LIMNOLOGY and OCEANOGRAPHY

ASLO

Limnol. Oceanogr 64, 2019, S22–S33
© 2018 Association for the Sciences of Limnology and Oceanography
doi: 10.1002/lno.11012

Long-term studies and reproducibility: Lessons from whole-lake experiments

M. L. Pace [1*] **S. R. Carpenter** [2] **G. M. Wilkinson** [3]

[1]Department of Environmental Sciences, University of Virginia, Charlottesville, Virginia
[2]Center for Limnology, University of Wisconsin, Madison, Wisconsin
[3]Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, Iowa

Abstract

An important tenet of science is establishing the reproducibility of findings. While long-term studies may seem ill-suited to this goal, here we provide an example of reproducible results from repeated nutrient additions to a lake. We added nitrogen and phosphorus to Peter Lake in 9 yr of a 33-yr study. For seven of these nine additions, phytoplankton biomass, as measured by seasonal mean chlorophyll *a*, increased in proportion to the rate of nutrient loading. Additionally, for these seven additions, similar nutrient loading rates resulted in mean Chl *a* concentrations within a roughly twofold range—an outcome within expectation given uncontrolled sources of variation in a whole-lake manipulation. However, for two of the nine nutrient additions, Chl *a* concentrations were well below expected concentrations. The low chlorophyll responses co-occurred with years having the highest water color (absorbance of light at 440 nm). The number of years of nutrient additions was too limited to strongly test the influence of color at the scale of seasonal mean values. We, therefore, tested for the effect of phosphorus load and color, on Chl *a* using time series models of weekly data. At the weekly scale, there was a strong negative effect of color on chlorophyll concentration. Overall, the repeated nutrient additions provided a confirmation of existing models at the whole-lake scale and demonstrated an interesting exception to these models. Including repeated manipulations as part of long-term studies is an important way to test generalizations and to identify unexpected outcomes that raise new questions.

A hallmark of science is reproducibility. Methods must be repeatable and results reproducible by others. However, the results of surveys of scientists on reproducibility as well as attempts to reproduce experiments suggest a significant fraction of contemporary research may not be reproducible because of failings in standards and practices resulting in suspect findings and uncertain application (Open Science Collaboration 2015; Baker 2016). This "reproducibility crisis" has led to calls for: better description of methods, full documentation of computer programs, free sharing of data, and cooperative distribution of research materials (McNutt et al. 2016). Other researchers have expressed needs to: improve experimental procedures, replicate study findings, and avoid reliance on *p* value thresholds for assessing significance (Begley and Ioannidis 2015; Amrhein et al. 2017; Nosek and Errington 2017; McShane et al. 2018). For the discipline of ecology, Houlahan et al. (2017) argued that the field should focus on evaluating reproducibility in the context of testing model predictions against independent data.

How do long-term studies of ecosystems address the problem of reproducibility? In terms of repeatability of methods and data sharing, many long-term studies make data available online, as for example, the Long Term Ecological Studies program of the National Science Foundation (Peters et al. 2014). This requirement includes documenting methods so that data are interpretable. Hence, many long-term ecosystem studies meet the criteria of repeatability of methods and sharing of data. A second criterion of reproducing findings is more difficult. Long-term studies rely on observations in time and space, field experiments, models, and large-scale manipulations to interpret ecosystem change. Long-term studies, like much of ecological research, use a weight of evidence approach that typically involves models and comparisons of models to data (Hilborn and Mangel 1997; Anderson et al. 2000). These approaches contrast with tightly controlled and replicated experiments common to some areas of ecology and in other sciences. Do long-term studies necessarily fail the second requirement of reproducibility that study findings be repeatable? Put differently, are findings from long-term studies singular, meaning do they represent a particular pattern and

*Correspondence: pacem@virginia.edu

Special Issue: Long-term Perspectives in Aquatic Research
Edited by: Stephanie Hampton, Matthew Church, John Melack and Mark Scheuerell.

plausible explanation for a specific system only over the timescale of the study? More importantly, how can generality emerge from long-term studies? These are critical questions especially in application of the results of long-term studies to management and policy. We propose that reproducibility of results is addressable in long-term research in particular through ongoing or repeated large-scale manipulations. Repeatable ecosystem manipulations provide important tests at critical scales (e.g., an entire lake over a season). However, ecosystem manipulations can cause changes that permanently alter conditions (e.g., regime shifts) or create long recovery times. Hence, the application of large-scale manipulations may not be reproducible if an ecosystem has changed or failed to recover. In addition, manipulations may not be possible where ecosystems are protected or provide a critical ecosystem service.

Here, we explore reproducibility using results from a series of whole lake, nutrient additions in concert with long-term observations (1984–2016) of Peter Lake (described below). Other whole lake, nutrient manipulation studies find large increases in phytoplankton biomass with inputs of inorganic nitrogen and phosphorus (Lewis et al. 2011) and general phosphorus–chlorophyll relationships are known based on among-lake data (Dillon and Rigler 1974; Filstrup et al. 2014). While comparative studies provide important frameworks and the potential to evaluate specific models (Downing et al. 2001; Hamilton et al. 2016), ecosystem manipulations are also critical tests for addressing questions and evaluating results from long-term studies (Carpenter 1998).

Here, we consider nutrient additions in multiple years at the scale of the same lake. We focus on the question: Did Peter Lake respond consistently to nutrient loading? We provide evidence that nutrient additions in different years were independent and thereby repeated manipulations. In addition to nutrient loading, our prior whole-lake manipulations indicated that grazing especially by large cladocerans is important as is the concentration of chromophoric dissolved organic matter (cDOM) in determining primary production and phytoplankton biomass (Carpenter et al. 1998; Carpenter et al. 2001). We consider these variables and other environmental conditions in assessing consistency of response. We find the answer to the question about consistent response to nutrients is both "yes" and "no" and each of those answers contributes to learning. We argue for greater focus on the issue of reproducibility in long-term studies through, where possible, sustained and repeated manipulations of ecosystems.

Methods

Study site

Long-term measurements and nutrient additions were carried out in Peter Lake located at the University of Notre Dame Environmental Research Center in Gogebic County, Michigan, USA (46°11′N, 89°30′W). The lake is an oligotrophic to mesotrophic, softwater lake surrounded by temperate forest and fringing bog. The lake is small (area = 2.7 ha) and deep for its area (mean depth = 5.7 m) with circumneutral pH. Beginning in the 1980s and continuing through 2015, whole-lake manipulations of Peter Lake altered fish communities, nutrients, and/or carbon stable isotope ratios. Here, we focus on nutrient manipulations. Summaries of the other manipulations are available in Carpenter and Kitchell (1993), Carpenter et al. (2005), and Carpenter et al. (2011).

Nutrient additions

We added nutrients to the lake in the years 1993–1997, 2002, and 2013–2015. Phosphoric acid (H_3PO_4) and ammonium nitrate (NH_4NO_3) were mixed with lake water to create stock concentrations that were then added to the lake to achieve a targeted load. Daily P-loads varied from 1.2 to 6.1 mg m^{-2} d^{-1} and N : P molar ratios were either > 25 to promote P limitation (relative to N) or near the Redfield ratio to favor cyanobacteria (Table 1). In most years, we made daily additions by pumping the nutrient mixture into the propeller wash of a moving boat. In 3 yr (1995–1997) nutrients were

Table 1. Summary of experimental loading amounts and methods for each year of nutrient addition. Natural baseline loads were 0.3 mg P m^{-2} d^{-1}.

Year	P-load (mg P m^{-2} d^{-1})	N-load (mg N m^{-2} d^{-1})	Cumulative added P (mg P m^{-2})	Cumulative added N (mg N m^{-2})	N : P (molar)	Addition method	Addition frequency	Duration (d)
1993	3.1	23.9	223	1719	17	Constant	Daily	71
1994	2.1	28.9	164	2256	30	Constant	Daily	77
1995	1.2	16.2	109	1474	30	Constant	Weekly	84
1996	6.1	95.1	555	8650	34	Constant	Weekly	84
1997	3.2	45.7	291	4155	34	Constant	Weekly	83
2002	3.4	37.8	283	3178	25	Pulse+constant	Daily	83
2013	2.6	9.9	219	832	15	Ramp	Daily	84
2014	3.0	20.3	267	1810	15	Constant	Daily	87
2015	3.0	20.3	87	590	15	Constant	Daily	28

added weekly. We also varied the loads in 2 yr either with an initial pulse of nutrients to promote phytoplankton growth (2002) or with a weekly stepwise increase in the loading rate (2013). In all other cases, the loading rates within a year were constant (Table 1). Nutrient additions covered the summer period from on or near 01 June to late August except in 2015 when we halted the nutrient addition on 30 June.

Natural, baseline P-loading was estimated using a Vollenweider equation (Carpenter et al. 2001), where P-input is a function of total phosphorus (TP) concentrations, hydraulic load, and mean depth. We measured TP, hydraulic load (Cole and Pace 1998), and mean depth. Estimated baseline loads were 0.3 mg m^{-2} d^{-1}. This estimate was corroborated by independent estimates of P-inputs based on dated sediment cores and phosphorus budgets (Houser et al. 2000). We assigned the value 0.3 mg m^{-2} d^{-1} to the three non-nutrient addition years that preceded each nutrient addition period (*see* Data analysis).

Overall, the nutrient additions produced both a range of loading rates and repetition of a similar loading rate (Table 1). Responses to nutrient loading were determined by measuring chlorophyll *a*—an index of algal biomass.

Chlorophyll analysis

We measured chlorophyll in years with and without nutrient additions over the period 1984–2016. Water samples were taken weekly from mid-May to late August or early September (typically n = 15–17) except in 2000 (n = 4) and 2004 (n = 7) where sampling was less frequent. Water samples were taken at previously determined light depths of 100%, 50%, 25%, 10%, 5%, and 1% of surface irradiance. Water samples were filtered (GF/F filters), and filters were frozen. The filters were later extracted for 24 h at 4°C in 25 mL of 100% methanol. Chlorophyll concentrations with corrections for phaeopigments were determined on the methanol extracts by fluorometry (Marker et al. 1980).

Additional variables

Total phosphorus was measured after persulfate digestion using an autoanalyzer. We made at least weekly TP measurements in all nutrient addition years. The first samples of the year were taken between 15 May and 30 May (before any nutrient additions) and used as a measure of spring TP concentration. TP was measured in all nutrient addition years and in 15 yr when nutrients were not added. Water color (light absorbance at 440 nm) and dissolved organic carbon (DOC) were measured weekly for epilimnetic water samples passed through GF/F filters. Color samples were stored at 4°C and analyzed within 7 d. We measured light absorbance compared to a reference (distilled water) using a spectrophotometer with a 10 cm cuvette (Cuthbert and del Giorgio 1992). Water color was measured in all years from 1996 to 2016 except 2000. DOC samples were acidified to a pH of 2 with 2 N H_2SO_4. The non-purgeable DOC (Findlay et al. 2010) was measured with a Shimadzu high-temperature organic carbon analyzer except prior to 1993 when a different instrument was used. The two instruments gave similar results (Pace and Cole 2000). DOC was measured in all years after 1987. Zooplankton length was determined using standardized net haul samples (80 μm mesh) from which abundance and lengths of crustacean taxa were measured. Length–weight relationships were used to convert abundance to biomass. Lengths were measured on the most abundant taxa and indices of mean lengths calculated from mean taxa lengths weighted by biomass. Zooplankton lengths were measured from 1984 to 2016 except for the years 2000, 2004–2006.

Data analysis

Chlorophyll concentrations were determined for the surface layer by averaging the upper three light depths (100%, 50%, and 25% surface light) to derive a volumetric value for the epilimnion. Chlorophyll concentrations were also integrated to the 10% light depth to derive areal values (i.e., per m^{-2}). Epilimnetic chlorophyll is of interest because it is a measure of surface-water quality. Assessments of eutrophication and management actions are often based on this value. Areal chlorophyll is indicative of planktonic primary production and autotrophic organic matter supporting the food web (Carpenter et al. 1999). The response of epilimnetic and areal chlorophyll to nutrient additions for a season were analyzed by averaging the respective chlorophyll values for the period from the third week of sampling (after nutrient additions had started) to the end of the season except in 2015 when values were averaged to the end of the addition.

Statistical analysis

To test reproducibility of the nutrient loading responses, we calculated linear regressions between the daily P-loading rate and the two measures of seasonally averaged chlorophyll (i.e., epilimnetic and areal). We also considered the cumulative-P added over the season as a predictor mainly to account for the shorter loading period of the 2015 addition (Table 1). In addition, we evaluated possible covariates including cladoceran length, DOC, and water color. Because the number of years was limited as further described below, we developed a time series model to test the significance of P-loading and water color using weekly data.

The data for the regression analysis consisted of the years where nutrients were added (1993–1997, 2002, and 2013–2015) and 3 yr where nutrients were not added immediately prior to each fertilization period (1992, 2001, and 2012). There were two reasons for using only three non-nutrient addition years instead of all the years of background loading. First, we only directly estimated baseline P-loading for a few years (Houser et al. 2000; Carpenter et al. 2001). Second, using all non-nutrient addition years would provide little additional information because P-loading and chlorophyll were similar low values. Hence, using the three non-nutrient addition years, our regression analysis consisted of 12 yr. We tested this data for an expected pattern between

nutrient loading and chlorophyll and then further analyzed model residuals against two additional variables—cladoceran length and color. As noted above, cladoceran grazing (represented by length) and water color, which impacts light and possibly nutrient availability, may diminish chlorophyll response to nutrient loading.

We previously established that Peter Lake is primarily phosphorus limited based on N : P ratios and on the accumulation of dissolved inorganic N but not dissolved inorganic P at high loading rates (Carpenter et al. 2001). Nitrogen (N) loading is a predictor of chlorophyll but regressions models using N had slightly lower R^2 values. Hence, we do not consider N-loading effects further in the current study.

The analysis of covariates based on seasonal means is limited because of the low number of years particularly for color where fewer years of measurement were available. To extend the analysis and specifically test the potential negative effect of color on chlorophyll, we analyzed the weekly data thereby gaining greater power. We fit time series regressions to weekly Peter Lake data from field seasons of 1996–1998, 2001–2002, and 2007–2016. In other years, color data or other weekly values were not available. This dataset included 208 observations of all variates of interest: epilimnetic and areal chlorophyll, color, and P-loading rate, where each set of observations was separated by 1 week. Data were fit to the model:

$$y(t+1) = b_0 + \varphi y(t) + b_1 x_1(t) + b_2 x_2(t) + + \varepsilon(t+1) \quad (1)$$

where y is the response variate (epilimnetic or areal chlorophyll, mg m^{-3} or mg mg^{-2}, respectively), the b_i are regression coefficients, φ is the autoregression coefficient (lag 1), x_1 is color (m^{-1}), x_2 is P-loading rate (mg m^{-2} d^{-1}), and ε are residuals assumed to be independently and identically distributed normal with mean 0 and standard deviation estimated from the data. We also tested a similar model that included an additional variable, cladoceran length, as a time series adding the term $b_3 x_3(t)$ to Eq. 1. Prior to fitting the models, chlorophyll variates were log-transformed to obtain more nearly normally distributed residuals, and all variates were centered.

We fitted the models using the lm() function of R 3.3.2 (R Core Team 2016). Equation 1 provided a better fit to the data than three alternative models. The alternatives were: (1) replace each $x_i(t)$ with its average on the interval from week t to week $t + 1$; (2) replace each $x_i(t)$ with its first difference $x_i(t + 1) - x_i(t)$; and (3) remove the autoregressive term and use $y(t + 1) - y(t)$ as the dependent variable. For models of epilimnetic and areal chlorophyll based on Eq. 1, we confirmed that the autocorrelation and partial autocorrelation functions have no remaining autocorrelation in chlorophyll and that quantile–quantile plots of the residuals were approximately normal. We also fitted Eq. 1 using maximum likelihood and obtained similar parameter values.

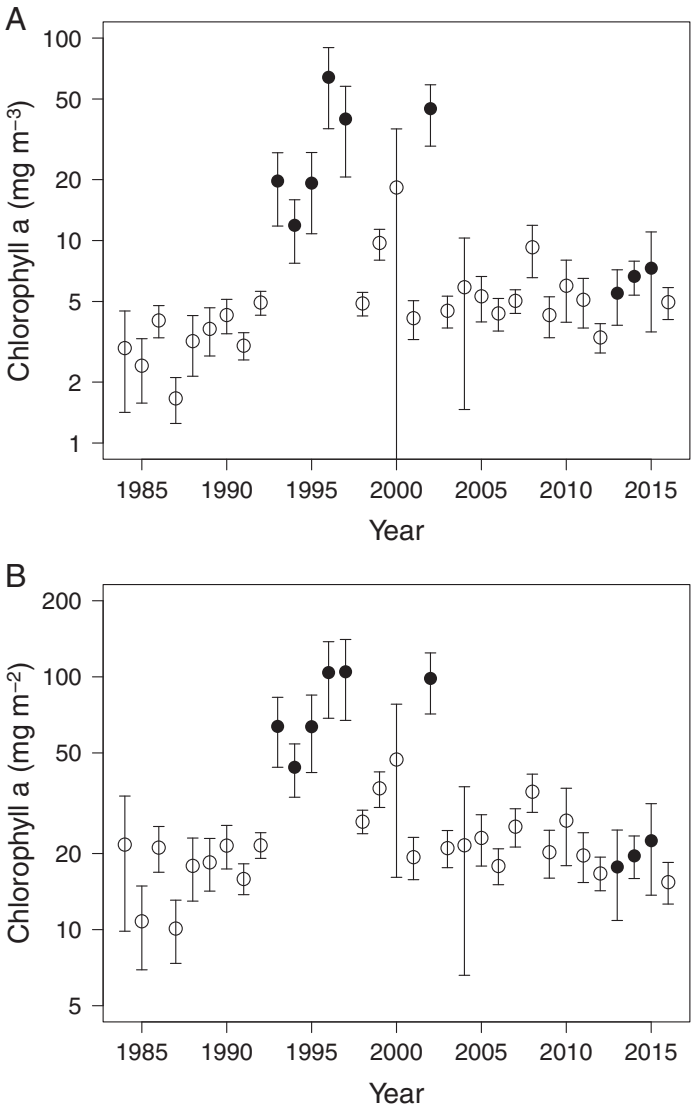

Fig. 1. Peter Lake (**A**) epilimnetic and (**B**) areal chlorophyll from 1984 to 2016. Values are seasonal means with 95% confidence intervals. Open circles are years without nutrient additions; closed circles are years with nutrient additions.

Results

Mean Peter Lake epilimnetic chlorophyll concentrations mainly ranged from 1.7 to 9.7 mg m^{-3} among years without nutrient addition from 1984 to 2016 (Fig. 1A). Associated 95% confidence intervals were typically in the range of ± 0.5–-2 mg m^{-3}. Exceptions to these ranges for means and confidence intervals were 2000 when the mean and variability were high and 2004 when variability was high. In both of these years sample number was low (*see* Methods) and a few high values contributed to the uncertainty. Overall, nominal chlorophyll concentrations indicated an oligo-mesotrophic status for this lake. With nutrient additions, chlorophyll means were > 10 mg m^{-3} (Fig. 1A) representing eutrophic

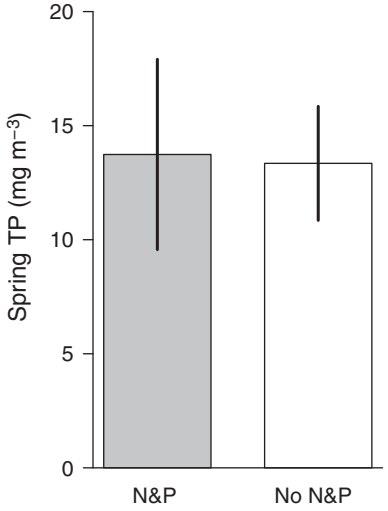

Fig. 2. Spring total phosphorus (TP) for years after nutrients were added (N&P, $n = 15$) and for years after no nutrient additions (no N&P, $n = 9$). Error bars are 95% confidence intervals.

conditions except for the most recent fertilization period (2013–2015, discussed below). Areal chlorophyll concentrations ranged from 20 to 155 mg m^{-2} and followed a pattern similar to epilimnetic chlorophyll (Fig. 1B).

Initial spring TP concentrations following a year of nutrient addition ($n = 9$) had a mean value (± 95% CI) of 13.7 ± 4.2 mg m^{-3}, while mean spring TP following years ($n = 15$) without nutrient addition was 13.3 ± 2.5 mg m^{-3} (Fig. 2). The similar means and broadly overlapping confidence intervals indicate there was no evidence for carry-over of nutrients. Therefore, we treated years as independent in the context of nutrient loading and chlorophyll response.

Mean chlorophyll concentrations increased with nutrient loading as expected (Fig. 3). The baseline nutrient load years (1992, 2001, 2012) had low and overlapping values of chlorophyll (Fig. 3, open circles). Chlorophyll concentrations increased in relation to P-loading for the nutrient addition years (Fig. 3, closed circles); however, two nutrient addition years, 2013 and 2014 had low-chlorophyll concentrations relative to P-loading (Fig. 3A). Excluding those points and rerunning the regression improved the model fit (Fig. 3B) with R^2 increasing from 0.64 to 0.82 (Table 2). For the year 2015 nutrient loading was limited to 28 d instead of the whole season for purposes of the manipulation conducted that year (Pace et al. 2017). The mean seasonal epilimnetic chlorophyll was relatively low (15.6 mg m^{-3}) given the loading rates of 3.0 mg P m^{-2} d^{-1}. Using cumulative-P addition as a predictor, mean epilimnetic, and areal chlorophyll for 2015 were close to the regression lines (Fig. 4) while values for 2013 and 2014 remained low relative to other years (Fig. 4). Using cumulative loading as a predictor is somewhat problematic because the mean chlorophyll values are temporal averages not a cumulative measure of biomass or production. Nonetheless, mean chlorophyll for the shorter loading period of 2015 was better represented relative to other years using the cumulative-P. Overall, chlorophyll responded as expected relative to P-loading in most years including 2015. However, the years 2013 and 2014 were exceptional with an unexpected low response to nutrient inputs.

Why was mean chlorophyll so low in 2013 and 2014 despite nutrient additions? Cladoceran body size was elevated and greater than values observed in recent years but lower than observed at the beginning of the study (Fig. 5A). DOC in 2013 and 2014 was slightly elevated relative to long-term data

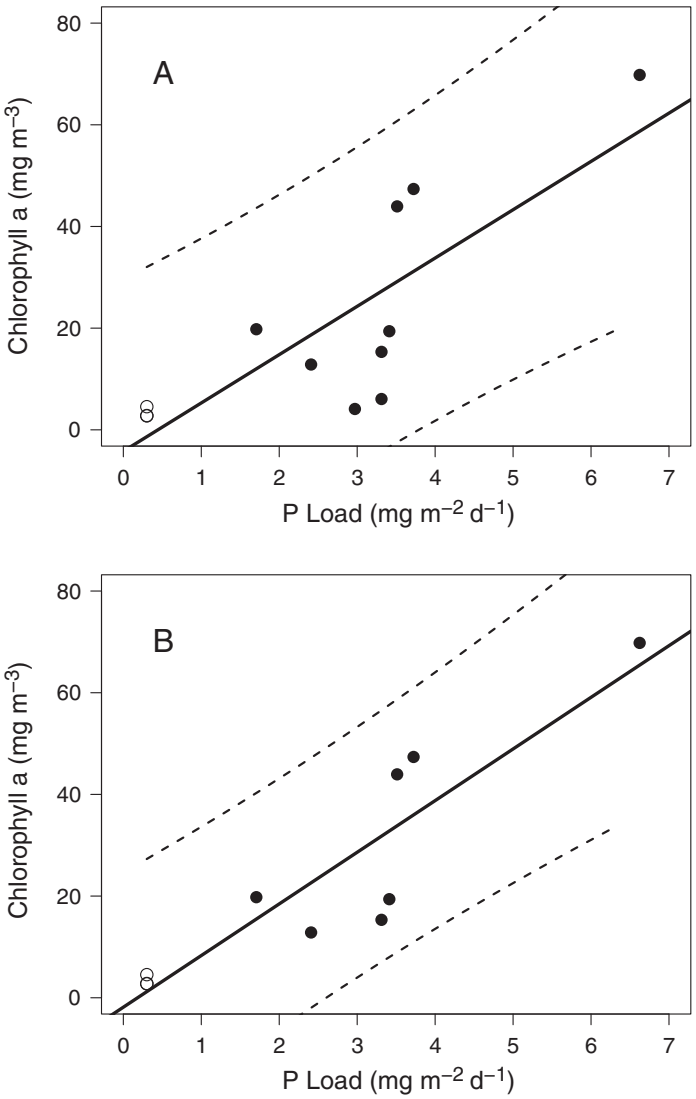

Fig. 3. Daily phosphorus load (P-load) vs. seasonal mean epilimnetic chlorophyll for: (**A**) all years including nutrient addition years and non-nutrient addition years and (**B**) with the nutrient addition years 2013 and 2014 excluded. . Closed circles are nutrient addition years (1993–1997, 2002, 2013–2015) and open circles are years prior to the initiation of nutrient additions (1992, 2001, 2012, note two points overlap). Solid line is the regression and dashed lines are 95% prediction intervals.

Table 2. Regression statistics for models of two phosphorus loading metrics and seasonal mean chlorophyll *a* (mg m^{-3}). For the first two rows of the table, daily P-loading (mg m^{-2} d^{-1}) is the independent variable and the dependent variable is either chlorophyll for all years (n = 12) or excluding the years 2013 and 2014 (n = 10). Similarly, for the third and fourth rows the total cumulative phosphorus load (mg P m^{-2}) is the independent variable and the dependent variable is either chlorophyll for all years (n = 12) or excluding the years 2013 and 2014 (n = 10). SE is the standard error for the intercept and slope. p represents the probability the intercept or slope is not different from 0. RSE is the residual standard error of the regression model. R^2 represents the proportion of variance explained.

P-loading metric	Chlorophyll years	Intercept±SE	*p*-Intercept	Slope±SE	*p*-Slope	RSE	R^2
Daily	All years	–4.2±7.1	0.56	9.5±2.2	0.002	13.5	0.64
Daily	Exclude 2013 and 2014	–1.9±5.4	0.74	10.2±1.7	<0.001	10.2	0.82
Cumulative	All years	–1.7±6.3	0.79	0.11±0.02	0.001	13.1	0.67
Cumulative	Exclude 2013 and 2014	19.5±10.3	0.09	0.21±0.04	<0.001	20.6	0.77

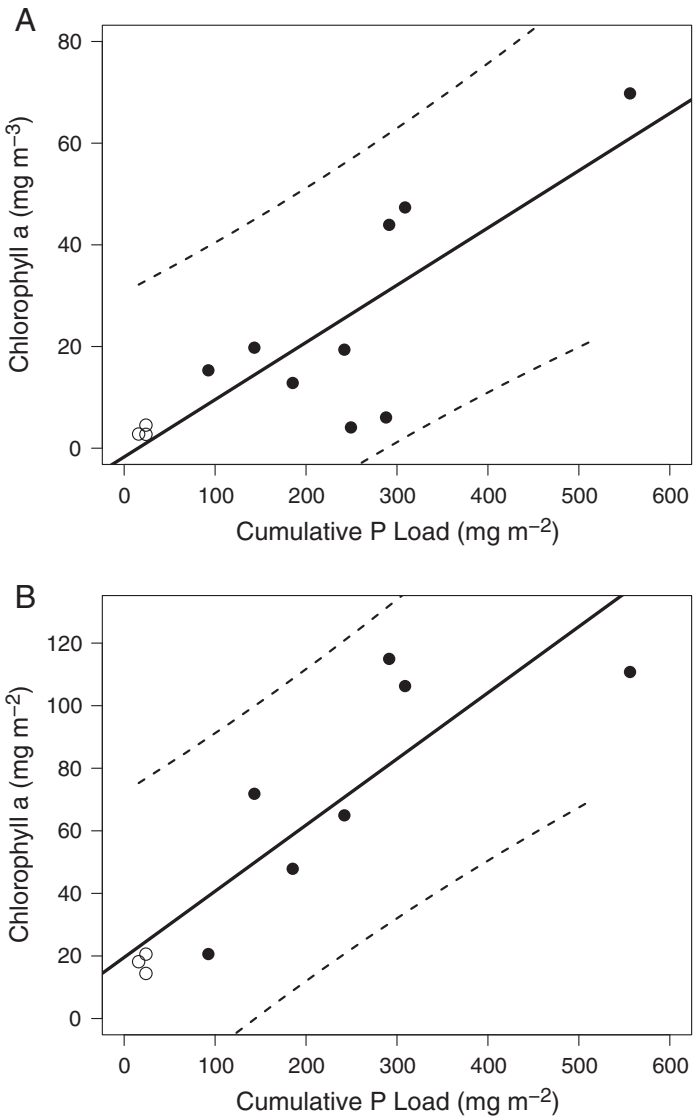

Fig. 4. Cumulative phosphorous load (P-load) vs. seasonal mean epilimnetic chlorophyll for: (**A**) all years including nutrient addition years and non-nutrient addition years and (**B**) with the nutrient addition years 2013 and 2014 excluded. Symbols and lines are as in Fig. 3.

(Fig. 5B), but it was also within the range observed for other years considered in this analysis. However, color was much higher in 2013 and 2014 relative to all prior years with means exceeding 2 m^{-1}(Fig. 5C). Temperature, stratification, and other physical and climatic conditions were not unusual in 2013 and 2014 (Wilkinson et al. 2018). Phytoplankton communities changed in response to the nutrient additions typically, but not always, with dominance by cyanobacteria (Cottingham 1999; Wilkinson et al. 2018). These patterns for phytoplankton were also observed in 2013 and 2014 (Wilkinson et al. 2018).

We examined the possible effects of grazer size and color on the chlorophyll residuals from the regression lines presented in Fig. 3. These values were plotted against mean cladoceran length and color for both epilimnetic and areal chlorophyll (Fig. 6). Cladocerans were relatively large in 2013 and 2014, but there were also some relatively large negative residuals associated with years when cladoceran size was moderate. Regressions for residuals vs. cladoceran length had probabilities > 0.1. Large negative residuals were also associated with water color for the years 2013 and 2014 (points lower right of Fig. 6B,D) and in one case a regression between residuals and water color had a marginally significant probability (epilimnetic p = 0.09). Overall, the residual analysis was limited by the small number of years especially in the case of water color (n = 8).

The weekly time series models tested the specific hypothesis that color had a negative relationship with chlorophyll dynamics while accounting for the expected strongly positive effect of P-loading. A time series model that included cladoceran length did not improve the fit and results for this model are not presented. The models of epilimnetic and areal chlorophyll were similar in several respects: autoregressive terms and effects of color and P-load were substantially larger than their standard errors (Table 3). Both models fit the data rather closely as indicated by the high R^2 and low residual standard errors (Table 3). In both models, P-load had a positive effect and color had a negative effect consistent with the hypothesis.

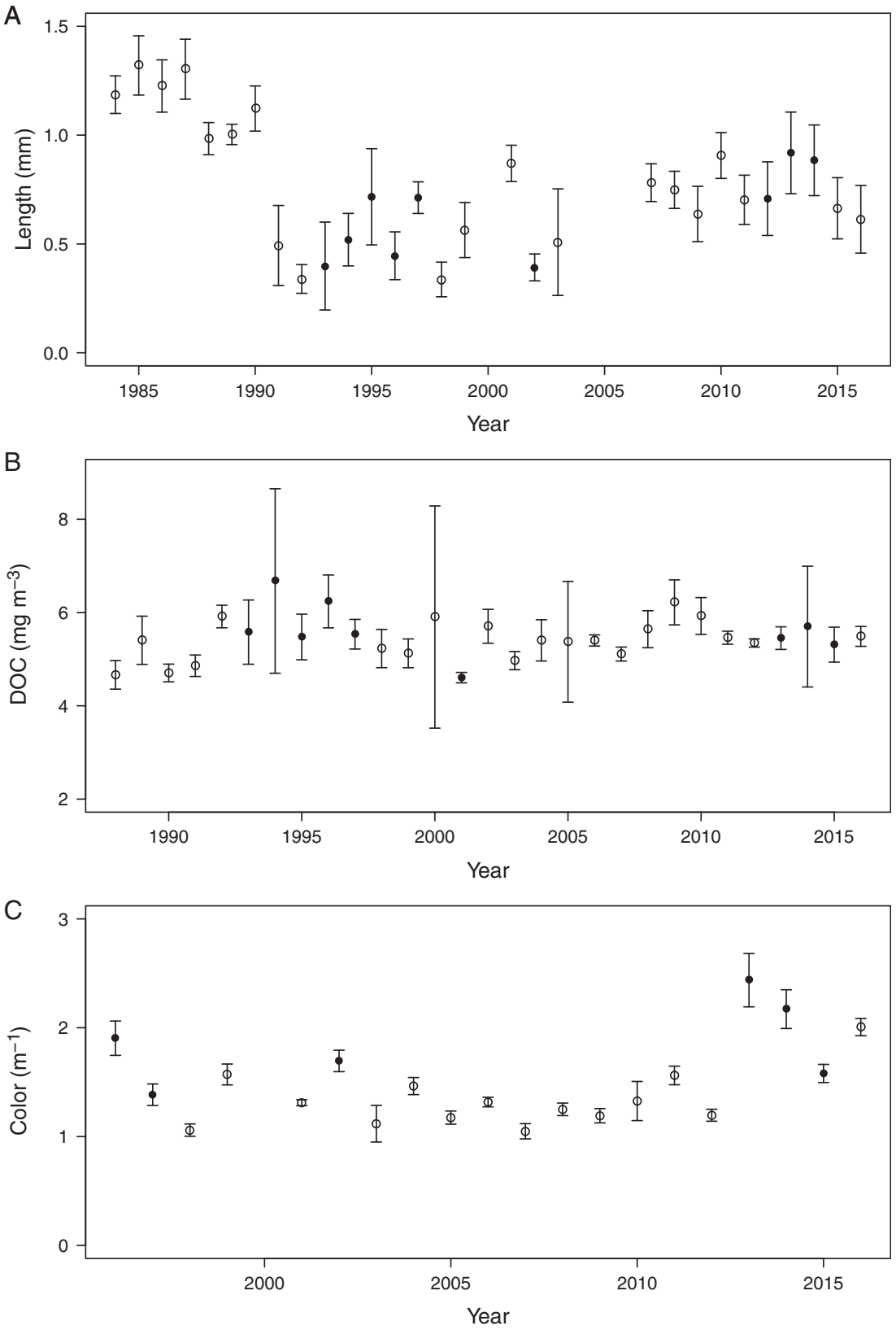

Fig. 5. Annual means and 95% confidence intervals for (**A**) cladoceran length, (**B**) dissolved organic carbon (DOC), and (**C**) water color (absorbance at 440 nm) in Peter Lake. Open circles are years without nutrients; closed circles are year with nutrient additions.

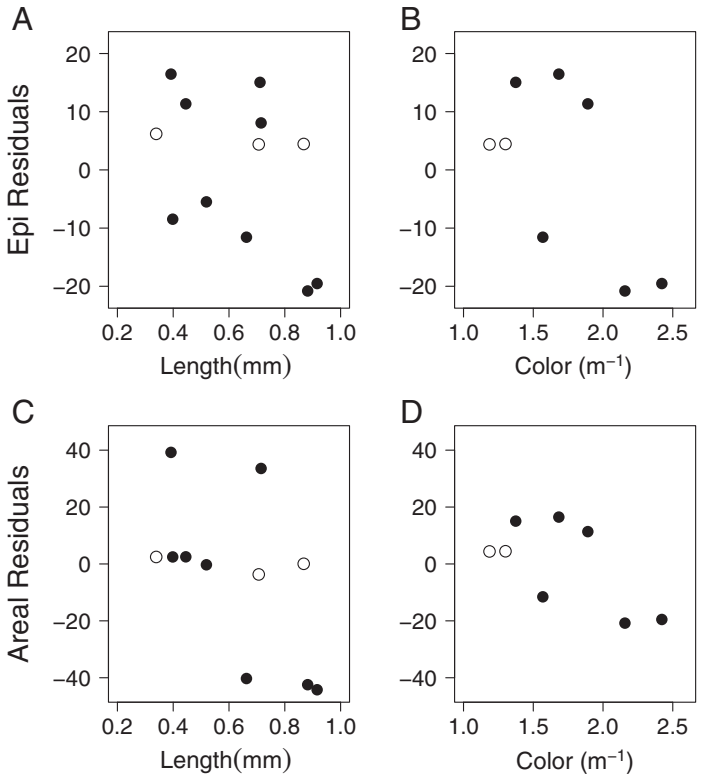

Fig. 6. Residuals for the P-loading regressions plotted vs. mean cladoceran length and color for the relationships presented in Fig. 3. (**A**) and (**B**) are for the eplimnetic chlorophyll relationship ("Epi Residuals"); (**C**) and (**D**) are for the areal chlorophyll relationship ("Areal Residuals"). Closed circles are nutrient addition years; open circles are non-nutrient addition years.

Table 3. Parameters, standard errors (SE), probabilities (*p*), and model statistics for least square fits of models for epilimnetic and areal chlorophyll. Autoregression (lag 1), color, and P-load are the predictor variables as defined in Eq. 1. RSE is residual standard error. Model df represents model degrees of freedom. Models were based on 208 time points.

Response variable	Coefficient/ statistic	Estimate	SE	*p*
Epilimnetic chlorophyll	Intercept	0.00362	0.0308	0.907
	Autoregression	0.795	0.0432	<0.0001
	Color	–0.164	0.0734	0.0264
	P-load	0.0768	0.0168	<0.0001
	RSE	0.428	—	—
	Model df	189	—	—
	Model P	<0.0001	—	—
	Adjusted R^2	0.753	—	—
Areal chlorophyll	Intercept	–0.00718	0.0290	0.805
	Autoregression	0.767	0.0455	<0.0001
	Color	–0.248	0.0722	7.49×10^{-4}
	P-load	0.0677	0.0155	<0.0001
	RSE	0.421	—	—
	Model df	189	—	—
	Model P	<0.0001	—	—
	Adjusted R^2	0.714	—	—

Discussion

For most years, nutrient loads led to a predictable response of Chl *a* consistent with general phosphorus loading models (Vollenweider 1968; Reckhow and Chapra 1983). In years with similar loading rates of 3.4 (1993), 3.5 (1997), and 3.7 (2001) mg P m^{-2} d^{-1}, mean epilimnetic chlorophyll concentrations were 20, 44, and 47 mg m^{-3}, respectively. This slightly greater than twofold variation is within the range of expectation given the inter-annual and seasonal differences in environmental and food web conditions. Hence, the nutrient additions gave results qualitatively consistent with prior models of nutrient loading that are based on log-transformed data and have a wide spread at a specific loading rate. The Peter Lake results provide an example of reproducibility of findings using large-scale manipulations and represent an important form of confirmation.

Formal comparison of our results with established models of nutrient loading are difficult because those models are based on annual loading (Reckow and Chapra 1983) while we only measured daily loading over seasonal time scales. However, based on earlier comparisons, phosphorus loading vs. chlorophyll for Peter Lake falls within the range of nonexperimental, comparative lake data (Carpenter 2002). There is considerably more variance in the nonexperimental lakes data even on log scales as these lakes include a wide variety of physical, chemical, and biological conditions.

Chlorophyll concentrations in relation to nutrient additions in 2013 and 2014 were well below expectations. In 2013, we started nutrient additions at a low rate and subsequently increased the rate to provide sufficient daily observations for early warning statistical analysis of blooms (Wilkinson et al. 2018). However, a bloom never materialized and chlorophyll concentrations were only slightly higher than a nearby, unfertilized lake. In 2014, we observed surface-scums of cyanobacteria during mid-summer, and chlorophyll concentrations were consistently elevated relative to the reference lake (Wilkinson et al. 2018). Nevertheless, phytoplankton biomass was low in both years given the nutrient loading and prior responses of Peter Lake to similar loads.

Color and P-load had significant effects on chlorophyll responses based on model fits for weekly data. Using the weekly model at average P-loading and color, a one-unit increase in color would cause declines of 55% for epilimnetic and 65% for areal chlorophyll, respectively. We cannot compare these estimates directly to the regression models based on seasonal means (Table 2) because of differences in the scale and structure of the models as well as the lack of a term for

color in the seasonal means model. Nonetheless, the analyses indicate a strong effect of color on chlorophyll in the years when color was high.

Cladoceran length did not have a significant negative effect in trial analyses of the weekly models (results not presented). However, potential effects of grazing should not be rejected on the basis of these results. Relative to the length of the long-term study, models presented in this article use a limited amount of data because color was not measured until 1996. Therefore, some of the strongest food web manipulations of Peter Lake (e.g., note the large changes in cladoceran mean length around 1990, Fig. 5) could not be included in this analysis (Carpenter and Kitchell 1993). Other time-series regressions for Peter and neighboring lakes have shown strong and significant effects of cladoceran or crustacean length on areal and volumetric chlorophyll (Carpenter et al. 1996, 1998, 2001). Other papers have documented significant effects on chlorophyll of grazer body size in comparative analyses (Pace 1984, Carpenter et al. 1991; Sarnelle 1992; Jeppesen et al. 2007).

Water color reflects variation of light-absorbing cDOM (Molot 2009) with higher color reducing light transmission. cDOM is the primary determinant of light absorption in Peter Lake and several other lakes even when lakes are fertilized (Carpenter et al. 1998). The years 2013 and 2014 had unusually high color (> 2 m^{-1}) relative to other years while DOC concentrations were not unusual (Fig. 5). We do not know the explanation for the higher color values but they may have been associated with late ice-off, high DOC inputs due to spring and summer precipitation events (Zwart et al. 2017), and possibly higher concentrations of iron (Dillon and Molot 2005; Kritzberg and Ekstrom 2012; Weyhenmeyer et al. 2017). We speculate that a combination of these factors lead to sustained high cDOM concentrations in the 2 yr.

High concentrations of cDOM may also interact with nutrients and limit or control their availability for phytoplankton growth (Jones 1992; Williamson et al. 1999; Thrane et al. 2014; Seekell et al. 2015). While light absorption and release of nutrients from DOM has been demonstrated experimentally (Vähätalo et al. 2003), the net effect of such changes on ecosystem primary production is not known. However, a lake manipulation that increased terrestrial-DOC inputs measured higher gross primary production probably due to higher nutrient concentrations (Zwart et al. 2016; Corman et al. 2018). Analyses of phosphorus–chlorophyll patterns indicate spatial variation in water color effects where in some lakes the effect is positive, others negative, and many insignificant (Fergus et al. 2016). This result partially contrasts with positive impacts of color on phosphorus–chlorophyll relationships observed by Nürnberg and Shaw (1998) for an among-lake analysis that did not account for spatial variation. The temporal variation of color in Peter Lake is informative in revealing a strong negative impact on chlorophyll. Our results for Peter Lake suggest light limitation, nutrient–DOM interactions, or both effects limit phytoplankton response to nutrient inputs. Our nutrient additions along with prior comparative lake studies (Nurnberg and Shaw 1998; Fergus et al. 2016) suggest the need for further study of the interactions of DOM, light and nutrients on phytoplankton.

The manipulations of Peter Lake provided a series of independent years to test the reproducibility of phytoplankton biomass increases in relation to nutrient additions. Note our manipulations were not replicates because of some variations in the nutrient additions (Table 1) and many sources of uncontrolled variation (e.g., weather, food web dynamics). However, these repeated tests at a whole-lake scale, in part, confirm a generality developed from theory, bioassay experiments, comparative studies, long-term studies, and prior whole-lake manipulations (Hutchinson 1957; Vollenweider 1968; Edmondson 1970; Elser et al. 1990; Lewis et al. 2011). While the confirmation is not surprising, reproducibility is important scientifically and for management applications. The capacity to conduct repeated and continuous large-scale manipulations is an important feature that contributes to the unique value of long-term studies for testing theory and establishing generality.

Another value of repeated manipulations is the insight that comes from unexpected results. The limited chlorophyll response of Peter Lake in 2013 and 2014 was a surprise. The high color of the lake during these 2 yr points to hypotheses of light and/or nutrient effects that need additional study. These hypotheses take on additional importance because lakes in some regions are "browning"—that is, have increasing concentrations of cDOM (Monteith et al. 2007; Kritzberg 2017). Our results suggest that brown lakes are more resilient to nutrient inputs and thus may resist eutrophication.

The main advantages of long-term research are establishing trends, observing surprises, and assessing data in relation to models (Carpenter 1998; Lindenmayer et al. 2012). In addition, long-term studies often include sustained experiments allowing observations of changing ecosystem responses (Schindler et al. 2008; Knapp et al. 2012; Melillo et al. 2017; Reich et al. 2018). Natural experiments may also occur during long-term studies allowing evaluation of ecosystem response before and after an event (Strayer et al. 2014). Long-term studies also contribute to reproducibility through documentation of methods and data as well as creating sample archives for retrospective analyses. The example presented in this article indicates another underappreciated aspect of long-term studies—repeated manipulations. Repeating studies in time and space is feasible and contributes to confirmation of general models as well as to identifying exceptions. Long-term studies should consider the prospect and potential benefits of including repeated manipulations as ways of testing reproducibility.

References

Amrhein, V., F. Korner-Nievergelt, and T. Roth. 2017. The earth is flat (p > 0.05): Significance thresholds and the crisis of unreplicable research. PeerJ **5**: e3544. doi:10.7717/peerj.3544

Anderson, D. R., K. P. Burnham, and W. L. Thompson. 2000. Null hypothesis testing: Problems, prevalence, and an alternative. J. Wildlife Manage. **64**: 912–923. doi:10.2307/3803199

Baker, M. 2016. Is there a reproducibility crisis? Nature **533**: 452–454. doi:10.1038/533452a

Begley, C. G., and J. P. A. Ioannidis. 2015. Reproducibility in science: improving the standard for basic and preclinical research. Circ. Res. **116**: 116–126. doi:10.1161/circresaha.114.303819

Carpenter, S. R. 1998. The need for large-scale experiments to assess and predict the response of ecosystems to perturbation, p. 287–312. *In* M. L. Pace and P. M. Groffman [eds.], Successes, limitations, and frontiers in ecosystem science. Springer. doi:10.1007/978-1-4612-1724-4_12

Carpenter, S. R. 2002. Ecological futures: Building an ecology of the long now. Ecology **83**: 2069–2083. doi:10.2307/3072038

Carpenter, S. R., and others. 1991. Patterns of primary production and herbivory in 25 north American lake ecosystems, p. 67–96. *In* J. J. Cole, S. Findlay, and G. Lovett [eds.], Comparative analyses of ecosystems: Patterns, mechanisms and theories. Springer.

Carpenter, S. R., and J. F. Kitchell. 1993. The trophic cascade in lakes. Cambridge.

Carpenter, S. R., J. F. Kitchell, K. L. Cottingham, D. E. Schindler, D. L. Christensen, D. M. Post, and N. Voichick. 1996. Chlorophyll variability, nutrient input, and grazing: Evidence from whole-lake experiments. Ecology **77**: 725–735. doi:10.2307/2265497

Carpenter, S. R., J. J. Cole, J. F. Kitchell, and M. L. Pace. 1998. Impact of dissolved organic carbon, phosphorus and grazing on phytoplankton biomass and production in lakes. Limnol. Oceanogr. **43**: 73–80. doi:10.4319/lo.1998.43.1.0073

Carpenter, S. R., J. F. Kitchell, J. J. Cole, and M. L. Pace. 1999. Predicting responses of chlorophyll and primary production to changes in phosphorus, grazing and dissolved organic carbon (reply to Nürnberg). Limnol. Oceanogr. **44**: 1179–1182. doi:10.4319/lo.1999.44.4.1179

Carpenter, S. R., and others. 2001. Trophic cascades, nutrients and lake productivity: Whole-lake experiments. Ecol. Monogr. **71**: 163–186. doi:10.1890/0012-9615(2001)071[0163:TCNALP]2.0.CO;2

Carpenter, S. R., and others. 2005. Ecosystem subsidies: Terrestrial support of aquatic food webs from ^{13}C addition to contrasting lakes. Ecology **86**: 2737–2750. doi:10.1890/04-1282

Carpenter, S. R., and others. 2011. Early warnings of regime shifts: A whole-ecosystem experiment. Science **332**: 1079–1082. doi:10.1126/science.1203672

Cole, J. J., and M. L. Pace. 1998. Hydrologic variability of small, northern lakes measured by the addition of tracers. Ecosystems **1**: 310–320. doi:10.1007/s100219900024

Corman, J. R., B. L. Bertolet, N. J. Casson, S. D. Sebestyen, R. K. Kolka, and E. H. Stanley. 2018. Nitrogen and phosphorus loads in temperate seepage lakes associated with allochthonous dissolved organic carbon loads. Geophys. Res. Lett. **45**: 5481–5490. doi:10.1029/2018GL077219

Cottingham, K. L. 1999. Nutrients and zooplankton as multiple stressors of phytoplankton communities: Evidence from size structure. Limnol. Oceanogr. **44**: 810–827. doi:10.1002/lno.v44.5

Cuthbert, I. D., and P. del Giorgio. 1992. Toward a standard method for measuring color in freshwater. Limnol. Oceanogr. **37**: 1319–1326. doi:10.4319/lo.1992.37.6.1319

Dillon, P. J., and F. H. Rigler. 1974. The phosphorus-chlorophyll relationship in lakes. Limnol. Oceanogr. **19**: 767–773. doi:10.4319/lo.1974.19.5.0767

Dillon, P. J., and L. A. Molot. 2005. Long-term trends in catchment export and lake retention of dissolved organic carbon, dissolved organic nitrogen, total iron, and total phosphorus: The Dorset, Ontario, study, 1978-1998. Eur. J. Vasc. Endovasc. Surg. **110**. doi:10.1029/2004JG000003

Downing, J. A., S. B. Watson, and E. McCauley. 2001. Predicting cyanobacteria dominance in lakes. Can. J. Fish. Aquat. Sci. **58**: 1905–1908. doi:10.1139/f01-143

Edmondson, W. T. 1970. Phosphorus, nitrogen, and algae in Lake Washington after diversion of sewage. Science **169**: 690–691. doi:10.1126/science.169.3946.690

Elser, J. J., E. R. Marzolf, and C. R. Goldman. 1990. Phosphorus and nitrogen limitation of phytoplankton growth in the freshwaters of North America: A review and critique of experimental enrichments. Can. J. Fish. Aquat. Sci. **47**: 1468–1477. doi:10.1139/f90-165

Fergus, C. E., A. O. Finley, P. A. Soranno, and T. Wagner. 2016. Spatial variation in nutrient and water color effects on lake chlorophyll at macroscales. PLoS One **11**: e0164592. doi:10.1371/journal.pone.0164592

Filstrup, C. T., T. Wagner, P. A. Soranno, E. H. Stanley, C. A. Strow, K. E. Webster, and J. A. Downing. 2014. Regional variability among nonlinear chlorophyll-phosphorus relationships in lakes. Limnol. Oceanogr. **59**: 1691–1703. doi:10.4319/lo.2014.59.5.1691

Findlay, S., W. H. McDowell, D. Fischer, M. L. Pace, N. Caraco, S. S. Kaushal, and K. C. Weathers. 2010. Total carbon analysis may overestimate organic carbon content of fresh waters in the presence of dissolved inorganic carbon. Limnol. Oceanogr.: Methods **8**: 196–201. doi:10.4319/lom.2010.8.196

Hamilton, D. P., N. Salmaso, and H. W. Paerl. 2016. Mitigating harmful cyanobacteria blooms: Strategies for control of nitrogen and phosphorus loads. Aquat. Ecol. **50**: 351–355. doi:10.1007/s10452-016-9594-z

Hilborn, R., and M. Mangel. 1997. The ecological detective: Confronting models with data. Princeton.

Houlahan, J. T., S. T. McKinney, T. M. Anderson, and B. J. McGill. 2017. The priority of prediction in ecological understanding. Oikos **126**: 1–7. doi:10.1111/oik.03726

Houser, J. N., S. R. Carpenter, and J. J. Cole. 2000. Food web structure and nutrient enrichment: Effects of sediment phosphorus retention in whole-lake experiments. Can. J. Fish. Aquat. Sci. **57**: 1524–1533. doi:10.1139/f00-075

Hutchinson, G. E. 1957. A treatise on limnology, volume 1: Geography, physics, and chemistry. Wiley.

Jeppesen, E., M. Søndergaard, M. Meerhoff, T. L. Lauridsen, and J. P. Jensen. 2007. Shallow lake restoration by nutrient loading reduction—some recent findings and challenges ahead. Hydrobiologia **584**: 239–252. doi:10.1007/s10750-007-0596-7

Jones, R. I. 1992. The influence of humic substances on lacustrine food chains. Hydrobiologia **229**: 73–91. doi:10.1007/BF00006992

Knapp, A. K., and others. 2012. Past, present, and future roles of long-term experiments in the LTER network. Bioscience **42**: 377–389.

Kritzberg, E. S. 2017. Centennial-long trends of lake browning show major effect of afforestation. Limnol. Oceanogr. Lett. **2**: 105–112. doi:10.1002/lol2.10041

Kritzberg, E. S., and S. M. Ekstrom. 2012. Increasing iron concentrations in surface waters—a factor behind brownification? Biogeosciences **9**: 1465–1478. doi:10.5194/bg-9-1465-2012

Lewis, W. M., Jr., W. A. Wurtsbaugh, and H. W. Paerl. 2011. Rationale for control of anthropogenic nitrogen and phosphorus to reduce eutrophication of inland waters. Environ. Sci. Technol. **45**: 10300–10305. doi:10.1021/es202401p

Lindenmayer, D. B., and others. 2012. Value of long-term studies. Austral Ecol. **37**: 745–757. doi:10.1111/j.1442-9993.2011.02351.x

Marker, A. F. H., C. A. Crowther, and R. J. M. Gunn. 1980. Methanol and acetone as solvents for estimating chlorophyll a and phaeopigments by spectrophotometry. Arch. Hydrobiol. Beih. Ergebn. Limnol. **14**: 52–69.

McNutt, M., K. Lehnert, B. Hanson, B. A. Nosek, A. M. Ellison, and J. L. King. 2016. Liberating field samples and data. Science **351**: 1024–1026. doi:10.1126/science.aad7048

McShane, B. B., D. Gal, A. Gelman, C. Robert, and J. L. Tackett. 2018. Abandon statistical significance. arXiv: 1709.07588v2. doi:10.1016/j.pupt.2018.03.005

Melillo, J. M., S. D. Frey, K. M. DeAngelis, W. J. Werner, M. J. Bernard, F. P. Bowles, G. Pold, and A. S. Grandy. 2017. Long-term pattern and magnitude of soil carbon feedback to the climate system in a warming world. Science **358**: 101–105. doi:10.1126/science.aan2874

Molot, L. A. 2009. Color of aquatic ecosystems, p. 657–663. *In* G. E. Likens [ed.], Encyclopedia of inland waters. Elsevier.

Monteith, D. T., and others. 2007. Dissolved organic carbon trends resulting from changes in atmospheric deposition chemistry. Nature **450**: 537–541. doi:10.1038/nature06316

Nosek, B. A., and T. H. Errington. 2017. Reproducibility in cancer biology: Making sense of replication. Elife **6**: e23383. doi:10.7554/elife.23383

Nürnberg, G. K., and M. Shaw. 1998. Productivity of clear and humic lakes: Nutrients, phytoplankton, bacteria. Hydrobiologia **382**: 97–112. doi:10.1023/A:1003445406964

Open Science Collaboration. 2015. Promoting an open research culture. Science **348**: 1422–1425. doi:10.1126/science.aab2374

Pace, M. L. 1984. Zooplankton community structure, but not biomass, influences the phosphorus–chlorophyll a relationship. Can. J. Fish. Aquat. Sci. **41**: 1089–1096. doi:10.1139/f84-128

Pace, M. L., and J. J. Cole. 2000. Effects of whole lake manipulations of nutrient loading and food web structure on planktonic respiration. Can. J. Fish. Aquat. Sci. **57**: 487–496. doi:10.1139/f99-279

Pace, M. L., R. D. Batt, C. D. Buelo, S. R. Carpenter, J. J. Cole, J. T. Kurtzweil, and G. M. Wilkinson. 2017. Reversal of a cyanobacterial bloom in response to early warnings. Proc. Natl. Acad. Sci. USA **114**: 352–357. doi:10.1073/pnas.1612424114

Peters, D. P. C., H. W. Loescher, M. D. SanClements, and K. Havstad. 2014. Taking the pulse of a continent: Expanding site-based research infrastructure for regional to continental-scale ecology. Ecosphere **5**: art29. doi:10.1890/ES13-00295.1

R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Reckhow, K. H., and S. C. Chapra. 1983. Engineering approaches for lake management, v. **1**. Butterworth.

Reich, P. B., S. E. Hobbie, T. D. Lee, and M. A. Pastore. 2018. Unexpected reversal of C_3 versus C_4 grass response to elevated CO_2 during a 20-year field experiment. Science **360**: 317–320. doi:10.1126/science.aas9313

Sarnelle, O. 1992. Nutrient enrichment and grazer effects on phytoplankton in lakes. Ecology **73**: 551–560. doi:10.2307/1940761

Schindler, D. W., and others. 2008. Eutrophication of lakes cannot be controlled by reducing nitrogen input: Results of a 37-year whole-ecosystem experiment. Proc. Nat. Acad. Sci. USA **105**: 11254–11258. doi:10.1073/pnas.0805108105

Seekell, D. A., J.-F. Lapierre, J. Ask, A.-K. Bergström, A. Deininger, P. Rodriguez, and J. Karlsson. 2015. The influence of dissolved organic carbon on primary production in northern lakes. Limnol. Oceanogr. **60**: 1276–1285. doi:10.1002/lno.10096

Strayer, D. L., J. J. Cole, S. E. G. Findlay, D. T. Fischer, J. A. Gephart, H. M. Malcom, M. L. Pace, and E. J. Rosi-Marshall. 2014. Decadal-scale change in a large-river ecosystem. Bioscience **64**: 496–510. doi:10.1093/biosci/biu061

Thrane, J.-E., D. O. Hessen, and T. Andersen. 2014. The absorption of light in lakes: Negative impacts of dissolved organic carbon on primary production. Ecosystems **17**: 1040–1052. doi:10.1007/s10021-014-9776-2

Vähätalo, A. V., K. Salonen, U. Münster, M. Järvinen, and R. G. Wetzel. 2003. Photochemical transformation of

allochthonous organic mater provides bioavailable nutrients in a humic lake. Arch. Hydrobiol. **156**: 287–314. doi: 10.1127/0003-9136/2003/0156-0287

Vollenweider, R. A. 1968. Scientific fundamentals of the eutrophication of lakes and flowing waters, with particular reference to nitrogen and phosphorus as factors in eutrophication. Technical Report DAS/CSI/68.27. Paris Organization for Economic and Cooperative Development.

Weyhenmeyer, G. A., Y. T. Prairie, and L. J. Tranvik. 2017. Browning of boreal freshwaters coupled to carbon-iron interactions along the aquatic continuum. PLoS One **9**: e88104. doi:10.1371/journal.pone.0088104

Wilkinson, G. M., S. R. Carpenter, J. J. Cole, M. L. Pace, R. D. Batt, C. D. Buelo, and J. T. Kurtzweil. 2018. Early warning signals precede cyanobacterial blooms in multiple whole-lake experiments. Ecol. Monogr. **88**: 183–208. doi:10.1002/ecm.1286

Williamson, C. E., D. P. Morris, M. L. Pace, and O. G. Olson. 1999. Dissolved organic carbon and nutrients as regulators of lake ecosystems: Resurrection of a more integrated paradigm. Limnol. Oceanogr. **44**: 795–803. doi:10.4319/lo.1999.44.3_part_2.0795

Zwart, J. A., N. Craig, P. T. Kelly, S. D. Sebestyen, C. T. Solomon, B. C. Weidel, and S. E. Jones. 2016. Metabolic and physiochemical responses to a whole-lake experimental increase in dissolved organic carbon in a north-temperate lake. Limnol. Oceanogr. **61**: 723–734. doi: 10.1002/lno.10248

Zwart, J. A., S. D. Sebestyen, C. T. Solomon, and S. E. Jones. 2017. The influence of hydrologic residence time on lake carbon cycling dynamics following extreme precipitation events. Ecosystems **20**: 1000–1014. doi:10.1007/s10021-016-0088-6

Acknowledgments

U.S. National Science Foundation grants (DEB-1455461, DEB-1456151) supported this work. We thank the many collaborators of the Cascade project, who over 33 yr, collected the data that allowed this synthesis.

Conflict of Interest

None declared.

Submitted 15 January 2018
Revised 8 June 2018
Accepted 26 June 2018

Associate editor: Stephanie Hampton