

# Large Uncertainty in Estimating $p\text{CO}_2$ from Carbonate Equilibria in Lakes

Malgorzata Golub<sup>1,2</sup>, Ankur R. Desai<sup>1,3,\*</sup>, Galen A. McKinley<sup>1,3</sup>, Christina K. Remucal<sup>4</sup>, Emily H. Stanley<sup>5</sup>

<sup>1</sup> Center for Climatic Research, University of Wisconsin – Madison, Wisconsin, USA

<sup>2</sup> Institute of Environmental Protection – National Research Institute, Warsaw, Poland

<sup>3</sup> Dept. of Atmospheric and Oceanic Sciences, University of Wisconsin-Madison, Madison, Wisconsin, USA

<sup>4</sup> Dept. of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, Wisconsin, USA

<sup>5</sup> Center for Limnology, University of Wisconsin-Madison, Madison, Wisconsin, USA

\*Corresponding author: Ankur Desai, AOSS 1549, 1225 W Dayton St., Phone: +1-608-520-0305, Fax: +1-608-262-0166, Email: [desai@aos.wisc.edu](mailto:desai@aos.wisc.edu)

## Key Points:

- $p\text{CO}_2$  random error varies by input parameter pairs and lake alkalinity groups
- Systematic uncertainty is larger than random error in  $p\text{CO}_2$  estimation
- Uncertain historical data produce  $p\text{CO}_2$  insensitive to detect long-term change

## Running title:

Uncertainty in freshwater  $p\text{CO}_2$  estimates

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/2017JG003794

## Abstract

Most estimates of carbon dioxide ( $\text{CO}_2$ ) evasion from freshwaters rely on calculating partial pressure of aquatic  $\text{CO}_2$  ( $p\text{CO}_2$ ) from two out of three  $\text{CO}_2$ -related parameters using carbonate equilibria. However, the  $p\text{CO}_2$  uncertainty has not been systematically evaluated across multiple lake types and equilibria. We quantified random errors in pH, dissolved inorganic carbon, alkalinity, and temperature from the North Temperate Lakes Long-Term Ecological Research site in four lake groups across a broad gradient of chemical composition. These errors were propagated onto  $p\text{CO}_2$  calculated from three carbonate equilibria, and for overlapping observations, compared against uncertainties in directly measured  $p\text{CO}_2$ . The empirical random errors in  $\text{CO}_2$ -related parameters were mostly below 2% of their median values. Resulting random  $p\text{CO}_2$  errors ranged from  $\pm 3.7\%$  to  $\pm 31.5\%$  of the median depending on alkalinity group and choice of input parameter pairs. Temperature uncertainty had a negligible effect on  $p\text{CO}_2$ . When compared with direct  $p\text{CO}_2$  measurements, all parameter combinations produced biased  $p\text{CO}_2$  estimates with less than one third of total uncertainty explained by random  $p\text{CO}_2$  errors, indicating that systematic uncertainty dominates over random error. Multi-decadal trend of  $p\text{CO}_2$  was difficult to reconstruct from uncertain historical observations of  $\text{CO}_2$ -related parameters. Given poor precision and accuracy of  $p\text{CO}_2$  estimates derived from virtually any combination of two  $\text{CO}_2$ -related parameters, we recommend direct  $p\text{CO}_2$  measurements where possible. To achieve consistently robust estimates of  $\text{CO}_2$  emissions from freshwater components of terrestrial carbon balances, future efforts should focus on improving accuracy and precision of  $\text{CO}_2$ -related parameters (including direct  $p\text{CO}_2$ ) measurements and associated  $p\text{CO}_2$  calculations.

## **Plain Language Summary**

Lakes, ponds, streams, and rivers process a large amount of organic matter, some of which is emitted to the atmosphere as global-warming greenhouse gases like carbon dioxide or methane. The rates and amounts of these emissions influence how quickly atmospheric greenhouse gas levels, and hence climate change, progresses. Though it is currently not feasible to make direct measurements of carbon dioxide in most lakes, it is possible to estimate carbon dioxide quantities from other sets of commonly taken aquatic chemistry measurements. However, we lack information if these estimates are close to direct measurements, and hence, trustworthy. Here, we analyze a unique three-decade long data of the key measurements across a gradient of 11 lakes to show that even pretty similar repeated measurements can sometimes cause large scatter of estimated carbon dioxide. Further, we compare a subset of observations to direct measurements of carbon dioxide, and find estimated quantities to be also largely different from measured carbon dioxide quantities. Our results suggest that we need to learn more about factors causing data differences if we want to trust our estimates of freshwater carbon dioxide emissions. We provide a set of recommendations to advance carbon dioxide estimates and measurements.

## **INDEX TERMS:**

Chemistry of fresh water (1806); Uncertainty quantification (1873); Biogeochemical kinetics and reaction (0412); Carbon cycling (0428); Biosphere/atmosphere interactions (0426)

## **KEYWORDS:**

carbonate equilibria, carbon dioxide system, random error, systematic error, error analysis, freshwaters

## 1 Introduction

Outgassing of carbon dioxide (CO<sub>2</sub>) from inland waters has been estimated to offset approximately 20% of net uptake of carbon into the terrestrial biosphere [Ciais *et al.*, 2013]. However, this calculation is based on estimates of source strength at the air-water interface that are highly uncertain [Raymond *et al.*, 2013]. One of the largest unknowns is the accuracy and precision of freshwater partial pressure of CO<sub>2</sub> (*p*CO<sub>2</sub>) estimates. Improving understanding of *p*CO<sub>2</sub> observational uncertainties is therefore a key step towards achieving robust estimates of CO<sub>2</sub> emissions from inland freshwaters, leading to confidence in detection of long-term change relevant to global terrestrial carbon cycling.

The net air-water CO<sub>2</sub> exchange is calculated as a product of the CO<sub>2</sub> gas transfer velocity (*k*), the CO<sub>2</sub> solubility constant (*K*<sub>0</sub>) and the gradient between *p*CO<sub>2</sub> in the atmosphere and water ( $\Delta p$ CO<sub>2</sub>). The aquatic component of  $\Delta p$ CO<sub>2</sub> in current estimates of carbon evasion from inland waters relies on calculating *p*CO<sub>2</sub> using carbonate equilibria due to scarcity of direct *p*CO<sub>2</sub> measurements at regional and global scales [Butman and Raymond, 2011; McDonald *et al.*, 2013; Raymond *et al.*, 2013]. Given the high accuracy and precision of atmospheric *p*CO<sub>2</sub> measurements [Andrews *et al.*, 2014], uncertainties attributed to measurement errors in aquatic carbon system parameters are likely the largest source of uncertainty in  $\Delta p$ CO<sub>2</sub>.

Carbonate equilibria use temperature and the combination of two of three CO<sub>2</sub>-related parameters (i.e., pH, alkalinity (ALK), dissolved inorganic carbon (DIC)) to calculate *p*CO<sub>2</sub> [Parkhurst and Appelo, 1999; van Heuven *et al.*, 2011]. Thus, *p*CO<sub>2</sub> errors in *p*CO<sub>2</sub> directly arise from measurement errors in these parameters. These measurement errors include systematic errors and random errors. Systematic errors (e.g., instrument limitations and methodological errors) affect the accuracy of the measurements [Skoog *et al.*, 2014] and lead to directional (i.e., positive or negative) biases in the measurements of pH, ALK, and DIC

concentration [Herczeg and Hesslein, 1984; French et al., 2002; Lozovik, 2005]. Systematic errors are likely to cause biased  $p\text{CO}_2$  estimates in surface waters [Herczeg and Hesslein, 1984; Butman and Raymond, 2011; Abril et al., 2015] and their contributions to regional and global  $\text{CO}_2$  emissions from freshwaters have not yet been quantified [Raymond et al., 2013].

While targeted efforts can help minimize systematic errors, random errors will always be present in observations of carbonate system parameters, and thus must be considered during data analysis, model-data comparison, and interpretation of trends [Richardson et al., 2012]. Difficult to control factors (e.g., fluctuations of temperature or barometric pressure) or insufficient understanding of errors in analytical procedures cause data to scatter around the mean values and affect parameter precision [Skoog et al., 2014]. Measurement precision can be characterized by estimating the standard deviation from multiple measurements collected under different conditions (reproducibility) or from a pair of independent measurements made under identical conditions (repeatability), assuming normally distributed errors [ISO, 2004]. While a few studies address random uncertainties in measurements of carbonate parameters [Wilkinson et al., 1992; French et al., 2002; Phillips et al., 2015] and direct and calculated  $p\text{CO}_2$  [Herczeg and Hesslein, 1984; Baehr and DeGrandpre, 2004], none of existing studies evaluates how parameter uncertainties propagate on random uncertainty in  $p\text{CO}_2$  calculated from multiple carbonate equilibria, identifies key parameters contributing to  $p\text{CO}_2$  errors, or tests if existing uncertainties would allow for detecting long-term change.

Oceanographers have made significant efforts to standardize and reduce these errors, resulting in thermodynamically consistent measurements of  $\text{CO}_2$ -related parameters, highly precise and accurate estimates of the seawater carbonate system [Lueker et al., 2000; Millero, 2007], and thus of the ocean sink for anthropogenic carbon [Sabine et al., 2004; Ciais et al., 2013b; Khatiwala et al., 2013]. In contrast, acceptable accuracy and precision levels have not been standardized for freshwater systems. Additional challenges stem from diverse chemical

composition of inland waters [Dickson and Riley, 1978], using data originally designed to monitor other characteristics of ecosystems, not  $p\text{CO}_2$ , and general lack of published error estimates to assess the uncertainty of  $\text{CO}_2$  source strength from aquatic systems.

In view of growing interest in integrating aquatic and terrestrial components of carbon balances and persistent use of carbonate equilibria to estimate  $p\text{CO}_2$  and C flux from freshwater systems, we asked: what uncertainties are attributed to  $p\text{CO}_2$  calculations from carbonate equilibria using two  $\text{CO}_2$ -related parameters? We answered this question with a comprehensive error analysis of three-decade long historical observations at the North Temperate Lake Long-Term Ecological Research (NTL-LTER) site. We quantified random errors in pH, DIC, ALK, and temperature measurements and propagated these to estimate uncertainties in  $p\text{CO}_2$  calculated from three carbonate equilibria for four lake groups across a broad gradient of water chemical composition. We also compared uncertainties in a subset of these  $p\text{CO}_2$  observations with directly measured  $p\text{CO}_2$  to determine the relative importance of random to systematic  $p\text{CO}_2$  errors.

## **2 Materials and Methods**

### *2.1 Study site and data collection*

We quantified random error using observations from the NTL-LTER data set for years 1986 – 2011 ([NTL-LTER website](#)). Carbonate system parameters have been measured since 1986 in seven lakes located in northern Wisconsin, USA, and in four lakes in southern Wisconsin since 1996. The northern lakes are located in the Northern Highland Lake District (NHLD) which has a mosaic of mixed, hardwood, and coniferous forests (~53 % of total area), wetlands (28%), lakes (13%), and other land coverages [Buffam *et al.*, 2010]. Soils in the NHLD are dominated by sandy gravel and gravelly sand with dominance of silicate over carbonate [Attig, 1985]. The southern lakes are located in the Yahara River Lake District

(YRLD), which is dominated by agriculture (65%), urban (20%) land uses and the remainder for forest, wetland, or water bodies [Carpenter *et al.*, 2007]. Soils in YRLD are dominated by glacial tills, most commonly sand, silt and clay accumulated over dolomite and limestone parent geology [Clayton and Attig, 1997]. Differences in soil composition are reflected in heterogeneous water chemical composition of lakes studied (Table 1).

The carbon system parameters: pH, total alkalinity (ALK), and dissolved inorganic carbon (DIC), and water temperature (WT) were measured biweekly (WT), monthly (pH and DIC), or quarterly (ALK). Depending on the lake maximum depth and thermocline depth, the samples were taken from one to six sampling depths. On each sampling occasion, blind-paired samples were collected for all variables except water temperature from a randomly selected depth. To ensure valid comparison across three combinations of input parameters, we used only data with paired measurements for all three CO<sub>2</sub>-related parameters. This limited the analysis to quarterly measurements at one depth per lake.

To prevent CO<sub>2</sub> loss or entrainment, water samples for determination of CO<sub>2</sub>-related parameters were gently collected, avoiding splashing and air exposure. The bottles were rinsed with the water sampled, then filled to the top including overflow, and carefully screwed on the displacement cap. Bottles were checked for the presence of air bubbles by inverting the bottles. Water samples were discarded and refilled again if bubbles were present.

Water samples for pH measurements were collected with a peristaltic pump and tubing to 20 ml scintillation vials with displacement caps to exclude air from the vial. In this study, we used the air-excluded pH samples only. The samples were stored in a cold and dark container to minimize biological activity until just before analysis, and then warmed up in the same container to room temperature. The pH samples were analyzed the same day using a potentiometric method in two laboratories: Hasler Lab in Madison (samples from YRLD) and

Trout Lake Station Lab (samples from NHLD). The electrode syringe barrel sealed with teflon tape around the electrode was conditioned with lake water to be analyzed for at least 15 minutes. After uncovering the electrode filling solution hole, the conditioning solution was removed from the barrel using the three-way valve and aspiration system. The electrode chamber was flushed in and out for several times with 2-ml water samples to be measured. The bottles for pH determination were opened just before analysis to draw of 2-ml aliquots of the water sample for several runs of measurements. The measurements were repeated until two consecutive millivolt readings were  $\pm 1$  mV. Last mV reading was recorded. After analyzing all samples, the mV readings for three buffer solutions: pH 10.00, 7.00, and 4.00 were obtained. The recorded buffer and sample mV values were used to calculate pH values. The pH meters were changed from PHM84 Research pH meter to Orion model 720 pH meter in 1988. Since July 2010, pH was measured using a Radiometer combination pH electrode and Orion 4Star pH meter. The dates of pH electrode replacement were unavailable. The relative accuracy of all pH meters was  $\pm 0.002$  according to manufacturers' specifications.

Dissolved inorganic carbon (DIC) samples were collected with the peristaltic pump, tubing and in-line filtered through 0.40-micron polycarbonate filter into 24 ml glass vials capped with septa, leaving no head space. The samples were not poisoned prior storage and analysis. The samples were refrigerated at 4°C and sent in the shipper to Hasler Lab via FedEx overnight delivery (NHLD samples only). The samples were stored refrigerated and analyzed within two-to-three weeks. After phosphoric acid addition, the samples were analyzed with OI Model 700 Carbon Analyzer (before May 2006) or a Shimadzu TOC-V-csh Total Organic Carbon Analyzer (to date). The detection limit for DIC was 0.15 ppm for the analytical measurement range of 60 ppm. The accuracy and precision of Shimadzu's Analyzer was 1.5% following manufacturer' specification.



Total alkalinity (ALK) samples were collected with the peristaltic pump and tubing to 20 ml HDPE plastic containers with conical caps. The ALK samples were not poisoned prior storage and analysis. The samples were stored refrigerated at 4°C and then sent to Hasler Lab in Madison via Fed Ex overnight delivery (NHLD samples only). The sample were stored refrigerated at 4°C until determination, which typically occurred within two weeks. Prior to analytical determination, samples were brought to room temperature. ALK was determined by titrating water samples to an endpoint pH of circa 3.557 by adding 10  $\mu\text{L}$  increments of 0.05N HCL to 16-ml sample from southern lakes or 0.01N HCL to 4-ml sample from northern lakes. Between February 1986 and November 2001, the alkalinity measurements in four lakes (Trout Lake, Sparkling Lake, Allequash Lake and Big Muskellunge Lake) were made using a Brinkmann 636 Titroprocessor. The detection limit for the gran alkalinity titration was approximately 5  $\mu\text{eq. L}^{-1}$ , for the analytical range spanning to 4000  $\mu\text{eq. L}^{-1}$ . The accuracy of manual alkalinity titration is unavailable.

Water temperature measurements were taken using a YSI Model 58 temporDO meter (before 2011) and a YSI Pro-ODO temporDO meter. The accuracy was  $\pm 0.2^\circ\text{C}$  according to manufacturer's specifications.

## 2.2 Random errors in $\text{CO}_2$ -related parameters

To minimize the impact of outlying observations on distribution and statistical properties of random errors, we removed paired measurements with chemical composition differences larger than 15% following the NTL LTER QA/QC protocol. Many of removed observations were already flagged for different quality control reasons. Quality control led to removal of 8% (58/709) of pH observations in antilog scale, 9.5% (68/709) of ALK measurements, and 2.4% (14/709) of DIC observations.

Because carbonate chemistry data in NTL LTER lakes varied over 1-3 orders of magnitude (Table 2.1), the lakes were grouped into four groups based on ALK and dissolved

organic (DOC) concentrations: two bog lakes with low ALK but high DOC (hitherto called “LB<sub>ALK</sub>”), one clearwater lake with low ALK and low DOC (“LC<sub>ALK</sub>”), four lakes with moderate ALK and low DOC (“M<sub>ALK</sub>”), and four lakes with high ALK and moderate DOC (“H<sub>ALK</sub>”). Grouping lakes also enlarged populations of paired observations to generate reasonable resampling distribution [Chernick and LaBudde, 2011] for error analysis.

To quantify random uncertainties from paired samples, we followed the approach described in Hollinger and Richardson [2005]. For a given parameter ( $P_i$ ) we used a pair of independent measurements ( $X_1, X_2$ ) that were made under identical conditions. Because every measurement ( $X_i$ ) is subject to uncertainties, each parameter value represents the best estimate of the measured constituent plus the random ( $\varepsilon$ ) and systematic ( $\delta$ ) errors. Since no information on systematic errors in CO<sub>2</sub>-related parameters at NTL-LTER site were available for most of records, we initially focused on the effect of random errors only while neglecting the effect of systematic uncertainties on the pair. Thus,  $\varepsilon$  approximated the random variable with mean 0 and standard deviation  $\sigma(\varepsilon)$ . Since the mean difference between two independent measurements ( $X_1 - X_2$ ) was close to zero and two-sample Kolmogorov-Smirnov test showed that random uncertainties were independent and identically distributed, the standard deviation  $\sigma(\varepsilon)$  can be determined from equation 1:

$$\sigma(\varepsilon) = \frac{1}{\sqrt{2}} \sigma(X_1 - X_2) \quad (\text{Eq. 1})$$

Therefore, random errors of each parameter,  $\varepsilon(P_i)$ , were quantified as the standard deviation of the difference of repeated pairs of measurements. We used Shapiro-Wilk test to evaluate the normality of distributions. To understand the impact of non-normality on error distribution, we additionally characterized distributions of parameter errors by fitting the probability density functions (PDFs) using the fitdist function in Matlab R2014b and open-

source codes. For each pdf, the mean, scaling, and shape (if applicable) parameters, and skewness and kurtosis were calculated.

### 2.3 Random errors in $p\text{CO}_2$ estimated from carbonate equilibria

To assess random uncertainty attributed to  $p\text{CO}_2$  estimation, we propagated errors onto  $p\text{CO}_2$  derived from two  $\text{CO}_2$ -related parameters. We used three combinations of two input parameters: pH and DIC ( $p\text{CO}_2$ -pH-DIC equilibrium), pH and ALK ( $p\text{CO}_2$ -pH-ALK equilibrium), and ALK and DIC ( $p\text{CO}_2$ -ALK-DIC equilibrium).

The mass-conservation equation for DIC calculations was defined as:

$$[\text{DIC}] = [\text{H}_2\text{CO}_3] + [\text{HCO}_3^-] + [\text{CO}_3^{2-}] \quad (\text{Eq. 2})$$

Where  $\text{H}_2\text{CO}_3$  is the sum of aqueous  $\text{CO}_2$  and carbonic acid ( $\text{H}_2\text{CO}_3$ ). The alkalinity equation neglected the contribution of non- $\text{CO}_2$  species and was defined as:

$$[\text{ALK}] = [\text{HCO}_3^-] + 2[\text{CO}_3^{2-}] + [\text{OH}^-] - [\text{H}^+] \quad (\text{Eq. 3})$$

For  $p\text{CO}_2$  calculations, we used *in-situ* water temperature, the dissociation constants for freshwaters after *Millero* [1979], and barometric pressure at 1 atmosphere. The influence of ionic strength was neglected and all calculations were performed in pH NBS scale. Calculations were performed with the MATLAB-version of the  $\text{CO}_2$  System Calculations [i.e., CO2SYS, *van Heuven et al.*, 2011]. The sets of equations for three parameter pairs are described in [*Dickson et al.*, 2007].

A bootstrap method was used to propagate parameter errors onto carbonate equilibria equations to estimate random  $p\text{CO}_2$  errors. This approach uses empirical data and does not introduce any assumptions about error population distributions [*Chernick and LaBudde*, 2011]. It also provides more realistic estimates of random uncertainty because allows partial cancelation of errors. At each iteration, the random error for each parameter were bootstrapped with substitution from the error pools specific for each alkalinity group. The parameters' error terms were simultaneously applied to the entire population of observations

within each ALK group representing a full spectrum of chemical and temperature ranges observed in lakes. Applying errors to all observations at once permitted to investigate errors propagating through equations, not errors in individual observations [Yanai *et al.*, 2010]. The covariance between errors in parameter pairs was close to zero except for ALK and DIC errors in  $M_{ALK}$  and  $H_{ALK}$  groups, however, all error pairs were statistically uncorrelated. Since parameter measurements were independent and uncorrelated, parameter-specific random errors were propagated independently without covariance term added.

We propagated random parameter errors 10000 times, and at each iteration, computed the population  $pCO_2$  median. For highly-skewed or heavy-tailed distributions, like distributions found for parameters and random errors (Table 2, Fig. S.1), the sample median is considered a good measure of central tendency [Chernick and LaBudde, 2011]. From distribution of 10000 population  $pCO_2$  medians, we inferred the properties of random  $pCO_2$  error distribution. The population median described the best estimate of  $pCO_2$  within each ALK group, while the standard error of the medians population described  $pCO_2$  uncertainty. For each ALK group, we propagated random errors in three runs, one run with errors in both parameters, and two runs with errors in each parameter separately. The uncertainty attributed to bootstrapping accounted for <1% in all three carbonate equilibria.

Unlike chemical components of water, no duplicate observations of water temperature exist in NTL-LTER dataset. To calculate temperature repeatability, we used high-frequency buoy temperature observations to randomly draw 10000 temperature pairs spaced over a short period of time and under similar climatic conditions. The random temperature error was  $\pm 0.23^\circ C$ . Since a long-term behavior of temperature random was unknown, we used sensitivity analysis to quantify upper bounds of uncertainty due to temperature. Long-term means of carbonate parameters were kept constant while propagating error over temperature ranges of  $0-25^\circ C$  to estimate temperature effect for isochemical water.

Although our goal was to demonstrate uncertainty propagating onto  $p\text{CO}_2$  derived from two carbonate parameters, we also acknowledge that using just two  $\text{CO}_2$ -related parameters may result in spurious  $p\text{CO}_2$  estimates in some ALK groups. Therefore, we ran additional simulations: after correcting organic acids contribution to total alkalinity in the  $\text{LB}_{\text{ALK}}$  group and after considering ionic strength influences in the  $\text{H}_{\text{ALK}}$  group. In the  $\text{LB}_{\text{ALK}}$  group,  $1 \mu\text{M}$  of ALK was subtracted for each  $1 \text{ mg L}^{-1}$  of DOC before running ALK-based equilibria. The pool of observations for  $p\text{CO}_2$  calculations decreased by 30% as negative ALK observations were removed before simulations. Additionally, we corrected thermodynamic constant for influences of ionic strength before  $p\text{CO}_2$  calculations in the  $\text{H}_{\text{ALK}}$  group. Only one hardwater lake (Lake Wingra) had historical major ions measurements and the same ionic strength was applied to all observations within  $\text{H}_{\text{ALK}}$  group. Estimated ionic strength was calculated from a Debye-Hückel equation [Brezonik and Arnold, 2011] and accounted for  $I=0.0091 \text{ M}$ . The activity coefficients were calculated from an extended Debye-Hückel equation [Brezonik and Arnold, 2011].

The  $p\text{CO}_2$  values estimated from  $p\text{CO}_2$ -ALK-DIC equilibrium in  $\text{LC}_{\text{ALK}}$ ,  $\text{M}_{\text{ALK}}$ , and  $\text{H}_{\text{ALK}}$  groups were biased relative  $p\text{CO}_2$  calculated from  $p\text{CO}_2$ -pH-DIC and  $p\text{CO}_2$ -pH-ALK equilibria because observations with DIC concentrations smaller than ALK concentrations ( $\text{DIC} < \text{ALK}$ ) were removed to enable  $p\text{CO}_2$ -ALK-DIC equilibrium to solve for pH over 10000 iterations in these ALK groups. Removing  $\text{DIC} < \text{ALK}$  (more alkaline) observations overestimated the median  $p\text{CO}_2$  by 22, 123, 1448  $\mu\text{atm}$  in the  $\text{LC}_{\text{ALK}}$ ,  $\text{M}_{\text{ALK}}$ , and  $\text{H}_{\text{ALK}}$  groups, respectively. Removed observations accounted for 4 (4.2%) observations in the  $\text{LC}_{\text{ALK}}$  group, 77 (11.7%) in the  $\text{M}_{\text{ALK}}$  group, and 101 (34.6%) in the  $\text{H}_{\text{ALK}}$  group (Text S.2, Fig. S.2). Negative ALK observations were removed before propagating errors in  $p\text{CO}_2$ -pH-ALK and  $p\text{CO}_2$ -ALK-DIC equilibria.

#### 2.4 Comparison of uncertainties in direct and indirect $p\text{CO}_2$ measurements

To determine how much uncertainty between direct and indirect measurements can be explained by random  $p\text{CO}_2$  errors, we took direct measurements of  $p\text{CO}_2$  together with carbonate chemistry measurements for a limited number of observations ( $n=21$ ). The measurements of  $\text{CO}_2$ -related parameters were taken according sampling and handling protocols described above. The mole fraction of  $\text{CO}_2$  (ppmv) at 0.1 m depth was directly measured using a non-dispersive infrared Vaisala CARBOCAP  $\text{CO}_2$  probe enclosed in waterproof, gas permeable polytetrafluoroethylene (PTFE) membrane following the *Johnson et al.*, [2010] approach. Each measurement was taken over minimum 16 minutes, first allowing the probe to equilibrate to environmental conditions (which generally occurred within 15 minutes), and then taking a one-minute measurement at one second intervals.

The probes were calibrated against gas standards in the laboratory before each field campaign to evaluate if the probe performs within the manufacturers' accuracy specifications and to identify potential sensor drifts. The equilibration time lasted approximately 15 minutes. Once probe was equilibrated, the average of the last 60 records was assumed to represent the measured  $\text{CO}_2$  concentrations. These values were linearly fitted to calibration curves. The probe accuracy and precision were 1.5% of the range and 2% of the reading at 25° C and 1 atmosphere for a range of 0-3000 ppm according to manufacturer specifications. However, the median lab accuracy was -4.2% of gas standards at 22°C and 1013 hPa. The median precision under repeatability conditions [ISO, 2004] was quantified from randomly selected 10000 of 60-second intervals since equilibration and repeatability standard deviation accounted for  $\pm 1\%$  (lab) and  $\pm 4.2\%$  (field).

The post-measurement corrections of calibrated values were applied to compensate temperature and barometric pressure differences relative to manufacturer's factory settings (i.e., 1013 hPa at 25°C), 1.5 ppmv  $\text{CO}_2$  increase per 1 hPa barometric pressure decrease and 3

ppmv CO<sub>2</sub> decline per 1°C water temperature decrease. These corrections were derived empirically by the manufacturer. Additionally, the CO<sub>2</sub> values were lowered by 14.7 ppm to compensate for pressure on probes at 10 cm measurement depth [Johnson *et al.*, 2010]. The partial pressure of CO<sub>2</sub> ( $p\text{CO}_2$ ) in  $\mu\text{atm}$  were calculated as a product of mole fraction and barometric pressure at 1 atmosphere.

Finally, for each data point, we used estimated random temperature error and the combination of corresponding paired measurements of pH, DIC, and ALK to calculate  $p\text{CO}_2$ . The random parameter errors specific for each alkalinity group were propagated onto each  $p\text{CO}_2$  observation over 10000 iterations. Each  $p\text{CO}_2$  estimates and its random error constituted the median and standard deviation of 10000 medians.

### 3 Results

#### 3.1 Random errors in CO<sub>2</sub>-related parameters

The analysis of nearly 600 paired samples showed that most random errors in CO<sub>2</sub>-related parameters measurements were relatively small relative to parameters' medians when pooled by alkalinity group (Table 1-2). The random error standard deviation ( $\sigma_n$ ), the estimate of measurements' precision, in pH measurements was  $\pm 0.02$  across all ALK groups and was below 0.4% of the median. Unlike errors in pH, the random errors in ALK and DIC measurements increased with the magnitude of parameters' measurements. The random ALK errors ranged from  $\pm 0.3 \mu\text{M}$  in  $\text{LC}_{\text{ALK}}$  group to  $\pm 56.4 \mu\text{M}$  in  $\text{H}_{\text{ALK}}$  group. Similarly, the smallest random DIC errors were in  $\text{LC}_{\text{ALK}}$  group,  $\pm 3.4 \mu\text{M}$ , while the largest errors were in  $\text{H}_{\text{ALK}}$  group,  $\pm 83 \mu\text{M}$ . However, when expressed in relative measures, the random uncertainty was the largest in low ALK groups, accounting for 15.94% of the median ALK in  $\text{LB}_{\text{ALK}}$  group and 5.5% of the median DIC in  $\text{LC}_{\text{ALK}}$  group. The parameter error magnitudes were also independent of the season, year, water temperature, and measurement depth (Fig. S.3).

The empirical distributions of random parameter errors were generally symmetrical around the mean with skewness values close to zero (Table 2, Fig. S.1). Although the kurtosis for pH across all ALK groups was close to kurtosis values observed in normally distributed data (typically  $<3$ ), the errors in ALK and DIC were strongly leptokurtic with characteristic high peaks near the mean difference and heavy tails. Gaussian distributions were confirmed for random pH errors in  $LB_{ALK}$  and  $LC_{ALK}$  groups only (Shapiro-Wilk test,  $p < 0.05$ ). Hence,  $\sigma_n$  would inadequately characterize the parameters' error dispersion in ALK and DIC measurements.

A *t* location-scale distribution (*tLocat*) best characterized distribution of random errors in ALK and DIC (Fig. S.1). This distribution has an additional, shape parameter ( $\nu$ ), where small values indicate heavy tails in error distributions and sensitivity to outliers. Low  $\nu$  values ( $<2$ ) indicates undefined variance and were found in DIC error distributions across all ALK groups and in ALK error distributions in moderate to high ALK groups (Table 2). For the pH uncertainties, the *tLocat* distribution provided only a slightly better fit compared to normal distribution and approached a normal distribution in the  $H_{ALK}$  group. Other probability density functions accepting negative values of random parameter errors did not improve the data fit (data not shown).

### 3.2 Random errors in $pCO_2$ estimated from carbonate equilibria

Random  $pCO_2$  errors propagating through ALK-based equilibria were always higher than  $pCO_2$  errors propagating through  $pCO_2$ -pH-DIC equilibrium (Table 3). The  $pCO_2$ -pH-ALK and  $pCO_2$ -ALK-DIC equilibria showed 6 and 0.2 times higher sensitivities to random parameter errors than  $pCO_2$ -pH-DIC equilibrium in the  $LB_{ALK}$  group. In the  $LC_{ALK}$  group, the random  $pCO_2$  ( $pH$ -ALK) and  $pCO_2$  (ALK-DIC) errors were nearly twofold higher compared to random  $pCO_2$  ( $pH$ -DIC) errors. While differences between random  $pCO_2$  ( $pH$ -DIC) and  $pCO_2$  ( $pH$ -ALK) errors were within 10% of the median  $pCO_2$  in  $M_{ALK}$  and  $H_{ALK}$  groups, the random



$p\text{CO}_2$  (ALK-DIC) errors were 5 times (in the  $M_{\text{ALK}}$  group) and 30 times (in the  $H_{\text{ALK}}$  group) higher relative to  $p\text{CO}_2$  (pH-DIC) errors. Random temperature error additionally contributed  $<1 \mu\text{atm}$  to random  $p\text{CO}_2$  errors across all alkalinity groups and carbonate equilibrium used (Fig. S.5).

Single-parameter propagation revealed that random DIC error was a dominant source of uncertainty in  $p\text{CO}_2$ -pH-DIC equilibrium in acidic ALK groups and  $p\text{CO}_2$ -ALK-DIC equilibrium across all ALK groups (Table 3). In contrast,  $p\text{CO}_2$  calculated from pH-based equilibria were mostly affected by random pH errors in alkaline ALK groups. Random ALK error dominated random  $p\text{CO}_2$  (pH-ALK) errors in the  $LC_{\text{ALK}}$  group.

Regardless of carbonate equilibrium used, random  $p\text{CO}_2$  errors were exponentially proportional to the median  $p\text{CO}_2$  (Table 3, Fig. S.4). While random  $p\text{CO}_2$  (pH-DIC) and  $p\text{CO}_2$  (pH-ALK) errors showed similar pattern of uncertainties with the highest errors in the  $LB_{\text{ALK}}$  group and similar in the remainder of ALK groups, the random  $p\text{CO}_2$  (ALK-DIC) errors were higher at both ends of alkalinity gradients (in  $LB_{\text{ALK}}$  and  $H_{\text{ALK}}$  groups) with the lowest  $p\text{CO}_2$  error in the  $LC_{\text{ALK}}$  group. Extreme  $p\text{CO}_2$  errors corresponded to unrealistically high estimates of median  $p\text{CO}_2$ :  $15225 \pm 1026 \mu\text{atm}$  ( $10999 \pm 935 \mu\text{atm}$  after adjustment) derived from  $p\text{CO}_2$ -pH-ALK equilibrium in the  $LB_{\text{ALK}}$  group, and  $3725 \pm 1156 \mu\text{atm}$  ( $3129 \pm 985 \mu\text{atm}$  after adjustment) calculated from  $p\text{CO}_2$ -ALK-DIC equilibrium in the  $H_{\text{ALK}}$  group (Table 3).

Adjustments for contribution of organic acids to total alkalinity in humic lakes ( $LB_{\text{ALK}}$  group) and for ionic strength in highly buffered lakes ( $H_{\text{ALK}}$  group) led to decreased random  $p\text{CO}_2$  errors (Table 3, Fig. S.4). The random  $p\text{CO}_2$  (pH-ALK) error declined by 9%, from  $\pm 1026 \mu\text{atm}$  to  $\pm 935 \mu\text{atm}$  in the  $LB_{\text{ALK}}$  group, however, adjusted values were still fivefold higher compared to random  $p\text{CO}_2$  (pH-DIC) errors. Adjusted  $p\text{CO}_2$  (ALK-DIC) errors in this group nearly doubled. After correcting for ionic strength in the  $H_{\text{ALK}}$  groups, the random  $p\text{CO}_2$  errors decreased by 15% across three carbonate equilibria.

### 3.3 Comparison of uncertainties in direct and indirect $p\text{CO}_2$ measurements

Random error deviation in direct  $p\text{CO}_2$  measurements expressed here as mean absolute deviation (MAD) were 14, 5, 10, and 3  $\mu\text{atm}$ , in  $\text{LB}_{\text{ALK}}$ ,  $\text{LC}_{\text{ALK}}$ ,  $\text{M}_{\text{ALK}}$ , and  $\text{H}_{\text{ALK}}$  groups, respectively (Table S.1). MAD for indirect  $p\text{CO}_2$  measurements accounted for at least 32, 10, 11, 35  $\mu\text{atm}$  in these four ALK groups. The random  $p\text{CO}_2$  errors between direct and indirect observations cumulatively explained from 3.6% to 32% of the root-mean-square error (RMSE) between direct and indirect  $p\text{CO}_2$  observations.

Although carbonate parameter measurements were collected simultaneously with direct  $p\text{CO}_2$  measurements, the calculated  $p\text{CO}_2$  failed to reproduce direct  $p\text{CO}_2$  except for median  $p\text{CO}_2$  derived from pH and DIC in the  $\text{LC}_{\text{ALK}}$  group (Fig.1, Table S.1). Equilibria  $p\text{CO}_2$ -pH-DIC and  $p\text{CO}_2$ -pH-ALK tended to overestimate  $p\text{CO}_2$ , while  $p\text{CO}_2$ -DIC-ALK equilibrium generally underestimated  $p\text{CO}_2$  (Fig. 1). The largest mismatch with directly measured  $p\text{CO}_2$  occurred for observations in the  $\text{LB}_{\text{ALK}}$  group calculated from  $p\text{CO}_2$ -pH-ALK equilibrium and in the  $\text{H}_{\text{ALK}}$  group calculated from three equilibria. These discrepancies persisted even after adjusting  $p\text{CO}_2$  values for influences of organics acids and ionic strength.

## 4 Discussion

### 4.1 Random errors in $\text{CO}_2$ -related parameters

Most random errors were relatively small relative to median parameter values when pooled by lake type (Table 1-2). The standard deviations of carbonate parameters derived for a normal distribution ( $\sigma_n$ ) were compared with published precision values given  $\sigma_n$  is typically used to characterize random uncertainties. The pH precision values presented in Table 2 were among the most precise potentiometric pH measurements reported for freshwaters, which ranged from  $\pm 0.01$  to  $\pm 0.17$  pH units [Herczeg and Hesslein, 1984; French et al., 2002; Phillips et al., 2015]. The reported DIC and ALK uncertainties in low to moderate ALK groups agreed with published precision values  $< 12 \mu\text{M}$  [Wilkinson et al., 1992; Baehr and

*DeGrandpre, 2002; Abril et al., 2015*]. However, random errors in highly buffered waters ( $H_{ALK}$  group) were higher by approximately four-fold for ALK and sevenfold for DIC (Table 2) than ever reported.

Although the tLocat distribution provided a better fit over Gaussian distribution in ALK and DIC observation within  $M_{ALK}$  to  $H_{ALK}$  groups, neither of these two distributions nor any other probability density functions sufficiently characterized error distribution in some ALK groups (i.e. pH in  $M_{ALK}$ , Fig. S.1, Table 2). Our results agree with  $CO_2$ -related studies showing non-Gaussian distributions of random errors [*Richardson and Hollinger, 2005; Ciais et al., 2013b; Cueva et al., 2015*], and imply that errors derived from normal distribution will underestimate both small and large random errors. Furthermore, the variety of error distributions limits the use of statistical and modeling techniques (i.e., assuming normal distribution) in characterizing random parameter uncertainties and propagating them onto  $pCO_2$  derived from carbonate equilibria.

The heavy tails in the PDFs of random parameter errors might also be indicative of a quality control problem that warrants further evaluation of NTL LTER data. The outlying observations with large concentration differences between duplicates, expressed as high kurtosis and low tLocat shape parameter  $\nu$  (Table 2), were present despite removing duplicate pairs that differed more than 15% (Fig. S.1). The concentration differences suggest that water samples' chemical composition can significantly change between sample collection and analytical analysis. Potential sources behind changing constituents' composition in duplicate samples include lack of sample poisoning to stop biological activity [*Dickson et al., 2007; Åberg and Wallin, 2014*], taking unfiltered ALK samples with substantial quantity of acid-neutralizing particles [*Abril et al., 2015*], DOC interference with pH electrode [*Herczeg and Hesslein, 1984*], sample transport to another lab, or long shelving time. The listed potential errors are systematic, so unlike random uncertainties, cannot be evaluated from duplicate

samples. This finding warrants further targeted efforts towards quantifying and reducing errors in NTL LTER site.

Larger uncertainty of certain measurements in some ALK groups might also indicate that the behavior of systematic errors may vary significantly under certain conditions. For example, the measurements of ALK in humic lakes ( $LB_{ALK}$  group) and pH in highly buffered and productive lakes ( $H_{ALK}$  group) were particularly vulnerable, with 46% and 17% of paired observations failing the QA/QC criterion. Furthermore, even though the random parameter errors were generally below 2% of the median, the uncertainties in ALK and DIC measurements exceeded 5% in low ALK groups (Table 1-2). These results may suggest the presence of systematic biases in the measurements in these groups (and likely in other ALK groups) and potential challenges for correcting historical observations for these biases.

#### 4.2 *Random errors in $pCO_2$ estimated from carbonate equilibria*

The cumulative effect of random parameter errors on  $pCO_2$  calculations across the alkalinity gradient showed that  $pCO_2$  sensitivity to parameter errors varied by the choice of input parameter pairs and alkalinity group. Although the parameter errors were generally below  $\pm 2\%$  of parameters' median values across all ALK groups (Table 1-2), unadjusted  $pCO_2$  errors ranged from  $\pm 3.7\%$  to  $\pm 31.5\%$ , depending on parameter pairs and lake ALK group (Table 3).

Among the three equilibrium models, the  $pCO_2$ -pH-DIC equilibrium was consistently the least sensitive to random parameter errors (Table 3), while  $pCO_2$  estimates calculated from ALK and DIC were the most uncertain. Since ALK-based equilibria require an additional step of calculating DIC or pH, random errors essentially propagated multiple times through the non-linear equations, unlike in the  $pCO_2$ -pH-DIC equilibrium, where they only propagate once.  $pCO_2$ -ALK-DIC equilibrium was additionally prone to errors due to similar DIC and ALK values [Dickson and Riley, 1978]. Nonetheless, the highest attainable precision

of  $p\text{CO}_2$  estimates was  $\pm 36 \mu\text{atm}$  (5.5%) in the  $\text{LC}_{\text{ALK}}$  group and  $\pm 45 \mu\text{atm}$  (3.7%) in the  $\text{M}_{\text{ALK}}$  group given random uncertainty of input parameter pairs.

Our repeatability estimates for  $p\text{CO}_2$ -pH-DIC ( $\pm 3.7\%$ - $5.5\%$ ) were within range of the few existing studies for freshwaters, where the precision of calculated  $p\text{CO}_2$  ranged from 3% to 5% [Herczeg and Hesslein, 1984; Baehr and DeGrandpre, 2004]. Since our lowest  $p\text{CO}_2$  errors were at least 20 times higher than those reported for seawaters [Millero, 2007], the results imply low precision of  $p\text{CO}_2$  estimates based on historical pH, DIC and ALK data in our study lakes.

The random  $p\text{CO}_2$  errors were generally proportional to median  $p\text{CO}_2$  rather than magnitudes of random error in input parameter pairs (Fig. S.4, Table 3). Similar random pH errors ( $\pm 0.02$ ) across all ALK groups (Table 2) contributed to 27%, 25%, 75%, and 90% to random  $p\text{CO}_2$  ( $\text{pH-DIC}$ ) errors in  $\text{LB}_{\text{ALK}}$ ,  $\text{LC}_{\text{ALK}}$ ,  $\text{M}_{\text{ALK}}$ , and  $\text{H}_{\text{ALK}}$  groups, respectively (Table 3). Since median  $p\text{CO}_2$  depends more on the median parameter values (where pH redistributes DIC species accordingly), the improved accuracy of input parameter pairs will have greater impact on increasing the  $p\text{CO}_2$  estimates' precision than improving the parameter's precision.

Single-parameter errors contributed to random  $p\text{CO}_2$  errors nonlinearly in different ALK groups (Table 3). Random  $p\text{CO}_2$  ( $\text{pH-DIC}$ ) and  $p\text{CO}_2$  ( $\text{pH-ALK}$ ) errors were more prone to DIC and ALK errors in acidic waters whereas to pH errors in alkaline waters (Table 1, Table 3). Similar exponentially increasing sensitivity to ALK errors with declining buffering capacity were reported for  $p\text{CO}_2$ -pH-ALK equilibrium [Abril *et al.*, 2015]. Thus, a priority should be placed on closer evaluation of parameters that are key sources of  $p\text{CO}_2$  uncertainty in acidic/alkaline ALK lake groups.

The random error temperature effects on random  $p\text{CO}_2$  errors was negligible and below detection limit relative to random parameter uncertainties contributing to  $p\text{CO}_2$  (Fig. S.5, Table 3). While the maximum random temperature effect was  $1 \mu\text{atm}$ , the minimum

effects were 9  $\mu\text{atm}$ , 14  $\mu\text{atm}$  (adjusted), 12  $\mu\text{atm}$  (adjusted) in pH, DIC, and ALK, respectively. The logarithmic coefficients of temperature effect on  $p\text{CO}_2$  ( $\frac{\partial \ln p\text{CO}_2}{\partial T}$ ) in our lakes were 0.0108-0.0263 $^{\circ}\text{C}^{-1}$  ( $p\text{CO}_2$ -pH-DIC equilibrium), 0.0274-0.0276 $^{\circ}\text{C}^{-1}$  ( $p\text{CO}_2$ -pH-ALK equilibrium), and 0.0170-0.0197 $^{\circ}\text{C}^{-1}$  ( $p\text{CO}_2$ -ALK-DIC equilibrium), low compared to published values 0.038-0.0384 $^{\circ}\text{C}^{-1}$  for freshwater [Lynch *et al.*, 2010; Atilla *et al.*, 2011] and 0.0423 $^{\circ}\text{C}^{-1}$  for seawater [Takahashi *et al.*, 2002] systems. Our results imply inability to detect temperature effects on  $p\text{CO}_2$  estimated from historical observations of  $\text{CO}_2$ -related parameters.

#### 4.3 Systematic errors in $p\text{CO}_2$ estimation

Lack of agreement between three median  $p\text{CO}_2$  values within each ALK group (Table 3) indicates the presence of systematic biases in input parameters affecting  $p\text{CO}_2$  calculations. Furthermore, these  $p\text{CO}_2$  discrepancies generally corresponded to biases between measured and modeled input parameters. For example, the biases in DIC calculated from pH and ALK accounted for -727  $\mu\text{M}$  (-584  $\mu\text{M}$  adj.) and -20  $\mu\text{M}$  in  $\text{LB}_{\text{ALK}}$  and  $\text{LC}_{\text{ALK}}$  groups, respectively. Similarly, biases in pH derived from ALK and DIC were +0.40 in the  $\text{LB}_{\text{ALK}}$  group and -0.60 in the  $\text{H}_{\text{ALK}}$  groups. Our results imply internal inconsistency of field measurements of  $\text{CO}_2$ -related parameters [Millero, 2007] likely leading to thermodynamically inconsistent  $p\text{CO}_2$  estimates.

The comparison of directly and indirectly measured  $p\text{CO}_2$  revealed poor precision and accuracy of virtually all  $p\text{CO}_2$  estimated values (Fig. 1, Table S.1). Random uncertainty attributed to  $p\text{CO}_2$  estimation was at least severalfold higher than uncertainty associated with direct  $p\text{CO}_2$  measurements, except for two cases ( $p\text{CO}_2$  ( $\text{pH-ALK}$ ) errors in the  $\text{M}_{\text{ALK}}$  group and  $p\text{CO}_2$  ( $\text{pH-DIC}$ ) in the  $\text{LC}_{\text{ALK}}$  group). Moreover, most observations significantly deviated from 1:1 line and resultant RMSEs between measured and modeled  $p\text{CO}_2$  were only partly explained by random uncertainty. Interestingly, indirect observations of  $p\text{CO}_2$  derived from

DIC-based equilibria in acidic waters aligned well along the 1:1 line and had relatively low random errors (Fig. 1). In these systems, we conclude that indirect  $p\text{CO}_2$  can be used in absence of direct  $p\text{CO}_2$  observations.

The relative  $p\text{CO}_2$  probe's field repeatability used in this study were within the reported precision values of direct  $p\text{CO}_2$  measurements ranged from 0.003% to 15% of readings [Herczeg and Hesslein, 1984; Baehr and deGrandpe, 2002, Lynch et al., 2010; Wallin et al., 2014; Abril et al., 2015] for freshwaters. However, these values were at least several times higher than precision for direct  $\text{CO}_2$  measurements for seawater [Millero, 2007]. In the case of reliability of  $p\text{CO}_2$  estimated from carbonate equilibria, earlier studies showed indirect observations to be overestimated [Herczeg and Hesslein, 1984; Gelbrecht et al, 1998; Butman and Raymond et al., 2011; Abril et al., 2015], overestimated under low ALK conditions [Wallin et al., 2014], unbiased [Cole et al., 1994], underestimated [Riera et al., 1999], or within 8% agreement (no directionality given) [Baehr and DeGrandpre, 2004] relative to direct  $p\text{CO}_2$  measurements.

Potential factors contributing to observed mismatches between direct and indirect  $p\text{CO}_2$  measurements include: changing parameter's concentrations between sampling and analytical determination, biased potentiometric pH measurements [Herczeg and Hesslein, 1984; Metcalf et al., 1989; French et al., 2002], probe degradation with time, oversimplification of the formulas for  $p\text{CO}_2$  calculations from carbonate equilibria (e.g., ignoring non- $\text{CO}_2$  acid-bases in ALK and calcium forming ions inference with DIC) [Gelbrecht et al., 1998], errors associated with dissociation constants [Lueker et al., 2000; Millero, 2007], biotic activity [Atilla et al., 2011], and multi-temporal variability.

Though we did show improved reliability of  $p\text{CO}_2$  calculations after adjusting for organic acids and ionic strength, the adjusted  $p\text{CO}_2$  values remained biased relative to direct  $p\text{CO}_2$  measurements and  $p\text{CO}_2$  calculated from other input parameter pairs (Table 3, Fig. 1).

The interferences of organic acids with ALK in low alkalinity and brown waters, and calcium forming ions to DIC in waters with  $\text{pH} > 7$  were previously linked to large  $p\text{CO}_2$  overestimation [Gelbrecht *et al.*, 1998; Hunt *et al.*, 2011; Wang *et al.*, 2013; Abril *et al.*, 2015]. Because of unreliable  $p\text{CO}_2$  estimates in observations with  $\text{pH} < 5.4$  [Raymond *et al.*, 2013],  $\text{pH} < 6$  [Jones *et al.*, 2003], and  $\text{ALK} < 40 \mu\text{M}$  [Cole *et al.*, 1994], a fraction of  $p\text{CO}_2$  estimates were excluded from analysis in earlier studies. In line with Abril *et al.*, [2015], our results also imply large overestimation of  $p\text{CO}_2$  in hard water.

At present, we cannot elucidate which systematic errors contributed most to observed mismatch between direct and indirect  $p\text{CO}_2$  measurements. Our results warrant the need to better quantify systematic errors and other sources of uncertainty attributed to different parameter measurements. While further investigation on a larger pool of observations is necessary to validate the accuracy of  $p\text{CO}_2$  calculations, our results definitively demonstrate that the effect of random parameter errors on  $p\text{CO}_2$  was rather small compared to systematic errors.

#### 4.4 Implications for estimating trends in $p\text{CO}_2$ and C flux

Despite using consistent methodology of sample collection, handling, analytical determination, and quality control, the results from NTL-LTER site indicate large uncertainties arising from  $p\text{CO}_2$  estimation from carbonate equilibria. To test if attainable levels of uncertainty are sufficient to detect long-term change at NTL site, we propagated random and systematic uncertainties onto a  $p\text{CO}_2$  ( $\text{pH-DIC}$ ) time-series at Crystal Lake (LC<sub>ALK</sub> group) (Fig. 2). This lake and carbonate equilibrium consistently showed least uncertain  $p\text{CO}_2$  estimates, hence represents “a best-case scenario” (Fig. 1-2, Fig. S.5, Table 2). Uncertainty around median  $p\text{CO}_2$  ranged from  $\pm 24$  to  $\pm 40 \mu\text{atm}$ , and in some cases, made it impossible to determine whether the lake was classified as undersaturated or supersaturated with respect to atmospheric  $p\text{CO}_2$ . Applying a simple linear fit indicated an insignificant



$p\text{CO}_2$  decrease of  $-2 \pm 26 \mu\text{atm}$  per year regardless of choice of start and end year, leading to a result that is not interpretable given the magnitude of random uncertainty. Moreover, the observed warming water temperature trend of  $0.59^\circ\text{C decade}^{-1}$  [O'Reilley *et al.*, 2015] increased  $p\text{CO}_2$  by  $0.0014 \mu\text{atm}$ , indicating temperature-mediated increase below detection limit given current parameter uncertainties. Our results highlight the challenge of inferring past  $p\text{CO}_2$  variability and change, and detecting lake-carbon responses to warming climate from historical observations of carbonate parameters.

Given all sources of uncertainty attributed to  $p\text{CO}_2$  calculations, we recommend using direct  $p\text{CO}_2$  measurements to constrain  $\text{CO}_2$  flux magnitudes and trends from NTL lakes. Further targeted effort is essential to identify the sources and behavior of systematic errors. Such information is necessary to correct for systematic biases in  $p\text{CO}_2$  calculations from historical observations of pH, DIC, and ALK, and improve the accuracy of measurements and  $p\text{CO}_2$  predictions in the future. Also, the lack of duplicate water temperature and quantifying the accuracies and precisions of laboratory equipment and resultant parameter uncertainties need to be addressed. Establishing protocols for routine direct  $p\text{CO}_2$  measurements is urgently needed.

#### 4.5 Implications for estimating $p\text{CO}_2$ from carbonate equilibria in freshwaters

Given the overall scarcity of direct  $p\text{CO}_2$  data from freshwater systems [i.e., 1%; Raymond *et al.*, 2013], providing confidence levels for  $p\text{CO}_2$  estimates (whether directly or indirectly measured) is critical to assessing the uncertainty on the contributions of aquatic  $\text{CO}_2$  systems to regional and global C cycles. These uncertainty values are essential also for comparing data from different sources and/or derived from multiple carbonate equilibria to include  $p\text{CO}_2$  estimates from more chemically diverse ecosystem types in many regions and account for potential temporal fluctuations of carbonate system.

To aid with estimating random uncertainty in other studies, the probability density functions fitted in this study (Table 2) could help reconstruct characteristics of random parameter errors to be used in Monte Carlo propagation on datasets with chemical characteristics similar to NTL lakes (Table 1). Likewise, the uncertainty data supporting this study are available with this manuscript to assist bootstrap error propagation. The relative errors of carbonate parameters and  $p\text{CO}_2$  estimates (Tables 2-3, Table S.1) could also provide upper bounds of uncertainty. We encourage all reports to provide levels of uncertainty in  $p\text{CO}_2$  and C flux whenever possible.

Limited direct  $p\text{CO}_2$  observations combined with a growing interest in assessing a role of freshwaters in regional and global scales has resulted in studies relying on routinely measured  $\text{CO}_2$ -related parameters to estimate  $p\text{CO}_2$  and carbon flux from inland waters at global [Cole *et al.*, 1994; Cole *et al.*, 2007; Tranvik *et al.*, 2009; Aufdenkampe *et al.*, 2011; Raymond *et al.*, 2013] and regional scales [Buffam *et al.*, 2010; Butman and Raymond, 2011; McDonald *et al.*, 2013; Lapierre *et al.*, 2017], and evaluating temporal trends of inorganic carbon species [Jones, 2003; Raymond and Cole, 2003; Seekel and Gudas, 2016; Nydahl *et al.*, 2017]. We demonstrate however that field measurements of pH, ALK, and DIC might be insufficient to provide robust estimates of mean  $p\text{CO}_2$  and question if they are sensitive enough to detect long-term change in chemically heterogeneous lakes (Fig. 1-2; Table S.1). Previous studies also implied limited value of estimating  $p\text{CO}_2$  from monitoring data [French *et al.* 2002; Abril *et al.*, 2015; Phillips *et al.*, 2015]. However, only reproducing this study approach over larger datasets may help elucidate if insensitivity of reconstructed  $p\text{CO}_2$  to long-term change is site-specific or represents a more widespread methodological issue.

To decrease uncertainty of C flux from freshwaters and enable prediction of future  $\text{CO}_2$  changes, we see two paths forward: reducing uncertainty of current measurements of

CO<sub>2</sub>-related parameters to achieve thermodynamically consistent  $p\text{CO}_2$  estimates and/or using direct CO<sub>2</sub> measurements.

If using carbonate equilibria parameters continues to be the most common method of estimating  $p\text{CO}_2$  in freshwater systems, the freshwater community must determine the acceptable levels of precision and accuracy of carbonate parameter measurements to achieve  $p\text{CO}_2$  estimates sensitive to detect long-term changes. Overall uncertainty of atmospheric CO<sub>2</sub> measurement account for <0.2 ppm [Andrews *et al.*, 2014] hence freshwater component of  $\Delta p\text{CO}_2$  is a dominant source of uncertainty. For comparison, current laboratory measurements of seawater have precision and accuracy  $\pm 1 \mu\text{mol kg}^{-1}$  and  $\pm 2 \mu\text{mol kg}^{-1}$  for DIC,  $\pm 1 \mu\text{mol kg}^{-1}$  and  $\pm 3 \mu\text{mol kg}^{-1}$  for ALK, and  $\pm 0.0004$  and  $\pm 0.002$  for pH measurements to produce fugacity CO<sub>2</sub> ( $f\text{CO}_2$  values are a few  $\mu\text{atm}$  lower than  $p\text{CO}_2$  after accounting for non-ideal nature of gas phase) estimates with uncertainty  $\pm 6 \mu\text{atm}$  or higher [Millero, 2007]. While above analytical uncertainties are incomparable with errors estimates in our lakes (Tables 2-3), they serve as a gold standard that sets the bar for improving freshwater measurements of CO<sub>2</sub>-related parameters.

Adapting solutions already developed for seawater could potentially advance methodological improvements of the CO<sub>2</sub> system measurements in freshwater systems. Using the certified reference materials (CRM) of CO<sub>2</sub> measurements contributed most towards development of fully calibrated dataset with uniformly calculated estimates [Key *et al.*, 2004; Sabine *et al.*, 2004]. The CRM samples were prepared in one certified laboratory and distributed among laboratories to serve as an independent measurement quality test. Furthermore, a unified quality assurance and quality control procedure was applied to compare the results from different research groups and identify laboratories having problems with accuracy and precision. Finally, an unambiguous guide on best practices on CO<sub>2</sub> measurements [Dickson *et al.*, 2007] provided up-to-date information on the chemistry of the

CO<sub>2</sub> system in seawater, well-tested analytical methods of analyzing parameters, and standard operating procedures.

An alternative approach, though not mutually exclusive, would rely on direct CO<sub>2</sub> measurements with supporting measurements of carbonate parameters characterizing chemical composition of freshwater systems. A growing number of recent studies already rely on direct CO<sub>2</sub> measurements [Wallin *et al.*, 2014; Borges *et al.*, 2015; Holgerson and Raymond, 2016]. Given direct CO<sub>2</sub> measurements always produce least uncertain CO<sub>2</sub> in seawater [Dickson and Riley, 1978; Millero, 2007], directly measured *p*CO<sub>2</sub> is recommended in studies aiming at constraining aquatic CO<sub>2</sub> system. Still, reported uncertainties for freshwater measurements suggest the need for significant improvement of the precision and accuracy of direct measurements relative to seawater ( $\pm 0.5 \mu\text{atm}$  precision and  $\pm 2 \mu\text{atm}$  accuracy CO<sub>2</sub> estimates) [Millero *et al.*, 2007]. While we acknowledge difficulty in getting direct *p*CO<sub>2</sub> data at broader spatiotemporal scales, uncertain and insensitive to change past CO<sub>2</sub> estimates imply limited use for predicting future of freshwater CO<sub>2</sub> system.

## 5 Conclusions

We evaluated uncertainties in a widely used method of estimating *p*CO<sub>2</sub> from uncertainties attributed to *p*CO<sub>2</sub> estimation from carbonate equilibria using two of three CO<sub>2</sub>-related parameters. Random parameter errors were typically low for pH, dissolved inorganic carbon, and total alkalinity measurements in lakes of all types of water chemical composition (humic, poorly buffered clearwater, moderate alkalinity, and hardwater). However, we found that these relatively low random parameter errors could still produce random *p*CO<sub>2</sub> error approaching one third of estimated median *p*CO<sub>2</sub>, depending on the choice of input parameter pairs and lake alkalinity group. Further, the comparison of direct and indirect *p*CO<sub>2</sub> observations revealed that nearly all parameter combinations produces biased *p*CO<sub>2</sub> estimates with systematic errors greatly exceeding random *p*CO<sub>2</sub> errors. Past *p*CO<sub>2</sub> time series in a

"best-case scenario" demonstrated undetectable long-term change owing high overall uncertainty.

As virtually no choice of input parameter pairs consistently provided reliable and reproducible  $p\text{CO}_2$  estimates, we recommend direct  $p\text{CO}_2$  measurements in studies aiming for precise and accurate estimation of  $p\text{CO}_2$  and carbon flux from inland waters. While we acknowledge limitations associated with getting direct  $\text{CO}_2$  observations at broader spatiotemporal scales, the further use of indirect  $p\text{CO}_2$  estimates without reducing uncertainty in current measurements of  $\text{CO}_2$ -related parameters will significantly hinder predictions of future freshwater  $p\text{CO}_2$  in face of anthropogenic pressures on aquatic systems. Additional work is needed to quantify key systematic errors and overall uncertainty, and determine acceptable levels of precision and accuracy to achieve robust  $p\text{CO}_2$  estimates enabling detection of temporal changes. Given bounds of uncertainty in global C emissions from lakes and reservoirs ( $0.32 \text{ Pg C yr}^{-1}$ ) range from  $0.06$  to  $0.84 \text{ Pg C yr}^{-1}$ , and from rivers and streams ( $1.8 \text{ Pg C yr}^{-1}$ ) vary from  $1.5$  to  $2.1 \text{ Pg C yr}^{-1}$  [Raymond *et al.*, 2013], freshwater researchers must make significant efforts to standardize and reduce errors in  $p\text{CO}_2$  predictions.

### **Acknowledgments**

We are grateful to the North Temperate Lakes Long-term Ecological Research (LTER) staff that has collected and analyzed data over the past 26 years. We also thank J. Thom, K. Xu, L. Loken for their assistance in the field measurements and calculations. We are also grateful for the critical and useful comments of two anonymous reviewers. This research was supported by funding from the National Science Foundation LTER Program for North Temperature Lakes LTER (NSFDEB-1440297, NTL LTER) and NSF DEB-0845166, and University of Wisconsin Anna Grant Birge Memorial Award For Graduate Research awarded to M. Golub. Supporting information, including data files used in this study are included in the electronic

supplement. All raw data used in this study are available from the NTL-LTER data portal (<https://lter.limnology.wisc.edu/data>).

## References

- Åberg, J., & Wallin, M. B. (2014). Evaluating a fast headspace method for measuring DIC and subsequent calculation of  $p\text{CO}_2$  in freshwater systems. *Inland Waters*, 4, 157–166. doi:10.5268/IW-4.2.694
- Abril, G., Bouillon, S., Darchambeau, F., Teodoru, C. R., Marwick, T. R., Tamooh, F., ... & Borges, A. V. (2015). Technical Note: Large overestimation of  $p\text{CO}_2$  calculated from pH and alkalinity in acidic, organic-rich freshwaters. *Biogeosciences*, 12, 67–78. doi:10.5194/bg-12-67-2015
- Andrews, A. E., Kofler, J. D., Trudeau, M. E., Williams, J. C., Neff, D. H., Masarie, K. A., ... & Tans, P. P. (2014).  $\text{CO}_2$ ,  $\text{CO}$ , and  $\text{CH}_4$  measurements from tall towers in the NOAA Earth System Research Laboratory's Global Greenhouse Gas Reference Network: instrumentation, uncertainty analysis, and recommendations for future high-accuracy greenhouse gas monitoring efforts. *Atmospheric Measurement Techniques*, 7(2), 647–687. doi:10.5194/amt-7-647-2014
- Atila, N., McKinley, G. A., Bennington, V., Baehr, M., Urban, N., DeGrandpre, M., Desai, A. R., & Chin, W. (2011). Observed variability of Lake Superior  $p\text{CO}_2$ . *Limnology and Oceanography*, 56(3), 775–786. doi:10.4319/lo.2011.56.3.0775
- Attig, J. W. (1985). *Pleistocene geology of Vilas County*, Wisconsin Geological and Natural History Survey, Information Circular 50, Wisconsin.
- Aufdenkampe, A. K., Mayorga, E., Raymond, P. A., Melack, J. M., Doney, S. C., Alin, S. R., Aalto, R. E., & Yoo, K. (2011). Riverine coupling of biogeochemical cycles between land, oceans, and atmosphere. *Frontiers in Ecology and the Environment*, 9(1), 53–60, doi:10.1890/100014
- Baehr, M. M., & DeGrandpre, M. D. (2002). Under-ice  $\text{CO}_2$  and  $\text{O}_2$  variability in a freshwater lake. *Biogeochemistry*, 61(1), 95–113. doi:10.1023/A:1020265315833
- Baehr, M. M., & DeGrandpre, M. D. (2004). In situ  $p\text{CO}_2$  and  $\text{O}_2$  measurements in a lake during turnover and stratification: Observations and modeling. *Limnology and Oceanography*, 49(2), 330–340. doi:10.4319/lo.2004.49.2.0330.
- Borges, A. V., Darchambeau, F., Teodoru, C. R., Marwick, T. R., Tamooh, F., Geeraert, N., ... & Bouillon, S. (2015). Globally significant greenhouse-gas emissions from African inland waters. *Nature Geoscience*, 8, 637–642. doi:10.1038/ngeo2486
- Brezonik, P. L., & Arnold, W. A. (2011). *Water Chemistry: An Introduction to the Chemistry of Natural and Engineered Aquatic Systems*, New York, NY: Oxford University Press.
- Buffam, I., Turner, M. G., Desai, A. R., Hanson, P. C., Rusak, J. A., Lottig, N. R., Stanley, E. H., & Carpenter, S. R. (2011). Integrating aquatic and terrestrial components to construct a complete carbon budget for a north temperate lake district. *Global Change Biology*, 17, 1193–1211. doi:10.1111/j.1365-2486.2010.02313.x
- Butman, D., & Raymond, P. A. (2011). Significant efflux of carbon dioxide from streams and rivers in the United States. *Nature Geoscience*, 4, 839–842. doi:10.1038/ngeo1294
- Carpenter, S. R., Benson, B. J., Biggs, R., Chipman, J. W., Foley, J. A., Golding, S. A., ... & Yuan, H. (2007). Understanding Regional Change: A Comparison of Two Lake Districts. *Bioscience*, 57(4), 323–335. doi:10.1641/B570407
- Chernick, M. R., & LaBudde, R. A. (2011). *An Introduction to Bootstrap Methods with Applications to R*. Hoboken, NJ: John Wiley.
- Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., ... & Thornton, P. (2013). Carbon and Other Biogeochemical Cycles. In: T.F., Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, (pp. 465-570), Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. doi:10.1017/CBO9781107415324.014
- Clayton, L., & Attig, J. W. (1997). *Pleistocene Geology of Dane County Wisconsin*, Wisconsin Geological and

Natural History Survey, Bulletin 95, Madison, WI.

- Cole, J. J., Caraco, N. F., Kling, G. W., & Kratz, T. K. (1994). Carbon dioxide supersaturation in the surface waters of lakes. *Science*, 265(5178), 1568–1570. doi:10.1126/science.265.5178.1568
- Cole, J. J., Prairie, Y. T., Caraco, N. F., McDowell, W. H., Tranvik, L. J., Striegl, R. G., ...& Melack, J. (2007). Plumbing the global carbon cycle: Integrating inland waters into the terrestrial carbon budget. *Ecosystems*, 10(1), 171–184. doi:10.1007/s10021-006-9013-8.
- Cueva, A., Bahn, M., Litvak, M., Pumpanen J., & R. Vargas (2015), A multisite analysis of temporal random errors in soil CO<sub>2</sub> efflux, *J. Geophys. Res. Biogeosciences*, n/a–n/a, doi:10.1002/2014JG002690.
- Dickson, A. G., and J. P. Riley (1978), The effect of analytical error on the evaluation of the components of the aquatic carbon-dioxide system, *Mar. Chem.*, 6(1), 77–85, doi:10.1016/0304-4203(78)90008-7.
- Dickson, A.G., Sabine, C.L., Christian, J. R. (2007), *Guide to best practices for ocean CO<sub>2</sub> measurements*, PICES Special Publication 3.
- French, C. R., J. J. Carr, E. M. Dougherty, L. A. . K. Eidson, J. C. Reynolds, and M. D. DeGrandpre (2002), Spectrophotometric pH measurements of freshwater, *Anal. Chim. Acta*, 453(1), 13–20, doi:10.1016/S0003-2670(01)01509-4.
- Gelbrecht, J., M. Fait, M. Dittrich, and C. Steinberg (1998), Use of GC and equilibrium calculations of CO<sub>2</sub> saturation index to indicate whether freshwater bodies in north-eastern Germany are net sources or sinks for atmospheric CO<sub>2</sub>, *Fresenius. J. Anal. Chem.*, 361(1), 47–53, doi:10.1007/s002160050832.
- Herczeg, A. L., and R. H. Hesslein (1984), Determination of hydrogen ion concentration in softwater lakes using carbon dioxide equilibria, *Geochim. Cosmochim. Acta*, 48(4), 837–845, doi:10.1016/0016-7037(84)90105-4.
- Hollinger, D. Y., and A. D. Richardson (2005), Uncertainty in eddy covariance measurements and its application to physiological models., *Tree Physiol.*, 25(7), 873–85.
- Hunt, C. W., J. E. Salisbury, and D. Vandemark (2011), Contribution of non-carbonate anions to total alkalinity and overestimation of pCO<sub>2</sub> in New England and New Brunswick rivers, *Biogeosciences*, 8(10), 3069–3076, doi:10.5194/bg-8-3069-2011.
- ISO (2004), ISO/TS 21748:2004(E): Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation, ISO, Geneva.
- Johnson, M. S., M. F. Billett, K. J. Dinsmore, M. Wallin, K. E. Dyson, and R. S. Jassal (2010), Direct and continuous measurement of dissolved carbon dioxide in freshwater aquatic systems-method and applications, *Ecohydrology*, n/a–n/a, doi:10.1002/eco.95.
- Jones, J. B., E. H. Stanley, and P. J. Mulholland (2003), Long-term decline in carbon dioxide supersaturation in rivers across the contiguous United States, *Geophys. Res. Lett.*, 30(10), n/a–n/a, doi:10.1029/2003GL017056.
- Key, R. M., A. Kozyr, C. L. Sabine, K. Lee, R. Wanninkhof, J. L. Bullister, R. A. Feely, F. J. Millero, C. Mordy, and T. H. Peng (2004), A global ocean carbon climatology: Results from Global Data Analysis Project (GLODAP), *Global Biogeochem. Cycles*, 18(4), n/a–n/a, doi:10.1029/2004GB002247.
- Khatiwala, S. et al. (2013), Global ocean storage of anthropogenic carbon, *Biogeosciences*, 10(4), 2169–2191, doi:10.5194/bg-10-2169-2013.
- Lapierre, J. F., D. A. Seekell, C. T. Filstrup, S. M. Collins, C. Emi Fergus, P. A. Soranno, and K. S. Cheruvilil (2017), Continental-scale variation in controls of summer CO<sub>2</sub> in United States lakes, *J. Geophys. Res. Biogeosciences*, 122(4), 875–885, doi:10.1002/2016JG003525.
- Lozovik, P. A. (2005), Contribution of Organic Acid Anions to the Alkalinity of Natural Humic Water, *J. Anal. Chem.*, 60(11), 1000–1004, doi:10.1007/s10809-005-0226-3.
- Lueker, T. J., A. G. Dickson, and C. D. Keeling (2000), Ocean pCO<sub>2</sub> calculated from dissolved inorganic carbon, alkalinity, and equations for K<sub>1</sub> and K<sub>2</sub>: Validation based on laboratory measurements of CO<sub>2</sub> in gas and seawater at equilibrium, *Mar. Chem.*, 70, 105–119, doi:10.1016/S0304-4203(00)00022-0.
- Lynch, J. K., C. M. Beatty, M. P. Seidel, L. J. Jungst, and M. D. DeGrandpre (2010), Controls of riverine CO<sub>2</sub> over an annual cycle determined using direct, high temporal resolution p CO<sub>2</sub> measurements, *J. Geophys. Res.*, 115(G3), G03016, doi:10.1029/2009JG001132.
- 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,

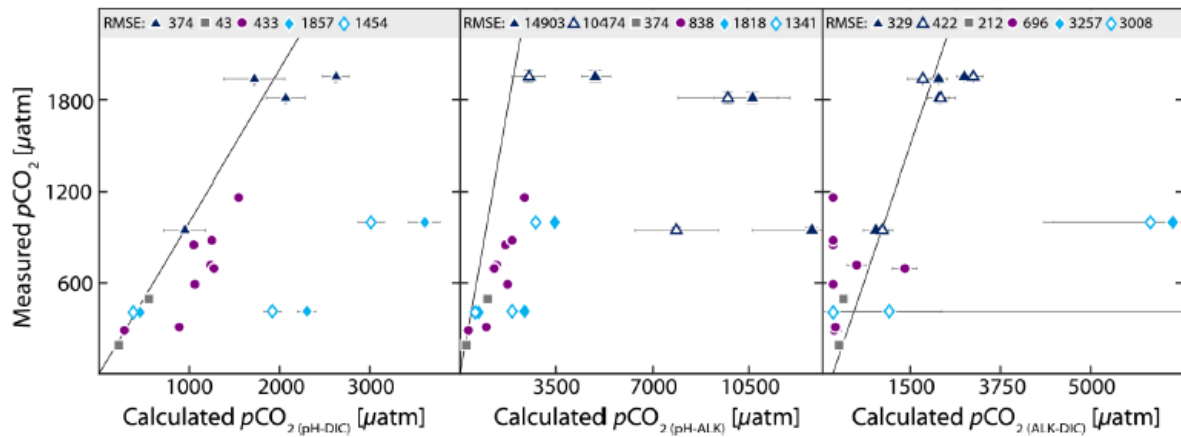
- Global Biogeochem. Cycles*, 27(2), 285–295, doi:10.1002/gbc.20032.
- Metcalf, R. C., D. V. Peck, and a. J. Lori (1989), The influence of dissolved organic carbon on pH measurements of low solute content waters, *Geochim. Cosmochim. Acta*, 53(4), 773–784, doi:10.1016/0016-7037(89)90023-9.
- Millero, F. J. (1979), The thermodynamics of the carbonate system in seawater, *Geochim. Cosmochim. Acta*, 43(10), 1651–1661, doi:10.1016/0016-7037(79)90184-4.
- Millero, F. J. (2007), The marine inorganic carbon cycle, *Chem. Rev.*, 107(2), 308–41, doi:10.1021/cr0503557.
- Nydahl, A. C., M. B. Wallin, and G. A. Weyhenmeyer (2017), No long-term trends in  $p\text{CO}_2$  despite increasing organic carbon concentrations in boreal lakes, streams and rivers, *Global Biogeochem. Cycles*, 1–11, doi:10.1002/2016GB005539.
- O'Reilly, C. M., R. J. Rowley, P. Schneider, J. D. Lenters, P. B. McIntyre, and B. M. Kraemer (2015), Rapid and highly variable warming of lake surface waters around the globe, *Geophys. Res. Lett.*, 1–9, doi:10.1002/2015GL066235.
- Parkhurst, B. D. L., and C. a J. Appelo (1999), *User's Guide To PHREEQC (version 2) — a Computer Program for Speciation, and Inverse Geochemical Calculations*, Water-Resources Investigations Report 99-4259.
- Phillips, J. C., G. A. McKinley, V. Bennington, H. A. Bootsma, D. J. Pilcher, R. W. Sterner, and N. R. Urban (2015), The potential for CO<sub>2</sub>-induced acidification in freshwater: A Great Lakes case study, *Oceanography*, 28(2), 136–145, doi:10.5670/oceanog.2015.37.
- Raymond, P. A. et al. (2013), Global carbon dioxide emissions from inland waters., *Nature*, 503(7476), 355–9, doi:10.1038/nature12760.
- Richardson, A. D., and D. Y. Hollinger (2005), Statistical modeling of ecosystem respiration using eddy covariance data: Maximum likelihood parameter estimation, and Monte Carlo simulation of model and parameter uncertainty, applied to three simple models, *Agric. For. Meteorol.*, 131(3-4), 191–208, doi:10.1016/j.agrformet.2005.05.008.
- Richardson, A. D., Aubinet, M., Barr, A. G., Hollinger, D. Y., Ibrom, A., Lasslop, G., & Reichstein, M. (2012). Uncertainty quantification. In M. Aubinet, T. Vesala, & D. Papale (Eds.), *Eddy Covariance: A Practical Guide to Measurement and Data Analysis*, (pp. 173-209), Springer.
- Riera, J. L., J. E. Schindler, and T. K. Kratz (1999), Seasonal dynamics of carbon dioxide and methane in two clear-water lakes and two bog lakes in northern Wisconsin, U.S.A., *Can J Fish Aquat Sci*, 274, 1–10, doi:10.1139/f98-182.
- Sabine, C. L. et al. (2004), The oceanic sink for anthropogenic CO<sub>2</sub>., *Science*, 305(5682), 367–371, doi:10.1126/science.1097403.
- Seekell, D. A., and C. Gudas (2016), Long-term  $p\text{CO}_2$  trends in Adirondack Lakes, *Geophys. Res. Lett.*, 43(10), 5109–5115, doi:10.1002/2016GL068939.
- Skoog, D. A., D. M. West, F. J. Holler, and S. R. Crouch (2014), *Fundamentals of Analytical Chemistry*, 9th ed., Brooks/Cole, Belmont, CA.
- Takahashi, T., S. Sutherland, C. Sweeney, A. Poisson, et. al (2002), Global Sea-Air CO<sub>2</sub> Flux Based on Climatological Surface Ocean  $p\text{CO}_2$  and Seasonal Biological and Temperature Effects, *Deep. Res. Pt. II*, 49, 1601.
- Tranvik, L. J., J. a. Downing, J. B. Cotner, S. a. Loiselle, R. G. Striegl, T. J. Ballatore, P. Dillon, K. Finlay, K. Fortino, and L. B. Knoll (2009), Lakes and reservoirs as regulators of carbon cycling and climate, *Limnol. Oceanogr.*, 54(6\_part\_2), 2298–2314, doi:10.4319/lo.2009.54.6\_part\_2.2298.
- van Heuven, S., Pierrot, D., Rae, J.W.B., Lewis, E., Wallace, D. V. R. (2011), Program Developed for CO<sub>2</sub> System Calculations. ORNL/CDIAC-105b. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee., , doi:10.3334/CDIAC/otg.CO2SYS\_MATLAB\_v1.1.
- Wallin, M. B., S. Löfgren, M. Erlandsson, and K. Bishop (2014), Representative regional sampling of carbon dioxide and methane concentrations in hemiboreal headwater streams reveal underestimates in less systematic approaches, *Global Biogeochem. Cycles*, 28(4), 465–479, doi:10.1002/2013GB004715.
- Wang, Z. A., D. J. Bienvenu, P. J. Mann, K. A. Hoering, J. R. Poulsen, R. G. M. Spencer, and R. M. Holmes (2013), Inorganic carbon speciation and fluxes in the Congo River, *Geophys. Res. Lett.*, 40(3), 511–516, doi:10.1002/grl.50160.



Wilkinson, K. J., H. G. Jones, P. G. C. Campbell, and M. Lachance (1992), Estimating organic acid contributions to surface water acidity in Quebec (Canada), *Water, Air, Soil Pollut.*, 61(1-2), 57–74, doi:10.1007/BF00478366.

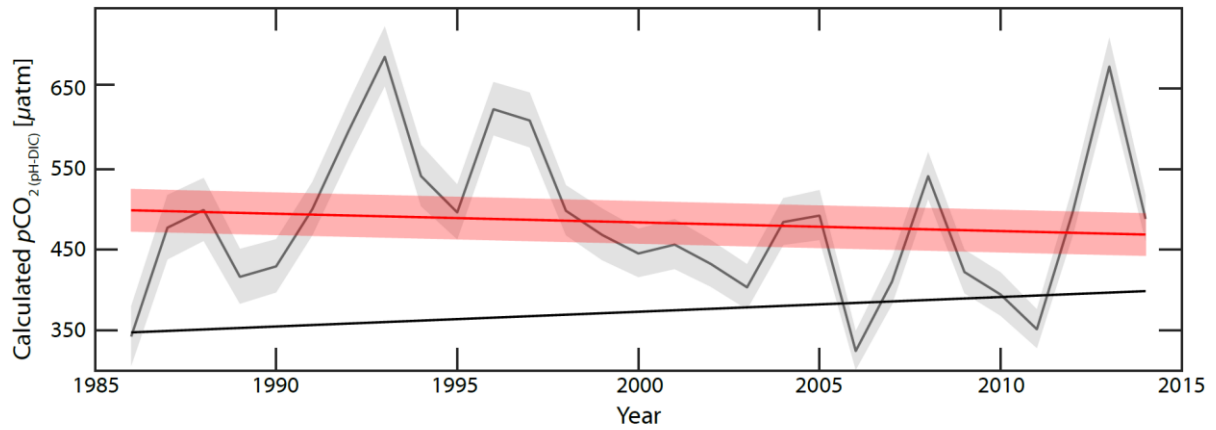
Yanai, R. D., J. J. Battles, A. D. Richardson, C. A. Blodgett, D. M. Wood, and E. B. Rastetter (2010), Estimating Uncertainty in Ecosystem Budget Calculations, *Ecosystems*, 13(2), 239–248, doi:10.1007/s10021-010-9315-8.

Accepted Article



**Figure 1.** The comparison of directly measured  $p\text{CO}_2$  and  $p\text{CO}_2$  estimated from three equilibria:  $p\text{CO}_2$ -pH-DIC (left),  $p\text{CO}_2$ -pH-ALK (middle), and  $p\text{CO}_2$ -DIC-ALK (right), shows high uncertainty and directionality of biases attributed to  $p\text{CO}_2$  estimation. Horizontal and vertical error bars represent standard error of the median. The global root mean square error (RMSE) accounted for 1087 (987 after adjustment), 6944 (5021), and 1688 (1611)  $\mu\text{atm}$ , respectively. The diagonal line represents a theoretical 1:1 relationship. The observations coding:  $L_{\text{ALK}}$  group – navy triangles;  $L_{\text{ALK}}$  group adjusted – open triangles;  $LC_{\text{ALK}}$  group – grey squares;  $M_{\text{ALK}}$  group - purple circles;  $H_{\text{ALK}}$  group – blue diamonds;  $H_{\text{ALK}}$  group adjusted – open diamonds. One extreme outlier estimated from  $p\text{CO}_2$ -pH-ALK is outside of a figure frame.

Accepted



**Figure 2.** The time-series of  $p\text{CO}_2$  estimates with propagated uncertainties attributed to random and systematic errors and different beginnings and endings of time series for Crystal Lake demonstrates scatter around best  $p\text{CO}_2$  estimates (median of 10000 median values) and fitted trend line (median of 10000 fitted linear regression). Confidence intervals represent standard error of the median for near-monthly observation (light gray) and trend (light red). Black solid line represents mean annual atmospheric  $\text{CO}_2$  at 1 atmosphere. Slope line indicates statistically insignificant decline of  $2 \pm 26 \mu\text{atm}$  per year.

Accepted

**Table 1.** Chemical characteristics within lake alkalinity groups (low to high); the values represent the median and 5<sup>th</sup> and 95<sup>th</sup> percentiles

| Variable                 | LB <sub>ALK</sub> <sup>1</sup>     | LC <sub>ALK</sub> <sup>2</sup>   | M <sub>ALK</sub> <sup>3</sup>    | H <sub>ALK</sub> <sup>4</sup>    |
|--------------------------|------------------------------------|----------------------------------|----------------------------------|----------------------------------|
| pH                       | 5.03 (4.58-5.52)<br>(n = 2136)     | 6.32 (5.62 – 7.06)<br>(n = 1543) | 7.44 (6.65 – 8.38)<br>(n = 6016) | 8.31 (7.44 – 9.08)<br>(n = 3631) |
| ALK (µM)                 | 15 (-16 - 102)<br>(n = 754)        | 29 (11 – 77)<br>(n = 498)        | 797 (388 – 997)<br>(n = 2025)    | 3524 (2608–4196)<br>(n = 1021)   |
| DIC (µM)                 | 263 (45-644)<br>(n = 2060)         | 62 (30-292)<br>(n = 1508)        | 855 (429 – 1227)<br>(n = 5826)   | 3612 (2475 - 4499)<br>(n = 3149) |
| DOC (mgL <sup>-1</sup> ) | 16.36 (8.09 – 28.37)<br>(n = 2055) | 1.90 (1.36 – 2.54)<br>(n = 1492) | 3.38 (2.48 – 4.73)<br>(n = 5782) | 6.12 (4.65 – 8.56)<br>(n = 3129) |
| TP (µgL <sup>-1</sup> )  | 8 (3 - 32)<br>(n = 2107)           | 23 (9-106)<br>(n = 1446)         | 11 (4 - 90)<br>(n = 5840)        | 61 (13 - 463)<br>(n = 4477)      |
| TN (µgL <sup>-1</sup> )  | 193 (105 - 463)<br>(n = 2112)      | 721 (360 – 1889)<br>(n = 1537)   | 311 (166 - 885)<br>(n = 5948)    | 920 (610 – 2540)<br>(n = 1537)   |

<sup>1</sup> – Lakes grouped in LB<sub>ALK</sub> group are: Crystal Bog, Trout Bog

<sup>2</sup> –LC<sub>ALK</sub> group includes Crystal Lake

<sup>3</sup> – Lakes grouped in M<sub>ALK</sub> group are: Allequash Lake, Big Muskellunge Lake, Sparkling Lake, Trout Lake

<sup>4</sup> – Lakes grouped in H<sub>ALK</sub> group are: Fish Lake, Lake Mendota, Lake Monona, Lake Wingra

Accepted Article

**Table 2.** Statistical properties of random errors in carbon system parameters within four alkalinity groups calculated from paired observations. The reported values were rounded according to significant figure convention<sup>1</sup>.

| Lake group                        | Parameter       | Normal distribution parameters |                | Skewness | Kurtosis | t location-scale distribution parameters |                |           |
|-----------------------------------|-----------------|--------------------------------|----------------|----------|----------|--|----------------|-----------|
|                                   |                 | [ $\mu$ ]                      | [ $\sigma_n$ ] |          |          | [ $\mu$ ]                                | [ $\sigma_i$ ] | [ $\nu$ ] |
| <b>LB<sub>ALK</sub></b><br>(n=55) | pH <sup>2</sup> | 0.002                          | 0.018          | 0.021    | 3.2      | 0.002                                    | 0.016          | 10.3      |
|                                   | ALK ( $\mu$ M)  | 0.1                            | 2.4            | 2.0      | 12.9     | 0.2                                      | 1.1            | 2.1       |
|                                   | DIC ( $\mu$ M)  | 1.4                            | 12.3           | 1.8      | 7.7      | -0.9                                     | 5.5            | 1.8       |
| <b>LC<sub>ALK</sub></b><br>(n=48) | pH <sup>2</sup> | 0.002                          | 0.017          | -0.114   | 3.7      | 0.002                                    | 0.013          | 4.3       |
|                                   | ALK ( $\mu$ M)  | 0.3                            | 0.3            | 0.282    | 4.4      | 0.2                                      | 1.1            | 3.3       |
|                                   | DIC ( $\mu$ M)  | 0.9                            | 3.4            | 1.62     | 7.7      | 0.5                                      | 1.3            | 1.6       |
| <b>M<sub>ALK</sub></b> (n=331)    | pH              | 0.003                          | 0.015          | -0.143   | 4.8      | 0.003                                    | 0.009          | 2.6       |
|                                   | ALK ( $\mu$ M)  | 0.6                            | 11.7           | 1.4      | 12.8     | -0.3                                     | 3.6            | 1.4       |
|                                   | DIC ( $\mu$ M)  | -0.3                           | 13.8           | -0.4     | 9.1      | -0.4                                     | 3.9            | 1.2       |
| <b>H<sub>ALK</sub></b><br>(n=150) | pH              | 0.003                          | 0.019          | 0.022    | 3.0      | 0.003                                    | 0.019          | 56.2      |
|                                   | ALK ( $\mu$ M)  | 6.8                            | 56.4           | 0.517    | 13.0     | 4.6                                      | 13.7           | 1.2       |
|                                   | DIC ( $\mu$ M)  | 6.1                            | 83.0           | -0.44    | 13.0     | 4.3                                      | 34.7           | 1.9       |

<sup>1</sup> the significant figures convention reports all certain digits plus the first uncertain digit.

<sup>2</sup> the variable is normally distributed at significance level 0.05 according to Shapiro-Wilk test

Accepted A

**Table 3.** The random parameter error effect on  $p\text{CO}_2$  calculations using three carbonate equilibria in four alkalinity groups shows that  $p\text{CO}_2$  sensitivity to parameter errors depends on the choice of input parameter pairs and lake group. Values reported represent best estimates of the median  $p\text{CO}_2$  at 1 atmosphere with error propagated over 10000 simulations. The random errors of median  $p\text{CO}_2$  are expressed as standard error of the median (SE, standard deviation of 10000 medians) and relative standard error of the median (RSE). Abbreviation n/a means not applicable.

| Equilibrium             | Input Parameter | $\text{LB}_{\text{ALK}}$           |                        |              | $\text{LC}_{\text{ALK}}$           |                        |              | $\text{M}_{\text{ALK}}$            |                        |              | $\text{H}_{\text{ALK}}$            |                        |              |
|-------------------------|-----------------|------------------------------------|------------------------|--------------|------------------------------------|------------------------|--------------|------------------------------------|------------------------|--------------|------------------------------------|------------------------|--------------|
|                         |                 | $p\text{CO}_2$ [ $\mu\text{atm}$ ] | SE [ $\mu\text{atm}$ ] | RSE [%]      | $p\text{CO}_2$ [ $\mu\text{atm}$ ] | SE [ $\mu\text{atm}$ ] | RSE [%]      | $p\text{CO}_2$ [ $\mu\text{atm}$ ] | SE [ $\mu\text{atm}$ ] | RSE [%]      | $p\text{CO}_2$ [ $\mu\text{atm}$ ] | SE [ $\mu\text{atm}$ ] | RSE [%]      |
| $p\text{CO}_2$ -pH-DIC  | Both            | <b>3554</b>                        | <b>±166</b>            | <b>±4.7</b>  | <b>653</b>                         | <b>±36</b>             | <b>±5.5</b>  | <b>1212</b>                        | <b>±45</b>             | <b>±3.7</b>  | <b>795</b>                         | <b>±38</b>             | <b>±4.8</b>  |
|                         | Both adj.       | n/a                                | n/a                    | n/a          | n/a                                | n/a                    | n/a          | n/a                                | n/a                    | n/a          | <b>661<sup>3</sup></b>             | <b>±32</b>             | <b>±4.8</b>  |
|                         | pH              |                                    | ±46                    | ±1.3         |                                    | ±9                     | ±1.4         |                                    | ±34                    | ±2.8         |                                    | ±34                    | ±4.3         |
|                         | pH adj.         |                                    | n/a                    | n/a          |                                    | n/a                    | n/a          |                                    | n/a                    | n/a          |                                    | ±28                    | ±5.1         |
|                         | DIC             |                                    | ±165                   | ±4.6         |                                    | ±35                    | ±5.4         |                                    | ±28                    | ±2.3         |                                    | ±17                    | ±2.1         |
|                         | DIC adj.        |                                    | n/a                    | n/a          |                                    | n/a                    | n/a          |                                    | n/a                    | n/a          |                                    | ±14                    | ±2.1         |
| $p\text{CO}_2$ -pH-ALK  | Both            | <b>15225</b>                       | <b>±1026</b>           | <b>±6.7</b>  | <b>900</b>                         | <b>±69</b>             | <b>±7.7</b>  | <b>1281</b>                        | <b>±50</b>             | <b>±3.9</b>  | <b>866</b>                         | <b>±41</b>             | <b>±4.7</b>  |
|                         | Both adj.       | <b>10999<sup>2</sup></b>           | <b>±935</b>            | <b>±8.5</b>  | n/a                                | n/a                    | n/a          | n/a                                | n/a                    | n/a          | <b>719<sup>3</sup></b>             | <b>±35</b>             | <b>±4.9</b>  |
|                         | pH              |                                    | ±716                   | ±4.7         |                                    | ±36                    | ±4.0         |                                    | ±43                    | ±3.4         |                                    | ±38                    | ±4.4         |
|                         | pH adj.         |                                    | ±694                   | ±6.3         |                                    | n/a                    | n/a          |                                    | n/a                    | n/a          |                                    | ±31                    | ±4.3         |
|                         | ALK             |                                    | ±629                   | ±4.1         |                                    | ±58                    | ±6.4         |                                    | ±25                    | ±2.0         |                                    | ±14                    | ±1.6         |
|                         | ALK adj.        |                                    | ±638                   | ±5.8         |                                    | n/a                    | n/a          |                                    | n/a                    | n/a          |                                    | ±12                    | ±1.7         |
| $p\text{CO}_2$ -ALK-DIC | Both            | <b>2546</b>                        | <b>±198</b>            | <b>±7.8</b>  | <b>580<sup>1</sup></b>             | <b>±59</b>             | <b>±10.2</b> | <b>1437<sup>1</sup></b>            | <b>±230</b>            | <b>±16.0</b> | <b>3725<sup>1</sup></b>            | <b>±1156</b>           | <b>±31.0</b> |
|                         | Both adj.       | <b>2867<sup>2</sup></b>            | <b>±391</b>            | <b>±13.6</b> | n/a                                | n/a                    | n/a          | n/a                                | n/a                    | n/a          | <b>3129<sup>1,3</sup></b>          | <b>±985</b>            | <b>±31.5</b> |
|                         | ALK             |                                    | ±68                    | ±2.7         |                                    | ±25                    | ±4.3         |                                    | ±166                   | ±11.6        |                                    | ±870                   | ±23.4        |
|                         | ALK adj.        |                                    | ±375                   | ±13.1        |                                    | n/a                    | n/a          |                                    | n/a                    | n/a          |                                    | ±866                   | ±27.7        |
|                         | DIC             |                                    | ±198                   | ±7.8         |                                    | ±55                    | ±9.5         |                                    | ±207                   | ±14.4        |                                    | ±1031                  | ±27.7        |
|                         | DIC adj.        |                                    | ±285                   | ±9.9         |                                    | n/a                    | n/a          |                                    | n/a                    | n/a          |                                    | ±1027                  | ±32.8        |

<sup>1</sup> presented  $p\text{CO}_2$  estimates are only for observations when DIC concentrations were larger than ALK concentrations (see Material and Method, and Supporting Information for more details)

<sup>2</sup> calculations after correcting for organic acids contribution to total alkalinity

<sup>3</sup> calculations after adjusting thermodynamic constants for ionic