Supplement of Biogeosciences, 20, 545–572, 2023
https://doi.org/10.5194/bg-20-545-2023-supplement

Biogeosciences Open Access EGU

Supplement of

Meteorological responses of carbon dioxide and methane fluxes in the terrestrial and aquatic ecosystems of a subarctic landscape

Lauri Heiskanen et al.

Correspondence to: Lauri Heiskanen (lauri.heiskanen@helsinki.fi)

S.1 Land cover classification

The land cover classification into ten land cover types (LCTs) was conducted following a geographic object-based image analysis approach (for class descriptions see Räsänen and Virtanen 2019). A 0.5 m spatial resolution aerial orthophoto acquired by National Land Survey of Finland on 27 June 2016 was segmented with an average and minimum segment size of 50 and 10 m^2, respectively, using full lambda schedule segmentation in Erdas Imagine (Hexagon Geospatial, Stockholm, Sweden). For each segment, a total of 55 spectral, topographic and vegetation structure features were calculated from the following remote sensing data sources: aerial orthophoto, aerial laser scanning data by National Land Survey of Finland, WorldView-2 satellite imagery (Digital Globe, Westminster, CO, USA), and four PlanetScope satellite images (Planet Labs, San Francisco, CA, USA). A training data for the classification was obtained with transect-based field inventory (see Räsänen and Virtanen 2019 for details) and visual interpretation of drone and aerial orthophotos. The data were classified into land cover classes with random forest classification (Breiman 2001) with package randomForest in R (Liaw and Wiener 2002). The classification accuracy was calculated using a plot-based field inventory as validation data (see Räsänen and Virtanen 2019 for details) and the overall classification accuracy was 69 %.

For the purposes of this study, five open peatland microforms were merged into one open peatland LCT class and two forest classes were merged into one forest LCT. Conversely, the pine bog LCT was divided into two treed and sparsely treed pine bog LCTs by utilising aerial laser scanning data. The pine bog area that had mean tree height < 0.5 m was classified into sparsely treed class and the area with mean tree height > 0.5 m into treed pine bog. The treed pine bog ecosystem is mainly located between the forest and fen continuum, while sparsely treed pine bog is observed as patches surrounded by the fen ecosystem. Additionally, the lakes in the landscape were divided into two categories based on their sediment type, namely to organic and mineral sediment lakes. Both information about the water flow in the area and the notion that sandy and peaty bottoms have a different albedo were used to distinguish between the lake types. Organic sediment lakes are located downstream of the surrounding peatland areas, while the mineral sediment lakes are found adjacent to sandy podzol soils, through which the water flows to the lakes. Due to the shallowness of the water bodies in the area, the sediment type could be visually distinguished from aerial orthophoto and Planetscope satellite images, with the sandy bottoms showing up as light-coloured and the peaty bottoms as dark-coloured segments.

S.2 Soil and sediment samples and analyses

In the pine bog and fen ecosystems, peat samples of approximately 5 cm × 5 cm × 5 cm were cut out of the soil at 0–5 cm (i.e. straight under the litter layer) and 15–20 cm depth in June 2017. The samples were dried for bulk density estimates and ground in a ball mill to measure their C and N concentrations using a CNS-2000 analyser (LECO Corporation, Saint Joseph, MI,

USA). Sample C and N contents (g dm^{-3}) were then calculated using the bulk density and C and N concentrations of the samples. The pH was measured in the field using water collected from the bottom of a 30 cm deep hole.

In the pine forest ecosystem, soil cores (diameter 3 cm) were collected from the organic layer, eluvial layer, top and bottom of illuvial layer, and 50 cm and 100 cm below the organic layer in June 2017. The cores from the organic and eluvial layer were vertical and their length was the depth of the layer. The other cores were horizontal and their length was 5 cm. Soil cores were dried, weighed and their bulk density and C and N content analysed as described above for peat samples. Soil C and N contents were corrected by stone volume, i.e. the estimates are for entire soil layers including the stones. For pine forest soil, the pH was measured in the laboratory in water solution with 10 ml distilled water and 0.5 g (organic layer sample) or 2 g (mineral layer sample at 30 cm depth) fresh mass soil.

In the lake ecosystem, 20 cm long surface cores of sediments were collected in March 2017 (MS lake southern basin), August 2017 (MS lake northern basin) and June 2018 (OS lake) using a HTH gravity corer (tube diameter 8.5 cm). The MS lake northern basin cores were cut into 0–5, 5–10, 10–15 and 15–20 cm layers, dried, weighted and their bulk density and C and N content analysed as described above for peat samples. The other surface cores were subsampled immediately after retrieval at 0.25 to 1 cm intervals and placed into Minigrip® bags, and subsequently transported to the laboratory and freeze-dried. Elemental contents of these lake sediments were analysed at 2 cm intervals with an elemental analyser at the dating laboratory of LUOMUS, University of Helsinki.

S.3 Eddy covariance data coverage

The EC data coverage especially at the upland pine forest site was low. This was mainly due to necessary wind sector exclusions and unavoidable equipment failures. The pine forests within the area are patchy and grow mostly on narrow eskers. Thus, the EC tower had to be installed on an edge of the forested area to guarantee sufficient (> 80 %) flux footprint coverage over the forest-facing wind sector. This placement led to exclusion of 48 % of the EC flux data, as the other wind sectors covered peatland and lake ecosystems. While our data coverage is admittedly low at this site, it should be noted that, in practice, EC data sets are generally far from being complete. For example, in the widely used global Fluxnet2015 database, comprising 1532 site-years, 68 % of the data are missing (Pastorello et al., 2020; Zhu et al., 2022). Our fen data had a better coverage than this.

Furthermore, the annual balance is most affected by the flux uncertainties when the flux magnitude is largest, i.e. during the growing season (Richardson and Hollinger, 2007). The longest data gaps in the forest data took place in October 2017 – February 2018, May–June 2018 and April–May 2019, i.e. outside the peak growing season, and fortunately the data coverage during the growing seasons was better than the annual average (Fig. S1).

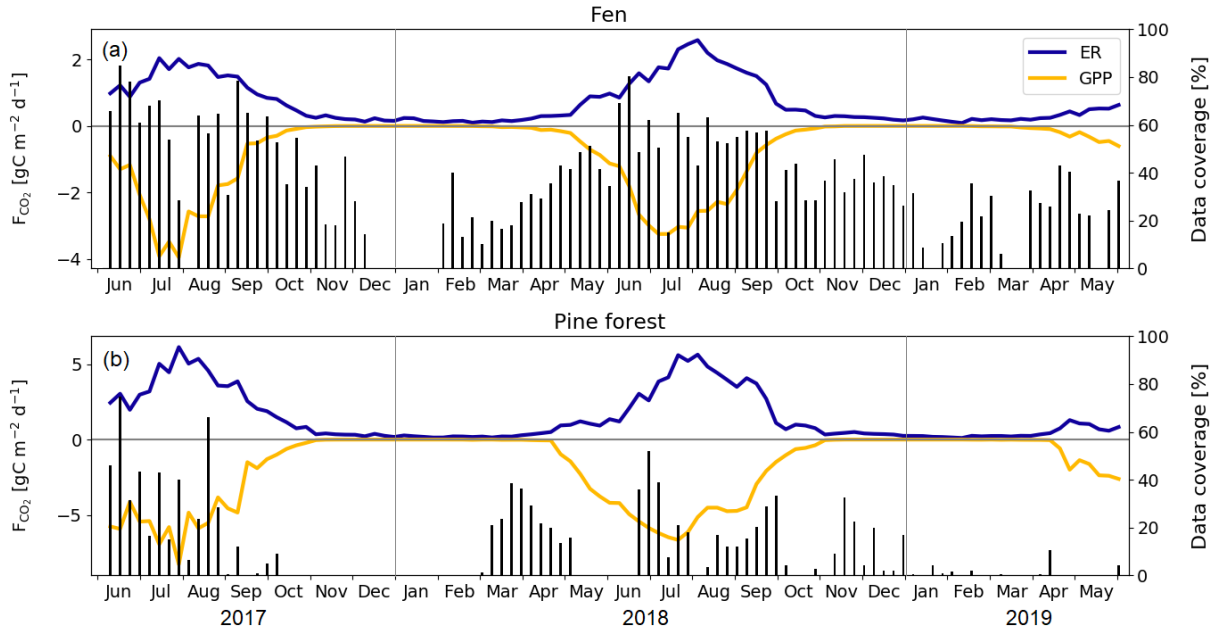

Figure S1. Fen (a) and pine forest (b) EC data coverage and weekly average ER and GPP fluxes.

Table S1. Eddy covariance data coverage after each quality control filtering step.

	Gaps after each step		
Filtering criterion	Pine forest CO_2	Fen CO_2	Fen CH_4
1 Equipment failure	39 %	13 %	14 %
2 Wind direction	73 %	30 %	31 %
3 No. of recorded data, spikes and variance outliers	74 %	31 %	32 %
4 Flux stationarity	82 %	51 %	54 %
5 u_* threshold	85 %	57 %	58 %
6 Gas mixing ratio	89 %	64 %	69 %

To assess the implications of the low data coverage on the estimation of annual balances, we created continuous, synthetic time series, inserted the original gaps into them and gap-filled the resulting time series using XGBoost. We made a total of 50 synthetic data sets of 30 min CO_2 fluxes whose statistical characteristics were similar to the original measured data (Fig. S2). The synthetic data sets were generated using an artificial neural network (ANN). We used a sequential model with four hidden layers and three different activation functions: linear, hyperbolic tangent and rectified linear activation. The mean squared error (MSE) was used as the loss function. The ANN was implemented using the Keras library (Keras team, 2022). We utilized

all the available measurement data to train the ANN and, after modelling for all 30 min periods, added noise to the modelled NEE. This was done by binning the residuals in 10 bins based on NEE and by selecting a residual for each 30 min modelled NEE randomly from the appropriate bin. The original data gaps were placed in the synthetic data sets and the resulting time series were gap-filled with the XGBoost method. We found that our gap-filling procedure resulted in an accurate replication of the CO_2 balances calculated from the full time series, even though the data coverage was low (Figs. S3 and S4). Therefore, we concluded that XGBoost can be used to obtain reliable gap-filling results also for the original measurement data.

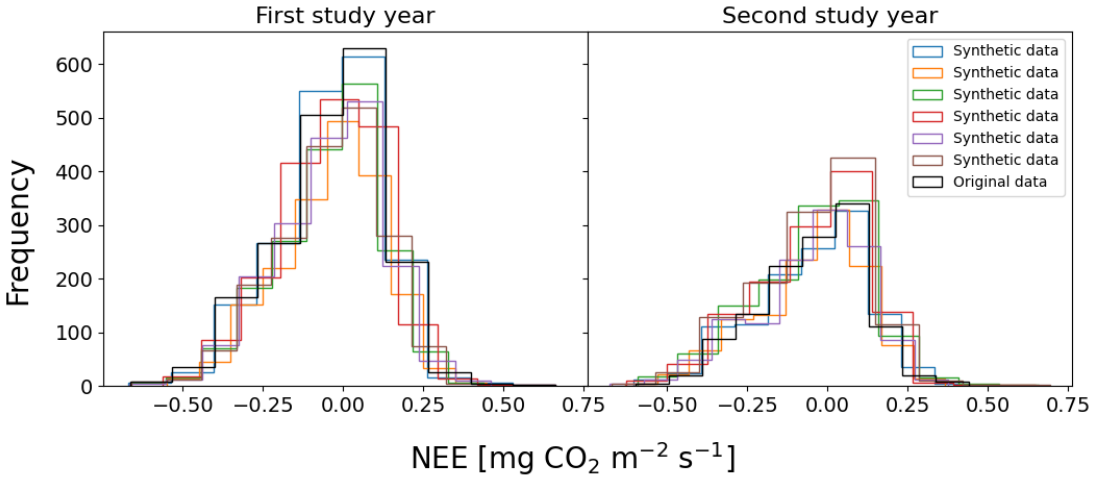

Figure S2. Distribution of the original, measured CO_2 flux data and representative samples of synthetic data used for the validation of the gap-filling method. The synthetic data include here only the values for which a corresponding measurement is available.

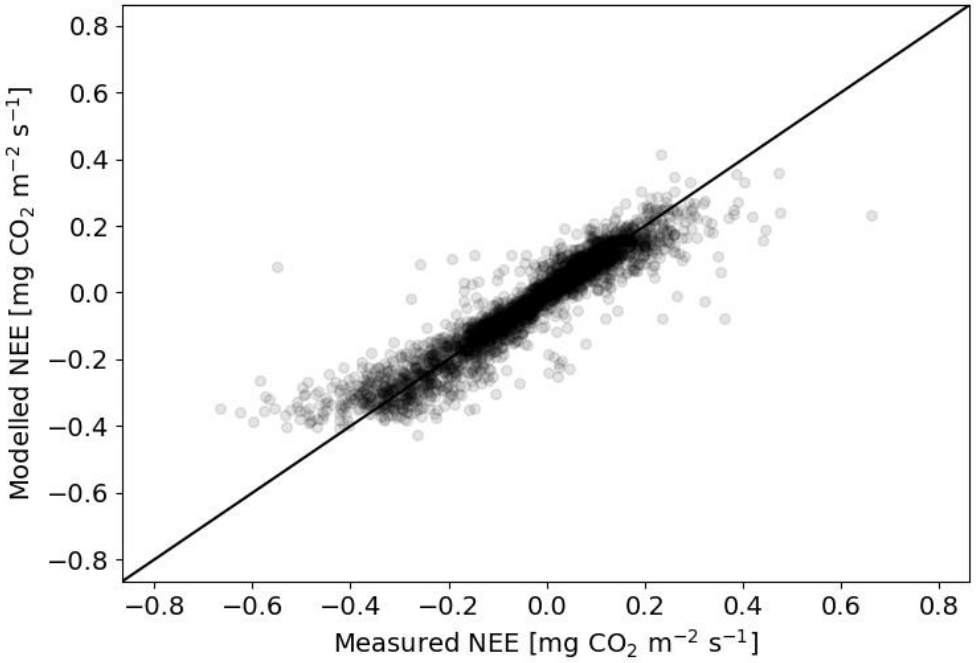

Figure S3. The 30 min NEE measured in the pine forest versus the corresponding NEE modelled using XGBoost. The modelled values are from the 10-fold cross validation results. R^2 = 0.88 ± 0.02, MSE = 0.003 ± 0.0006 mg CO_2 m^{-2} s^{-1}.

85

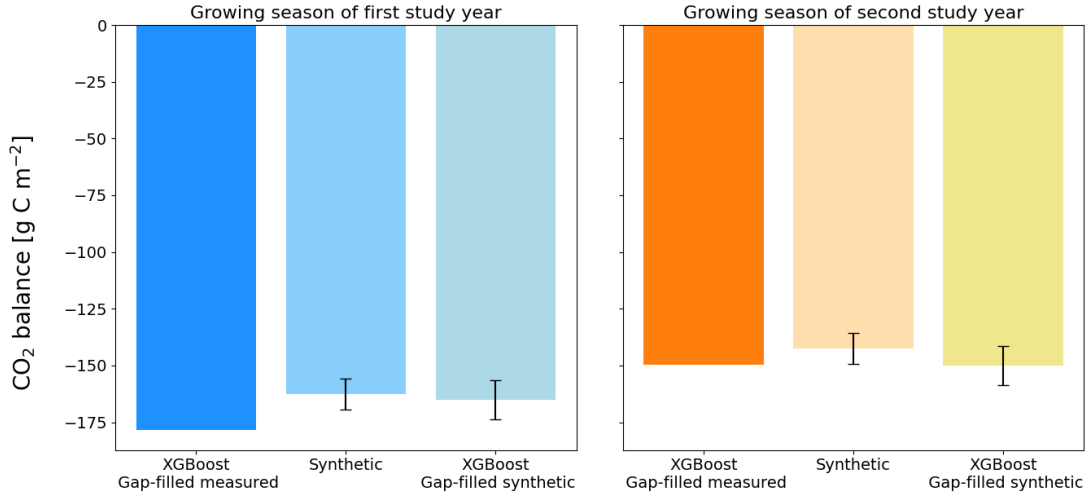

Figure S4. Growing season CO_2 balances estimated from the measured data using XGBoost for gap-filling (left), the mean and standard deviation of the balances of full synthetic time series (middle) and the mean and standard deviation of synthetic data with the original gaps gap-filled using XGBoost (right). The synthetic data were generated 50 times.

S.4 Vegetation sampling and measurement methods

Biomass and leaf area index (LAI) were calculated and modelled based on field measurements. Specifically, field data was collected in 130 circular plots with 5 m radius (71 random plots, 59 plots in transects). In these plots, percentage coverage and average height of different plants and plant functional types (PFT) in ground and field layer were measured. For mosses and lichen, only percentage coverage was measured. The number of trees in each plot, and their average height, canopy percentage, and diameter at breast height were also measured.

Ground and field layer samples in quadrats with 50 cm side lengths were harvested to measure biomass and LAI. Vascular plant samples in 57 plots, moss samples in 18 plots, and lichen samples in 13 plots were collected. The plots were located both in peatland and upland areas. Biomass was measured from oven-dried samples and LAI was measured from photographed samples. Ground and field layer biomass and LAI for the 130 circular plots were then predicted separately for different plant functional types (evergreen dwarf shrubs, evergreen tall shrubs, deciduous dwarf shrubs, deciduous tall shrubs, herbaceous plants, graminoids, mosses, and lichens) using linear regressions. In the regressions, biomass or LAI was the dependent variable, and four different independent variable options were tested: (1) %-coverage, (2) height, (3) %-coverage and height, (4) volume (%-coverage × height). The best regression model was chosen by minimising the root mean square error (RMSE) value. PFT-specific biomass and LAI estimates were then summed to get the total ground and field layer biomass or LAI in each plot. Pine biomass was calculated with an allometric equation (model 2a in Repola 2009).

References

Breiman, L.: Random forests. *Machine learning*, *45*(1), 5-32, 2001.

Keras team: Keras, version 2.10.0 [code]. https://github.com/keras-team/keras, 2022.

Liaw, A., and Wiener, M.: Classification and regression by randomForest. *R news*, *2*(3), 18-22, 2002.

Pastorello, G., Trotta, C., Canfora, E. et al.: The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Sci Data* 7, 225, https://doi.org/10.1038/s41597-020-0534-3, 2020.

Räsänen, A. and Virtanen, T.: Data and resolution requirements in mapping vegetation in spatially heterogeneous landscapes, Remote Sens. Environ., 230, 111207, https://doi.org/10.1016/j.rse.2019.05.026, 2019.

Repola, J.: Biomass equations for Scots pine and Norway spruce in Finland. Silva Fennica, 43(4), 625-647, ISSN 0037-5330, 2009.

Richardson, A. D. and Hollinger, D. Y. A method to estimate the additional uncertainty in gap-filled NEE resulting from long gaps in the CO2 flux record. *Agricultural and Forest Meteorology*, *147*(3-4), 199-208, 2007.

Zhu, S., Clement, R., McCalmont, J., Davies, C. A., and Hill, T. Stable gap-filling for longer eddy covariance data gaps: A globally validated machine-learning approach for carbon dioxide, water, and energy fluxes. *Agricultural and Forest Meteorology*, *314*, 108777, 2022.