ENVIRONMENTAL RESEARCH
LETTERS

LETTER

A boreal forest model benchmarking dataset for North America: a case study with the Canadian Land Surface Scheme Including Biogeochemical Cycles (CLASSIC)

Bo Qu[1,2,*], **Alexandre Roy**[2,3], **Joe R Melton**[4], **T Andrew Black**[5], **Brian Amiro**[6], **Eugénie S Euskirchen**[7], **Masahito Ueyama**[8], **Hideki Kobayashi**[9], **Christopher Schulze**[1,10], **Gabriel Hould Gosselin**[1,11], **Alex J Cannon**[4], **Matteo Detto**[12] and **Oliver Sonnentag**[1,2]

[1] Département de Géographie, Université de Montréal, Montréal, Canada
[2] Centre d'Études Nordiques, Université Laval, Québec, Canada
[3] Centre de Recherche Sur les Interactions Bassins Versants-écosystèmes Aquatiques (RIVE), Université du Québec à Trois-Rivières, Trois-Rivières, Canada
[4] Climate Research Division, Environment and Climate Change Canada, Victoria, Canada
[5] Biometeorology and Soil Physics Group, University of British Columbia, Vancouver, Canada
[6] Department of Soil Science, University of Manitoba, Winnipeg, Canada
[7] Institute of Arctic Biology, University of Alaska Fairbanks, Fairbanks, AK, United States of America
[8] Graduate School of Agriculture, Osaka Metropolitan University, Osaka 599-8531, Japan
[9] Research Institute for Global Change, Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan
[10] Department of Renewable Resources, University of Alberta, Edmonton, Canada
[11] Department of Geography, Wilfrid Laurier University, Waterloo, Canada
[12] Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, United States of America
[*] Author to whom any correspondence should be addressed.

E-mail: bo.qu@umontreal.ca

OPEN ACCESS

RECEIVED
24 February 2023

REVISED
26 June 2023

ACCEPTED FOR PUBLICATION
3 July 2023

PUBLISHED
18 July 2023

Keywords: benchmarking dataset, boreal forest, terrestrial ecosystem model, eddy covariance, CLASSIC

Supplementary material for this article is available online

Abstract

Climate change is rapidly altering composition, structure, and functioning of the boreal biome, across North America often broadly categorized into ecoregions. The resulting complex changes in different ecoregions present a challenge for efforts to accurately simulate carbon dioxide (CO_2) and energy exchanges between boreal forests and the atmosphere with terrestrial ecosystem models (TEMs). Eddy covariance measurements provide valuable information for evaluating the performance of TEMs and guiding their development. Here, we compiled a boreal forest model benchmarking dataset for North America by harmonizing eddy covariance and supporting measurements from eight black spruce (*Picea mariana*)-dominated, mature forest stands. The eight forest stands, located in six boreal ecoregions of North America, differ in stand characteristics, disturbance history, climate, permafrost conditions and soil properties. By compiling various data streams, the benchmarking dataset comprises data to parameterize, force, and evaluate TEMs. Specifically, it includes half-hourly, gap-filled meteorological forcing data, ancillary data essential for model parameterization, and half-hourly, gap-filled or partitioned component flux data on CO_2 (net ecosystem production, gross primary production [GPP], and ecosystem respiration [ER]) and energy (latent [LE] and sensible heat [H]) and their daily aggregates screened based on half-hourly gap-filling quality criteria. We present a case study with the Canadian Land Surface Scheme Including Biogeochemical Cycles (CLASSIC) to: (1) demonstrate the utility of our dataset to benchmark TEMs and (2) provide guidance for model development and refinement. Model skill was evaluated using several statistical metrics and further examined through the flux responses to their environmental controls. Our results suggest that CLASSIC tended to overestimate GPP and ER among all stands. Model performance regarding the

energy fluxes (i.e., LE and H) varied greatly among the stands and exhibited a moderate correlation with latitude. We identified strong relationships between simulated fluxes and their environmental controls except for H, thus highlighting current strengths and limitations of CLASSIC.

1. Introduction

The boreal biome is an integral component of the Earth's climate system, exchanging energy and matter (e.g., carbon and water) with the atmosphere (Bonan 2008). A large portion of the boreal biome is underlain by permafrost (perennially frozen ground; Gruber 2012), which increases in areal extent from south to north, from isolated (⩽10%), sporadic (>10%–50%) and discontinuous (>50%–90%), to continuous permafrost (>90%). In North America, forests occupy about 60% of the boreal biome along with lake, river, and wetland ecosystems (Brandt 2009). The presence of permafrost exerts a critical control over many physical, ecological, and biogeochemical processes governing boreal forest composition, structure, and functioning (Schuur and Mack 2018). Climate change over the Arctic–boreal region has led to unprecedented changes in boreal forests caused, for example, by amplified warming and associated permafrost thaw (Schuur and Abbott 2011) and an intensification of disturbance regimes (Foster *et al* 2022). Consequences of these changes include a potential release of large quantities of greenhouse gases to the atmosphere due to increased microbial activity in organic-rich soils and peats that are characteristic for large swaths of boreal forests (Koven *et al* 2011, Schuur *et al* 2015). However, how amplified climate warming changes boreal forests at the regional scale depends greatly on the complex interplay of permafrost, soils, vegetation, and natural and anthropogenic disturbance regimes (Kurz *et al* 2013, Gauthier *et al* 2015). The landscape diversity across the boreal biome has been broadly classified into distinctive ecoregions (e.g., Boreal Plain and Taiga Plain; United States Environmental Protection Agency Level II Ecoregions of North America). Understanding boreal forests' complex responses to climate change across ecoregions is important to evaluate the current and future role of boreal forests in the climate system (Nazarbakhsh *et al* 2020, Walker *et al* 2020).

Terrestrial ecosystem models (TEMs), including those integrated as the land component in Earth system models, are crucial tools to estimate boreal forests' responses to climate warming (Fisher *et al* 2018, Birch *et al* 2021). Recent decades have seen a rapid development in TEM skill to simulate carbon, water and energy fluxes and balances of terrestrial ecosystems (Fisher *et al* 2014, Bonan and Doney 2018). However, uncertainties in simulations remain large (Huntzinger *et al* 2012, Schwalm *et al* 2019). In particular, the Arctic–boreal region has been regarded as a large source of uncertainty in simulations of the Earth's climate system (Fisher *et al* 2018, Braghiere *et al* 2023). Model benchmarking datasets harmonizing different data streams to parameterize, force, and evaluate TEMs can help with efforts addressing uncertainties through more comprehensive evaluation of model strengths and limitations (Fisher *et al* 2018, Stofferahn *et al* 2019). Data collection efforts in North America's boreal biome were launched decades ago and have increased consistently since then. Prominent examples include the Boreal Ecosystem-Atmosphere Study (BOREAS) conducted in the 1990s (Sellers *et al* 1995), and the Fluxnet-Canada Research Network (FCRN) and the Canadian Carbon Program (CCP) in the 2000s (Coursolle *et al* 2012). BOREAS and FCRN/CCP included and focused on tower-based carbon dioxide (CO_2) and energy flux measurements made with the eddy covariance technique (Baldocchi and Meyers 1998). Numerous eddy covariance towers have been established and operated over different boreal ecoregions of North America (e.g., Helbig *et al* 2020), some through BOREAS and FCRN/CCP, but also by individual site investigators (Bergeron *et al* 2007, Euskirchen *et al* 2014, Ueyama *et al* 2014, Ikawa *et al* 2015, Helbig *et al* 2017). Many of the flux-tower teams contribute their eddy covariance and supporting measurements to network initiatives such as AmeriFlux and FLUXNET (e.g., Pastorello *et al* 2020). However, boreal forests are not well represented in those networking initiatives or previously published model benchmarking datasets (Williams *et al* 2009, Ukkola *et al* 2021).

Several challenges have been encountered by modellers attempting to use eddy covariance and supporting measurements. Examples include gaps in the meteorological forcing data and lack of convenient access to ancillary data to adequately parameterize soil, permafrost, and vegetation characteristics (e.g., near-surface organic soil depth, soil texture, and percentage covers of vegetation types). Eddy covariance networks have been pivotal to provide TEM communities with access to curated datasets (Boden *et al* 2013, Pastorello *et al* 2020). The FLUXNET2015 dataset is a prominent example that provides meteorological data, gap-filled using a global reanalysis dataset. The AmeriFlux Biological, Ancillary, Disturbance, and Metadata (BADM) comprise ancillary data describing soils and vegetation for each

tower site. However, up to now, FLUXNET2015 contains only a limited number of sites representative of North America's boreal forests. The AmeriFlux BADM is incomplete at many sites, lacking some important data required to parameterize, for example, 'percentage covers of plants in understory and ground cover' (Turetsky *et al* 2012), and 'permeable soil depth' (Lawrence *et al* 2008). Sustained efforts are essential to provide the modelling communities with consistently gap-filled and harmonized model benchmarking datasets for North America's boreal forests. These efforts require close collaboration between modelling and measurement communities (Fisher *et al* 2018).

In this study, we describe a boreal forest model benchmarking dataset comprising various data streams including eddy covariance and supporting measurements collected at eight mature boreal forest stands, located in six boreal ecoregions of the North America (Level II Ecoregions of North America; table S1, figure S1). First, we describe the curation of the dataset comprising harmonized and gap-filled: (1) half-hourly CO_2 and energy fluxes and their daily aggregates, (2) half-hourly meteorological data, and (3) ancillary data describing soil, permafrost, and vegetation characteristics. Second, we use the dataset to evaluate the performance of the Canadian Land Surface Scheme Including Biogeochemical Cycles (CLASSIC).

2. Data sources

2.1. Study sites

The study sites are eight boreal forest stands representative of a broad range of stand characteristics, disturbance history, climate, permafrost conditions and soil properties (figure 1, tables S1 and S2). All eight sites are mature boreal forest stands (>70 years old), dominated by black spruce (*Picea mariana*). Some forest stands also contain small percentages of tamarack (*Larix laricina*) and/or jack pine (*Pinus banksiana*). Permafrost extent varies among sites following a south-to-north gradient. Located near the southern limit of permafrost, the southern three sites (⩽60 °N), CA-Obs, CA-Qfo and CA-Man, are permafrost-free, closed forest stands with approximately 9- to 14 m-tall black spruce trees (Bergeron *et al* 2007; BADM). Four more northern sites (>60 °N), CA-SMC, US-BZS, US-Uaf, and US-Prr, are in the discontinuous permafrost zone, described as relatively open forest stands with shorter stature black spruce trees and more abundant understory and ground cover vegetation compared to their southern counterparts (Euskirchen *et al* 2014, Ueyama *et al* 2015, El-Amine *et al* 2022). One site, CA-HPC, is in the continuous permafrost zone and representative of the forest–tundra ecotone (Callaghan *et al* 2002), characterized by scattered, stunted, mostly black spruce, trees over continuous understory and ground cover vegetation (Martin *et al* 2022).

2.2. Eddy covariance measurements

We compiled and harmonized available half-hourly eddy covariance flux measurements from multiple sources (table S1). The fluxes include net ecosystem CO_2 exchange (NEE), its two partitioned component fluxes, gross primary production (GPP) and ecosystem respiration (ER), and latent (LE) and sensible heat (H). Data were available through FLUXNET2015 or were requested from site investigators. FLUXNET2015 provides NEE filtered for low-turbulence conditions using two thresholds for friction velocity (u_*) (Pastorello *et al* 2020): a variable u_* threshold and a constant u_* threshold (CUT). We chose the latter, 'NEE_CUT_REF', partitioned into GPP and ER following both nighttime-based (Reichstein *et al* 2005; 'GPP_NT_CUT_REF' and 'RECO_NT_CUT_REF') and daytime-based methods (Lasslop *et al* 2010; 'GPP_DT_CUT_REF' and 'RECO_DT_CUT_REF'). This choice assumes negligible changes in forest stand structures over the measurement periods and in the geometric characteristics of the instrumental set-ups at each site (e.g., height of the eddy covariance systems above the canopy). Marginal distribution sampling (MDS) following Reichstein *et al* (2005) was used by FLUXNET2015 and most site investigators to fill gaps in half-hourly NEE, LE and H (Euskirchen *et al* 2014, Ikawa *et al* 2015, Helbig *et al* 2017). We obtained LE and H gap-filled with MDS ('LE_F_MDS' and 'H_F_MDS') for the sites in FLUXNET2015. For other sites, we requested gap-filled NEE, LE and H from site investigators.

Daily flux aggregates include the sums of CO_2 fluxes (net ecosystem production [NEP = −NEE], GPP and ER in g C $m^{-2}\,d^{-1}$) and the averages of energy fluxes (LE and H in W m^{-2}), computed using half-hourly gap-filled (NEE, LE and H) and partitioned fluxes (GPP, ER). Daily GPP and ER from nighttime-based and daytime-based flux partitioning methods were averaged to account for the uncertainty in flux partitioning methods. Uncertainties arising from data collection and post-processing, including gap-filling (Richardson and Hollinger 2007), propagate into temporal flux aggregates (e.g., daily and monthly) (Soloway *et al* 2017) and might hamper model evaluation (Williams *et al* 2009). Thus, daily flux aggregates based on less than 80% high-quality (i.e., MDS quality flag 0 ['measured'] or 1 ['good quality gapfill']) half-hourly data (i.e., 39 out of 48 half-hours) were excluded. Similarly, daily aggregates computed from half-hourly fluxes not gap-filled with MDS (US-Uaf) were screened by excluding aggregates for which less than 40% high-quality (i.e., measured) half-hourly data (i.e., 20 out

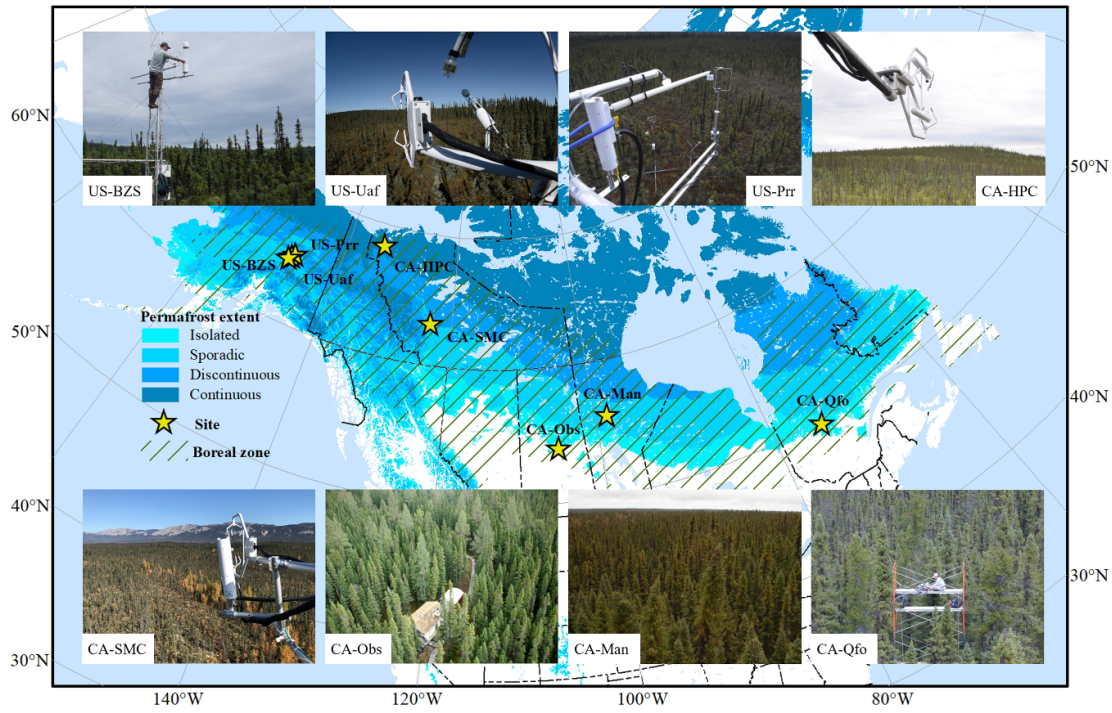

Figure 1. Locations of the study sites in the boreal biome (Brandt 2009) and permafrost zone (Gruber 2012) of North America. Permafrost is isolated (⩽10% in areal extent), sporadic (>10%–50%), discontinuous (>50%–90%), and continuous (>90%). Sites names refer to AmeriFlux ID: CA-Qfo (Quebec—Eastern Boreal, Mature Black Spruce), CA-Obs (Saskatchewan—Western Boreal, Mature Black Spruce [in the former BOREAS Southern Study Area]), CA-Man (Manitoba—Northern Old Black Spruce [in the former BOREAS Northern Study Area]), CA-SMC (Smith Creek), US-BZS (Bonanza Creek Black Spruce), US-Uaf (University of Alaska, Fairbanks), US-Prr (Poker Flat Research Range Black Spruce Forest), and CA-HPC (Havikpak Creek). Site photos were taken by the authors and obtained from Hank Margolis (CA-Qfo).

of 48 half-hours) were available. Growing season start, end, and length for each site-year were delineated using daily GPP and a double-logistic function fitted to the daily GPP time series (Gonsamo *et al* 2013). The first maximum and the last minimum of the third derivative of the fitted function were used to determine the growing season start and end, respectively. The growing season length was defined as the total days from the growing season start to end (figure S2).

2.3. Meteorological measurements

Half-hourly meteorological measurements include downwelling shortwave radiation (SW, W m^{-2}), downwelling longwave radiation (LW, W m^{-2}), total precipitation (sums of rain- and snowfall [as snow water equivalent], mm), air temperature (TA, °C), vapour pressure deficit (VPD, hPa), wind speed (m s^{-1}), and atmospheric pressure (PA, kPa). Specific humidity (Q, kg kg^{-1}) was estimated using TA, PA, and VPD variables (Monteith and Unsworth 2013). FLUXNET2015 provides gap-filled meteorological data on the basis of downscaled ERA-Interim reanalysis data (Vuichard and Papale 2015, Pastorello *et al* 2020). Based on the partially gap-filled data from site investigators (table S1), we applied a downscaled, bias-adjusted and temporally disaggregated global reanalysis dataset, GSWP3-W5E5-ERA5 (Melton and Arora 2016, Cannon 2018, Lange 2019, Meyer *et al* 2021), to fill any remaining gaps in all variables and replace the medium-to-poor quality gap-filled data (quality flag of 2 and 3) in the variables that were gap-filled using MDS.

2.4. Ancillary data

Ancillary data describing soil, permafrost, and vegetation characteristics including disturbance history are essential to adequately represent ecosystems in TEMs through parameterization (Lawrence *et al* 2008, Launiainen *et al* 2015). Near-surface soil data include soil texture (e.g., sandy loam) and the proportion of sand, silt, and clay (%) comprising the mineral fraction, in addition to soil organic layer thickness (cm) and organic carbon content (g [100 g]$^{-1}$), soil permeable depth (m) and drainage class (table 1). For three sites, CA-Obs, CA-Man, and CA-Qfo, most of the data were available through AmeriFlux BADM (table S1). For the remaining five sites, most of these data were obtained from the literature or from site investigators (unpublished data). In addition, soil data were extracted from the World Soil Information Service (Batjes *et al* 2020) when no published or unpublished data were available. A circular buffer of 20 km radius was defined around each site to select

Table 1. Soil, permafrost, and vegetation characteristics. Site names refer to AmeriFlux ID (AMF-ID) and are ordered by latitude, from south (CA-Qfo) to north (CA-HPC; figure 1). Soil and permafrost characteristics are near-surface soil organic layer thickness (SOL) and organic carbon content (SOC), soil permeable depth (SPD), and soil drainage class (SDC), and active layer thicknesses (ALT), respectively. Vegetation is overstory tree crown closure (OCC), stand height (SH), stand age (SA), and percentage cover of plant functional types in overstory and understory (PFT-OU) and ground cover (PFT-G). SDC is well (W), moderately well (MW), imperfectly (I), poorly (P) and very poorly drained (VP; https://sis.agr.gc.ca/cansis/nsdb/soil/v2/snt/drainage.html). PFT-OU are evergreen needleleaf forest (ENF), deciduous needleleaf forest (DNF), evergreen broadleaf shrub (EBS), deciduous broadleaf shrub (DBS), C3 grass (C3G), sedge (SDG), and forb (FORB). Some site characteristics are not applicable (NA).

Site (AMF-ID)	CA-Qfo	CA-Obs	CA-Man	CA-SMC	US-BZS	US-Uaf	US-Prr	CA-HPC
Lat, Long (°)	49.693, −74.342	53.987, −105.118	55.880, −98.481	63.153, −123.252	64.696, −148.324	64.866, −147.856	65.124, −147.488	68.320, −133.519
Near-surface SOL (cm)	15–40	20–30	30–50	100–200	100	50–100	39	30–60
Near-surface SOC (g $[100\ g]^{-1}$)	47	44–46	40–47	43–49	24–41	45	45	47–49
Soil texture	Sandy loam	Sandy loam to sand	Clay	Silty clay to silty clay loam	Silt	Silt loam	Silt loam	Silty clay
Sand (%)	51–52	65–96	1–3	10–12	9	22–31	29–38	10
Silt (%)	44–45	2–29	4–23	45–53	88	64–66	56–66	40
Clay (%)	4	2–7	76–95	35–45	3	5–12	4–6	50
SPD (m)	31	32	36	46	16	16	5	7
SDC	W to MW	P to VP	W to MW	P to VP	P	P to VP	P to VP	I
ALT (cm)	NA	NA	NA	62–85	50–90	40–50	43	40–80
OCC	Closed	Closed	Closed	Open	Open	Open	Open	Open
SH (m)	13.8	11.0	9.1	4.6	4.5	3.0	2.4	2.4
SA (years)	100 in 2005	111 in 2005	155 in 1994	141 in 2019	100 in 2014	85 in 2014	67 in 2012	109 in 2019
PFT-OU	90% ENF, 10% DBS	90% ENF, 10% DNF	90% ENF, 5% EBS, 5% DBS	22% ENF, 34% EBS, 12% DBS, 15% SDG, 4% FORB	21% ENF, 75% EBS, 22% DBS, 39% SDG, 13% FORB	20% ENF, 45% EBS, 10% DBS, 25% SDG	20% ENF, 14% EBS, 13% DBS, 24% SDG	15% ENF, 58% EBS, 5% DBS, 4% C3G, 3% FORB
PFT-G	13% *Sphagnum*, 46% feather moss, 6% lichen	10% *Sphagnum*, 70% feather moss, 10% lichen	34% *Sphagnum*, 53% feather moss, 17% brown moss	36% *Sphagnum*, 25% feather moss, 15% lichen	6% *Sphagnum*, 73% feather moss, 10% lichen	76% *Sphagnum*, 14% feather moss, 10% lichen	14% *Sphagnum*, 13% feather moss, 13% lichen	4% brown moss, 9% *Sphagnum*, 18% feather moss, 42% lichen

WoSIS soil profiles closest to each site. From these, the soil profile with the shortest radial distance to, and the most similar in near-surface organic layer thickness compared to the respective site was selected to extract missing soil data.

Soil water drainage is a key property controlling the 'wetness' of the land surface, and is closely associated with plant composition and distribution, soil carbon storage, and disturbance history (Wickland *et al* 2010). To provide consistent data on soil water drainage for each site, soil drainage class from the Soil Survey Geographic database and the Soil Landscapes of Canada database was extracted for the sites in Alaska and in Canada, respectively (tables 1 and S1). Soil permeable depth was extracted as 'soil and sedimentary deposit thickness' from a global 1 km gridded soil thickness dataset (Pelletier *et al* 2016). Consistent soil permeable depth for each site was reported as the mean estimate of the pixel containing the respective site and its eight neighbouring pixels. The seasonal freeze–thaw cycle in the active layer regulates soil moisture and thermal conditions for plant roots and microbes in permafrost (Schuur and Mack 2018). Active layer thickness, commonly year-round maximum thaw depth describing growing-season permafrost conditions (Fisher *et al* 2016), was compiled based on estimates made by site investigators.

Vegetation data include the percentage covers (%) of plant functional types (PFTs) in overstory, understory and ground cover vegetation, stand age (years) and stand height (m; table 1). Data sources were published in-situ field measurements, e.g., Kobayashi *et al* (2013) and Euskirchen *et al* (2014), and site investigator assessments (unpublished data, table S1). Overstory and understory plants were aggregated into seven commonly used PFTs (Groenendijk *et al* 2011, Rogers *et al* 2021) based on the information on common life-forms, e.g., growth form (herbs, shrubs, and trees), leaf (needleleaf and broadleaf), and leaf status (evergreen and deciduous). Ground cover composition distinguishes *Sphagnum*, brown moss, feather moss, and lichen (Turetsky *et al* 2012).

3. Modelling protocol

We used our model benchmarking dataset to evaluate the performance of the Canadian Land Surface Scheme Including Biogeochemical Cycles (CLASSIC; Melton *et al* 2020). We demonstrated the usefulness of our dataset to benchmark TEMs by examining biases and uncertainties in simulated CO_2 and energy fluxes. CLASSIC is the successor to the coupled model framework of the Canadian Land Surface Scheme (CLASS; Verseghy 2017) and the Canadian Terrestrial Ecosystem Model (CTEM; Melton and Arora 2016). In CLASSIC, CTEM simulates biogeochemistry including the photosynthetic CO_2 uptake and respiratory CO_2 loss based on five carbon pools in litter, soils, leaves, stems, and roots (Arora and Boer 2005). CLASS solves the model physics including land surface energy and water fluxes using vegetation structural properties (e.g., leaf area index) simulated by CTEM (Verseghy 2017). The two models, CLASS and CTEM, have been under continuous development since they were introduced in the 1980s and 2000s, respectively. CLASSIC has recently been extensively benchmarked at several sites and tested at global scale (Seiler *et al* 2021).

3.1. Site parameterization

Vegetation in CLASSIC is represented by nine default PFTs including five forest (evergreen needleleaf forest, deciduous needleleaf forest, evergreen broadleaf forest, deciduous cold broadleaf forest and deciduous dry broadleaf forest), two grass (C_3 and C_4 grass), and two crop (C_3 and C_4 crop) PFTs (Melton and Arora 2016). Recently, a version with three additional PFTs (evergreen broadleaf shrub, deciduous broadleaf shrub and sedge) was developed for simulations in Low-Arctic tundra (Meyer *et al* 2021). Our stand-level case study used a CLASSIC version with a total of 12 PFTs. We parameterized CLASSIC according to the compiled soil and vegetation data (section 2.4 and table S3). Over each site, the soil profile was stratified into 20 layers with a total ground column depth of 61.4 m (Melton *et al* 2019). Soil layers below the soil permeable depth were simulated as hydrologically inactive bedrock. The organic layers were assumed to be peat and assigned to be fibric for the first layer, hemic for the second and third layers, and sapric for any lower layers, using the thermal and hydraulic parameters described in Letts *et al* (2000). The permeable soil layers below soil horizon B were all assumed to be parent material (i.e., soil horizon C) by using the properties of soil horizon C. We used the default parameterization for PFTs, e.g., visible and near-infrared albedo and rooting depth (Melton and Arora 2016, Meyer *et al* 2021).

The gap-filled meteorological data were used to force CLASSIC. We applied a spin-up procedure by repeating the forcing data until the model reached an equilibrium, which is defined by the soil carbon pools varying by less than 0.1% compared to the last loop. CLASSIC simulations were performed by forcing the model one more time after the spin-up.

3.2. Performance evaluation

We used a suite of statistical metrics to evaluate CLASSIC performance for NEP, GPP, ER, LE and H: bias, coefficient of determination (R^2), Pearson correlation coefficient (r; Taylor diagrams), and root mean square error (RMSE) (Taylor 2001, Tramontana *et al* 2016). The performance evaluation was based on screened daily fluxes and restricted for GPP to the delineated growing seasons (as defined in section 2.2).

Model performance was evaluated among sites and ecoregions, respectively. Model performance was also investigated by examining the responses of the modelled fluxes to different environmental controls including TA, SW and VPD (Schaefer *et al* 2012, Tramontana *et al* 2016). The locally estimated scatterplot smoothing (LOESS), a nonparametric regression function based on local polynomial fitting (Cleveland and Loader 1996), was applied to fit response curves for the modelled and observed fluxes, respectively. We analysed the shape characteristics of LOESS curves, e.g., slope, transition, and summit, to identify limitations and strengths of CLASSIC simulations (Schaefer *et al* 2012, Rogers *et al* 2017).

4. Results and discussion

4.1. Model benchmarking dataset

The model benchmarking dataset comprises mature boreal forest stands dominated by black spruce, one of the most widespread tree species in the boreal forests of North America (Kurz *et al* 2013). The stands are located in six Level II Ecoregions of North America: Boreal Plain (CA-Obs), Softwood Shield (CA-Qfo and CA-Man), Taiga Plain (CA-SMC), Alaska Boreal Interior (US-BZS and US-Uaf), Boreal Cordillera (US-Prr) and Southern Arctic (CA-HPC) (figure S1; table S1). The dataset contains a total of 68 site-years distributed broadly over the 'boreal forest' and 'woodland/shrubland' classes defined by the Whittaker biomes (Whittaker 1970; figure 2(a)). Specifically, the sites span numerous gradients including, for example, wetness (East–West) in the Softwood Shield from CA-Qfo to CA-Man, and TA and permafrost extent (South–North) from the Boreal Plain (CA-Obs), though Taiga Plain (CA-SMC), to Southern Arctic (CA-HPC) (table S2). Other important variations across the sites include soil properties, overstory tree crown coverage, and understory and ground cover composition (table 1).

Growing season start, end, and length varied across ecoregions (figure 2(b)). The growing season lengths in the Boreal Plain and Softwood Shield (CA-Obs, CA-Qfo, and CA-Man) tended to be longer than other ecoregions because of both an earlier start and a later end. In the Southern Arctic (CA-HPC) the growing seasons were substantially shorter. As an indicator of the forest-stand photosynthetic activity (Gonsamo *et al* 2013), the estimated growing seasons were closely linked to abiotic and biotic factors, such as dominant plant species, climate, soil freeze–thaw seasonality, and photoperiod (Bauerle *et al* 2012, Barichivich *et al* 2013).

4.2. Flux data screening

Data with high uncertainty in daily aggregates were effectively excluded by our screening procedures (section 2.2). For example, at US-Prr, the screening excluded most of the data in early to mid-2010, late 2012 and mid-2013 (figure S3). During those periods, numerous longer gaps (e.g., weeks to seasons) in half-hourly data existed due to, e.g., instrument failures and insufficient quality. Gap-filling uncertainty tends to increase with gap length (Richardson and Hollinger 2007), and the associated uncertainty might propagate into temporal aggregates (e.g., daily and monthly), increasing the uncertainty of model validations and then predictions (Williams *et al* 2009).

4.3. CLASSIC performance

The performance of CLASSIC was evaluated for sites (figure 3) and ecoregions (figure 4). Simulated daily GPP, ER, and LE were well correlated with observations ($r \geqslant 0.78$), suggesting that the model well captured seasonal flux variations (Schaefer *et al* 2012), although the three fluxes tended to be overestimated. NEP was generally well simulated (e.g., low RMSE), but the Pearson correlations varied greatly among sites and ecoregions ($r = 0.55$–0.89).

Model performance varied greatly over the months for all fluxes (figures S4 and S5). Similar seasonal patterns were found among sites and ecoregions, e.g., high RMSE (all fluxes) and high R^2 (LE and H) in summer (from June to August). The high summer RMSE provided a major contribution to the overall model performance. This could be explained by the fact that model-observation discrepancies for CO_2 and energy fluxes tend to be higher with higher fluxes in summer (Kuppel *et al* 2012, Birch *et al* 2021). The monthly biases indicate that the model overestimated the seasonal amplitudes in NEP among sites except for US-Prr.

Model performance variations among sites and ecoregions were remarkable. Comparing among ecoregions, Boreal Cordillera (i.e., US-Prr) was the best simulated with the lowest RMSE for all fluxes (figure 4). A clear latitudinal variation was shown in model performance of energy fluxes. Boreal Plain and Softwood Shield tended to have a higher RMSE in LE than northern ecoregions. The seasonal bias of H had distinct differences from Boreal Plain and Softwood Shield, though Taiga Plain, to other northern ecoregions (figure S5). Strong negative summer biases of H were found in the Boreal Plain and Softwood Shield.

The model generally performed well in simulating the seasonality of GPP, ER, LE, and H across site-years (high r) but was challenged to simulate the year-to-year variations in the magnitudes of these fluxes (large variations in RMSE) (figure S6). The model tended to overestimate annual values of GPP, ER and LE (figure S7), consistent with the performance evaluated at daily scales. Small year-to-year variations in RMSE were found for NEP, while model performance of the seasonality in NEP tended to be low (low r; figure S6). Regarding annual values of NEP, model performance varied greatly across site-years, depending on simulated GPP and ER (figure S7). However, the overall quality of model performance regarding

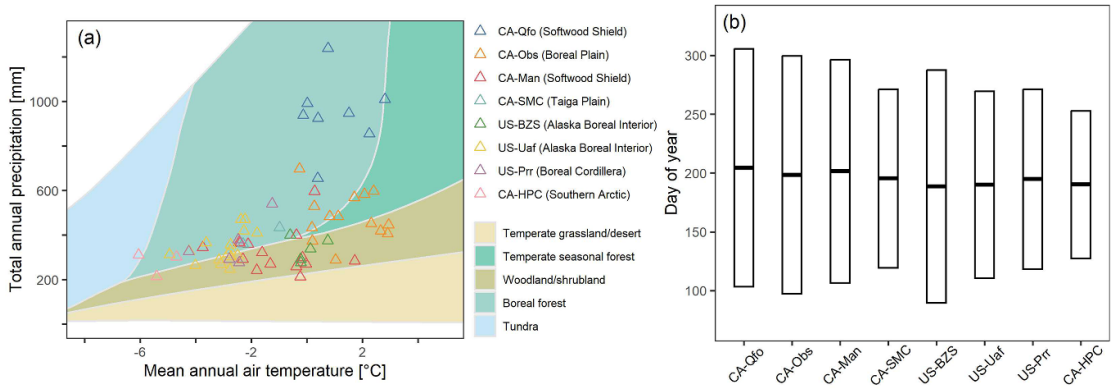

Figure 2. Dataset characteristics: (a) site-year climates ($n = 68$) over the Whittaker biomes (Whittaker 1970), (b) start, end, and length of growing seasons (section 2.2 and figure S2). The sites are displayed in a latitudinal order from south (CA-Qfo) to north (CA-HPC; figure 1).

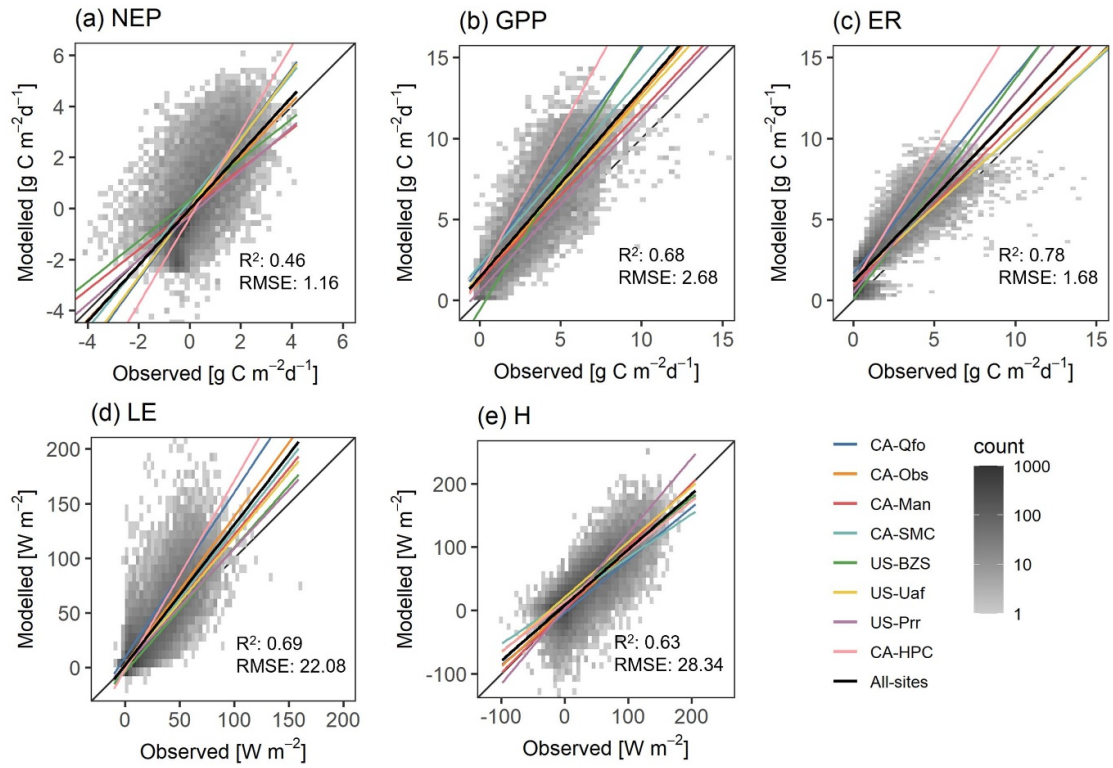

Figure 3. Comparison of CO_2 and energy fluxes simulated with CLASSIC against observations. Daily fluxes are net ecosystem production (NEP), gross primary production (GPP), ecosystem respiration (ER), and latent (LE) and sensible (H) heat fluxes. The GPP comparison was limited to growing seasons (section 2.2). Scatter density plots and regression statistics are based on data from all sites and details of regression statistics are shown in table S4. The sites are ordered by latitude, from south (CA-Qfo) to north (CA-HPC; figure 1).

annual values was certainly influenced by the uncertainty in gap-filled fluxes (Soloway *et al* 2017).

4.4. Flux responses to environmental controls

The response of modelled NEP to TA generally captured magnitudes and variations in observations (figure 5(a)). However, CLASSIC underestimated NEP between −5 °C and 5 °C and overestimated NEP between 15 °C and 25 °C. These mismatches are consistent with the amplified NEP seasonality simulated by CLASSIC. For example, the underestimated NEP response from −5 °C to 5 °C was primarily due to the overestimated temperature response of ER at the same TA (figure 5(b)), which is possibly related to model parameters or formulations for autotrophic and/or heterotrophic respiration, e.g., Q_{10} (Tuomi *et al* 2008, Piao *et al* 2010), or biases in simulated carbon pools (Schmid *et al* 2006) and soil thermal dynamics (Exbrayat *et al* 2013). The NEP overestimations from 15 °C to 25 °C, in contrast, were difficult to interpret because various responses of GPP and ER to environmental controls could possibly be involved,

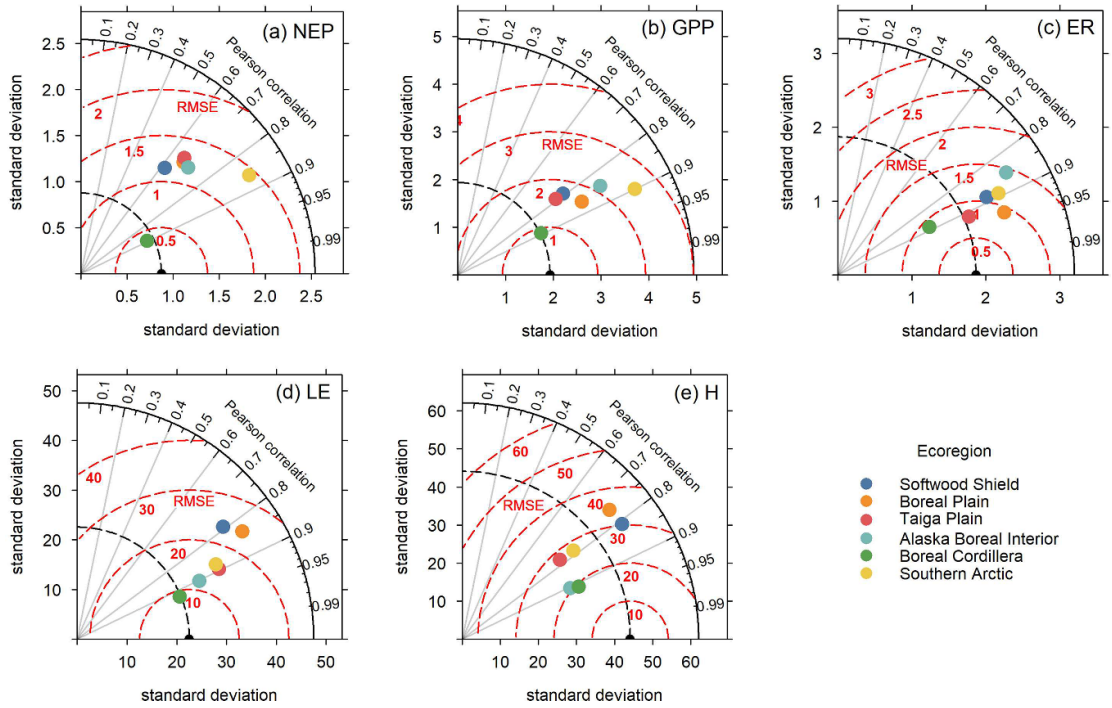

Figure 4. Taylor diagrams of CLASSIC performance across ecoregions. The units of standard deviations and root mean square error (RMSE) are g C m^{-2} d^{-1} for CO_2 fluxes (net ecosystem production [NEP], gross primary production [GPP] and ecosystem respiration [ER]) and W m^{-2} for energy fluxes (latent [LE] and sensible heat [H]). The GPP comparison was limited to growing seasons (section 2.2).

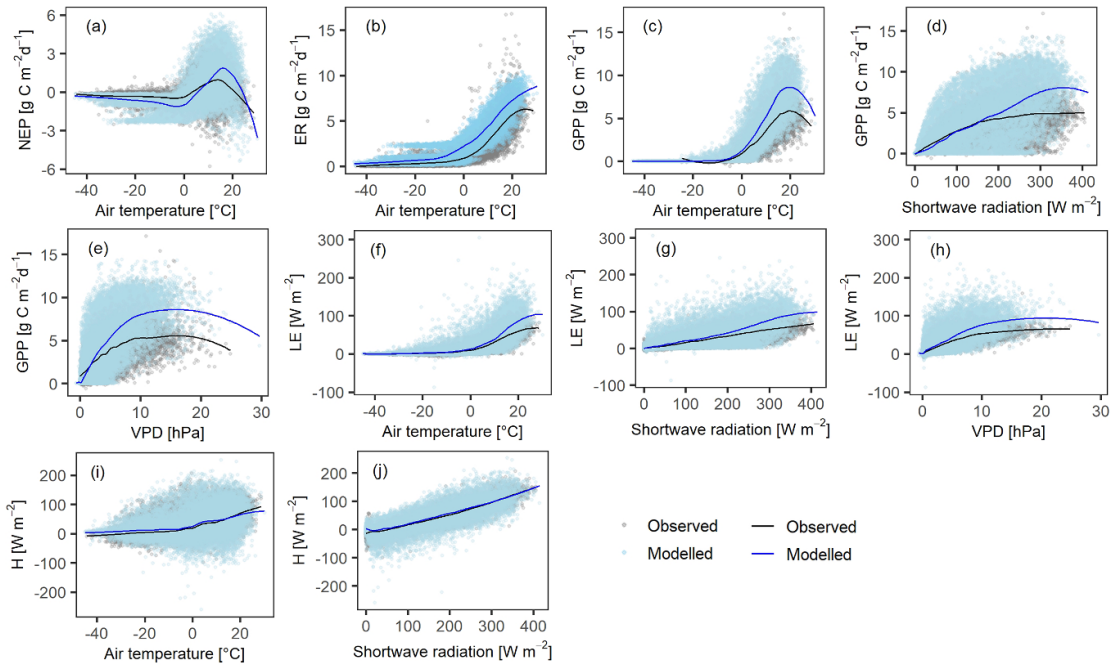

Figure 5. Response curves of observed and modelled fluxes to environmental controls by locally estimated scatterplot smoothing (Cleveland and Loader 1996). The estimates were based on daily observations and simulations from all sites. The GPP comparison was limited to growing seasons (section 2.2). (a) Net ecosystem production (NEP) to air temperature (TA), (b) ecosystem respiration (ER) to TA, (c) gross primary production (GPP) to TA, (d) GPP to downwelling shortwave radiation (SW), (e) GPP to vapour pressure deficit (VPD), (f) latent heat (LE) to TA, (g) LE to SW, (h) LE to VPD, (i) sensible heat (H) to TA and (j) H to SW.

e.g., GPP or ER to TA and GPP to TA, SW or VPD (Yuan *et al* 2008, Schaefer *et al* 2012).

The model closely matched observed GPP responses to TA and SW when TA is less than 5 °C (figure 5(c)) and SW is less than 150 W m^{-2} (figure 5(d)). However, clear overestimations in the modelled responses occurred when TA and SW increase, which explains the large biases of GPP during summer months and may likely be a major contributor to the NEP overestimations from 15 °C to 25 °C. Possible reasons for the GPP overestimations may include biases in photosynthetic parameters, in particular the maximum photosynthetic carboxylation rate (V_{cmax}), which determines modelled photosynthetic capacity under light saturation and controls the GPP responses to SW and TA (Rogers *et al* 2017). Although there were overestimations, the model still captured variations in the response of GPP to TA (e.g., the optimal temperature of GPP).

VPD reflects the moisture status of the atmosphere (Monteith and Unsworth 2013). The observed GPP increased with increasing VPD and reached a maximum at around 15 hPa (figure 5(e)). A main cause of this response might be that transpiration generally does not continually increase in proportion to VPD because high VPD (i.e., a dry atmosphere) leads to reduced stomatal conductance and photosynthesis (Schaefer *et al* 2012). CLASSIC overestimated GPP response to VPD, especially at >10 hPa, indicating that it simulated a smaller atmospheric humidity stress compared to the observations when accounting for stomatal conductance impacts on photosynthesis. Because photosynthesis and transpiration are tightly coupled through a stomatal conductance model in CLASSIC (Verseghy 2017), the GPP responses to environmental controls are generally closely associated with the responses of LE, e.g., LE and GPP to VPD (figures 5(e) and (h)).

In contrast to LE, CLASSIC closely matched observed responses of H to TA and SW (figures 5(i) and (j)). An important reason might be that H was not directly coupled to the stomatal conductance in CLASSIC (Verseghy 2017). However, the close matches were also possibly due to the contrasting performance in H among sites, where negative biases at some sites (e.g., CA-Qfo and CA-Obs) were offset by positive biases at others (e.g., US-Uaf and US-Prr; figure S5).

4.5. Limitations and perspectives

Eddy covariance and supporting measurements are affected by uncertainties which may hamper model parameterization and evaluation attempts (Williams *et al* 2009). Although we carefully screened the flux data, errors from data collection and processing may remain (Richardson *et al* 2006, Soloway *et al* 2017), in particular, GPP and ER, which were not directly measured but derived from measured NEE (Schaefer *et al* 2012). The uncertainty of LE and H may come from energy balance closure (Pastorello *et al* 2020). The compiled data did not include corrected LE and H products. CLASSIC enforces energy balance closure (Verseghy 2017). Correcting energy balance closure would help further examine model performance (Ukkola *et al* 2021).

Carefully isolating the most important model parameters or formulations for characterizing model performance is a challenge due to equifinality in model parameterization (Bonan *et al* 2011). Several possible reasons may be involved in CLASSIC model limitations. First, our analysis is based on ecosystem-level fluxes, but plant species as well as their aggregations by PFTs generally differ greatly in their responses to environment (Lloyd and Bunn 2007, Pearson *et al* 2013). In models, parameters and/or formulations generally vary with PFTs (Melton and Arora 2016), interacting with each other in ecosystem-level simulations. Second, the impacts of moss and lichen on the ground are currently not simulated in CLASSIC, even though ground cover vegetation can contribute substantially to ecosystem-level CO_2 and energy fluxes of boreal forest stands (Heijmans *et al* 2004, Gaumont-Guay *et al* 2014). The absence of moss and lichen ground cover in CLASSIC might be compensated by the parameterization of overstory and understory PFTs, but biases could still be introduced because moss and lichen have environmental responses distinctly different from overstory and understory plants (Nilsson and Wardle 2005, Turetsky *et al* 2012). Third, the soil physics component of TEMs, e.g., soil thermal dynamics and soil moisture, is important for modelling CO_2 and energy fluxes as well, especially in the forest stands underlain by permafrost (Ju *et al* 2006, Melton *et al* 2019, Andresen *et al* 2020). Uncertainty in the soil physics component has been identified as an important contributor to uncertainty in modelling CO_2 and energy fluxes in boreal forests (Ju *et al* 2006, Jiang *et al* 2016).

Despite the above limitations, our simulations with CLASSIC demonstrate the capacity of our model benchmarking dataset to parameterize, force, and evaluate TEMs. Our dataset provides numerous new opportunities for understanding and addressing model shortcomings and associated uncertainties in North America's boreal forests. In addition, our dataset has the potential to be a powerful tool for further model–data fusion attempts. For example, a parameter optimization (e.g., on V_{cmax}) could be conducted using our dataset to improve GPP performance (Kuppel *et al* 2014). Other observations at the studied boreal forest stands, e.g., ground CO_2 fluxes by soil chambers, can be easily combined with our dataset to constrain TEMs with both ecosystem-level and ground fluxes (Purdy *et al* 2016). Our dataset can be combined with satellite-derived datasets, e.g., leaf area index and GPP, for model development (Seiler *et al* 2021).

Our dataset can be utilized to study CO_2 and energy fluxes of black spruce-dominated, mature boreal forests, e.g., the magnitudes and links to the variability of environmental variables (Bergeron *et al* 2007), and for flux upscaling through, for example, machine learning (Jung *et al* 2020).

5. Conclusions

We compiled and harmonized eddy covariance and supporting measurements made at eight sites in a model benchmarking dataset for North America's boreal forests. We present a case study of the dataset to initialize, force, and evaluate CLASSIC. This case study allowed us to identify strengths and limitations of CLASSIC regarding its skill to reproduce CO_2 and energy fluxes across boreal forest stands. Based on the estimated responses of CO_2 and energy fluxes to their environmental controls, our case study highlights possible avenues for the improvement of CLASSIC through, e.g., V_{cmax} parameterization. While model structures and parameters differ widely among TEMs, our analysis for CLASSIC provides an example of how similar studies could be performed with other TEMs to evaluate and improve their performance for North America's boreal forests.

Data availability statements

The data that support the findings of this study are openly available at the following URL/DOI: https://doi.org/10.5281/zenodo.7266010. Variable descriptions and units are provided in table 1 (ancillary data) and table S5 (meteorological and flux data).

Author contributions

B Q compiled and processed the dataset and conducted model evaluation with inputs from O S, J R M, and A R. O S, G H G, C S, E S E, A J C and M U provided eddy-covariance and/or meteorological data. T A B, B A, E S E, M U, H K, C S, M D, and O S provided expert assessments and/or unpublished ancillary data. B Q prepared the manuscript with contributions from all authors.

Acknowledgments

B Q acknowledges funding from the China Scholarship Council and Centre d'Études Nordiques. T A B acknowledges the support for CA-Obs through grants from the Fluxnet Canada Research Network (FCRN) and the Canadian Carbon Program (CCP) and by an NSERC (Climate Change and Atmospheric Research) Grant to the Changing Cold Regions Network, and an NSERC Discovery Grant. US-Uaf and US-Prr teams acknowledge funding through the Arctic Challenge for Sustainability (ArCS; JPMXD1300000000) and ArCS II (JPMXD1420318865). The US-BZS team acknowledges funding from the US Geological Survey, Research Work Order 224, the National Science Foundation (NSF DEB-1026415, DEB-1636476), and the NSF Long-Term Research in Environmental Biology Program (NSF LTREB 2011276). O S acknowledges funding through the Canada Research Chairs, Canada Foundation for Innovation Leaders Opportunity Fund, and Natural Sciences and Engineering Research Council Discovery Grant programs. We acknowledge the use of FLUXNET data, and the work of Hank Margolis and Carole Coursolle (FCRN/CCP) at CA-Qfo. We thank the flux-tower teams and CLASSIC team for providing support.

ORCID iDs

Bo Qu https://orcid.org/0000-0002-0973-7274
Joe R Melton https://orcid.org/0000-0002-9414-064X
Masahito Ueyama https://orcid.org/0000-0002-4000-4888
Christopher Schulze https://orcid.org/0000-0002-6579-0360
Alex J Cannon https://orcid.org/0000-0002-8025-3790
Matteo Detto https://orcid.org/0000-0003-0494-188X

References

Andresen C G *et al* 2020 Soil moisture and hydrology projections of the permafrost region—a model intercomparison The *Cryosphere* **14** 445–59
Arora V K and Boer G J 2005 A parameterization of leaf phenology for the terrestrial ecosystem component of climate models *Glob. Change Biol.* **11** 39–59
Baldocchi D and Meyers T 1998 On using eco-physiological, micrometeorological and biogeochemical theory to evaluate carbon dioxide, water vapor and trace gas fluxes over vegetation: a perspective *Agric. For. Meteorol.* **90** 1–25
Barichivich J, Briffa K R, Myneni R B, Osborn T J, Melvin T M, Ciais P, Piao S and Tucker C 2013 Large-scale variations in the vegetation growing season and annual cycle of atmospheric CO_2 at high northern latitudes from 1950 to 2011 *Glob. Change Biol.* **19** 3167–83
Batjes N H, Ribeiro E and Van Oostrum A 2020 Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019) *Earth Syst. Sci. Data* **12** 299–320
Bauerle W L, Oren R, Way D A, Qian S S, Stoy P C, Thornton P E, Bowden J D, Hoffman F M and Reynolds R F 2012 Photoperiodic regulation of the seasonal pattern of photosynthetic capacity and the implications for carbon cycling *Proc. Natl Acad. Sci.* **109** 8612–7
Bergeron O, Margolis H A, Black T A, Coursolle C, Dunn A L, Barr A G and Wofsy S C 2007 Comparison of carbon dioxide fluxes over three boreal black spruce forests in Canada *Glob. Change Biol.* **13** 89–107
Birch L, Schwalm C R, Natali S, Lombardozzi D, Keppel-Aleks G, Watts J, Lin X, Zona D, Oechel W and Sachs T 2021 Addressing biases in Arctic–boreal carbon cycling in the Community Land Model Version 5 *Geosci. Model Dev.* **14** 3361–82
Boden T A, Krassovski M and Yang B 2013 The AmeriFlux data activity and data system: an evolving collection of data

management techniques, tools, products and services *Geosci. Instrum. Methods Data Syst.* **2** 165–76
Bonan G B 2008 Forests and climate change: forcings, feedbacks, and the climate benefits of forests *Science* **320** 1444–9
Bonan G B and Doney S C 2018 Climate, ecosystems, and planetary futures: the challenge to predict life in Earth system models *Science* **359** eaam8328
Bonan G B, Lawrence P J, Oleson K W, Levis S, Jung M, Reichstein M, Lawrence D M and Swenson S C 2011 Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data *J. Geophys. Res. Biogeosci.* **116** G02014
Braghiere R K, Fisher J B, Miner K R, Miller C E, Worden J R, Schimel D S and Frankenberg C 2023 Tipping point in North American Arctic-boreal carbon sink persists in new generation Earth system models despite reduced uncertainty *Environ. Res. Lett.* **18** 025008
Brandt J P 2009 The extent of the North American boreal zone *Environ. Rev.* **17** 101–61
Callaghan T V, Crawford R M, Eronen M, Hofgaard A, Payette S, Rees W G, Skre O, Sveinbjörnsson B, Vlassova T K and Werkman B R 2002 The dynamics of the tundra-taiga boundary: an overview and suggested coordinated and integrated approach to research *Ambio* **Spec No 12** 3–5
Cannon A J 2018 Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables *Clim. Dyn.* **50** 31–49
Cleveland W S and Loader C 1996 Smoothing by local regression: principles and methods *Statistical Theory and Computational Aspects of Smoothing. Contributions to Statistics* ed W Härdle and M G Schimek (Physica-Verlag HD) (https://doi.org/10.1007/978-3-642-48425-4_2)
Coursolle C, Margolis H, Giasson M-A, Bernier P-Y, Amiro B, Arain M, Barr A, Black T, Goulden M and McCaughey J 2012 Influence of stand age on the magnitude and seasonality of carbon fluxes in Canadian forests *Agric. For. Meteorol.* **165** 136–48
El-Amine M, Roy A, Koebsch F, Baltzer J L, Barr A, Black A, Ikawa H, Iwata H, Kobayashi H and Ueyama M 2022 What explains the year-to-year variation in growing season timing of boreal black spruce forests? *Agric. For. Meteorol.* **324** 109113
Euskirchen E S, Edgar C W, Turetsky M R, Waldrop M P and Harden J W 2014 Differential response of carbon fluxes to climate in three peatland ecosystems that vary in the presence and stability of permafrost *J. Geophys. Res. Biogeosci.* **119** 1576–95
Exbrayat J-F, Pitman A, Zhang Q, Abramowitz G and Wang Y-P 2013 Examining soil carbon uncertainty in a global model: response of microbial decomposition to temperature, moisture and nutrient limitation *Biogeosciences* **10** 7095–108
Fisher J B, Hayes D J, Schwalm C R, Huntzinger D N, Stofferahn E, Schaefer K, Luo Y, Wullschleger S D, Goetz S and Miller C E 2018 Missing pieces to modeling the Arctic-Boreal puzzle *Environ. Res. Lett.* **13** 020202
Fisher J B, Huntzinger D N, Schwalm C R and Sitch S 2014 Modeling the terrestrial biosphere *Annu. Rev. Environ. Resour.* **39** 91–123
Fisher J P, Estop-Aragonés C, Thierry A, Charman D J, Wolfe S A, Hartley I P, Murton J B, Williams M and Phoenix G K 2016 The influence of vegetation and soil characteristics on active-layer thickness of permafrost soils in boreal forest *Glob. Change Biol.* **22** 3127–40
Foster A C, Wang J A, Frost G V, Davidson S J, Hoy E, Turner K W, Sonnentag O, Epstein H, Berner L T and Armstrong A H 2022 Disturbances in North American boreal forest and Arctic tundra: impacts, interactions, and responses *Environ. Res. Lett.* **17** 113001
Gaumont-Guay D, Black T, Barr A, Griffis T, Jassal R, Krishnan P, Grant N and Nesic Z 2014 Eight years of forest-floor CO_2 exchange in a boreal black spruce forest: spatial integration and long-term temporal trends *Agric. For. Meteorol.* **184** 25–35
Gauthier S, Bernier P, Kuuluvainen T, Shvidenko A and Schepaschenko D 2015 Boreal forest health and global change *Science* **349** 819–22
Gonsamo A, Chen J M and D'Odorico P 2013 Deriving land surface phenology indicators from CO_2 eddy covariance measurements *Ecol. Indic.* **29** 203–7
Groenendijk M, Dolman A, van der Molen M, Leuning R, Arneth A, Delpierre N, Gash J, Lindroth A, Richardson A and Verbeeck H 2011 Assessing parameter variability in a photosynthesis model within and between plant functional types using global Fluxnet eddy covariance data *Agric. For. Meteorol.* **151** 22–38
Gruber S 2012 Derivation and analysis of a high-resolution estimate of global permafrost zonation The *Cryosphere* **6** 221–33
Heijmans M M, Arp W J and Chapin F S III 2004 Carbon dioxide and water vapour exchange from understory species in boreal forest *Agric. For. Meteorol.* **123** 135–47
Helbig M, Chasmer L E, Desai A R, Kljun N, Quinton W L and Sonnentag O 2017 Direct and indirect climate change effects on carbon dioxide fluxes in a thawing boreal forest–wetland landscape *Glob. Change Biol.* **23** 3231–48
Helbig M, Waddington J M, Alekseychik P, Amiro B D, Aurela M, Barr A G, Black T A, Blanken P D, Carey S K and Chen J 2020 Increasing contribution of peatlands to boreal evapotranspiration in a warming climate *Nat. Clim. Change* **10** 555–60
Huntzinger D N, Post W M, Wei Y, Michalak A, West T O, Jacobson A, Baker I, Chen J M, Davis K and Hayes D 2012 North American Carbon Program (NACP) regional interim synthesis: terrestrial biospheric model intercomparison *Ecol. Modelling* **232** 144–57
Ikawa H, Nakai T, Busey R C, Kim Y, Kobayashi H, Nagai S, Ueyama M, Saito K, Nagano H and Suzuki R 2015 Understory CO_2, sensible heat, and latent heat fluxes in a black spruce forest in interior Alaska *Agric. For. Meteorol.* **214** 80–90
Jiang Y, Zhuang Q, Sitch S, O'Donnell J A, Kicklighter D, Sokolov A and Melillo J 2016 Importance of soil thermal regime in terrestrial ecosystem carbon dynamics in the circumpolar north *Glob. Planet. Change* **142** 28–40
Ju W, Chen J M, Black T A, Barr A G, Liu J and Chen B 2006 Modelling multi-year coupled carbon and water fluxes in a boreal aspen forest *Agric. For. Meteorol.* **140** 136–51
Jung M, Schwalm C, Migliavacca M, Walther S, Camps-Valls G, Koirala S, Anthoni P, Besnard S, Bodesheim P and Carvalhais N 2020 Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach *Biogeosciences* **17** 1343–65
Kobayashi H, Suzuki R, Nagai S, Nakai T and Kim Y 2013 Spatial scale and landscape heterogeneity effects on FAPAR in an open-canopy black spruce forest in interior Alaska *IEEE Geosci. Remote Sens. Lett.* **11** 564–8
Koven C D, Ringeval B, Friedlingstein P, Ciais P, Cadule P, Khvorostyanov D, Krinner G and Tarnocai C 2011 Permafrost carbon-climate feedbacks accelerate global warming *Proc. Natl Acad. Sci.* **108** 14769–74
Kuppel S, Peylin P, Chevallier F, Bacour C, Maignan F and Richardson A 2012 Constraining a global ecosystem model with multi-site eddy-covariance data *Biogeosciences* **9** 3757–76
Kuppel S, Peylin P, Maignan F, Chevallier F, Kiely G, Montagnani L and Cescatti A 2014 Model–data fusion across ecosystems: from multisite optimizations to global simulations *Geosci. Model Dev.* **7** 2581–97
Kurz W A, Shaw C, Boisvenue C, Stinson G, Metsaranta J, Leckie D, Dyk A, Smyth C and Neilson E 2013 Carbon in Canada's boreal forest—a synthesis *Environ. Rev.* **21** 260–92
Lange S 2019 Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BASD (v1.0) *Geosci. Model Dev.* **12** 3055–70

Lasslop G, Reichstein M, Papale D, Richardson A D, Arneth A, Barr A, Stoy P and Wohlfahrt G 2010 Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation *Glob. Change Biol.* **16** 187–208

Launiainen S, Katul G G, Lauren A and Kolari P 2015 Coupling boreal forest CO_2, H_2O and energy flows by a vertically structured forest canopy–soil model with separate bryophyte layer *Ecol. Modelling* **312** 385–405

Lawrence D M, Slater A G, Romanovsky V E and Nicolsky D J 2008 Sensitivity of a model projection of near-surface permafrost degradation to soil column depth and representation of soil organic matter *J. Geophys. Res. Earth Surf.* **113** F02011

Letts M G, Roulet N T, Comer N T, Skarupa M R and Verseghy D L 2000 Parametrization of peatland hydraulic properties for the Canadian Land Surface Scheme *Atmos.-Ocean* **38** 141–60

Lloyd A H and Bunn A G 2007 Responses of the circumpolar boreal forest to 20th century climate variability *Environ. Res. Lett.* **2** 045013

Martin M R, Kumar P, Sonnentag O and Marsh P 2022 Thermodynamic basis for the demarcation of Arctic and alpine treelines *Sci. Rep.* **12** 1–14

Melton J R and Arora V K 2016 Competition between plant functional types in the Canadian Terrestrial Ecosystem Model (CTEM) v. 2.0 *Geosci. Model Dev.* **9** 323–61

Melton J R, Arora V K, Wisernig-Cojoc E, Seiler C, Fortier M, Chan E and Teckentrup L 2020 CLASSIC v1. 0: the open-source community successor to the Canadian Land Surface Scheme (CLASS) and the Canadian Terrestrial Ecosystem Model (CTEM)–part 1: model framework and site-level performance *Geosci. Model Dev.* **13** 2825–50

Melton J R, Verseghy D L, Sospedra-Alfonso R and Gruber S 2019 Improving permafrost physics in the coupled Canadian Land Surface Scheme (v.3.6.2) and Canadian Terrestrial Ecosystem Model (v.2.1) (CLASS-CTEM) *Geosci. Model Dev.* **12** 4443–67

Meyer G, Humphreys E R, Melton J R, Cannon A J and Lafleur P M 2021 Simulating shrubs and their energy and carbon dioxide fluxes in Canada's Low Arctic with the Canadian Land Surface Scheme Including Biogeochemical Cycles (CLASSIC) *Biogeosciences* **18** 3263–83

Monteith J and Unsworth M 2013 *Principles of Environmental Physics: Plants, Animals, and the Atmosphere* (Academic)

Nazarbakhsh M, Ireson A M and Barr A G 2020 Controls on evapotranspiration from jack pine forests in the Boreal Plains Ecozone *Hydrol. Process.* **34** 927–40

Nilsson M-C and Wardle D A 2005 Understory vegetation as a forest ecosystem driver: evidence from the northern Swedish boreal forest *Front. Ecol. Environ.* **3** 421–8

Pastorello G, Trotta C, Canfora E, Chu H, Christianson D, Cheah Y-W, Poindexter C, Chen J, Elbashandy A and Humphrey M 2020 The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data *Sci. Data* **7** 225

Pearson R G, Phillips S J, Loranty M M, Beck P S, Damoulas T, Knight S J and Goetz S J 2013 Shifts in Arctic vegetation and associated feedbacks under climate change *Nat. Clim. Change* **3** 673–7

Pelletier J D, Broxton P D, Hazenberg P, Zeng X, Troch P A, Niu G-Y, Williams Z, Brunke M A and Gochis D 2016 A gridded global data set of soil, intact regolith, and sedimentary deposit thicknesses for regional and global land surface modeling *J. Adv. Modelling Earth Syst.* **8** 41–65

Piao S, Luyssaert S, Ciais P, Janssens I A, Chen A, Cao C, Fang J, Friedlingstein P, Luo Y and Wang S 2010 Forest annual carbon cost: a global-scale analysis of autotrophic respiration *Ecology* **91** 652–61

Purdy A, Fisher J, Goulden M and Famiglietti J 2016 Ground heat flux: an analytical review of 6 models evaluated at 88 sites and globally *J. Geophys. Res. Biogeosci.* **121** 3045–59

Qu B *et al* 2022 A boreal forest model benchmarking dataset for North America: a case study with the Canadian Land Surface Scheme including Biogeochemical Cycles (CLASSIC) [Data set] *Zenodo* (https://doi.org/10.5281/zenodo.7266010)

Reichstein M, Falge E, Baldocchi D, Papale D, Aubinet M, Berbigier P, Bernhofer C, Buchmann N, Gilmanov T and Granier A 2005 On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm *Glob. Change Biol.* **11** 1424–39

Richardson A D and Hollinger D Y 2007 A method to estimate the additional uncertainty in gap-filled NEE resulting from long gaps in the CO_2 flux record *Agric. For. Meteorol.* **147** 199–208

Richardson A D, Hollinger D Y, Burba G G, Davis K J, Flanagan L B, Katul G G, Munger J W, Ricciuto D M, Stoy P C and Suyker A E 2006 A multi-site analysis of random error in tower-based measurements of carbon and energy fluxes *Agric. For. Meteorol.* **136** 1–18

Rogers A, Medlyn B E, Dukes J S, Bonan G, Caemmerer S, Dietze M C, Kattge J, Leakey A D, Mercado L M and Niinemets Ü 2017 A roadmap for improving the representation of photosynthesis in Earth system models *New Phytol.* **213** 22–42

Rogers A, Serbin S P and Way D A 2021 Reducing model uncertainty of climate change impacts on high latitude carbon assimilation *Glob. Change Biol.* **28** 1222–47

Schaefer K, Schwalm C R, Williams C, Arain M A, Barr A, Chen J M, Davis K J, Dimitrov D, Hilton T W and Hollinger D Y 2012 A model-data comparison of gross primary productivity: results from the North American Carbon Program site synthesis *J. Geophys. Res. Biogeosci.* **117** G03010

Schmid S, Thürig E, Kaufmann E, Lischke H and Bugmann H 2006 Effect of forest management on future carbon pools and fluxes: a model comparison *For. Ecol. Manage.* **237** 65–82

Schuur E A and Abbott B 2011 High risk of permafrost thaw *Nature* **480** 32–33

Schuur E A and Mack M C 2018 Ecological response to permafrost thaw and consequences for local and global ecosystem services *Annu. Rev. Ecol. Evol. Syst.* **49** 279–301

Schuur E A, McGuire A D, Schädel C, Grosse G, Harden J, Hayes D, Hugelius G, Koven C, Kuhry P and Lawrence D 2015 Climate change and the permafrost carbon feedback *Nature* **520** 171–9

Schwalm C R, Schaefer K, Fisher J B, Huntzinger D, Elshorbany Y, Fang Y, Hayes D, Jafarov E, Michalak A M and Piper M 2019 Divergence in land surface modeling: linking spread to structure *Environ. Res. Commun.* **1** 111004

Seiler C, Melton J R, Arora V K and Wang L 2021 CLASSIC v1. 0: the open-source community successor to the Canadian Land Surface Scheme (CLASS) and the Canadian Terrestrial Ecosystem Model (CTEM)–part 2: global benchmarking *Geosci. Model Dev.* **14** 2371–417

Sellers P, Hall F, Margolis H, Kelly B, Baldocchi D, den Hartog G, Cihlar J, Ryan M G, Goodison B and Crill P 1995 The Boreal Ecosystem–Atmosphere Study (BOREAS): an overview and early results from the 1994 field year *Bull. Am. Meteorol. Soc.* **76** 1549–77

Soloway A D, Amiro B D, Dunn A L and Wofsy S C 2017 Carbon neutral or a sink? Uncertainty caused by gap-filling long-term flux measurements for an old-growth boreal black spruce forest *Agric. For. Meteorol.* **233** 110–21

Stofferahn E, Fisher J B, Hayes D J, Schwalm C R, Huntzinger D N, Hantson W, Poulter B and Zhang Z 2019 The Arctic-Boreal vulnerability experiment model benchmarking system *Environ. Res. Lett.* **14** 055002

Taylor K E 2001 Summarizing multiple aspects of model performance in a single diagram *J. Geophys. Res. Atmos.* **106** 7183–92

Tramontana G, Jung M, Schwalm C R, Ichii K, Camps-Valls G, Ráduly B, Reichstein M, Arain M A, Cescatti A and Kiely G

2016 Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms *Biogeosciences* **13** 4291–313
Tuomi M, Vanhala P, Karhu K, Fritze H and Liski J 2008 Heterotrophic soil respiration—comparison of different models describing its temperature dependence *Ecol. Modelling* **211** 182–90
Turetsky M R, Bond-Lamberty B, Euskirchen E, Talbot J, Frolking S, McGuire A D and Tuittila E-S 2012 The resilience and functional role of moss in boreal and arctic ecosystems *New Phytol.* **196** 49–67
Ueyama M, Iwata H and Harazono Y 2014 Autumn warming reduces the CO_2 sink of a black spruce forest in interior Alaska based on a nine-year eddy covariance measurement *Glob. Change Biol.* **20** 1161–73
Ueyama M, Kudo S, Iwama C, Nagano H, Kobayashi H, Harazono Y and Yoshikawa K 2015 Does summer warming reduce black spruce productivity in interior Alaska? *J. For. Res.* **20** 52–59
Ukkola A M, Abramowitz G and De Kauwe M G 2021 A flux tower dataset tailored for land model evaluation *Earth Syst. Sci. Data Discuss.* **14** 449–61
Verseghy D L 2017 CLASS—the Canadian land surface scheme (v. 3.6.2) (Climate Research Division, Science and Technology Branch, Environment Canada)
Vuichard N and Papale D 2015 Filling the gaps in meteorological continuous data measured at FLUXNET sites with ERA-interim reanalysis *Earth Syst. Sci. Data* **7** 157–71
Walker X J, Baltzer J L, Bourgeau-Chavez L, Day N J, Dieleman C M, Johnstone J F, Kane E S, Rogers B M, Turetsky M R and Veraverbeke S 2020 Patterns of ecosystem structure and wildfire carbon combustion across six ecoregions of the North American boreal forest *Front. For. Glob. Change* **3** 87
Whittaker R H 1970 Communities and ecosystems *Communities and ecosystems*
Wickland K P, Neff J C and Harden J W 2010 The role of soil drainage class in carbon dioxide exchange and decomposition in boreal black spruce (*Picea mariana*) forest stands *Can. J. For. Res.* **40** 2123–34
Williams M, Richardson A D, Reichstein M, Stoy P C, Peylin P, Verbeeck H, Carvalhais N, Jung M, Hollinger D Y and Kattge J 2009 Improving land surface models with FLUXNET data *Biogeosciences* **6** 1341–59
Yuan F, Arain M A, Barr A G, Black T A, Bourque C P A, Coursolle C, Margolis H A, McCaughey J H and Wofsy S C 2008 Modeling analysis of primary controls on net ecosystem productivity of seven boreal and temperate coniferous forests across a continental transect *Glob. Change Biol.* **14** 1765–84