Evaluating remote sensing and modeling approaches for estimating net ecosystem exchange in Canadian peatlands

by

E. Kathryn Hettinga

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirements for the degree of

Master of Science

in

Earth Sciences

Waterloo, Ontario, Canada, 2025

© E. Kathryn Hettinga 2025

Author's declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of contributions

Chapter 2 of this thesis consists of a co-authored, manuscript-format paper. Chapter 3 of this thesis has also been prepared in a manuscript-format paper and will be modified for submission to a journal. As the first author of both chapters, I was primarily responsible for the study design, execution, data analysis, and writing. The contributions of the listed co-authors for chapter 2 are as follows: E. Kathryn Hettinga, Fereidoun Rezanezhad, Bhaleka D. Persaud, Stephanie Slowinski, Maria Strack, Haojie Liu, Bernd Lennartz, and Philippe Van Cappellen conceptualized the study. Elyn Humphreys, Jianghua Wu, June Skeeter, Sara H. Knox, Haley Alcock, and Oliver Sonnentag provided the data. E. Kathryn Hettinga conducted the data analysis. Maria Strack, Fereidoun Rezanezhad, and Philippe Van Cappellen acquired funding for the project. E. Kathryn Hettinga drafted the original manuscript while all other authors provided critical feedback and contributions to shaping the research, analysis, and the manuscript. The contributions of the listed co-authors for chapter 3 are as follows: E. Kathryn Hettinga, Fereidoun Rezanezhad, Ali Reza Shahvaran, Mir Amir Mohammad Reshadi, Maria Strack, and Philippe Van Cappellen conceptualized the study. Elyn Humphreys, Phaedra Cowden, Michelle Garneau, Pierre Taillardat, Tim Papakyriakou, Brian Amiro, Mario Tenuta, Kyle Arndt, Susan Natali, Patrick Murphy, Sara H. Knox, Rosemary Howard, Andreas Christen, Peter Lafleur, Jianghua Wu, Oliver Sonnentag, Haley Alcock, Gabriel Hould Gosselin, Lawrence B. Flanagan, June Skeeter, Bruce Johnson, Matt Morison, Kelly A. Nugent, and Ian Strachan provided the data. E. Kathryn Hettinga conducted the data analysis with the assistance of Ali Reza Shahvaran and Mir Amir Mohammad Reshadi. Maria Strack, Fereidoun Rezanezhad, and Philippe Van Cappellen acquired funding for the project. E. Kathryn Hettinga drafted the original manuscript, and all other authors will review the manuscript before submission to the journal. The author list for the manuscript will be updated before publication to reflect each sites data use policy.

Abstract

Peatlands represent a type of wetland, that has accumulated a layer of organic material or peat, resulting in high organic carbon (C) accumulation in this ecosystem. Peatlands hold up to one-third of the global organic C stock, and specifically peatlands located at high latitudes in the Northern Hemisphere, or northern peatlands, store a large proportion of the of the total peatland organic C stock. However, this organic C may be in jeopardy, as warmer temperatures may lead to increases in both the C uptake and output. There is some uncertainty as to how northern peatlands C sink function will be impacted by climate warming, and conventional models of peatland C cycling have been constrained to site-specific applications rather than national-scale analyses. Moreover, *in-situ* measurements are limited in northern peatlands, due to the remoteness of these sites, but also due to equipment limitations under low temperature and light conditions. In this thesis, I specifically looked at one component of C flux, the net ecosystem exchange (NEE) of CO_2. To enable forecasting of NEE of CO_2 fluxes in peatlands under future scenarios, or to generate real-time estimates where no *in-situ* measurements exist, machine learning algorithms trained on *in-situ* CO_2 fluxes from the eddy covariance (EC) technique must be applied. Remotely sensed or gridded climate data products represent potentially important inputs to these modeling applications, as they are widespread both geographically and temporally enabling flux estimation for broader geographic domains.

In Chapter 2, I explored the possibility of using the remotely sensed and modeled Soil Moisture Active Passive Level 4 Global Daily EASE-Grid Carbon NEE (SMAP-NEE) data product to determine NEE in Canadian peatlands. I acquired nine years (2015–2023) of SMAP-NEE data for five peatlands. I also acquired a subset of year-round eddy covariance NEE (EC-NEE) measurements within this time frame at each of the five peatland sites. The analyses showed that

the SMAP-NEE data product reports a stronger growing season (GS) sink and a weaker non-growing season (NGS) source than the EC-NEE measurements. As a result of this finding, I used the relationship between SMAP-NEE and EC-NEE to produce a Corrected-SMAP-NEE dataset, which provides an estimate of seasonal and annual CO_2 budgets. The data analyses of the Corrected-SMAP-NEE dataset showed that NGS CO_2 emissions represent a variable proportion (33%–256%) of the GS CO_2 uptake, and when these NGS emissions were accounted for, the annual CO_2 sink strength was reduced proportionally. Furthermore, this study showed that longer growing seasons were consistent with greater annual net CO_2 uptake at these five peatland sites from 2015-2023. The findings highlight the importance of considering the NGS when evaluating annual northern peatland C budgets. This chapter also provides evidence that existing algorithms leveraging remotely sensed and gridded climate data products to model NEE need improvement for peatlands.

In Chapter 3, I compiled year-round measurements of EC-NEE from 15 Canadian peatland sites and coupled these target data with 34 hydroclimatic predictor variables (features) from remote sensing and gridded climate data products. The models were trained, validated, and tested using four algorithms: ElasticNet Regression (EN), Light Gradient-Boosting Machine (LGBM), Random Forest Regression (RFR), and Support Vector Regression (SVR). A comprehensive feature importance and selection workflow including hierarchical clustering, Gini importance, and minimum redundancy maximum relevance (mRMR) analysis was followed. Model performance stabilized at eight features, which were (relative importance shown in parentheses): evapotranspiration (40%), shortwave radiation (19%), burn area index (11%), normalized difference snow index (10%), snow water equivalent (7%), climate water deficit (4%), wind speed (4%), and soil moisture (4%). I found the best performing model to be the RFR model with these

eight features (R^2 = 0.76; RMSE = 0.31 g C m^{-2} day^{-1}). I also assessed the generalizability and transferability of the top-performing model via a leave-one-ecoregion-out sensitivity analysis as well as on six external validation sites. The RFR model had the highest generalizability within the Taiga Plains ecoregion and the lowest generalizability within the Boreal Plain ecoregion. When testing the model on the external validation sites, performance metrics were comparable to the internal testing data for sites outside of the ecoregions represented in the training data. The findings demonstrate that the eight-feature models can be confidently upscaled to national extents, offering a clear pathway to improve Canada's spatially explicit CO_2 emission inventories.

Acknowledgements

I would like to thank my supervisors Dr. Fereidoun Rezanezhad and Dr. Philippe Van Cappellen for their support and encouragement throughout my studies. I am especially grateful for the many opportunities they provided whether it was presenting at conferences, pursuing professional certifications, or participating in workshops that enriched my academic experience. Thank you also for your flexibility and trust in me as I completed my research. I will always fondly remember sharing maple taffy during our visit to Quebec City. I would also like to thank my committee members Dr. Maria Strack and Dr. Jenine McCutcheon for their valuable guidance, thoughtful suggestions, and crucial feedback on my work. The successful completion of this thesis is very much owing to my committee and their excellence at reigning in my more ambitious ideas. On top of that, being part of the Can-Peat project has been a true privilege. I'm grateful to have met so many inspiring individuals through the network and annual meetings. From exploring restored peatlands in Quebec City to that borderline vertical hike in Kananaskis, the memories and connections I've made will stay with me. It's been incredibly motivating to witness so many people letting their passion for peatlands shine, and it encourages me to do the same, even if it means receiving the same anti-wetland terminology meme again and again.

This research would not have been possible without the support of our funding sources: This research is part of the project Can-Peat: Canada's peatlands as nature-based climate solutions (https://uwaterloo.ca/can-peat). This project was undertaken with the financial support of the Government of Canada. Ce projet a été réalisé avec l'appui financier du gouvernement du Canada. I also acknowledge funding from the Advancing Climate Change Science in Canada program (ACCPJ 536050-18), Global Water Futures (GWF) Winter Soil Processes in Transition project funded under the Canada First Excellence Research Fund, Natural Sciences and Engineering

Research Council (NSERC) Discovery Grant (RGPIN-2022-03334), Canada Research Chairs Program (CRC-2019-00299) and DFG Research Training Group Baltic TRANSCOAST (GRK2000). Also, the Natural Sciences and Engineering Research Council Alexander Graham Bell Canada Graduate Scholarship.

I respectfully acknowledge that much of my work takes place on the traditional territory of the Neutral, Anishinaabeg, and Haudenosaunee peoples. The main campus of the University of Waterloo is situated on the Haldimand Tract, the land granted to the Six Nations that includes six miles on either side of the Grand River. In addition, all peatlands in Canada are located on Indigenous territory. As such all data collected from these lands should be recognized as Indigenous data. As someone who benefits from data sourced from these territories, I am committed to acknowledging its origins. Upon publication, I will upload these findings to Local Contexts to enable the appropriate communities to claim ownership and assert their rights over the data.

Finally, thank you to my fellow members of the Ecohydrology Research Group for always asking thoughtful questions and providing valuable feedback during group meetings. I'm also grateful for all the bubble tea runs and trips to the Grad House over the years, and I will never forget our kayaking adventure. To my wonderful friends, thank you for always being there for me and for commenting on my numerous LinkedIn posts. A heartfelt thanks to my family for their unwavering support, for sitting through the same presentation time and time again, and for always coming up with new and insightful questions to help me prepare. Finally, this thesis is dedicated in loving memory to my sister Sophie.

Table of Contents

Author's declaration ... ii
Statement of contributions ... iii
Abstract ... iv
Acknowledgements ... vii
List of figures ... xi
List of tables ... xiv
List of abbreviations ... xvi
List of symbols ... xx
1 Introduction ... 1
1.1 Carbon cycling in peatlands ... 1
1.2 Important controls on the peatland carbon cycle ... 2
1.3 Climate change in peatlands ... 2
1.4 The eddy covariance technique ... 3
1.5 Remote sensing and gridded climate data ... 4
1.6 Current modeling approaches for NEE ... 5
1.7 Thesis objectives ... 6
1.8 Thesis outline ... 7
2 Non-growing season contributions to the annual net carbon dioxide budget in Canadian peatlands ... 9
2.1 Introduction ... 9
2.2 Materials and Methods ... 13
2.2.1 Site descriptions ... 13
2.2.2 Data acquisition and preprocessing ... 16
2.2.3 Data analysis ... 20
2.3 Results ... 22
2.3.1 Direct comparison of daily SMAP-NEE and EC-NEE ... 22
2.3.2 NGS emissions compared to GS uptake for Corrected-SMAP-NEE ... 27
2.3.3 Inter-seasonal and site differences for Corrected-SMAP-NEE ... 28
2.3.4 Effect of season duration on annual Corrected-SMAP-NEE ... 29
2.4 Discussion ... 30
2.4.1 Correcting SMAP-NEE is valuable permitting data availability ... 30
2.4.2 NGS CO_2 emissions account for a variable proportion of GS CO_2 uptake ... 32

2.4.3 Longer GS results in greater CO_2 uptake as annual NEE 34
2.4.4 Limitations of the SMAP-NEE data product 35
2.5 Conclusions 37
3 Prediction of the net ecosystem exchange of carbon dioxide in Canadian peatlands by integrating remote sensing and gridded climate data products 39
3.1 Introduction 39
3.2 Materials and Methods 42
3.2.1 Data collection and preprocessing 42
3.2.2 Applied machine learning models 52
3.2.3 Modeling workflow 56
3.2.4 External validation and model application 59
3.3 Results 60
3.3.1 Distribution of NEE 60
3.3.2 Distribution of predictors 61
3.3.3 Model performance 62
3.3.4 Model interpretation 64
3.3.5 Model application 65
3.4 Discussion 67
3.4.1 Model performance and applicability 67
3.4.2 Contributions of variables 69
3.4.3 Uncertainties and limitations 71
3.5 Conclusion 72
4 Conclusion 74
4.1 Summary of key findings 74
4.2 Recommendations for future research 77
References 79
Appendices 96
Appendix I: Supplementary Materials for Chapter 2 96
Appendix II: Supplementary Materials for Chapter 3 100

List of figures

Figure 2-1: Map of five peatland sites located in Canada and associated climate stations within Ecoregions of Canada. The five peatland sites include: Delta Burns Bog (CA-DBB), Kinoje Lake Peatland (CA-KLP), Robinsons Natural Bog (CA-RPn), Robinsons Peatland Pasture (CA-RPp), Scotty Creek Bog (CA-SCB). 14
Figure 2-2: Scatterplots of Eddy Covariance NEE (g C m^{-2} day^{-1}) vs Corrected-SMAP-NEE (g C m^{-2} day^{-1}) by site with their linear regression relationships and associated R^2 and p-values. The top row represents the GS, whereas the bottom row represents the NGS. The dashed black line represents the 1:1 line. Sites are represented by colour and shape. 22
Figure 2-3: Boxplots of Corrected-SMAP-NEE (g C m^{-2} $period^{-1}$) throughout the study period (2015-2023). The fluxes are divided into the growing season (GS) (green), non-growing season (NGS) (blue), and annual (purple). The white points represent the mean. The red line across 0 represents the CO_2 sink or source line. 29
Figure 2-4: Scatterplots of cumulative NEE vs seasonal durations. These include annual cumulative Corrected-SMAP-NEE (g C m^{-2} $year^{-1}$) vs GS duration (number of days) (a), NGS cumulative Corrected-SMAP-NEE (g C m^{-2} $period^{-1}$) vs GS duration (number of days) (b), and GS cumulative Corrected-SMAP-NEE (g C m^{-2} $period^{-1}$) vs NGS duration (number of days) (c) with their regression lines and associated R^2 and p-values. Sites are represented by colour and shape. 30
Figure 3-1: Workflow schematic illustrating the four key stages of this study: (i) data collection and preprocessing of eddy covariance (EC) fluxes and gridded predictors, (ii) training and evaluation of four machine learning models: ElasticNet Regression (EN), Light Gradient-Boosting Machine (LGBM), Random Forest Regression (RFR), Support Vector Regression (SVR), (iii) SHAP (SHapley Additive exPlanations)-based assessment of predictor importance, and (iv) external validation and application across independent peatland sites. 43
Figure 3-2: Geographic distribution of the 15 selected peatland sites across Canada, overlaid on the U.S. EPA Level II Ecoregions of North America (Environmental Protection Agency, 2024).46
Figure 3-3: Ridgeline distributions of daily eddy covariance net ecosystem exchange (EC-NEE; g C m^{-2} day^{-1}) at each peatland site (x-axis labels), with negative values (net CO_2 uptake) in blue and positive values (net CO_2 release) in red. Horizontal dashed line at zero distinguishes uptake versus release. 49
Figure 3-4: Pairwise correlations among monthly-mean predictor variables and net ecosystem exchange (NEE). The upper triangle displays Pearson's r, the lower triangle Spearman's ρ. Cell colors range from blue (−1 for strong negative) to red (+1 for strong positive). Asterisks indicate significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. 57
Figure 3-5: Seasonal cycles of monthly-mean net ecosystem exchange (NEE) and the eight key predictor variables: evapotranspiration (ET_M), shortwave radiation (srad_DM), burn area index (BAI), normalized difference snow index (NDSI), snow water equivalent (swe_DM), climate water deficit (cwd), wind speed (ws), and soil moisture (sm), with 95% confidence intervals (CI). Lines show the average of all site-level monthly means; shaded ribbons denote the 2.5th–97.5th percentiles across sites. Monthly means were computed from daily averages (requiring ≥70% half-hourly coverage per day and ≥10 days per month). 61

Figure 3-6: Random Forest (RF) predictions versus eddy covariance (EC) observations of daily CO_2 net ecosystem exchange (NEE) for (a) the training set (n = 462) and (b) the internal test set (n = 155). Colored points show site-level monthly averages, the dashed gray line is the 1:1 reference, and the solid black line is the fitted regression. Key performance metrics (R^2, MAE, MSE, RMSE, MAPE) are listed in the inset for each panel. 63
Figure 3-7: Beeswarm summary of SHAP-based local explanations for the selected eight features: evapotranspiration (ET_M), shortwave radiation (srad_DM), burn area index (BAI), normalized difference snow index (NDSI), snow water equivalent (swe_DM), climate water deficit (cwd), wind speed (ws), and soil moisture (sm), ranked by descending influence on NEE. Point colors (blue to red) show feature magnitude (low to high, respectively), the x-axis shows SHAP values (positive = higher predicted NEE), and each point is an individual monthly average record illustrating feature impacts on model predictions..... 64
Figure 3-8: Random Forest (RF) predictions versus eddy covariance (EC) observations of daily CO_2 net ecosystem exchange (NEE) at six external validation sites (n = 293). Colored points indicate site-level data, the dashed gray line is the 1:1 reference, and the solid black line is the fitted regression. Key performance metrics (R^2, MAE, MSE, RMSE, MAPE) are listed in the inset. 66
Figure AI-1: Boxplots of measures of daily NEE (g C m^{-2} day^{-1}) at a) CA-DBB, b) CA-KLP, c) CA-RPn, d) CA-RPp, and e) CA-SCB. Notable periods of EC-NEE data gaps are represented in Table 2-2. 98
Figure AI-2: Measures of annual NEE (g C m^{-2} $year^{-1}$) at a) CA-DBB, b) CA-RPn, c) CA-RPp, and d) CA-SCB. Only years without large EC-NEE data gaps (Table 2-2) are represented. 98
Figure AI-3: Scatterplots of Eddy Covariance NEE (EC-NEE; g C m^{-2} day^{-1}) vs SMAP-NEE (g C m^{-2} day^{-1}) in the GS (left) and NGS (right) with their regression line and associated R^2 and p-value. The dashed black line represents the 1:1 line..... 99
Figure AI-4: Scatterplots of EC-GPP (g C m^{-2} day^{-1}) vs SMAP-GPP (g C m^{-2} day^{-1}) with their regression line and associated R^2 and p-value at a) CA-DBB in the GS, b) CA-RPn in the GS, c) CA-RPp in the GS, d) CA-SCB in the GS, e) CA-DBB in the NGS, f) CA-RPn in the NGS, g) CA-RPp in the NGS, h) CA-SCB in the NGS. The dashed black line represents the 1:1 line..... 99
Figure AII-1: Time series of half-hourly eddy covariance (EC) net ecosystem exchange (NEE) (grey dots), FFT (fast Fourier transform)-smoothed fluxes (green line), and their corresponding monthly averages (red = raw; blue = smoothed) for each peatland site. The congruence of red and blue monthly lines (appearing purple) demonstrates that FFT filtering does not alter the aggregated monthly NEE estimates. 103
Figure AII-2: Light Gradient-Boosting Machine (LGBM, top row), Support Vector Regression (SVR, middle row), and ElasticNet Regression (EN, bottom row) predictions versus eddy covariance observations of daily CO_2 net ecosystem exchange (NEE) for (a) the training set (n = 462) and (b) the internal test set (n = 155). Colored points show site-level monthly averages; the gray dashed line is the 1:1 reference; and the solid black line is the fitted regression. Key performance metrics (R^2, MAE, MSE, RMSE, MAPE) are listed in each inset. 104
Figure AII-3: Random Forest (RF) predictive performance for daily CO_2 net ecosystem exchange (NEE). Observed EC-NEE (blue) and RF predictions (green) are plotted across

validation instances, with the 95% prediction interval shown in orange (mean width = 1.56 g C m^{-2} day^{-1}). Pink stars mark monthly averages falling outside the interval. 105
Figure AII-4: Bar chart of relative SHAP feature importance for eight predictors: evapotranspiration (ET_M), shortwave radiation (srad_DM), burn area index (BAI), normalized difference snow index (NDSI), snow water equivalent (swe_DM), climate water deficit (cwd), wind speed (ws), and soil moisture (sm). 105
Figure AII-5: SHAP (SHapley Additive exPlanations) summary heatmap displaying the impact of eight predictors: evapotranspiration (ET_M), shortwave radiation (srad_DM), burn area index (BAI), normalized difference snow index (NDSI), snow water equivalent (swe_DM), climate water deficit (cwd), wind speed (ws), and soil moisture (sm), on predicted eddy covariance net ecosystem exchange (NC-NEE) of CO_2 ($f(x)$). Observations (bottom axis) are ordered by decreasing predicted NC-NEE; cell colors denote SHAP values (red = positive, blue = negative), the right-hand bar chart ranks features by mean absolute SHAP importance, and the top line plot traces $f(x)$ for each instance relative to its average. 106
Figure AII-6: Partial dependence plots from the Random Forest model illustrating how each of the eight key predictors: evapotranspiration (ET_M), shortwave radiation (srad_DM), burn area index (BAI), normalized difference snow index (NDSI), snow water equivalent (swe_DM), climate water deficit (cwd), wind speed (ws), and soil moisture (sm), influences daily CO_2 net ecosystem exchange, with tick marks indicating the distribution of observed values..... 107
Figure AII-7: SHAP (SHapley Additive exPlanations) interaction networks from the Random Forest regression model for (a) the most negative (blue) and (b) the most positive (red) predicted net ecosystem exchange (NEE) samples. Each feature (evapotranspiration (ET_M), shortwave radiation (srad_DM), burn area index (BAI), normalized difference snow index (NDSI), snow water equivalent (swe_DM), climate water deficit (cwd), wind speed (ws), and soil moisture (sm)) is shown as a node, its size proportional to its SHAP contribution to NEE, and edges connect features with line width scaled by their SHAP interaction strength. 108
Figure AII-8: Light Gradient-Boosting Machine (LGBM, top row), Support Vector Regression (SVR, middle row), and ElasticNet Regression (EN, bottom row) predictions versus eddy covariance observations of daily CO_2 net ecosystem exchange (NEE) at the external validation sites (n = 293). Colored points denote site-level monthly averages; the gray dashed line is the 1:1 reference; and the solid black line is the fitted regression. Key performance metrics (R^2, MAE, MSE, RMSE, MAPE) are listed in each inset. 109

List of tables

Table 2-1: Site descriptions including climate types and historical meteorological parameters (Christen & Knox, 2021; Environment and Climate Change Canada, 2023; Environmental Protection Agency, 2024; Humphreys *et al.*, 2014; Sonnentag & Quinton, 2021; Wang *et al.*, 2018). 15
Table 2-1 Continued: Site descriptions including historical meteorological parameters, dominant vegetation, peatland classification, and peat depth (Christen & Knox, 2021; Environment and Climate Change Canada, 2023; Environmental Protection Agency, 2024; Humphreys *et al.*, 2014; Sonnentag & Quinton, 2021; Wang *et al.*, 2018). 16
Table 2-2: Eddy covariance setup characteristics and data processing information. 17
Table 2-2 Continued: Eddy covariance setup characteristics and data processing information. 18
Table 2-3: The percentage of the SMAP pixel represented by the land use land cover of the EC tower (Government of Canada, 2024). 19
Table 2-4: Direct comparison of average SMAP-NEE and EC-NEE (g C m^{-2} day^{-1}) from five study sites. Negative values indicate underestimation, whereas positive values indicate overestimation. The *p*-value represents the comparison between SMAP-NEE and EC-NEE for each site within that season. 23
Table 2-4 Continued: Direct comparison of average Corrected-SMAP-NEE and EC-NEE (g C m^{-2} day^{-1}) from five study sites. Negative values indicate underestimation, whereas positive values indicate overestimation. The *p*-value represents the comparison between Corrected-SMAP-NEE and EC-NEE for each site within that season. 23
Table 2-5: Derived daily and cumulative non-growing season (NGS), growing season (GS) and annual Corrected-SMAP-NEE fluxes of CO_2 for the five peatland sites (2015-2023) (Kimball *et al.*, 2022; ORNL DAAC 2018a-e; Satriawan *et al.*, 2023). Different superscript letters indicate statistical significance between sites within that period. 26
Table 2-5 Continued: Literature reported cumulative non-growing season (NGS), growing season (GS), and annual NEE fluxes of CO_2 for the five peatland sites. 27
Table 3-1: Characteristics of the selected 15 peatland study sites: site identifier, geographic coordinates, EPA Level II ecoregion, Köppen–Geiger climate classification, mean annual temperature and precipitation, peatland hydrological type, and primary data citation (ECCC, 2025; AmeriFlux, 2025). 45
Table 3-2: Eddy covariance setup and data characteristics. Different superscript letters indicate statistical significance between 15 selected sites (Kruskal-Wallis, $p<0.05$) (NEE = Net Ecosystem Exchange). 47
Table 3-3: Summary of 32 candidate predictor variables extracted from Google Earth Engine (GEE). The remaining two (in total 34) are the peatland and ecoregion classifications presented in Table 1. GEE datasets are accessed via https://developers.google.com/earth-engine/datasets/catalog/ concatenated with the respective Image Collection ID. 51
Table 3-4: Summary of validation site descriptions: site identifier, geographic coordinates (latitude and longitude), EPA Level II ecoregion, Köppen–Geiger climate classification, mean annual temperature and precipitation, peatland classification, and primary data citation (ECCC, 2025). 60

Table 3-5: Performance metrics on internal testing (and training in parentheses) data for each machine learning model. Bold values indicate the optimal performance for each metric among all models. The top-performing model, considering all metrics, is RF (EN = ElasticNet Regression; LGBM = Light Gradient-Boosting Machine; RFR = Random Forest Regression; SVR = Support Vector Regression). 63
Table AI-1: Coefficients of variation (%) for each measure of net ecosystem exchange. 97
Table AI-2: Direct comparison of Peatland-Corrected-SMAP-NEE and EC-NEE (g C m^{-2} day^{-1}). Negative values indicate underestimation, whereas positive values indicate overestimation. 97
Table AII-1: Monthly means (± 95% CI) of daily-averaged predictor variables and net ecosystem exchange (NEE), stratified by US EPA Level II ecoregion. Values bearing different superscript letters within each column differ significantly ($p < 0.05$, Kruskal-Wallis). Predictor variables are defined as follows: ET_M = total evapotranspiration, cwd = climate water deficit, srad_DM = incident shortwave radiation flux density, NDSI = normalized difference snow index, ws = wind speed at 10 m height, swe_DM = snow water equivalent, sm = soil moisture, and BAI = burn area index. 102
Table AII-2: Performance metrics for the leave-one-ecoregion-out sensitivity analysis. 102

List of abbreviations

ANOVA	Analysis of variance
BAI	Burn area index
C	Carbon
CA-ARB	Attawapiskat River bog
CA-ARF	Attawapiskat River fen
CA-BDB	Bois-des-Bel peatland
CA-BOU	Bouleau peatland
CA-BP	Burned peatland
CA-CF1	Churchill Fen 1
CA-CF2	Churchill Fen 2
CA-CF3	Churchill Fen 3
CA-DB2	Delta burns bog 2
CA-DBB	Delta burns bog
CA-FI	Fish island peatland
CA-KLP	Kinoje lake peatland
CA-Mer	Mer Bleue bog
CA-RPp	Robinsons peatland pasture
CA-RPn	Robinsons natural bog
CA-SCB	Scotty creek bog
CA-SHF	Sandhill Fen
CA-SMC	Smith Creek peatland
CA-UBP	Unburned peatland

CA-WP1	Alberta – Western Peatland – LaBiche River, Black Spruce/Larch Fen
CA-WP2	Alberta – Western Peatland – Poor Fen (*Sphagnum* Moss)
CA-WP3	Alberta – Western Peatland – Rich Fen (*Carex*)
CI	Confidence interval
CO_2	Carbon dioxide
Corrected-SMAP-NEE	SMAP-NEE corrected on a site-by-site basis
CV	Coefficient of variability
CWD	Climate water deficit
DOY	Day of year
EC	Eddy covariance
EC-GPP	Eddy covariance derived gross primary productivity
EC-NEE	Eddy covariance measured net ecosystem exchange
EDA	Exploratory data analysis
EFB	Exclusive feature bundling
EN	ElasticNet regression
ER	Ecosystem respiration
EVI	Enhanced vegetation index
FFT	Fast Fourier transform
fPAR	Fraction of photosynthetically active radiation
GBDT	Gradient-boosting decision trees
GEE	Google earth engine
GHG	Greenhouse gas

GOSS	Gradient-based one-side sampling
GPP	Gross primary production
GS	Growing season
HP	Hudson plain
HR	Heterotrophic respiration
HST	High smooth threshold
IRGA	Infrared gas analyser
LAI	Leaf area index
Lasso	Least absolute shrinkage and selection operator
LGBM	Light gradient-boosting machine
LST	Low smooth threshold
LST	Land surface temperature
LULC	Land use land cover
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MODIS	Moderate resolution imaging spectroradiometer
MSE	Mean square error
mRMR	Minimum redundancy maximum relevance
MURMSE	Mean unbiased root mean square error
MWCF	Marine west coast forest
NDSI	Normalized difference snow index
NDVI	Normalized difference vegetation index
NDWI	Normalized difference water index

NEE	Net ecosystem exchange
NGS	Non-growing season
Peatland-Corrected-SMAP-NEE	SMAP-NEE corrected using data from the five peatland sites
PI	Principal investigator
RFR	Random forest regression
RH	Relative humidity
RMSE	Root mean square error
SHAP	SHapley Additive exPlanations
SOC	Soil organic carbon
SS	Softwood shield
SMAP-GPP	Soil moisture active passive modeled gross primary productivity
SMAP-NEE	Soil moisture active passive modeled net ecosystem exchange
SVM	Support vector machine
SVR	Support vector regression
SWE	Snow water equivalent
Ta	Air temperature
TP	Taiga plain
VLR	Variable length regression

List of symbols

n	Number
y_i	Estimated value
$\hat{y}_i$	Measured value
I	Given data point
L1	Lasso penalty
L2	Ridge penalty
Y	Target variable
X	Matrix of input variables
β	Vector of coefficients
ε	Error term
k	Ridge parameter
y	Vector of observed values
w	Coefficient vector
$\|\|y - Xw\|\| \frac{2}{2}$	Residual sum of squares
$\|\|w\|\|_1$	L1 norm of the coefficient
$\|\|w\|\|_2^2$	L2 norm of the coefficient
α	Confidence level
θ	Mixing parameter
x	SVR input vector
$\in$	SVR input vector
$\mathbb{R}^n$	SVR input vector

$\phi(x)$	Nonlinear function
ω	Weight vector
b	Bias term
$\|\omega\|^2$	Euclidian norm of the weight vector
m	Number of training samples
C	Regularization parameter
ϵ	Error tolerance
$L_\epsilon(e) = \max\{0, \|e\| - \epsilon\}$	ϵ-Insensitive loss
ρ	Correlation coefficient
F	ANOVA test statistic
p	Probability value
R^2	Coefficient of determination
$\sum$	Sum
Med	Median
W	Wilcoxon test statistic
u^*	Friction velocity

1 Introduction

1.1 Carbon cycling in peatlands

Peatlands are a type of wetland with an accumulated layer of partially decomposed organic material, or peat, of 40 cm or greater by Canadian definition (National Wetlands Working Group, 1988). This organic material is high in organic carbon (C), making peatlands critical C sinks. Northern (boreal and subarctic) peatlands are especially important organic C stores, as despite representing 57% of the global distribution of peatlands, they are estimated to store up to 88% of the total peatland C stock (UNEP, 2022; Yu *et al*., 2010). Carbon dioxide (CO_2) represents a major component of the C flux in peatland ecosystems, and they function as CO_2 sinks when CO_2 fixation from the atmosphere by vegetation via gross primary production (GPP), exceeds CO_2 loses via ecosystem respiration (ER), which is the sum of plant and microbial respiration (Chapin *et al.*, 2006). GPP in northern peatlands is relatively low compared to other ecosystem types, but northern peatlands remain a moderate C sink over time resulting from similarly low ER, due to anoxia, low temperatures, acidity, and nutrient limitations (Santoni *et al*., 2023; Umigalski and Bayley, 1996). The difference between ER and GPP is the net ecosystem exchange (NEE), (NEE=ER-GPP) (Chapin *et al.*, 2006). NEE varies strongly seasonally, typically switching from negative (net CO_2 uptake) (Coffer & Hestir, 2019; Lund *et al.*, 2010) during the growing season (GS) to positive (net CO_2 release) during the non-growing season (NGS) (Lafleur *et al.*, 2003; Natali *et al*., 2019; Rafat *et al.*, 2021 and 2022). However, due to the variability in C cycling processes, both uptake and release of CO_2 could occur in either the GS or NGS (Lafleur *et al.*, 2003; Lund *et al.*, 2010). Importantly, peatland CO_2 uptake typically outweighs emissions of methane (CH_4) and dissolved or particulate organic C, reinforcing their role as small but enduring C sinks (Chapin *et al.*, 2006; Wu & Roulet, 2014).

1.2 Important controls on the peatland carbon cycle

The main controls on peatland C cycling are temperature, vegetation community composition, water table height, nutrient availability, and sunlight (Blodau, 2002). ER is mostly governed by factors like temperature, moisture, and the availability of oxygen, but the availability and quality of organic material also play a role (Li *et al.*, 2021b). Alternatively, GPP is primarily controlled by light availability, temperature, and water availability. Vegetation community composition influences organic matter quality for ER but also impacts GPP due to phenological differences (Corbett *et al.,* 2013; Piechl *et al.*, 2018). Plant functional types control C inputs, notably in the case of *Sphagnum*, where they contribute to high rates of peat accumulation through their inhibition of decomposition via the production of phenolic compounds and organic acids (Aerts *et al.*, 1999; Laine *et al.*, 2022; Palozzi and Lindo, 2017; Zeh *et al.,* 2019). Generally, under higher temperatures both ER and GPP increase (Dorrepaal *et al*., 2009; Flanagan and Syed, 2011; Li *et al.*, 2021a). Water availability limits the extent to which increasing temperatures can increase both components of NEE (Ratcliffe *et al.*, 2019; Thayamkottu *et al.*, 2025). The height of the water table limits microbial degradation by limiting the amount of oxygen available to microorganisms. This helps preserve organic material and generally with a higher water table there is lower ER (Zhong *et al.*, 2020).

1.3 Climate change in peatlands

As a result of the temperature and water availability sensitivities of C cycling processes in peatlands there is some concern over peatlands' future CO_2 sink capabilities (Limpens *et al.*, 2008; Liu *et al.*, 2019; Rafat *et al.*, 2022; Ward *et al.*, 2019). While there is a wealth of literature

examining the impacts of temperature-based effects on CO_2 fluxes in northern peatlands, the outcomes have been highly variable across a range of peatland types. In some cases, greater precipitation and a longer GS may increase GPP, resulting in greater annual net CO_2 uptake in northern peatlands (Coffer & Hestir, 2019; Gallego-Sala *et al.*, 2018; Liu *et al.*, 2019; Lunt *et al.*, 2019). However, rising temperatures may also accelerate soil organic matter decomposition, thereby increasing C losses by respiration, resulting in reduced annual net CO_2 uptake or net release of CO_2 (Coffer & Hestir, 2019; Lunt *et al.*, 2019). GPP and ER are also interdependent on nutrient availability, as faster SOC decomposition makes soil nutrients more available and most of the vegetation nutrient supply comes from recycling plant detritus, fostering GPP (Hobbie & Kling, 2014). It is generally accepted that both GPP and ER will increase with climate change, but their relative changes and, in turn, the change in the balance between them will ultimately impact how peatlands' annual NEE fluxes respond to climate warming (Gallego-Sala *et al.*, 2018).

1.4 The eddy covariance technique

One of the primary methods used to measure C fluxes (CO_2 and CH_4) at the ecosystem scale is via the eddy covariance (EC) technique. The theory of the EC technique dates to 1895 (Reynolds, 1895), but advances in this field were limited until equipment and technologies were refined (Baldocchi *et al.*, 2003). Today, this micrometeorological technique relies on calculating the covariance of fluctuations of vertical wind velocity and fluctuations of CO_2 concentration (Baldocchi *et al.*, 2003). The equipment uses a sampling frequency of 10 Hz, or collects 10 samples per second, then averages these fluxes over 30-minute periods (Baldocchi *et al.*, 2003). The setup includes a 3D sonic anemometer that measures wind speed and direction, as well as an infrared gas analyser (IRGA) that measures the concentrations of gases of interest, in this case CO_2 (Baldocchi *et al.*, 2003). Sonic anemometers measure wind speed by sending a sound pulse back

and forth from transmitter to receiver, whereas IRGA functions by determining the absorption of an emitted infrared light source through a gas sample. The EC setup could consist of either an open-path or closed-path IRGA (Haslwanter *et al.*, 2009). Open-path analysers are typically located in the same location as the sonic anemometer, measuring gas concentrations at that location, whereas, closed-path analysers are typically located inside a weather-proof enclosure, measuring gas concentrations that come down from an intake near the location of the sonic anemometer (Haslwanter *et al.*, 2009). Open-path analysers, while requiring less power and maintenance, are exposed to the elements influencing measurement accuracy and resulting in more data gaps when the sensor is dirty or wet (Haslwanter *et al.*, 2009). One study that evaluated the differences between open and closed-path analysers at a mountain grassland site in Austria found that after post-processing open-path and closed-path sensors had very good 1:1 agreement for CO_2 flux measurements (Haslwanter *et al.*, 2009). There are some limitations to the EC method as it has the highest accuracy over flat terrain, under steady environmental conditions, and in a relatively homogenous vegetation landscape both within and outside of the flux tower's footprint (Baldocchi *et al.*, 2003). Across uneven terrain and a heterogenous landscape, due to the variability in the flux tower's footprint, the fluxes could be mostly dominated by contributions outside of the vegetation community or ecosystem of interest (Hernandez Rodriguez *et al.*, 2023).

1.5 Remote sensing and gridded climate data

Broadly, remote sensing detects the radiation reflected or emitted from the Earth's surface and translates this into surface characteristics. The majority of the remote sensing data products used throughout this thesis come from the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument on the Terra (launch date: December 18, 1999) and Aqua (launch date: May 4, 2002)

satellites (Salomonson *et al.*, 1989). These instruments collect data from 36 spectral bands ranging in wavelengths of 0.4 μm to 14.4 μm and cover the entire planet every 1-2 days (Salomonson *et al.*, 1989). The spatial resolution for different bands (and therefore different data products) ranges from 250 m, 500 m, to 1000 m (Salomonson *et al.*, 1989). Gridded climate data products use observations from existing weather stations to inform efforts of scaling these variables to the respective geographic domain. For the Daymet products, they use observations from meteorological stations to interpolate at a 1 km spatial resolution across North America using a spatial regression with corrections for elevation (Thornton *et al.*, 2022). TerraClimate downscales existing observations and reanalysis data to the 4 km resolution across the entire globe (Abatzoglou *et al.*, 2018). The NASA's Soil Moisture Active Passive Level 4 Carbon Net Ecosystem Exchange (SMAP-NEE) data product used in Chapter 2 of this thesis provides daily estimates of NEE, GPP, and heterotrophic respiration (HR) derived from a terrestrial process-based C flux model informed by satellite data at a 9 km spatial resolution (Kimball *et al*., 2022).

1.6 Current modeling approaches for NEE

A growing body of work demonstrates the power of combining EC observations with satellite-derived products to model CO_2 fluxes across landscapes (Papale & Valentini, 2003; Wylie *et al.*, 2007; Xiao *et al.*, 2008; Yamaji *et al.*, 2008). Additionally, site-scale, data-driven methods have proven effective at simulating NEE across a variety of ecosystems (Cai *et al.*, 2020; Dou & Yang, 2018; He *et al.*, 2006; Melesse & Hanley, 2005). In peatlands specifically, several modeling paradigms have been applied at individual tower sites: process-based models (Wu *et al.*, 2011; Wu & Roulet, 2014), purely data-driven approaches (Rafat *et al.*, 2021), and semi-empirical formulations (Balogun *et al.*, 2023; Balogun *et al.*, 2024). Meanwhile, in other biomes, site-level

EC observations have been expanded to regional analyses using machine learning and regression techniques (Beer *et al.*, 2010; Bodesheim *et al.*, 2018; Jung *et al.*, 2011; Tramontana *et al.*, 2016; Virkkala *et al.*, 2021; Xiao *et al.*, 2008), with similar upscaling efforts in wetlands (McNicol *et al.*, 2023; Peltola *et al.*, 2019). By contrast, few studies have attempted to apply these strategies to peatland NEE, either across the Northern Hemisphere (Qiu *et al.*, 2020) or within Canada (Webster *et al.* 2018).

1.7 Thesis objectives

The main objective of this thesis is to assess the capabilities of remotely sensed and gridded climate data products to model the NEE of CO_2 across Canadian peatlands in both the growing and non-growing season for potential upscaling efforts. Specific objectives designed to meet this overarching objective were to:

- Compare the performance of a globally-produced NEE data product to the measured *in-situ* data at Canadian peatland sites
- Train, validate, and test four machine learning algorithms in their efficacy in modeling peatland NEE
- Execute a comprehensive feature importance and selection workflow to determine key environmental drivers in modeling peatland NEE
- Apply the top-performing model to independent validation sites to quantify predictive accuracy and assess regional transferability

1.8 Thesis outline

This thesis is comprised of four chapters, including the current Chapter 1. Chapter 2 is a manuscript on which I am the first author and is currently under review for publication in *Journal of Geophysical Research: Biogeosciences*. The chapter describes the performance of the remotely sensed and modeled Soil Moisture Active Passive Level 4 Global Daily EASE-Grid Carbon NEE data product at five Canadian peatland sites in both the growing and non-growing seasons. Chapter 3 is also a manuscript that presents performance metrics for four different machine learning algorithms across fifteen Canadian peatland sites. Chapter 4 summarizes research findings from this thesis and proposes recommendations for future research. Appendices I and II contain supplementary materials for Chapters 2 and 3, respectively.

Non-growing season contributions to the annual net carbon dioxide budget in Canadian peatlands

E. Kathryn Hettinga[1,*], Fereidoun Rezanezhad[1,2], Bhaleka D. Persaud[1,2], Stephanie Slowinski[1], Maria Strack[2,3], Haojie Liu[4], Bernd Lennartz[4], Elyn Humphreys[5], Jianghua Wu[6], June Skeeter[7,8], Sara H. Knox[8,9], Haley Alcock[10], Oliver Sonnentag[10,11], Matteo Detto[12], and Philippe Van Cappellen[1,2]

[1]Ecohydrology Research Group, Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, Canada

[2]Water Institute, University of Waterloo, Waterloo, Canada

[3]Department of Geography and Environmental Management, University of Waterloo, Waterloo, Canada

[4]Faculty of Agricultural and Environmental Sciences, University of Rostock, Rostock, Germany

[5]Department of Geography & Environmental Studies, Carleton University, Ottawa, Canada

[6]School of Science and the Environment, Memorial University of Newfoundland, Corner Brook, Canada

[7] Geologic Survey of Canada, Natural Resource Canada, Vancouver, Canada

[8] Department of Geography, University of British Columbia, Vancouver, Canada

[9] Department of Geography, McGill University, Montréal, Canada

[10] Département de géographie, Université de Montréal, Montréal, Canada

[11] Department of Geography and Environmental Studies, Wilfrid Laurier University, Waterloo, Canada

[12] Department of Ecology and Evolutionary Biology, Princeton University, Princeton, USA

[*]Corresponding author: E. Kathryn Hettinga

k2hettin@uwaterloo.ca

2 Non-growing season contributions to the annual net carbon dioxide budget in Canadian peatlands

2.1 Introduction

Peatlands are wetlands that have accumulated a layer of organic soil, or peat, of 40 cm or deeper by Canadian definition (NWWG, 1988). Despite occupying only 3% of global terrestrial land area, they are critical stores of organic carbon (C), storing 15–30% of the terrestrial soil organic C pool (Gorham, 1991; Post *et al.*, 1982; UNEP, 2022). Specifically, northern (boreal and subarctic) peatlands, representing 57% of the global distribution of peatlands (UNEP, 2022), are estimated to store up to 88% of the total peatland C stock (Yu *et al*., 2010), with yearly C sequestration capacity in this region as high as 96 Tg C yr^{-1} (Gallego-Sala *et al.*, 2018). Organic C accumulation in peatland ecosystems occurs because carbon dioxide (CO_2) fixation from the atmosphere by vegetation via gross primary production (GPP) exceeds CO_2 loss via ecosystem respiration (ER), which is the sum of plant (autotrophic) respiration and respiration by heterotrophic microorganisms (Baldocchi *et al.*, 2018; Chapin *et al.*, 2006; Lees *et al.*, 2018). The difference between ER and GPP is the net ecosystem exchange (NEE), (NEE=ER-GPP) (Baldocchi *et al.*, 2018; Chapin *et al.*, 2006; Lees *et al.*, 2018). NEE varies seasonally, typically switching from negative (net CO_2 uptake) during the growing season (GS) (Coffer & Hestir, 2019; Lund *et al.*, 2010) to positive (net CO_2 release) during the non-growing season (NGS; *i.e.,* fall, winter, and spring) (Lafleur *et al.*, 2003; Natali *et al*., 2019; Rafat *et al.*, 2021; Rafat *et al.*, 2022). However, due to variability in C cycling processes, both uptake and release of CO_2 could occur in either the GS or NGS (Lafleur *et al.*, 2003; Lund *et al.*, 2010). An understanding of C dynamics in both the GS and NGS, as well as their interaction, is crucial in predicting how peatland CO_2 emissions will respond to a changing climate.

With climate change, the rate of increase in air temperature in the Northern hemisphere is greater than the global average (Hu *et al.*, 2021; Jeong *et al.*, 2012). As a result, there is concern about northern peatlands' future CO_2 sink strength due to the strong temperature sensitivity of C cycling processes (Limpens *et al.*, 2008; Liu *et al.*, 2019; Rafat *et al.*, 2022; Ward *et al.*, 2019). Recent climate projections for boreal peatlands indicate higher air temperatures of about 2.5–7.5 °C by the year 2100 (Jensen *et al.*, 2019). These rising temperatures and the associated response of peatland ecosystems could result in substantial releases of CO_2 (Limpens *et al.*, 2008; Lunt *et al.*, 2019). At these high latitudes, warming is more pronounced during the NGS, while, at the same time, the duration of the NGS is decreasing (Barnard *et al.*, 2018; Liu *et al.*, 2018; Rafat *et al.*, 2022). While there is a wealth of literature examining the impacts of temperature-based effects on CO_2 fluxes in northern peatlands, the outcomes have been highly variable across a range of peatland types. In some cases, increased precipitation and a longer GS may increase GPP, resulting in greater annual net CO_2 uptake in northern peatlands (Coffer & Hestir, 2019; Gallego-Sala *et al.*, 2018; Liu *et al.*, 2019; Lunt *et al.*, 2019). However, rising temperatures may also accelerate soil organic C (SOC) decomposition, thereby increasing C losses by respiration, resulting in reduced annual net CO_2 uptake or net release of CO_2 (Coffer & Hestir, 2019; Lunt *et al.*, 2019). GPP and ER are also interdependent on nutrient availability, as faster SOC decomposition makes soil nutrients more available and most of the vegetation nutrient supply comes from recycling plant detritus, fostering GPP (Hobbie & Kling, 2014). It is generally accepted that both GPP and ER will increase with climate change, but their relative changes and, in turn, the change in the balance between them will ultimately impact how peatlands' annual NEE fluxes respond to climate warming (Gallego-Sala *et al.*, 2018).

One way to determine NEE *in-situ* is through the eddy covariance (EC) technique. EC is a micrometeorological technique that provides a means to derive nearly continuous net CO_2 exchange (*i.e.,* NEE) by calculating the covariance between the fluctuations of vertical wind velocity and the CO_2 mixing ratio (Baldocchi *et al.*, 2018). There are inherent limitations to data acquisition using the EC method, however, especially in northern peatland sites, which are typically remote where collecting EC data may be both expensive and logistically challenging, notably during harsh winter conditions (Coffer & Hestir, 2019; Limpens *et al.*, 2008; Lunt *et al.*, 2019). Additionally, direct measurements of NGS NEE are still relatively rare due to the challenges of using open-path EC equipment in the winter (Dutch *et al.*, 2024).

To combat these challenges, in this study, I tested NASA's Soil Moisture Active Passive Level 4 Carbon Net Ecosystem Exchange (SMAP-NEE) data product (Kimball *et al.*, 2022). This data product provides daily estimates of NEE, GPP, and heterotrophic respiration (HR) from a satellite data-driven terrestrial C flux model at a 9 km spatial resolution (Kimball *et al.*, 2022). The following satellite-measured variables are input into the algorithm to produce the derived C data product: daily surface meteorology (NASA Goddard Earth Observing System, Version 5), Moderate Resolution Imaging Spectroradiometer (MODIS) MOD12Q1 land cover, and MOD16A2 eight-day composite observations of fraction of photosynthetically active radiation (Jones *et al.*, 2017; Kimball *et al.*, 2022). The model was calibrated using a global network of *in-situ* EC CO_2 flux measurements (2000-2008) from the FLUXNET La Thuile Collection (Endsley *et al.*, 2023; Kimball *et al.*, 2022). The targeted accuracy for the SMAP-NEE data product is a mean unbiased root-mean-square error (MURMSE) of 1.6 g C m^{-2} day^{-1}, or 30 g C m^{-2} $year^{-1}$, which they report as similar to the accuracy of EC observations (Endsley *et al.*, 2023). The SMAP-

NEE data product has been used to estimate NEE in northern regions where in-situ measurements are relatively sparse, particularly in western Siberia where there is extensive peatland coverage (Karlsson *et al*., 2021; Zolkos *et al.*, 2022). They used the SMAP-NEE data product to estimate NEE in this region and found that monthly mean NEE ranged from -1.88 to 0.64 g C m^{-2} day^{-1} in the western Siberian lowlands, demonstrating its promise for the applicability of the data product to peatland sites (Karlsson *et al*., 2021). Similarly, Zolkos *et al.* (2022), used the SMAP-NEE data product to estimate NEE in the Mackenzie Lowlands and Travaillant Uplands, in the Northwest Territories in Canada, where peatlands are also widespread. They found these modeled estimates to range from -0.60 to -0.82 g C m^{-2} day^{-1} for the period of June through August of 2016 (Zolkos *et al.*, 2022). They highlight however, that SMAP-NEE has higher applicability at multi-kilometer spatial scales, evidenced by their comparison of mean summertime SMAP-NEE to EC flux tower observations (Zolkos *et al.*, 2022). Moreover, the SMAP-NEE data product has been evaluated against EC towers in the United States, and Australia, and recently the Arctic Boreal Zone encompassing some peatland sites (Jones *et al.*, 2017; Madelon *et al.*, 2025); however, the assessment in peatlands broadly remains limited.

In this study, I analyzed observations of net CO_2 fluxes from five Canadian peatland sites, spanning four ecoclimate zones, to determine: (1) the difference between daily SMAP-NEE estimates and daily eddy covariance NEE (EC-NEE) measurements, (2) the proportion of the net GS CO_2 uptake offset by net NGS emissions, (3) interseasonal and intersite differences in NEE trends, and (4) the effect of seasonal duration on seasonal and annual NEE fluxes. The quantification of NGS and GS CO_2 fluxes will improve our understanding of annual NEE dynamics in peatland ecosystems and how resilient their CO_2 sink status may be to changes in climatic conditions.

2.2 Materials and Methods

2.2.1 Site descriptions

I searched the AmeriFlux and Fluxnet databases for peatland sites across Canada with available EC-NEE data for any period aligning with the SMAP-NEE data product's collection period (2015-2023) for both the GS and NGS. An additional consideration in the site selection process was a diversity of ecoregions and biomes within Canada. I selected four sites within the boreal biome, representing the traditional range of northern peatlands, and one site from the temperate biome (Delta Burns Bog (CA-DBB)) to incorporate diversity. Further, as climate change progresses, the range of ecoregions and biomes are predicted to shift northwards (Dobrowski *et al.*, 2021). Peatlands located in the boreal biome have high transition potential and may experience climatic conditions more characteristic of the temperate biome in the future. All peatland sites were part of the AmeriFlux Network and include Kinoje Lake Peatland (CA-KLP) in Ontario, Robinsons Peatland Pasture (CA-RPp) in Newfoundland, Robinsons Natural Bog (CA-RPn) in Newfoundland, CA-DBB in British Columbia, and Scotty Creek Bog (CA-SCB) in the Northwest Territories (Figure 2-1, Table 2-1). Although 98% of peatlands in Canada exist in a relatively disturbance-free state, there are large uncertainties in this number, and natural resource extraction and land use change represent threats to natural peatlands (UNEP, 2022). To represent sites with a history of disturbance, I elected to include two sites with a history of anthropogenic disturbance, that have had efforts of restoration. The CA-RPp site was previously an undisturbed bog but was drained in the 1970s (Wang *et al.*, 2018). Around 1985, pasture forage grasses (*Phalaris arundinacea*) were introduced, and the site was fertilized and used to graze cattle (Wang *et al.*, 2018). After 10 years of active pasture and fertilization, the site was abandoned for over 20 years and left to regenerate naturally. The current state of the site is unmanaged to the best of our knowledge. Agriculture is the main activity leading to the degradation of peatlands in Canada,

with around 15% of degraded peatlands having been drained for conversion to pastureland (UNEP, 2022). The CA-DBB site is part of a larger peatland complex within the Burns Bog Ecological Conservancy Area (Christen *et al.*, 2016; Christen & Knox, 2021; Lee *et al.*, 2017). The site was subject to peat extraction from 1957–1963, but the site was restored and rewetted in 2007 via ditch blocking (Christen *et al.*, 2016; Christen & Knox, 2021; D'Acunha *et al.*, 2019; Lee *et al.*, 2017). In Canada, up to 2% of degraded peatlands have been impacted by the horticultural peat industry (UNEP, 2022). The other sites (CA-KLP, CA-SCB, and CA-RPn) are generally disturbance-free to the best of our knowledge (Helbig *et al.*, 2017; Humphreys *et al.*, 2014; Wang *et al.*, 2018).

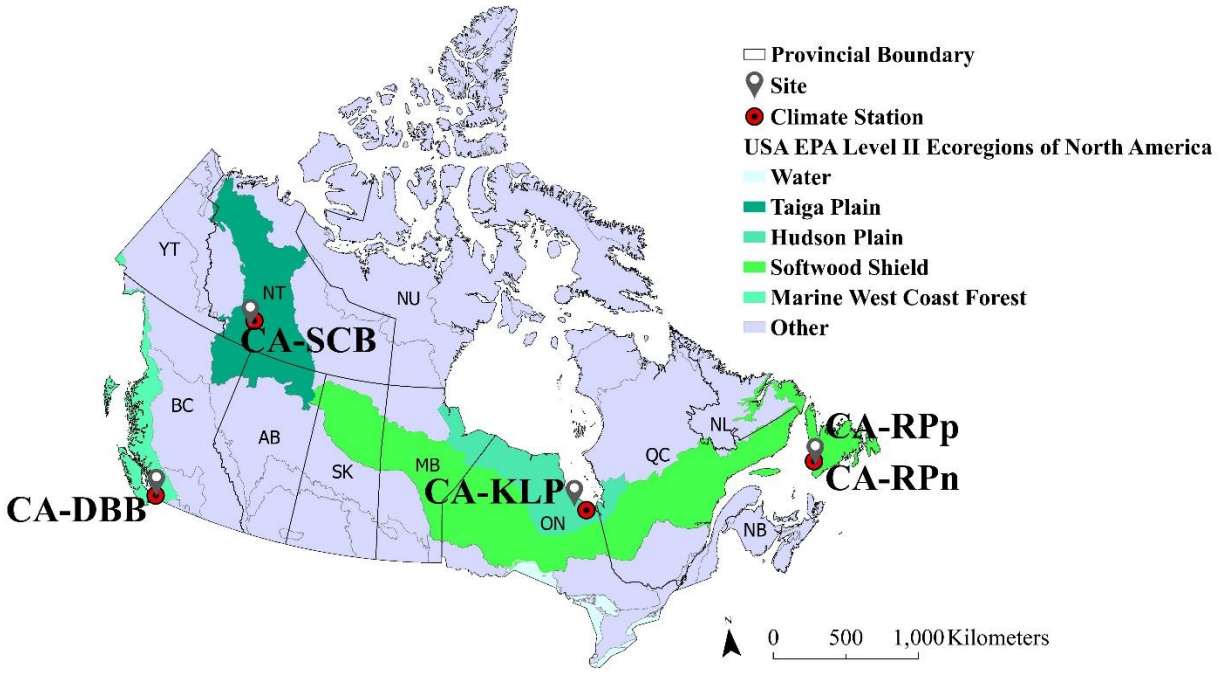


Figure 2-1: Map of five peatland sites located in Canada and associated climate stations within Ecoregions of Canada. The five peatland sites include: Delta Burns Bog (CA-DBB), Kinoje Lake Peatland (CA-KLP), Robinsons Natural Bog (CA-RPn), Robinsons Peatland Pasture (CA-RPp), Scotty Creek Bog (CA-SCB).

Table 2-1: Site descriptions including climate types and historical meteorological variables (Christen & Knox, 2021; Environment and Climate Change Canada, 2023; Environmental Protection Agency, 2024; Humphreys *et al.*, 2014; Sonnentag & Quinton, 2021; Wang *et al.*, 2018).

Site Full Name	Site Shortform	Coordinates	Ecoregion**	Koppen climate classification	Nearest weather station (km from site)
Delta Burns Bog	CA-DBB	49.1293°N, 122.9849°W	MWCF	Csb	Vancouver Int'l A * (16)
Kinoje Lake Peatland	CA-KLP	51.5902°N, 81.7684°W	HP	Dfb	Moosonee (85)
Robinsons Natural Bog	CA-RPn	48.2604°N, 58.6632°W	SS	Dfb	Stephenville A * (31)
Robinsons Peatland Pasture	CA-RPp	48.2632°N, 58.6670°W	SS	Dfb	Stephenville A * (31)
Scotty Creek Bog	CA-SCB	61.3089°N, 121.2984°W	TP	Dfc	Fort Simpson A * (50)

*Indicates a station that meets the United Nations World Meteorological Organization standards

**Shortforms for the ecoregions are as follows: Taiga Plain (TP), Hudson Plain (HP), Softwood Shield (SS), and Marine West Coast Forest (MWCF) (EPA, 2024).

Table 2-1 Continued: Site descriptions including historical meteorological variables, dominant vegetation, peatland classification, and peat depth (Christen & Knox, 2021; Environment and Climate Change Canada, 2023; Environmental Protection Agency, 2024; Humphreys *et al.*, 2014; Sonnentag & Quinton, 2021; Wang *et al.*, 2018).

Site	Mean daily air temperature (°C)	Mean annual rainfall (mm)	Mean annual snowfall (mm)	Dominant vegetation	Peatland classification	Peat depth (m)	Citations
CA-DBB	10.4	1153	38	*White beakrush, Sphagnum mosses*	Disturbed and rewetted bog	5.83-5.9; 4-7, 1-9	Christen *et al*., (2016); Lee *et al*., (2016); D'Acunha *et al*., (2019)
CA-KLP	-0.5	503	227	*Sphagnum* mosses, lichen, sedges, shrubs	Low shrub bog	1.4-2.7	Humphreys *et al., (*2014)
CA-RPn	5.0	995	393	*Sphagnum mosses, ericaceous shrubs, sedges*	Natural Bog	1.8-2.4	Wang *et al.,* (2018)
CA-RPp	5.0	995	393	Reed canary grass*, sedges*, *Ranunculus* spp*., Hieracium* spp.	Disturbed bog	1.7-2.2	Wang *et al.,* (2018)
CA-SCB	-2.8	239	187	*Sphagnum* mosses, shrubs, isolated black spruce and tamarack	Collapse-scar bog	>3	Helbig *et al.*, (2017)

2.2.2 Data acquisition and preprocessing

The eddy covariance (EC) CO_2 flux data were acquired directly from site principal investigators and the Ameriflux network. The EC instrumentation utilized at each site as well as the data processing procedure differed between sites (see Table 2-2). Processing of the EC data for CA-SCB and CA-DBB was performed according to the method outlined in Wang *et al*. (2018), as this was also how the data for CA-RPp and CA-RPn were treated prior to use for this study. This method was adopted to standardize the quality control, gap-filling, and flux partitioning procedure. The gap-filled datasets were used to calculate daily average fluxes for each site, using only days

where 70% or more of the half-hourly fluxes were present. In the case of CA-KLP, the dataset was acquired as daily average fluxes pre-processed according to Humphreys *et al.* (2014), and as such were only post-processed by removing days with <70% of the half-hourly fluxes present. Notable data gaps, as well as the timespans used in further calculations, are presented in Table 2-2. The SMAP-NEE data (2015-2023) were acquired from the Oak Ridge National Laboratory Distributed Active Archive Center using NASA's Fixed Sites Subset Tools and their satellite remote sensing and modelled data products (Kimball *et al.*, 2022; ORNL DAAC, n.d.). For each site, the SMAP pixel NEE value was assigned according to the location of the EC tower. As a result of the data product's spatial scale, ecosystem types other than peatlands could be represented in the pixel. To account for these spatial differences, I identified the land cover class (Government of Canada, 2024) of the EC tower and determined the fractional area of this land cover class within the respective SMAP pixel (Table 2-3). Then I multiplied the NEE value for the SMAP pixel by the fractional area resulting in the final SMAP-NEE variable.

Table 2-2: Eddy covariance setup characteristics and data processing information.

Site	Source of data	Time frame	Source of information on pre-processing methodology	Instrumentation	Open vs. closed path	Height of instrumentation (m)
CA-DBB	Ameriflux	Half-hourly	D'Acunha *et al.* (2019)	IRGA: LI-7200, IRGA: LI-7700	Closed-path, open-path	1.8
CA-KLP	Site PI	Daily	Humphreys *et al.* (2014)	IRGA: LI-7200	Closed-path	3
CA-RPn	Site PI	Half-hourly	Wang *et al.* (2018)	IRGA: LI-7200	Closed-path	3.21
CA-RPp	Site PI	Half-hourly	Wang *et al.* (2018)	IRGA: LI-7200	Closed-path	3.54
CA-SCB	Site PI	Half-hourly	Helbig *et al.*, (2017)	IRGA: EC150	Open-path	1.9

Table 2-2 Continued: Eddy covariance setup characteristics and data processing information.

Site	Post-processing	Large data gaps (>25 days) after post-processing	Years without large data gaps	Time span used in testing exercise
CA-DBB	Filling meteorological data (RH and Ta) with nearby climate station data, REddyProc (*u** filtering, gap-filling, partitioning), Computation of daily fluxes, Removal of daily fluxes with <70% observations	2015/12/01–2016/02/20	2016, 2017, 2018, 2019	06/2015–12/2020
CA-KLP	Removal of daily fluxes with <70% observations	2015/08/22–2015/12/31, 2016/01/10–2016/02/23, 2016/10/17–2017/04/14, 2017/11/20–2018/02/21, 2018/02/24–2018/03/21, 2018/09/28–2018/10/23, 2018/10/27–2019/03/21, 2019/03/24–2019/06/11, 201910/02–2019/11/09	N/A	05/2015–12/2019
CA-RPn	Computation of daily fluxes, Removal of daily fluxes with <70% observations	N/A	2015	05/2015–04/2016
CA-RPp	Computation of daily fluxes, Removal of daily fluxes with <70% observations	N/A	2015	05/2015–04/2016
CA-SCB	Filling meteorological data (rH and Ta) with nearby climate station data, REddyProc (*u** filtering, gap-filling, partitioning), Computation of daily fluxes, Removal of daily fluxes with <70% observations	2016/12/11–2017/03/27, 2017/11/06–2018/03/30, 2018/11/01–2019/03/27, 2019/06/22–2021/04/13, 2021/04/21–2021/07/09, 2021/09/18–2022/04/08	2015	06/2015–09/2022

Table 2-3: The percentage of the SMAP pixel represented by the land use land cover of the EC tower (Government of Canada, 2024).

Site	% of SMAP pixel represented by the LULC of the EC tower
CA-DBB	19%
CA-KLP	67%
CA-RPn	21%
CA-RPp	25%
CA-SCB	35%

The NGS was initially defined using a method based on Rafat *et al.* (2022), where the start of the NGS was defined as the first day of three consecutive days when the daytime air temperature fell below 1°C, and the end of the NGS was defined as the first day of three consecutive days where the temperature rose above 1°C. Using this method, I found that the temperate site (CA-DBB) did not have a NGS for each year of study. Since there were periods with lower temperatures and solar radiation aligning with periods of CO_2 release occurring during the winter, I adapted our definition as follows.

I used SMAP-NEE to determine the GS start and end dates, based on when NEE transitions from positive to negative (Barnard *et al.*, 2018). The period inside these dates represents the GS, whereas the period outside these dates represents the NGS. To determine these dates, I used three different methods and then took the mean (Barnard *et al.*, 2018; Grolemund & Wickham, 2011; Wickham *et al.*, 2023a; Wickham *et al.*, 2023b; Zeileis & Grothendieck, 2005). The first method was the variable length regression (VLR) method which calculates linear regressions on subsets of daily SMAP-NEE from three to seven days in length during Spring (March 1 to June 21) and Fall (September 1 to December 1). The day with the steepest positive slope (SMAP-NEE to the day of year (DOY)) marks the start of the GS, whereas the day with the shallowest negative slope (SMAP-NEE to DOY) marks the end of the GS. The second method is the low smooth threshold method

(LST), which marks the start of the GS as the day when an interval averaged NEE crosses and remains below a threshold (zero) across five days, whereas the end of the GS is marked when the interval averaged NEE remains above the threshold (zero) for five days, after the last possible GS start of the year. The third method is the high smooth threshold method (HST), which is functionally identical to the LST method, but for ten days. The nine-year time series (2015–2023) was separated into years from the first GS start to the GS start of the following year, including only one GS and one NGS per year.

2.2.3 Data analysis

All data analyses were performed in *R* (R Core Team, 2023). After splitting the data into the NGS and GS, based on SMAP-NEE, the dataset was split into training and testing subsets using a stratified random sampling approach stratified by site and season, using only the dates where EC-NEE was available (no gaps) (Kuhn, 2008). The training data represented 75% of the EC-NEE data without gaps, whereas the testing subset represented 25%. The training data was used to run a linear regression to identify the relationship between SMAP-NEE and EC-NEE on a site-by-site basis separated seasonally. The resulting equations were applied to the original SMAP-NEE dataset resulting in the Corrected-SMAP-NEE dataset. The testing data was used to determine how well the SMAP-NEE/Corrected-SMAP-NEE data product estimates EC-NEE, by first calculating the difference between daily SMAP-NEE/Corrected-SMAP-NEE and daily EC-NEE for all available days and then taking the average. As an additional validation metric, the root mean squared error (RMSE) was calculated, given the large sample size, to compare with the target mean unbiased RMSE (MURMSE) for the SMAP-NEE/Corrected-SMAP-NEE data product using:

$$RSME = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (2.1)$$

where n is the number of data points, y_i is the estimated value (SMAP-NEE and Corrected-SMAP-NEE) for the i^{th} data point, and $\hat{y}_i$ is the measured value (EC-NEE) for the i^{th} data point. Additionally, the Kruskal Wallis and Wilcoxon comparisons were used to determine the differences between different measures of daily NEE, partitioned by site, and year.

The resulting Corrected-SMAP-NEE dataset was used for the remainder of the analyses where the cumulative fluxes of the NGS to the GS were compared by calculating the percentage of net GS NEE to net NGS NEE ((net NGS NEE/net GS NEE) × 100%). The Kruskal Wallis and Wilcoxon tests were also used to determine if there were differences in annual and seasonal estimates of Corrected-SMAP-NEE between sites and to determine if there were differences in Corrected-SMAP-NEE between seasons. Statistical significance was assessed at $p<0.05$. Finally, both NGS and GS duration were plotted against annual NEE, and NEE from the respective seasons were also plotted against one another to determine the relationship between these variables. To evaluate the variability of respective measures of NEE, I calculated coefficients of variability (CV), which are defined as the standard deviation divided by the mean and reported as a percentage. I then used the Kruskal-Wallis and pairwise Wilcoxon tests to determine if there were differences between these measures.

2.3 Results

2.3.1 *Direct comparison of daily SMAP-NEE and EC-NEE*

Comparing the daily GS SMAP-NEE to the daily GS EC-NEE (Figure 2-2a-e), with all sites considered, on average SMAP-NEE values were greater than EC-NEE values by 0.37 g C m^{-2} day^{-1} (95% confidence interval (CI)=(-0.10,0.84)) (Table 2-4). Greater SMAP-NEE values relative to the EC-NEE values were also found at each of the five sites (Table 2-4). When comparing the daily NGS SMAP-NEE to the daily NGS EC-NEE (Figure 2-2f-j), with all sites considered, on average SMAP-NEE values were smaller than EC-NEE values by 0.20 g C m^{-2} day^{-1} (95% CI= (-0.41,0.02)) (Table 2-4). All sites during the NGS, except CA-KLP, share this finding of SMAP-NEE values being smaller than EC-NEE values, whereas at CA-KLP, SMAP-NEE values were greater on average than EC-NEE values by 0.02 g C m^{-2} day^{-1} (95% CI=(-0.08,0.12)) (Table 2-4). With all sites considered the RMSE for the daily fluxes for SMAP-NEE during the GS was 1.06 g C m^{-2} day^{-1}, whereas during the NGS 0.62 g C m^{-2} day^{-1}. Both RMSE values are within SMAP's target mean unbiased RMSE indicating that the model performs within their target.

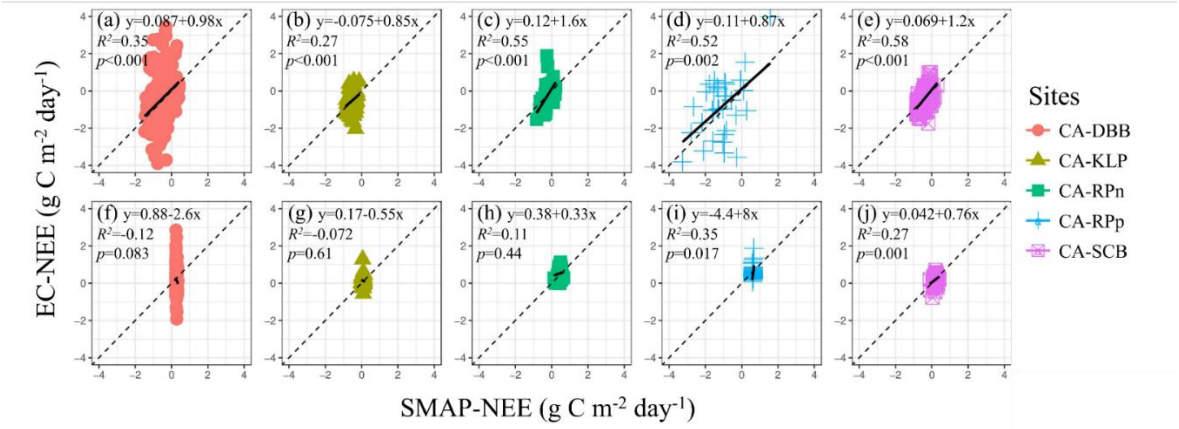


Figure 2-2: Scatterplots of Eddy Covariance NEE (g C m^{-2} day^{-1}) vs Corrected-SMAP-NEE (g C m^{-2} day^{-1}) by site with their linear regression relationships and associated R^2 and p-values. The top row represents the GS, whereas the bottom row represents the NGS. The dashed black line represents the 1:1 line. Sites are represented by colour and shape.

Table 2-4: Direct comparison of average SMAP-NEE and EC-NEE (g C m^{-2} day^{-1}) from five study sites. Negative values indicate underestimation, whereas positive values indicate overestimation. The *p*-value represents the comparison between SMAP-NEE and EC-NEE for each site within that season.

Site	Sample size (*n*)		SMAP-NEE compared to EC-NEE, mean (95% CI) (g C m^{-2} day^{-1})			
	GS	NGS	GS	*p*-value	NGS	*p*-value
All Sites	5	5	0.37 (-0.10,0.84)	0.065	-0.20 (-0.41,0.02)	0.065
CA-DBB	268	232	0.38 (0.22,0.54)	<0.001	-0.19 (-0.30,-0.08)	<0.001
CA-KLP	155	51	0.28 (0.18,0.37)	<0.001	0.02 (-0.08,0.12)	0.012
CA-RPn	36	47	0.11 (-0.11,0.32)	0.085	-0.39 (-0.47,-0.30)	<0.001
CA-RPp	36	46	1.01 (0.42,1.60)	0.002	-0.34 (-0.44,-0.25)	<0.001
CA-SCB	179	149	0.07 (0.01,0.13)	0.078	-0.09 (-0.12,-0.06)	<0.001

Table 2-4 Continued: Direct comparison of average Corrected-SMAP-NEE and EC-NEE (g C m^{-2} day^{-1}) from five study sites. Negative values indicate underestimation, whereas positive values indicate overestimation. The *p*-value represents the comparison between Corrected-SMAP-NEE and EC-NEE for each site within that season.

Site	Sample size (*n*)		Corrected-SMAP-NEE compared to EC-NEE, mean (95% CI) (g C m^{-2} day^{-1})			
	GS	NGS	GS	*p*-value	NGS	*p*-value
All Sites	5	5	-0.01 (-0.05,0.03)	0.062	-0.002 (-0.08,0.07)	0.062
CA-DBB	268	232	-0.04 (-0.19,0.12)	0.9	-0.01 (-0.12,0.10)	<0.001
CA-KLP	155	51	-0.01 (-0.08,0.06)	0.46	-0.02 (-0.09,0.05)	0.85
CA-RPn	36	47	0.03 (-0.19,0.24)	0.40	-0.08 (-0.16,0.01)	0.09
CA-RPp	36	46	-0.05 (-0.57,0.47)	0.59	0.09 (-0.01,0.19)	0.005
CA-SCB	179	149	0.01 (-0.05,0.07)	0.82	0.01 (-0.02,0.04)	0.96

The correction (Corrected-SMAP-NEE) performed on a site-by-site basis greatly reduced the variability of the SMAP-NEE fluxes, particularly during the NGS. After the correction, with all sites considered for the GS, on average Corrected-SMAP-NEE were smaller than EC-NEE values by 0.01 g C m^{-2} day^{-1} (95% CI= (-0.05,0.03)) (Table 2-4). CA-KLP, CA-RPp, and CA-DBB also share this tendency, whereas at CA-RPn and CA-SCB Corrected-SMAP-NEE values were greater than EC-NEE values by 0.03 (95% CI=(-0.19,0.24)), and 0.01 (95% CI=(-0.05,0.07)) g C m^{-2} day^{-1}, respectively (Table 2-4). After the correction, with all sites considered for the NGS, on average Corrected-SMAP-NEE were smaller than EC-NEE values by 0.002 g C m^{-2} day^{-1} (95% CI= (-0.08,0.07)). CA-KLP, CA-RPn, and CA-DBB share this tendency, whereas at CA-RPp, and CA-SCB Corrected-SMAP-NEE values were greater than EC-NEE values by on average 0.09 (95% CI=(-0.01,0.19)), and 0.01 (95% CI=(-0.02,0.04)) g C m^{-2} day^{-1}, respectively (Table 2-4). Additionally, the RMSE with all sites considered for the daily fluxes for the Corrected-SMAP-NEE during the GS was 0.94 g C m^{-2} day^{-1}, whereas during the NGS 0.57 g C m^{-2} day^{-1}. These RMSE values are lower than those for the SMAP-NEE dataset, representing a more accurate model estimation of measured fluxes. For the coefficients of variation, I found that SMAP-NEE after correcting for LULC within the pixel had the highest CV followed by unprocessed SMAP-NEE (Table AI-1). The variability of the SMAP-NEE dataset was greatly reduced when I implemented the site-by-site correction, but it better reflected the variability of the EC-NEE dataset (Table AI-1).

The target mean unbiased RMSE for SMAP-NEE is less than or equal to 1.6 g C m^{-2} day^{-1}, which they report as similar to the accuracy of EC observations (Endsley *et al.*, 2023). The unprocessed SMAP-NEE dataset meets this target where the RMSE for the GS is 1.06 g C m^{-2} day^{-1}, but for the NGS is 0.62 g C m^{-2} day^{-1}. However, given the lower productivity of peatlands, reaching these targets could still lead to erroneous source or sink attributions at these sites. I found that based on mean daily NEE, SMAP-NEE generally captured the interannual variability in shifting from sink to source (Figure AI-1). Based on annual NEE SMAP-NEE, these differences in measures of NEE have a cumulative effect in which SMAP-NEE does not capture the sink or source trend of that site in that year (Figure AI-2). The correction (Corrected-SMAP-NEE) reduced the RMSE in the GS to 0.94 g C m^{-2} day^{-1} and to 0.57 g C m^{-2} day^{-1} for the NGS. Overall, the correction reduced the difference between the SMAP-NEE and EC-NEE fluxes, resulting in seasonal and annual budgets that better reflect the *in-situ* environment (Table 2-5).

Table 2-5: Derived daily and cumulative non-growing season (NGS), growing season (GS) and annual Corrected-SMAP-NEE fluxes of CO_2 for the five peatland sites (2015-2023) (Kimball *et al.*, 2022; ORNL DAAC 2018a-e; Satriawan *et al.*, 2023). Different superscript letters indicate statistical significance between sites within that period.

Site	**Mean season duration (95% CI) (days)**		**Mean daily NEE (95% CI) (g C m^{-2} day^{-1})**			**Mean cumulative NEE (95% CI) (g C m^{-2} $period^{-1}$)**		
	GS	NGS	GS	NGS	Annual	GS	NGS	Annual
All Sites	146 (121,171)	218 (192,244)	-0.61 (-1.07,-0.15)	0.33 (0.07,0.59)	-0.05 (-0.23,0.12)	-90.5 (-163.1,-17.9)	70.7 (16.6,124.8)	-19.8 (-82.5, 42.5)
CA-DBB	177^{a} (164,189)	188^{c} (175,202)	-0.63^{d} (-0.70,-0.57)	0.27^{c} (0.27,0.27)	-0.17^{c} (-0.19,-0.15)	-111.4^{c} (-117.5,-105.3)	50.7^{c} (47.2,54.2)	-60.7^{c} (-66.8, -54.7)
CA-KLP	132^{bc} (118,147)	234^{a} (219,248)	-0.55^{c} (-0.57,-0.53)	0.10^{e} (0.10,0.10)	-0.14^{d} (-0.17,-0.11)	-73.7^{b} (-85.0,-64.3)	23.0^{d} (20.5,25.4)	-50.7^{c} (-62.0, -39.5)
CA-RPn	149^{b} (141,156)	214^{b} (207,220)	-0.26^{a} (-0.29,-0.23)	0.45^{b} (0.44,0.47)	0.16^{a} (0.14,0.18)	-38.8^{a} (-43.8,-33.9)	97.2^{b} (91.5,102.8)	58.4^{a} (52.2,64.6)
CA-RPp	150^{b} (143,157)	212^{b} (205,220)	-1.21^{e} (-1.35,-1.07)	0.62^{a} (0.62,0.62)	-0.14^{e} (-0.2,-0.07)	-181.9^{d} (-206.7,-157.1)	132.2^{a} (127.2, 137.3)	-49.7^{c} (-74.0, -25.3)
CA-SCB	123^{c} (116,130)	241^{a} (233, 250)	-0.38^{b} (-0.42,-0.34)	0.21^{d} (0.20,0.22)	0.01^{b} (-0.01,0.03)	-46.6^{a} (-52.7, -40.6)	50.2^{c} (46.5,53.9)	3.6^{b} (-3.5,10.6)

Table 2-5 Continued: Literature reported cumulative growing season (GS), non-growing season (NGS), and annual NEE fluxes of CO_2 for the five peatland sites.

Site	Literature reported cumulative NEE (g C m^{-2} $period^{-1}$)			Method and citation
	GS	NGS	Annual	
All Sites	-98.4 (-174.7,-22.1)*	47.3 (15.1, 79.4)*	-51.4 (-105.5, 2.8)*	95% CI
CA-DBB	-62.3 ± 17.3	50.0 ± 29.1	-12.3 ± 20.4	95% CI (2016-2021) (Satriawan *et al.,* 2023)
CA-KLP	-59±3	9±3*	-51 ± 15	95% CI (2011, 2012) (Humphreys *et al.,* 2014)
CA-RPn	-105 ± 60	59 ± 69	-46 ± 36	Mean ± standard deviation (2014-2016) (Wang *et al.,* 2018)
CA-RPp	-203 ± 78	79 ± 89	-124 ± 56	Mean ± standard deviation (2014-2016) (Wang *et al.,* 2018)
CA-SCB	-62.8*	39.3	-23.5 (-19.6, -35.1)	95% CI (2015-2016) (Helbig *et al.,* 2017)

*Indicates a calculated value

2.3.2 NGS emissions compared to GS uptake for Corrected-SMAP-NEE

The cumulative NEE for the entire study period, as well as the mean cumulative and daily Corrected-SMAP-NEE fluxes of CO_2 for the five peatland sites are presented in Table 2-5. These fluxes are further divided into NGS, GS, and annual fluxes. The GS overall consistently represented a period of CO_2 uptake, whereas the NGS overall consistently represented a period of CO_2 output, despite there being some variability at the daily scale (Table 2-5). At CA-KLP, net NGS emissions accounted for on average 33.0% (95% CI=(22.5,43.7)) of net GS uptake resulting in the site functioning as an annual CO_2 sink for all years studied. At the CA-RPn site, net NGS emissions

accounted for on average 255.9% (95% CI=(223.4,288.4)) of net GS uptake, resulting in the site functioning as an annual CO_2 source for all years studied. Moreover, at the CA-RPp site, net NGS emissions accounted for on average 74.5% (95% CI=(65.2,83.8)) of net GS uptake, resulting in the site functioning as an annual CO_2 sink for all years studied. At CA-SCB, net NGS emissions accounted for on average 110.7% (95% CI=(93.5,127.9)) of net GS uptake, and the site functioned as an annual CO_2 sink from 2015-2018, then functioned as an annual CO_2 sink from 2018-2023. Finally, at CA-DBB, net NGS emissions accounted for on average 45.6% (95% CI=(42.2,49.0)) of net GS uptake, resulting in the site functioning as an annual CO_2 sink for all years studied.

2.3.3 Inter-seasonal and site differences for Corrected-SMAP-NEE

The boxplots for Corrected-SMAP-NEE separated by period (GS, NGS, and annual fluxes) for the five peatland sites are presented in Figure 2-3. Some notable patterns were that Corrected-SMAP-NEE during the GS was always negative, whereas Corrected-SMAP-NEE during the NGS was always positive. However, annual Corrected-SMAP-NEE fluctuated between being positive or negative (Figure 2-3). Notably, with all sites considered, the Corrected-SMAP-NEE fluxes were always positive during the NGS (70.7 g C m^{-2} $period^{-1}$ (95% CI=(58.7,82.6))), compared to the negative values during the GS (-90.5 g C m^{-2} $period^{-1}$ (95% CI=(-107.1,-73.9))) (Wilcoxon rank sum, p<0.001, W=0). There were significant differences in Corrected-SMAP-NEE between sites during both the GS and NGS, as well as annually (Kruskal-Wallis and Pairwise Wilcox Test: p<0.0001, F=39.7; p<0.0001, F=40.2; and p<0.0001, F=35.0, respectively). The differences between Corrected-SMAP-NEE fluxes by site are presented in Table 2-5.

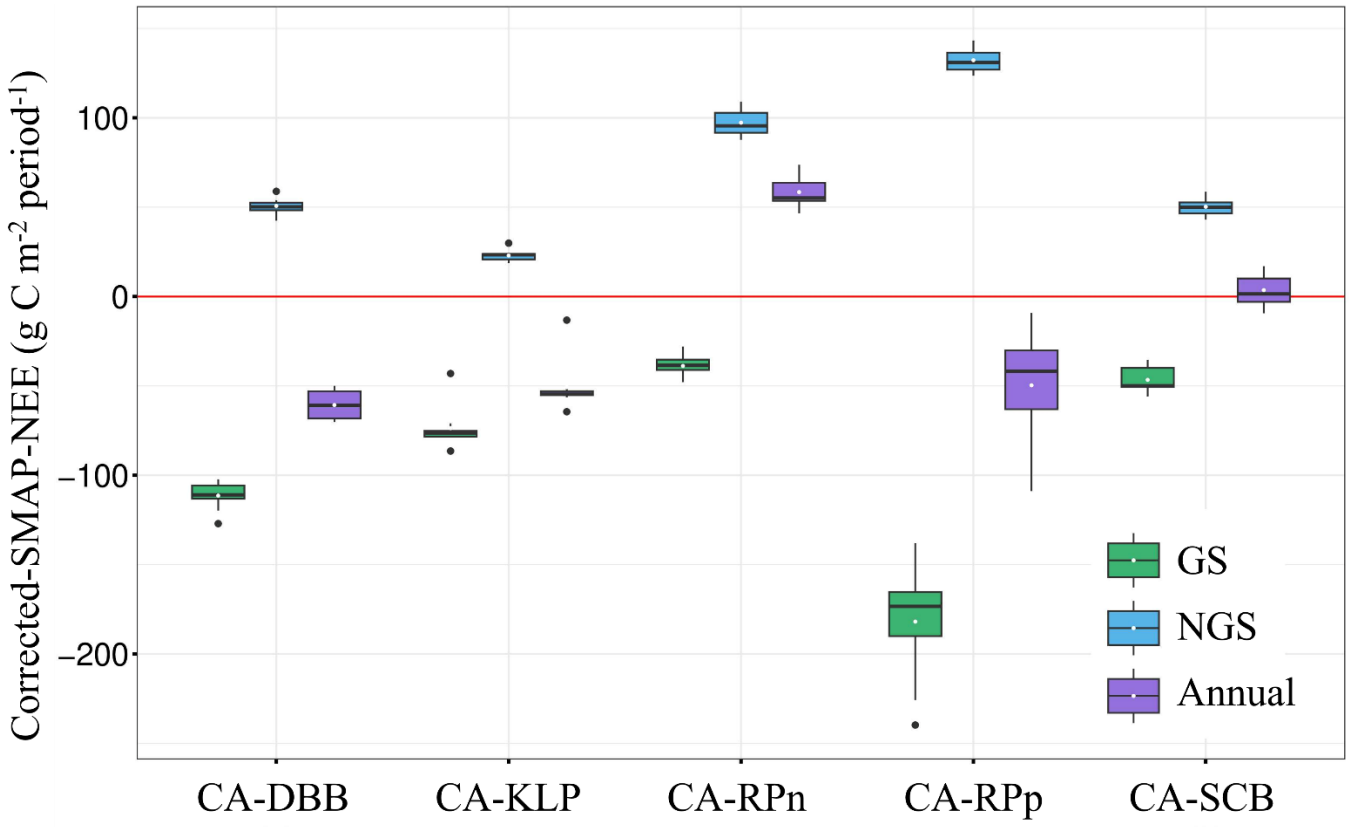


Figure 2-3: Boxplots of Corrected-SMAP-NEE (g C m^{-2} $period^{-1}$) throughout the study period (2015-2023). The fluxes are divided into the growing season (GS) (green), non-growing season (NGS) (blue), and annual (purple). The white points represent the mean, whereas the black points represent outliers. The red line across 0 represents the CO_2 sink or source line.

2.3.4 Effect of season duration on annual Corrected-SMAP-NEE

Both the mean GS duration and NGS duration varied substantially among sites with all years considered (ANOVA: $p<0.001$, F=20.89, and $p<0.001$, F=20.53). The differences between GS and NGS duration by site are presented in Table 2-5. From the Pearson's correlation analysis, with all sites considered, there was a significant negative relationship between GS duration and annual NEE ($p<0.04$). Additionally, there was a slight positive relationship between GS duration and subsequent NGS NEE ($p=0.29$) and a significant positive relationship between NGS duration and

subsequent GS NEE (p<0.014). Scatterplots demonstrating these relationships are presented in Figure 2-4.

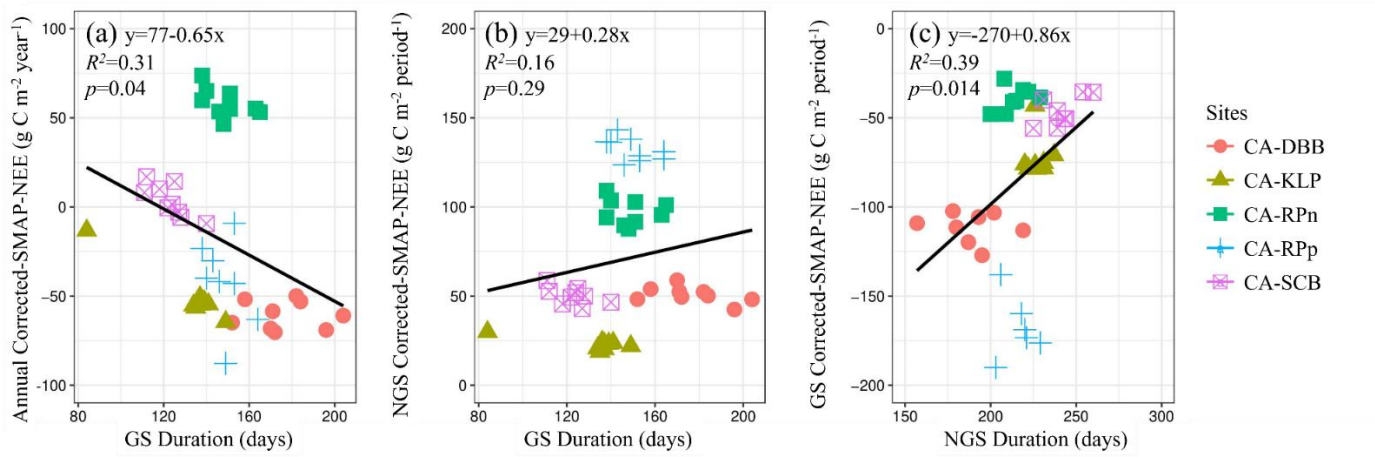


Figure 2-4: Scatterplots of cumulative NEE vs seasonal durations. These include annual cumulative Corrected-SMAP-NEE (g C m^{-2} year^{-1}) vs GS duration (number of days) (a), NGS cumulative Corrected-SMAP-NEE (g C m^{-2} period^{-1}) vs GS duration (number of days) (b), and GS cumulative Corrected-SMAP-NEE (g C m^{-2} period^{-1}) vs NGS duration (number of days) (c) with their regression lines and associated R^2 and p-values. Sites are represented by colour and shape.

2.4 Discussion

2.4.1 Correcting SMAP-NEE is valuable permitting data availability

While the correction of the SMAP data improved the accuracy of the SMAP dataset, it is likely that several years of *in-situ* data per site would be needed to have the correction be meaningful. If there were several years of EC-NEE data, as well as additional meteorological information resulting from the EC method, a better route for long-term assessment of C cycling trends would be to generate models on a site-by-site basis, permitting data availability. An area for future extension would be, upon the inclusion of additional peatland sites, using the relationship between EC-NEE and SMAP-NEE at all sites to correct SMAP-NEE (Appendix I, Table AI-2, and Figure

AI-3), and extrapolating this correction relationship to additional peatland sites. The relative similarity of the linear regression parameters between the EC-NEE and SMAP-NEE used for the SMAP data correction across the 5 sites suggests that such a general correction relationship for peatland sites may be possible. However, the agreement between the Corrected-SMAP-NEE and EC-NEE was better on a site-by-site basis. Additionally, I recommend implementing machine learning methods to evaluate the relationship between SMAP-NEE and EC-NEE, as the linear relationships after the correction had relatively low explanatory power, albeit being higher than the unprocessed dataset.

Looking more broadly at the SMAP-NEE data products' performance in other ecosystems, a study conducted in watersheds in the Canadian arctic examined SMAP-NEE summertime (June-August 2016) fluxes at nine AmeriFlux sites with eddy covariance stations with available data and ecosystems similar to the ecosystems studied here (boreal forest and tundra) (Zolkos *et al.*, 2022). Their analysis shows that EC-NEE had a wider variation compared to SMAP-NEE, but also that SMAP-NEE tended to overestimate NEE reporting more positive values than their EC-NEE counterparts (Zolkos *et al.*, 2022). This finding is consistent with our finding that SMAP-NEE tends to be higher than EC-NEE during the GS at the five peatland sites. In another study looking at how land use impacts carbon fluxes in Texas, they evaluated the performance of SMAP-NEE compared to EC-NEE at an open-developed site, where they found reasonable agreement between the two datasets (Ray *et al.*, 2019). They found the correlation coefficient between SMAP-NEE and EC-NEE to be 0.42, which demonstrates better agreement compared to our average of all sites during the GS of 0.27 and during the NGS 0.16 (Figure AI-3) (Ray *et al.*, 2019). However, their calculated RMSE of 2.42 g C m^{-2} day^{-1} (Ray *et al*., 2019) was higher than the target accuracy of

SMAP-NEE, and higher than the value of 1.06 g C m^{-2} day^{-1} and 0.62 g C m^{-2} day^{-1}, in the GS and NGS, respectively that I found for the five peatland sites. They also found SMAP-NEE tended to report a greater CO_2 source during the winter, as well as a greater CO_2 sink during the summer compared to its EC-NEE counterpart (Ray *et al.*, 2019). This finding is somewhat consistent with our results, as I found SMAP-NEE tended to report a greater CO_2 sink during the GS, but may report a weaker NGS CO_2 source compared to EC-NEE. Additionally, the SMAP-NEE data product was assessed against CO_2 flux measurements at 26 core validation sites from the period of March 31, 2015, to December 31, 2015 (Jones *et al.*, 2017). Seven global MODIS plant functional type vegetation classes were represented by these core validation sites (Jones *et al.*, 2017). They found that the overall RMSE was 1.04 g C m^{-2} day^{-1}, which met their targeted accuracy (Jones *et al.*, 2017), and our calculation for the peatland sites of a RMSE of 1.06 g C m^{-2} day^{-1} during the GS and 0.62 during the NGS is consistent with their targeted accuracy as well. They found that SMAP GPP demonstrated the highest correlation with tower observations, which is consistent with our findings (Figure AI-4), followed by ER, then NEE showed the lowest correlation relative to tower observations (Jones *et al.*, 2017).

2.4.2 NGS CO_2 emissions account for a variable proportion of GS CO_2 uptake

The data analyses in this study showed that net NGS CO_2 release represented between 33% and 256% of the net GS CO_2 uptake. This range was higher than that reported by Zhao *et al.* (2016), who found that NGS CO_2 emissions accounted for 16–80% of GS CO_2 uptake in boreal peatlands, based on a review of four studies. Specifically, at CA-KLP, CA-RPp, and CA-DBB GS CO_2 uptake offset by NGS CO_2 emissions fell within the range reported by Zhao *et al.* (2016), whereas CA-SCB was slightly higher than their reported range (110.7%), and CA-RPn was over three times

higher than the maximum range value. The two Newfoundland Sites (CA-RPp and CA-RPn) had the highest daily NGS emissions (0.62, and 0.45 g C m^{-2} day^{-1}, respectively) despite having the shortest NGS period, I hypothesize this result comes from climatic differences as the two Newfoundland sites had substantially higher mean annual snowfall (Table 2-1). Snow can be insulating resulting in higher soil temperatures that facilitate higher microbial activity, and thereby CO_2 emissions (Gouttevin *et al.*, 2012; Zhao *et al.*, 2022). CA-RPp offset these high NGS emissions with large GS CO_2 uptake due to fast growing vegetation, resulting in a proportion that falls within the expected range; however, CA-RPn had typical bog vegetation and therefore resulted in the high emissions as they were not offset during the GS (-1.21, and -0.26 g C m^{-2} day^{-1}, respectively) (Wang *et al*., 2018).

Our estimated annual mean cumulative NEE ranged from -60.7 to 58.4 g C m^{-2} yr^{-1} across the 5 sites (Table 2-5), which are in line with measurements determined using EC towers in the range of -90 to 100 g C m^{-2} yr^{-1} reported by Limpens *et al.* (2008). Our estimates are also in line with the mean annual NEE of -48 g C m^{-2} yr^{-1} (95% CI=-171,75) calculated from EC measurements in high latitude peatlands (>45°N) (Strack *et al.*, 2023). Additionally, our analyses showed that the five peatlands differed in their annual CO_2 sink capacity throughout the study periods (Table 2-5 and Figure 2-3). This finding contrasts with a study by Lund *et al.* (2010) that synthesized EC data for 12 arctic wetland sites, maintaining that all sites acted as annual sinks for CO_2. In the present study, land use may be a factor influencing the sink or source status of these peatlands, as in the case of CA-DBB, CA-RPn, and CA-RPp, there have been various degrees of disturbance and restoration. Further, given the rate of climate change progression in the arctic, it is also possible that these sites may have shifted from sinks at the time of publication (2010) to potential sources. Our estimates

based on Corrected-SMAP-NEE were compared with estimates from the literature on a site-by-site basis are presented in Table 2-5. Generally, after correction, the derived values better reflect the *in-situ* environment, although the lower performance of the NGS Corrected-SMAP-NEE fluxes indicates that additional model validation exercises are likely needed for SMAP-derived NGS CO_2 exchange fluxes in peatland environments.

2.4.3 Longer GS results in greater CO_2 uptake as annual NEE

Our analysis showed a significant negative relationship between GS duration and annual NEE (Figure 2-4a), that is, the net CO_2 sink strength across our peatland sites and study years was greatest when the GS was longer and the NGS shorter. Our findings are consistent with most other studies in northern peatlands, which have shown that NEE has a negative relationship with GS duration and that GS duration is one of the key variables explaining interannual and inter-site variability in CO_2 flux parameters including NEE (Baldocchi *et al.*, 2018; Humphreys *et al.*, 2014; Lund *et al.*, 2010; Lunt *et al.*, 2019). In contrast, other studies have found that NEE had a positive, although sometimes non-significant correlation with GS duration (Lund *et al.*, 2010; Rafat *et al.*, 2022). In these cases, it's likely that the stimulating effect of an increased GS duration on GPP is offset by a concurrent increase in respiration (Lund *et al.*, 2010). Our observation that the GS duration had a positive, although not significant (p=0.29), correlation with the subsequent NGS cumulative NEE (Figure 2-4c) has also been observed by others (Zhao *et al.*, 2016). This correlation is likely the result of the cross-seasonal link in NEE dynamics, whereby NGS CO_2 dynamics are influenced by the amount and decomposability of the C substrates produced by GPP during the GS (Zhao *et al.*, 2016). Based on our data analysis, the lengthening of the growing season expected under climate change, may lead to increased CO_2 uptake, but understanding how

meteorological conditions may change alongside the lengthening of the growing season is a priority for future analysis.

2.4.4 Limitations of the SMAP-NEE data product

The SMAP data product is advantageous because it (1) enables the estimation of peatland C fluxes throughout both the GS and NGS, with no limitations due to equipment failure during the harsh conditions of the NGS in cold regions, and (2) allows for inexpensive data collection throughout the NGS and GS and across years in hard-to-access high latitude peatlands. However, our findings show that the C fluxes derived by the SMAP data product have their limitations.

Our findings highlight the discrepancy between *in-situ* C fluxes measured in northern peatlands and those estimated by the SMAP data product, suggesting that there are limitations to the application of the SMAP data product to northern peatlands. There are a number of uncertainties that could contribute to errors in the SMAP data product, including errors stemming from inputs and uncertainties associated with model parameterization, initialization, and calibration (Kimball *et al.*, 2022). The SMAP derived C fluxes have been derived by calibrating the SMAP carbon cycling model outputs via comparison against EC tower measurements from global FLUXNET sites representing major plant functional type classes (Kimball *et al.*, 2022). Version 7 of the validation assessment report for the SMAP data product (Endsley *et al.*, 2023) detailed that they specifically validated the data product in peatlands by comparing the derived fluxes to peatland FLUXNET tower sites. As a result, the newest version of the SMAP data generally reproduces the patterns of peatland sites, whereby peatlands have lower seasonal NEE amplitudes compared to non-peatland sites. Our finding that the SMAP-NEE values were higher than EC-NEE values during the GS and lower during the NGS demonstrates that using the uncorrected SMAP-NEE

dataset would result in peatlands being estimated to be stronger GS CO_2 sources and weaker NGS CO_2 sinks. I propose that further validation exercises of the SMAP-NEE data product leverage the wealth of publicly available globally distributed peatland EC-NEE data and make these comparisons available to end users.

On a global basis, derived SMAP-GPP is generally consistent with findings in the literature (Endsley *et al.*, 2023). In the peatland sites I studied, SMAP-GPP values during the GS and NGS had generally good agreement with the values of EC-GPP (Figure AI-4). There are several reasons why SMAP-NEE may not be performing as well as SMAP-GPP at the selected peatland sites in this study. Other land cover types may be represented within the 9 by 9 km pixel modelled by SMAP, and neither wetlands or peatlands are explicitly represented in the MODIS plant functional type classification that were used to calibrate SMAP (Jones *et al.*, 2017). It is also notable that wetlands and peatlands were not well represented in the SMAP-NEE calibration dataset (Jones *et al.*, 2017). These factors may contribute to larger differences in SMAP-NEE compared to EC-NEE that, when aggregated, lead to erroneous sink or source attributions (Figure AI-2). Additionally, it has been reported that SMAP data product has increased the peatland soil porosity compared to previous versions of this data product, which resulted in potentially stronger modeled water limitations relative to other ecosystems (Endsley *et al.*, 2023). The SMAP data products have also highlighted that they have modeled peatlands to be equal or greater summertime CO_2 sinks relative to non-peatlands, which may not necessarily be the case, especially in low productivity bogs (Endsley *et al.*, 2023). While this may be the case, nutrient poor peatlands will still have limitations on respiration fluxes, especially during the winter. Despite these limitations, the application of the SMAP data products to estimate CO_2 flux data in these remote cold regions is valuable. Further

validation for peatland sites of the SMAP algorithms for calculating NEE will be necessary to increase our confidence in the SMAP-derived values for monitoring future change.

2.5 Conclusions

I used the remotely sensed and modeled SMAP-NEE data product to explore NEE for five Canadian peatland sites corrected based on a subset of eddy covariance data. This study is the first study to evaluate the performance of SMAP-NEE across a variety of peatland types within Canada. I found the NGS to represent a period of net CO_2 emission, whereas the GS represented a period of net CO_2 uptake. The proportion of NGS emissions compared to GS uptake varied substantially on a site-by-site basis ranging from 33% to 256%. Our focus on seasonality and identifying different correction factors for either season represents a novel approach, and our findings highlight the importance of considering the NGS when evaluating annual peatland carbon budgets. Furthermore, while peatlands are likely sinks in the long term, their annual sink status is dynamic, and multi-year measurements are crucial for reducing the uncertainty in the estimates of a site's source/sink status. I also found that longer growing seasons were consistent with greater annual net CO_2 uptake at these five peatland sites from 2015-2023. This finding highlights the potential for increased CO_2 uptake with a longer GS expected under climate change scenarios. However, I recommend that future studies consider other gaseous carbon emissions, such as methane (CH_4), as well as fluvial export, as these processes can be significant in peatlands and are crucial in determining the overall C sink status of these environments.

Prediction of the net ecosystem exchange of carbon dioxide in Canadian peatlands by integrating remote sensing and gridded climate data products

E. Kathryn Hettinga[1,*], Fereidoun Rezanezhad[1,2], Ali Reza Shahvaran[1], Mir Amir Mohammad Reshadi[1], Maria Strack, and Philippe Van Cappellen[1,2]

[1]Ecohydrology Research Group, Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, Canada

[2]Water Institute, University of Waterloo, Waterloo, Canada

[*]Corresponding author: E. Kathryn Hettinga

k2hettin@uwaterloo.ca

3 Prediction of the net ecosystem exchange of carbon dioxide in Canadian peatlands by integrating remote sensing and gridded climate data products

3.1 Introduction

Peatlands are globally significant organic carbon (C) reservoirs, storing up to one third of the terrestrial soil organic C pool despite covering just 3% of Earth's land surface (Gorham, 1991; Post, 1982; Yu *et al.*, 2010). This extraordinary accumulation reflects greatly reduced decomposition under the anoxic, water-saturated conditions characteristic of peat soils (Gorham, 1991; Wu & Roulet, 2014). In these ecosystems, gross primary productivity (GPP), the uptake of CO_2 by vegetation, generally exceeds ecosystem respiration (ER), the release of CO_2 by plants and microbes, yielding a negative net ecosystem exchange (NEE) and a net CO_2 sink (Chapin *et al.*, 2006; Wu & Roulet, 2014). Furthermore, peatland CO_2 uptake typically outweighs emissions of methane (CH_4) and dissolved or particulate organic C, reinforcing their role as small but enduring C sinks (Chapin *et al.*, 2006; Wu & Roulet, 2014). Northern peatlands in Canada alone sequester an estimated 20–60 g C m^{-2} yr^{-1} (Strack *et al.*, 2023; Wu & Roulet, 2014).

Two primary *in-situ* approaches are used to measure NEE of CO_2: chamber-based methods and the eddy covariance (EC) technique. In chamber measurements, a sealed chamber is placed over a small patch of vegetation and/or soil, therefore the measurements are highly influenced by vegetation type and local water table position and the rate of CO_2 concentration change inside the chamber is logged over minutes to hours, yielding high-precision and plot-scale fluxes that may not capture the full diversity of plant and soil conditions (Wang *et al.*, 2013; Webster *et al.*, 2018). In contrast, EC, a micrometeorological technique, uses fast-response sensors mounted above the

canopy to record rapid fluctuations in vertical wind speed and CO_2 concentration; by correlating these fluctuations, EC derives continuous, ecosystem-scale flux estimates over a larger region (Baldocchi, 2003). EC systems, comprising fast-response anemometers and CO_2 analyzers mounted on towers above the vegetation canopy, measure NEE continuously across a footprint that typically spans square kilometers (Baldocchi, 2003). Although EC systems require more elaborate setup and quality control processing compared to chamber-based techniques, and can have occasional data gaps under very stable atmospheric conditions (particularly in the non-growing season), their ability to integrate across whole ecosystems outweighs these considerations when upscaling measurements to larger domains (Dutch *et al.*, 2024; Limpens *et al.*, 2008; Wang *et al.*, 2013).

To estimate nation-wide CO_2 budgets and assess climate change impacts over broad regions, *in-situ* CO_2 flux measurements from multiple EC towers must be integrated across wider geographic regions. A growing body of previous studies demonstrates the power of combining EC observations with satellite-derived products to interpolate CO_2 fluxes across landscapes, for example: in deciduous broadleaf forests of Japan (Yamaji *et al.*, 2008), grasslands of the northern Great Plains (Wylie *et al.*, 2007), forests in Europe (Papale & Valentini, 2003), and diverse biomes throughout the continental United States (Xiao *et al.*, 2008). In these studies, a suite of remote-sensing-derived predictors such as surface reflectance, enhanced vegetation index (EVI), daytime and nighttime land surface temperature (LST), normalized difference water index (NDWI), fraction of absorbed photosynthetically active radiation (fPAR), and leaf area index (LAI) has proven highly effective for modeling NEE under varying climatic conditions (Xiao *et al.*, 2008).

Site-scale and data-driven methods have proven effective at simulating NEE across a variety of ecosystems, ranging from forests and grasslands to agricultural systems (Cai *et al.*, 2020; Dou & Yang, 2018; He *et al.*, 2006; Melesse & Hanley, 2005). In peatlands specifically, several modeling paradigms have been applied at individual tower sites, for example: process-based models (Wu *et al.*, 2011; Wu & Roulet, 2014), purely data-driven approaches (Rafat *et al.*, 2021), and semi-empirical formulations (Balogun *et al.*, 2023; Balogun *et al.*, 2024). Meanwhile, in other biomes, site-level EC observations have been expanded to regional analyses using machine learning and regression techniques, for example: across North American forests (Xiao *et al.*, 2008), global FLUXNET sites (Jung *et al.*, 2011; Tramontana *et al.*, 2016), and European carbon budgets (Beer *et al.*, 2010; Bodesheim *et al.*, 2018; Virkkala *et al.*, 2021), with similar upscaling efforts in wetlands (McNicol *et al.*, 2023; Peltola *et al.*, 2019). By contrast, few studies have attempted to apply these regional-scale data-driven strategies to peatland NEE, either across the Northern Hemisphere (Qiu *et al.*, 2020) or within Canada (Webster *et al.* 2018).

In this study, I integrated year-round EC-derived NEE of CO_2 measurements from 15 Canadian peatland sites spanning six ecoregions with 34 candidate predictors from remote sensing and gridded climate datasets. The main objectives were to: (i) train, validate, and test four machine learning algorithms: ElasticNet Regression (EN), Light Gradient-Boosting Machine (LGBM), Random Forest Regression (RFR), and Support Vector Regression (SVR) with their predictive performance being evaluated via R^2, Root Mean Square Error (RMSE), Mean Square Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percent Error (MAPE) (See *3.2.3.2*); (ii) execute a comprehensive feature importance and selection workflow including hierarchical clustering, Gini importance, and mRMR analysis to narrow inputs to the eight most influential drivers of NEE in

Canadian peatlands ; and (iii) apply the top-performing model to six independent validation peatland sites to quantify predictive accuracy and assess regional transferability. By achieving these objectives, this study provides a generalizable framework for upscaling NEE estimates across Canada's under-studied peatlands by leveraging a model trained on diverse ecoregions based on key satellite-derived and climate predictors. By pinpointing the eight most influential variables, this study shows how remote sensing and climate products can effectively complement limited, and costly, *in-situ* measurements, enabling seamless application to Canadian peatlands.

3.2 Materials and Methods

3.2.1 Data collection and preprocessing

The study proceeded in four stages: i) data collection and preprocessing of eddy covariance (EC) fluxes and gridded predictors (see 3.*2.1*), ii) training and evaluation of four machine learning models: ElasticNet Regression (EN), Light Gradient-Boosting Machine (LGBM), Random Forest Regression (RFR), Support Vector Regression (SVR) (see 3.*2.2*), iii) SHAP (SHapley Additive exPlanations) based assessment of predictor importance (see 3.*2.3*), and iv) external validation and application across independent peatland sites (see 3.*2.4).* A schematic of methodological flowchart is shown in Figure 3-1.

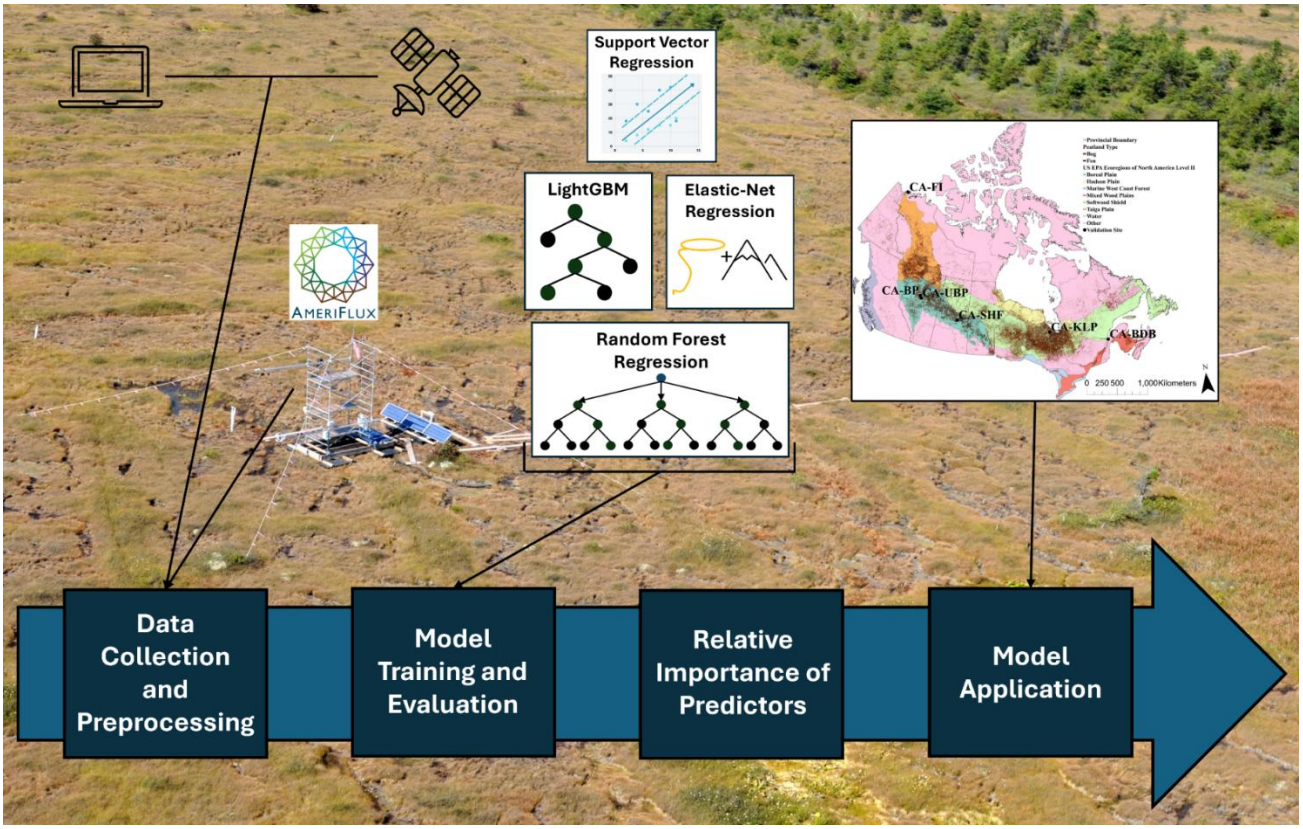


Figure 3-1: Workflow schematic illustrating the four key stages of this study: (i) data collection and preprocessing of eddy covariance (EC) fluxes and gridded predictors, (ii) training and evaluation of four machine learning models: ElasticNet Regression (EN), Light Gradient-Boosting Machine (LGBM), Random Forest Regression (RFR), Support Vector Regression (SVR), (iii) SHAP (SHapley Additive exPlanations)-based assessment of predictor importance, and (iv) external validation and application across independent peatland sites. The peatland pictured is CA-DBB (Photo Credit: Metro Vancouver).

3.2.1.1 Site selection

Half-hourly NEE flux data of 15 Canadian peatland sites (Figure 3-2 and Table 3-1) with continuous EC-derived CO_2 records was collected from AmeriFlux network (AmeriFlux, 2025). From a hydrological perspective, the selected peatland sites include eight bogs (ombrotrophic systems receiving water and nutrients solely from precipitation and dry deposition) and seven fens (receiving additional water and nutrient inputs from connections with the ground and surface-water) (National Wetlands Working Group, 1988), capturing the contrasting carbon-storage and CO_2-flux dynamics of each peatland type (Bridgham *et al.*, 2008). From a disturbance-gradient

perspective, two sites (CA-DBB and CA-DB2) within Burns Bog Ecological Conservancy Area were subject to peat extraction in 1957–1963 and land use change in the surrounding urban area and have undergone phased restoration since 2001 (Christen *et al.*, 2016; Christen & Knox, 2021; D'Acunha *et al.*, 2019; Lee *et al.*, 2017), while the remaining thirteen are undisturbed to the best of our knowledge. Collectively, these fifteen sites encompass a variety of different peatland type, and have some representations of pristine, disturbed, and restored sites, providing a robust foundation for training and evaluating our predictive models. EC system configurations for each site are detailed in Table 3-2.

Table 3-1: Characteristics of the selected 15 peatland study sites: site identifier, geographic coordinates, EPA Level II ecoregion, Köppen–Geiger climate classification, mean annual temperature and precipitation, peatland hydrological type, and primary data citation (ECCC, 2025; AmeriFlux, 2025).

AmeriFlux Site ID	Coordinates (°)		Ecoregion	Koppen Climate Classification*	Mean Annual Temperature (°C)	Mean Annual Precipitation (mm)	Peatland Classification	Reference
	Latitude	Longitude						
CA-ARB	52.6950	−83.9452	Hudson Plain	Dfb	−1.30	700.00	Bog	Helbig *et al.*, 2019
CA-ARF	52.7008	−83.9550	Hudson Plain	Dfb	−1.30	700.00	Fen	Helbig *et al.*, 2019
CA-BOU	50.5244	−63.2064	Softwood Shield	Cwc	1.50	1011.00	Bog	Primeau & Garneau, 2021
CA-CF1	58.6658	−93.8300	Hudson Plain	Dfc	−6.50	452.00	Fen	Hanis *et al.*, 2015
CA-CF2	58.6658	−93.8300	Hudson Plain	Dfc	−6.50	452.00	Fen	Hanis *et al.*, 2015
CA-CF3	58.6658	−93.8304	Hudson Plain	Dfc	−6.50	452.00	Fen	Arndt & Natali, 2025
CA-DB2	49.1190	−122.9951	Marine West Coast Forest	Csb	10.00	1128.00	Bog	D'Acunha *et al*., (2019)
CA-DBB	49.1293	−122.9849	Marine West Coast Forest	Csb	10.00	1128.00	Bog	Christen *et al.,* 2016
CA-Mer	45.4094	−75.5186	Mixed Wood Plains	Dfb	6.00	943.00	Bog	Lafleur *et al.*, 2001
CA-RPn	48.2604	−58.6632	Softwood Sheild	Dfb	5.00	1340.40	Bog	Wang *et al*., 2018
CA-SCB	61.3089	−121.2984	Taiga Plain	Dfc	−2.80	388.00	Bog	Helbig *et al.*, (2017)
CA-SMC	63.1534	−123.2522	Taiga Plain	Dfc	−2.80	388.00	Bog	Chartrand *et al.*, 2023
CA-WP1	54.9538	−112.4670	Boreal Plain	Dfc	1.10	461.07	Fen	Coursolle *et al.*, 2006
CA-WP2	55.5375	−112.3343	Boreal Plain	Dfc	1.20	428.30	Fen	Glenn *et al.*, 2006
CA-WP3	54.4700	−113.3200	Boreal Plain	Dfc	2.30	478.70	Fen	Glenn *et al.*, 2006

*Dfb = a continental climate with no dry season and warm summers; Cwc = a temperate climate with dry winters and cool summers; Dfc = a continental climate with no dry season and cool summers; Csb = a temperate climate with dry and warm summers.

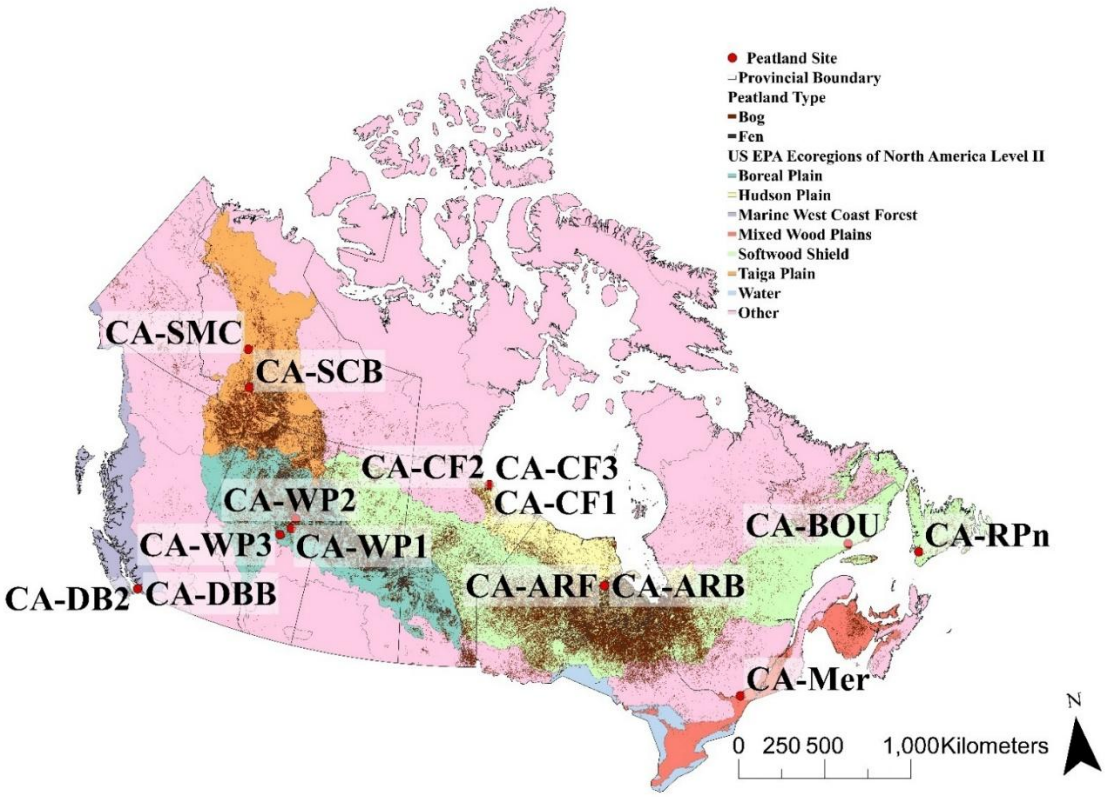


Figure 3-2: Geographic distribution of the 15 selected peatland sites across Canada, overlaid on the U.S. EPA Level II Ecoregions of North America (Environmental Protection Agency, 2024).

Table 3-2: Eddy covariance setup and data characteristics. Different superscript letters indicate statistical significance between 15 selected sites (Kruskal-Wallis, $p<0.05$) (NEE = Net Ecosystem Exchange).

Site	Data Source	Instrumentation	Path Type	Height (m)	Period	u^* Threshold (m s^{-1})	Half-Hourly Gap (%)	Monthly mean NEE (g C m^{-2} day; 95% CI)	Number of Observations	Reference
CA-ARB	Ameriflux	IRGA: LI-7200	Closed-path	7.5	2011–2015	0.28	33%	-0.19^{ab} (−0.30, −0.08)	57	Helbig *et al.*, 2019; Humphreys *et al.* (2014)
CA-ARF	Ameriflux	IRGA: LI-7200	Closed-path	7.5	2011–2015	0.19	30%	-0.13^{ab} (−0.30, 0.04)	54	Helbig *et al.*, 2019
CA-BOU	Ameriflux	IRGA: LI-7200	Closed-path	2.2	2018–2020	0.07	47%	-0.13^{a} (−0.25, −0.03)	17	Primeau & Garneau, 2021; Taillardat *et al.*, 2022
CA-CF1	Ameriflux	IRGA: LI-7500	Open-path	3.9	2007–2008	0.08	72%	0.10^{ab} (−0.18, 0.40)	7	Hanis *et al.*, 2013; Hanis *et al.*, 2015
CA-CF2	Ameriflux	IRGA: LI-7500	Open-path	4.3	2008–2011	0.20	77%	0.07^{ab} (−0.23, 0.37)	18	Hanis *et al.*, 2015
CA-CF3	Ameriflux	IRGA: LI-7500; IRGA: LI-7200	Open-path; Closed-path	2.5	2022–2023	0.07	45%	-0.21^{abc} (−0.79, 0.37)	6	Arndt & Natali, 2025
CA-DB2	Ameriflux	IRGA: LI-7200	Closed-path	2.5	2019–2020	0.05	15%	-0.49^{abc} (−0.97, 0.001)	12	Christen *et al.*, (2016); Lee *et al.*, (2017); D'Acunha *et al.*, (2019)
CA-DBB	Ameriflux	IRGA: LI-7200	Closed-path	1.8	2014–2020	0.05	37%	-0.26^{abc} (−0.45, −0.07)	71	Christen *et al.*, (2016); Lee *et al.*, (2017); D'Acunha *et al.*, (2019)
CA-Mer	Ameriflux	IRGA: LI-7000	Closed-path	3.0	1998–2020	0.09	37%	-0.27^{abc} (−0.36, −0.18)	215	Humphreys *et al.* (2014)
CA-RPn	Site PI	IRGA: LI-7200	Closed-path	3.21	2014–2016	-*	-*	-0.10^{ab} (−0.37, 0.17)	22	Wang *et al.* (2018)
CA-SCB	Ameriflux	IRGA: EC150	Open-path	1.9	2014–2019	0.05	44%	0.02^{ab} (−0.09, 0.12)	31	Helbig *et al.*, (2017)
CA-SMC	Ameriflux	IRGA: LI-7200	Closed-path	14.9	2018	0.10	50%	0.30^{b} (0.04, 0.57)	5	Kuhn *et al.*, 2024; Schulze *et al.*, 2023
CA-WP1	Ameriflux	IRGA: LI-7000	Closed-path	9.0	2003–2009	0.16	43%	-0.28^{ab} (−0.51, −0.06)	72	Syed *et al.*, 2006
CA-WP2	Ameriflux	IRGA: LI-7500	Open-path	3.0	2004–2006	0.05	48%	-0.63^{c} (−0.90, −0.36)	16	Glenn *et al.*, 2006
CA-WP3	Ameriflux	IRGA: LI-7500	Open-path	3.0	2004–2006	0.15	56%	-0.43^{ab} (−0.86, −0.02)	14	Glenn *et al.*, 2006

*Datasets were preprocessed (quality filtered and gap-filled) by site principal investigators.

3.2.1.2 Observation-based NEE fluxes

Half-hourly NEE flux data of 15 selected peatland sites were also obtained from the AmeriFlux BASE archive and processed in REddyProcWeb (Wutzler *et al.*, 2018) to apply friction-velocity (*u**) filtering, gap-filling, and partitioning of NEE into GPP and ecosystem respiration (ER). Low-turbulence periods, when the EC method underestimates fluxes (Papale *et al.*, 2006), were identified and removed using seasonally calculated *u** thresholds, after which gaps were filled via meteorologically stratified lookup tables and temporal interpolation following Falge *et al.* (2001) and Reichstein *et al.* (2005). Processed data were then aggregated to daily and monthly averages under completeness criteria (≥70% half-hourly data per day; ≥10 days per month). This workflow adheres to established protocols (Papale *et al.*, 2006; Wutzler *et al.*, 2018), with full methodological details provided in Appendix II.

Using the accepted daily and monthly averages, I assessed whether smoothing the high-frequency variability would alter these results, I also applied a fast Fourier transform (FFT) via numpy.fft.fft (Harris *et al.*, 2020) to the raw half-hourly time series, decomposing the signal into sine and cosine components, zeroing frequencies above the 24-hour threshold (low-pass filtering), and reconstructing the smoothed series with an inverse FFT, before aggregating those values to monthly averages. Comparison of the two approaches showed that our completeness-based aggregation already effectively removed high-frequency noise: unfiltered and FFT-smoothed monthly NEE agreed within MAE 0.00–0.01 g C m^{-2} day^{-1}, RMSE 0.00–0.02 g C m^{-2} day^{-1}, and MAPE 1–6% (Figure AII-1), rendering the additional FFT step unnecessary.

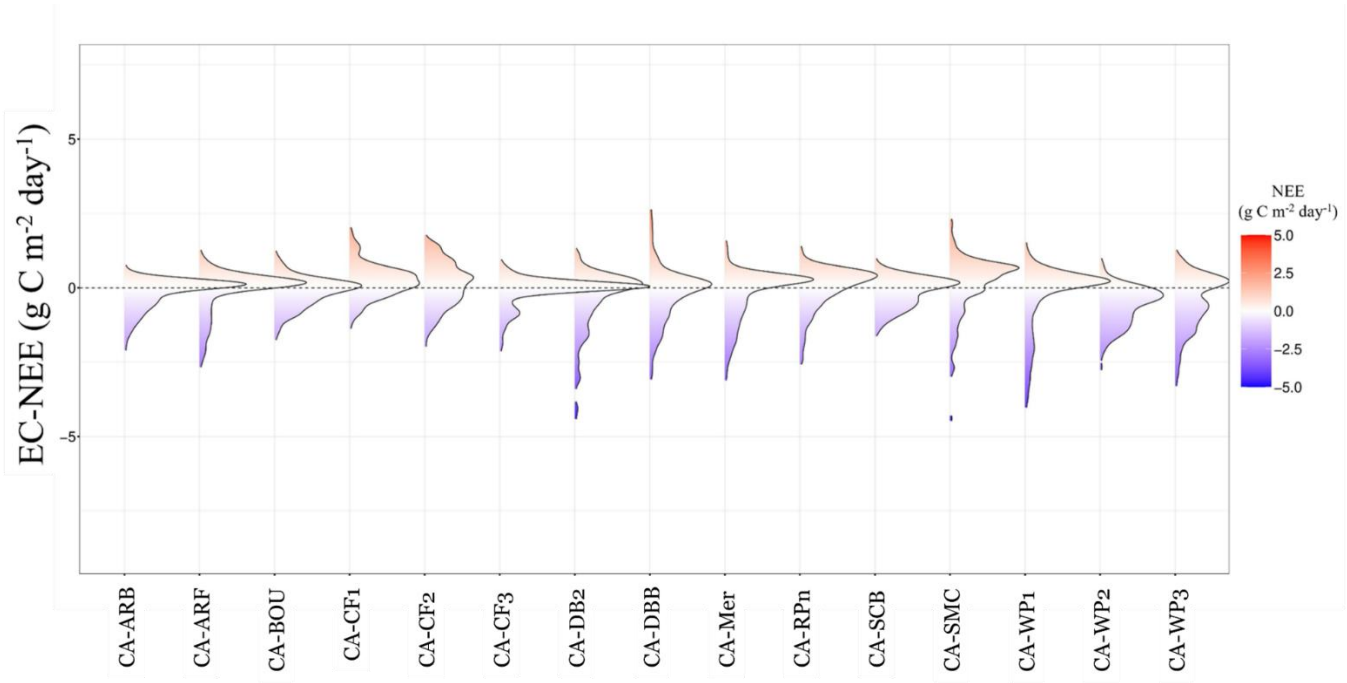


Figure 3-3: Ridgeline distributions of daily eddy covariance net ecosystem exchange (EC-NEE; g C m^{-2} day^{-1}) at each peatland site (*x*-axis labels), with negative values (net CO_2 uptake) in blue and positive values (net CO_2 release) in red. Horizontal dashed line at zero distinguishes uptake versus release.

3.2.1.3 Gridded predictor products

Thirty-four predictor variables were selected based on their relevance to modeling NEE of CO_2 in peatlands (see Table 3-3). To capture the high temporal variability of peatland CO_2 fluxes, static predictors or those with coarser-than-monthly resolution were excluded from collection. All predictor variables are accessible via Google Earth Engine (GEE) and Table 3-3 summarizes each predictor's temporal cadence, spatial resolution, unit, and source. Given that EC flux tower footprints can span <1–3 km^2 (Baldocchi, *et al.*, 1997; Running *et al.*, 2004), predictor data were extracted by averaging values within a 1 km buffer around the tower coordinates. For coarser resolution datasets (>5 km), the pixel value containing the flux tower was used directly. Data extraction followed each variable's native cadence. The algorithm matched each daily averaged NEE observation to the closest available predictor value within a ±4-day window, selecting the

date with the smallest offset when multiple candidates existed. Additionally, pre-processing steps (unit conversions and merging dual-platform products, *e.g.,* MODIS Terra and Aqua) were applied to ensure consistency across datasets before model development. Once all daily predictor matchups were compiled and harmonized, they were aggregated to monthly averages, yielding a final dataset of monthly-mean environmental predictors paired with monthly-mean NEE as the target variable. Data were aggregated to monthly-mean to make use of the wealth of variables in the TerraClimate dataset (monthly resolution), but also to dampen extreme daily values. Moreover, Kruskal Wallis and Wilcoxon comparisons were used to determine the differences between different measures of daily NEE and the gridded-predictor variables, partitioned by peatland type and ecoregion.

Table 3-3: Summary of 32 candidate predictor variables extracted from Google Earth Engine (GEE). The remaining two (in total 34) are the peatland and ecoregion classifications presented in Table 3-1. GEE datasets were accessed via https://developers.google.com/earth-engine/datasets/catalog/ concatenated with the respective Image Collection ID.

Predictor Variable	Shortform	Unit	Cadence	Spatial Resolution (km)	GEE Image Collection ID	Time Span
Fraction of Photosynthetically Active Radiation (fPAR) absorbed by vegetation	fPAR	-	4 Days	0.5	MODIS_061_MCD15A3H	2002–2025
Leaf Area Index (LAI)	LAI					
Average Daytime Land Surface Temperature (LST)	DLST	°C	1 Day	1	MODIS_061_MOD21C1 MODIS_061_MYD21C1	2000–2025
Average Nighttime Land Surface Temperature (LST)	NLST				MODIS_061_MOD21C1 MODIS_061_MYD21C1	
Total Evapotranspiration (ET)	ET_M	mm day^{-1}	8 Days	0.5	MODIS_061_MOD16A2	2001–2024
Total Evapotranspiration (ET) gap-filled*	ET_M				MODIS_061_MOD16A2GF	
Gross Primary Product (GPP)	GPP	gC m^{-2} day^{-1}	8 Days	0.5	CAS_IGSNRR_PML_V2_v018	2001–2024
Normalized Difference Water Index (NDWI)	NDWI	-	1 Day	~0.5	MODIS_MCD43A4_006_NDWI	2000–2023
Normalized Difference Snow Index (NDSI)*	NDSI				MODIS_MCD43A4_006_NDSI	
Normalized Difference Vegetation Index (NDVI)	NDVI				MODIS_MCD43A4_006_NDVI	
Enhanced Vegetation Index (EVI)	EVI				MODIS_MCD43A4_006_EVI	
Burn Area Index (BAI)*	BAI				MODIS_MCD43A4_006_BAI	
Duration of the daylight period	Dayl	hr	1 Day	1	NASA_ORNL_DAYMET_V4	1980–2024
Daily total precipitation	prcp_DM	mm				
Incident shortwave radiation flux density*	Srad_DM	W m^{-2}				
Snow water equivalent*	swe_DM	kg m^{-2}				
Daily maximum 2-meter air temperature	Tmax_DM	°C				
Daily minimum 2-meter air temperature	Tmin_DM	°C				
Daily average partial pressure of water vapor	vp_DM	kPa				
Actual evapotranspiration	aet	mm	Monthly	~5	IDAHO_EPSCOR_TERRACLIMATE	1958–2024
Climate water deficit*	cwd	mm				
Palmer Drought Severity Index (PDSI)	PDSI	-				
Potential evapotranspiration	pet	mm				
Precipitation accumulation	prcp_TC	mm				
Runoff	ro	mm				
Soil moisture*	sm	mm				
Downward surface shortwave radiation	srad_TC	W m^{-2}				
Snow water equivalent	swe_TC	mm				
Maximum temperature	Tmax_TC	°C				
Minimum temperature	Tmin_TC	°C				
Vapor pressure	vp_TC	kPa				
Vapor pressure deficit	vpd	kPa				
Wind-speed at 10m*	ws	m s^{-1}				

* Indicates selected feature.

3.2.2 Applied machine learning models

3.2.2.1 ElasticNet (EN) regression

ElasticNet (EN) is a highly interpretable multiple linear regression model that brings together two regularization methods: the Least absolute shrinkage and selection operator (Lasso) (L1) and Ridge (L2) penalties (Zou & Hastie, 2005). Its strength lies in its simplicity while retaining the flexibility of more complex approaches (Kujawska *et al.*, 2025). This model has been used in a variety of studies, such as Harrison *et al.* (2021) for selecting predictors in models of CO_2 and CH_4 exchange from reservoir systems and Kujawska *et al.* (2025) for evaluating and predicting CO_2 flux from agricultural soils under various treatments. The EN model is defined by:

$$Y = X\beta + \varepsilon \qquad (3.1)$$

where Y represents the target variable (NEE), X is the matrix of input variables (Table 3-3), β is the vector of coefficients, and ε represents the error term (Kujawska *et al.*, 2025).

The first regularization component, Lasso, imposes an L1 penalty on coefficient magnitudes, constraining the sum of absolute values and forcing some coefficients exactly to zero, thereby performing variable selection in addition to reducing multicollinearity (Tibshirani, 1996). The second regularization component in EN is the ridge (L2) penalty, originally devised to mitigate multicollinearity in ordinary least-squares regression by Hoerl and Kennard (1970). When predictor variables are highly correlated, coefficient estimates can become unstable, degrading model performance; ridge regression addresses this by adding a small bias term, denoted k, to the sum of squared coefficients, effectively shrinking them toward zero and reducing their variance, thereby enhancing predictive accuracy (Hoerl & Kennard, 1970). In implementation for this study, the optimal ridge parameter k was selected via scikit-learn's ElasticNetCV, which evaluates

multiple values through cross-validation and chooses the one that minimizes prediction error (Pedregosa *et al.*, 2011). In the EN framework, this ridge parameter is incorporated into the term $\alpha(1 - \text{l1_ratio})\|w\|_2^2$. Importantly, ridge regularization retains all variables in the model. By combining these two penalties, EN balances coefficient shrinkage and variable selection to improve predictive accuracy. Formally, EN minimizes the following objective (Zou & Hastie, 2005):

$$\min_w \frac{1}{2n}\|y - Xw\|_2^2 + \alpha\theta\|w\|_1 + \frac{1}{2}\alpha(1-\theta)\|w\|_2^2 \qquad (3.2)$$

where n is the number of samples, y is the vector of observed values, X is the matrix of input variables (Table 3-3), w is the coefficient vector, $\|y - Xw\|\frac{2}{2}$ is the residual sum of squares, $\|w\|_1$ and $\|w\|_2^2$ are the L1 and L2 norms of the coefficients, α is the overall regularization strength, and θ is the mixing parameter that controls the relative weight of L1 versus L2 penalties (Pedregosa *et al.*, 2011; Zou & Hastie, 2005).

3.2.2.2 Light Gradient-Boosting Machine (LGBM)

Light Gradient-Boosting Machine (LGBM) is an optimized implementation of gradient-boosting decision trees (GBDT) designed for high efficiency and scalability (Ke *et al.*, 2017). Like conventional GBDT, it builds decision trees sequentially, with each tree correcting the errors of its predecessors to achieve high predictive accuracy (Ke *et al.*, 2017). However, traditional GBDT approaches can be computationally expensive, so LGBM incorporates two key techniques: gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB) to reduce both the number of data instances and features used during training (Ke *et al.*, 2017). GOSS leverages gradient magnitudes, where larger gradients indicate data points with higher error contribution to

preferentially retain those informative instances and randomly down-sample lower-gradient instances, thereby shrinking the training set without sacrificing key information (Ke *et al.*, 2017). EFB, in turn, bundles mutually exclusive features, those never simultaneously non-zero, into single composite features with further reducing dimensionality (Ke *et al.*, 2017). Moreover, LGBM adopts a leaf-wise growth strategy, prioritizing the growth of leaf nodes minimizing loss at each step rather than growing trees level-by-level, which accelerates convergence relative to traditional tree-based methods (Wu *et al.*, 2024). LGBM has been demonstrated to achieve comparable accuracy with other tree-based machine learning methods while lowering computational costs (Ke *et al.*, 2017). This method along with random forest regression has been used to model oceanic CO_2 fluxes (Zeng *et al.*, 2022) and CO_2 fluxes from underground coal fires (Wang *et al.*, 2025).

3.2.2.3 Random Forest Regression (RFR)

Random forest regression (RFR) is an ensemble learning method that constructs a multitude of decision trees using bootstrap samples of the training data and random subsets of predictor variables at each split (Breiman, 2001). By injecting randomness in both the sampling of observations and the selection of features, individual trees become decorrelated, dramatically reducing overfitting compared to single-tree models (Breiman, 2001). During prediction, the forest's output is obtained by averaging the predictions of all constituent trees, thereby stabilizing variance and improving generalization. Beyond its use in conjunction with LGBM (see 3.*2.2.2*), random forest algorithms have been successfully applied to model greenhouse gas (GHG) fluxes in wetland environments, including CH_4 emissions in boreal and temperate peatlands (McNicol *et al.*, 2023; Peltola *et al.*, 2019), as well as other C exchange studies across diverse ecosystem types (Cai *et al.*, 2020; Virkkala *et al.*, 2021).

3.2.2.4 Support Vector Regression (SVR)

Support Vector Regression (SVR) is a powerful extension of the Support Vector Machine (SVM) framework, repurposed for regression rather than classification. In SVR, input vectors $x \in \mathbb{R}^n$ are first mapped into a high-dimensional feature space via a nonlinear function $\phi(x)$, enabling the model to capture complex, nonlinear relationships in the data. The prediction function is then defined as:

$$f(x) = \omega^\top \phi(x) + b \tag{3.3}$$

where ω is the weight vector that determines the orientation of the regression hyperplane in feature space, and b is the bias term that shifts this hyperplane to best fit the data. Together, ω and b form the parameters of the regression function, balancing the model's capacity to follow the data's underlying pattern against its complexity (Cortes & Vapnik, 1995).

The training of an SVR model involves solving a convex optimization problem that simultaneously minimizes model complexity and penalizes deviations beyond a user-specified tolerance ε. This is formulated as:

$$\min_{\omega,b} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{m} L_\epsilon(f(x_i) - y_i) \tag{3.4}$$

where $\|\omega\|^2$ measures the squared Euclidean norm of the weight vector (thus controlling model smoothness), m is the number of training samples, and $C > 0$ is the regularization parameter that governs the trade-off between the flatness of the regression function and the penalty for errors. The function $L_\epsilon(e) = \max\{0, |e| - \epsilon\}$ is the ϵ-insensitive loss, meaning deviations $|f(x_i) - y_i|$ smaller than ϵ incur no penalty, while larger deviations are penalized linearly. Slack variables embedded in L_ϵ quantify the extent to which points lie outside the ϵ-tube, and C determines how heavily these slack variables are weighted in the overall risk minimization (Drucker *et al.*, 1996).

This dual objective ensures robust generalization even with limited data and enables SVR to adapt to both linear and highly nonlinear trends through the choice of kernel in $\phi(x)$.

3.2.3 Modeling workflow

3.2.3.1 Feature selection

Hierarchical clustering-based Exploratory Data Analysis (EDA) revealed substantial redundancy among predictors, underscoring the need for dimensionality reduction. In this study, it began by computing Spearman's pairwise correlations and flagging feature pairs with $\rho \geq 0.85$ (Figure 3-4). A random forest regressor was then trained to assess Gini importance for each feature (Cai *et al.*, 2020); within each highly correlated pair, the feature with lower importance was discarded. This procedure reduced the predictor set from 34 to 13 variables. Next, the minimum redundancy maximum relevance (mRMR) algorithm was applied to the remaining 13 features using mrmr_selection library. Originally developed for gene-selection tasks (Ding & Peng, 2003; Peng *et al.*, 2005), mRMR selects the feature with highest relevance (to the target variable in this study; NEE) first, then incrementally adds variables that maximize relevance while minimizing redundancy, producing a ranked list by mRMR score. mRMR has been adapted to environmental modeling (Bouzgou & Gueymard, 2017; Cushman *et al.*, 2024; Reshadi *et al.*, 2025; Wu *et al.*, 2023). Categorical predictors (Ecoregion and Peatland Type) were one-hot encoded and treated numerically; for continuous predictors, relevance to NEE was evaluated via F-statistics (Analysis of Variance; ANOVA), and for binary variables via t-tests. Redundancy among all variable pairs was quantified by (i) absolute Pearson correlation, (ii) Euclidean distance, and (iii) inverse distance (Ding & Peng, 2003).

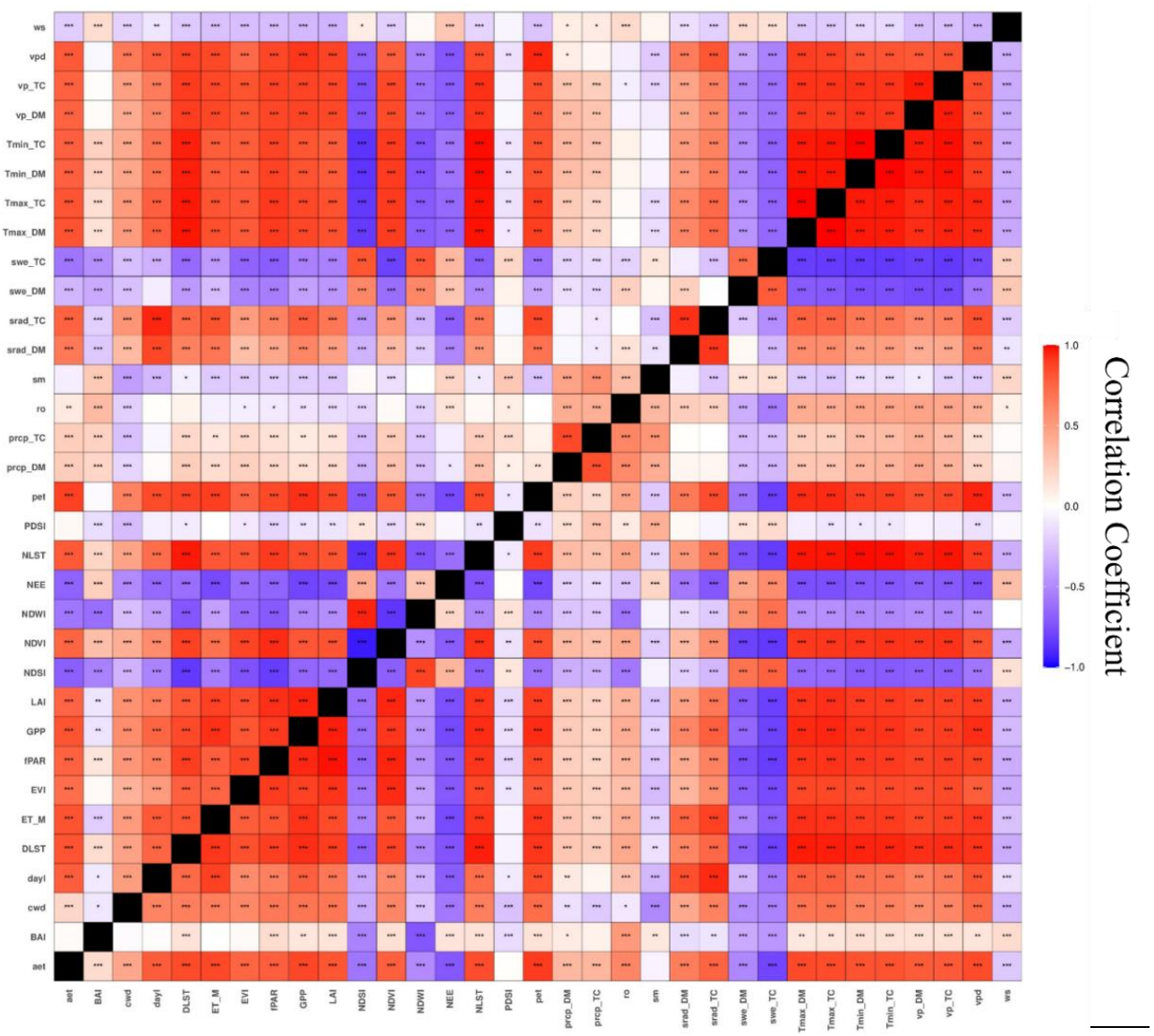


Figure 3-4: Pairwise correlations among monthly-mean predictor variables and net ecosystem exchange (NEE). The upper triangle displays Pearson's $\boldsymbol{r}$, the lower triangle Spearman's $\boldsymbol{\rho}$. Cell colors range from blue (−1 for strong negative) to red (+1 for strong positive). Asterisks indicate significance levels: * $\boldsymbol{p} < 0.05$, ** $\boldsymbol{p} < 0.01$, *** $\boldsymbol{p} < 0.001$. Variable shortforms are expressed in Table 3-3.

To identify the optimal feature count, a sensitivity analysis was performed using ordinary least squares regression on subsets ranging from 2 to 13 features. Model performance improved sharply up to eight predictors, beyond which performance metrics plateaued. Consequently, the top eight mRMR-ranked features: evapotranspiration, shortwave radiation, burn area index (BAI),

normalized difference snow index (NDSI), snow water equivalent (SWE), climate water deficit (CWD), wind speed, and soil moisture, were selected for subsequent model development (see 3.*3.4* for detailed results).

3.2.3.2 Model development and evaluation

The modeling workflow used in this study was built on Reshadi *et al.* (2025) with two main adaptations: EN was included alongside LGBM, RFR, and SVR, and interpretation was expanded with a more comprehensive SHAP (SHapley Additive exPlanations)-based interpretation. For each model, the data were split into a training set (75%) and a hold-out internal test set (25%), stratified by site. Hyperparameters were tuned using five-fold cross-validation on the training data, which was further split into five folds, each fold serving once as a validation set while the remaining four folds were used for fitting, and the best combination was chosen by averaging performance across folds. Model performance on the unseen internal test data was assessed using five statistical metrics (MAE = Mean Absolute Error, MSE = Mean Squared Error, RMSE = Root Mean Squared Error, MAPE = Median Absolute Percentage Error, R^2 = Coefficient of Determination):

$$MAE = 10^{\frac{1}{n}\sum|\log(y_i)-\log(\hat{y}_i)|} \tag{3.5}$$

$$MSE = \frac{1}{n}\sum(y_i - \hat{y}_i)^2 \tag{3.6}$$

$$RMSE\ (g\ C\ m^{-2}\ day^{-1}) = \sqrt{\frac{1}{n}\sum(y_i - \hat{y}_i)^2} \tag{3.7}$$

$$MAPE\ (\%) = Med\left(\frac{|y_i-\hat{y}_i|}{\hat{y}_i}\right) \times 100 \tag{3.8}$$

$$R^2 = 1 - \frac{\sum(y_i-\hat{y}_i)^2}{\sum(y_i-\bar{\hat{y}}_i)^2} \tag{3.9}$$

where y_i and $\hat{y}_i$ denote the measured and predicted NEE of the i-th data point, respectively. The $\sum$ represents summation from $i = 1$ to n, where n is the total number of data points, and Med represents the median.

To quantify uncertainty and guard against overfitting, the Jackknife+ resampling technique was applied (Barber *et al.*, 2021), which iteratively leaves out one observation, fits the model on the remaining $n - 1$ samples, and predicts the held-out point; these n predictions form a distribution from which a prediction interval is constructed.

3.2.3.3 Model interpretation

To elucidate each predictor's influence on NEE of CO_2 estimates, Shapley additive explanation (SHAP) analysis was applied (Lundberg & Lee, 2017), which assigns contribution values that capture both local (individual prediction) and global (overall model) feature importance (Wu *et al.*, 2024). In addition, a leave-one-ecoregion-out sensitivity analysis was conducted by iteratively omitting each ecoregion from the training data, retraining the model, and evaluating performance changes to assess the model's reliance on specific regional data.

3.2.4 External validation and model application

The Web of Science database was searched for additional peatland sites with publicly accessible or Principal Investigator (PI)-provided NEE data for external validation. From this search, datasets for six external validation sites were acquired (see Table 3-4). For each site, the same eight predictor variables were collected from GEE following the procedure described in *3.2.1.3*. Half-hourly flux data for Bois-des-Bel (CA-BDB) and Sandhill Fen (CA-SHF) sites were processed using the workflow described in *3.2.1.2*. Data for the Burned Peatland (CA-BP) and Unburned

Peatland (CA-UBP) sites were supplied by the site PIs and processed according to Morison *et al.* (2020). The Fish Island Peatland (CA-FI) site dataset, as described by Skeeter *et al.* (2022), was likewise subjected to the same post-processing protocol outlined in *3.2.1.2*.

Table 3-4: Summary of external validation site descriptions: site identifier, geographic coordinates (latitude and longitude), EPA Level II ecoregion, Köppen–Geiger climate classification, mean annual temperature and precipitation, peatland classification, and primary data citation (ECCC, 2025).

Site ID	Coordinates (°)		Ecoregion	Koppen Climate Classification*	Mean Annual Temperature (°C)	Mean Annual Precipitation (mm)	Peatland Classification	Reference
	Latitude	Longitude						
CA-BDB	47.9672	−69.4287	Atlantic Highlands	Dfb	3.5	963.5	Bog	He *et al.*, 2025; Nugent *et al.*, 2018
CA-BP	56.1070	−115.5610	Boreal Plain	Dfc	1.7	462.1	Fen	Morison *et al.*, 2020
CA-FI	69.3723	−134.8810	Southern Arctic	ET	−10.1	160.7	Bog	Skeeter *et al.*, 2022
CA-UBP	55.8000	−115.1000	Boreal Plain	Dfc	1.7	462.1	Fen	Morison *et al.*, 2020
CA-KLP	51.5902	−81.7684	Hudson Plain	Dfb	−0.5	703.6	Bog	Humphreys *et al.*, 2014
CA-SHF	53.8021	−104.6180	Boreal Plain	Dfb	0.9	427.3	Fen	Helgason *et al.*, 2024

*Dfb = a continental climate with no dry season and warm summers; Dfc = a continental climate with no dry season and cool summers; ET = polar, tundra.

3.3 Results

3.3.1 Distribution of NEE

At the daily timescale, average NEE fluxes varied widely across sites (Figure 3-3), ranging from −8.37 to 6.94 g C m^{-2} day^{-1} (mean = −0.22 g C m^{-2} day^{-1}). Aggregating to monthly averages both dampened extreme values and preserved the overall mean, yielding a range of −2.74 to 2.09 g C m^{-2} day^{-1} (mean = −0.22 g C m^{-2} day^{-1}). These monthly averages closely matched those derived from FFT-smoothed data (MAE: 0.00–0.01 g C m^{-2} day^{-1}; RMSE: 0.00–0.02 g C m^{-2} day^{-1}; MAPE: 1–6%; see Figure AII-1). Although some site-level means differed significantly (Table 3-2), there were no significant differences between bogs (mean = −0.22 g C m^{-2} day^{-1}; 95% CI: −0.28 to −0.16; n = 430) and fens (mean = −0.22 g C m^{-2} day^{-1}; 95% CI: −0.34 to −0.12; n = 187;

Wilcoxon W = 37992, p = 0.27), nor among ecoregions (Kruskal–Wallis F = 9.96, p = 0.07; pairwise Wilcoxon). NEE displayed pronounced seasonality: strong CO_2 uptake (negative NEE) occurred from May through October, while net emission (positive NEE) prevailed during the remainder of the year (January–April and October–December) (Figure 3-5).

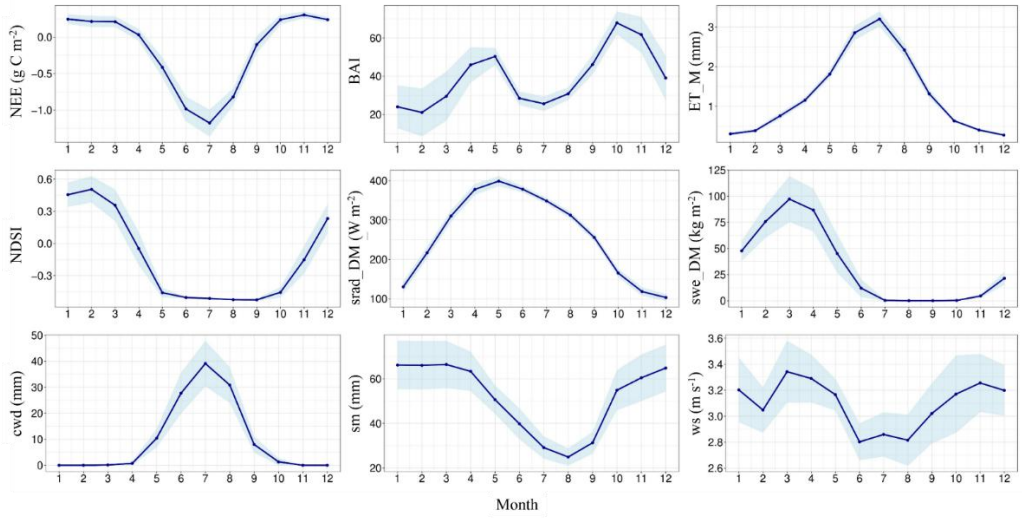


Figure 3-5: Seasonal cycles of monthly-mean net ecosystem exchange (NEE) and the eight key predictor variables: evapotranspiration (ET_M), shortwave radiation (srad_DM), burn area index (BAI), normalized difference snow index (NDSI), snow water equivalent (swe_DM), climate water deficit (cwd), wind speed (ws), and soil moisture (sm), with 95% confidence intervals (CI). Lines show the average of all site-level monthly means; shaded ribbons denote the 2.5th–97.5th percentiles across sites. Monthly means were computed from daily averages (requiring ≥70% half-hourly coverage per day and ≥10 days per month).

3.3.2 Distribution of predictors

Of the 34 candidate features initially assembled (Table 3-3), our feature selection workflow indicated that model performance stabilized once eight predictors were retained: evapotranspiration, CWD, shortwave radiation, NDSI, wind speed, SWE, soil moisture, and BAI. Seasonally, evapotranspiration, shortwave radiation, and CWD all peaked during summer months, while soil moisture reached its minimum at that time (Figure 3-5). BAI and wind speed exhibited

early-season peaks around April–May, declined through summer, then rose again to maxima in October–November (Figure 3-5). Both NDSI and SWE were highest in January–March before tapering off until November (Figure 3-5). Substantial differences among all eight variables were evident at both the site and ecoregion levels (Table AII-1).

3.3.3 Model performance

Performance was broadly comparable across RFR, LGBM, and SVR, with EN lagging (Figure 3-6 and Figure AII-2). RFR achieved the best internal test-set performance (R^2 = 0.76 and RMSE = 0.31 g C m^{-2} day^{-1}) followed by LGBM, SVR, and EN (Table 3-5 and Figure 3-6). All models exhibited a slight bias: they tended to overestimate (*i.e.,* predict less negative NEE) when true EC-NEE values were strongly negative, and underestimate (*i.e.,* predict less positive NEE) when true values were strongly positive, with this bias increasing modestly from training to testing data. To quantify uncertainty, prediction intervals were computed at multiple confidence levels and evaluated coverage (the fraction of test observations whose true values fell within the interval) and mean interval width. At the 95% confidence level (α = 0.05), RFR achieved a coverage of 0.97, meaning 97% of true values lay within its intervals, with an average interval width of 1.56 g C m^{-2} day^{-1} (Figure AII-3).

Table 3-5: Performance metrics on internal testing (and training in parentheses) data for each machine learning model. Bold values indicate the optimal performance for each metric among all models. The top-performing model, considering all metrics, is RF (EN = ElasticNet Regression; LGBM = Light Gradient-Boosting Machine; RFR = Random Forest Regression; SVR = Support Vector Regression).

	EN	LGBM	RFR	SVR
MAE (g C m^{-2} day^{-1})	0.25	0.23	**0.21**	0.23
	(0.28)	(0.05)	(0.11)	(0.17)
MSE (g^2 C^2 m^{-2} day^{-2})	0.13	**0.10**	**0.10**	0.11
	(0.18)	(0.01)	(0.03)	(0.08)
RMSE (g C m^{-2} day^{-1})	0.36	0.32	**0.31**	0.34
	(0.42)	(0.08)	(0.17)	(0.28)
MAPE (%)	1.36	1.40	**1.30**	1.60
	(1.55)	(0.34)	(0.50)	(0.86)
R^2	0.68	0.75	**0.76**	0.71
	(0.63)	(0.98)	(0.93)	(0.83)

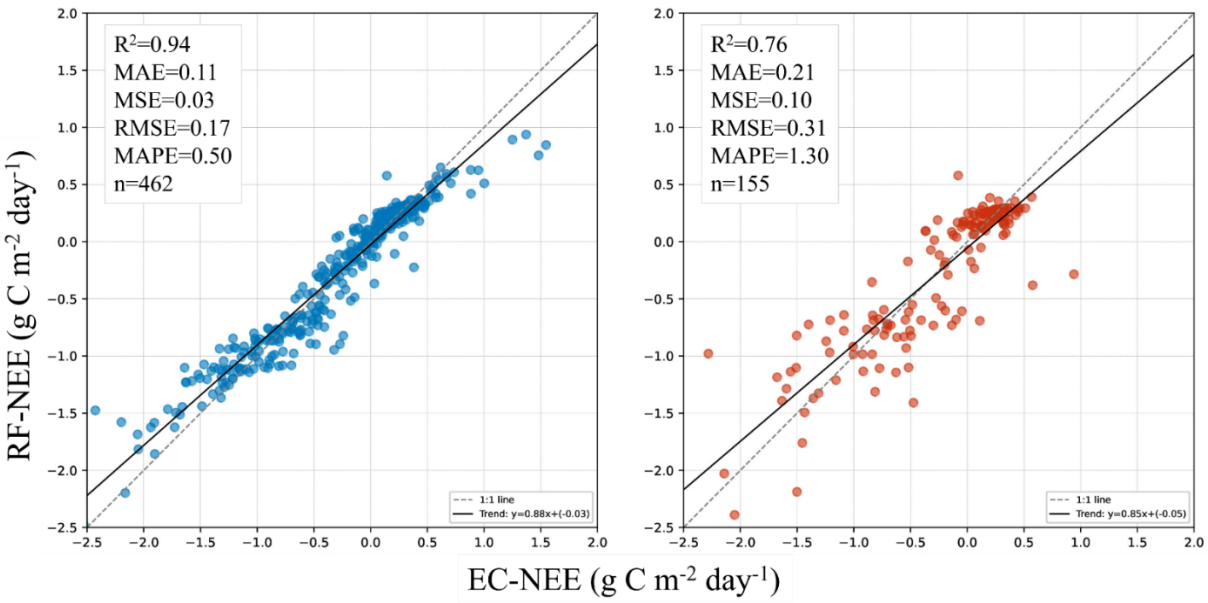


Figure 3-6: Random Forest (RF) predictions versus eddy covariance (EC) observations of daily CO_2 net ecosystem exchange (NEE) for (a) the training set (n = 462) and (b) the internal test set (n = 155). Colored points show site-level monthly averages, the dashed gray line is the 1:1 reference, and the solid black line is the fitted regression. Key performance metrics (R^2, MAE, MSE, RMSE, MAPE) are listed in the inset for each panel.

3.3.4 Model interpretation

SHAP analysis identified the top four drivers of NEE as evapotranspiration (SHAP value=0.29; normalized relative contribution=40%), shortwave radiation (0.14; 19%), BAI (0.08; 11%), and NDSI (0.07; 10%). The remaining four predictors (SWE, CWD, wind speed, and soil moisture) each contributed less than 7% collectively (cumulative SHAP value=0.14) (Figure AII-4). The SHAP analysis also showed the contributions of variables on the prediction (Figure 3-7). Generally, high values of evapotranspiration, shortwave radiation, CWD, and soil moisture had a positive contribution on the prediction, while low values had a negative contribution. Conversely, low values of BAI, NDSI, SWE, and wind speed had a negative contribution on the prediction, and high values had a positive contribution. In the SHAP heatmap (Figure AII-5), it is shown that all features have strongly non-linear effects on the prediction.

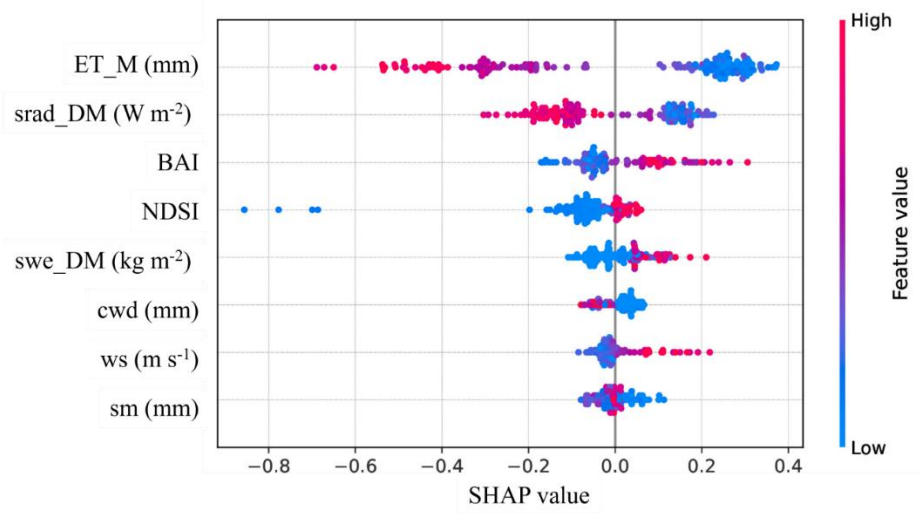


Figure 3-7: Beeswarm summary of SHAP-based local explanations for the selected eight features: evapotranspiration (ET_M), shortwave radiation (srad_DM), burn area index (BAI), normalized difference snow index (NDSI), snow water equivalent (swe_DM), climate water deficit (cwd), wind speed (ws), and soil moisture (sm), ranked by descending influence on NEE. Point colors (blue to red) show feature magnitude (low to high, respectively), the *x*-axis shows SHAP values (positive = higher predicted NEE), and each point is an individual monthly average record illustrating feature impacts on model predictions.

Partial dependence plots (Figure AII-6) showed that higher evapotranspiration, shortwave radiation, and CWD corresponded to more negative (greater CO_2 uptake) NEE, whereas elevated BAI, NDSI, and SWE drove predictions toward more positive NEE. Wind speed and soil moisture exhibited strongly non-linear effects: wind speed produced a threshold response with predicted NEE increasing sharply above ~3.5 m s^{-1}, and soil moisture first enhanced uptake before, past a certain point, reducing it. SHAP interaction networks (Figure AII-7) further revealed that feature interconnections were strongest in samples with the most negative NEE, suggesting these predictors collectively capture growing-season dynamics most effectively.

Our leave-one-ecoregion-out sensitivity analysis quantified generalizability by omitting each ecoregion in turn and retraining the RFR model. The lowest RMSE occurred when Taiga Plains (n = 27) was excluded (RMSE = 0.22 g C m^{-2} day^{-1}), followed by Mixed Wood Plains (n = 161; RMSE = 0.29 g C m^{-2} day^{-1}) and Hudson Plain (n = 106; RMSE = 0.37 g C m^{-2} day^{-1}). Excluding Marine West Coast Forest (n = 62; RMSE = 0.61 g C m^{-2} day^{-1}), Softwood Shield (n = 29; RMSE = 0.51 g C m^{-2} day^{-1}), or Boreal Plain (n = 77; RMSE = 0.67 g C m^{-2} day^{-1}) all increased error above the overall mean RMSE (0.46 g C m^{-2} day^{-1}) (Table AII-2).

3.3.5 Model application

An external validation was conducted on six independent peatland sites outside our training dataset: CA-UBP, CA-BP, and CA-SHF (all in the Boreal Plains ecoregion where transferability was weakest), CA-KLP (Hudson Plains), and CA-FI and CA-BDB (ecoregions not represented during training). Unlike the development phase, here LGBM appeared as the top performer (R^2 = 0.53; RMSE = 0.56 g C m^{-2} day^{-1}), while RFR ranked as a close second (R^2 = 0.51; RMSE = 0.57

g C m^{-2} day^{-1}; Figure 3-8 and Figure AII-8). Across these external sites, the characteristic bias, overestimating (less negative) fluxes under strong CO_2 uptake and underestimating (less positive) fluxes under net emission, was amplified relative to the internal test set. Performance peaked at CA-KLP site, followed by CA-SHF and CA-BDB. However, at the other three sites with limited observations (CA-FI, CA-UBP, CA-BP), the model failed to fully capture EC-NEE variability, likely reflecting their very small sample sizes (3–9 flux observations). These results affirm the model's strong transferability to both represented and novel ecoregions for external validation, while highlighting that sites with sparse data may require caution or targeted data collection. Furthermore, the model performance metrics on these external validation sites were heavily influenced by CA-SHF, the site with the highest sample size. Our model tended to overestimate the extreme summer uptake values observed at this site, further skewing overall performance, and caution should be used when applying the model to sites with extreme values.

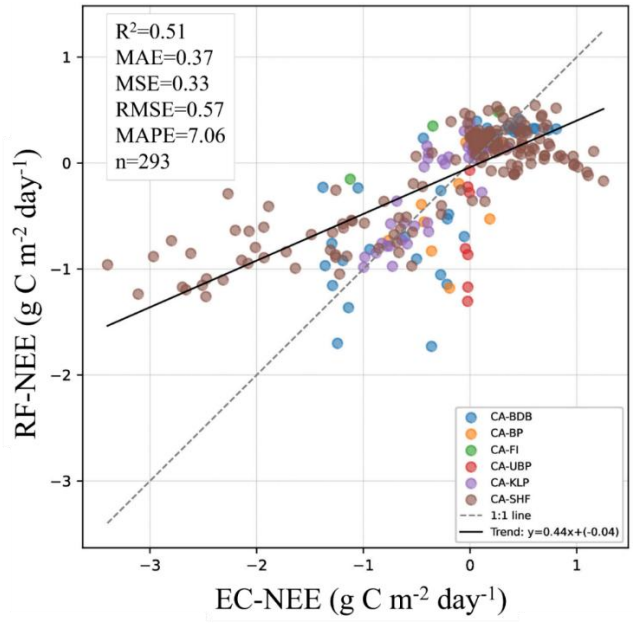


Figure 3-8: Random Forest (RF) predictions versus eddy covariance (EC) observations of daily CO_2 net ecosystem exchange (NEE) at six external validation sites (n = 293). Colored points indicate site-level data, the dashed gray line is the 1:1 reference, and the solid black line is the fitted regression. Key performance metrics (R^2, MAE, MSE, RMSE, MAPE) are listed in the inset.

3.4 Discussion

3.4.1 Model performance and applicability

The RFR model outperformed all other methods tested in predicting NEE of CO_2 at our peatland sites (see *3.3.3*). RFR has also demonstrated strong predictive skill in various previous studies for peatland applications: microplastic concentration (Tran *et al.*, 2023); water table depth (Toca *et al.*, 2023); GPP (Zhou *et al.*, 2023); fire-hotspot (Kurniawan *et al.*, 2025); EC-measured CH_4 flux (Li *et al.*, 2024); and water quality (Saravani *et al.*, 2024) prediction. These studies benchmarked RFR against alternatives such as Least-Squares Support Vector Machines, simple and multiple linear regressions, various neural networks, gradient boosting, and decision-tree-based models and consistently found RFR to be among the top performers. Its capacity to capture complex nonlinear relationships via averaging across decorrelated decision trees likely explains its superior performance for peatland NEE. In our external validation (see *3.3.5*), RFR also demonstrated strong predictive accuracy (ranking second), narrowly behind LGBM. Cai *et al.* (2020) evaluated the performance of GBM and RF for predicting NEE and highlighted that the best performance was obtained using an ensemble approach, suggesting that integrating these models into a combined ensemble could leverage the strengths of both methods, an area worth investigating in future work.

The leave-one-ecoregion-out sensitivity analysis revealed that omitting individual ecoregions from the training set led to only modest increases in prediction error, underscoring the model's resilience to regional gaps (see *3.3.4*). However, it is crucial to highlight that there were different sample sizes represented in each ecoregion, and it is possible that sample size effects and site-specific influences may be impacting the results of the sensitivity analysis. Moreover, when the external

validation was performed on six independent sites, including sites in ecoregions that were not among those represented in the training dataset, the model continued to deliver reliable NEE of CO_2 estimates (see *3.3.5*). Together, these results affirm the potential for this workflow to be applied to under- or un-represented peatland regions that share similar environmental characteristics with the original training sites.

The RF model used in this study exhibited a modest bias: it tended to underpredict the magnitude of strong CO_2 uptake (*i.e.,* predicting less negative NEE than observed) and to underpredict emissions when EC-NEE was strongly positive (Figure 3-6). A similar pattern of bias has been reported in other studies for the SMAP-NEE data product (Soil Moisture Active Passive Level 4 Global Daily EASE-Grid Carbon NEE) which showed attenuated extremes at peatland sites (Chapter 2: Hettinga *et al.*, in review; Kimball *et al.*, 2022). Crucially, however, the RF model outperforms SMAP-NEE: on the internal test dataset it achieves $R^2 = 0.76$ and RMSE = 0.31 g C m^{-2} day^{-1}, compared to $R^2 < 0.60$ and RMSE > 0.57 g C m^{-2} day^{-1} for the Corrected-SMAP-NEE product reported by Hettinga *et al.* (in review). Moreover, Peng *et al.* (2015) evaluated nine terrestrial ecosystem models and found that all tended to overestimate NEE (*i.e.,* predict less CO_2 uptake) during peak growing-season uptake and, to a lesser extent, overestimate emissions in winter. In this study, RF model displays a similar, but much reduced, seasonal bias: slight underestimation of uptake in summer and minimal overestimation of emissions in winter. These small discrepancies likely reflect remaining process omissions or underrepresented extreme conditions in the training data, but overall, the RF model delivers more accurate and tightly bounded NEE estimates than existing global products. These discrepancies for more extreme

values could be mitigated by the inclusion of additional sites with high CO_2 uptake or output in the model's training data.

3.4.2 Contributions of variables

The seasonal cyclicity of the final eight predictors mirrors that of peatland CO_2 exchange, and five of these: evapotranspiration, CWD, SWE, NDSI, and soil moisture, directly relate to water availability. Because the water cycle tightly couples to C cycling, water stress can strongly limit both photosynthesis and ecosystem respiration (Gentine *et al.*, 2019) and it is unsurprising that the SHAP analysis identified evapotranspiration and CWD as the two most important drivers in predicting NEE. The partial dependence analysis showed that higher evapotranspiration corresponds to more negative (stronger uptake) NEE (Figure AII-6), consistent with greater latent-heat losses under well-watered conditions enhancing plant productivity (reduced GPP). Likewise, increasing CWD (the difference between potential and actual evapotranspiration, which gives a better measure of water stress) indicates soil moisture deficit and generally corresponds to reduced CO_2 uptake (Helbig *et al.*, 2020) (Figure AII-6).

Two complementary snow metrics including NDSI (which detects snow cover) and SWE (which quantifies liquid water in the snowpack) were also incorporated to capture winter and shoulder-season dynamics that are often underrepresented in year-round models, leaving them susceptible to underestimating fluxes without sufficient climatic information during the non-growing season (NGS) (Natali *et al.*, 2019; Rafat *et al.*, 2021). Both variables help explain fluxes during periods of snow cover when other predictors may lack sensitivity. Soil moisture completed the suite of

water-related features, providing a direct measure of peat water content and further improving model fidelity to hydrologic controls on NEE.

Shortwave radiation was the second-most influential predictor, with higher radiation driving greater GPP and thus more negative NEE (Figure AII-6), a well-known relationship in C-flux modeling resulting from its inherent relationship with GPP (Cai *et al.*, 2020; Dou & Yang, 2018; He *et al.*, 2006; Melesse & Hanley, 2005; Rafat *et al.*, 2021; Tramontana *et al.*, 2016; Virkkala *et al.*, 2021; Xiao *et al.*, 2008). This aligns with Cai *et al.* (2020), who found latent-heat flux (evapotranspiration multiplied with latent heat of vaporization (constant)) and radiation among the top four predictors of NEE in forest ecosystems, with addition of soil temperature and photosynthetically active radiation (PAR) also identified as important features.

Finally, BAI ranked third overall in importance which reflects its sensitivity to surface-burn signatures captured in the red and near-infrared bands, analogous to NDVI's detection of vegetation browning resulting from its use of these bands in common with vegetation indices (Myers-Smith *et al.*, 2020; Tucker, 1979; Virkkala *et al.*, 2021). Additionally, since BAI was designed to evaluate the post-burn charcoal signatures, it's also possible that it may mislabel spaces with low reflectance, *i.e.,* water bodies or the presence of water as burned area (Chuvieco *et al.*, 2002). This finding aligns much more closely with the seasonal cycle we observed for BAI, which shares similarities with a peatland's annual hydrograph. Several training sites (*e.g.,* CA-DBB, CA-DB2), lying in fire-susceptible landscapes (Khosravi *et al.*, 2025), had experienced wildfires in the past; and an external validation site (CA-BP) experienced wildfire disturbance 2–3 years prior to

measurement. Although the model under-predicted NEE at CA-BP likely due to small sample size effects, BAI still emerged as a key driver with a distinctly positive partial dependence relationship in predicting NEE (Figure AII-6), highlighting the value of fire-history indicators in peatland C modeling. To our knowledge, BAI has not been included as a feature for predicting NEE in peatlands in previous studies.

3.4.3 Uncertainties and limitations

Data collected via the EC method are derived from high-frequency measurements of turbulence and CO_2 concentrations and require extensive post-processing rather than representing direct flux readings (Lees *et al.*, 2018). Although EC provides near-continuous half-hourly flux estimates, instrumentation constraints and quality-control procedures resulted in gaps amounting to 15–77% of the data at 15 selected sites (Table 3-2). To fill these gaps, empirical gap-filling was applied using *in-situ* measurements including shortwave radiation, air temperature, and vapor pressure deficit, following established protocols. After feature selection, only radiation remained among these as an important driver of NEE, while air temperature and vapor pressure deficit were excluded due to high collinearity with other variables. Tramontana *et al.* (2016) implemented identical quality-control and gap-filling workflows and reported lower bias when only a small fraction of data were gap-filled; Moffat *et al.* (2007) similarly reported negligible bias from filling short gaps in forest ecosystems. In this study, potential bias was further minimized by confining gap-filling to short intervals and by aggregating the resulting half-hourly fluxes first to daily, then to monthly, averages.

Furthermore, EC footprint size and position shift dynamically with turbulence and wind direction (Baldocchi *et al.*, 1997; Lees *et al.*, 2018). The predictor layers in this study, spanned spatial resolutions from 0.5 to 5 km, and the values were extracted within a 1 km radius around each tower location in GEE. Xiao *et al.* (2008) and Tramontana *et al.* (2016) instead averaged over a central 3 km window within 7 km MODIS scenes to capture larger footprint extents. However, expanding the buffer radius introduces a trade-off: while a larger buffer can better encompass footprint heterogeneity, it also increases geolocation uncertainty and mixed-pixel effects that reduce spatial precision. By using a 1 km buffer for finer-resolution datasets, the representativeness was balanced of the EC footprint against the risk of including non-peatland pixels.

Moreover, each data product carries its own uncertainties and constraints. Remotely sensed datasets in particular are prone to noise, biases, and data gaps, and their retrievals can be severely impacted by clouds and snow cover (Tramontana *et al.*, 2016). In high-latitude peatlands, persistent cloudiness poses a major obstacle to continuous observation; however, sensors with high revisit frequency, such as MODIS, provide enough repeat observations to effectively gap-fill cloud-obscured intervals (Lees *et al.*, 2018).

3.5 Conclusion

In this study, I demonstrated that machine learning models, trained on year-round EC CO_2 fluxes from 15 Canadian peatland sites and 34 remote sensing and gridded climate predictors, can accurately reproduce NEE of CO_2 across diverse ecoregions. Among the four algorithms tested (ElasticNet (EN), Light Gradient Boosted Machine (LGBM), Random Forest Regression (RFR), and Support Vector Regression (SVR)), RFR emerged as the top performer (R^2 = 0.76; RMSE =

0.31 g C m^{-2} day^{-1}) using just eight key features: evapotranspiration, shortwave radiation, burn area index (BAI), normalized difference snow index (NDSI), snow water equivalent (SWE), climate water deficit (CWD), wind speed, and soil moisture. Crucially, this eight-feature RF model maintained comparable skill when applied to six external validation sites, both within and beyond the ecoregions represented in the training data, highlighting its robustness and transferability. By pinpointing the most influential predictors, our approach not only reduces reliance on costly *in-situ* measurements but also provides a clear framework for scaling peatland CO_2-NEE estimates to the national level. These results offer a practical pathway to enhance Canada's national GHG inventories but can also deliver spatially explicit insights to inform policy and management decisions for peatland conservation and climate change mitigation across Canada.

4 Conclusion

4.1 Summary of key findings

The main objective of this thesis was to assess the capabilities of remotely sensed and gridded climate products to model the net ecosystem exchange (NEE) of carbon dioxide (CO_2) across Canadian peatlands for potential upscaling efforts. In Chapter 2, a previously existing remotely sensed and modeled data product was assessed, whereas in Chapter 3 a random forest regressor algorithm was developed to model NEE across a variety of peatland sites.

In Chapter 2, the performance of the globally produced remotely sensed and modeled Soil Moisture Active Passive Level 4 Global Daily EASE-Grid Carbon NEE (SMAP-NEE) data product was assessed at five Canadian peatland sites. These sites represented a variety of peatland types and were geographically widespread. Year-round eddy covariance data was used to validate the performance of this modeled data product. SMAP-NEE tended to model peatlands as stronger CO_2 sinks during the growing season (GS) and weaker CO_2 sources during the non-growing season (NGS). As a result, a Corrected-SMAP-NEE dataset was produced for these peatland sites to provide an estimate of seasonal and annual CO_2 budgets. While the Corrected-SMAP-NEE dataset better matched the EC-NEE fluxes, this correction conducted on a site-by-site basis was questionable for applicability to potential upscaling efforts and still maintained the strong bias of the original SMAP-NEE data product. However, some meaningful conclusions were found about the seasonality of the modeled data. Primarily, the NGS CO_2 emissions represented a highly variable proportion of GS uptake, resulting in annual CO_2 sink strengths that were reduced accordingly when the NGS emissions were included in the annual calculations. This finding underscores the importance of having NGS measurements and not drawing conclusions about the peatlands CO_2 sink strength or even status from solely GS measurements. Moreover, the agreement between SMAP-NEE and EC-NEE was better during the GS, highlighting that the model more

effectively captures GS dynamics, compared to NGS dynamics. This finding further highlights the necessity of collecting data during the NGS, to better understand the processes occurring during this season, and thereby better represent these processes in the satellite informed process-based model like SMAP-NEE. Moreover, the longer GS were consistent with higher annual CO_2 uptake at the five peatland sites evaluated in Chapter 2. This finding highlights that as the GS length increases, a higher annual CO_2 uptake might be observed, but it is important to consider the variables that may limit these increases under different climate change scenarios. While the SMAP-NEE data product overall was highly interpretable and user friendly it's notably mismatch with the EC-NEE data at these five northern peatland sites demonstrate its limitations for the application of this data product in national upscaling efforts.

In Chapter 3, pre-existing publicly available peatland CO_2 NEE data collected via the eddy covariance (EC) method was used to train, validate, and test four machine learning algorithms (ElasticNet Regression, Light Gradient-Boosting Machine (LGBM), Random Forest Regression (RFR), and Support Vector Regression (SVR) in their efficacy in modeling peatland NEE based on remotely sensed and gridded climate predictor variables. After a comprehensive feature importance and selection workflow, model performance increased up to the inclusion of eight features. These eight features included evapotranspiration, shortwave radiation, burn area index, normalized difference snow index, snow water equivalent, climate water deficit, wind speed, and soil moisture were required to produce accurate estimates of NEE after a comprehensive feature importance and selection workflow. The remote sensing derived indices burn area index and normalized difference snow index must be interpreted with caution as they may be more representative of the spectral bands they measure, rather than the indices they are meant to represent. After training and testing the four machine learning algorithms, the best performing

model was the RFR. Interestingly, the model exhibited a modest bias underpredicting the magnitude of strong CO_2 uptake and emissions, similar to SMAP-NEE. However, the RFR model outperformed SMAP-NEE in all performance metrics. This bias was amplified moving from training to testing to the external validation datasets. The training data was clustered around NEE~0, meaning that the model did not have many instances of extreme uptake or release to learn, potentially resulting in these biases. Similar to the SMAP-NEE data product, the RFR model captured GS dynamics with a higher accuracy than NGS dynamics, based on our supplemental SHAP analyses. The potential for the model to be applied in upscaling efforts was evaluated by conducting a leave-one-ecoregion-out sensitivity analysis and applying the RFR model to external validation sites. The results of these analyses demonstrated moderate generalizability with the model performance metrics having lower than average error in ecoregions with high generalizability (Taiga Plains, Hudson Plains, and Mixed Wood Plains), while having slightly larger than average error for ecoregions with low generalizability (Marine West Coast Forest, Softwood Shield, and Boreal Plain). The model achieved comparable performance for ecoregions not represented in the training data to those with low generalizability. These results affirm the model's strong transferability, while highlighting that areas with low generalizability may require caution or targeted data collection.

The results from both Chapters 2 and 3 highlight challenges and strengths in modeling NEE in peatlands. Proper representation of input features contributing to the highly dynamic daily NEE fluxes is crucial and despite the best efforts to ensure representation of features governing the dynamics in both the growing and non-growing seasons it is likely that not all effects were captured. Importantly, additional monitoring in the NGS is recommended to make data available for validation during that season as well. Additionally, while remotely sensed and gridded climate

products are simplistic and highly generalizable to broad geographic domains required for upscaling, each product comes with its own sets of biases and uncertainties and may not necessarily be representative of the conditions of the site. Moreover, indices developed for a specific purpose may be representative of the characteristic they aim to estimate, but caution in over-interpreting these spectral indices is advised as they may mask other important characteristics or the models used to derive remotely sensed data products may integrate several different sources of information. Importantly data driven models informed by machine learning rely heavily on statistical relationships, where correlation may not always equate to causation and fundamental processes may not be captured, and again the caution should be made for over-interpretation of the input features. Despite these limitations and uncertainties, efforts to upscale peatland processes to the national level are crucial in constraining GHG budgets and evaluating peatlands potential to contribute to nature-based solutions to climate change.

4.2 Recommendations for future research

In Chapter 2 of this thesis, I experimented with a peatland correction factor for SMAP-NEE data product but found that the results were substantially more accurate on a site-by-site basis. Future work, upon the availability of more data could experiment further with generating a peatland correction factor for the model. Additionally, peatlands are not explicitly represented within the plant functional types in the SMAP-NEE model; future work should aim to include peatland sites in model training and validation steps, as well as refine peatland C cycling processes.

In Chapter 3 of this thesis, four machine learning algorithms were trained and tested on their accuracy in predicting NEE across Canadian peatland sites. Future research should use this model framework to upscale estimates of NEE based on the eight selected features. This research should prioritize the spatial component, while also maintaining a functional scale of less than 5 km

(highest resolution of data). The outcome would be the first production of a spatially explicit national peatland NEE dataset at high resolution. This work would directly support Can-Peat objectives by reducing uncertainty in GHG emissions estimates, and future projects aimed at determining GHG emission changes in response to climate change or disturbance regimes. By incorporating the spatial component, it also allows for more localized communities and government agencies to use this information to guide their decisions. Moreover, given that peatlands C sink status is not only determined by the CO_2 flux (NEE), additional C fluxes (CH_4, POC, DOC) should also be considered. The measurement of these fluxes should be prioritized and future work could apply a similar modeling framework to attempt to integrate these fluxes into an overarching peatland C flux model.

References

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, 5, 170191. https://doi.org/10.1038/sdata.2017.191

AmeriFlux. (2025). *AmeriFlux*. https://ameriflux.lbl.gov/

Arndt, K., & Natali, S. (2025), AmeriFlux BASE CA-CF3 Churchill Fen Site 3, Ver. 2-5, AmeriFlux AMP, (Dataset). https://doi.org/10.17190/AMF/2335573

Baldocchi, D. (1997). Flux Footprints Within and Over Forest Canopies. *Boundary-Layer Meteorology*, *85*(2), 273–292. https://doi.org/10.1023/A:1000472717236

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., ... & Wofsy, S. (2001). FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. Bulletin of the American Meteorological society, 82(11), 2415-2434. https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2

Baldocchi, D. D. (2003). Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. *Global Change Biology*, *9*(4), 479–492. https://doi.org/10.1046/j.1365-2486.2003.00629.x

Baldocchi, D., Chu, H., & Reichstein, M. (2018). Inter-annual variability of net and Gross Ecosystem Carbon Fluxes: A Review. *Agricultural and Forest Meteorology*, 249, 520–533. https://doi.org/10.1016/j.agrformet.2017.05.015

Balogun, O., Bello, R., & Higuchi, K. (2023). Terrestrial CO2 exchange diagnosis using a peatland-optimized vegetation photosynthesis and respiration model (VPRM) for the Hudson Bay Lowlands. *The Science of the Total Environment*, *875*, 162591–162591. https://doi.org/10.1016/j.scitotenv.2023.162591

Balogun, O., Bello, R., & Higuchi, K. (2024). Long-Term Response of Peatland Carbon Exchange to Climatic Changes in the Hudson Bay Lowlands. *Journal of Geophysical Research. Biogeosciences*, *129*(5). https://doi.org/10.1029/2023JG007873

Barber, R. F., Candès, E. J., Ramdas, A., & Tibshirani, R. J. (2021). Predictive inference with the jackknife. *The Annals of Statistics*, *49*(1), 486–507. https://doi.org/10.1214/20-AOS1965

Barnard, D. M., Knowles, J. F., Barnard, H. R., Goulden, M. L., Hu, J., Litvak, M. E., & Molotch, N. P. (2018). Reevaluating growing season length controls on net ecosystem production in evergreen conifer forests. *Scientific Reports*, 8(1), 17973–10. https://doi.org/10.1038/s41598-018-36065-0

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Arain, M. A., Baldocchi, D., Bonan, G. B., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K. W., Roupsard, O., Veenendaal, E., Viovy, N.,

Williams, C., Woodward, F,A,, & Papale, D. (2010). Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate. *Science (American Association for the Advancement of Science)*, *329*(5993), 834–838. https://doi.org/10.1126/science.1184984

Blodau, C. (2002). Carbon cycling in peatlands — A review of processes and controls. *Environmental Reviews*, *10*(2), 111–134. https://doi.org/10.1139/a02-004

Bodesheim, P., Jung, M., Gans, F., Mahecha, M. D., & Reichstein, M. (2018). Upscaled diurnal cycles of land–atmosphere fluxes: a new global half-hourly data product. *Earth System Science Data*, *10*(3), 1327–1365. https://doi.org/10.5194/essd-10-1327-2018

Bouzgou, H., & Gueymard, C. A. (2017). Minimum redundancy – Maximum relevance with extreme learning machines for global solar radiation forecasting: Toward an optimized dimensionality reduction for solar time series. *Solar Energy*, *158*, 595–609. https://doi.org/10.1016/j.solener.2017.10.035

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Bridgham, S. D., Pastor, J., Dewey, B., Weltzin, J. F., & Updegraff, K. (2008). Rapid Carbon Response of Peatlands to Climate Change. *Ecology (Durham)*, *89*(11), 3041–3048. https://doi.org/10.1890/08-0279.1

Cai, J., Xu, K., Zhu, Y., Hu, F., & Li, L. (2020). Prediction and analysis of net ecosystem carbon exchange based on gradient boosting regression and random forest. *Applied Energy*, *262*, 114566-. https://doi.org/10.1016/j.apenergy.2020.114566

Chapin, F. S., Woodwell, G. M., Randerson, J. T., Rastetter, E. B., Lovett, G. M., Baldocchi, D. D., Clark, D. A., Harmon, M. E., Schimel, D. S., Valentini, R., Wirth, C., Aber, J. D., Cole, J. J., Goulden, M. L., Harden, J. W., Heimann, M., Howarth, R. W., Matson, P. A., McGuire, A. D., Melillo, J.M., Mooney, H.A., Neff, J.C., Houghton, R.A., Pace, M.L., Ryan, M.G., Running, S.W., Sala, O.E., Schlesinger, W.H., & Schulze, E.-D. (2006). Reconciling Carbon-Cycle Concepts, Terminology, and Methods. *Ecosystems (New York)*, 9(7), 1041–1050. https://doi.org/10.1007/s10021-005-0105-7

Chartrand, P.G., Sonnentag, O., Sanderson, N. K., & Garneau, M. (2023). Recent peat and carbon accumulation on changing permafrost landforms along the Mackenzie River valley, Northwest Territories, Canada. *Environmental Research Letters*, *18*(9), Article 095002. https://doi.org/10.1088/1748-9326/ace9ed

Christen, A., Jassal, R. S., Black, T. A., Grant, N. J., Hawthorne, I., Johnson, M. S., Lee, S.-C., & Merkens, M. (2016). Summertime greenhouse gas fluxes from an urban bog undergoing restoration through rewetting. *Mires and Peat*, 17(3), 1–24. https://doi.org/10.19189/MaP.2015.OMB.207

Christen, A., Knox, S. (2021), AmeriFlux BASE CA-DBB Delta Burns Bog, Ver. 2-5, AmeriFlux AMP, (Dataset). https://doi.org/10.17190/AMF/1543378

Chuvieco, E., Martín, M. P., & Palacios, A. (2002). Assessment of different spectral indices in the red-near-infrared spectral domain for burned land discrimination. *International Journal of Remote Sensing*, *23*(23), 5103–5110. https://doi.org/10.1080/01431160210153129

Coffer, M. M., & Hestir, E. L. (2019). Variability in trends and indicators of CO_2 exchange across arctic wetlands. *Journal of Geophysical Research: Biogeosciences*, 124(5), 1248–1264. https://doi.org/10.1029/2018jg004775

Corbett, E.J., Burdige, D.J., Tfaily, M.M., Dial, A.R., Cooper, W.T., Glaser, P.H., & Chanton, J.P. (2013). Surface production fuels deep heterotrophic respiration in northern peatlands. *Global Biogeochemical Cycles*, *27*(4), 1163–1174. https://doi.org/10.1002/2013GB004677

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1023/A:1022627411411

Coursolle, C., Margolis, H. A., Barr, A. G., Black, T. A., Amiro, B. D., McCaughey, J. H., Flanagan, L. B., Lafleur, P. M., Roulet, N. T., & Bourque, C. P. A. (2006). Late-summer carbon fluxes from Canadian forests and peatlands along an east-west continental transect. *Canadian Journal of Forest Research*, *36*(3), 783–800. https://doi.org/10.1139/x05-270

Cushman, S.A., Kaszta, Z.M., Burns, P., Hakkenberg, C.R., Jantz, P., Macdonald, D.W., Brodie, J.F., Deith, M.C.M. & Goetz, S. (2024). Simulating multi-scale optimization and variable selection in species distribution modeling Simulating multi-scale optimization and variable selection in species distribution modeling. *Ecological Informatics*, *83*, 102832-.

D′Acunha, B., Morillas, L., Black, T. A., Christen, A., & Johnson, M. S. (2019). Net Ecosystem Carbon Balance of a Peat Bog Undergoing Restoration: Integrating CO_2 and CH4 Fluxes From Eddy Covariance and Aquatic Evasion With DOC Drainage Fluxes. *Journal of Geophysical Research. Biogeosciences*, 124(4), 884–901. https://doi.org/10.1029/2019JG005123

Ding, C. & Peng, H. Minimum redundancy feature selection from microarray gene expression data in Computational Systems Bioinformatics: Proceedings of the 2003 IEEE Bioinformatics Conference. 523–528 (IEEE Comput. Soc, 2003). doi:10.1109/CSB.2003.1227396.

Dobrowski, S. Z., Littlefield, C. E., Lyons, D. S., Hollenberg, C., Carroll, C., Parks, S. A., Abatzoglou, J. T., Hegewisch, K., & Gage, J. (2021). Protected-area targets could be undermined by climate change-driven shifts in ecoregions and biomes. *Communications Earth & Environment*, *2*(1), 1–11. https://doi.org/10.1038/s43247-021-00270-z

Dorrepaal, E., Toet, S., Logtestijn, R. S. P. van, Swart, E., Weg, M. J. van de, Callaghan, T. V., & Aerts, R. (2009). Carbon respiration from subsurface peat accelerated by climate warming in the subarctic. *Nature (London)*, *460*(7255), 616–619. https://doi.org/10.1038/nature08216

Dou, X., & Yang, Y. (2018). Estimating forest carbon fluxes using four different data-driven techniques based on long-term eddy covariance measurements: Model comparison and evaluation. *The Science of the Total Environment*, *627*, 78–94. https://doi.org/10.1016/j.scitotenv.2018.01.202

Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A., Vapnik, V.: Support vector regression machines. In: M. Mozer, M. Jordan, T. Petsche (eds.) Advances in Neural Information Processing Systems 9, pp. 155–161. The MIT Press, Denver, USA (1996). https://proceedings.neurips.cc/paper_files/paper/1996/file/d38901788c533e8286cb6400b40b386d-Paper.pdf

Dutch, V. R., Rutter, N., Wake, L., Sonnentag, O., Hould Gosselin, G., Sandells, M., Derksen, C., Walker, B., Meyer, G., Essery, R., Kelly, R., Marsh, P., Boike, J., & Detto, M. (2024). Simulating net ecosystem exchange under seasonal snow cover at an Arctic tundra site. *Biogeosciences*, 21(3), 825–841. https://doi.org/10.5194/bg-21-825-2024

Endsley, K. A., Kimball, J. S., Kundig, T., Reichle, R. H., Ardizzone, and J.V. 2023. Validation Assessment for the Soil Moisture Active Passive (SMAP) Level 4 Carbon (L4_C) Data Product Version 7. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://nsidc.org/sites/default/files/documents/technical-reference/nasa_smap_l4c_assessment_v7.pdf.

Environmental Protection Agency. (2024). *Ecoregions of North America*. EPA. https://www.epa.gov/eco-research/ecoregions-north-america

Environment and Climate Change Canada. (2023, July 14). *1981-2010 Climate Normals & Averages*. Climate. https://climate.weather.gc.ca/climate_normals/index_e.html

Falge, E., Baldocchi, D., Olson, R., Anthoni, P., Aubinet, M., Bernhofer, C., Burba, G., Ceulemans, R., Clement, R., Dolman, H., Granier, A., Gross, P., Grünwald, T., Hollinger, D., Jensen, N.-O., Katul, G., Keronen, P., Kowalski, A., Lai, C. T., Law, B.E., Meyers, T., Moncrieff, J., Moors, E., Munger, J.W., Pilegaard, K., Rannik, U., Rebmann, C., Suyker, A., Tenhunen, J., Tu. K., Verma, S., Vesala, T., Wilson, K., & Wofsy, S. (2001). Gap filling strategies for defensible annual sums of net ecosystem exchange. *Agricultural and Forest Meteorology*, *107*(1), 43–69. https://doi.org/10.1016/S0168-1923(00)00225-2

Flanagan, L.B., & Syed, K.H. (2011). Stimulation of both photosynthesis and respiration in response to warmer and drier conditions in a boreal peatland ecosystem. *Global Change Biology*, *17*(7), 2271–2287. https://doi.org/10.1111/j.1365-2486.2010.02378.x

Gallego-Sala, A. V., Charman, D. J., Brewer, S., Page, S. E., Prentice, I. C., Friedlingstein, P., Moreton, S., Amesbury, M. J., Beilman, D. W., Björck, S., Blyakharchuk, T., Bochicchio, C., Booth, R. K., Bunbury, J., Camill, P., Carless, D., Chimner, R. A., Clifford, M., Cressey, E., Courtney-Mustaphi, C., De Vleeschouwer, F., de Jong, R., Fialkiewicz-Koziel, B., Finkelstein, S. A., Garneau, M., Githumbi, E., Hribjlan, J., Holmquist, J., Hughes, P. D. M., Jones, C., Jones, M. C., Karofeld, E., Klein, E. S., Kokfelt, U., Korhola, A., Lacourse, T., Le Roux, G., Lamentowicz, M., Large, D., Lavoie, M., Loisel, J., Mackay, H., MacDonald, G. M., Makila, M., Magnan, G., Marchant, R., Marcisz, K., Martínez Cortizas, A., Massa, C., Mathijssen, P., Mauquoy, D., Mighall, T., Mitchell, F. J. G., Moss, P., Nichols, J., Oksanen, P. O., Orme, L., Packalen, M. S., Robinson, S., Roland, T. P., Sanderson, N. K., Sannel, A. B. K., Silva-Sánchez, N., Steinberg, N., Swindles, G. T., Turner, T. E., Uglow, J., Väliranta, M., van Bellen, S., van der Linden, M., van Geel, B., Wang, G., Yu, Z., Zaragoza-Castells, J., & Zhao, Y. (2018). Latitudinal limits to the

predicted increase of the peatland carbon sink with warming. *Nature Climate Change*, 8(10), 907–913. https://doi.org/10.1038/s41558-018-0271-1

Gentine, P., Green, J. K., Guérin, M., Humphrey, V., Seneviratne, S. I., Zhang, Y., & Zhou, S. (2019). Coupling between the terrestrial carbon and water cycles-a review. *Environmental Research Letters*, *14*(8), Article 083003. https://doi.org/10.1088/1748-9326/ab22d6

Glenn, A. J., Flanagan, L. B., Syed, K. H., & Carlson, P. J. (2006). Comparison of net ecosystem CO_2 exchange in two peatlands in western Canada with contrasting dominant vegetation, Sphagnum and Carex. *Agricultural and Forest Meteorology*, *140*(1–4), 115–135. https://doi.org/10.1016/j.agrformet.2006.03.020

Gorham, E. (1991). Northern Peatlands: Role in the Carbon Cycle and Probable Responses to Climatic Warming. *Ecological Applications*, 1(2), 182–195. https://doi.org/10.2307/1941811

Gouttevin, I., Menegoz, M., Dominé, F., Krinner, G., Koven, C., Ciais, P., Tarnocai, C., & Boike, J. (2012). How the insulating properties of snow affect soil carbon distribution in the continental pan-Arctic area. *Journal of Geophysical Research: Biogeosciences*, *117*(G2). https://doi.org/10.1029/2011JG001916

Government of Canada (2024). 2020 Land Cover of Canada. Open Government Portal. https://open.canada.ca/data/en/dataset/ee1580ab-a23d-4f86-a09b-79763677eb47

Grolemund, G., Wickham, H. (2011). Dates and Times Made Easy with lubridate. Journal of Statistical Software, 40(3), 1-25. URL https://www.jstatsoft.org/v40/i03/.

Hanis, K. L., Tenuta, M., Amiro, B. D., & Papakyriakou, T. N. (2013). Seasonal dynamics of methane emissions from a subarctic fen in the Hudson Bay Lowlands. *Biogeosciences*, *10*(7), 4465–4479. https://doi.org/10.5194/bg-10-4465-2013

Hanis, K. L., Amiro, B. D., Tenuta, M., Papakyriakou, T., & Swystun, K. A. (2015). Carbon exchange over four growing seasons for a subarctic sedge fen in northern Manitoba, Canada. *Arctic Science*, *1*(2), 27–44. https://doi.org/10.1139/as-2015-0003

Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... & Oliphant, T. E. (2020). Array programming with NumPy. Nature, 585(7825), 357-362. https://doi.org/10.1038/s41586-020-2649-2

Harrison, J. A., Prairie, Y. T., Mercier-Blais, S., & Soued, C. (2021). Year-2020 Global Distribution and Pathways of Reservoir Methane and Carbon Dioxide Emissions According to the Greenhouse Gas From Reservoirs (G-res) Model. *Global Biogeochemical Cycles*, *35*(6), Article e2020GB006888. https://doi.org/10.1029/2020GB006888

Haslwanter, A., Hammerle, A., & Wohlfahrt, G. (2009). Open-path vs. closed-path eddy covariance measurements of the net ecosystem carbon dioxide and water vapour exchange: A long-term perspective. *Agricultural and Forest Meteorology*, *149*(2), 291–302. https://doi.org/10.1016/j.agrformet.2008.08.011

He, H., Yu, G., Zhang, L., Sun, X., & Su, W. (2006). Simulating CO2 flux of three different ecosystems in ChinaFLUX based on artificial neural networks. *Science in China. Series D, Earth Sciences*, *49*(2), 252–261. https://doi.org/10.1007/s11430-006-8252-z

He, H., Strachan, I. B., & Roulet, N. T. (2025). Simulating ecosystem carbon dioxide fluxes and their associated influencing factors for a restored peatland. *Biogeosciences*, *22*(5), 1355–1368. https://doi.org/10.5194/bg-22-1355-2025

Helbig, M., Chasmer, L. E., Desai, A. R., Kljun, N., Quinton, W. L., & Sonnentag, O. (2017). Direct and indirect climate change effects on carbon dioxide fluxes in a thawing boreal forest–wetland landscape. *Global Change Biology*, 23(8), 3231–3248. https://doi.org/10.1111/gcb.13638

Helbig, M., Humphreys, E. R., & Todd, A. (2019). Contrasting Temperature Sensitivity of CO2 Exchange in Peatlands of the Hudson Bay Lowlands, Canada. *Journal of Geophysical Research. Biogeosciences*, *124*(7), 2126–2143. https://doi.org/10.1029/2019JG005090

Helbig, M., Waddington, J. M., Alekseychik, P., Amiro, B. D., Aurela, M., Barr, A. G., Black, T. A., Blanken, P. D., Carey, S. K., Chen, J., Chi, J., Desai, A. R., Dunn, A., Euskirchen, E. S., Flanagan, L. B., Forbrich, I., Friborg, T., Grelle, A., Harder, S., Heliasz, M., Humphreys, E.R., Ikawa, H., Isabelle, P.-E., Iwata, H., Jassal, R., Korkiakoski, M., Kurbatova, J., Kutzbach, L., Lindroth, A., Lofenius, M.O., Lohila, A., Mammarella, I., Marsh, P., Maximov, T., Melton, J.R., Moore, P.A., Nadeau, D.F., Nicholls, E.M., Nilsson, M.B., Ohta, T., Peichl, M., Petrone, R.M., Petrov, R., Prokushkin, A., Quinton, W.L., Reed, D.E., Roulet, N.T., Runkle, B.R.K., Sonnentag, O., Strachan, I.B., Taillardat, P., Tuittila, E.-S., Tuovinen, J.-P., Turner, J., Ueyama, M., Varlagin, A., Wilmking, M., Wofsy, S.C., & Zyrianov, V. (2020). Increasing contribution of peatlands to boreal evapotranspiration in a warming climate. *Nature Climate Change*, *10*(6), 555–560. https://doi.org/10.1038/s41558-020-0763-7

Helgason, W., Johnson, B., David, C., Barr, A., Black, A. (2024). Long-term meteorological and carbon, water and energy flux data from the Boreal Ecosystem Research and Monitoring Sites, Saskatchewan, Canada. Federated Research Data Repository. https://doi.org/10.20383/103.01063

Hettinga, E.K., Rezanezhad, F., Persaud, B.D., Slowinski, S., Strack, M., Liu, H., Lennartz, B., Humphreys, E., Wu, J., Skeeter, J., Knox, S., Alcock, H., Sonnentag, O., Detto, M., and Van Cappellen, P. (in review). Non-growing season contributions to the annual net carbon dioxide budget in Canadian peatlands. *Journal of Geophysical Research: Biogeosciences.* Manuscript #2025JG009173

Hernandez Rodriguez, L. C., Goodwell, A. E., & Kumar, P. (2023). Inside the flux footprint: The role of organized land cover heterogeneity on the dynamics of observed land-atmosphere exchange fluxes. *Frontiers in Water*, 5. https://doi.org/10.3389/frwa.2023.1033973

Hobbie, John E., and George W. Kling, eds. *Alaska's changing Arctic: Ecological consequences for tundra, streams, and lakes*. Oxford University Press, 2014.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, *12*(1), 55–67. https://doi.org/10.2307/1267351

Hu, G., Zhao, L., Wu, T., Wu, X., Park, H., Fedorov, A., Wei, Y., Li, R., Zhu, X., Sun, Z., Ni, J., & Zou, D. (2021). Spatiotemporal variations and regional differences in air temperature in the permafrost regions in the Northern Hemisphere during 1980–2018. *The Science of the Total Environment*, *791*, 148358–148358. https://doi.org/10.1016/j.scitotenv.2021.148358

Humphreys, E. R., Charron, C., Brown, M., & Jones, R. (2014). Two bogs in the Canadian Hudson Bay Lowlands and a temperate bog reveal similar annual Net Ecosystem Exchange of CO_2. *Arctic, Antarctic, and Alpine Research*, 46(1), 103–113. https://doi.org/10.1657/1938-4246.46.1.103

Jensen, A. M., Warren, J. M., King, A. W., Ricciuto, D. M., Hanson, P. J., & Wullschleger, S. D. (2019). Simulated projections of boreal forest peatland ecosystem productivity are sensitive to observed seasonality in leaf physiology. *Tree Physiology*, 39(4), 556–572. https://doi.org/10.1093/treephys/tpy140

Jeong, J.-H., Kug, J.-S., Kim, B.-M., Min, S.-K., Linderholm, H. W., Ho, C.-H., Rayner, D., Chen, D., & Jun, S.-Y. (2012). Greening in the circumpolar high-latitude may amplify warming in the growing season. *Climate Dynamics*, *38*(7–8), 1421–1431. https://doi.org/10.1007/s00382-011-1142-x

Jones, L. A., Kimball, J. S., Reichle, R. H., Madani, N., Glassy, J., Ardizzone, J. V., Colliander, A., Cleverly, J., Desai, A. R., Eamus, D., Euskirchen, E. S., Hutley, L., Macfarlane, C., & Scott, R. L. (2017). The SMAP Level 4 Carbon Product for Monitoring Ecosystem Land-Atmosphere CO2 Exchange. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(11), 6517–6532. https://doi.org/10.1109/TGRS.2017.2729343

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B. E., Lindroth, A., Merbold, L., Montagnani, L., Moors, E. J., Papale, D., Sottocornola, M., Vaccari, F., and Williams, C.: Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, J. Geophys. Res., 116, G00J07, https://doi.org/10.1029/2010JG001566, 2011.

Karlsson, J., Serikova, S., Vorobyev, S. N., Rocher-Ros, G., Denfeld, B., & Pokrovsky, O. S. (2021). Carbon emission from Western Siberian inland waters. *Nature Communications*, *12*(1), 825–825. https://doi.org/10.1038/s41467-021-21054-1

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, *30*, 3146–3154.

Khosravi, K., Mosallanejad, A., Bateni, S. M., Kim, D., Jun, C., Shahvaran, A. R., ... & Ali, M. (2025). Assessing Pan-Canada wildfire susceptibility by integrating satellite data with novel hybrid deep learning and black widow optimizer algorithms. Science of the Total Environment, 977, 179369. https://doi.org/10.1016/j.scitotenv.2025.179369

Kimball, J. S., L. A. Jones, A. Endsley, T. Kundig, and R. Reichle. 2022. SMAP L4 Global Daily 9 km EASE-Grid Carbon Net Ecosystem Exchange, Version 7. Boulder, Colorado USA. NASA

National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/3K9F0S1Q5J2U

Kuhn, M. (2008). Building Predictive Models in *R* Using the caret package. Journal of Statistical Software, 28(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Kuhn, M. A., Varner, R. K., McCalley, C. K., Perryman, C. R., Aurela, M., Burke, S. A., Chanton, J. P., Crill, P. M., DelGreco, J., Deng, J., Heffernan, L., Herrick, C., Hodgkins, S. B., Jones, C. P., Juutinen, S., Kane, E. S., Lamit, L. J., Larmola, T., Lilleskov, E., Olefeldt, D., Palace, M.W., Rich, V.I., Schulze, C., Shorter, J.H., Sullivan, F.B., Sonnentag, O., Turetsky, M.R., & Waldrop, M. P. (2024). Controls on Stable Methane Isotope Values in Northern Peatlands and Potential Shifts in Values Under Permafrost Thaw Scenarios. *Journal of Geophysical Research. Biogeosciences*, *129*(7), Article e2023JG007837. https://doi.org/10.1029/2023JG007837

Kujawska, J., Kulisz, M., Cel, W., Kwiatkowski, C. A., Harasim, E., & Bandura, L. (2025). Evaluating and predicting CO2 flux from agricultural soils treated with organic amendments: a comparative study of ANN and ElasticNet models. *Journal of Soils and Sediments*, *25*(3), 864–882. https://doi.org/10.1007/s11368-025-03971-6

Kurniawan, R., Muttaqien, F. H., Syafarina, I., Belgaman, H. A., Dewi, S., & Latifah, A. L. (2025). Data-driven model for predicting peatland fire hotspots and carbon emissions in South Kalimantan. *Wetlands Ecology and Management*, *33*(3), Article 44. https://doi.org/10.1007/s11273-025-10058-z

Lafleur, P. M., Roulet, N. T., Admiral, S. W. (2001) Annual Cycle Of CO2 Exchange At A Bog Peatland, Journal Of Geophysical Research: Atmospheres, 106(D3), 3071-3081. https://doi.org/10.1029/2000jd900588

Lafleur, P. M., Roulet, N. T., Bubier, J. L., Frolking, S., & Moore, T. R. (2003). Interannual variability in the peatland-atmosphere carbon dioxide exchange at an ombrotrophic bog. *Global Biogeochemical Cycles*, 17(2). https://doi.org/10.1029/2002gb001983

Lee, S. C., Christen, A., Black, A. T., Johnson, M. S., Jassal, R. S., Ketler, R., Nesic, Z., & Merkens, M. (2017). Annual greenhouse gas budget for a bog ecosystem undergoing restoration by rewetting. *Biogeosciences*, 14(11), 2799–2814. https://doi.org/10.5194/bg-14-2799-2017

Lees, K. J., Quaife, T., Artz, R. R. E., Khomik, M., & Clark, J. M. (2018). Potential for using remote sensing to estimate carbon fluxes across northern peatlands – a review. *Science of The Total Environment*, 615, 857–874. https://doi.org/10.1016/j.scitotenv.2017.09.103

Li, Q., Gogo, S., Leroy, F., Guimbaud, C., & Laggoun-Défarge, F. (2021a). Response of Peatland CO2 and CH4 Fluxes to Experimental Warming and the Carbon Balance. *Frontiers in Earth Science (Lausanne)*, *9*, Article 631368. https://doi.org/10.3389/feart.2021.631368

Li, Q., Leroy, F., Zocatelli, R., Gogo, S., Jacotot, A., Guimbaud, C., & Laggoun-Défarge, F. (2021b). Abiotic and biotic drivers of microbial respiration in peat and its sensitivity to temperature change. *Soil Biology & Biochemistry*, *153*, Article 108077. https://doi.org/10.1016/j.soilbio.2020.108077

Li, X., Ge, J., Liu, Z., Yang, S., Wang, L., & Liu, Y. (2024). Estimating the methane flux of the Dajiuhu subalpine peatland using machine learning algorithms and the maximal information coefficient technique. *The Science of the Total Environment*, *916*, Article 170241. https://doi.org/10.1016/j.scitotenv.2024.170241

Limpens, J., Berendse, F., Blodau, C., Canadell, J. G., Freeman, C., Holden, J., Roulet, N., Rydin, H., & Schaepman-Strub, G. (2008). Peatlands and the carbon cycle: From local processes to global implications – A synthesis. *Biogeosciences*, *5*(5), 1475–1491. https://doi.org/10.5194/bg-5-1475-2008

Liu, L., Chen, H., Liu, X., Yang, Z., Zhu, D., He, Y., & Liu, J. (2019). Contemporary, modern and ancient carbon fluxes in the zoige peatlands on the Qinghai-Tibetan Plateau. *Geoderma*, 352, 138–149. https://doi.org/10.1016/j.geoderma.2019.06.008

Liu, Q., Piao, S., Janssens, I. A., Fu, Y., Peng, S., Lian, X., Ciais, P., Myneni, R. B., Peñuelas, J., & Wang, T. (2018). Extension of the growing season increases vegetation exposure to frost. *Nature Communications*, *9*(1), 426–428. https://doi.org/10.1038/s41467-017-02690-y

Lund, M., Lafleur, P. M., Roulet, N. T., Lindroth, A., Christensen, T. R., Aurela, M., Chojnicki, B. H., Flanagan, L. B., Humphreys, E. R., Laurila, T., Oechel, W. C., Olejnik, J., Rinne, J., Schubert, P., & Nilsson, M. B. (2010). Variability in exchange of CO_2 across 12 northern peatland and Tundra Sites. *Global Change Biology*. https://doi.org/10.1111/j.1365-2486.2009.02104.x

Lundberg, S., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions*. https://doi.org/10.48550/arxiv.1705.07874

Lunt, P. H., Fyfe, R. M., & Tappin, A. D. (2019). Role of recent climate change on carbon sequestration in Peatland Systems. *Science of The Total Environment*, 667, 348–358. https://doi.org/10.1016/j.scitotenv.2019.02.239

Madelon, R., Kimball, J. S., Endsley, K. A., De Lannoy, G. J. M., Sonnentag, O., Alcock, H., Detto, M., Virkkala, A. M., Rogers, B. M., Watts, J. D., Mavrovic, A., Williamson, S. N., Humphreys, E., Colliander, A., Mialon, A., & Roy, A. (2025). Assessing the SMAP Level-4 Carbon Product Over the Arctic and Subarctic Zones. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *18*, 9850–9864. https://doi.org/10.1109/JSTARS.2025.3555850

McNicol, G., Fluet-Chouinard, E., Ouyang, Z., Knox, S., Zhang, Z., Aalto, T., Bansal, S., Chang, K. Y., Chen, M., Delwiche, K., Feron, S., Goeckede, M., Liu, J., Malhotra, A., Melton, J. R., Riley, W., Vargas, R., Yuan, K., Ying, Q., Zhu, Q., Alekseychik, P., Aurela, M., Billesbach, D.P., Campbell, D.I., Chen, J., Chu, H., Desai, A.R., Euskirchen, E., Goodrich, J., Griffis, T., Helbig, M., Hirano, T., Iwata, H., Jurasinski, G., King, J., Koebsch, F., Kolka, R., Krauss, K., Lohila, A., Mammarella, I., Nilson, M., Noormets, A., Oechel, W., Peichl, M., Sachs, T., Sakabe, A., Schulze, C., Sonnentag, O., Sullivan, R.C., Tuittila, E.-S., Ueyama, M., Vesala, T., Ward, E., Wille, C., Wong, G.X., Zona, D., Windham-Myers, L., Poulter, B., & Jackson, R. B. (2023). Upscaling Wetland Methane Emissions From the FLUXNET-CH4 Eddy Covariance Network (UpCH4 v1.0):

Model Development, Network Assessment, and Budget Comparison. *AGU Advances*, *4*(5). https://doi.org/10.1029/2023AV000956

Melesse, A. M., & Hanley, R. S. (2005). Artificial neural network application for multi-ecosystem carbon flux simulation. *Ecological Modelling*, *189*(3), 305–314. https://doi.org/10.1016/j.ecolmodel.2005.03.014

Moffat, A. M., Papale, D., Reichstein, M., Hollinger, D. Y., Richardson, A. D., Barr, A. G., Beckstein, C., Braswell, B. H., Churkina, G., Desai, A. R., Falge, E., Gove, J. H., Heimann, M., Hui, D., Jarvis, A. J., Kattge, J., Noormets, A., & Stauch, V. J. (2007). Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. *Agricultural and Forest Meteorology*, *147*(3), 209–232. https://doi.org/10.1016/j.agrformet.2007.08.011

Morison, M. Q., Petrone, R. M., Wilkinson, S. L., Green, A., & Waddington, J. M. (2020). Ecosystem scale evapotranspiration and CO_2 exchange in burned and unburned peatlands: Implications for the ecohydrological resilience of carbon stocks to wildfire. *Ecohydrology*, *13*(2). https://doi.org/10.1002/eco.2189

Myers-Smith, I. H., Kerby, J. T., Phoenix, G. K., Bjerke, J. W., Epstein, H. E., Assmann, J. J., John, C., Andreu-Hayles, L., Angers-Blondin, S., Beck, P. S. A., Berner, L. T., Bhatt, U. S., Bjorkman, A. D., Blok, D., Bryn, A., Christiansen, C. T., Cornelissen, J. H. C., Cunliffe, A. M., Elmendorf, S. C., Forbes, B.C., Goetz, S.J., Hollister, R.D., de Jong, R., Loranty, M.M., Macias-Fauria, M., Maseyk, K., Normand, S., Olofsson, J., Parker, T.C., Parmentier, F.-J.W., Post, E., Shaepman-Strub, G., Stordal, F., Sullivan, P.F., Thomas, H.J.D., Tommervik, H., Treharne, R., Tweedie, C.E., Walker, D.A., Wilmking, M., & Wipf, S. (2020). Complexity revealed in the greening of the Arctic. *Nature Climate Change*, 10, 106–117. https://doi.org/10.1038/s41558-019-0688-1

Natali, S. M., Watts, J. D., Rogers, B. M., Potter, S., Ludwig, S. M., Selbmann, A.-K., Sullivan, P. F., Abbott, B. W., Arndt, K. A., Birch, L., Bjorkman, M. P., Bloom, A. A., Celis, G., Christensen, T. R., Christiansen, C. T., Commane, R., Cooper, E. J., Crill, P., Czimczik, C., Davydov, S., Du, J., Egan, J. E., Elberling, B., Euskirchen, E. S., Friborg, T., Genet, H., Gockede, M., Goodrich, J. P., Grogan, P., Helbig, M., Jafarov, E. E., Jastrow, J. D., Kalhori, A. A. M., Kim, Y., Kimball, J. S., Kutzbach, L., Lara, M. J., Larsen, K. S., Lee, B-Y., Liu, Z., Loranty, M. M., Lund, M., Lupascu, M., Madani, N., Malhotra, A., Matamala, R., McFarland, J., McGuire, A. D., Michelsen, A., Minions, C., Oechel, W. C., Olefeldt, D., Parmentier, F-J. W., Pirk, N., Poulter, B., Quinton, W., Rezanezhad, F., Risk, D., Sachs, T., Schaefer, K., Schmidt, N. M., Schuur, E. A. G., Semenchuk, P. R., Shaver, G., Sonnentag, O., Starr, G., Treat, C. C., Waldrop, M. P., Wang, Y., Welker, J., Wille, C., Xu, X., Zhang, Z., Zhuang, Q., & Zona, D. (2019). Large loss of CO2 in winter observed across the northern permafrost region (vol 9, pg 852, 2019). *Nature Climate Change*, *9*(12), 1005–1005. https://doi.org/10.1038/s41558-019-0644-0

NWWG - National Wetlands Working Group, 1988. Wetlands of Canada. Ecological Land Classification Series No. 24. Sustainable Development Branch, Environment Canada, Ottawa, Ontario, Canada. Polyscience Publications, Inc., Montreal, Quebec.

Nugent, K. A., Strachan, I. B., Strack, M., Roulet, N. T., and Rochefort, L.: Multi-year net ecosystem carbon balance of a restored peatland reveals a return to carbon sink, Glob Chang Biol, 24, 5751-5768, 10.1111/gcb.14449, 2018.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'Amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., Kassem, K. R. 2001. Terrestrial ecoregions of the world: a new map of life on Earth. Bioscience 51(11):933-938.

ORNL DAAC. (n.d.). *Fixed sites subsets tool*. MODIS/VIIRS Subsets. https://modis.ornl.gov/sites

ORNL DAAC. 2018a. Fixed Sites Subsetting and Visualization Tool. ORNL DAAC, Oak Ridge, Tennessee, USA. Accessed October 11, 2024. Subset obtained for SPL4CMDL product at site id 'ca_britishcolumbia_delta_burns_bog'. https://doi.org/10.3334/ORNLDAAC/1567

ORNL DAAC. 2018b. Fixed Sites Subsetting and Visualization Tool. ORNL DAAC, Oak Ridge, Tennessee, USA. Accessed September 08, 2023. Subset obtained for SPL4CMDL product at site id 'ca_ontario_kinoje_lake_peatland'. https://doi.org/10.3334/ORNLDAAC/1567

ORNL DAAC. 2018c. Fixed Sites Subsetting and Visualization Tool. ORNL DAAC, Oak Ridge, Tennessee, USA. Accessed September 08, 2023. Subset obtained for SPL4CMDL product at site id 'ca_nl_robinsons_peatland_pasture'. https://doi.org/10.3334/ORNLDAAC/1567

ORNL DAAC. 2018d. Fixed Sites Subsetting and Visualization Tool. ORNL DAAC, Oak Ridge, Tennessee, USA. Accessed October 11, 2024. Subset obtained for SPL4CMDL product at site id 'ca_nl_robinsons_natural_bog'. https://doi.org/10.3334/ORNLDAAC/1567

ORNL DAAC. 2018e. Fixed Sites Subsetting and Visualization Tool. ORNL DAAC, Oak Ridge, Tennessee, USA. Accessed October 11, 2024. Subset obtained for SPL4CMDL product at site id 'ca_northwestterritories_scotty_creek_bog'. https://doi.org/10.3334/ORNLDAAC/1567

Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., Longdoz, B., Rambal, S., Valentini, R., Vesala, T., & Yakir, D. (2006). Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation. *Biogeosciences*, *3*(4), 571–583. https://doi.org/10.5194/bg-3-571-2006

Papale, D., & Valentini, R. (2003). A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization. *Global Change Biology*, *9*(4), 525–535. https://doi.org/10.1046/j.1365-2486.2003.00609.x

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830. https://doi.org/10.5555/1953048.2078195

Peichl, M., Gažovič, M., Vermeij, I., de Goede, E., Sonnentag, O., Limpens, J., & Nilsson, M. B. (2018). Peatland vegetation composition and phenology drive the seasonal trajectory of maximum

gross primary production. *Scientific Reports*, *8*(1), Article 8012. https://doi.org/10.1038/s41598-018-26147-4

Peltola, O., Vesala, T., Gao, Y., Räty, O., Alekseychik, P., Aurela, M., Chojnicki, B., Desai, A. R., Dolman, A. J., Euskirchen, E. S., Friborg, T., Göckede, M., Helbig, M., Humphreys, E., Jackson, R. B., Jocher, G., Joos, F., Klatt, J., Knox, S. H., Kowalska, N., Kutzbach, L., Lienert, S., Lohila, A., Mammarella, I., Nadeau, D.F., Nilsson, M.B., Oechel, W.C., Peichl, M., Pypker, T., Quinton, W., Rinne, J., Sachs, T., Samson, M., Schmid, H.P., Sonnentag, O., Wille, C., Zona, D., & Aalto, T. (2019). Monthly gridded data product of northern wetland methane emissions based on upscaling eddy covariance observations. *Earth System Science Data*, *11*(3), 1263–1289. https://doi.org/10.5194/essd-11-1263-2019

Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(8), 1226–1238. https://doi.org/10.1109/TPAMI.2005.159

Peng, S., Ciais, P., Chevallier, F., Peylin, P., Cadule, P., Sitch, S., Piao, S., Ahlström, A., Huntingford, C., Levy, P., Li, X., Liu, Y., Lomas, M., Poulter, B., Viovy, N., Wang, T., Wang, X., Zaehle, S., Zeng, N., Zhao, F., & Zhao, H. (2015). Benchmarking the seasonal cycle of CO2 fluxes simulated by terrestrial ecosystem models. *Global Biogeochemical Cycles*, *29*(1), 46–64. https://doi.org/10.1002/2014GB004931

Primeau, G., & Garneau, M. (2021). Carbon accumulation in peatlands along a boreal to subarctic transect in eastern Canada. *Holocene (Sevenoaks)*, *31*(5), Article 0959683620988031. https://doi.org/10.1177/0959683620988031

Post, W. M., Emanuel, W. R., Zinke, P. J., & Stangenberger, A. G. (1982). Soil carbon pools and world life zones. *Nature*, 298(5870), 156–159. https://doi.org/10.1038/298156a0

Qiu, C., Zhu, D., Ciais, P., Guenet, B., Peng, S., Xu, X., & Xu, X. (2020). The role of northern peatlands in the global carbon cycle for the 21st century. *Global Ecology and Biogeography*, *29*(5), 956–973. https://doi.org/10.1111/geb.13081

R Core Team (2023). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

Rafat, A., Rezanezhad, F., Quinton, W. L., Humphreys, E. R., Webster, K., & Van Cappellen, P. (2021). Non-growing season carbon emissions in a northern peatland are projected to increase under global warming. *Communications Earth & Environment*, 2(1). https://doi.org/10.1038/s43247-021-00184-w

Rafat, A., Byun, E., Rezanezhad, F., Quinton, W. L., Humphreys, E. R., Webster, K., & Van Cappellen, P. (2022). The definition of the non-growing season matters: A case study of net ecosystem carbon exchange from a Canadian peatland. *Environmental Research Communications*, 4(2), 021003. https://doi.org/10.1088/2515-7620/ac53c2

Ratcliffe, J. L., Campbell, D. I., Clarkson, B. R., Wall, A. M., & Schipper, L. A. (2019). Water table fluctuations control CO2 exchange in wet and dry bogs through different mechanisms. The

Science of the Total Environment, 655, 1037–1046. https://doi.org/10.1016/j.scitotenv.2018.11.151

Ray, R. L., Ibironke, A., Kommalapati, R., & Fares, A. (2019). Quantifying the Impacts of Land-Use and Climate on Carbon Fluxes Using Satellite Data across Texas, U.S. *Remote Sensing (Basel, Switzerland)*, 11(14), 1733. https://doi.org/10.3390/rs11141733

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., ... & Valentini, R. (2005). On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. Global change biology, 11(9), 1424-1439. https://doi.org/10.1111/j.1365-2486.2005.001002.x

Reshadi, M. A. M., Rezanezhad, F., Shahvaran, A. R., Ghajari, A., Kaykhosravi, S., Slowinski, S., & Van Cappellen, P. (2025). Assessment of environmental and socioeconomic drivers of urban stormwater microplastics using machine learning. *Scientific Reports*, *15*(1), Article 6299. https://doi.org/10.1038/s41598-025-90612-0

Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., & Hashimoto, H. (2004). A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production. *Bioscience*, *54*(6), 547–560. https://doi.org/10.1641/0006-3568(2004)054[0547:ACSMOG]2.0.CO;2

Salomonson, V. V., Barnes, W. L., Maymon, P. W., Montgomery, H. E., & Ostrow, H. (1989). MODIS: Advanced facility instrument for studies of the Earth as a system. *IEEE Transactions on Geoscience and Remote Sensing*, 27(2), 145–153. https://doi.org/10.1109/36.20292

Saravani, M. J., Saadatpour, M., & Shahvaran, A. R. (2024). A web GIS based integrated water resources assessment tool for Javeh Reservoir. Expert Systems with Applications, 252, 124198. https://doi.org/10.1016/j.eswa.2024.124198

Satriawan, T. W., Nyberg, M., Lee, S.-C., Christen, A., Black, T. A., Johnson, M. S., Nesic, Z., Merkens, M., & Knox, S. H. (2023). Interannual variability of carbon dioxide (CO_2) and methane (CH_4) fluxes in a rewetted temperate bog. *Agricultural and Forest Meteorology*, *342*, 109696. https://doi.org/10.1016/j.agrformet.2023.109696

Schulze, C., Sonnentag, O., Voigt, C., Thompson, L., Delden, L., Heffernan, L., Hernandez-Ramirez, G., Kuhn, M., Lin, S., & Olefeldt, D. (2023). Nitrous Oxide Fluxes in Permafrost Peatlands Remain Negligible After Wildfire and Thermokarst Disturbance. *Journal of Geophysical Research. Biogeosciences*, *128*(4), Article e2022JG007322. https://doi.org/10.1029/2022JG007322

Skeeter, J., Christen, A., & Henry, G. H. R. (2022). Controls on carbon dioxide and methane fluxes from a low-center polygonal peatland in the Mackenzie River Delta, Northwest Territories. *Arctic Science*, *8*(2), 471–497. https://doi.org/10.1139/as-2021-0034

Strack, M., Lupascu, M., Byrne, K.A., Roulet, N., Turetsky, M., Waddington, J.M. 2023. Peatlands, greenhouse gas exchange and climate change. In: *Peatlands and Climate Change: 2nd Edition,* International Peatland Society.

Sonnentag, O., and Quinton, W. (2021), AmeriFlux BASE CA-SCB Scotty Creek Bog, Ver. 2-5, AmeriFlux AMP, (Dataset). https://doi.org/10.17190/AMF/1498754

Thayamkottu, S., Masta, M., Skeeter, J., Pärn, J., Knox, S. H., Smallman, T. L., & Mander, Ü. (2025). Dual controls of vapour pressure deficit and soil moisture on photosynthesis in a restored temperate bog. *The Science of the Total Environment*, *963*, Article 178366. https://doi.org/10.1016/j.scitotenv.2024.178366

Syed, K. H., Flanagan, L. B., Carlson, P. J., Glenn, A. J., & Van Gaalen, K. E. (2006). Environmental control of net ecosystem CO_2 exchange in a treed, moderately rich fen in northern Alberta. *Agricultural and Forest Meteorology*, *140*(1–4), 97–114. https://doi.org/10.1016/j.agrformet.2006.03.022

Taillardat, P., Bodmer, P., Deblois, C. P., Ponçot, A., Prijac, A., Riahi, K., Gandois, L., Giorgio, P. A., Bourgault, M. A., Tremblay, A., & Garneau, M. (2022). Carbon Dioxide and Methane Dynamics in a Peatland Headwater Stream: Origins, Processes and Implications. *Journal of Geophysical Research. Biogeosciences*, *127*(7), Article e2022JG006855. https://doi.org/10.1029/2022JG006855

Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58, 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Thornton, M. M., Shrestha, R., Wei, Y., Thornton, P. E. (2022). *Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 4*. ORNL DAAC. https://doi.org/10.3334/ORNLDAAC/2129

Toca, L., Artz, R. R. E., Smart, C., Quaife, T., Morrison, K., Gimona, A., Hughes, R., Hancock, M. H., & Klein, D. (2023). Potential for Peatland Water Table Depth Monitoring Using Sentinel-1 SAR Backscatter: Case Study of Forsinard Flows, Scotland, UK. *Remote Sensing (Basel, Switzerland)*, *15*(7), Article 1900. https://doi.org/10.3390/rs15071900

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M. A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S., & Papale, D. (2016). Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. *Biogeosciences*, *13*(14), 4291–4313. https://doi.org/10.5194/bg-13-4291-2016

Tran, H.-T., Hadi, M., Nguyen, T. T. H., Hoang, H. G., Nguyen, M.-K., Nguyen, K. N., & Vo, D.-V. N. (2023). Machine learning approaches for predicting microplastic pollution in peatland areas. *Marine Pollution Bulletin*, *194*, Article 115417. https://doi.org/10.1016/j.marpolbul.2023.115417

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127–150. https://doi.org/10.1016/0034-4257(79)90013-0

UNEP (2022). Global Peatlands Assessment – The State of the World's Peatlands: Evidence for action toward the conservation, restoration, and sustainable management of peatlands. Summary for Policy Makers. Global Peatlands Initiative. United Nations Environment Programme, Nairobi.

Virkkala, A., Aalto, J., Rogers, B. M., Tagesson, T., Treat, C. C., Natali, S. M., Watts, J. D., Potter, S., Lehtonen, A., Mauritz, M., Schuur, E. A. G., Kochendorfer, J., Zona, D., Oechel, W., Kobayashi, H., Humphreys, E., Goeckede, M., Iwata, H., Lafleur, P. M., Euskirchen, E.S., Bokhorst, S., Marushchak, M., Martikainen, P.J., Elberling, B., Voigt, C., Biasi, C., Sonnentag, O., Parmentier, F.-J.W., Ueyama, M., Celis, G., St.Louis, V.L., Emmerton, C.A., Peichl, M., Chi, J., Jarveoja, J., Nilsson, M.B., Oberbauer, S.F., Torn, M.S., Park, S.-J., Dolman, H., Mammarella, I., Chae, N., Poyatos, R., Lopes-Blanco, E., Christensen, T.R., Kwon, M.J., Torsten, S., Holl, D., & Luoto, M. (2021). Statistical upscaling of ecosystem CO2 fluxes across the terrestrial tundra and boreal domain: Regional patterns and uncertainties. *Global Change Biology*, *27*(17), 4040–4059. https://doi.org/10.1111/gcb.15659

Wang, J. M., Murphy, J. G., Geddes, J. A., Winsborough, C. L., Basiliko, N., & Thomas, S. C. (2013). Methane fluxes measured by eddy covariance and static chamber techniques at a temperate forest in central Ontario, Canada. *Biogeosciences*, *10*(6), 4371–4382. https://doi.org/10.5194/bg-10-4371-2013

Wang M., Wu J., Lafleur P., Luan J., Chen H., Zhu X. (2018). Can abandoned peatland pasture sequestrate more carbon dioxide from the atmosphere than an adjacent pristine bog in Newfoundland, Canada? *Agricultural and Forest Meteorology*, 248: 91-108. doi: https://doi.org/10.1016/j.agrformat.2017.09.010

Wang, Y., Guo, M., Vo Thanh, H., Zhang, H., Liu, X., Zheng, Q., Zhang, X., Daoud, M. S., & Abualigah, L. (2025). Advanced machine learning schemes for prediction CO2 flux based experimental approach in underground coal fire areas. *Journal of Advanced Research*, *70*, 587–601. https://doi.org/10.1016/j.jare.2024.10.034

Ward, E. J., Warren, J. M., McLennan, D. A., Dusenge, M. E., Way, D. A., Wullschleger, S. D., & Hanson, P. J. (2019). Photosynthetic and respiratory responses of two bog shrub species to whole ecosystem warming and elevated CO_2 at the boreal-temperate ecotone. *Frontiers in Forests and Global Change*, 2. https://doi.org/10.3389/ffgc.2019.00054

Webster, K. L., Bhatti, J. S., Thompson, D. K., Nelson, S. A., Shaw, C. H., Bona, K. A., Hayne, S. L., & Kurz, W. A. (2018). Spatially-integrated estimates of net ecosystem exchange and methane fluxes from Canadian peatlands. *Carbon Balance and Management*, *13*(1), 16–21. https://doi.org/10.1186/s13021-018-0105-5

Wickham, H., François, R., Henry, L., Müller, K., Vaughan D. (2023a). dplyr: A Grammar of Data Manipulation. *R* package version 1.1.2, https://CRAN.R-project.org/package=dplyr

Wickham, H., Vaughan, D., Girlich, M. (2023b). tidyr: Tidy Messy Data. *R* package version 1.3.0, https://CRAN.R-project.org/package=tidyr

Wu, J., & Roulet, N. T. (2014). Climate change reduces the capacity of northern peatlands to absorb the atmospheric carbon dioxide: The different responses of bogs and fens. *Global Biogeochemical Cycles*, *28*(10), 1005–1024. https://doi.org/10.1002/2014GB004845

Wu, J., Roulet, N. T., Moore, T. R., Lafleur, P., & Humphreys, E. (2011). Dealing with microtopography of an ombrotrophic bog for simulating ecosystem-level CO_2 exchanges. *Ecological Modelling*, *222*(4), 1038–1047. https://doi.org/10.1016/j.ecolmodel.2010.07.015

Wu, H., Yang, T., Li, H., & Zhou, Z. (2023). Air quality prediction model based on mRMR–RF feature selection and ISSA–LSTM. *Scientific Reports*, *13*(1), Article 12825. https://doi.org/10.1038/s41598-023-39838-4

Wu, Y., Luo, M., Ding, S., & Han, Q. (2024). Using a Light Gradient-Boosting Machine–Shapley Additive Explanations Model to Evaluate the Correlation Between Urban Blue–Green Space Landscape Spatial Patterns and Carbon Sequestration. *Land (Basel)*, *13*(11), Article 1965. https://doi.org/10.3390/land13111965

Wutzler T, Lucas-Moffat A, Migliavacca M, Knauer J, Sickel K, Sigut, Menzer O & Reichstein M (2018) Basic and extensible post-processing of eddy covariance flux data with REddyProc. Biogeosciences, Copernicus, 15, doi: 10.5194/bg-15-5015-2018

Wylie, B. K., Fosnight, E. A., Gilmanov, T. G., Frank, A. B., Morgan, J. A., Haferkamp, M. R., & Meyers, T. P. (2007). Adaptive data-driven models for estimating carbon fluxes in the Northern Great Plains. *Remote Sensing of Environment*, *106*(4), 399–413. https://doi.org/10.1016/j.rse.2006.09.017

Xiao, J., Zhuang, Q., Baldocchi, D. D., Law, B. E., Richardson, A. D., Chen, J., Oren, R., Starr, G., Noormets, A., Ma, S., Verma, S. B., Wharton, S., Wofsy, S. C., Bolstad, P. V., Burns, S. P., Cook, D. R., Curtis, P. S., Drake, B. G., Falk, M., Fischer, M.L., Foster, D.R., Gu, L., Hadley, J.L., Hollinger, D.Y., Katul, G.G., Litvak, M., Martin, T.A., Matamala, R., McNulty, S., Meyers, T.P., Monson, R.K., Munger, J.W., Oechel, W.C., Tha Paw U, K., Schmid, H.P., Scott, R.L., Sun, G., Suyker, A.E., & Torn, M. S. (2008). Estimation of net ecosystem carbon exchange for the conterminous United States by combining MODIS and AmeriFlux data. *Agricultural and Forest Meteorology*, *148*(11), 1827–1847. https://doi.org/10.1016/j.agrformet.2008.06.015

Yamaji, T., Sakai, T., Endo, T., Baruah, P. J., Akiyama, T., Saigusa, N., Nakai, Y., Kitamura, K., Ishizuka, M., & Yasuoka, Y. (2008). Scaling-up technique for net ecosystem productivity of deciduous broadleaved forests in Japan using MODIS data. *Ecological Research*, *23*(4), 765–775. https://doi.org/10.1007/s11284-007-0438-0

Yu, Z., Loisel, J., Brosseau, D. P., Beilman, D. W., & Hunt, S. J. (2010). Global peatland dynamics since the Last Glacial Maximum. Geophysical Research Letters, 37(13). https://doi.org/10.1029/2010GL043584

Zeileis, A., Grothendieck, G. (2005). zoo: S3 Infrastructure for Regular and Irregular Time Series. Journal of Statistical Software, 14(6), 1-27. https://doi.org/10.18637/jss.v014.i06

Zeng, J., Matsunaga, T., and Shirai, T.: A new estimate of oceanic CO_2 fluxes by machine learning reveals the impact of CO_2 trends in different methods, Earth Syst. Sci. Data Discuss. [preprint], https://doi.org/10.5194/essd-2022-71, 2022.

Zhao, J., Peichl, M., Öquist, M., & Nilsson, M. B. (2016). Gross primary production controls the subsequent Winter CO_2 exchange in a boreal peatland. *Global Change Biology*, 22(12), 4028–4037. https://doi.org/10.1111/gcb.13308

Zhao, Z., De Frenne, P., Peñuelas, J., Van Meerbeek, K., Fornara, D. A., Peng, Y., Wu, Q., Ni, X., Wu, F., & Yue, K. (2022). Effects of snow cover-induced microclimate warming on soil physicochemical and biotic properties. *Geoderma*, *423*, 115983-. https://doi.org/10.1016/j.geoderma.2022.115983

Zhong, Y., Jiang, M., & Middleton, B. A. (2020). Effects of water level alteration on carbon cycling in peatlands. *Ecosystem Health and Sustainability*, *6*(1), Article 1806113. https://doi.org/10.1080/20964129.2020.1806113

Zhou, Y., Sachs, T., Li, Z., Pang, Y., Xu, J., Kalhori, A., Wille, C., Peng, X., Fu, X., Wu, Y., & Wu, L. (2023). Long-term effects of rewetting and drought on GPP in a temperate peatland based on satellite remote sensing data. *The Science of the Total Environment*, *882*, Article 163395. https://doi.org/10.1016/j.scitotenv.2023.163395

Zolkos, S., Tank, S. E., Kokelj, S. V., Striegl, R. G., Shakil, S., Voigt, C., Sonnentag, O., Quinton, W. L., Schuur, E. A. G., Zona, D., Lafleur, P. M., Sullivan, R. C., Ueyama, M., Billesbach, D., Cook, D., Humphreys, E. R., & Marsh, P. (2022). Permafrost Landscape History Shapes Fluvial Chemistry, Ecosystem Carbon Balance, and Potential Trajectories of Future Change. *Global Biogeochemical Cycles*, *36*(9). https://doi.org/10.1029/2022GB007403

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Appendices

Appendix I: Supplementary Materials for Chapter 2

Introduction. This section contains Supplementary Materials including: Supplementary Results (Peatland-Corrected-SMAP-NEE Results); Table AI-1 (Coefficients of variation (%) for each measure of net ecosystem exchange); Table AI-2 (Direct comparison of Peatland-Corrected-SMAP-NEE and EC-NEE (g C m^{-2} day^{-1})); Figure AI-1 **(**Boxplots of measures of daily NEE (g C m^{-2} day^{-1}); Figure AI-2 (Measures of annual NEE (g C m^{-2} $year^{-1}$); Figure AI-3 (Scatterplots of Eddy Covariance NEE (g C m^{-2} day^{-1}) vs SMAP-NEE (g C m^{-2} day^{-1}); Figure AI-4 (Scatterplots of EC-GPP (g C m^{-2} day^{-1}) vs SMAP-GPP (g C m^{-2} day^{-1}) with their regression line and associated R^2 and *p*-value.

Supplementary Results: Peatland-Corrected-SMAP-NEE Results

The correction incorporating data from all sites (Peatland-Corrected-SMAP-NEE) greatly reduced the variability of the SMAP-NEE fluxes. After the correction, with all sites considered for the GS, on average Corrected-SMAP-NEE were greater than EC-NEE values by 0.06 g C m^{-2} day^{-1} (95% CI= (-0.36,0.47)) (Table AI-1). After the correction, with all sites considered for the NGS, on average Corrected-SMAP-NEE were smaller than EC-NEE values by 0.04 g C m^{-2} day^{-1} (95% CI= (-0.25,0.18)) (Table AI-1). Additionally, the RMSE with all sites considered for the daily fluxes for the Peatland-Corrected-SMAP-NEE during the GS was 1.02 g C m^{-2} day^{-1}, whereas during the NGS 0.59 g C m^{-2} day^{-1}. These RSME values are lower than those for the SMAP-NEE dataset, but higher than those for the correction done on a site-by-site basis.

Table AI-1: Coefficients of variation (%) for each measure of net ecosystem exchange.

Site	Unprocessed SMAP-NEE	SMAP-NEE corrected by LULC	Corrected-SMAP-NEE	EC-NEE
All Sites	32.3745	50.64	12.541	4.9036
CA-DBB	4.400102	4.400806	3.321648	5.624485
CA-KLP	30.1431	30.14768	2.364711	1.290648
CA-RPn	25.47214	25.47605	2.581086	3.485249
CA-RPp	12.57595	12.57789	9.1532	6.195177
CA-SCB	8.826266	8.827629	31.81852	3.998044

Table AI-2: Direct comparison of Peatland-Corrected-SMAP-NEE and EC-NEE (g C m^{-2} day^{-1}). Negative values indicate underestimation, whereas positive values indicate overestimation.

	Sample Size, *n*		**Peatland-Corrected-SMAP-NEE compared to EC-NEE, mean (95% CI) (g C m^{-2} day^{-1})**			
	GS	NGS	GS	*p*-value	NGS	*p*-value
All Sites	5	5	0.06 (-0.36,0.47)	0.06	-0.04 (-0.25,0.18)	0.06
CA-KLP	159	55	-0.01 (-0.09,0.06)	0.18	0.16 (0.09,0.23)	<0.001
CA-RPp	40	50	0.62 (0.07,1.17)	0.06	-0.21 (0.30,-0.120	<0.001
CA-RPn	40	51	-0.19 (-0.40,0.005)	1.08	-0.21 (-0.29,-0.13)	<0.001
CA-SCB	183	153	-0.19 (-0.25,-0.13)	<0.001	0.08 (0.05,0.11)	<0.001
CA-DBB	272	236	0.06 (-0.10,0.22)	0.25	-0.001 (-0.11,0.10)	<0.001

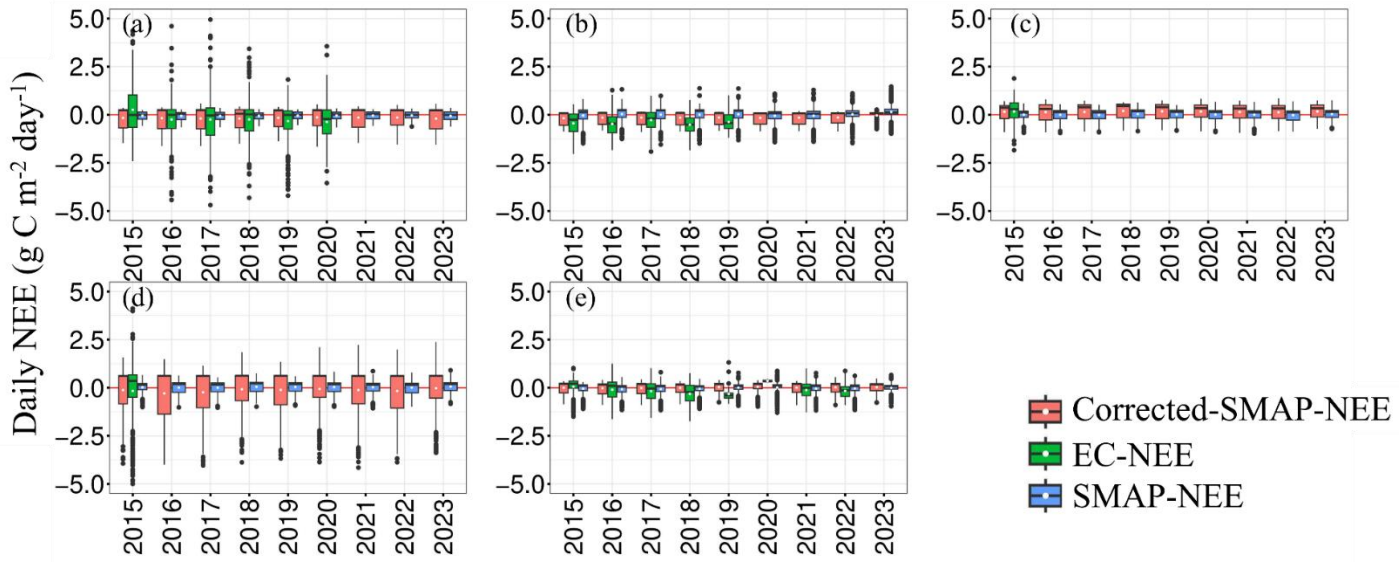


Figure AI-1: Boxplots of measures of daily NEE (g C m^{-2} day^{-1}) at a) CA-DBB, b) CA-KLP, c) CA-RPn, d) CA-RPp, and e) CA-SCB. Notable periods of EC-NEE data gaps are represented in Table 2-2.

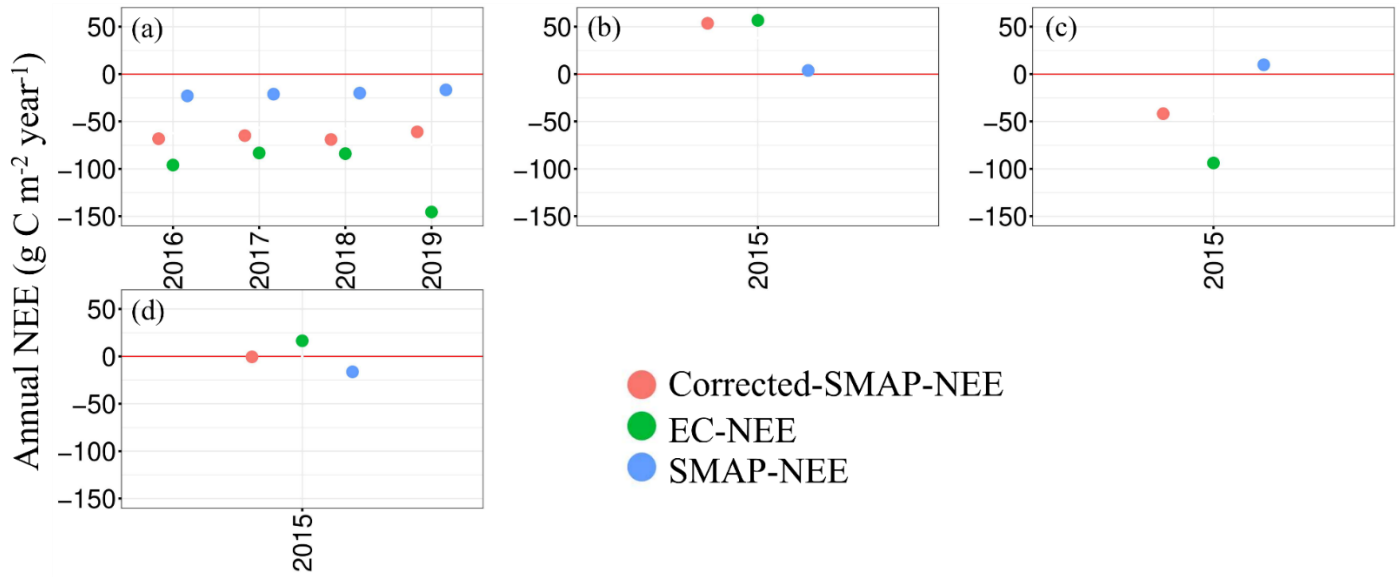


Figure AI-2: Measures of annual NEE (g C m^{-2} $year^{-1}$) at a) CA-DBB, b) CA-RPn, c) CA-RPp, and d) CA-SCB. Only years without large EC-NEE data gaps (Table 2-2) are represented.

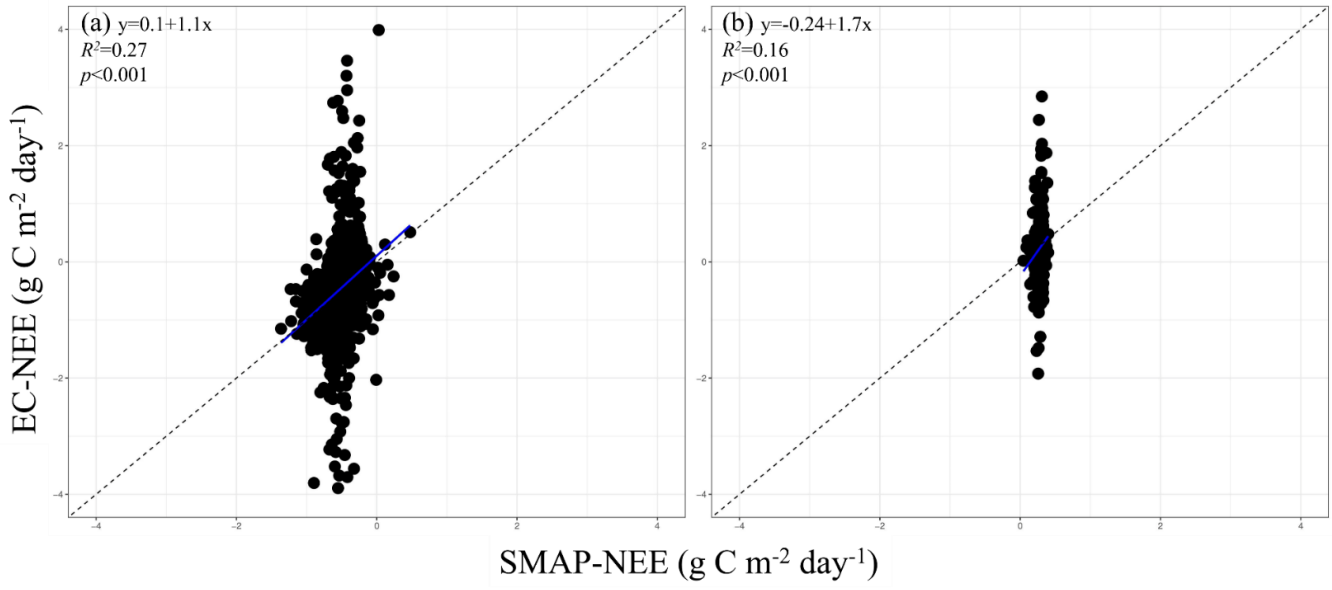


Figure AI-3: Scatterplots of Eddy Covariance NEE (EC-NEE; g C m^{-2} day^{-1}) vs SMAP-NEE (g C m^{-2} day^{-1}) in the GS (left) and NGS (right) with their regression line and associated R^2 and p-value. The dashed black line represents the 1:1 line.

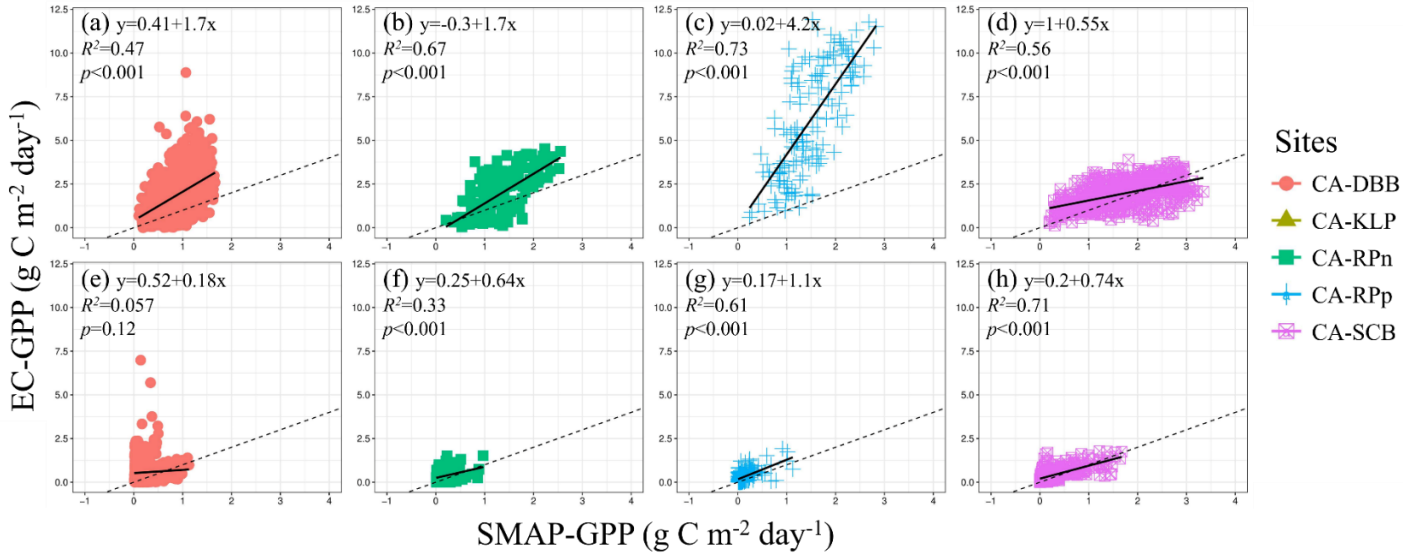


Figure AI-4: Scatterplots of EC-GPP (g C m^{-2} day^{-1}) vs SMAP-GPP (g C m^{-2} day^{-1}) with their regression line and associated R^2 and p-value at a) CA-DBB in the GS, b) CA-RPn in the GS, c) CA-RPp in the GS, d) CA-SCB in the GS, e) CA-DBB in the NGS, f) CA-RPn in the NGS, g) CA-RPp in the NGS, h) CA-SCB in the NGS. The dashed black line represents the 1:1 line.

Appendix II: Supplementary Materials for Chapter 3
Observation-Based NEE Fluxes (Further Details)

Half-hourly net ecosystem exchange (NEE) data for each peatland site were downloaded from the AmeriFlux BASE repository (AmeriFlux, 2025) and imported into the REddyProcWeb online platform (Wutzler *et al.*, 2018) to perform three key steps: friction-velocity (u^*) filtering, gap-filling, and partitioning of NEE into gross primary productivity (GPP) and ecosystem respiration (ER). In the eddy covariance (EC) approach, fast-response instruments record turbulent fluctuations in vertical wind speed and CO_2 concentration at 10–20 Hz, which are block-averaged to half-hourly fluxes under the assumption of fully developed turbulence (Papale *et al.*, 2006). However, this assumption breaks down during low-turbulence conditions, most prevalent at night, leading to underestimation of NEE. To correct for this bias, REddyProcWeb implements a u^*-filtering algorithm: the year is divided into four overlapping three-month windows (Mar–May, Jun–Aug, Sep–Nov, and Dec–Feb, with the December window extending into the following calendar year to capture continuous winter conditions) (Wutzler *et al.*, 2018). Within each window, observations are first binned into six temperature quantiles and then each quantile's data are subdivided into twenty u^* classes (Papale *et al.*, 2006). Any u^* class containing fewer than three nighttime observations is excluded to ensure statistical robustness (Wutzler *et al.*, 2018). For each temperature quantile, the u^* threshold is defined as the lowest u^* class whose mean nocturnal NEE reaches >99% of the mean flux observed in all higher-u^* classes (Reichstein *et al.*, 2005; Papale *et al.*, 2006). To prevent spurious thresholds, an auxiliary check identifies if two consecutive u^* classes exhibit mean flux ≥0.95 × the average of the next ten higher-u^* classes; if so, the lower of the two classes is selected as the threshold (Wutzler *et al.*, 2018). Candidate thresholds are then accepted only if the correlation coefficient between u^* and air temperature is weak ($|r| < 0.4$), as recommended by Papale *et al.* (2006); if no suitable threshold emerges, the 90th percentile of the

*u** distribution is applied. The final site- and season-specific *u** threshold is taken as the median of the six quantile-based values (Papale *et al.*, 2006).

After removing all fluxes measured below the *u** threshold, gaps in the resulting half-hourly NEE time series are filled using a three-condition hierarchical lookup-table method adapted from Falge *et al.* (2001) and Reichstein *et al.* (2005). Condition 1 applies when only NEE is missing but global shortwave radiation (Rg), air temperature (Tair), and vapor pressure deficit (VPD) are available, in which case missing fluxes are estimated from the average of similar meteorological conditions within a ±14-day window. Under Condition 2, when NEE and either Tair or VPD are missing but Rg is available, the same lookup approach is used within a ±7-day window. Under Condition 3, when all meteorological inputs are missing, flux gaps are linearly interpolated from adjacent hours. If a gap cannot be filled under one condition, the window is expanded until the gap is filled or until a 60-day limit is reached, beyond which gaps remain unfilled (this procedure does not fill pre-existing gaps that are 60 days or longer) (Wutzler *et al.*, 2018).

Following gap-filling, REddyProcWeb partitions NEE into GPP and ER using its standard algorithms and then computes daily averages when ≥70% of half-hourly observations are present, and monthly averages when ≥10 daily values exist. To quantify uncertainty in *u** thresholds, 200 seasonal pseudoreplicate datasets are generated per three-month window, and thresholds are re-estimated for each; the 5^{th}, 50^{th}, and 95^{th} percentiles of these estimates define the uncertainty range (Wutzler *et al.*, 2018). This end-to-end processing ensures that our half-hourly EC fluxes are rigorously quality-controlled, gap-filled, and partitioned in accordance with best practices.

Table AII-1: Monthly means (± 95% CI) of daily-averaged predictor variables and net ecosystem exchange (NEE), stratified by US EPA Level II ecoregion. Values bearing different superscript letters within each column differ significantly ($p < 0.05$, Kruskal-Wallis). Predictor variables are defined as follows: ET_M = total evapotranspiration, cwd = climate water deficit, srad_DM = incident shortwave radiation flux density, NDSI = normalized difference snow index, ws = wind speed at 10 m height, swe_DM = snow water equivalent, sm = soil moisture, and BAI = burn area index.

Ecoregion	Number of Observations	NEE (g C m^{-2} day^{-1})	ET_M (mm)	cwd (mm)	srad_DM (W m^{-2})	NDSI	ws (m s^{-1})	swe_DM (kg m^{-2})	sm (mm)	BAI
Boreal Plain	102	-0.36^{a} (-0.53, -0.19)	1.30^{ac} (1.11, 1.50)	20.1^{a} (15.0, 25.3)	253.8^{a} (233.3, 274.4)	-0.21^{a} (-0.29, -0.13)	3.18^{b} (3.11, 3.25)	14.6^{a} (9.57, 19.6)	22.8^{f} (20.0, 25.7)	40.6^{ad} (36.3, 44.9)
Hudson Plain	142	-0.12^{a} (-0.21, -0.03)	1.14^{a} (0.99, 1.28)	9.06^{b} (5.59, 12.5)	252.5^{a} (233.0, 271.9)	-0.07^{ab} (-0.16, 0.02)	3.35^{a} (3.18, 3.52)	34.3^{e} (26.3, 42.4)	26.7^{e} (25.0, 28.3)	38.3^{d} (33.3, 43.3)
Marine West Coast Forest	83	-0.29^{a} (-0.46, -0.11)	1.47^{bc} (1.28, 1.65)	20.8^{a} (13.4, 28.2)	253.5^{a} (229.7, 277.3)	-0.51^{d} (-0.55, -0.48)	3.01^{a} (2.91, 3.11)	0.31^{d} (-0.13, 0.74)	80.7^{d} (72.1, 89.3)	75.8^{c} (68.5, 83.0)
Mixed Wood Plains	215	-0.27^{a} (-0.36, -0.18)	1.63^{bc} (1.45, 1.81)	6.56^{b} (4.68, 8.43)	288.0^{b} (275.1, 301.0)	-0.17^{c} (-0.24, -0.10)	2.79^{c} (2.72, 2.85)	32.1^{ce} (25.3, 38.8)	71.6^{c} (68.7, 74.5)	28.2^{b} (24.9, 31.4)
Softwood Shield	39	-0.12^{a} (-0.27, 0.03)	1.81^{b} (1.48, 2.14)	4.33^{b} (1.01, 7.65)	324.5^{b} (288.1, 360.8)	-0.25^{bc} (-0.40, -0.10)	4.20^{d} (3.83, 4.56)	82.1^{bce} (46.6, 117.8)	34.3^{b} (30.8, 37.8)	22.1^{b} (18.4, 25.8)
Taiga Plain	36	0.06^{a} (-0.04, 0.16)	0.86^{a} (0.70, 1.02)	11.0^{ab} (4.29, 17.7)	214.9^{a} (177.7, 252.1)	-0.08^{a} (-0.26, 0.09)	2.36^{e} (2.22, 2.50)	29.6^{abce} (15.9, 43.4)	17.8^{a} (13.9, 21.8)	49.9^{a} (39.6, 60.1)

Table AII-2: Performance metrics for the leave-one-ecoregion-out sensitivity analysis.

Ecoregion	Number of Observations	MAE (g C m^{-2} day^{-1})	MSE (g C m^{-2} day^{-1})	RMSE (g C m^{-2} day^{-1})	MAPE (%)	R^2
Boreal Plain	77	0.41	0.45	0.67	1.29	0.41
Hudson Plain	106	0.25	0.13	0.37	1.89	0.53
Marine West Coast Forest	62	0.43	0.37	0.61	2.40	0.49
Mixed Wood Plains	161	0.21	0.08	0.29	0.58	0.80
Softwood Shield	29	0.41	0.26	0.51	5.41	−0.07
Taiga Plain	27	0.16	0.05	0.22	0.68	0.52

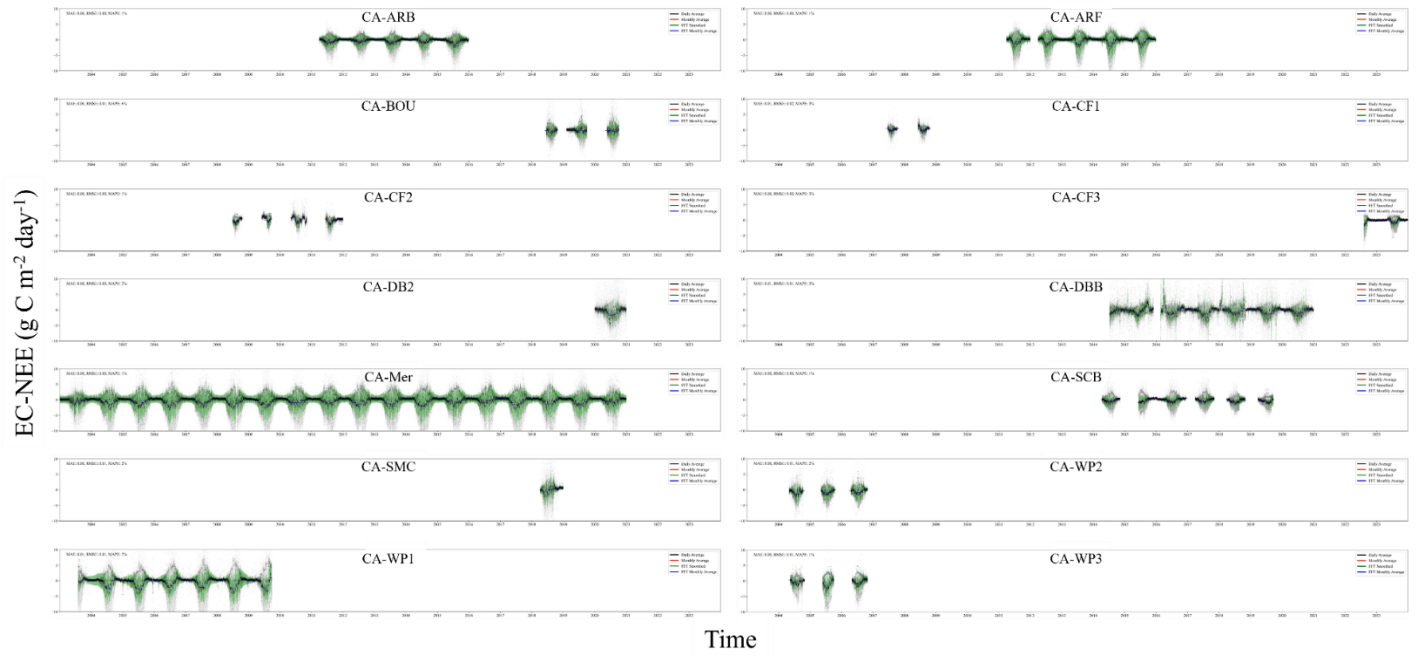


Figure AII-1: Time series of half-hourly eddy covariance (EC) net ecosystem exchange (NEE) (grey dots), FFT (fast Fourier transform)-smoothed fluxes (green line), and their corresponding monthly averages (red = raw; blue = smoothed) for each peatland site. The congruence of red and blue monthly lines (appearing purple) demonstrates that FFT filtering does not alter the aggregated monthly NEE estimates.

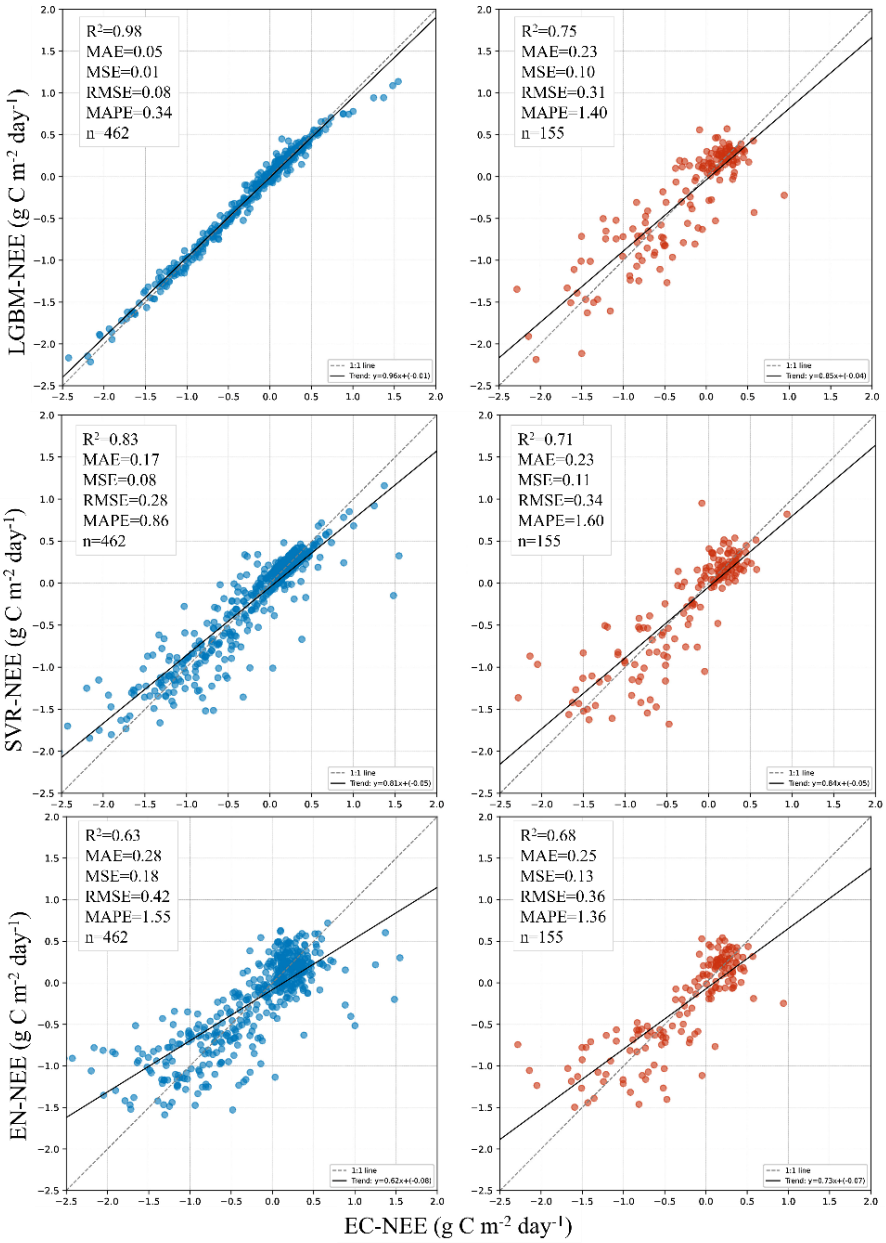


Figure AII-2: Light Gradient-Boosting Machine (LGBM, top row), Support Vector Regression (SVR, middle row), and ElasticNet Regression (EN, bottom row) predictions versus eddy covariance observations of daily CO_2 net ecosystem exchange (NEE) for (a) the training set (n = 462) and (b) the internal test set (n = 155). Colored points show site-level monthly averages; the gray dashed line is the 1:1 reference; and the solid black line is the fitted regression. Key performance metrics (R^2, MAE, MSE, RMSE, MAPE) are listed in each inset.

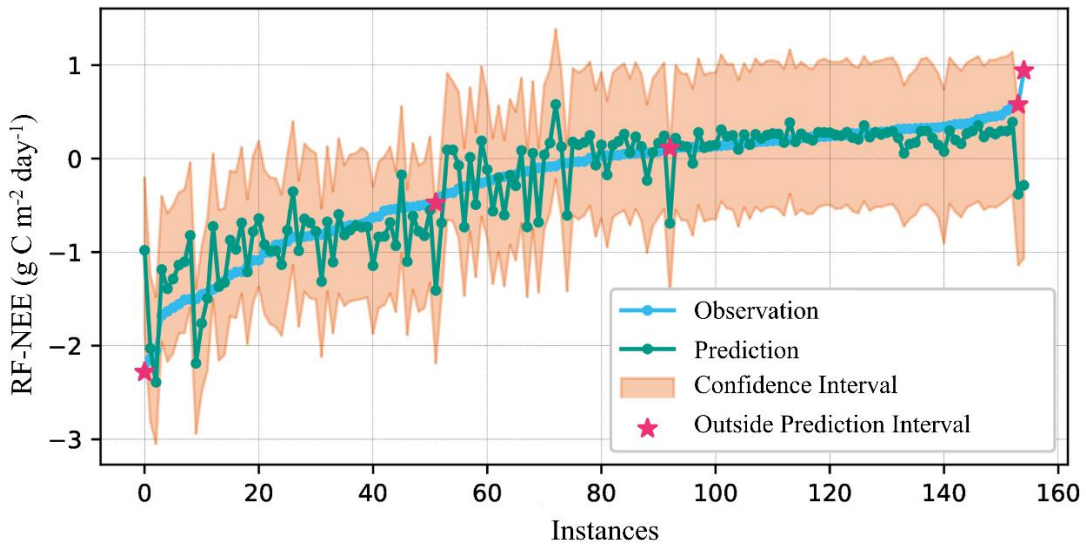


Figure AII-3: Random Forest (RF) predictive performance for daily CO_2 net ecosystem exchange (NEE). Observed EC-NEE (blue) and RF predictions (green) are plotted across validation instances, with the 95% prediction interval shown in orange (mean width = 1.56 g C m^{-2} day^{-1}). Pink stars mark monthly averages falling outside the interval.

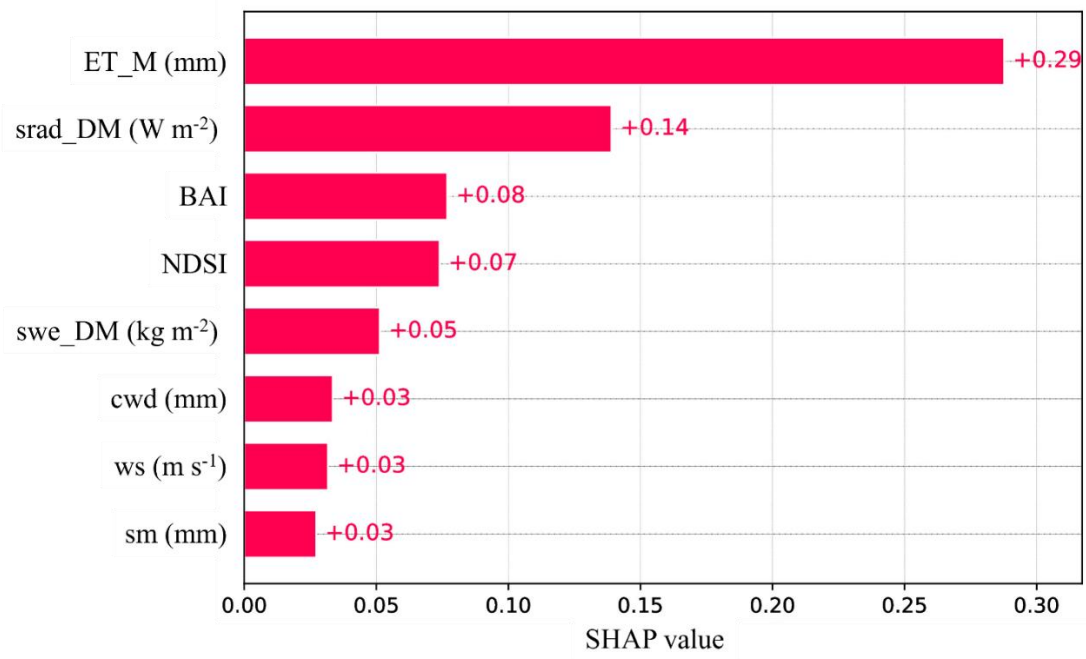


Figure AII-4: Bar chart of relative SHAP feature importance for eight predictors: evapotranspiration (ET_M), shortwave radiation (srad_DM), burn area index (BAI), normalized difference snow index (NDSI), snow water equivalent (swe_DM), climate water deficit (cwd), wind speed (ws), and soil moisture (sm).

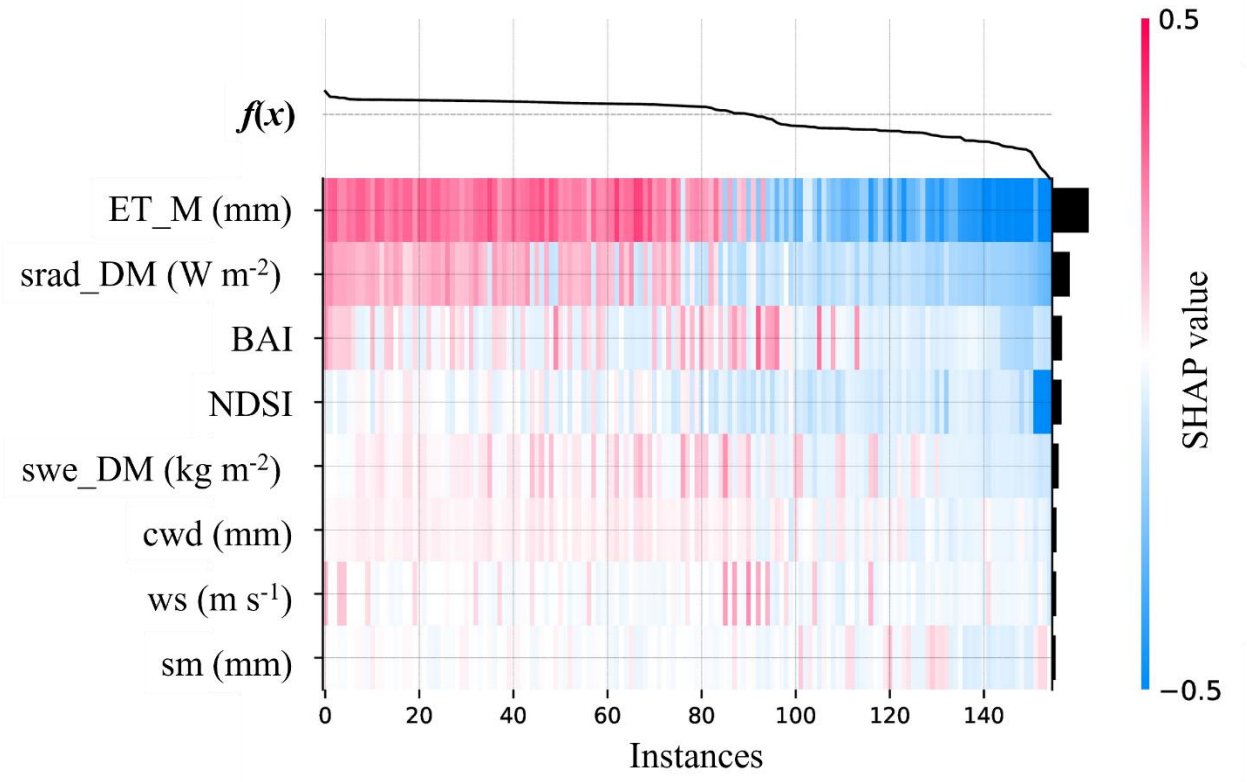


Figure AII-5: SHAP (SHapley Additive exPlanations) summary heatmap displaying the impact of eight predictors: evapotranspiration (ET_M), shortwave radiation (srad_DM), burn area index (BAI), normalized difference snow index (NDSI), snow water equivalent (swe_DM), climate water deficit (cwd), wind speed (ws), and soil moisture (sm), on predicted eddy covariance net ecosystem exchange (NC-NEE) of CO_2 ($\boldsymbol{f(x)}$). Observations (bottom axis) are ordered by decreasing predicted NC-NEE; cell colors denote SHAP values (red = positive, blue = negative), the right-hand bar chart ranks features by mean absolute SHAP importance, and the top line plot traces $\boldsymbol{f(x)}$ for each instance relative to its average.

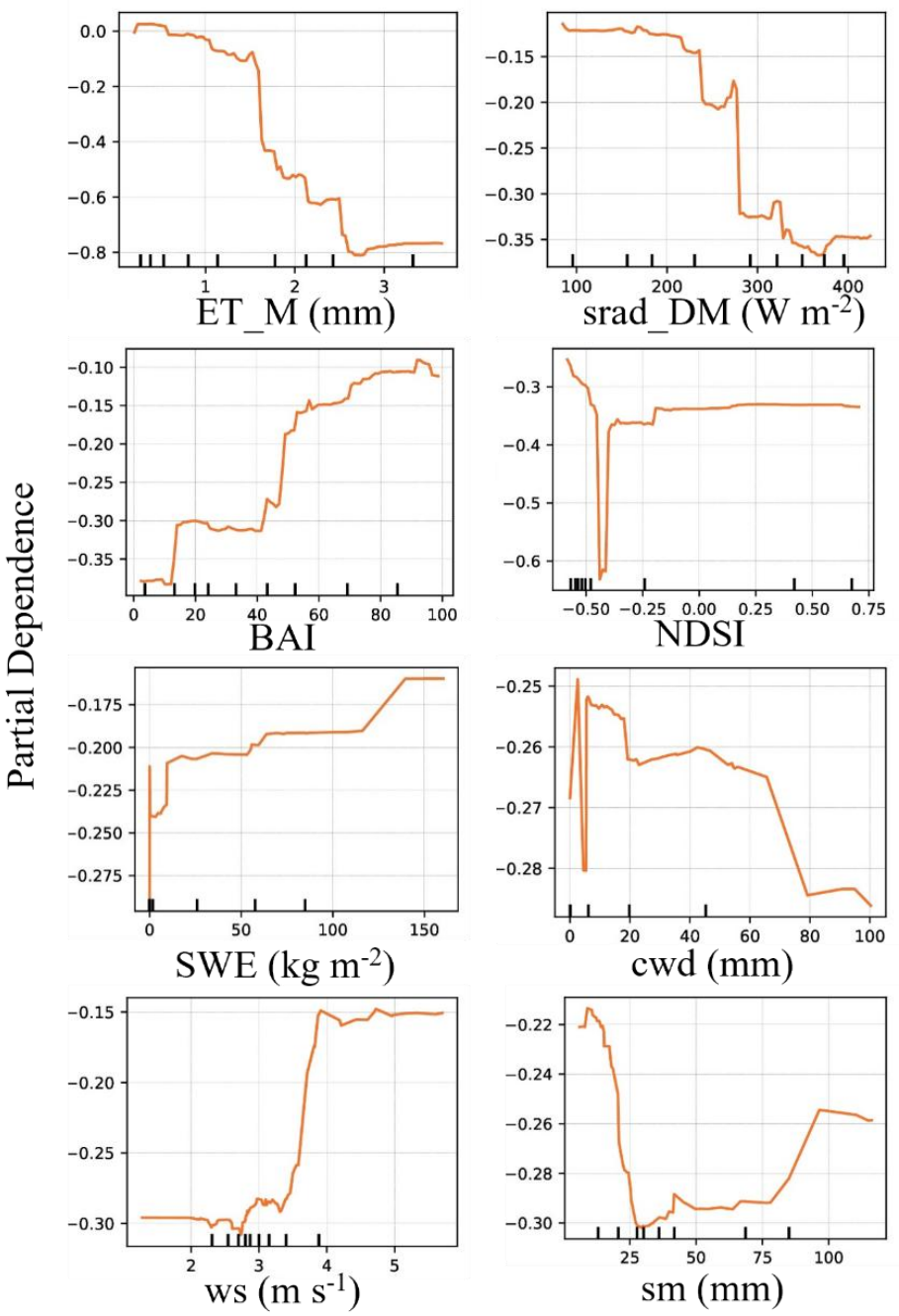


Figure AII-6: Partial dependence plots from the Random Forest model illustrating how each of the eight key predictors: evapotranspiration (ET_M), shortwave radiation (srad_DM), burn area index (BAI), normalized difference snow index (NDSI), snow water equivalent (swe_DM), climate water deficit (cwd), wind speed (ws), and soil moisture (sm), influences daily CO_2 net ecosystem exchange, with tick marks indicating the distribution of observed values.

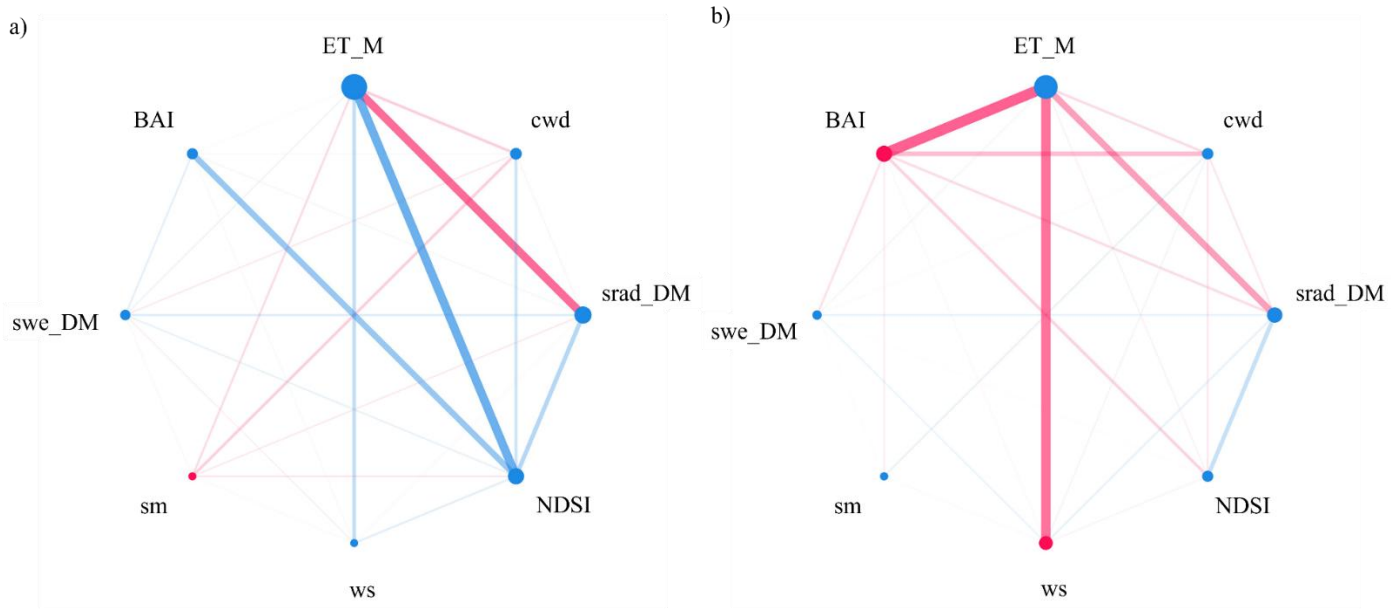


Figure AII-7: SHAP (SHapley Additive exPlanations) interaction networks from the Random Forest regression model for (a) the most negative (blue) and (b) the most positive (red) predicted net ecosystem exchange (NEE) samples. Each feature: evapotranspiration (ET_M), shortwave radiation (srad_DM), burn area index (BAI), normalized difference snow index (NDSI), snow water equivalent (swe_DM), climate water deficit (cwd), wind speed (ws), and soil moisture (sm), is shown as a node, its size proportional to its SHAP contribution to NEE, and edges connect features with line width scaled by their SHAP interaction strength.

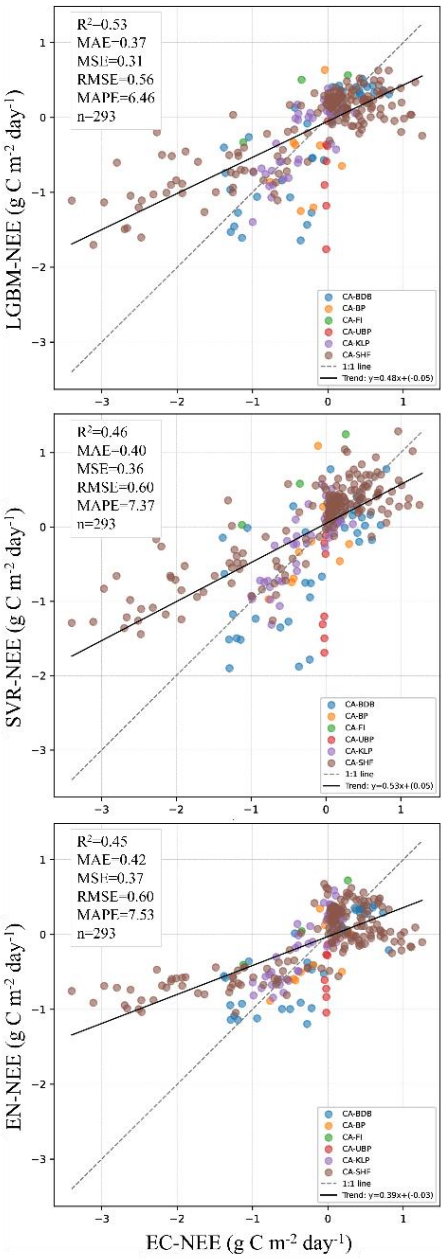


Figure AII-8: Light Gradient-Boosting Machine (LGBM, top row), Support Vector Regression (SVR, middle row), and ElasticNet Regression (EN, bottom row) predictions versus eddy covariance observations of daily CO_2 net ecosystem exchange (NEE) at the external validation sites (n = 293). Colored points denote site-level monthly averages; the gray dashed line is the 1:1 reference; and the solid black line is the fitted regression. Key performance metrics (R^2, MAE, MSE, RMSE, MAPE) are listed in each inset.