Generalized additive models of climatic and metabolic controls of sub-annual variation in pCO_2 in productive hardwater lakes

E. Wiik[1,2], H. A. Haig[1], N. M. Hayes[1], K. Finlay[1], G. L. Simpson[3], R. J. Vogt[4], P. R. Leavitt[1,3]

[1]Limnology Laboratory, Department of Biology, University of Regina, Regina, Saskatchewan, S4S 0A2, Canada
[2]School of Environment, Natural Resources and Geography, Bangor University, Bangor, Gwynedd, LL57 2UW, UK
[3]Institute of Environmental Change and Society, University of Regina, Regina, Saskatchewan, S4S 0A2, Canada
[4]Pavillon des sciences biologiques (SB), Université du Québec à Montréal, Montréal (Québec), H2X 1Y4, Canada

Key Points:

- In Canadian hardwater prairie lakes, calculated CO_2 fluxes correlate mostly with pH, not DIC
- Intra-annual CO_2 correlates with algal abundance (-CO_2) and prolonged clearwater phases (+CO_2)
- CO_2 influx increases with drier weather conditions, and is reduced with extreme N loading

Corresponding author: Emma Wiik, `e.wiik@bangor.ac.uk`, `e.wiik@ucl.ac.uk`

Abstract

Spatio-temporal variation in climate and weather, allochthonous carbon loads, and autochthonous factors such as lake metabolism (photosynthesis and respiration) interact to regulate atmospheric CO_2 exchange of lakes. Understanding this interplay in diverse basin types at different timescales is required to adequately place lakes into the global carbon cycle, and predict CO_2 flux through space and time. We analyzed 18 years of data from seven moderately hard lakes in an agricultural prairie landscape in central Canada. We applied generalized additive models and sensitivity analyses to evaluate the roles of metabolic and climatic drivers in regulating CO_2 flux at the intra-annual scale. In all basins, at mean conditions with respect to other predictors, metabolic controls resulted in uptake of atmospheric CO_2 when surface waters exhibited moderate primary production, but released CO_2 only when primary production was very low ($5 - 13\ \mu g\,L^{-1}$) or when dissolved nitrogen was elevated ($>2000\ \mu g\,L^{-1}$), implying that respiratory controls offset photosynthetic CO_2 uptake under these conditions. Climatically, dry conditions increased the likelihood of ingassing, likely due to evaporative concentration of base cations and/or reduced allochthonous carbon loads. While more research is required to establish the relative importance of climate and metabolism at other time scales (diel, autumn/winter), we conclude that these hard fresh waters characteristic of continental interiors are mainly affected by metabolic drivers of pCO_2 at daily-monthly timescales, are climatically controlled at interannual intervals, and are more likely to in-gas CO_2 for a given level of algal abundance, than are softwater, boreal ecosystems.

1 Introduction

It is widely accepted that lakes are important nodes that process terrestrial carbon (C) and influence global C fluxes [*Cole et al.*, 2007; *Downing et al.*, 2008; *Tranvik et al.*, 2009]. However, improved understanding of regulatory mechanisms which underlie trends and variability among lentic systems is needed to improve predictions of how lakes will both contribute and respond to future climate change [*Prairie*, 2008; *Tranvik et al.*, 2009]. In particular, there remains high regional and temporal variation in the mechanisms regulating lake pCO_2, despite increasing efforts to synthesize and upscale in-lake CO_2 levels and greenhouse gas fluxes. In part, this variability reflects the wide range of analytical methods and study time frames, varying from instantaneous estimates of regional lakes [*Duarte et al.*, 2008; *Lapierre and del Giorgio*, 2012] to decadal analyses of individual sites [*Finlay et al.*, 2015; *Perga et al.*, 2016]. Furthermore, certain lake types (e.g., hardwater and saline) are understudied relative to softwater boreal systems. Variability in the importance of contrasting regulatory mechanisms (e.g., broad-scale climatic drivers vs. local metabolic factors) across temporal and spatial scales can obscure the hierarchical relationships among control processes, which in turn limits insights derived from upscaled, ecosystem-level comparisons and global estimates.

Interannual and decadal trends in lake pCO_2 are modulated by many interacting variables, primarily acting at the landscape scale through climatic and meterological drivers. For example, changes in precipitation affects transport of solutes such as dissolved organic (DOC) and inorganic carbon (DIC), which in turn alter lake water CO_2 content [*Ojala et al.*, 2011]. In the case of organic forms of carbon, higher substrate supply tends to elevate microbial respiration [*Maberly et al.*, 2013; *Ducharme-Riel et al.*, 2015], whereas increased DIC can either increase or reduce in situ pCO_2 in hardwater systems, depending on ambient pH and alternate buffering mechanisms [e.g., *Baehr and DeGrandpre*, 2004; *Knoll et al.*, 2013]. Additionally, landscape-scale variation in irradiance (e.g., cloud cover) or air temperature [*O'Reilly et al.*, 2015] can lead to evaporative concentration of lakes [*Pham et al.*, 2009] and consequent changes in parameters regulating pCO_2 (DIC, DOC, nutrients, etc.). For example, in continental Canadian hardwater lakes, interannual variability in both temperature and precipitation has affected pH and CO_2 flux via effects of ice-off timing [*Finlay et al.*, 2015], DIC content [*Pham et al.*, 2009], and regional hydrology [*Bonsal and Shabbar*, 2008; *van der Kamp et al.*, 2008].

Metabolic processes are likely to be paramount in regulating atmospheric exchange of greenhouse gases at scales of hours to days. For example, water-column pCO_2 typically increases overnight as photosynthesis becomes light-limited and respiration continues [*Raymond et al.*, 2013; *Liu et al.*, 2016]. In softwater reservoirs, these diel metabolic patterns can account for ca. 30% of total variation in CO_2 flux over a summer season [*Morales-Pineda et al.*, 2014]. In general, larger diel amplitudes of CO_2 content are found as lake productivity increases [*Hanson et al.*, 2003; *Shao et al.*, 2015; *Morales-Pineda et al.*, 2014], suggesting that multiple temporal scales may be needed to evaluate CO_2 regulation in productive lakes.

At intermediate timescales, trends in lake pCO_2 are likely to be regulated by a combination of metabolic and climatic mechanisms [*Morales-Pineda et al.*, 2014]. For example, metabolic controls underlie seasonal trends in dimictic temperate lakes when, in winter, CO_2 accumulates under ice [*Denfeld et al.*, 2015], causing springtime efflux of CO_2 during ice melt and lake overturn. Reduced pCO_2 occurs in summer when the water column is stable and primary production increases, whereas pCO_2 increases during fall as respiratory products in the hypolimnion are mixed into surface waters [*Alin and Johnson*, 2007; *Stets et al.*, 2009; *Ducharme-Riel et al.*, 2015; *Marcé et al.*, 2015]. These seasonal patterns can be disrupted by climatic or meteorological events such as passing storms or heat waves [*Maberly*, 1996; *Klug et al.*, 2012; *Audet et al.*, 2017], or be dampened in polymictic lakes where CO_2 exhibits more limited seasonal variation [*Jonsson et al.*, 2003].

While metabolic controls of CO_2 also operate at seasonal scales in hardwater lakes [*Striegl and Michmerhuizen*, 1998], their influence can be overrun by landscape-level controls of solute loading [*Anderson et al.*, 1999; *Sobek et al.*, 2005; *Christensen et al.*, 2013; *Knoll et al.*, 2013; *Marcé et al.*, 2015]. For example, lakes with strong groundwater influences can have high allocthonous supplies of DIC and exhibit super-saturation of CO_2, particularly in regions close to the groundwater entry points [*Stets et al.*, 2009]. On the other hand, the high pH and alkalinity of hardwater lakes also buffers against large fluctuations in pH [*Duston et al.*, 1986; *Hanson et al.*, 2003], leading to smaller amplitudes of both pH and CO_2 than exist in softwater lakes. Therefore, especially in polymictic hardwater lakes without strong stratification, hypolimnetic CO_2 accumulation should be relatively low and uniform throughout the year, with the net direction of atmospheric CO_2 exchange depending on climate effects on solute loading and metabolism. Thus, seasonal patterns of CO_2 content in hard-water lakes may contrast sharply from those known from dimictic boreal systems.

Here, we use generalized additive models (GAMs) and sensitivity analysis to quantify the effects of climatic and metabolic parameters in regulating intra-annual variability in pCO_2 of hardwater lakes in the sub-humid Canadian interior. Using bi-weekly data for 18 years in seven lakes, we sought to determine: 1) When and to what extent metabolic factors (photosynthesis and respiration) were regulating lakewater pCO_2 and CO_2 flux; 2) Whether local meteorology and global climatic factors contribute to intra-annual CO_2 flux variability, and; 3) How consistent the drivers of CO_2 flux were among study lakes that varied more than 10-fold in size, productivity, and catchment area. Improved understanding of the relative importance of biotic and abiotic controls of CO_2 flux in hardwater lake types is critical to achieving a predictive understanding of the role of freshwater ecosystems in global carbon cycles.

2 Methods

2.1 Study sites

The seven study sites are situated within the Qu'Appelle River catchment (ca. 52,000 km^2) in the northern Great Plains of southern Saskatchewan, Canada (Fig. 1). The region has a sub-humid continental climate and is hydrologically reliant on water originating from the Rocky Mountains as well as local snowmelt [*Bonsal and Shabbar*, 2008; *Pham et al.*, 2009]. The South Saskatchewan River feeds the Qu'Appelle River system via Lake Diefenbaker reservoir (D). Water flows eastward from the main reservoir through a chain of lakes includ-

ing Buffalo Pound (B), Pasqua (P), Katepwa (K), and Crooked (C) Lakes. Wascana (W) and Last Mountain (L) Lakes are situated on tributaries that feed into the Qu'Appelle river system upstream of Pasqua Lake. All lakes receive diffuse nutrient sources from agriculture, with the wastewater treatment plants from the cities of Regina and Moose Jaw acting as point sources of nutrients to Pasqua and eastern basins [*Hall et al.*, 1999]. All lakes are dammed to variable extent, and Buffalo Pound and Diefenbaker are actively managed reservoirs. For simplicity, we refer to all sites as lakes.

Median nutrient concentrations are generally elevated (Table 1), including total dissolved nitrogen (TDN) ($0.96\ \mathrm{mg\,N\,L^{-1}}$) and total dissolved phosphorus (TDP) ($106\ \mu\mathrm{g\,L^{-1}}$), resulting in high algal abundance as chlorophyll *a* (Chl *a*) (median $16\ \mu\mathrm{g\,L^{-1}}$) and mesotrophic to hypereutrophic conditions [*Hall et al.*, 1999; *Finlay et al.*, 2009]. Compared with saline lakes worldwide [e.g., *Duarte et al.*, 2008], Qu'Appelle lakes have moderate DIC (median = $45\ \mathrm{mg\,L^{-1}}$) and conductivity (median = $1050\ \mu\mathrm{S\,L^{-1}}$), but rather high pH (median = 8.8) (Fig. 2a). DOC concentrations are moderate (median $11.5\ \mathrm{mg\,L^{-1}}$). Temporal variation in many major chemical variables such as pH is highly synchronous across the sites (Fig. 2b; *Vogt et al.* [2011]) (see Fig. A.1 for intra-annual variability in variables relating to nutrient status and lake metabolism).

2.2 Long-term limnological sampling

Biweekly limnological sampling of pH, temperature, dissolved oxygen, conductivity, salinity, DIC, DOC, Chl *a*, TDN and metabolic bioassay estimates (primary production, respiration) followed methods outlined in *Finlay et al.* [2009]. Briefly, pH was measured at the lake surface, while oxygen, temperature, conductivity and salinity were recorded at 1 m depth using YSI-85 multi-probe meters (YSI, Inc., Yellow Springs, OH). DIC, DOC, Chl *a*, TDN and metabolic bioassay samples used depth-integrated water samples pooled from 2-L Van Dorn sampler casts taken at 0.5 m intervals.

Filtered water ($0.45\ \mu\mathrm{m}$ pore size) was used for DIC and DOC analyses using a total carbon analyser (Shimadzu 500A), while TDN was measured by photocombustion, both following Environment Canada protocols [*Environment Canada*, 1979]. Chl *a* was determined trichromatically from particulate organic matter (POM) collected on $1.2\ \mu\mathrm{m}$ pore Whatman GF/C glass fiber filters following *Jeffrey and Humphrey* [1975] and following extraction using 80% acetone: 20% methanol, by volume. The wavelength-specific absorbance was quantified using a Hewlett Packard model 8452A photodiode array spectrophotometer (1996–2004) or an Agilent model 8453 UV-Visible spectrophotometer (2005–2014).

Metabolic estimates of gross primary production, net primary production and respiration were based on changes in oxygen concentration following incubation of whole water samples in light and dark glass bottles [*Finlay et al.*, 2009]. All analyses were run in triplicate using screened ($243\ \mu\mathrm{m}$ mesh), depth-integrated water following *Howarth and Michaels* [2000]. Incubations occurred for 24 h at ambient lake temperature and under a 12-hour light/dark cycle with $450\ \mu\mathrm{mol}$ quanta $\mathrm{m^{-2}\ s^{-1}}$, comparable to that recorded in situ at Secchi depth using a profiling radiometer [*Finlay et al.*, 2009].

Sampling occurred primarily from May 1st to August 31st between the hours of 0900 and 1300, with ca. 5% of sampling dates occurring earlier in spring or later in autumn. This long-term ecological research program began sampling in 1994, but for reasons related to data availability, we restricted this study to data from either 1996 (most lakes) or 2004 (Pasqua) to 2014, inclusive.

2.3 CO_2 flux calculation

In the absence of direct measurements of CO_2, we relied on calculated fluxes which approximate real values particularly well in high-alkalinity lakes [*Abril et al.*, 2015] (such as our study sites), where there are strong chemical relationships between pH and dissolved

CO_2 [*Soumis et al.*, 2004, R^2 = 0.81]. Calculated values are widely applied in the absence of measurements, particularly when long-term or broad spatial data are being examined (e.g. *Duarte et al.* [2008]; *Seekell and Gudasz* [2016])

The procedure for calculating CO_2 fluxes and pCO_2 followed *Finlay et al.* [2009]. Briefly, CO_2 concentrations ($[CO_2]$) were calculated based on DIC concentrations (depth-integrated samples) and pH (surface), with correction for ionic strength and water temperature measured at 1 m depth [*Stumm and Morgan*, 1996]. Partial pressure of CO_2 (Pa) was estimated using Henry's Law constant [*Kling et al.*, 1992], and chemically enhanced CO_2 flux (mmol $m^{-2}\,d^{-1}$) was calculated following *Cole et al.* [1998]:

$$\text{net daily } CO_2 \text{ flux} = \alpha k([CO_{2_{\text{lake}}}] - [CO_{2_{\text{sat}}}]) \tag{1}$$

where: in-lake CO_2 concentration for $[CO_{2_{lake}}]$ refers to surface water; saturation levels $[CO_{2_{sat}}]$ refer to equilibrium with the atmosphere; α is the chemical enhancement of CO_2 flux at high pH [*Hoover and Berkshire*, 1969], calculated following *Wanninkhof and Knox* [1996], and; k is piston velocity (cm h^{-1}) following *Cole et al.* [1998], relating k to wind speed and temperature [*Wanninkhof*, 1992].

The effect of an alternative piston velocity was evaluated by including the effect of lake surface area on piston velocity and therefore CO_2 flux in our sensitivity analysis (See Statistical methods) [equations for k derived from Table 2, Model B; *Vachon and Prairie*, 2013]. We did not have data to account for wind direction, which would plausibly incur errors in lake area-based estimates of gas transfer for e.g. Katepwa (North-South orientation) vs Pasqua (West-East orientation). Overall, however, the influence of lake area on chemically enhanced flux was subsidiary to pH and therefore not considered further in this paper.

Complete data for calculating CO_2 flux were available from 1996 for all lakes except Pasqua at which sampling began in 2004. Variables included temperature, pH, conductivity, salinity, DIC, wind speed, air pressure and atmospheric pCO_2. Observations with any one missing variable were omitted, leaving 991 data points for modeling. Hourly wind speed and air pressure were acquired from publicly available Environment Canada (EC) data (`http://climate.weather.gc.ca/`) using Regina stations 4016560 and 4016566 (Climate IDs) which had complete records for the study period. Using one weather station location for all lakes was deemed acceptable as existing records from other weather stations were found highly correlated. Two-week average wind speed was calculated to smooth out brief effects of extreme weather events. Monthly averages of air pressure (EC), and Mauna Loa atmospheric pCO_2 (Earth System Research Laboratory, `http://www.esrl.noaa.gov/gmd/ccgg/trends/data.html`) were used.

2.4 Statistical methods

All statistical analyses were performed using R version 3.2.5 [*R Development Core Team*, 2016], using packages mgcv [*Wood*, 2011, 2017] and pse [*Chalom and de Prado*, 2016]. R code is available at `https://github.com/simpson-lab/jgr-co2-flux`.

Our analytical approach follows a few key underlying considerations. Since CO_2 flux was estimated from water chemistry and physical variables and not measured directly, we avoided any approach that would circularly include these 'calculation variables' as metabolic or climatic proxy predictors of CO_2 flux. Furthermore, we were specifically interested in which of these calculation variables correlate the most with CO_2 flux in our study region. In this regard we note that, although the real, rather than estimated, relationship between these variables and CO_2 flux is unknown, this step can identify which variable is key to proxy CO_2 flux in our region (and conversely, which variables are not). Therefore, we first quantified the influence of the calculation variables on estimated CO_2 flux ('influence' here used in the regression sense of changes in x influencing estimates of y, rather than a directional causal

sense). Secondly, we regressed our metabolic and climatic variables of interest against the variable that accounted for most of this variation. The second step allowed us to use a measured, rather than estimated, response variable, reducing the amount of imprecision in our regression values. We were then able to relate these values back into CO_2 flux estimates using the results from the first step, thereby avoiding presenting misleadingly precise results for CO_2 flux itself.

2.4.1 Variable selection

Metabolic variables were selected from various estimates of lake production and respiration to achieve the greatest availability over the data period. In the case of highly correlated variables, we modelled only a single variable, so in our case, respiration (R) was selected over net and gross primary production, whereas TDN was retained over TDP (at most times at most study sites, N limitation exceeds P limitation [*Patoine et al.*, 2006]). Ultimately, five metabolic variables were selected for modeling, including in situ O_2 (respiration/photosynthesis), DOC (potential effects on respiration), Chl *a* (algal biomass or production), R (respiration), and TDN (nutrient availability). Chl *a*, TDN, and DOC were log10-transformed to approximate a normal error distribution.

To capture the major climatic processes most likely to influence lake CO_2 via solute and nutrient loading (hydrological processes, evapotranspiration), we included both broad drivers of intra-annual climate and more local, instantaneous proxies for evaporation-precipitation balance. Variables included the Southern Oscillation Index (SOI) and Pacific Decadal Oscillation (PDO), metrics of climate systems which strongly influence regional precipitation and temperature patterns, either alone or in combination [*Bonsal and Shabbar*, 2008; *Pham et al.*, 2009; *Shabbar and Yu*, 2012]. Both indices were included as three-month averages, six months prior to sample collection, to account for the lags between the regions of observation and effect [*Pomeroy et al.*, 2007; *Shabbar et al.*, 2011]. Monthly values were obtained from the National Oceanic and Atmospheric Administration (NOAA) (`http://www.cpc.noaa.gov/data/indices/soi`) and the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) (`http://research.jisao.washington.edu/pdo/PDO.latest`). Because regional precipitation is highly localized (lake-specific; *Vogt et al.* [2011]) and weather stations were not adjacent to our study sites, we did not attempt to use data from weather stations to estimate rainfall. Instead, Standardized Precipitation Evapotranspiration Index (SPEI) values for each site (0.5 degree spatial resolution) were obtained from the Consejo Superior de Investigaciones Cientficas (CSIC) Global SPEI database (`http://sac.csic.es/spei/database.html`) [*Vicente-Serrano et al.*, 2016]. Index values were calculated using a two-month 'memory' (autocorrelation) to account for temporal variation in soil drying and hydration.

2.4.2 Sensitivity analysis

Given the absence of direct measurements, we analysed data to select the best proxy of CO_2 in our climatic-metabolic model by simulating the sensitivity of calculated CO_2 flux to changes in pH, conductivity, salinity, water temperature, DIC, wind speed, atmospheric pCO_2, and local air pressure. A sensitivity analysis was used for this purpose because it shows the magnitude of individual variable contributions to estimate CO_2 flux for multiple combinations of variables and values. Further, this method allows us to perform multi-step calculations while controlling for underlying data correlations [*Chalom and de Prado*, 2015].

Differences among lakes in the relative contribution of variables to calculated CO_2 flux were tested by comparing an analysis conducted for all lakes combined, with those for each lake individually. Specifically, we used a latin hypercube sampling (LHS) approach [*Chalom and de Prado*, 2015] and generated realistic data variations of all variables for each lake based on their observed variation over the sampling period (n = 500 per simulation). Rank correlations were selected, rather than a linear analysis among variables, to account for potential

nonlinear relationships between predictors and responses. The output metric (partial rank correlation coefficient: PRCC), for any one variable, controls for the effect of all other variables by reflecting the correlation between the unexplained part of the outcome, given all other variables, and the unexplained part of one variable, given all other variables (i.e., a correlation between residuals).

2.4.3 Generalized additive models

pH was the strongest correlate with calculated CO_2 flux based on sensitivity analysis (see Results) and, therefore, was carried forward to evaluate the effects of selected metabolic and climatic variables on CO_2 flux. Here, we applied generalized additive models (GAM), which account for nonlinear relationships between predictors and responses [*Hastie and Tibshirani*, 1990; *Wood*, 2017] (Section 2.4.4). GAMs also allowed us to include Year and Lake as random effects to account for between-lake and inter-annual variations known to be important [*Finlay et al.*, 2009, 2015]. The resolutions of all other predictors also link with the resolution of variability they are able to explain: e.g., biweekly predictors can explain pH variation at a within-month scale, while monthly predictors can only explain pH variation occurring at a between-month scale. Temporal structure within the climatic-metabolic model was visualised by plotting term contributions to pH against time.

The first model, which evaluated the degree to which lakes differed in their relationship between CO_2 and pH was formulated as follows, for y = CO_2 flux,

$$y = \beta_0 + f(\text{pH}) + f_{\text{lake}}(\text{pH}) + \alpha_{\text{lake}} + \gamma_{\text{year}} + \varepsilon \tag{2}$$

Here, the effect of pH was modelled both globally ($f(\text{pH})$) and by lake ($f_{\text{lake}}(\text{pH})$, while terms α and γ were random effects of lake and year, respectively, and ε was the error term. The global and lake-specific effects of pH were identified via different orders of quadratic penalties on their respective basis expansions. The global function of pH ($f(\text{pH})$) was subject to the usual second-order penalty whereby the wiggliness penalty was on the second derivative of a fitted spline. First-order penalties were used for the lake-specific splines so that the penalty applied to departure from a flat or zero function. This approach had the effect of making each $f_{\text{lake}}(\text{pH})$ represent the departure of each lake from the global pH effect. Smoothness parameters for f and f_{lake} were chosen using restricted maximum likelihood (REML) selection [*Wood*, 2011]. Lake-specific effects of pH on CO_2, ($f_{\text{lake}}(\text{pH})$), were only retained when they were assessed to be significantly different from a zero (flat) function. Therefore, lake-specific splines retained reflect regional heterogeneity (objective 3) between the study sites. pH was selected for a combined metabolic and climatic GAM to explore sub-annual controls of CO_2 flux.

The second model, which quantified the influence of climatic and metabolic variables on pH, followed the principles outlined above for the first model. For y = pH,

$$y = \beta_0 + \sum_{j=1}^{J} [f(x_j) + f_{\text{lake}}(x_j)] + f(\text{PDO}, \text{SOI}) + \alpha_{\text{lake}} + \gamma_{\text{year}} + \varepsilon \tag{3}$$

where: x_j is the jth metabolic (TDN, DOC, Chl a, O_2) or climatic (SPEI) covariate, $f(\text{PDO}, \text{SOI})$ is a 2-D tensor product spline combining the main and interactive effects of PDO and SOI; α and γ are random effects of lake and year, and; ε is the error term. As above, the unique effects of the x_j for each lake were incorporated through inclusion of separate difference splines for each lake ($f_{\text{lake}}(x_j)$) employing first-order wiggliness penalties. REML smoothness selection was used as described above. Where model terms were marginally significant, likelihood ratio tests were used to determine whether a model including the terms was justifiable.

Preliminary runs suggested that co-linearity between DOC and TDN was sufficient to confound results, and argued for retaining only one predictor (DOC), based on both internal model Wald tests and Akaike and Bayesian Information Criteria (AIC; BIC). However, due to TDN being a significant correlate absent from the final model, the model replacing DOC with TDN is also used in this paper to portray the relationship between TDN and pH.

3 Results

3.1 The sensitivity of CO_2 flux to variables used in its calculation

Sensitivity analysis showed that pH explained the greatest amount of variation in CO_2 flux (PRCC = -0.96) followed by DIC (PRCC = 0.51) for all lakes (Table 2, Fig. B.1). This sequence was also retained in the simulations for individual lakes; however, DIC was more influential in some lakes (B, C, D, L) than in others (K, P, W). Overall, the importance of DIC was small (Table 2) and sensitive to which simulation data were used for analysis (not shown).

Generalized additive modeling echoed the results of the sensitivity analysis and showed that pH was the main correlate of CO_2 flux (Fig. 3). This model explained 97% of deviance in CO_2 flux, while the use of DIC as an additional term only explained a further 1% of variation (and an equivalent model with DIC, not pH, explained only 30% of flux variation; not shown).

Lakes were predicted to in-gas atmospheric CO_2 above a pH of 8.8, the median pH over the whole data set, while no net atmospheric exchange occurred around pH 8.7. Generally more productive lakes (K,P,W) were significantly different from less productive sites (B, C, D, L) based on GAM analysis of the relationship between pH and CO_2, primarily at the high and low ends of pH (<10% of all observations). These groups of lake also differed in the extent to which DIC content tended to influence sensitivity analyses (Table 2).

3.2 Metabolic and climatic regulation of pH

GAM analysis explained 43% of historical deviance in pH, mainly due to climatic and metabolic parameters (Figs 4 – 6). Significant predictors of pH included Chl *a* ($p < 0.001$), PDO∗SOI ($p < 0.001$), Lake + Year ($p < 0.001$), oxygen ($p = 0.0108$), DOC ($p = 0.0137$) and SPEI ($p = 0.0122$). The only variable for which individual lake splines were significant was Chl *a*. In all cases, R was insignificant and removed from the model. The ranges of pH over which the metabolic and climatic variables exerted control were variable, and in decreasing order included PDO∗SOI (ca 8.5-10), Chl *a* (8.5-9.6), oxygen (8.6-8.9), DOC (8.75-8.9), and SPEI (8.9-9.1), approximately (see uncertainties at the edge of prediction: Figs 4, 6). Using all measured combinations of our predictors, i.e. the empirical data, our model pH predictions encompass a range from 7.8 to 10 ($\pm$ errors), which does not capture the full range of observed pH (7 to 10.9) (Figs C.1, C.2).

Concentrations of Chl *a* were correlated positively with pH, with low algal abundance ($< 5 - 13\ \mu\mathrm{g\,L^{-1}}$) occurring when depressed pH correlates with out-gassing of CO_2 when all other predictors were held at their mean (Fig. 4). Results from the two small, shallow lakes (W, B) were significantly different from other basins in that both increases and declines in Chl *a* had comparatively strong relationships with pH. In general, pH increased with oxygen saturation, with CO_2 in-gassing at supersaturated oxygen concentrations $>9 - 10\ \mathrm{mg\,L^{-1}}$). CO_2 efflux occurred only at low oxygen concentrations (<5% of all observations which were lower than ca $5\ \mathrm{mg\,L^{-1}}$ when all other predictors held at their mean). Finally, DOC was positively correlated with pH, particularly in the range where elevated pH favoured influx of CO_2.

In the alternative model where DOC was replaced with TDN, TDN had a slight positive relation with pH up to concentrations of ca $1100\ \mu\mathrm{g\,N\,L^{-1}}$ above which pH declined consistently (Fig. 5). Uncertainties in the effect of TDN on pH were high at both ends of

the range due to low observation frequency; however, extremely high values of TDN ($> 2000 - 6500\ \mu\mathrm{g\,N\,L^{-1}}$) co-occurred with pH values that correspond with CO_2 efflux.

Broader-scale climate variables PDO and SOI had stronger relationships with pH than did SPEI. The highest pH values were associated with the most negative SOI and positive PDO (Figs 4 a-c), which typically indicate warm and dry conditions. In contrast variation in SPEI had a limited effect on pH (ca 0.2 units) and was associated with above-mean pH at the low and high end of its range (Fig. 6). Low pH was particularly common when PDO was low and wet conditions predominate [Bonsal and Shabbar, 2008]. PDO had a more complex multi-modal relationship with pH than did SOI, which was more linear (Figs 4 b-c). For a given PDO, increasing SOI shifted the position of the spline. In general, SOI had a positive relationship with pH except at high PDO when high pH occurred also at low SOI values (Figs 4 b-c). Overall, the range in climatic index values during the observation period was similar to that recorded during the past century (PDO mostly within -2,2; SOI mostly within -2.5, 2.5, SPEI mostly within -2,2).

Consistent long-term intra-annual trends were apparent for the metabolic variables Chl *a*, and oxygen (Fig. 7), but not DOC or the climatic variables SPEI and PDO$*$SOI. Chl *a* increased in positive effect on pH over the summer in most lakes except during the clear-water phase in June. Below-average pH at low Chl *a* occurred consistently at the least productive site, Lake Diefenbaker. Oxygen effects in four lakes (C, K, B, D) were most negative towards the end of the summer.

4 Discussion

Given the importance of climate and ice-cover duration in determining annual mean pH and CO_2 flux in these hard-water lakes [*Finlay et al.*, 2015], we sought to determine whether metabolic factors would emerge as a driving factor at an intra-annual timescale. While we found similar controls also at sub-annual timescales (high coherence within the region, pH the most significant predictor of CO_2) (Table 2, Fig. 3), metabolic controls were important in determining the balance between high likelihoods of influx (pH>8.8) and efflux (pH<8.7) of CO_2 (Fig. 4). Lake metabolism, as measured using algal abundance (Chl *a*), was a key parameter controlling whether lakes acted as C sources or sinks within any given year.

4.1 The role of lake metabolism in directing pH and CO_2 flux

There was strong evidence for metabolic control of pH and thereby CO_2 flux both at the high and low ends of a gradient of nutrient concentration when either primary production was insufficient to sequester CO_2 or it seemed offset by high levels of inferred respiration. Further, metabolic effects exhibited a strong intra-annual pattern, stressing the importance of short term controls of pH and thereby CO_2 flux in these lakes in calculating the annual CO_2 budget.

Elevated algal abundance increased the likelihood of net CO_2 uptake from the atmosphere. Specifically, we found that CO_2 under average conditions (all other predictors at mean) was in-gassing at moderate to high primary production (Chl $a > 15\ \mu\mathrm{g\,L^{-1}}$) while lower levels of productivity (Chl $a\ 5 - 10\ \mu\mathrm{g\,L^{-1}}$) could result in a net heterotrophic state and CO_2 efflux. Such low productivity values were found most frequently in the mesotrophic Lake Diefenbaker, while strongly positive relationships between pH and Chl *a* occurred often in the most shallow lakes (Wascana, Buffalo Pound; Fig. 4). In general, the observed Chl *a* concentrations needed for net CO_2 release were low ($7 - 15\ \mu\mathrm{g\,L^{-1}}$) relative to those found in other eutrophic lakes where out-gassing may predominate even under the most productive conditions (Chl $a > 40\ \mu\mathrm{g\,L^{-1}}$) [*Huttunen et al.*, 2003; *Reis and Barbosa*, 2014], although outgassing was predicted even in our sites at similar algal production provided other predictors were set to values favouring outgassing (e.g. low oxygen, high TDN).

Both Chl *a* and pH increased through the summer in most lakes suggesting a progressive increase in the importance of metabolic controls. However, these trends were not monotonic, particularly in the more productive lakes. In early summer, the more productive lakes have consistent clear-water periods [*Dröscher et al.*, 2009] caused by strong zooplankton grazing on phytoplankton, thus increasing pCO_2 and subsequently decreasing pH (Fig. 7). Conversely, in late summer, the more productive lakes exhibit reduced oxygen concentrations ($<5\ \mathrm{mg\,L^{-1}}$) indicative of increased respiration of organic material, which favours release of CO_2 to the atmosphere (Fig. 7). More intensive evaluation of fall metabolism is required to establish whether this trend continues through to ice formation in late October or November.

Similar to results of annual mean data [*Finlay et al.*, 2009], rising DOC content tended to co-occur with increasing pH at moderate to high DOC levels (DOC: $5 - 25\ \mathrm{mg\,L^{-1}}$). These patterns are contrary to studies from boreal lakes which tend to show that DOC mineralization increases pCO_2 and reduces pH [*Balmer and Downing*, 2011]. Although speculative, the observed positive relationship between pH and DOC may reflect recalcitrant DOC which is not respired [*Ostapenia et al.*, 2009], autochthonously derived DOC during high primary production [*Søndergaard et al.*, 2000], and/or a positive correlation between DOC and nutrient influx [*Osburn et al.*, 2011]. The latter two are most likely given the positive correlation between TDN and DOC in our study lakes, however further research is required to distinguish among these explanations.

The unimodal relationship of TDN and pH (peak ca. $1100\ \mu\mathrm{g\,N\,L^{-1}}$) suggests that there is a limit to the fertilising effect of nutrients on primary production and in turn pH. Such a limit may reflect a consistent rise in bacterial decomposition of organic matter along the production gradient, leading to a paramount effect of respiration under highly eutrophic conditions [*Hollander and Smith*, 2001]. In our case, TDN itself may be directly utilised by heterotrophs, as most ($>80\%$) dissolved N in these lakes is in organic forms of TDN not available to autotrophs [*Bogard et al.*, 2012]. Consistent with this idea, we note that addition of organic N (as urea) to mesocosm experiments in Wascana Lake increased respiration and decreased pH corresponding with CO_2 efflux [*Bogard et al.*, 2017]. Finally, we infer that the negative correlation between high TDN and pH does not reflect a change in the nutrient limitation status of the lakes, as only Diefenbaker and to a lesser extent Buffalo Pound show evidence of P limitation [*Vogt et al.*, 2015; *Quiñones-Rivera et al.*, 2015] and these sites generally exhibit low TDN values relative to other, more definitively N-limited systems [*Leavitt et al.*, 2006; *Patoine et al.*, 2006].

While we observed a predictable positive relationship between pH and O_2 concentration when oxygen was below saturation, the relationship reversed direction when waters were super-saturated with oxygen (Fig. 4b). We speculate that there are times when there may be simultaneous supersaturation of oxygen and CO_2 thereby decoupling the relationships between oxygen and pH, as observed in other hardwater systems where excess allochthonous carbon coincides with high primary production [*Stets et al.*, 2009; *McDonald et al.*, 2013].

4.2 Climatic regulation of pH

The strength of the relationship between climatic variables and pH was comparable to that of metabolism and pH (Figs 4, 6), a pattern which suggests that climatic mechanisms may also influence intra-annual variation in regional CO_2 flux. For example, dry and warm conditions (very high PDO and very low SOI) as well as high drought index values were associated with elevated pH and increased concentrations of base cations in these and other lakes lakes [*Pham et al.*, 2009; *Lake*, 2011]. Similarly, this pattern is consistent with findings of *Finlay et al.* [2015] who demonstrated that spring and summer pH is elevated during years when short duration of ice cover reduces under-ice respiration and favours increased pH in spring and summer. The most likely drivers of climatic effects on pH are increased base cation concentrations due to evaporative concentration [*Evans and Prepas*, 1996; *Pham et al.*, 2009], elevated residence time [*Knoll et al.*, 2013], reduced allochthonous DIC loads

due to longer transit times [*Stets et al.*, 2017], and higher reliance on groundwater contributions [*Lake*, 2011]. However, further research will be required to better refine these possibilities, including spatial studies relating geology, landscape position, external loading and groundwater supply to seasonality of lake chemistry.

Despite strong and significant results from our modelling exercise, our statistical approach captured only ca. 43% of the deviance in pH, leaving a considerable proportion to be accounted for by other factors. Because model residuals were random and normal, they provided little indication of model deficiencies. In principle, model prediction might be improved through distinction of DOC providence via spectrophotometric or compound-specific analyses to better estimate its effect on respiration [*Koehler et al.*, 2012], while quantification of physico-chemical processes such as convection and mixing may be important in identifying additional controls of pH, such as seen elsewhere [*Maberly*, 1996; *Morales-Pineda et al.*, 2014; *Liu et al.*, 2016]. Thirdly, the use of more finely resolved taxonomic data (e.g., algal groups) in place of coarse metrics of planktonic metabolism (Chl *a*, R) may help refine how the importance of biotic controls varies along long limnological gradients [*Felip and Catalan*, 2000; *George and Heaney*, 1978; *Zhang et al.*, 2010]. Finally, we have not been able to account for alkalinity affecting the buffering capacity and thus the lakes' responsiveness in pH to changes in metabolic CO_2. However, the lack of overall correlation between pH and metabolic covariates suggests that alkalinity changes should be investigated for their potential contribution to pH and CO_2 flux.

4.3 Regional coherence and implications for upscaling CO_2 fluxes

Predicting CO_2 fluxes in these hard-water systems was simplified by the fact that DIC concentrations varied little across the lakes, and that all lakes behaved similarly with regards to metabolic and physical relationships with pH over broad spatial scales. While the lakes varied substantially in salinity and conductivity (Fig. 2), these parameters had relatively low impact on CO_2 fluxes in their respective ranges (Table 2). Conversely, while DIC concentrations are predicted to have substantial effects on atmospheric CO_2 exchange in other lake regions [*Cumming et al.*, 1995; *Doctor et al.*, 2008; *Duarte et al.*, 2008], in our study DIC levels were comparatively low, and also correlated weakly and negatively with changes in pH ($p < 0.001$, R^2=0.014) which implied an absence of negative effects of high DIC on CO_2 influx at high pH.

We found an unexpectedly strong effect of lake morphology on the role of algal abundance (as Chl *a*) as a determinant of pH, with the effect of Chl *a* being much greater in very shallow Buffalo Pound and Wascana lakes (<4 m mean depth) than deeper lakes, particularly at very high pH values (Fig. 4). We speculate that shallow lakes are more likely to exhibit whole-lake responses to photic-zone metabolism, and may have less vertical structure than even deep polymictic lakes (*Zhang et al.* [2010], but see *George and Heaney* [1978]). Fortunately, most prairie lakes are of a similar depth, many being shallow [*Last*, 1989], suggesting that variation in morphology will not unduly affect efforts to estimate regional CO_2 fluxes [*Finlay et al.*, 2015]. Overall, the high level of coherence among basins in terms of high pH and moderate DIC suggest that many lakes will act as CO_2 sinks during much of the summer, provided they are moderately to highly productive ($> 15 - 20\ \mu g\ L^{-1}$ Chl *a*) and are not under extreme (organic) TDN loads.

Metabolic control of CO_2 flux in these hardwater lakes does not appear to be as strong as that observed in boreal or softwater regions where microbial metabolism of DOC [*Sobek et al.*, 2005; *Lapierre and del Giorgio*, 2012] or photosynthesis [*Maberly*, 1996; *Reis and Barbosa*, 2014] regulates pCO_2, albeit with variable allochthonous contributions of respired or otherwise derived DIC [*Weyhenmeyer et al.*, 2015; *Bogard and del Giorgio*, 2016]. These results fit within the larger matrix of lake types along gradients of DIC, DOC, nutrients and alkalinity, and suggest that moderately hardwater lakes are more likely to capture atmospheric CO_2 at a given level of productivity than would dilute lakes [*Reis and Barbosa*, 2014], those with high

DOC loads [*Huttunen et al.*, 2003], or hardwater systems with chronic oversaturation of DIC [*Marcé et al.*, 2015]. Further, because such systems often co-incide with intensively fertilized agricultural regions, there exists the possibility that many of these systems will fall below the global average estimate of lake CO_2 flux [*Raymond et al.*, 2013].

5 Conclusions

Based on advanced time series analysis using GAMs, we found that both metabolic and climatic factors strongly influenced factors related to pH and that variation in DIC was of only secondary importance in affecting CO_2 content. Overall, a modest degree of eutrophication was required for high rates of CO_2 uptake from the atmosphere and some less productive lakes exhibited a release of CO_2 from surface waters. These agricultural areas often exhibit high allochthonous loads of organic carbon and nitrogen which are likely to fertilize the lake. This increases the likelihood of CO_2 influx, but the balance may switch in favour of respiration at extreme nitrogen loads. Overall, climate appeared to have an effect on gas exchange mainly during extremes, such as regional drought, when evaporative concentration of base cations and elevated pH may favour regional influx of CO_2 into lakes. These results aid in our ability to understand and predict how future human-mediated changes to nutrient loading and climate change will impact carbon cycling in lakes.

6 Tables, and figure captions

Table 1. Summary data of study lakes, showing median, minimum-to-maximum (in parentheses) values of monitoring data over the sampling period, as well as mean depth and residence time.

Lake	Residence time (yr)	Mean depth (m)	TDN ($\mu g\,N\,L^{-1}$)	Chl *a* ($\mu g\,L^{-1}$)	DOC ($mg\,L^{-1}$)	TDP ($\mu g\,P\,L^{-1}$)
B	0.7	3	491 (218-1350)	20.1 (1.5-319)	6.1 (0.5-31)	23 (9-132)
C	0.5	8	920 (450-2090)	18 (0-237)	12 (0-41)	126 (16-650)
D	1.3	33	401 (107-1440)	4.7 (0.8-26)	4.8 (0-29)	9 (0.4-295)
K	1.34	14	1152 (418-2390)	21 (1.5-117)	12 (3.7-37)	159 (40-690)
L	12.6	8	999 (482-1510)	13 (2.3-49)	13 (0-82)	31 (14-470)
P	0.71	6	1420 (171-3100)	22 (1.2-287)	12 (0-56)	162 (5-662)
W	0.7	1.5	1309 (600-6400)	27 (2.2-309)	16 (4.8-53)	318 (33-830)

Fig. 1: The seven study sites lie along the Qu'Appelle River (SK, Canada) flowing west to east, with the exception of Wascana (south tributary) and Last Mountain (north tributary).

Fig. 2: a: Box plots for limnological data used to calculate carbon dioxide flux in the lakes, showing medians, upper and lower quartiles, 1.5×inter-quartile ranges, and 'outliers'. b: Major patterns of annual variation in pH in all lakes, based on a generalised additive model of pH by Lake, Year, and Day of Year. Rug: annual means of pH observed over time.x

Fig. 3: GAM splines for pH with lake splines significantly different (see section 2.4.3) from the global spline indicated by color/hue and linetype. Dotted lines: means of y and x; Shaded area: Middle 90% of all observations. Rug: Data points

Fig. 4: a-c: GAM splines for significant metabolic variables. Dotted lines: means of y and x; Shaded area: Middle 90% of all observations. Rug: Data points. a: GAM splines for chlorophyll *a*, with lakes with significantly different splines to the global spline (see section 2.4.3) indicated by color/hue and linetype. b: GAM spline of oxygen, with standard errors indicated by shading. c: GAM spline of DOC, with standard errors indicated by shading.

Table 2. Partial Rank Correlation Coefficients (PRCCs) following Latin Hypercube sensitivity analysis for all variables and all lakes (left panel) and the most important two variables for individual lakes (right panel).

Variable	PRCC (all lakes)	Lake	PRCC (pH)	PRCC (DIC)
pH	-0.96	Last Mountain	-0.98	0.74
DIC	0.51	Crooked	-0.99	0.69
Temperature	-0.28	Diefenbaker	-0.99	0.68
Conductivity	-0.26	Buffalo Pound	-0.99	0.65
Wind	0.20	Pasqua	-0.99	0.64
Salinity	0.10	Katepwa	-0.99	0.57
Air pressure	0.10	Wascana	-0.99	0.56
Air pCO_2	-0.09			

Table 3. Summary of the climatic-metabolic model of pH, showing the estimated effects of the predictors. EDF=estimated degrees of freedom, DF=degrees of freedom. Deviance explained: 43.2%, n=991.

Predictor spline	EDF	DF	chi^2	p value
Chlorophyll *a* (global)	0.979	9	134.366	$\ll 0.0001$
Chlorophyll *a* (Katepwa)	0.000159	4	0	0.47556
Chlorophyll *a* (Last Mountain)	0.0000767	4	0	1
Chlorophyll *a* (Buffalo Pound)	1.80	4	11.168	0.01886
Chlorophyll *a* (Crooked)	0.277	4	0.433	0.22987
Chlorophyll *a* (Diefenbaker)	0.0380	4	0.05	0.28051
Chlorophyll *a* (Wascana)	2.65	4	66.947	$\ll 0.0001$
Chlorophyll *a* (Pasqua)	0.000168	4	0	0.49175
DOC	1.40	9	39.519	0.01285
Oxygen	3.07	9	28.417	0.00772
PDO*SOI	10.8	24	567	$\ll 0.0001$
SPEI	1.41	2	16.342	0.01158
Lake*Year	105	128	532.24	$\ll 0.0001$

Fig. 5: GAM spline for TDN in the alternative model without DOC. Dotted lines: means of y and x; Shaded area: Middle 90% of all observations. Rug: Data points. Standard errors are indicated by shading.

Fig. 6: a-c: GAM interactions of PDO and SOI. a: Heatmap with data points. Dashed lines indicate cross sections for b-c, which show GAM splines for pH for selected combinations of SOI (b) and PDO (c) values. Missing line segments reflect uncertainties in prediction. d: GAM spline of SPEI, with standard errors indicated by shading. Rug: Data points.

Fig. 7: Contributions of each predictor to pH summarised over the months of highest data availability, averaged across lakes for weather and climate indices which were homogenous through the study region. Box plots show medians, upper and lower quartiles, 1.5×interquartile ranges, and 'outliers'. Shaded area: $\pm$ 0.05 regions to aid comparison of magnitudes across predictors.

Fig. A1: Intra-annual variability expressed as median absolute deviation (i.e. the median of the absolute deviations from the median) of key metabolic and/or nutrient status variables over the LTER period over the months of most frequent observations (May-September). The data are superimposed such that the lakes with the lowest variability appear toward the centre of the figure, and lakes with higher variability contain the variability of the more central lakes plus the additional value indicated by the colouring.

Fig. B1: The relationship between calculated carbon dioxide flux and simulated data sets (N=500) of input variables for sensitivity analysis.

Fig. C1: R output for main model diagnostics.

Fig. C2: Measured vs predicted pH over time in the study sites, displayed as monthly means over the months of the most frequent observations.

7 Figures

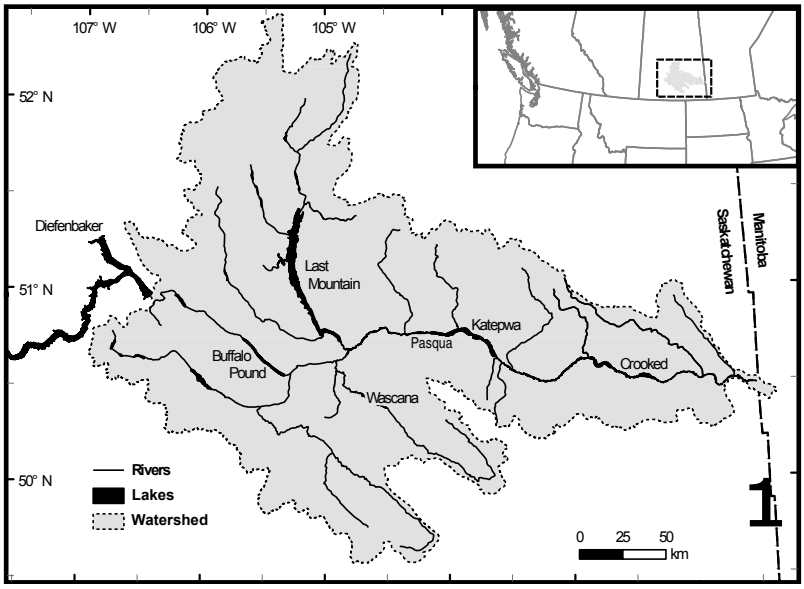

Figure 1. The seven study sites lie along the Qu'Appelle River (SK, Canada) flowing west to east, with the exception of Wascana (south tributary) and Last Mountain (north tributary).

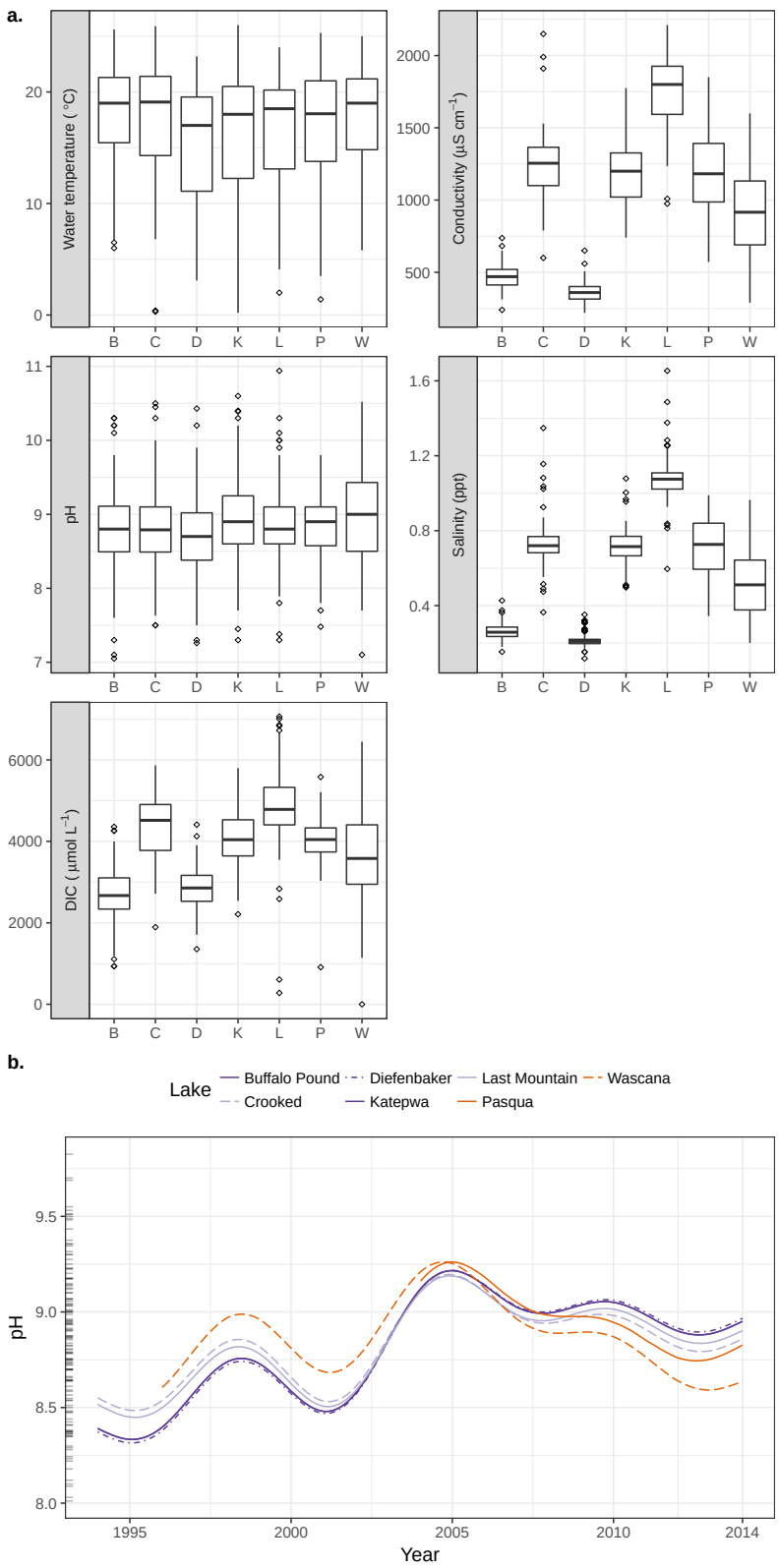

Figure 2. a: Box plots for limnological data used to calculate carbon dioxide flux in the lakes, showing medians, upper and lower quartiles, 1.5×inter-quartile ranges, and 'outliers'. b: Major patterns of annual variation in pH in all lakes, based on a generalised additive model of pH by Lake, Year, and Day of Year. Rug: annual means of pH observed over time.

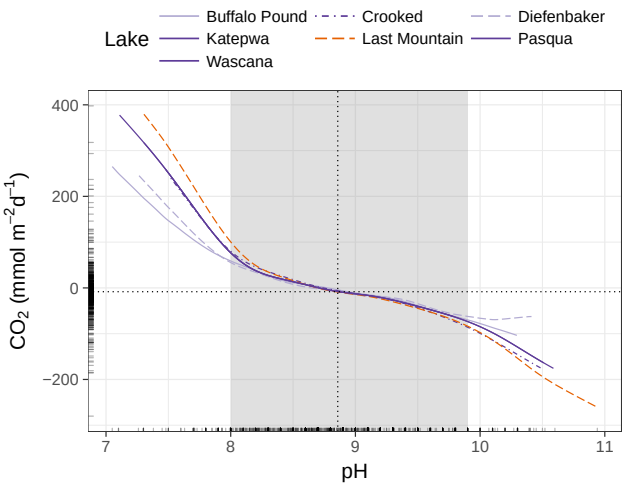

Figure 3. GAM splines for pH with lake splines significantly different (see section 2.4.3) from the global spline indicated by color/hue and linetype. Dotted lines: means of y and x; Shaded area: Middle 90% of all observations. Rug: Data points

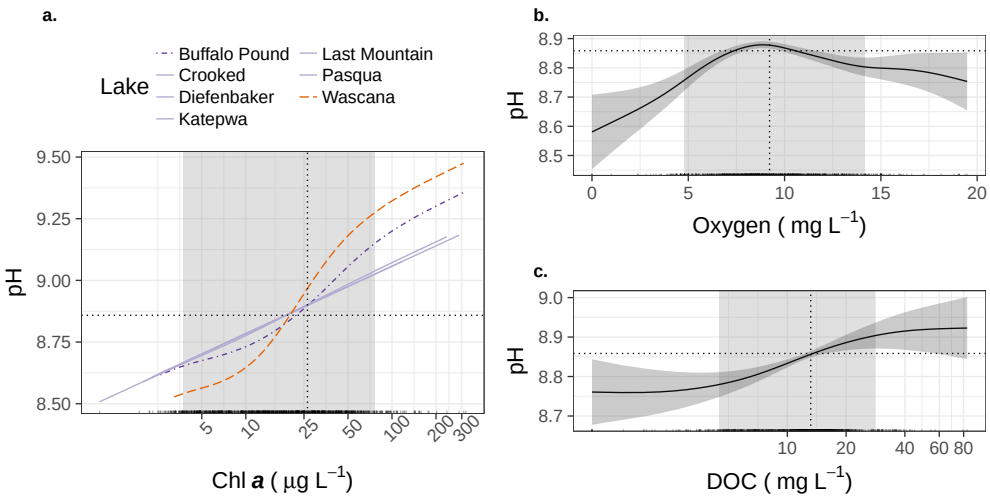

Figure 4. a-c: GAM splines for significant metabolic variables. Dotted lines: means of y and x; Shaded area: Middle 90% of all observations. Rug: Data points. a: GAM splines for chlorophyll *a*, with lakes with significantly different splines to the global spline (see section 2.4.3) indicated by color/hue and linetype. b: GAM spline of oxygen, with standard errors indicated by shading. c: GAM spline of DOC, with standard errors indicated by shading.

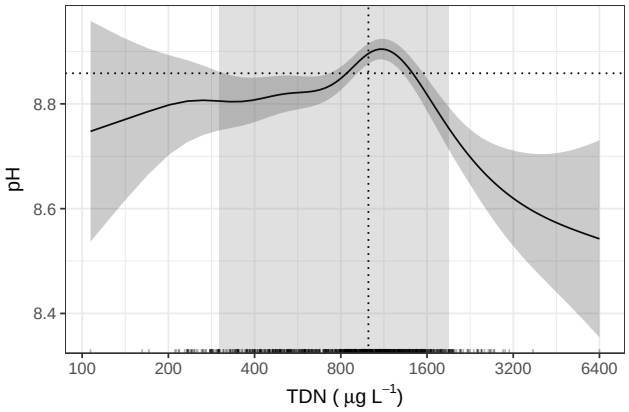

Figure 5. GAM spline for TDN in the alternative model without DOC. Dotted lines: means of y and x; Shaded area: Middle 90% of all observations. Rug: Data points. Standard errors are indicated by shading.

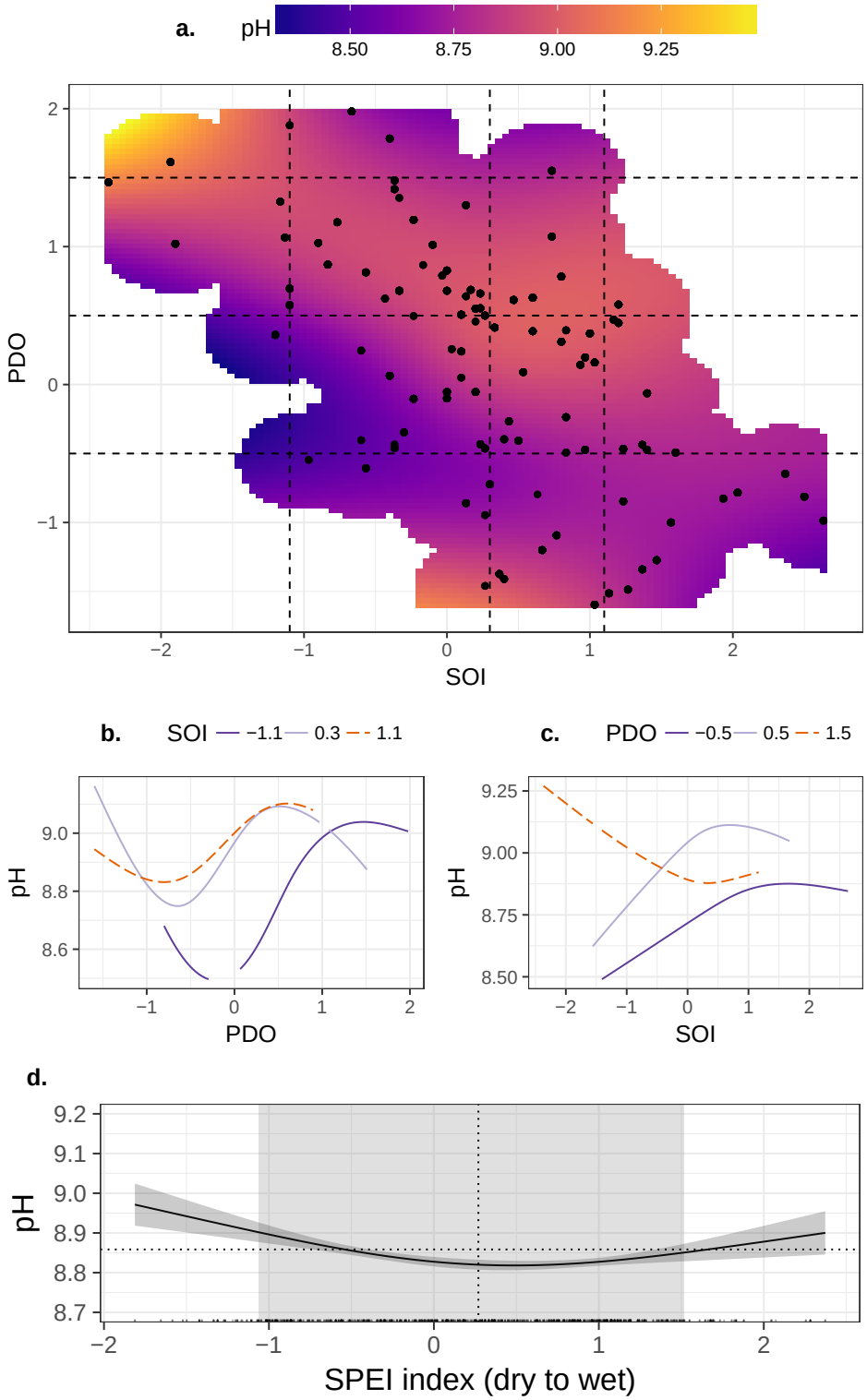

Figure 6. a-c: GAM interactions of PDO and SOI. a: Heatmap with data points. Dashed lines indicate cross sections for b-c, which show GAM splines for pH for selected combinations of SOI (b) and PDO (c) values. Missing line segments reflect uncertainties in prediction. d: GAM spline of SPEI, with standard errors indicated by shading. Rug: Data points.

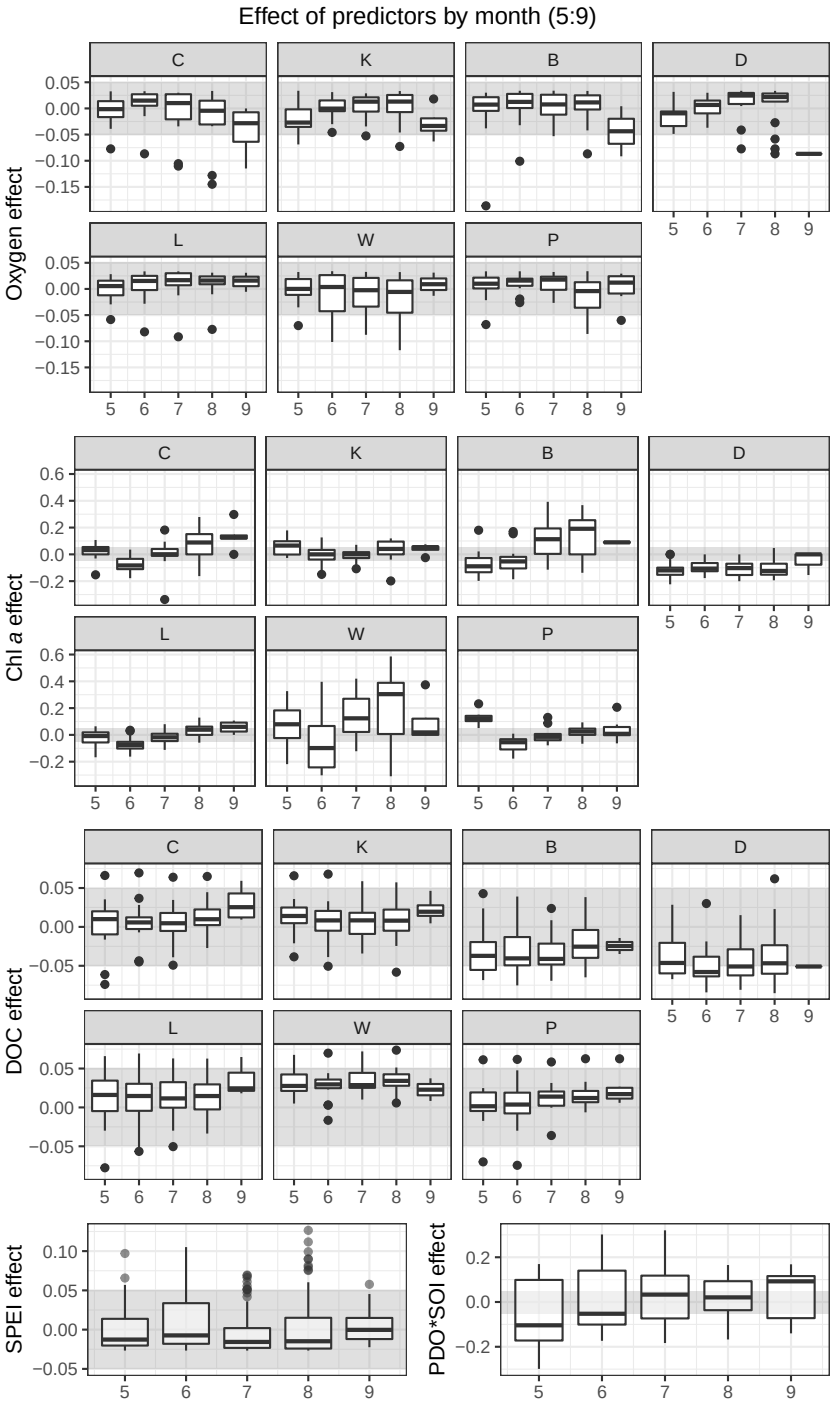

Figure 7. Contributions of each predictor to pH summarised over the months of highest data availability, averaged across lakes for weather and climate indices which were homogenous through the study region. Box plots show medians, upper and lower quartiles, 1.5×inter-quartile ranges, and 'outliers'. Shaded area: ± 0.05 regions to aid comparison of magnitudes across predictors.

A: Summary data for all lakes

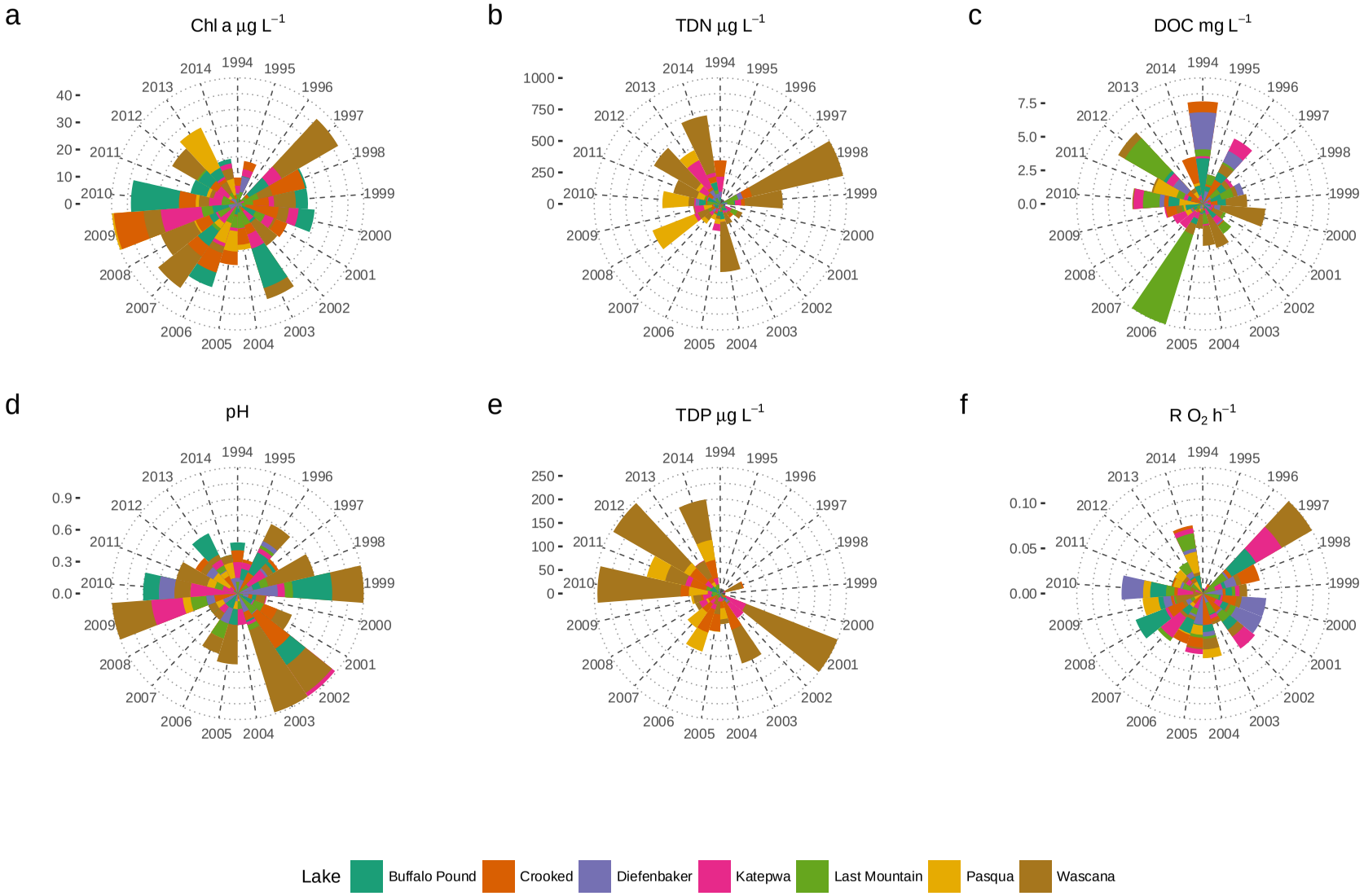

Figure A.1. Intra-annual variability expressed as median absolute deviation (i.e. the median of the absolute deviations from the median) of key metabolic and/or nutrient status variables over the LTER period over the months of most frequent observations (May-September). The data are superimposed such that the lakes with the lowest variability appear toward the centre of the figure, and lakes with higher variability contain the variability of the more central lakes plus the additional value indicated by the colouring.

B: Simulated relationships between predictors and CO_2 using sensitivity analysis

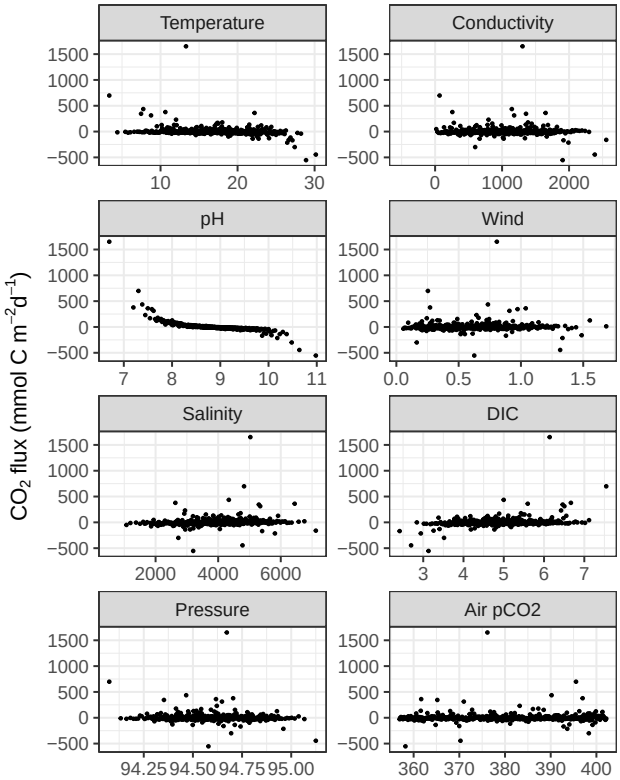

Figure B.1. The relationship between calculated carbon dioxide flux and simulated data sets (N=500) of input variables for sensitivity analysis.

C: Model summaries and diagnostic plots

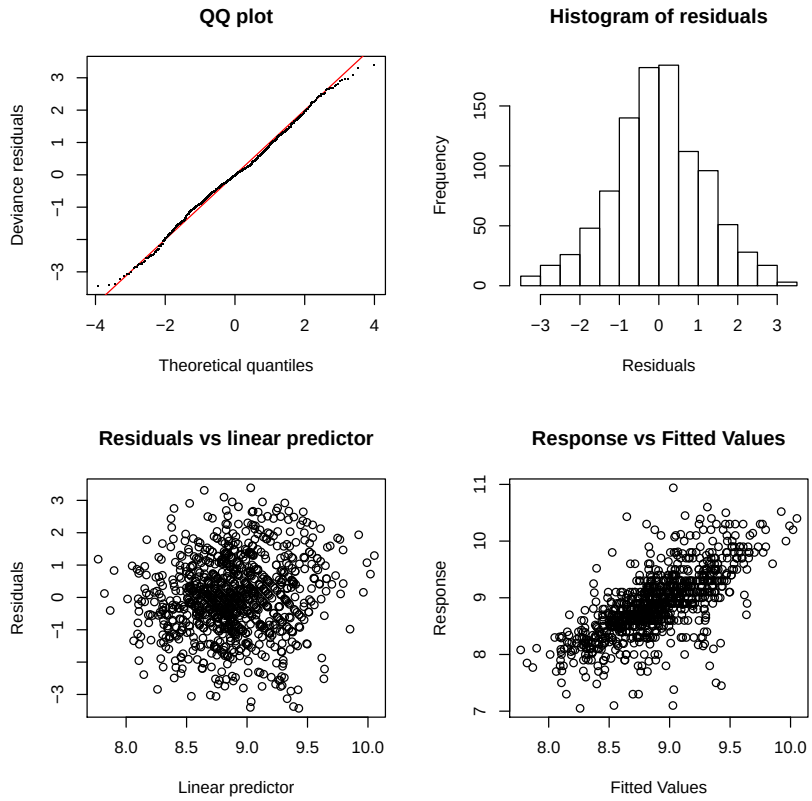

Figure C.1. R output for main model diagnostics.

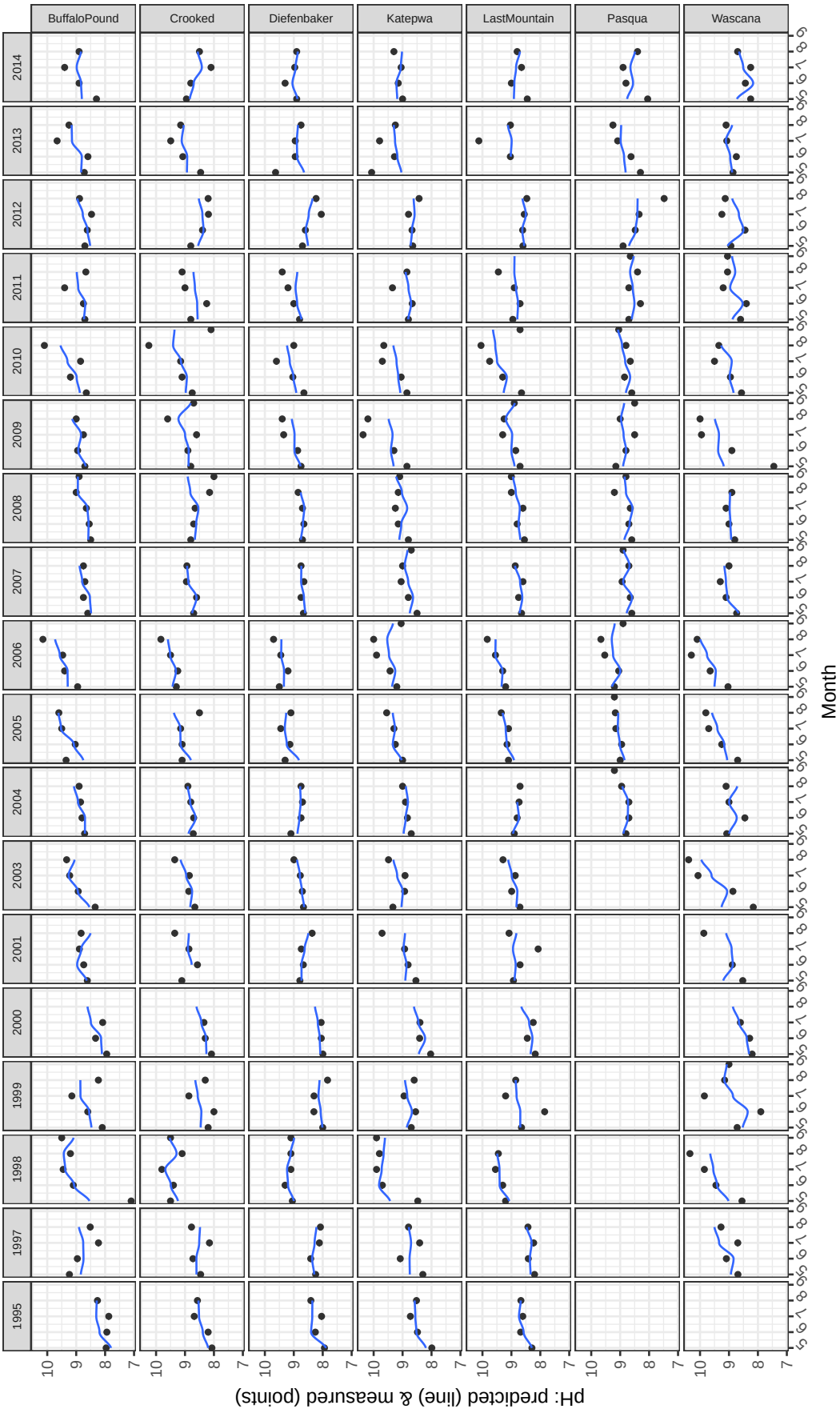

Figure C.2. Measured vs predicted pH over time in the study sites, displayed as monthly means over the months of the most frequent observations.

Acknowledgments

This work was funded by the NSERC Canada Discovery Grants program (PRL, GLS); and Canada Research Chairs, Canada Foundation for Innovation, the Province of Saskatchewan, and the University of Regina (PRL). This paper is a contribution to the Qu'Appelle Long-term Ecological Research Program.

All R code and open data for analysis and figures has been archived at `https://github.com/simpson-lab/jgr-co2-flux`. The version containing private data has been archived at `https://github.com/simpson-lab/jgr-co2-flux-private-data`. DOIs for both archives will be provided upon acceptance. Queries regarding private data belonging to the Qu'Appelle Long-term Ecological Research Program may be directed to Peter.Leavitt@uregina.ca.

We want to thank past and present members of the Limnology Laboratory for assistance with data collection since 1994, and A Chalom for his quick response to changing code on the R PSE package. B Tutolo provided very useful answers to some water chemistry questions.

References

Abril, G., S. Bouillon, F. Darchambeau, C. R. Teodoru, T. R. Marwick, F. Tamooh, F. Ochieng Omengo, N. Geeraert, L. Deirmendjian, P. Polsenaere, et al. (2015), Technical note: Large overestimation of pCO_2 calculated from ph and alkalinity in acidic, organic-rich freshwaters, *Biogeosciences*, *12*(1), 67–78.

Alin, S. R., and T. C. Johnson (2007), Carbon cycling in large lakes of the world: A synthesis of production, burial, and lake-atmosphere exchange estimates, *Global Biogeochemical Cycles*, *21*(3), doi:10.1029/2006GB002881.

Anderson, D. E., R. G. Striegl, D. I. Stannard, C. M. Michmerhuizen, T. A. McConnaughey, and J. W. LaBaugh (1999), Estimating lake-atmosphere CO_2 exchange, *Limnology and Oceanography*, *44*(4), 988–1001.

Audet, J., É. M. Neif, Y. Cao, C. C. Hoffmann, T. L. Lauridsen, S. E. Larsen, M. Søndergaard, E. Jeppesen, and T. A. Davidson (2017), Heat-wave effects on greenhouse gas emissions from shallow lake mesocosms, *Freshwater Biology*, *62*(7), 1130–1142.

Baehr, M. M., and M. D. DeGrandpre (2004), In situ pCO_2 and O_2 measurements in a lake during turnover and stratification: observations and modeling, *Limnology and Oceanography*, *49*(2), 330–340.

Balmer, M. B., and J. A. Downing (2011), Carbon dioxide concentrations in eutrophic lakes: undersaturation implies atmospheric uptake, *Inland Waters*, *1*, 125–132.

Bogard, M. J., and P. A. del Giorgio (2016), The role of metabolism in modulating CO_2 fluxes in boreal lakes, *Global Biogeochemical Cycles*, *30*(10), 1509–1525, doi:10.1002/2016GB005463.

Bogard, M. J., D. B. Donald, K. Finlay, and P. R. Leavitt (2012), Distribution and regulation of urea in lakes of central north america, *Freshwater Biology*, *57*(6), 1277–1292.

Bogard, M. J., K. Finlay, M. J. Waiser, V. P. Tumber, D. B. Donald, E. Wiik, G. L. Simpson, P. A. del Giorgio, and P. R. Leavitt (2017), Effects of experimental nitrogen fertilization on planktonic metabolism and co_2 flux in a hypereutrophic hardwater lake, *PLOS ONE*, *12*(12), 1–19, doi:10.1371/journal.pone.0188652.

Bonsal, B., and A. Shabbar (2008), Impacts of Large-Scale Circulation Variability on Low Streamflows over Canada: A Review, *Canadian Water Resources Journal*, *33*(2), 137–154.

Chalom, A., and P. I. K. L. de Prado (2015), Parameter space exploration of ecological models, *https://arxiv.org/pdf/1210.6278.pdf*, pp. 1–46.

Chalom, A., and P. I. K. L. de Prado (2016), *pse: Parameter Space Exploration with Latin Hypercubes*.

Christensen, J. P. A., K. Sand-Jensen, and P. A. Staehr (2013), Fluctuating water levels control water chemistry and metabolism of a charophyte-dominated pond, *Freshwater Biology*,

58(7), 1353–1365.

Cole, J., J. Nina, and F. Caraco (1998), Atmospheric exchange of carbon dioxide in a low-wind oligotrophic lake measured by the addition of SF_6, *Limnology and Oceanography*, *43*, 647–656.

Cole, Y. T., J. J.and Prairie, N. F. Caraco, W. H. McDowell, L. J. Tranvik, R. G. Striegl, C. M. Duarte, P. Kortelainen, J. A. Downing, J. J. Middelburg, and J. Melack (2007), Plumbing the global carbon cycle: Integrating inland waters into the terrestrial carbon budget, *Ecosystems*, *10*(1), 172–185.

Cumming, B., S. Wilson, R. Hall, and J. P. Smol (1995), Diatoms from British Columbia (Canada) Lakes and their Relationship to Salinity, Nutrients and Other Limnological Variables, in *Bibliotheca Diatomologica*, vol. Band 31, edited by H. Lange-Bertalot, pp. 1–203, J Cramer, Berlin, Germany.

Denfeld, B. A., M. B. Wallin, E. Sahlée, S. Sobek, J. Kokic, H. E. Chmiel, and G. A. Weyhenmeyer (2015), Temporal and spatial carbon dioxide concentration patterns in a small boreal lake in relation to ice cover dynamics, *Boreal Environmental Research*, *20*, 667–78.

Doctor, D. H., C. Kendall, S. D. Sebestyen, J. B. Shanley, N. Ohte, and E. W. Boyer (2008), Carbon isotope fractionation of dissolved inorganic carbon (DIC) due to outgassing of carbon dioxide from a headwater stream, *Hydrological Processes*, *22*(14), 2410–2423.

Downing, J. A., J. J. Cole, J. J. Middelburg, R. G. Striegl, C. M. Duarte, P. Kortelainen, Y. T. Prairie, and K. A. Laube (2008), Sediment organic carbon burial in agriculturally eutrophic impoundments over the last century, *Global Biogeochemical Cycles*, *22*(1), n/a–n/a, doi: 10.1029/2006GB002854.

Dröscher, I., A. Patoine, K. Finlay, and P. R. Leavitt (2009), Climate control of the spring clear-water phase through the transfer of energy and mass to lakes, *Limnology and Oceanography*, *54*(6/2), 2469–2480.

Duarte, C. M., Y. T. Prairie, C. Montes, J. J. Cole, R. Striegl, J. Melack, and J. A. Downing (2008), CO_2 emissions from saline lakes: A global estimate of a surprisingly large flux, *Journal of Geophysical Research: Biogeosciences*, *113*(G4).

Ducharme-Riel, V., D. Vachon, P. A. del Giorgio, and Y. T. Prairie (2015), The relative contribution of winter under-ice and summer hypolimnetic CO_2 accumulation to the annual CO_2 emissions from northern lakes, *Ecosystems*, *18*(4), 547–559.

Duston, N. M., R. M. Owen, and B. H. Wilkinson (1986), Water chemistry and sedimentological observations in Littlefield Lake, Michigan: Implications for lacustrine marl deposition, *Environmental Geology and Water Sciences*, *8*(4), 229–236.

Environment Canada (1979), *Analytical method manual. Inland Waters Directorate NAQUADAT 15413*, Environment Canada.

Evans, J. C., and E. E. Prepas (1996), Potential effects of climate change on ion chemistry and phytoplankton communities in prairie saline lakes, *Limnology and Oceanography*, *41*(5), 1063–1076.

Felip, M., and J. Catalan (2000), The relationship between phytoplankton biovolume and chlorophyll in a deep oligotrophic lake: decoupling in their spatial and temporal maxima, *Journal of Plankton Research*, *22*(1), 91–106.

Finlay, K., P. Leavitt, B. Wissel, and Y. Prairie (2009), Regulation of spatial and temporal variability of carbon flux in six hard-water lakes of the northern great plains, *Limnology and Oceanography*, *54*(6), 2553.

Finlay, K., R. J. Vogt, M. J. Bogard, B. Wissel, B. M. Tutolo, G. L. Simpson, and P. R. Leavitt (2015), Decrease in CO_2 efflux from northern hardwater lakes with increasing atmospheric warming, *Nature*, *519*, 215–218.

George, D. G., and S. I. Heaney (1978), Factors Influencing the Spatial Distribution of Phytoplankton in a Small Productive Lake, *Journal of Ecology*, *66*(1), 133–155.

Hall, R. I., P. R. Leavitt, R. Quinlan, A. S. Dixit, and J. P. Smol (1999), Effects of agriculture, urbanization, and climate on water quality in the northern great plains, *Limnology and Oceanography*, *44*(3), 739–756.

Hanson, P. C., D. L. Bade, S. R. Carpenter, and T. K. Kratz (2003), Lake metabolism: Relationships with dissolved organic carbon and phosphorus, *Limnology and Oceanography*, *48*(3), 1112–1119.

Hastie, T., and R. Tibshirani (1990), *Generalized Additive Models*, Chapman and Hall/CRC.

Hollander, D. J., and M. A. Smith (2001), Microbially mediated carbon cycling as a control on the δ^{13}C of sedimentary carbon in eutrophic Lake Mendota (USA): New models for interpreting isotopic excursions in the sedimentary record, *Geochimica et Cosmochimica Acta*, *65*(23), 4321–4337.

Hoover, T. E., and D. C. Berkshire (1969), Effects of hydration on carbon dioxide exchange across an air-water interface, *Journal of Geophysical Research*, *74*(2), 456–464.

Howarth, R. W., and A. F. Michaels (2000), *The measurement of primary productivity in aquatic ecosystems*, pp. 72–85, Springer.

Huttunen, J. T., J. Alm, A. Liikanen, S. Juutinen, T. Larmola, T. Hammar, J. Silvola, and P. J. Martikainen (2003), Fluxes of methane, carbon dioxide and nitrous oxide in boreal lakes and potential anthropogenic effects on the aquatic greenhouse gas emissions, *Chemosphere*, *52*(3), 609–621.

Jeffrey, S., and G. F. Humphrey (1975), New spectrophotometric equations for determining chlorophylls a, b, c_1 and c_2 in higher plants, algae and natural phytoplankton, *Biochemie und Physiologie der Pflanzen*, *167*(S), 191–194.

Jonsson, A., J. Karlsson, and M. Jansson (2003), Sources of carbon dioxide supersaturation in clearwater and humic lakes in Northern Sweden, *Ecosystems*, *6*(3), 224–235.

Kling, G., G. Kipphut, and M. Miller (1992), The flux of CO_2 and CH_4 from lakes and rivers in arctic Alaska, *Hydrobiologia*, *240*(23), 23–36.

Klug, J. L., D. C. Richardson, H. A. Ewing, B. R. Hargreaves, N. R. Samal, D. Vachon, D. C. Pierson, A. M. Lindsey, D. M. O'Donnell, S. W. Effler, and K. C. Weathers (2012), Ecosystem Effects of a Tropical Cyclone on a Network of Lakes in Northeastern North America, *Environmental Science & Technology*, *46*(21), 11,693–11,701.

Knoll, L. B., M. J. Vanni, W. H. Renwick, E. K. Dittman, and J. A. Gephart (2013), Temperate reservoirs are large carbon sinks and small CO_2 sources: Results from high-resolution carbon budgets, *Global Biogeochemical Cycles*, *27*(1), 52–64.

Koehler, B., E. von Wachenfeldt, D. Kothawala, and L. J. Tranvik (2012), Reactivity continuum of dissolved organic carbon decomposition in lake water, *Journal of Geophysical Research: Biogeosciences*, *117*(G1), n/a–n/a, doi:10.1029/2011JG001793.

Lake, P. (2011), *Drought and Aquatic Ecosystems: Effects and Responses*, 1 ed., Wiley-Blackwell.

Lapierre, J.-F., and P. A. del Giorgio (2012), Geographical and environmental drivers of regional differences in the lake pCO_2 versus DOC relationship across northern landscapes, *Journal of Geophysical Research: Biogeosciences*, *117*(G3).

Last, W. M. (1989), Continental brines and evaporites of the northern Great Plains of Canada, *Sedimentary Geology*, *64*(4), 207–221, doi:https://doi.org/10.1016/0037-0738(89)90049-3.

Leavitt, P. R., C. S. Brock, C. Ebel, and A. Patoine (2006), Landscape-scale effects of urban nitrogen on a chain of freshwater lakes in central North America, *Limnology and Oceanography*, *51*(5), 2262–2277, doi:10.4319/lo.2006.51.5.2262.

Liu, H., Q. Zhang, G. G. Katul, J. J. Cole, F. S. C. III, and S. MacIntyre (2016), Large CO_2 effluxes at night and during synoptic weather events significantly contribute to CO_2 emissions from a reservoir, *Environmental Research Letters*, *11*(6), doi:10.1088/1748-9326/11/6/064001.

Maberly, S. C. (1996), Diel, episodic and seasonal changes in pH and concentrations of inorganic carbon in a productive lake, *Freshwater Biology*, *35*(3), 579–598.

Maberly, S. C., P. A. Barker, A. W. Stott, and M. M. De Ville (2013), Catchment productivity controls CO_2 emissions from lakes, *Nature Climate Change (Letter)*, *3*, 391–394.

Marcé, R., B. Obrador, J.-A. Morguí, J. L. Riera, P. López, and J. Armengol (2015), Carbonate weathering as a driver of CO_2 supersaturation in lakes, *Nature Geoscience*, *8*(2), 107–111.

McDonald, C. P., E. G. Stets, R. G. Striegl, and D. Butman (2013), Inorganic carbon loading as a primary driver of dissolved carbon dioxide concentrations in the lakes and reservoirs of the contiguous united states, *Global Biogeochemical Cycles*, *27*(2), 285–295, doi:10.1002/gbc.20032.

Morales-Pineda, M., A. Cózar, I. Laiz, B. Úbeda, and J. Á. Gálvez (2014), Daily, biweekly, and seasonal temporal scales of pCO_2 variability in two stratified mediterranean reservoirs, *Journal of Geophysical Research: Biogeosciences*, *119*(4), 509–520.

Ojala, A., J. L. Bellido, T. Tulonen, P. Kankaala, and J. Huotari (2011), Carbon gas fluxes from a brown-water and a clear-water lake in the boreal zone during a summer with extreme rain events, *Limnology and Oceanography*, *56*(1), 61–76, doi:10.4319/lo.2011.56.1.0061.

O'Reilly, C. M., S. Sharma, D. K. Gray, S. E. Hampton, J. S. Read, R. J. Rowley, P. Schneider, J. D. Lenters, P. B. McIntyre, B. M. Kraemer, G. A. Weyhenmeyer, D. Straile, B. Dong, R. Adrian, M. G. Allan, O. Anneville, L. Arvola, J. Austin, J. L. Bailey, J. S. Baron, J. D. Brookes, E. de Eyto, M. T. Dokulil, D. P. Hamilton, K. Havens, A. L. Hetherington, S. N. Higgins, S. Hook, L. R. Izmest'eva, K. D. Joehnk, K. Kangur, P. Kasprzak, M. Kumagai, E. Kuusisto, G. Leshkevich, D. M. Livingstone, S. MacIntyre, L. May, J. M. Melack, D. C. Mueller-Navarra, M. Naumenko, P. Noges, T. Noges, R. P. North, P.-D. Plisnier, A. Rigosi, A. Rimmer, M. Rogora, L. G. Rudstam, J. A. Rusak, N. Salmaso, N. R. Samal, D. E. Schindler, S. G. Schladow, M. Schmid, S. R. Schmidt, E. Silow, M. E. Soylu, K. Teubner, P. Verburg, A. Voutilainen, A. Watkinson, C. E. Williamson, and G. Zhang (2015), Rapid and highly variable warming of lake surface waters around the globe, *Geophysical Research Letters*, *42*(24), 10,773–10,781.

Osburn, C. L., C. R. Wigdahl, S. C. Fritz, and J. E. Saros (2011), Dissolved organic matter composition and photoreactivity in prairie lakes of the U.S. Great Plains, *Limnology and Oceanography*, *56*(6), 2371–2390.

Ostapenia, A. P., A. Parparov, and T. Berman (2009), Lability of organic carbon in lakes of different trophic status, *Freshwater Biology*, *54*(6), 1312–1323.

Patoine, A., M. D. Graham, and P. R. Leavitt (2006), Spatial variation of nitrogen fixation in lakes of the northern Great Plains, *Limnology and Oceanography*, *51*(4), 1665–1677, doi:10.4319/lo.2006.51.4.1665.

Perga, M.-E., S. C. Maberly, J.-P. Jenny, B. Alric, C. Pignol, and E. Naffrechoux (2016), A century of human-driven changes in the carbon dioxide concentration of lakes, *Global Biogeochemical Cycles*, *30*(2), 93–104.

Pham, S. V., P. R. Leavitt, S. McGowan, B. Wissel, and L. I. Wassenaar (2009), Spatial and temporal variability of prairie lake hydrology as revealed using isotopes of hydrogen and oxygen, *Limnology and Oceanography*, *54*(1), 101–118.

Pomeroy, J. W., D. de Boer, and L. Martz (2007), *Hydrology and water resources*, pp. 63–80, Canadian Plains Research Centre.

Prairie, Y. T. (2008), Carbocentric limnology: looking back, looking forward, *Canadian Journal of Fisheries and Aquatic Sciences*, *65*(3), 543–548.

Quiñones-Rivera, Z. J., K. Finlay, R. J. Vogt, P. R. Leavitt, and B. Wissel (2015), Hydrologic, metabolic and chemical regulation of water-column metabolism and atmospheric CO_2 exchange in a large continental reservoir during spring and summer, *Journal of Great Lakes Research*, *41*(Supp. 2), 144 – 154.

R Development Core Team (2016), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Raymond, P. A., J. Hartmann, R. Lauerwald, S. Sobek, C. McDonald, M. Hoover, D. Butman, R. Striegl, E. Mayorga, C. Humborg, P. Kortelainen, H. Dürr, M. Meybeck, P. Ciais, and P. Guth (2013), Global carbon dioxide emissions from inland waters, *Nature*, *503*, 355–359.

Reis, P., and F. Barbosa (2014), Diurnal sampling reveals significant variation in CO_2 emission from a tropical productive lake, *Brazilian Journal of Biology*, *74*(3), S113–S119.

Seekell, D. A., and C. Gudasz (2016), Long-term pCO_2 trends in Adirondack Lakes, *Geophysical Research Letters*, *43*(10), 5109–5115.

Shabbar, A., and B. Yu (2012), Intraseasonal Canadian Winter Temperature Responses to Interannual and Interdecadal Pacific SST Modulations, *Atmosphere-Ocean*, *50*(1), 109–121.

Shabbar, A., B. R. Bonsal, and K. Szeto (2011), Atmospheric and Oceanic Variability Associated with Growing Season Droughts and Pluvials on the Canadian Prairies, *Atmosphere-Ocean*, *49*(4), 339–355.

Shao, C., J. Chen, C. A. Stepien, H. Chu, Z. Ouyang, T. B. Bridgeman, K. P. Czajkowski, R. H. Becker, and R. John (2015), Diurnal to annual changes in latent, sensible heat, and CO_2 fluxes over a Laurentian Great Lake: A case study in Western Lake Erie, *Journal of Geophysical Research: Biogeosciences*, *120*(8), 1587–1604.

Sobek, S., L. J. Tranvik, and J. J. Cole (2005), Temperature independence of carbon dioxide supersaturation in global lakes, *Global Biogeochemical Cycles*, *19*(2), 1–15.

Søndergaard, M., N. H. Borch, and B. Riemann (2000), Dynamics of biodegradable DOC produced by freshwater plankton communities, *Aquatic Microbial Ecology*, *23*(1), 73–83.

Soumis, N., É. Duchemin, R. Canuel, and M. Lucotte (2004), Greenhouse gas emissions from reservoirs of the western United States, *Global Biogeochemical Cycles*, *18*, doi:10.1029/2003GB002197.

Stets, E. G., R. G. Striegl, G. R. Aiken, D. O. Rosenberry, and T. C. Winter (2009), Hydrologic support of carbon dioxide flux revealed by whole-lake carbon budgets, *Journal of Geophysical Research: Biogeosciences*, *114*(G1), n/a–n/a, doi:10.1029/2008JG000783.

Stets, E. G., D. Butman, C. P. McDonald, S. M. Stackpoole, M. D. DeGrandpre, and R. G. Striegl (2017), Carbonate buffering and metabolic controls on carbon dioxide in rivers, *Global Biogeochemical Cycles*, *31*(4), 663–677, doi:10.1002/2016GB005578.

Striegl, R. G., and C. M. Michmerhuizen (1998), Hydrologic influence on methane and carbon dioxide dynamics at two north-central Minnesota lakes, *Limnology and Oceanography*, *43*(7), 1519–1529.

Stumm, W., and J. J. Morgan (1996), *Aquatic chemistry: Chemical equilibria and rates in natural waters*, Wiley.

Tranvik, L. J., J. A. Downing, J. B. Cotner, S. A. Loiselle, R. G. Striegl, T. J. Ballatore, P. Dillon, K. Finlay, K. Fortino, L. B. Knoll, P. L. Kortelainen, T. Kutser, S. Larsen, I. Laurion, D. M. Leech, S. L. McCallister, D. M. McKnight, J. M. Melack, E. Overholt, J. A. Porter, Y. Prairie, W. H. Renwick, F. Roland, B. S. Sherman, D. W. Schindler, S. Sobek, A. Tremblay, M. J. Vanni, A. M. Verschoor, E. von Wachenfeldt, and G. A. Weyhenmeyer (2009), Lakes and reservoirs as regulators of carbon cycling and climate, *Limnology and Oceanography*, *54*(6part2), 2298–2314.

Vachon, D., and Y. T. Prairie (2013), The ecosystem size and shape dependence of gas transfer velocity versus wind speed relationships in lakes, *Canadian Journal of Fisheries and Aquatic Sciences*, *70*, 1757–1764.

van der Kamp, G., D. Keir, and M. S. Evans (2008), Long-Term Water Level Changes in Closed-Basin Lakes of the Canadian Prairies, *Canadian Water Resources Journal*, *33*(1), 23–38.

Vicente-Serrano, S. M., S. Beguería, and J. I. López-Moreno (2016), A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index, *Journal of Climate*, *23*, 1696–1718.

Vogt, R. J., J. A. Rusak, A. Patoine, and P. R. Leavitt (2011), Differential effects of energy and mass influx on the landscape synchrony of lake ecosystems, *Ecology*, *92*, 1104–1114.

Vogt, R. J., S. Sharma, and P. R. Leavitt (2015), Decadal regulation of phytoplankton abundance and water clarity in a large continental reservoir by climatic, hydrologic and trophic processes, *Journal of Great Lakes Research*, *41*, 81–90.

Wanninkhof, R. (1992), Relationship between wind speed and gas exchange over the ocean, *Journal of Geophysical Research: Oceans*, *97*(C5), 7373–7382.

Wanninkhof, R., and M. Knox (1996), Chemical enhancement of CO_2 exchange in natural waters, *Limnology and Oceanography*, *41*(4), 689–697.

Weyhenmeyer, G. A., S. Kosten, M. B. Wallin, L. J. Tranvik, E. Jeppesen, and F. Roland (2015), Significant fraction of CO_2 emissions from boreal lakes derived from hydrologic

inorganic carbon inputs, *Nature Geoscience*, *8*, 933–936.

Wood, S. (2011), Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models, *Journal of the Royal Statistical Society (B)*, *73*(1), 3–36.

Wood, S. (2017), *Generalized Additive Models: An Introduction with R*, 2. ed., Chapman and Hall/CRC.

Zhang, Y., L. Feng, J. Li, L. Luo, Y. Yin, M. Liu, and Y. Li (2010), Seasonal–spatial variation and remote sensing of phytoplankton absorption in Lake Taihu, a large eutrophic and shallow lake in China, *Journal of Plankton Research*, *32*(7), 1023–1037.