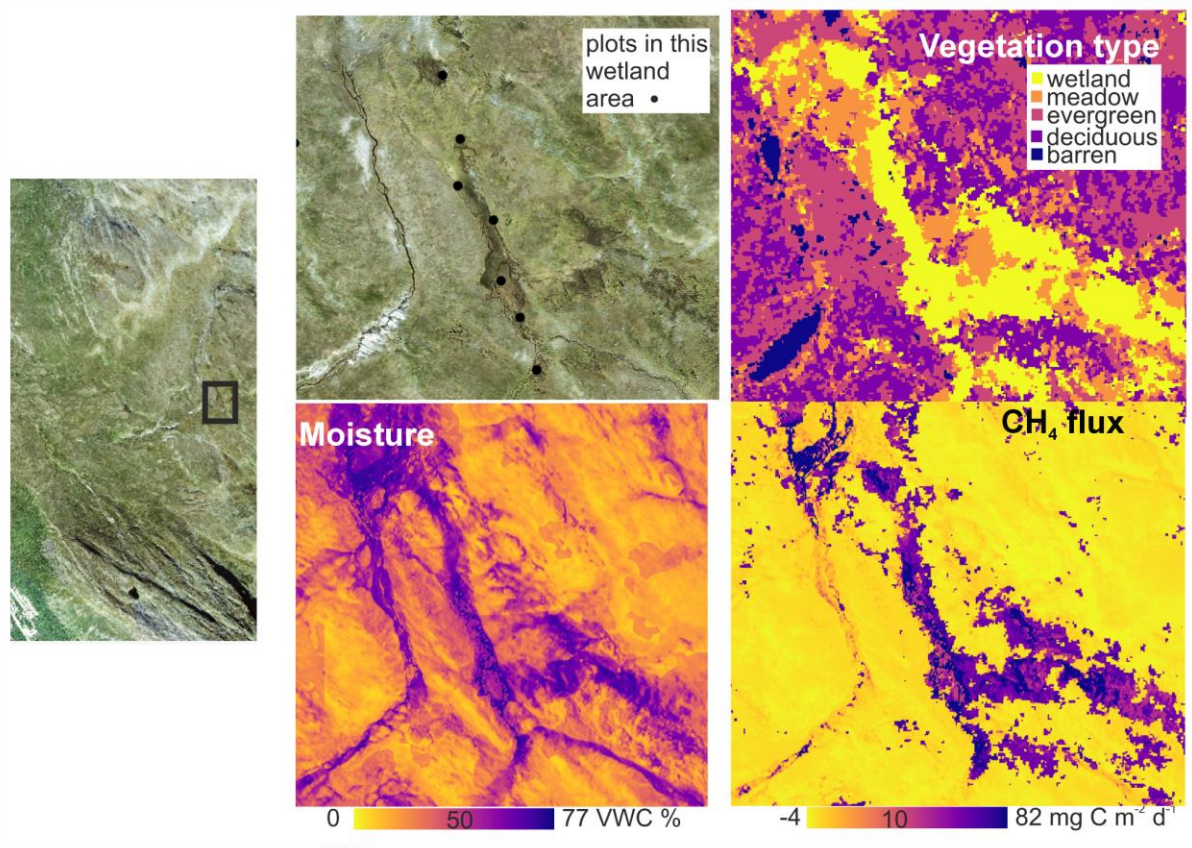
Fig. S7. A visualization of the fine-scale variation in soil moisture and the upscaled CH$_4$ flux within and across vegetation types, and in particular within the wetland class. Some parts of the wetland are wet (i.e. soil moisture around 70 % VWC) while others are drier (i.e. close to 50 % VWC), and these differences are important for the resulting CH$_4$ flux patterns. These are zoomed-in figures from Fig. 1 and 8.
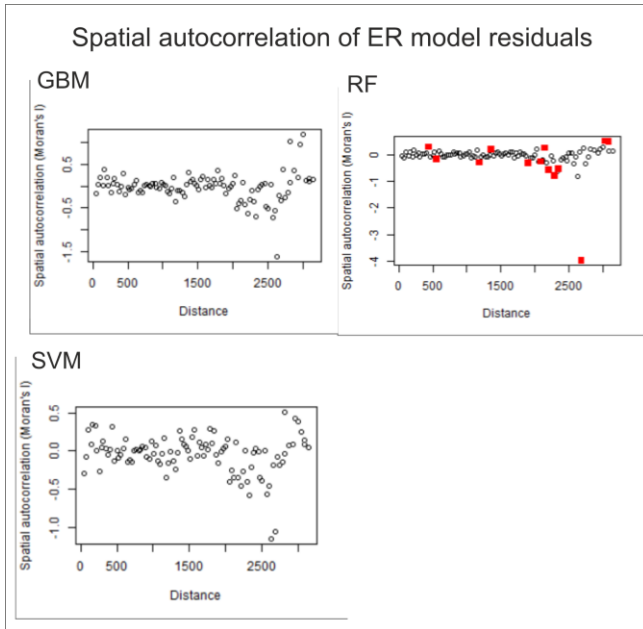
Fig. S8. Spatial autocorrelation of model residuals for ER, as an example. As indicated by the lack of clear high or low index values on the y axis, the models did not suffer from spatial autocorrelation that should have been acknowledged in the model evaluation phase.
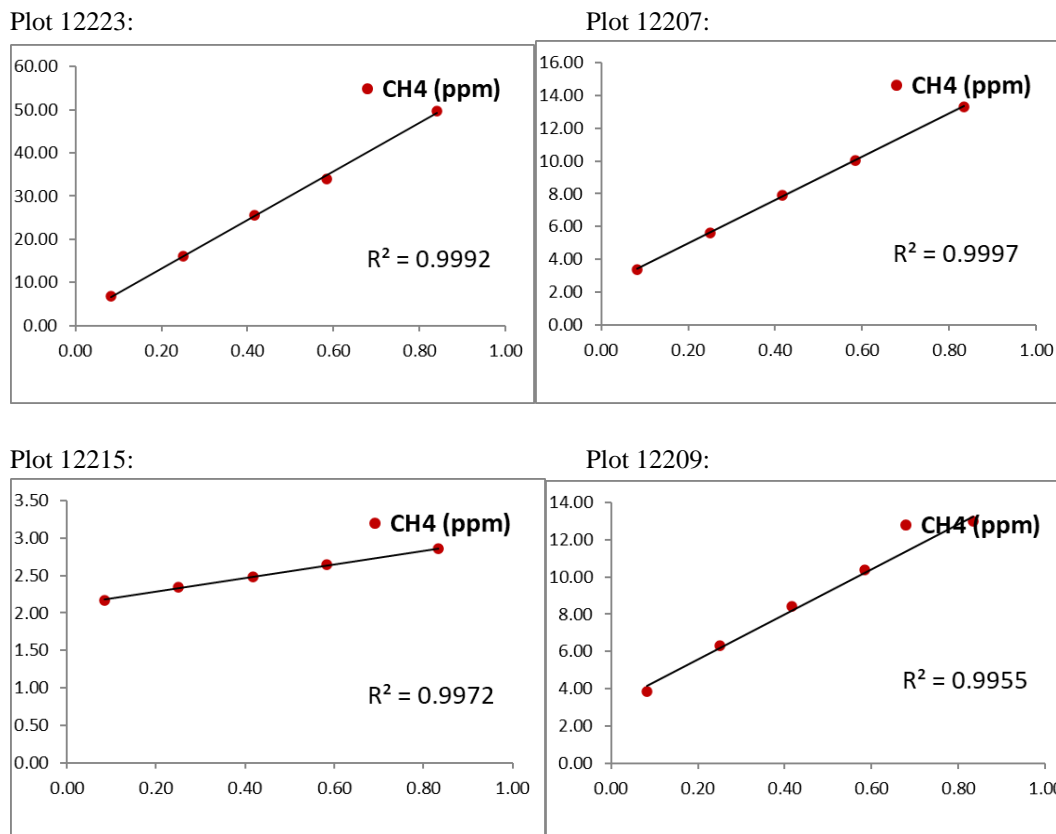

Fig S9. Example graphs showing the development of CH4 concentrations at some of the wetland sites during the 50-minute chamber closure.
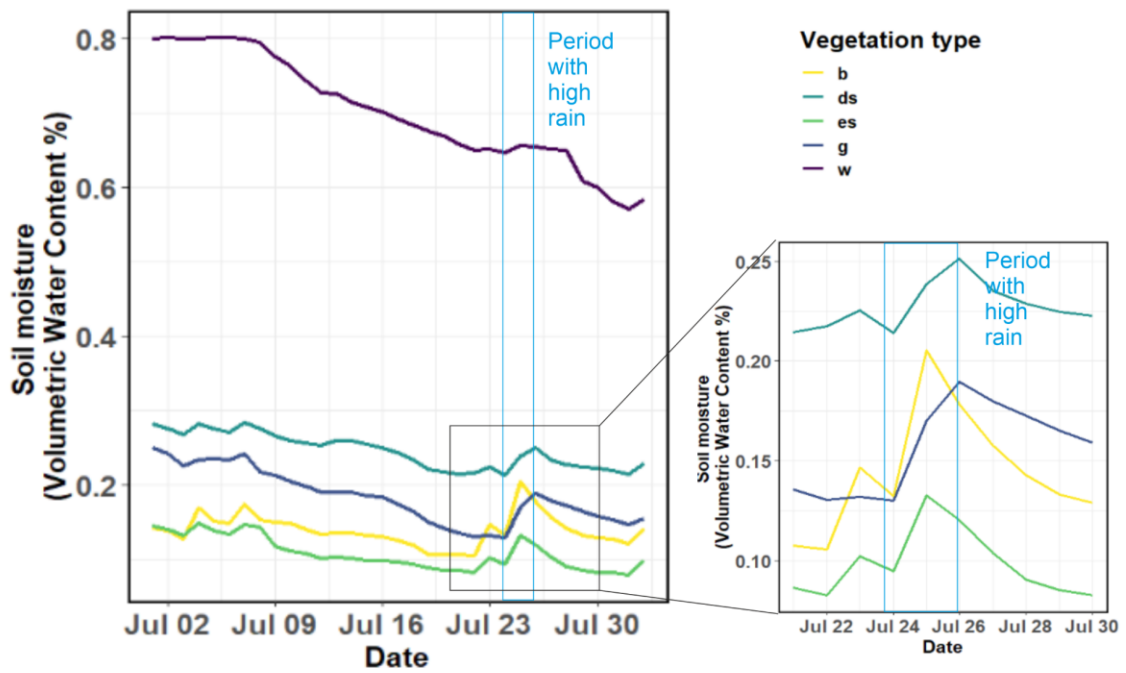
Fig S10. A figure showing the soil moisture variation during the study period from 5 example plots representing the vegetation types, and an example demonstration of a period with high rain. The subplot shows how soil moisture changes after the rain. Vegetation types are b=barren, ds=deciduous shrub, es=evergreen shrub, g=meadow, w=wetland.

References

Ågren, A. M., Lidberg, W., Strömgren, M., Ogilvie, J., and Arp, P. A.: Evaluating digital terrain indices for soil wetness mapping – a Swedish case study, https://doi.org/10.5194/hess-18-3623-2014, 2014.

Böhner, J. and Antonić, O.: Chapter 8 Land-Surface Parameters Specific to Topo-Climatology, https://doi.org/10.1016/s0166-2481(08)00008-1, 2009.

Breiman, L.: Random Forests, Mach. Learn., 45, 5–32, 2001.

Bürkner, P.-C.: brms: An R Package for Bayesian Multilevel Models Using Stan, Journal of Statistical Software, Articles, 80, 1–28, 2017.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A.: Stan: A probabilistic programming language, J. Stat. Softw., 76, 2017.

Drake, J. M., Randin, C., and Guisan, A.: Modelling ecological niches with support vector machines, J. Appl. Ecol., 43, 424–432, 2006.

Elith, J., Leathwick, J. R., and Hastie, T.: A working guide to boosted regression trees, J. Anim. Ecol., 77, 802–813, 2008.

Freeman, T. G.: Calculating catchment area with divergent flow based on a regular grid, Comput. Geosci., 17, 413–422, 1991.

Hastie, T: and Tibshirani, R.: Generalized Additive Models: Some Applications, Journal of the American Statistical Association, 82:398, 371-386, 1987.

Honaker, J., King, G., Blackwell, M., and Others: Amelia II: A program for missing data, J. Stat. Softw., 45, 1–47, 2011.

Kemppinen, J., Niittynen, P., Riihimäki, H., and Luoto, M.: Modelling soil moisture in a high-latitude landscape using LiDAR and soil data, Earth Surf. Processes Landforms, 43, 1019–1031, 2018.

Kuhn, M.: The caret package, R Foundation for Statistical Computing, Vienna, Austria. URL https://cran. r-project. org/package= caret, 2012.

Marushchak, M. E., Kerttula, J., Diáková, K., Faguet, A., Gil, J., Grosse, G., Knoblauch, C., Lashchinskiy, N., Martikainen, P. J., Morgenstern, A., Nykamb, M., Ronkainen, J. G., Siljanen, H. M. P., van Delden, L., Voigt, C., Zimov, N., Zimov, S., and Biasi, C.: Thawing Yedoma permafrost is a neglected nitrous oxide source, Nat. Commun., 12, 7107, 2021.

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

Voigt, C., Marushchak, M. E., Lamprecht, R. E., Jackowicz-Korczyński, M., Lindgren, A., Mastepanov, M., Granlund, L., Christensen, T. R., Tahvanainen, T., Martikainen, P. J., and Biasi, C.: Increased nitrous oxide emissions from Arctic peatlands after permafrost thaw, Proc. Natl. Acad. Sci. U. S. A., 114, 6238–6243, 2017.

Wang, L. and Liu, H.: An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling, Int. J. Geogr. Inf. Sci., 20, 193–213, 2006.

Williams, M., Street, L. E., van Wijk, M. T., and Shaver, G. R.: Identifying Differences in Carbon Exchange among Arctic Ecosystem Types, Ecosystems, 9, 288–304, 2006.

Zevenbergen, L. W. and Thorne, C. R.: Quantitative analysis of land surface topography, Earth Surf. Processes Landforms, 12, 47–56, 1987.