ARTICLE

Check for updates

# Non-growing season carbon emissions in a northern peatland are projected to increase under global warming

Arash Rafat[1,2], Fereidoun Rezanezhad [2✉], William L. Quinton[3], Elyn R. Humphreys[4], Kara Webster[5] & Philippe Van Cappellen[2]

Peatlands are important ecosystems that store approximately one third of terrestrial organic carbon. Non-growing season carbon fluxes significantly contribute to annual carbon budgets in peatlands, yet their response to climate change is poorly understood. Here, we investigate the governing environmental variables of non-growing season carbon emissions in a northern peatland. We develop a support-vector regression model using a continuous 13-year dataset of eddy covariance flux measurements from the Mer Blue Bog, Canada. We determine that only seven variables were needed to reproduce carbon fluxes, which were most sensitive to net radiation above the canopy, soil temperature, wind speed and soil moisture. We find that changes in soil temperature and photosynthesis drove changes in net carbon flux. Assessing net ecosystem carbon exchange under three representative concentration pathways, we project a 103% increase in peatland carbon loss by 2100 under a high emissions scenario. We suggest that peatland carbon losses constitute a strong positive climate feedback loop.

[1] Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, ON, Canada. [2] Ecohydrology Research Group, Department of Earth and Environmental Sciences and Water Institute, University of Waterloo, Waterloo, ON, Canada. [3] Cold Regions Research Centre, Wilfrid Laurier University, Waterloo, ON, Canada. [4] Department of Geography and Environmental Studies, Carleton University, Ottawa, ON, Canada. [5] Canadian Forest Service Great Lakes Forestry Centre—Natural Resources Canada, Sault Ste Marie, ON, Canada. ✉email: frezanez@uwaterloo.ca

While peatlands are estimated to cover only 3% of the continents[1], they store approximately 30% of land-based organic carbon (C) (300-450 Pg C)[2–5]. Northern peatlands are of particular interest because high latitude regions are warming at greater rates than the global average[6–10]. Furthermore, the greatest warming in cold regions, including northern peatlands, is occurring during the nongrowing season (NGS)[11,12]. Many previous studies have focused on peatland C dynamics during the growing season, given the inherent difficulties of measuring NGS carbon dioxide (CO_2) fluxes in remote cold regions. However, a growing body of literature shows that net ecosystem CO_2 exchange (NEE) from various landscapes during the NGS is nontrivial and may contribute significantly to annual ecosystem C budgets[12–20]. Despite these efforts, there is a lack of predictive understanding of how NGS CO_2 emissions from northern peatlands will change under an evolving and uncertain 21st-century climate.

Only a few studies have projected future NGS–NEE of peatland ecosystems under representative climate concentration pathways (RCPs) for northern regions[12,21–23]. In part, this reflects an incomplete understanding of the key drivers of NGS soil respiration. Previous work, however, shows that soil CO_2 effluxes during the NGS are modulated by the interactions between numerous environmental parameters, including soil temperature and moisture[12,24–26], snow depth[25,27–30], vegetation cover[31], availability of labile C substrates, and the soil microbial abundance and community structure[32–34].

Deterministic models are traditionally used for understanding environmental processes. The processes included, and their model representations, are derived from prior knowledge and theoretical considerations. The models in turn form the basis for hypothesis-driven experimentation with, as a possible unintended consequence, biases the observational data due to the underlying hypotheses[35]. As an example of this traditional deterministic approach, terrestrial ecosystem C cycling has been successfully simulated using process-based dynamic vegetation models. As an alternative approach, data-driven (rather than hypothesis-driven) modeling techniques, including machine- and deep-learning methods, are increasingly employed to derive quantitative relationships between environmental variables and ecosystem functions (e.g., soil respiration) based on patterns in the measured data.

Machine learning (ML) is the study of statistical models and algorithms[36] and has been extensively applied to the modeling of carbon fluxes measured at eddy covariance (EC) flux towers from a variety of ecosystem types[37–40]. The applicability of ML to modeling carbon fluxes was best demonstrated in the recent work by Tramontana et al.[41] who used a large ensemble of ML-based models in predicting CO_2, energy, and radiative fluxes across various plant-functional types. Similar studies have achieved great success in modeling CO_2 fluxes using ML techniques, as shown in the studies conducted by Cai et al.[36], Melesse and Hanley[39], and Xiao et al.[42] who used gradient boosting and random forest, multilayer artificial neural networks, and regression trees respectively to predict carbon fluxes measured at EC flux towers for various applications.

In this study, we train an ML model for NGS-NEE on a 13-year (1998–2010) continuous record of EC flux measurements at the Mer Bleue Bog located in Ottawa, Canada. The most important variables modulating the NGS–NEE at the research site are determined using a variable selection methodology and a moment-independent global sensitivity analysis (GSA) method. In addition, we project the NGS–NEE CO_2 emission rates over the remainder of the 21st century by considering how the key environmental variables will change under a low, moderate, and high radiative forcing scenario (RCP2.6, RCP4.5, RCP8.5,

respectively). The findings of this study provide novel insights into NGS CO_2 emissions from northern peatlands as they transition into a warmer world, with implications for future climate policy, evolving northern landscapes, and the associated hydro-meteorological, snow, and frozen-ground processes.

Results

Model performance and validation. Model performance, both validation and testing, of predicted NGS-NEE is illustrated in Fig. 1. Model validation and testing performance metrics varied slightly with each new training cycle. Averaged over $n = 1000$ model training cycles, the cross-validation results showed a mean RMSE = 0.09 and mean $r^2 = 0.60$, implying that the model explained most of the observed variability of NGS–NEE. Model testing results further implied good model performance across the testing data with RMSE = 0.08 and $r^2 = 0.65$.

Model dependencies. The effects of each of the seven predictor variables on the modeled NGS–NEE are shown in the partial dependence and individual conditional expectation plots of Fig. 2. In most simulations, the modeled NGS–NEE showed net positive responses to increases in wind speed, soil moisture content, as well as air and soil temperatures. Wind direction exhibited little to no net influence on modeled NGS–NEE and upwelling photosynthetic photon flux density (PPFD) a slightly negative effect. By contrast, net radiation above the canopy exhibited a strong negative effect on the predicted NGS–NEE. This negative control can be attributed to increased photosynthetic CO_2 uptake with increased net radiation.

Sensitivity analysis. The calculated SIs provide further insight into the relative influence of the seven predictor variables on NGS–NEE estimates (Fig. 3). The results implied that net radiation above the canopy has the greatest influence on modeled NGS–NEE with a mean SI of 0.98 (0.98–0.99, 95% confidence interval). Wind direction was the least sensitive variable, with a mean SI of 0.08, that is, only slightly larger than that for the dummy (control) variable, which had a mean SI of 0.03. Soil temperature exerted a pronounced influence on modeled NGS–NEE with a mean SI of 0.72, consistent with soil temperature as a major control on subsurface organic C mineralization and, hence, production of CO_2. Wind speed and soil moisture content were of moderate to high importance (mean SIs = 0.62 and 0.61, respectively) and a lower influence of upwelling PPFD. Table 1 summarizes the mean values and ranges of all the SIs.

Climate change predictions. Under all three emission scenarios, NGS–NEE CO_2 fluxes were predicted to increase over the remainder of the 21st century (Fig. 4). That is, the Mer Bleue Bog will act as a stronger source of CO_2 during the NGS, even under the lowest radiative forcing scenario (i.e., RCP2.6). As also shown in Fig. 4, the predicted future NGS–NEE trends for the three climate scenarios diverged most significantly after 2050.

Under the high radiative forcing scenario considered (RCP8.5), the mean NGS–NEE CO_2 emission rates at the end of the century would rise to 0.62 μmol m^{-2} s^{-1} (0.60–0.63; minimum and maximum values from $n = 1000$ runs of the model). Compared to the observed mean NGS–NEE of 0.31 μmol m^{-2} s^{-1} for the period 1998–2010, NGS–NEE under RCP8.5, therefore, exhibited roughly a 2-fold increase by 2100 (mean increase of 103%). If the stringent emission reduction measures and policy shifts are implemented to limit radiative forcing to 2.6 W m^{-2} by the year 2100 (RCP2.6), NGS–NEE values at the Mer Bleue Bog site would only increase by 17% (approximately 6.2 times less than under
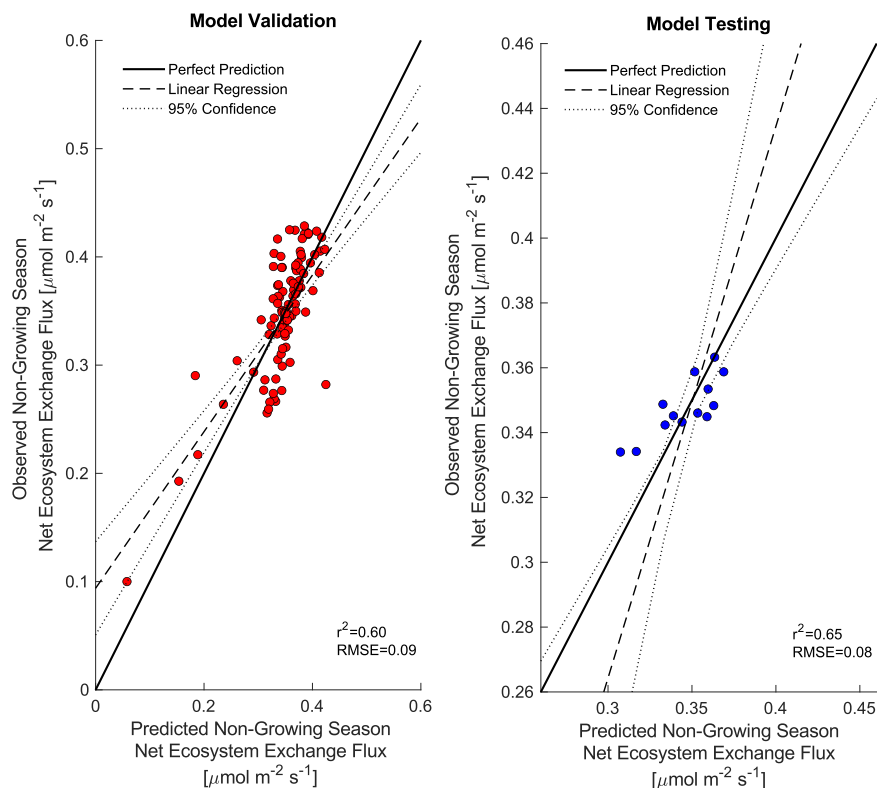
Fig. 1 Model validation and testing performance on predicted nongrowing season net ecosystem exchanges of CO$_2$ (NGS–NEE). Circles represent predicted vs. observed NGS–NEE fluxes. The solid black lines represent the 1:1 (perfect) agreement; dashed lines are linear regressions with dotted lines enclosing 95% confidence intervals.

RCP8.5). Under the stabilization climate scenario of RCP4.5, NGS would increase by 48% to 0.45 μmol m^{-2} s^{-1}. Table 2 summarizes both mid-century (2050) and end of the century (2100) predicted NGS–NEE CO$_2$ fluxes under the three RCPs.

Discussion

Our results for the Mer Bleue Bog site support the hypothesis that environmental changes accompanying climate warming have the potential to increase CO$_2$ emissions from peatlands during the NGS, hence potentially strengthening a positive climate feedback loop. As expected, the predicted NGS–NEE increase is highest for the climate scenario with the largest radiative forcing and vice versa. The predictor variables for NGS–NEE further implies that both subsurface and surface processes play important roles in modulating the net CO$_2$ fluxes during the NGS. The subsurface controls are likely linked to the decomposition of soil organic matter, which is known to strongly depend on soil temperature and moisture. Aboveground turbulent transport of CO$_2$ and energy balances, in turn, depend on meteorological conditions and snow cover dynamics[43,44]. In the following discussion special attention is given to the influence of snow on both surface and subsurface processes modulating NGS–NEE of CO$_2$.

Subsurface processes. According to the GSA, at the Mer Bleue Bog site soil temperature is the second most sensitive predictor variable of NGS–NEE (Fig. 3), with a positive influence on the net emission of CO$_2$ (Fig. 2). This positive effect is consistent with the positive temperature dependence of soil microbial respiration, even when soil temperatures approach or drop below zero during the winter[18,45,46]. In the climate projections considered in this study, soil temperature changes were assumed to be controlled by the predicted changes in air temperature, with an imposed 1 °C

attenuation (see "Climate projections"). The actual relationship between air and soil temperatures during the NGS may be far more complex, however, largely due to the influence of snow coverage on the belowground thermal regime[47,48]. Even at the shallow depth of 20 cm, Mer Bleue Bog peat rarely freezes despite mostly subzero winter temperatures (Supplementary Fig. S1). This can be explained by the insulating effect of the snowpack, with peak annual snow depths ranging from 30 to 120 cm for the 1998–2010 period. With the projected decreases in snow depth and fractional snow-cover for the Ottawa region[9], the more exposed soils could experience greater heat loss and, consequently, cooler peat temperatures. More detailed future projections of the changes in soil temperature will have to take into account the possible confounding effect of reduced snow cover.

Snow depth and coverage also alter the mechanisms and rates of soil-atmosphere gas exchanges, with variable contributions of diffusive and non-diffusion transport[49]. A lower fraction of snow-covered soil diminishes the gas transport barrier and enhances the release of CO$_2$ while, at the same time, facilitating the influx of molecular oxygen (O$_2$) which fuels the production of soil CO$_2$ by aerobic respiration. However, surface heterogeneities, including hummock and hollow peat microforms, may cause differential snow accumulation patterns that influence the overall effect on NGS–NEE of a reduction in snow cover. Such small-scale processes are a source of uncertainty not taken into account in our analysis.

Climate warming may also increase in the frequency of freeze–thaw events, especially at the start and near the end of the NGS. Many recent studies have linked freeze–thaw cycles to variations in CO$_2$ emissions[50–53]. Recurring freezing and thawing alter the physical and biological processes that contribute to winter soil respiration[54–56]. In addition, freezing causes ice layers to form in the snowpack and in the near-surface peat resulting in
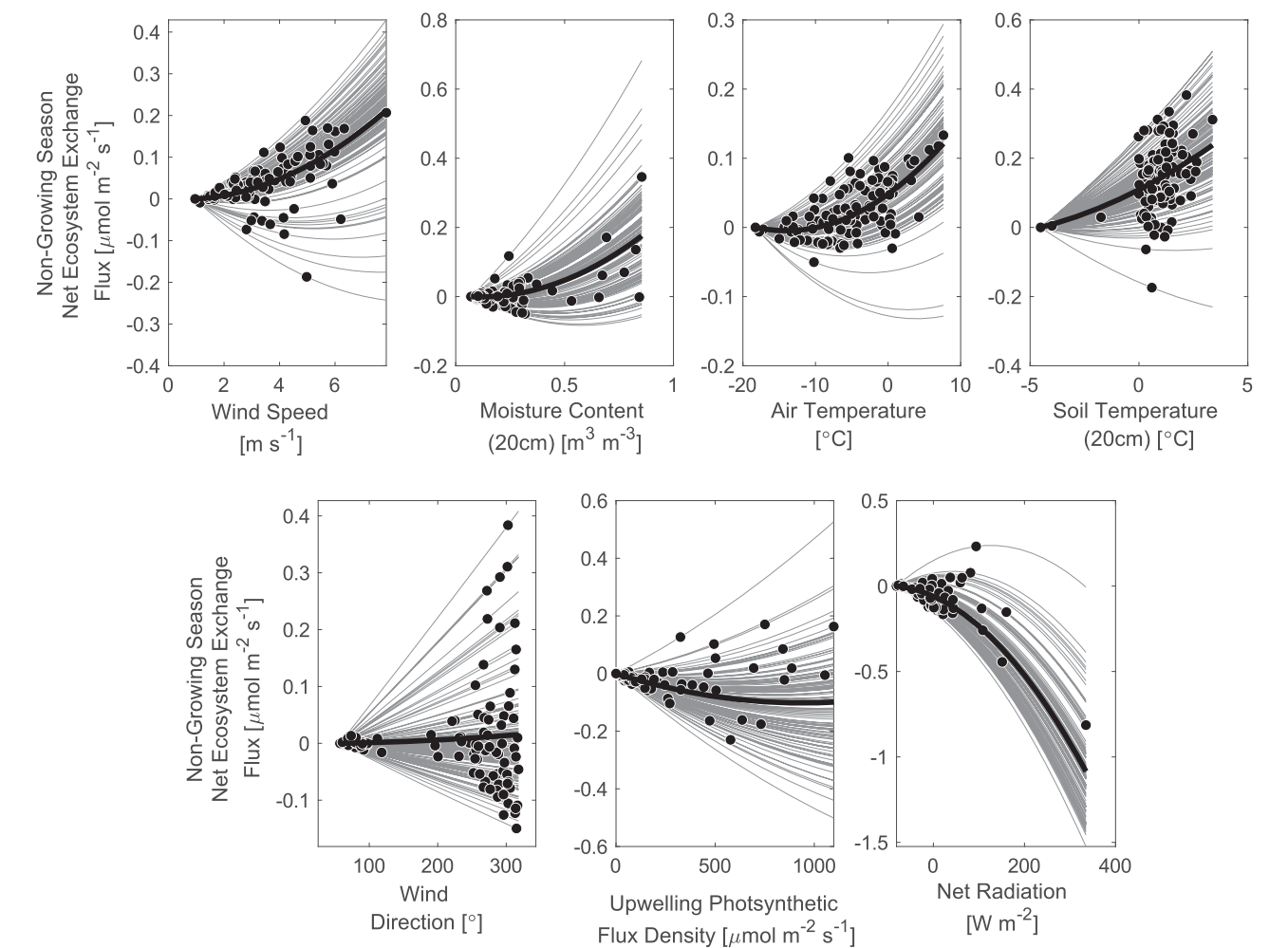
Fig. 2 Partial dependence and individual conditional expectations for modeled nongrowing season net ecosystem exchanges of CO$_2$ (NGS–NEE). The net dependence of the modeled NGS–NEE (y-axis, μmol m^{-2} s^{-1}) on a given predictor (x-axis) is shown as a solid black line. Circles represent individual observations for each predictor variable. Each grey line represents a single model simulation and displays the influence of the predictor on the predicted response.
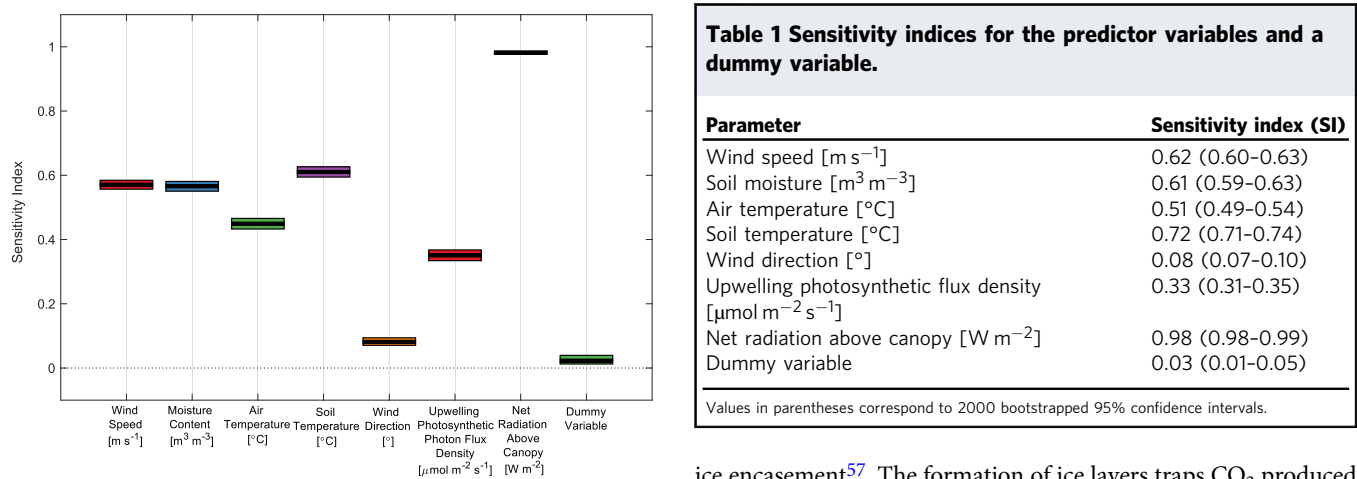

Fig. 3 PAWN global sensitivity analysis. The calculated sensitivity indices (SI) are moment-independent and computed based on the CDF of modeled nongrowing season net ecosystem exchanges of CO$_2$ (NGS–NEE). For each variable, the solid black horizontal line represents the mean SI value with the rectangular boxes surrounding this solid line representing a 2000 bootstrapped 95% confidence interval on the SI.

Table 1 Sensitivity indices for the predictor variables and a dummy variable.

Parameter	Sensitivity index (SI)
Wind speed [m s^{-1}]	0.62 (0.60–0.63)
Soil moisture [m^3 m^{-3}]	0.61 (0.59–0.63)
Air temperature [°C]	0.51 (0.49–0.54)
Soil temperature [°C]	0.72 (0.71–0.74)
Wind direction [°]	0.08 (0.07–0.10)
Upwelling photosynthetic flux density [μmol m^{-2} s^{-1}]	0.33 (0.31–0.35)
Net radiation above canopy [W m^{-2}]	0.98 (0.98–0.99)
Dummy variable	0.03 (0.01–0.05)

Values in parentheses correspond to 2000 bootstrapped 95% confidence intervals.

ice encasement[57]. The formation of ice layers traps CO$_2$ produced or stored in the peat and snowpack while thawing of the ice results in the release of the trapped CO$_2$ to the atmosphere.

Besides soil temperature, the modeled NGS–NEE CO$_2$ fluxes are sensitive to soil moisture recorded at 20 cm depth (Fig. 3). At this depth, NGS soil moisture in the Mer Bleue Bog remains quite low (typically less than 0.15 m^3 m^{-3}, Supplementary Fig. S1)

because of the relatively good drainage of the surface peat. Values of ~0.04 m³ m⁻³ when peat at 20 cm occasionally freezes indicating a continued availability of unfrozen water during winter, likely as thin films surrounding organic soil particles[58]. The positive and nonlinear correlation between soil moisture content and NGS–NEE at Mer Bleue (Fig. 2) agrees with similar positive relationships reported by Hirano[24], Liptzin et al.[25], and Schindlbacher et al.[59].

In our GSA and ML modeling, we rely on the soil temperature and moisture content data recorded at 20 cm depth. This is justified because most decomposition of plant debris (and hence CO_2 production) in peatlands occurs within the upper 10–20 cm[60]. In addition, the temporal records of both soil properties exhibited sufficient variability at 20 cm depth to train the ML model. At the Mer Bleue Bog site, temperature and moisture are higher deeper in the soil profile by the end of the NGS. However, the lack of fresh plant residues and O_2 at these depths limit microbial mineralization[61,62] and, thus, variations in temperature or moisture below 20 cm are unlikely to have a strong impact on the NGS–NEE CO_2 fluxes.

Because of its mid-latitude location (45°24′N), the Mer Bleue Bog site will continue to experience a high frequency of near freeze-thaw cycles over the remainder of the 21st century[63]. The implications of changes in freeze–thaw frequency and timing for the instantaneous and cumulative CO_2 emissions during the NGS, at this site and across peatlands in general, remain to be fully understood. Similarly, the length of time during winter when the soil is continuously frozen probably affects the NGS–NEE, in line with Humphreys et al.[64] who showed smaller winter CO_2 emissions for peatlands further north in the Hudson Bay Lowlands compared to the Mer Bleue Bog. A fully predictive understanding of CO_2 production and effluxes during the NGS will require additional characterization of the temporal variations in physical and biogeochemical properties and processes of the snowpack and underlying peat.

Surface processes. The most sensitive predictor variable for NGS–NEE is net radiation measured above the canopy (Fig. 3). With increasing net radiation, the modeled net CO_2 emissions decrease and NEE eventually switches to net CO_2 uptake. Net radiation in the NGS, particularly in early spring (February–March), strongly depends on the ground snow coverage. Upwelling PPFD also exerts a negative effect on NGS–NEE. With wavelengths between 400 and 700 nm, the PPFD encompasses part of the shortwave radiation spectrum and provides energy to support photosynthetic CO_2 fixation. Similar to net radiation, upwelling PPFD at Mer Bleue Bog is greatest in early spring as days get longer and a large fraction of incoming radiation is reflected from the snow surface.

The strong negative correlation between NGS–NEE and net radiation and, to a lesser extent, upwelling PPFD (Fig. 2) may be the result of patchy snow accumulation and melting, which exposes *Sphagnum*-covered hummocks to light thus enabling photosynthesis[65]. *Sphagnum* growth has been shown to depend on various environmental factors related to snow coverage, including the timing of snow onset and retreat, the amount of snow, mid-winter thawing, and drifting snow conditions[66,67]. Nonnegligible primary productivity by bryophytes may thus have important, but largely unexplored, consequences for estimating peatland NGS–NEE of CO_2.

With projected decreases in snow-water equivalents and snow-cover fractions at Mer Bleue Bog over the 21st century[9], interesting implications arise for advective sensible heat transfer between bare soils and snow patches[68,69]. Rain-on-snow events may similarly have significant impacts on NGS–NEE. With a greater proportion of NGS precipitation falling as rain[70–72], rain-on-snow events can result in wetter soils and reduced or removed snow coverage. Under these conditions, photosynthetic activity during the NGS may play an increasingly important role in the annual C budget of the Mer Bleue Bog and, by extension, other temperate and boreal peatlands[73,74].

The final sensitive predictor variable is wind speed which correlates positively with NGS–NEE (Figs. 2 and 3). A positive relationship is expected as turbulent exchanges of CO_2, heat, and moisture are driven by forced convection during the NGS when the surface is snow-covered and the atmosphere remains neutral or stable. In addition, wind-induced ventilation of the snowpack can cause pulse emissions of CO_2 stored in the snow and the porous and unsaturated near-surface peat[73,74]. Turbulent exchanges of heat and moisture play important roles in governing the snowpack energy balance[75,76]. Wind can, therefore, be indirectly linked to subsurface CO_2 production processes through its effects on energy fluxes into the ground.
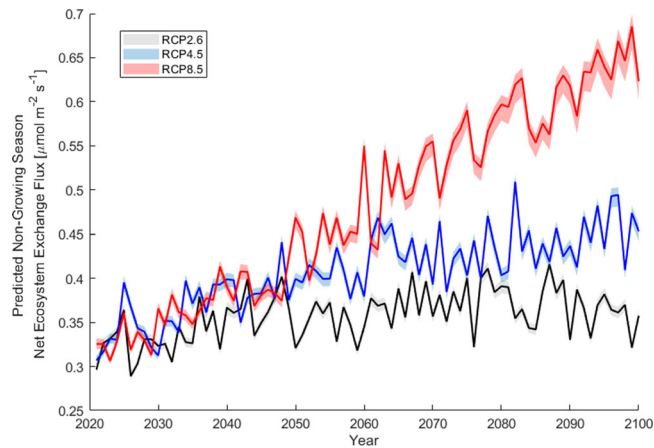

Fig. 4 Future 30-min averaged values of nongrowing season net ecosystem exchanges of CO₂ (NGS–NEE) under RCP2.6, RCP4.5, and RCP8.5. Solid lines represent mean NGS–NEE and shaded regions bound minimum and maximum predicted NEE over $n = 1000$ model runs.

Table 2 Predicted mean nongrowing season net ecosystem exchange fluxes in years 2050 and 2100.

Case	Mean NGS–NEE (year 2050) [μmol m⁻² s⁻¹]	% increase (year 2050)[a]	Mean NGS-NEE (year 2100) [μmol m⁻² s⁻¹]	% increase (year 2100)[a]
RCP 2.6	0.32 (0.32–0.33)	5.0 (3.1–6.5)	0.36 (0.35–0.36)	17 (14–19)
RCP 4.5	0.40 (0.39–0.41)	30 (28–33)	0.45 (0.44–0.46)	48 (44–50)
RCP 8.5	0.47 (0.46–0.48)	52 (49–56)	0.62 (0.60–0.63)	103 (96–107)

[a]Percentage increase relative to mean nongrowing season net ecosystem exchanges (NGS–NEE) for the period 1998–2010.
Values in parentheses represent the minimum and maximum values from 1000 runs of the model.

Some cautionary notes. As for ML in general, the development of the NGS–NEE model was only possible because the dataset for the Mer Bleue Bog site was large and diverse enough to enable the model learning process. Applying the same methodology to other locations is thus contingent on access to sufficient data records. In addition, a limitation of the ML approach is that only predictor variables included in the available dataset can be selected. Important variables not explicitly represented in the dataset may therefore remain hidden. Such variables may include, for example, the structure and pH of the peat, groundwater flow rates and pathways, and the timing and the frequency of rain-on-snow events.

Furthermore, the inferred sensitivity ranking and effects of the predictor variables, and the projected future changes in NGS–NEE, are site-specific and cannot be automatically extrapolated to other locations. Rather, the Mer Bleue Bog provides a reference site against which differences in the importance of surface and subsurface processes across peatlands can be hypothesized to lead to differences in NGS–NEE CO_2 fluxes and their sensitivity to environmental change. For instance, compared to the Mer Bleue Bog, extensive permafrost at higher latitudes results in unique thermal and hydrologic regimes regulating CO_2 exchanges of peatland ecosystems[64,77,78]. For these peatlands, the future trajectories of NGS–NEE will undoubtedly be impacted by permafrost thaw due to climate warming[79–81].

The results of the sensitivity analysis and ML modeling are also affected by how the NGS is defined. Our results actually point to non-negligible photosynthetic CO_2 fixation during the NGS, in order to explain the role of net radiation on the observed NGS–NEE. We speculate this may primarily reflect the onset of *Sphagnum* growth in early spring when snow cover is still present. To address the ambiguities related to the definition of the NGS, research should focus on a finer-resolution (say, monthly) analysis of the variations in NGS-NEE from late fall to early spring. The more detailed temporal trends of NGS–NEE should in turn yield more robust estimates of the contributions of the winter and shoulder seasons to the annual NEE of peatlands.

The projected future trajectories of NGS–NEE CO_2 emissions are based on estimating how the seven predictor variables will change under a set of radiative forcing scenarios. Future NGS–NEE trajectories will also be affected by other environmental drivers, including, for instance, future shifts in vegetation and fauna (including invasive species), and changes in land use and water management. Continued global warming may also increase the absorption of CO_2 by peatlands during the growing season while decreasing the length of the NGS. Together, the changing NEE of growing and non-growing seasons will ultimately determine the evolving status of a given peatland as either a net CO_2 sink or source.

Conclusions

The NGS–NEE CO_2 fluxes at Mer Bleue Bog over the period 1998–2010 can be reproduced by taking into account seven environmental variables: near-surface soil temperature and moisture, wind speed and direction, air temperature, net radiation above the canopy, and upwelling (i.e., reflected) PPFD. Of these seven predictor variables, net radiation and wind direction are the most and least influential, respectively. The significant effects of soil temperature and moisture are expected due to their roles in the subsurface production of CO_2. The effects of net radiation and upwelling PPFD is attributed to photosynthetic CO_2 uptake by *Sphagnum* mosses during the NGS, while wind speed controls the release and aboveground transport of soil CO_2. In turn, most of the predictor variables are themselves influenced by variations in the spatial distribution and depth of the snow cover. The mean

NGS–NEE CO_2 fluxes at Mer Bleue Bog site are projected to increase during the 2021–2100 period, reaching values by the end of the 21st century that is 17%, 48%, and 103% higher under the RCP2.6 (low), RCP4.5 (medium), and RCP8.5 (high) radiative forcing scenarios, respectively. Thus, in a warmer world, the Mer Bleue Bog site will act as a stronger source of CO_2 during the NGS.

Site description and methods

Site description. The Mer Bleue research site is a low shrub domed ombrotrophic bog located within a 2800 hectare wetland complex in Ottawa, Canada (45°24′N, 75°30′W)[64]. Originally part of the Peatland Carbon Simulator Project[82], the site has been continuously monitored for CO_2, energy (latent and sensible heat), radiative (long and shortwave radiation), and momentum fluxes since the construction of an EC tower in 1998. Mer Bleue has been part of a number of flux tower networks including the Fluxnet-Canada Research Network, the Canadian Carbon Program, Ameriflux, and FLUXNET.

The Mer Bleue site has a mean annual air temperature of 6.0 °C. Seasonal air temperatures between November to April and May to August of −3.7 and 18 °C, respectively (climate normal)[64]. Annual precipitation at Mer Bleue is 943.5 mm with 341.7 mm of this precipitation falling between May and August[64]. Soil temperatures below the surface are rarely frozen at Mer Bleue with mean NGS temperatures (1998–2010) of 0.27 °C at a 20 cm depth (see Supplementary Fig. S1). Vegetation at the Mer Bleue research site is comprised of a near-continuous cover of *Sphagnum capillifolium* and *Sphagnum magellanicum* with an overstory dominated by ericaceous shrubs including *Chamaedaphne calyculata*, *Kalmia angustifolia*, *Rhododendron groenlandicum*, and *Vaccinium myrtilloides*, along with some sedges (*Eriophorum vaginatum*) and herbs (*Maianthemum trifolium*)[83]. The area underwent bog formation starting approximately 7000 years ago with current peat depths ranging between <0.3 m at the margins to >5 m in the center of the bog[84]. The research site area has a typical hummock-hollow (70–30%) microtopography with a mean elevation difference of 0.25 m between hummock-tops and hollow-bottoms[85] with few scattered tree species (*Larix laricina*, *Betula populifolia*, and *Picea mariana*)[64].

Data acquisition. Fluxes and ancillary measurements of micrometeorological parameters recorded at the Mer Bleue site were obtained from the Fluxnet Canada Research Network via the Oak Ridge National Laboratory's Distributed Active Archive Center for Biogeochemical Dynamics (ORNL DAAC) (https://daac.ornl.gov). The 13-year dataset (1998–2010) was retrieved and filtered to only extract NGS-specific data. For this study, we define NGS as the period starting with the first day of the first 3 consecutive days with ground snow coverage and ending with the first day of the first three consecutive days with bare ground. The corresponding dates along with the associated NGS durations are provided in Supplementary Table S1.

Values of NEE were derived from CO_2 fluxes measured by EC with a three-dimensional sonic anemometer thermometer (model 1012R3 prior to September 1, 2000, and model R3-50 thereafter; Gill Instruments Ltd., Lymington, UK) and closed-path H_2O/CO_2 gas analyzer (model LI-6252 until September 1, 2000, LI-6262 until January 1, 2004, and LI-7000 thereafter; LI-COR Inc., Lincoln, NE) located 3 m above the bog surface[84]. High-frequency data (10 Hz prior to 2004, 20 Hz thereafter) were used to average fluxes over a 30-min period. Details on the flux calculation are given by Roulet et al.[84] they involved accounting

for air density fluctuations using either the WPL procedure[86] or high-frequency calculation of CO_2 and H_2O mixing ratios.

Although no spectral corrections were applied, a correction factor of 1.25 was applied to the CO_2 fluxes prior to 2004 to account for the changes in the EC system in 2004, which reduced high-frequency flux attenuation[84]. Thus, this CO_2 flux dataset retains some bias due to high and low-pass filtering effects. A subset of the flux data collected since 2004 was reprocessed using LI-COR's EddyPro v7.0.6 software and demonstrated that this underestimation is approximately 10%. Specifically, fluxes with analytic corrections of high-pass[87] and low-pass[88] filtering effects were 1.11 times greater than fluxes without spectral corrections (least-squares linear regression, intercept = 0.06 μmol m^{-2} s^{-1}, r^2 = 0.994, root mean square error (RMSE) = 0.31 μmol m^{-2} s^{-1}). Fluxes with analytic corrections of high-pass filtering[87], in-situ corrections of low-pass filtering[89], and corrections for instrument separation (crosswind and vertical)[90] were 1.08 times greater than CO_2 fluxes computed with no spectral corrections (intercept = 0.01 μmol m^{-2} s^{-1}, r^2 = 0.996, RMSE = 0.23 μmol m^{-2} s^{-1}). The analysis cut-off frequency used in Fratini et al.[89] was 1.28 Hz for CO_2 and between 0.33 and 0.70 Hz for H_2O. Further details of uncertainties associated with EC data acquisition are presented in Massman and Lee[91], including unreliable spectral nighttime corrections, and 2D and 3D advective effects.

At any given time, NEE is the sum of the turbulent CO_2 fluxes and the rate of change in CO_2 storage below the EC tower calculated from the concentrations of CO_2 measured by the EC instrumentation. We define positive NEE as the net release of CO_2 from the ecosystem while negative NEE represents net CO_2 uptake by the ecosystem. Thus, NEE represents the difference between gross primary productivity and heterotrophic plus autotrophic respiration (together referred to as ecosystem respiration). Note, that although the primary productivity of vascular plants during the NGS is usually negligible, this may not be the case for peat-forming mosses (see "Surface processes"). NEE fluxes were removed from the dataset in case of instrument malfunction, statistics outside of acceptable limits, and when the nighttime friction velocity dropped below 0.1 m s^{-1}.

Accompanying measurements of weather conditions and soil temperature and moisture was collected at varying frequencies (from 5 s to 30 min) and averaged over 30-min intervals (for further details, see Roulet et al.[84], and Lafleur et al.[85]). The measured variables are listed in Supplementary Table S2 and include radiation variables (variables #4–13, 57, 68), relative humidity (#14–15), air temperature (#16–18, 48), soil temperature (#19–37), soil moisture and water table depth (#50–56, 46), precipitation (#44–45, 49), wind speed and direction (#38–43), and atmospheric pressure (#47). In addition to these ancillary variables, the dataset includes friction velocity (#60), sensible (#59), and latent heat fluxes (#58) also measured based on EC and the rate of change in CO_2 storage (#61) and CO_2 flux (#62) as described above.

Variable selection. The initial dataset containing 68 variables (67 predictor variables and 1 response variable) was subjected to a four-step variable selection methodology (Fig. 5). The methodology was designed to reduce the number of variables represented in the dataset, leaving only those that exert the greatest influence on the NGS–NEE of CO_2. Only these remaining predictor variables were then used in the creation of the ML model.

Correlated variables. As a first step, collinearities between variables were examined by computing Pearson's linear coefficients of correlation (ρ) between pairs of variables. The correlation matrix (M) of ρ for each pair of variables (x, y) was obtained using the

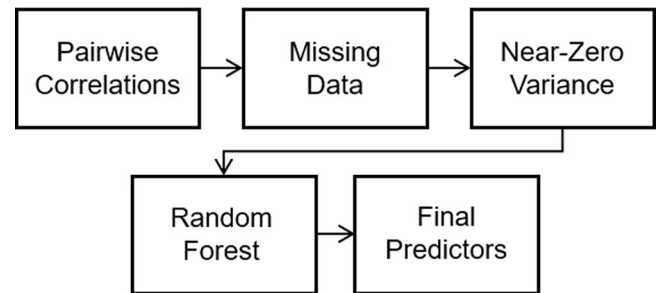## Flowchart Summarizing Variable Selection Methodology

Fig. 5 A flowchart summarizing the variable selection methodology. From 67 originally, predictor variables were selected down to 7: (1) soil temperature (at 20 cm depth) [°C], (2) soil moisture (at 20 cm depth) [m^3 m^{-3}], (3) air temperature [°C], (4) wind speed [m s^{-1}], (5) wind direction [°], (6) net radiation above canopy [W m^{-2}], and (7) upwelling (reflected) photosynthetic photon flux density (PPFD) [μmol m^{-2} s^{-1}].

following equations:

$$Cov = \sum_{i=1}^{N} \frac{(x_i - \bar{\bar{x}})(y_i - \bar{\bar{y}})}{N} \quad (1)$$

$$\rho = \frac{Cov(x, y)}{\sigma_x \sigma_y} \quad (2)$$

$$M = \begin{bmatrix} \rho(x, x) & \rho(x, y) \\ \rho(y, x) & \rho(y, y) \end{bmatrix} \quad (3)$$

where Cov represents the covariance between variables (x) and (y), $\bar{\bar{x}}$ and $\bar{\bar{y}}$ are the mean values of the variables (x) and (y), respectively, and σ is the standard deviation for each variable. Note that in M pairwise correlations between the same variables, i.e., $\rho(x, x)$ and $\rho(y, y)$, equal unity. Pairs of variables with $|\rho| \geq$ 0.75 were identified and labeled as correlated variable 1 (CV1) and correlated variable 2 (CV2). The average correlation between CV1 and the remaining 67 variables was then calculated. The process was repeated for CV2. The CV with the largest average correlation with the remaining 67 variables in the dataset was removed. Using this approach, 40 predictor variables were removed. Figure 6 provides a visual representation of variable collinearities prior to the application of the variable selection methodology. For variables in which the linear correlation coefficient could not be calculated due to missing observations or a standard deviation of zero, a Not a Number placeholder was assigned.

Missing sata and near zero variance. In order to reduce biases in the development of the ML model, no gap-filling or data imputation methods were applied. Variables with more than 50% of observational values missing were removed from the dataset. Supplementary Figure S2 presents an overview of the data coverage over the data record used in this study. With this threshold, nine variables were eliminated. The remaining 19 variables were screened and removed if their variance approached zero. Variables with minimal to no variation in their values contribute minimally to the model learning process. They are also likely to be of lesser relevance in predicting future NEE than variables prone to change[92]. Variables were therefore discarded if the fraction of their unique values was low, and the ratio of the mode to the second most common value (defined as the frequency factor) was high[93]. Specifically, in this study, predictors with <10% of their total observed values being unique and a frequency
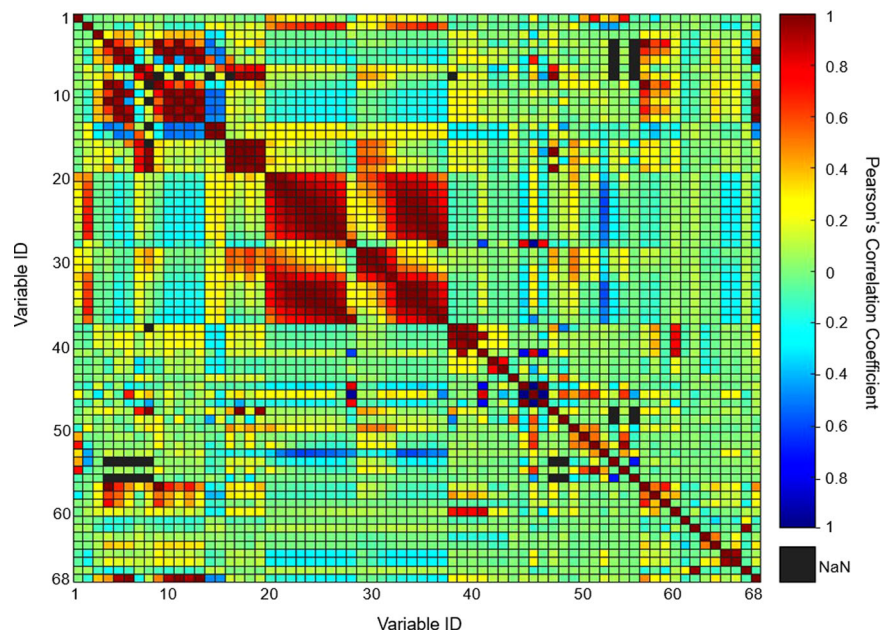
Fig. 6 Calculated Pearson's correlation coefficients (ρ) showing pairwise collinearities between 68 variables. Variables were labeled from 1 to 68 and aligned along each axis (see Supplementary Table S2), with variable 67 representing the non-growing season net ecosystem exchanges of CO_2 (NGS–NEE).

factor greater than 20 were removed, as proposed by Kuhn and Johnson[93]. Using these criteria, 1 predictor was removed leaving 18 remaining predictors.

Random forest permutation. A Random Forest Ensemble Model (RFM) was used to rank the importance of each of the variables remaining after the selection procedure. Random forest regression is a supervised learning algorithm that considers an ensemble learning method and bootstrap aggregation. This algorithm uses combinations of trained decision trees to obtain a predicted output. The reader is directed to Breiman[94] for a detailed discussion of random forests. In this study, the RFM consisted of an out-of-bag variable importance estimation algorithm modified for regression analysis according to Breiman[94]. The RFM was trained on the observations for each of the remaining variables (x_j) using 500 learner trees. Surrogate splits helped reduce biases due to the presence of missing observations. To further minimize bias, an interaction–curvature test was applied during the training process such that tree splits were made on predictors that minimize p-values from pairwise chi-squared (χ^2) tests of independence between the predictor and the response[95].

To estimate the importance of each variable, predictions based on the data used for the training of the RFM were compared to observations not used in the creation of the trees (i.e., the out-of-bag samples). The mean-square error (MSE) between predictions using the training data and the out-of-bag samples was determined and referred to as the out-of-bag error (ε_{oob}). The observations of each variable (x_j) were randomly permuted and a prediction .. obtained. The model error for each observation (ε_{mj}) was calculated by comparing y_j to the out-of-bag samples. The mean difference Δ_m between ε_{oob} and ε_{mj}, as well as the standard deviation (σ_j) of the differences in permuted values overall learners, were calculated. The predictor importance was then expressed as

$$I_j = \frac{\Delta_m}{\sigma_j} \qquad (4)$$

where I_j measures the importance of each predictor j: a larger value of I_j represents greater importance. Based on the value of I_j, soil temperature and moisture (at 20 cm depth below the hummock surface), wind direction and speed, air temperature, net radiation (above the canopy), and the upwelling PPFD were identified as the final predictor variables for use in developing the NGS–NEE ML model. Values of I_j are presented in Supplementary Table S3 for the reader reference.

NGS–NEE model development. Further data preprocessing standardized the values of each of the final predictor variables between 0 and 1. A K-means clustering algorithm was then applied to $n = 15,267$ observations to collapse and group the dataset into 122 representative centroids. The K-means clustering algorithm is an iterative unsupervised learning algorithm that aims to assign individual observations into K distinct clusters to minimize the intra-cluster sum of squares[96]. The centroids of these clusters contain the statistical information of individual clusters thus allowing for increased training speeds by reducing the number of data points[97]. The number of clusters used in this analysis was determined by applying the K-means clustering algorithm for increasing values of K ($K \in \mathbb{Z}^+, 1 \le K \le \lceil \sqrt{n} \rceil$) until 99% of the variance in inter-cluster Euclidean distances between data points and individual clusters was explained. The reader is directed to Telgarsky and Vattani[98] for a detailed description of the K-means clustering algorithm.

A wide range of regression algorithms (linear, tree-based, support vector, Gaussian process-based, and ensembles) was trained and validated using a tenfold cross-validation strategy on 85% of the entire dataset (training dataset). The tenfold cross-validation strategy randomly partitioned the data into ten segments of roughly equal size, and systematically trained data on nine of the ten partitioned folds while using the remaining tenth segment to estimate model performance. The process was repeated until all ten of the segments had been used to evaluate model performance. The 10-fold cross-validation strategy was run 1000 times, with each run consisting of a newly partitioned training (85%) and testing dataset (15%).

Model performance was evaluated based on the root-mean-square error (RMSE) and coefficient of determination (r^2) averaged over the 1000 runs of the 10-fold cross-validation

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (5)$$

$$r^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{\bar{y}})^2} \qquad (6)$$

where $\sum_{i=1}^{n}(x_i - y_i)^2$ represents the sum of squared errors between the predicted (x_i) and observed NEE (y_i) and $\bar{\bar{y}}$ represents the mean observed NEE flux over the entire data record.

Model testing using the remaining 15% of the dataset (testing dataset) was evaluated using the mean r^2 and RMSE values over 1000 runs, with each run consisting of a uniquely partitioned testing and training dataset. Comparing the computed RMSE and r^2 values for all trained models, an epsilon insensitive support vector machine regression (ε-SVM) model was selected because it had the lowest RMSE and highest r^2 values. An additional advantage of a ε-SVM model is its relatively simple structure and implementation. The performance of other regression algorithms used in this analysis for comparison is provided in Supplementary Fig. S3.

Support vector machines comprise a set of Kernel-based ML algorithms for classification and regression analyses. SVMs generally operate through the nonlinear mapping of input vectors into a higher dimensional space and solving optimization problems in this space[99]. In the case of classification, an optimization problem is solved to correctly classify the greatest number of observations through maximizing the margin separating the hyperplane, which is controlled by the presence of support vectors[100]. For regression analyses, this concept is extended to include a region surrounding the plane referred to as the ε-insensitive region. The goal of the SVM regression is to solve a convex optimization problem that aims to minimize a defined ε-insensitive loss function while maximizing the flatness of the ε-insensitive region which contains many of the training observations. The reader is directed to Cortes and Vapnik[99] and Awad et al.[100] for a detailed mathematical discussion of SVMs.

With few model parameters, SVMs have been shown to mimic the performance of the best artificial neural networks[101] that have been used extensively in modeling CO_2 fluxes[37,38]. Moreover, the computational complexity of SVM regressions does not depend on the dimensionality of the input space making for an attractive choice for large datasets with many variables[100]. The trained ε-SVM uses a quadratic kernel, with hyperparameters determined based on the interquartile range of the observed NGS-NEE. Training data used in the creation of the ε-SVM model are provided in Supplementary Data 1.

Sensitivity analysis. A moment-independent GSA was conducted based on PAWN[102]. Unlike other density-based sensitivity analyses, PAWN uses empirically computed cumulative density functions (CDFs) rather than probability density functions to calculate sensitivity indices (SI). Given the non-normal distribution of the observed NGS–NEE (Supplementary Fig. S1), the variance of the NGS–NEE distribution would not adequately capture the model uncertainty[103]. Hence, PAWN offers a more natural choice for the calculation of SIs.

We implemented PAWN using the SAFE MATLAB Toolbox, an open-source toolbox developed by F. Pianosi, F. Sarrazin, and T. Wagener (https://www.safetoolbox.info/). Modifications to this toolbox were made to handle the nonparametric distributions unique to the underlying distribution of the variables used in this study (Supplementary Fig. S1). SI was calculated as the distance between the conditional and unconditional CDFs. Here, this distance was approximated using the Kolmogorov-Smirnov statistic. Unconditional CDFs were calculated by varying all input variables simultaneously, while conditional CDFs were obtained through varying all input variables except one, which was allowed to vary within a specified range of values (conditioning intervals). Supplementary Figure S4 provides a visual representation of the empirically derived CDFs through a regional sensitivity analysis.

The GSA used $n = 8000$ samples obtained via Latin Hypercube Sampling in five conditioning intervals, yielding a 2000 boot-strapped 95% confidence interval. A dummy variable in the GSA served as a control against which the SIs of other variables were compared. SI values range from 0 to 1, with zero representing null effects (no sensitivity) and 1 representing the largest sensitivity to modeled NEE. For a more detailed description of PAWN, the reader is referred to Zadeh et al.[104], Pianosi and Wagener[102,105], and to Noacco et al.[106] for examples of the use of the SAFE toolbox.

Climate projections. Three future climate scenarios were considered: RCP2.6, RCP4.5, and RCP8.5. These scenarios describe the climate trajectories associated with stringent, moderate, and high degrees of radiative forcing, respectively[107]. Values of the most sensitive model variables (soil temperature and moisture, air temperature, wind speed and direction, upwelling PPFD, and net radiation above canopy; see "Sensitivity analysis") were altered according to their predicted changes under each RCP. These predicted changes were obtained from either previously conducted studies or outputs from global climate models and are described hereafter.

Winter air temperature projections for the Ottawa region under each RCP were obtained from statistically downscaled (10 km spatial resolution) predictions from 24 global climate models participating in the fifth phase of the Coupled Model Intercomparison Project (CMIP5). The obtained data were previously statistically downscaled using the Bias Correction/Constructed Analogs with Quantile mapping version 2 (BCCAQv2) algorithm[108,109]. Changes in soil temperatures were assumed to be 1 °C lower than the projected changes in air temperature based on the work of Zhang et al.[110]. and Wisser et al.[111]. Winter wind speeds and directions were kept constant for all RCPs as predictions by Jeong and Sushama[112] and McInnes et al.[113] showed little to no increase in wind speeds in the Ottawa region, in contrast to the predicted increases in regions at higher latitudes in Canada.

Predicted changes for upwelling PPFD and soil moisture were obtained as outputs from the Second Generation Canadian Earth System Model (CanESM2). Predicted reductions in upwelling shortwave radiation were used as a proxy for reductions in upwelling PPFD, given that photosynthetically active radiation falls partly in the shortwave radiation spectrum. It is recognized that changes in upwelling radiation are difficult to predict due to complex interactions between changing snow-cover fractions[9], sea ice-snow albedos[9,114], and aerosol concentrations[115].

Changes to net radiation were applied linearly between 2021 and 2100. The net change by the year 2100 corresponded directly to the radiative forcing associated with each RCP (+2.6, +4.5, +8.5 W/m^2 for RCP2.6, 4.5, and 8.5, respectively). Mean NGS–NEE CO_2 effluxes were then calculated with the ML model using the predicted temporal trends of the seven predictor variables for the 2021–2100 period. Net changes to the model predictor variables under each RCP by the end of the 21st century are presented in Table 3.

Table 3 Climate projections: predicted changes in input variables by the end of the 21st century, according to the three radiative forcing scenarios, representative concentration pathways (RCPs) 2.6, 4.5, and 8.5.

Time frame	Scenario	Air temperature [°C]	Soil temperature [°C]	Net radiation [W m^{-2}]	Upwelling PPFD [μmol m^{-2} s^{-1}]	Moisture content [m^3 m^{-3}]	Wind speed [m s^{-1}]	Wind direction [°]
2021-2100	RCP2.6	+1.8	+0.9	−2.6	−15	−0.02	–	–
	RCP4.5	+3.8	+2.8	−4.5	−58	−0.02	–	–
	RCP8.5	+6.8	+5.8	−8.5	−80	−0.07	–	–

The changes listed are relative to the corresponding 1986–2005 mean values.

Data availability

Code availability

References

1. Maltby, E. & Immirzi, P. Carbon dynamics in peatlands and other wetland soils regional and global perspectives. *Chemosphere* **27**, 999–1023 (1993).
2. Gorham, E. Northern peatlands: role in the carbon cycle and probable responses to climate warming. *Ecol. Appl.* **1**, 182–195 (1991).
3. Mitra, S., Wassmann, R. & Vlek, P. L. G. An appraisal of global wetland area and its organic carbon stock. *Curr. Sci.* **88**, 25–35 (2005).
4. Limpens, J. et al. Peatlands and the carbon cycle: from local processes to global implications—a synthesis. *Biogeosciences* **5**, 1475–1491 (2008).
5. Harenda, K. M., Lamentowicz, M., Samson, M. & Chojnicki, B. H. The role of peatlands and their carbon storage function in the context of climate change. in *GeoPlanet: Earth and Planetary Sciences*. 169–187 (Springer Verlag, 2018). https://doi.org/10.1007/978-3-319-71788-3_12.
6. Holland, M. M. & Bitz, C. M. Polar amplification of climate change in coupled models. *Clim. Dyn.* **21**, 221–232 (2003).
7. Serreze, M. C. & Barry, R. G. Processes and impacts of Arctic amplification: a research synthesis. *Glob. Planet. Change* **77**, 85–96 (2011).
8. Bintanja, R. & Van Der Linden, E. C. The changing seasonal climate in the Arctic. *Sci. Rep.* **3**, 1–8 (2013).
9. Derksen, C. et al. *Changes in Snow, Ice, and Permafrost Across Canada; Chapter 5 in Canada's Changing Climate Report*. (2018).
10. Campbell, J. L. Arctic loses carbon as winters wane. *Nat. Clim. Chang.* **9**, 806–807 (2019).
11. Koenigk, T. et al. Arctic climate change in 21st century CMIP5 simulations with EC-Earth. *Clim. Dyn.* **40**, 2719–2743 (2013).
12. Natali, S. M. et al. Large loss of CO$_2$ in winter observed across the northern permafrost region. *Nat. Clim. Chang.* **9**, 852–857 (2019).
13. Zimov, S. A. Winter biotic activity and production of CO$_2$ in Siberian soils: a factor in the greenhouse effect. *J. Geophys. Res.* **98**, 5017–5023 (1993).
14. Sommerfeld, R. A., Mosier, A. R. & Musselman, R. C. CO$_2$, CH$_4$ and N$_2$O flux through a Wyoming snowpack and implications for global budgets. *Nature* **361**, 140–142 (1993).
15. Sommerfeld, R. A., Massman, W. J., Musselman, R. C. & Mosier, A. R. Diffusional flux of CO$_2$ through snow: Spatial and temporal variability among alpine-subalpine sites. *Global Biogeochem. Cycles* **10**, 473–482 (1996).
16. Winston, G. C., Sundquist, E. T., Stephens, B. B. & Trumbore, S. E. Winter CO$_2$ fluxes in a boreal forest. *J. Geophys. Res. Atmos.* **102**, 28795–28804 (1997).
17. Fahnestock, J. T., Jones, M. H. & Welker, J. M. Wintertime CO$_2$ efflux from Arctic soils: implications for annual carbon budgets. *Global Biogeochem. Cycles* **13**, 775–779 (1999).
18. Panikov, N. S. & Dedysh, S. N. Cold season CH$_4$ And CO$_2$ emission from boreal peat bogs (West Siberia): winter fluxes and thaw activation dynamics. *Global Biogeochem. Cycles* **14**, 1071–1080 (2000).
19. Miao, Y. et al. Greenhouse gas emissions from different wetlands during the snow-covered season in Northeast China. *Atmos. Environ.* **62**, 328–335 (2012).
20. Webster, K. L. et al. Spatially-integrated estimates of net ecosystem exchange and methane fluxes from Canadian peatlands. *Carbon Balance Manag.* **13**, 16 (2018).
21. Wu, J. & Roulet, N. T. Climate change reduces the capacity of northern peatlands to absorb the atmospheric carbon dioxide: The different responses of bogs and fens. *Global Biogeochem. Cycles* **28**, 1005–1024 (2014).
22. Chaudhary, N. et al. Modelling past and future peatland carbon dynamics across the pan-Arctic. *Glob. Chang. Biol.* **26**, 4119–4133 (2020).
23. Qiu, C., Zhu, D., Ciais, P., Guenet, B. & Peng, S. The role of northern peatlands in the global carbon cycle for the 21st century. *Glob. Ecol. Biogeogr.* **29**, 956–973 (2020).
24. Hirano, T. Seasonal and diurnal variations in topsoil and subsoil respiration under snowpack in a temperate deciduous forest. *Global Biogeochem. Cycles* **19**, 1–10 (2005).
25. Liptzin, D. et al. Process-level controls on CO$_2$ fluxes from a seasonally snow-covered subalpine meadow soil, Niwot Ridge, Colorado. *Biogeochemistry* **95**, 151–166 (2009).
26. Schädel, C. et al. Potential carbon emissions dominated by carbon dioxide from thawed permafrost soils. *Nat. Clim. Chang.* **6**, 950–953 (2016).
27. Welker, J. M., Fahnestock, J. T. & Jones, M. H. Annual CO$_2$ flux in dry and moist Arctic tundra: Field responses to increases in summer temperatures and winter snow depth. *Clim. Change* **44**, 139–150 (2000).
28. Rogers, M. C., Sullivan, P. F. & Welker, J. M. Evidence of nonlinearity in the response of net ecosystem CO$_2$ exchange to increasing levels of Winter Snow Depth in the High Arctic of Northwest Greenland. *Arctic Antarct. Alp. Res.* **43**, 95–106 (2011).
29. Aanderud, Z. T., Jones, S. E., Schoolmaster, D. R., Fierer, N. & Lennon, J. T. Sensitivity of soil respiration and microbial communities to altered snowfall. *Soil Biol. Biochem.* **57**, 217–227 (2013).
30. Contosta, A. R., Burakowski, E. A., Varner, R. K. & Frey, S. D. Winter soil respiration in a humid temperate forest: the roles of moisture, temperature, and snowpack. *J. Geophys. Res. Biogeosci.* **121**, 3072–3088 (2016).
31. Grogan, P. Cold season respiration across a low arctic landscape: the influence of vegetation type, snow depth, and interannual climatic variation. *Arctic, Antarct. Alp. Res.* **44**, 446–456 (2012).
32. Michaelson, G. J. & Ping, C. L. Soil organic carbon and CO$_2$ respiration at subzero temperature in soils of Arctic Alaska. *J. Geophys. Res. Atmos.* **108**, 5–1 (2003). ALT.
33. Monson, R. K. et al. Winter forest soil respiration controlled by climate and microbial community composition. *Nature* **439**, 711–714 (2006).
34. Wang, T. et al. Controls on winter ecosystem respiration in temperate and boreal ecosystems. *Biogeosciences* **8**, 2009–2025 (2011).
35. Montáns, F. J., Chinesta, F., Gómez-Bombarelli, R. & Kutz, J. N. Data-driven modeling and learning in science and engineering. *C. R. Mec.* **347**, 845–855 (2019).
36. Cai, J., Xu, K., Zhu, Y., Hu, F. & Li, L. Prediction and analysis of net ecosystem carbon exchange based on gradient boosting regression and random forest. *Appl. Energy* **262**, 114566 (2020).
37. He, H., Yu, G., Zhang, L., Sun, X. & Su, W. Simulating CO$_2$ flux of three different ecosystems in ChinaFLUX based on artificial neural networks. *Sci. China Ser. D Earth Sci.* **49**, 252–261 (2006).
38. Jung, M. et al. Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach. *Biogeosciences* **17**, 1343–1365 (2020).
39. Melesse, A. M. & Hanley, R. S. Artificial neural network application for multi-ecosystem carbon flux simulation. *Ecol. Modell.* **189**, 305–314 (2005).
40. Dou, X., Yang, Y. & Luo, J. Estimating forest carbon fluxes using machine learning techniques based on eddy covariance measurements. *Sustainability* **10**, 203 (2018).

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

42. Xiao, J. et al. Estimation of net ecosystem carbon exchange for the conterminous United States by combining MODIS and AmeriFlux data. *Agric. For. Meteorol.* **148**, 1827–1847 (2008).

43. Dunne, T., Price, A. G. & Colbeck, S. C. The generation of runoff from subarctic snowpacks. *Water Resour. Res.* **12**, 677–685 (1976).

44. Diro, G. T., Sushama, L. & Huziy, O. Snow-atmosphere coupling and its impact on temperature variability and extremes over North America. *Clim. Dyn.* **50**, 2993–3007 (2018).

45. Panikov, N. S., Flanagan, P. W., Oechel, W. C., Mastepanov, M. A. & Christensen, T. R. Microbial activity in soils frozen to below −39 °C. *Soil Biol. Biochem.* **38**, 785–794 (2006).

46. Schaefer, K. & Jafarov, E. A parameterization of respiration in frozen soils based on substrate availability. *Biogeosciences* **13**, 1991–2001 (2016).

47. Gold, L. W. Influence of snow cover on the average annual ground temperature at Ottawa, Canada. *Int. Assoc. Sci. Hydrol.* **61**, 82–91 (1963).

48. Zhang, T. Influence of the seasonal snow cover on the ground thermal regime: an overview. *Rev. Geophys.* **43**, RG4002 (2005).

49. Jones, H. G., Pomeroy, J. W., Davies, T. D., Tranter, M. & Marsh, P. CO_2 in Arctic snow cover: landscape form, in-pack gas concentration gradients, and the implications for the estimation of gaseous fluxes. *Hydrol. Processes* **13**, 2977–2989 (1999).

50. Kurganova, I., Teepe, R. & Loftfield, N. Influence of freeze-thaw events on carbon dioxide emission from soils at different moisture and land use. *Carbon Balance Manag.* **2**, 2 (2007).

51. Matzner, E. & Borken, W. Do freeze-thaw events enhance C and N losses from soils of different ecosystems? A review. *Eur. J. Soil Sci.* **59**, 274–284 (2008).

52. Wang, J. et al. Effects of freezing-thawing cycle on peatland active organic carbon fractions and enzyme activities in the Da Xing'anling Mountains, Northeast China. *Environ. Earth Sci.* **72**, 1853–1860 (2014).

53. Yang, J., Zhou, W., Liu, J. & Hu, X. Dynamics of greenhouse gas formation in relation to freeze/thaw soil depth in a flooded peat marsh of Northeast China. *Soil Biol. Biochem.* **75**, 202–210 (2014).

54. Du, E. et al. Winter soil respiration during soil-freezing process in a boreal forest in Northeast China. *J. Plant Ecol.* **6**, 349–357 (2013).

55. Hayashi, M. The Cold Vadose Zone: hydrological and ecological significance of frozen-soil processes. *Vadose Zone J.* **12**, vzj2013.03.0064 (2013).

56. Pi, K. et al. The cold region critical zone in transition: responses to climate warming and land use change. *Annu. Rev. Environ. Resour.* https://doi.org/10.1146/annurev-environ-012220-125703 (2021).

57. Martz, F., Vuosku, J., Ovaskainen, A., Stark, S. & Rautio, P. The snow must go on: ground ice encasement, snow compaction and absence of snow differently cause soil hypoxia, CO_2 accumulation and tree seedling damage in boreal forest. *PLoS ONE* **11**, e0156620 (2016).

58. Romanovsky, V. E. & Osterkamp, T. E. Effects of unfrozen water on heat and mass transport processes in the active layer and permafrost. *Permafr. Periglac. Process* **11**, 219–239 (2000).

59. Schindlbacher, A., Zechmeister-Boltenstern, S., Glatzel, G. & Jandl, R. Winter soil respiration from an Austrian mountain forest. *Agric. For. Meteorol.* **146**, 205–215 (2007).

60. Clymo, R. The limits to peat bog growth. *Philos. Trans. R. Soc. Lond. Biol. Sci.* **303**, 605–654 (1984).

61. Blodau, C., Basiliko, N. & Moore, T. R. Carbon turnover in peatland mesocosms exposed to different water table levels. *Biogeochemistry* **67**, 331–351 (2004).

62. McCarter, C. P. R. et al. Pore-scale controls on hydrological and geochemical processes in peat: Implications on interacting processes. *Earth Sci. Rev.* **207**, 103227 (2020).

63. Prairie Climate Centre. *Climate Change in Canada.* https://climateatlas.ca/ (2019).

64. Humphreys, E., Charron, C., Brown, M. & Jones, R. Two bogs in the Canadian Hudson Bay Lowlands and a Temperate Bog reveal similar annual net ecosystem exchange of CO_2. *Arctic Antarct. Alp. Res.* **46**, 103–113 (2014).

65. Küttim, M., Küttim, L., Ilomets, M. & Laine, A. M. Controls of Sphagnum growth and the role of winter. *Ecol. Res.* **35**, 219–234 (2020).

66. Sonesson, M. Studies on Mire Vegetation in Tornetrask Area, Northern Sweden. 2. Winter conditions of poor mires. *Bot. Not.* **122**, 481 (1969).

67. Eurola, S. Snow and ground frost conditions of some Finnish mire types. *Ann. Bot. Fenn.* **12**, 1–16 (1975).

68. Shook, K. & Gray, D. M. Snowmelt resulting from advection. *Hydrol. Process.* **11**, 1725–1736 (1997).

69. Marsh, P. et al. Snowmelt processes and runoff at the arctic treeline: ten years of MAGS research. in *Cold Region Atmospheric and Hydrologic Studies. The Mackenzie GEWEX Experience.* vol. 2, 97–123 (Springer, Berlin, Heidelberg, 2008).

70. Screen, J. A. & Simmonds, I. Declining summer snowfall in the Arctic: causes, impacts and feedbacks. *Clim. Dyn.* **38**, 2243–2256 (2012).

71. Vincent, L. A. et al. Observed trends in Canada's climate and influence of low-frequency variability modes. *J. Clim.* **28**, 4545–4560 (2015).

72. Zhang, X. et al. Changes in temperature and precipitation across Canada. In *Canada's Changing Climate Report* (eds. Bush, E. & Lemmen, D. S.) Chapter 4, pp. 112–193 (Government of Canada, Ottawa, Ontario, 2019).

73. Lai, D. Y. F., Roulet, N. T., Humphreys, E. R., Moore, T. R. & Dalva, M. The effect of atmospheric turbulence and chamber deployment period on autochamber CO_2 and CH_4 flux measurements in an ombrotrophic peatland. *Biogeosciences* **9**, 3305–3322 (2012).

74. Graham, L. & Risk, D. Explaining CO_2 fluctuations observed in snowpacks. *Biogeosciences* **15**, 847–859 (2018).

75. Cline, D. W. Snow surface energy exchanges and snowmelt at a continental, midlatitude Alpine site. *Water Resour. Res.* **33**, 689–701 (1997).

76. Knox, S. H., Carey, S. K. & Humphreys, E. R. Snow surface energy exchanges and snowmelt in a shrub-covered bog in eastern Ontario, Canada. *Hydrol. Process.* **26**, 1876–1890 (2012).

77. Turetsky, M. R., Wieder, R. K. & Vitt, D. H. Boreal peatland C fluxes under varying permafrost regimes. *Soil Biol. Biochem.* **34**, 907–912 (2002).

78. O'Donnell, J. A. et al. The effects of permafrost thaw on soil hydrologic, thermal, and carbon dynamics in an Alaskan Peatland. *Ecosystems* **15**, 213–229 (2012).

79. Schuur, E. A. G. et al. The effect of permafrost thaw on old carbon release and net carbon exchange from tundra. *Nature* **459**, 556–559 (2009).

80. Mauritz, M. et al. Nonlinear CO_2 flux response to 7 years of experimentally induced permafrost thaw. *Glob. Chang. Biol.* **23**, 3646–3666 (2017).

81. Voigt, C. et al. Ecosystem carbon response of an Arctic peatland to simulated permafrost thaw. *Glob. Chang. Biol.* **25**, 1746–1764 (2019).

82. Frolking, S. et al. Modeling seasonal to annual carbon balance of Mer Bleue Bog, Ontario, Canada. *Global Biogeochem. Cycles* **16**, 4-1–4–21 (2002).

83. Bubier, J. L., Moore, T. R. & Crosby, G. Fine-scale vegetation distribution in a cool temperate peatland. *Can. J. Bot.* **84**, 910–923 (2006).

84. Roulet, N. T. et al. Contemporary carbon balance and late Holocene carbon accumulation in a northern peatland. *Glob. Chang. Biol.* **13**, 397–411 (2007).

85. Lafleur, P. M., Roulet, N. T., Bubier, J. L., Frolking, S. & Moore, T. R. Interannual variability in the peatland-atmosphere carbon dioxide exchange at an ombrotrophic bog. *Global Biogeochem. Cycles* **17**, 1036 (2003).

86. Webb, E. K., Pearman, G. I. & Leuning, R. Correction of flux measurements for density effects due to heat and water vapour transfer. *Q. J. R. Meteorol. Soc.* **106**, 85–100 (1980).

87. Moncrieff, J., Clement, R., Finnigan, J. & Meyers, T. Averaging, detrending, and filtering of eddy covariance time series. in *Handbook of Micrometeorology* 7–31 (Kluwer Academic Publishers, 2004). https://doi.org/10.1007/1-4020-2265-4_2.

88. Moncrieff, J. B. et al. A system to measure surface fluxes of momentum, sensible heat, water vapour and carbon dioxide. *J. Hydrol.* **188–189**, 589–611 (1997).

89. Fratini, G., Ibrom, A., Arriga, N., Burba, G. & Papale, D. Relative humidity effects on water vapour fluxes measured with closed-path eddy-covariance systems with short sampling lines. *Agric. For. Meteorol.* **165**, 53–63 (2012).

90. Horst, T. W. & Lenschow, D. H. Attenuation of scalar fluxes measured with spatially-displaced sensors. *Bound. Layer Meteorol.* **130**, 275–300 (2009).

91. Massman, W. J. & Lee, X. Eddy covariance flux corrections and uncertainties in long-term studies of carbon and energy exchanges. *Agric. For. Meteorol.* **113**, 121–144 (2002).

92. Dankers, F. J. W. M., Traverso, A., Wee, L. & van Kuijk, S. M. J. Prediction Modeling Methodology. in *Fundamentals of Clinical Data Science* 101–120 (Springer International Publishing, 2019). https://doi.org/10.1007/978-3-319-99713-1_8.

93. Kuhn, M. & Johnson, K. *Applied Predictive Modeling* https://doi.org/10.1007/978-1-4614-6849-3 (2013).

94. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).

95. Loh, W.-Y. *Regression Trees With Unbiased Variable Selection and Interaction Detection.* Stat. Sin. http://lib.stat.cmu (2002).

96. Hartigan, J. A. & Wong, M. A. Algorithm AS 136: A K-Means Clustering Algorithm. *Appl. Stat.* **28**, 100 (1979).

97. Zhifeng, H., Wen, W., Xiaowei, Y., Jie, L. & Guangquan, Z. A fast data preprocessing procedure for support vector regression. *Lect. Notes Comput. Sci.* **4224**, LNCS 48–LNCS 56 (2006).

98. Telgarsky, M. & Vattani, A. *Hartigan's Method: k-Means Clustering Without Voronoi.* (2010).

99. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995).

100. Awad, M., Khanna, R., Awad, M. & Khanna, R. Support vector regression. in *Efficient Learning Machines* 67–80 (Apress, 2015). https://doi.org/10.1007/978-1-4302-5990-9_4.

101. Hearst, M. A., Schölkopf, B., Dumais, S., Osuna, E. & Platt, J. Trends and controversies—support vector machines. *IEEE Intell. Syst.* **13**, 18–28 (1998).

102. Pianosi, F. & Wagener, T. A simple and efficient method for global sensitivity analysis based oncumulative distribution functions. *Environ. Model. Softw.* **67**, 1–11 (2015).

103. Saltelli, A. Sensitivity analysis for importance assessment. *Risk Anal.* **22**, 579–590 (2002).

104. Zadeh, F. et al. Comparison of the PAWN and Sobol' sensitivity analysis methods for a highly-parameterized hydrological model using SWAT. *E-proceedings 36th IAHR World Congr. 28 June–3 July, 2015, Hague, Netherlands.* 1–4 (2015).

105. Pianosi, F. & Wagener, T. Distribution-based sensitivity analysis from a generic input-output sample. *Software availability.* https://doi.org/10.1016/j.envsoft.2018.07.019 (2018).

106. Noacco, V., Sarrazin, F., Pianosi, F. & Wagener, T. Matlab/R workflows to assess critical choices in Global Sensitivity Analysis using the SAFE toolbox. *MethodsX* **6**, 2258–2280 (2019).

107. Moss, R. H. et al. The next generation of scenarios for climate change research and assessment. *Nature* **463**, 747–756 (2010).

108. Werner, A. T. & Cannon, A. J. Hydrologic extremes—an intercomparison of multiple gridded statistical downscaling methods. *Hydrol. Earth Syst. Sci.* **20**, 1483–1508 (2016).

109. Cannon, A. J., Sobie, S. R. & Murdock, T. Q. Bias correction of GCM precipitation by quantile mapping: how well do methods preserve changes in quantiles and extremes? *J. Clim.* **28**, 6938–6959 (2015).

110. Zhang, Y., Chen, W. & Riseborough, D. W. Modeling long-term dynamics of snow and their impacts on permafrost in Canada. in *Ninth international conference on permafrost, proceedings volume 2* (eds Kane, D. L. & Hinkel, K. M.) 2055–2060 (International Permafrost Association, 2008).

111. Wisser, D., Marchenko, S., Talbot, J., Treat, C. & Frolking, S. Soil temperature response to 21st century global warming: The role of and some implications for peat carbon in thawing permafrost soils in North America. *Earth Syst. Dyn.* **2**, 121–138 (2011).

112. Jeong, D. Il & Sushama, L. Projected changes to mean and extreme surface wind speeds for North America based on regional climate model simulations. *Atmospher.* **10**, 497 (2019).

113. McInnes, K. L., Erwin, T. A. & Bathols, J. M. Global Climate Model projected changes in 10 m wind speed and direction due to anthropogenic climate change. *Atmos. Sci. Lett.* **12**, 325–333 (2011).

114. Mudryk, L. R. et al. Canadian snow and sea ice: historical trends and projections. *Cryosphere* **12**, 1157–1176 (2018).

115. Liao, H., Chen, W. T. & Seinfeld, J. H. Role of climate change in global predictions of future tropospheric ozone and aerosols. *J. Geophys. Res. Atmos.* **111**, D12304 (2006).

Author contributions

A.R. and F.R. worked together in developing the variable selection methodology, creating the model, and led the preparation of the paper. E.R.H., P.V.C., W.L.Q. and K.W. discussed the results and contributed to writing and reviewing the paper and have approved the final version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43247-021-00184-w.

Correspondence and requests for materials should be addressed to F.R.

Peer review information Communications Earth & Environment thanks the anonymous reviewers for their contribution to the peer review of this work. Primary handling editors: Clare Davis. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.