Temporal variation in total phosphorus concentrations revealed from a multidecadal monitoring program on Big Platte Lake, Michigan

Travis O. Brenden • Renée Reilly • Edward Eisch • Aaron Switzer • Gary E. Whelan

Travis O. Brenden • Renée Reilly
Quantitative Fisheries Center, Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824 USA

Edward Eisch • Aaron Switzer • Gary E. Whelan
Fisheries Division, Michigan Department of Natural Resources, P.O. Box 30446, Lansing, MI 48909 USA

Corresponding author contact information: Travis O. Brenden; brenden@msu.edu; 517-930-2156

Abstract Effective water quality management depends on enactment of appropriately-designed monitoring programs to reveal current and forecasted conditions. Because water quality conditions are influenced by numerous factors, commonly measured attributes such as total phosphorus (TP) can be highly temporally varying. For highly varying processes, monitoring programs should be long term and periodic quantitative analyses are needed so that temporal trends can be distinguished from stochastic variation, which can yield insights into potential modifications to the program. Using generalized additive mixed modeling, we assessed temporal (yearly and monthly) trends and quantified other sources of variation (daily and subsampling) in TP concentrations from a multidecadal depth-specific monitoring program on Big Platte Lake, Michigan. Yearly TP concentrations decreased from the late 1980s to late 1990s before rebounding through the early 2000s. At depths of 2.29 to 13.72 m, TP concentrations have cycled around stationary points since the early 2000s, while at the surface and depths ≥ 18.29
concentrations have continued declining. Summer and fall peaks in TP concentrations were observed at most depths, with the fall peak at deeper depths occurring one month earlier than shallower depths. Daily sampling variation (i.e., variation within a given month and year) was greatest at shallowest and deepest depths. Variation in subsamples collected from depth-specific water samples constituted a small fraction of total variation. Based on model results, cost-saving measures to consider for the monitoring program include reducing subsampling of depth-specific concentrations and reducing the number of sampling depths given observed consistencies across the program period.

Key Words: water quality; monitoring program; generalized additive mixed model; Big Platte Lake
Introduction

Water quality monitoring, which entails the collection of physical, chemical, and/or biological characteristics of water through statistical sampling, is a fundamental component of effective water quality management (Ward et al. 1986; Dodds et al. 2012). Information derived from water quality monitoring can reveal the current condition of a system, as well as be used to forecast expected results stemming from alterations in management policies, effects of invasive species, or variations in climate and other large-scale processes, such as land use (Moore et al. 1976; Adrian et al. 2009; Glaser et al. 2009). In the United States, passage of the amended Clean Water Act in 1972, which mandated control of pollutants into navigable waters, prompted many agencies to enact monitoring programs so that compliance with regulations could be monitored (LaBeau et al. 2013). Similarly, European Union (EU) states intensified water quality monitoring after parliament adopted the Water Framework Directive in 2000, which committed EU states to achieving “good” water quality status in all water bodies (Fölster et al 2014). While federal statutes create the framework for many monitoring programs, monitoring efforts are often implemented in partnership with state and local agencies, private individuals, consulting firms, and non-governmental organizations.

Water quality goals and standards vary greatly across systems and states, but usually include protecting recreational uses of waters, ensuring consumable fish, protecting and restoring aquatic ecosystems, and ensuring safe drinking water and public health. Likewise, goals of water quality monitoring programs can be diverse, and include elements related to determination of trends, compliance with water quality standards, and/or assessment of environmental impacts (Whitfield 1988). Ideally, the sampling strategy associated with a particular monitoring program
is developed considering both water quality and monitoring goals in combination as they both influence whether collected data can actually determine whether goals have been met (Moore et al. 1976; Whitfield 1988). Critical design features of sampling strategies for water quality monitoring include measurement time span, measurement frequency, and method of measurement (Moore et al. 1976).

Water quality monitoring involves sampling a time-varying stochastic process (Loftis and Ward 1980), and a range of factors can affect the measured attribute including anthropogenic disturbance and/or management policies, climate, and instrumentation error/noise (Moore et al. 1976; Loftis and Ward 1980). Together, these factors can lead to a high degree of temporal variability in water quality attributes. Ecological and environmental processes that are characterized by high degree temporal variability require long-term monitoring programs so that process patterns (i.e., trends) can be separated from noise, and that the relative importance of different components of variation can be assessed (Hirsch et al. 1982; Franklin 1989; Pace and Cole 1989; Dodds et al. 2012). When short-term monitoring programs are used to characterize a process with high temporal variability, problems can arise because management decisions may be made based on anomalous random results (Dodds et al. 2012).

For water-quality management to benefit fully from a long-term monitoring program, periodic quantitative analysis of collected data also is necessary (Moore et al. 1976; Franklin 1989; Pace and Cole 1989). Ward et al. (1986) described water-quality monitoring as suffering from a “data-rich but information-poor” syndrome because of what they believed were inadequate attempts to extract meaningful information from collected data. This in turn can put monitoring programs at risk of termination because benefits cannot be easily communicated to members of the public, agency administrators, or government officials (Ward et al. 1986).
Quantitative analyses of the data resulting from long-term water-quality monitoring programs can be used to assess short- (seasonal) and long-term (annual) temporal variation (i.e., trends) in the attribute of interest, which can indicate whether management policies are having desired effects or require modification (Whitfield 1988). Quantitative analyses can also be used to assess other components of variation in the attribute of interest due to factors such as instrumentation noise and spatial variability, which may provide beneficial information for making improvements to the monitoring program (Moore et al. 1976; Beck 1987).

Big Platte Lake (44°41.48′N, 86°05.63′W) is a 1,020-ha lake located in the northwest region of the state of Michigan’s Lower Peninsula in Benzie County (Fig. 1). Since the late 1980s, total phosphorus (TP) in Big Platte Lake has been intensively monitored as part of litigation involving phosphorus discharge from the state of Michigan’s Platte River State Fish Hatchery (PRSFH) located upstream from Big Platte Lake. As part of this monitoring, TP concentrations have been measured at multiple depths from a single site approximately every two weeks with triplicate readings taken at each depth. The long-term monitoring of Big Platte Lake and the sampling strategy employed in the monitoring program (i.e., consistent measurement techniques employed over a regular schedule) provide a rather unique opportunity for assessing variation in TP from an inland lake (Hirsch et al. 1982). Prior research by Smith and Canale (2015) assessed volume-weighted averaged TP concentrations from Big Platte Lake using a subset (2005 to 2013) of data from the monitoring program for determining whether the sampling program was appropriate for assessing compliance with a numerical standard (see Site description). From this analysis, it was determined that the sampling program was more intensive than needed based on recent measurements and that reducing the number of readings per depth would still have high power for comparison against the numerical standard (Smith and
Canale 2015). An assessment of temporal trends in the depth-specific TP concentrations across the entire period of the monitoring program has not previously been conducted. The goal for this study was to quantify temporal (yearly, monthly) trends and assess other components of variation (daily, subsampling) variation in depth-specific TP concentrations from the multidecadal, depth-specific monitoring program from Big Platte Lake. A rigorous quantitative analysis decomposing temporal trends of TP concentrations in Big Platte Lake and the variability in daily and subsampling variations will offer insights into possible modifications to the lake’s water quality monitoring program and aid in the design of programs for other lakes in the region (Beck 1987). According to Pace and Cole (1989), dissemination of results on interannual variability in monitored attributes from long-term studies is important because the findings can have broad relevance.

Materials and methods

Site description

Big Platte Lake lies within the Platte River watershed, which has a total surface area of 49,840 ha (Fig. 1). Land use/cover in the watershed is predominantly upland and lowland forest (61.1%), followed by upland openland (16.9%), agriculture (9.4%), water (7.5%), and urban (2.7%) (Fig. 1). Mean and maximum depths of Big Platte Lake are 4.6 and 27.4 m, respectively (Tonello 2010). Shoreline development of Big Platte Lake is heavy with many homes and cottages located around the lake’s perimeter with the exception of the southeast shoreline.
The lake is considered oligotrophic with algal growth limited by phosphorus levels (Canale et al. 2004).

The PRSFH, which is operated by the Michigan Department of Natural Resources Fisheries Division (DNR), is located approximately 13 km upstream from the upper end of Big Platte Lake (Fig. 1). The PRSFH is the primary producer of Coho salmon (*Oncorhynchus kisutch*) for stocking in Michigan, although Chinook salmon (*Oncorhynchus tshawytscha*), Atlantic salmon (*Salmo salar*), and walleye (*Sander vitreus*) also are produced at the hatchery. Historically, the PRSFH used surface water from the Platte River for fish production with the water subsequently becoming enriched with phosphorus from fish egestion and unconsumed feed prior to its being discharged back into the river. In the 1970s, phosphorus loading from the PRSFH was estimated to be as high as 1,960 kg/yr (Canale et al. 2004), which prompted a lawsuit in the 1980s by local residents of Big Platte Lake (Platte Lake Improvement Association) against the DNR to reduce phosphorus discharge from the hatchery. In 2000, a settlement agreement between the parties was reached whereby phosphorus discharge from the hatchery after facility renovations would be reduced to a maximum of 79.5 kg/yr and no more than a total of 34.0 kg in any 3-month period (Canale et al. 2004). The settlement agreement also stipulated that volume-weighted averaged TP concentration of Big Platte Lake should be less than 8.0 μg/L 95% of the time (Canale et al. 2004). Facility renovations of the PRSFH were completed in 2004. Between 2000 and 2009, the PRSFH was occasionally out of compliance with the settlement agreement. Since summer 2010, phosphorus discharge from the PRSFH has complied with the settlement agreement.

Canale et al. (2010) constructed a phosphorus budget for Big Platte Lake using monitoring data collected to the mid 2000s. According to their analysis, based on typical loads
and lake inflow rates, 86% of the baseline total phosphorus load to Big Platte Lake originated from nonpoint sources (Canale et al. 2010). Other sources based on their analyses included atmospheric deposition (4%), discharge from the PRSFH (3%), and internal loading from sediment release (3.5%) (Canale et al. 2010).

Sampling methods

The description of the TP sampling in Big Platte Lake has previously been described in Canale et al. (2004, 2010) and Smith and Canale (2015) and is only briefly summarized here. TP concentrations have been measured at 8 depths (0.0, 2.29, 4.57, 9.14, 13.72, 18.29, 22.86, and 27.43 m below the surface) since 1989 from a single site located over the deepest portion of the lake, although sampling at the 2.29 m depth did not begin until early 1993. For this study, we used data collected from November 1989 to November 2014. Sampling has occurred approximately every 2 weeks, weather permitting. Monitoring during the winter months is sometimes difficult because it depends on ice conditions being suitable for safe sampling; the longest time span between successive samples was 105 days during winter 2002. Early in the monitoring period (pre 1999), sampling was sometimes conducted weekly. Across the entire monitoring program period, water samples were collected on average every 16.8 days.

Water samples are collected by lowering a Kemmerer water sampler to the desired depth and activating the sampler trip heads. A single water sample is collected at each depth, with triplicate subsamples taken from each sample for TP analysis. TP concentrations are measured using the acid persulfate digestion-ammonium molybdate method (Eaton et al. 2005). Laboratories that have conducted the TP analyses changed in 2002 and 2012. Because TP
concentrations in Big Platte Lake are near the detection limits for laboratory operations, several quality control measures are implemented to improve accuracy and precision of TP concentration measurements (Smith and Canale 2015). Occasionally, TP concentrations from a subsample are discarded because of presumed contamination. This was generally a rare occurrence as the average number of subsamples available during the course of the study ranged from 2.95 to 2.97 for the various depths. The time series of measured TP concentrations by depth from Big Platte Lake is shown in Fig. 2. In total, 12,488 TP concentration measurements were used for this study.

Statistical analyses

For analyses, TP concentrations were log$_e$ transformed to help stabilize variation in measurements across the time-series. A generalized additive mixed model was fit to the transformed concentrations that included depth-specific intercepts, smoothing components for sampling year, sampling month, and the tensor-product interaction (Wood 2017) between sampling year and month, and a sampling date random effect term that was unique to each measurement depth. The smoothing components for sampling year, sampling month, and the interaction between sampling year and month were intended to describe the temporal trends in TP while the sampling date random effect captured the short-term (i.e., daily) variation in TP concentrations. With this model, the residual component accounted for the variation among subsample concentrations across the sampling depths as well as other stochastic sources of variation. Smoothing components were based on penalized regression splines with the degree of smoothness estimated as part of the model fitting process. The number of knots for the spline
smoothing components was set at 24 knots for year, 12 knots for month, and 12 knots for the year×month interaction. Models were fit by maximum likelihood in R version 3.3.2 (R Core Team 2016) using the bam function from the mcgv library (Wood 2011). Because of the size of the dataset and the complexity of the model, model estimation was performed on Michigan State University high-performance compute clusters.

After fitting the generalized additive mixed model, Pearson residuals were calculated and autocorrelation in the depth-specific residuals was assessed in R using the acf function from the stats library (R Core Team 2016). Autocorrelation in the residuals by sampling depth was assessed using two ways: 1) by randomly sampling a single residual on each sampling date, and 2) by averaging the residuals for each sampling date. For the autocorrelation analysis based on random sampling, we repeated the analysis 1,000 times and calculated the average of the autocorrelation value across the iterations. We additionally conducted a breakpoint analysis of the depth-specific residuals using the cross-entropy method for normally distributed random variables described in Priyadarshana and Sofronov (2015). The purpose of the breakpoint analysis of the generalized additive mixed model time-series of residuals was to determine if there were points in the time series where the mean or variance of the residuals changed, which might suggest the presence of an influencing factor that was unaccounted for by the generalized additive mixed model. The breakpoint analysis was conducted in R using the breakpoint package (Priyadarshana and Sofronov 2016). Breakpoints in the mean of the residuals was determined using CE.Normal.Mean function, whereas breakpoints in the mean or variance of the residuals was determined using the CE.Normal.MeanVar function. The maximum number of possible breakpoints was set at 20 with the optimum number of breakpoint determined using Bayesian information criterion model selection (Priyadarshana and Sofronov 2016).
Results

Transformation of the TP concentrations resulted in more homogenous variation across the time series, although there remained some concentrations that might be considered as outliers at the shallowest and deepest measurement depths (Fig. 3). Plots of both the raw (Fig. 3) and transformed (Fig. 4) measurements suggest that TP concentrations declines early in the time series, followed by a rebound and subsequent periodicity in the concentrations. Visual determination of temporal trends in the TP concentrations is difficult because of the considerable amount of variation evident in measurements from the sampling program.

The generalized additive mixed model fit to the transformed TP concentrations converged on a solution, although it took nearly 65 hours for the model to be estimated even with analyses performed on Michigan State University high-performance compute clusters. The adjusted $R^2$ for the estimated model was 90.0%. The basis dimensions for the smoothing effects for year, month, and year×month interaction were appropriate based on residual randomization tests described in Wood (2017).

The depth-specific intercepts for the generalized additive mixed model indicated that TP concentrations increased with sampling depth with deeper areas having the largest differences between sampling depths (Table 1). In other words, there was a larger difference in TP concentrations between the 22.86 and 27.43 m sampling depths then between the 0.0 and 4.57 m sampling depths. The smoothing components for year were largely consistent across the different sampling depths and suggested generally declining TP concentrations from 1989 to the late 1990s followed by increasing concentrations from the late 1990s to the early 2000s (Fig. 4).
At sampling depths of 2.29 m to 13.72 m, TP concentrations exhibited some cycling from the early 2000s to 2014, whereas at the surface and sampling depths of 18.29 to 27.43 m TP concentrations steadily declined with more rapid declines at the deeper depths (Fig. 4).

The smoothing components for month were consistent for sampling depths ranging from 4.57 to 18.29 m with peak concentrations occurring in June and around November (Fig. 5). At a sampling depth of 0 m, a peak in transformed TP concentrations also occurred in November (Fig. 5), with a somewhat smaller peak around March. At the 22.86 and 27.43 m sampling depths, transformed TP concentrations peaked in June with a smaller peak in September. At the deepest sampling depth, there was another peak in concentrations around January (Fig. 5). For the 2.29 m sampling depth, the estimated smoothing component for month was linear and suggested generally increasing concentrations during the course of a year (Fig. 5).

The smoothing components for the year×month interactions suggested that for each sampling depth there were particular years where TP concentrations exhibited even greater monthly fluctuations than what was suggested from the estimated monthly smoothing component (Fig. 6). For example, across most sampling depths October to November was typically associated with peak TP concentrations based on the estimated monthly smoothing component. Based on the smoothing component for the interaction between year and month, in the early years of the sampling program there was a negative effect (i.e., TP concentrations were lower than what was predicted from the additive year and month effects) predicted from the year and month interaction whereas in later years there was a positive effect (i.e., TP concentrations were greater than what was predicted from the additive year and month effects) (Fig. 6). Conversely, the opposite was true (positive effect predicted for early in the time series and negative effect predicted for later in the time series) for the March and April sampling months (Fig. 6).
The standard deviation estimates for the sampling date random effect were the largest and nearly equal at the shallowest (0 m: 0.218) and deepest (27.43 m: 0.225) sampling depths meaning that these depths had the largest daily fluctuations in TP concentrations (Table 2). The standard deviation estimates for the sampling date random effort for the other depths ranged 0.155 to 0.187 (Table 2). The standard deviation for the residual component of the generalized additive mixed model, which accounts for all remaining unexplained variation in the data including factors such as variation among subsamples, was 0.090 (Table 2).

Examination of model predictions based only on the smooth terms for year, month, and year×month interactions (i.e., absent the predictions from the sampling date random effect), supported the general pattern from the visual examination of the transformed TP concentrations (i.e., initial decline early in the sampling period followed by somewhat of a rebound in the late 1990s and early 2000s) but also better revealed some of the seasonal trend in the concentrations (Fig. 7). Including the random effect predictions in the model predictions clearly demonstrated the extent of sampling date variation in the concentrations across the time series (Fig. 8).

The lag-1 autocorrelation when sampling date residuals were randomly sampled was less than 0.005 for each of the sampling depths. Conversely, when sampling date residuals were averaged, the lag-1 autocorrelations ranged from 0.103 to 0.235 for the sampling depths, suggesting there was some, although not strong, positive autocorrelation in TP concentrations across sampling dates that was not accounted for in the generalized additive mixed model fit to the observed data.

No mean breakpoints were detected from the breakpoint analyses of the residuals from the generalized additive mixed model at any of the sampling depths. When breakpoint analyses were allowed to account for changes in mean or variances, some breakpoints were identified for
each sampling depth (Fig. 9). The number of estimated breakpoints for each of the sampling depths ranged from two (0 m) to nine (27.43 m). All sampling depths except for the 27.43 m sampling depth had four or fewer estimated breakpoints. At the 0.0 and 2.29 m sampling depths, breakpoints were identified within a couple of months of when the first laboratory change occurred (Fig 9). Across all sampling depths, breakpoints were identified within 8 months of when the second laboratory occurred (Fig. 9).

Discussion

Maintaining water quality monitoring programs can be expensive and logistically challenging (Dodds et al. 2012; La Beau et al. 2013); consequently, many monitoring programs are characterized by short periods and irregular sampling (Whitfield 1988; Stow 1995). Oftentimes, monitoring programs are initiated to evaluate the success of a particular restoration project and consequently programs may have limited funding or have been instigated by a political directive (Lindemayer and Likens 2009), which likely contributes to the paucity of long-term monitoring programs. One proposed solution for dealing with limited funding to support monitoring is to establish endowments and use the earned interest to support the program (Steinman and Ogdahl 2004). Despite the associated challenges in maintaining long-term monitoring programs, their importance is widely recognized among ecologists and natural resource managers (Lindemayer and Likens 2009). Long-term monitoring programs are crucial for separating pattern from noise, and increase the chances of finding ecological “surprises” (i.e., unexpected outcomes that lead to major paradigm shifts in thinking) in the measured attribute (Lindenmayer et al. 2010; Dodds et al. 2012). As well, data arising from long-term monitoring
can prove useful for answering/testing future questions/hypotheses that were never foreseen when monitoring was initiated (Burt et al. 2014).

What length of time constitutes “long term” for a monitoring program is admittedly equivocal (Strayer et al. 1986). In a case-study review of long-term data sets, Dodds et al. (2012) evaluated six monitoring programs that ranged in duration from 10 to 80 yrs. Similarly, Lindenmayer et al. (2010) in a review of the types of ecological surprises that can result from long-term studies considered monitored programs with durations of at least 25 years. For this study, we analyzed a 25-year time series of TP concentrations from Big Platte Lake, which is in the range of time spans of the case studies evaluated by Lindenmayer et al. (2010) and Dodds et al. (2012). Dodds et al. (2012) noted that nearly every ecological study that involves some form of active monitoring covers only a small fraction of time from a paleoecological perspective. Nevertheless, they defined a long-term data set as one that is “measured through time using standardized methods that allow for the elucidation of ecological system responses to drivers (e.g., linear, lag, threshold, regime shift) to drivers, disturbances (e.g., presses or pulses) recovery from disturbances, and relevant interactions for a given hypothesis” (Dodds et al. 2012). Major drivers of phosphorus levels in lakes include point sources, nonpoint sources, and internal loading, with point sources tending to be temporally stable and nonpoint sources and internal loading tending to be temporally variable due to linkages with seasonal agricultural activities, irregular climate events, and anthropogenic activities (Carpenter et al. 1998; Orihel 2017). While a 25-year time span is perhaps not long enough to distinguish major land use/cover changes in the surrounding watershed or rare climatic events, it should be of sufficient duration for contrasting temporal variation at the scales of interest for this study (i.e., since major changes in the PRSFH operations were implemented).
As noted earlier, for management to benefit fully from a monitoring program periodic quantitative analysis of collected data is necessary (Moore et al. 1976; Franklin 1989; Pace and Cole 1989). One of the recognized benefits from quantitative analysis of long-term monitoring data is that it provides a framework for organizing information on the measured attributes conditioned by the assumed process and underlying statistical model (Stow 2015). According to Ward et al. (1986), water quality monitoring programs must have firmer scientific and systematic bases if they are to provide useful information for water quality management. Analyses similar to those conducted in this study can reveal the scale of variability in the attribute, which in turn can yield important information for how a monitoring program can be modified. Expansion of the type of continuous monitoring conducted at Big Platte Lake to more systems and watersheds across the state or larger spatial areas (e.g., Laurentian Great Lakes) would likely prove beneficial for providing key information on temporal and systematic changes in important water quality attributes.

Monitoring of TP concentrations in Big Platte Lake was initiated because of concerns and subsequent litigation regarding phosphorus discharge from the PRSFH. The PRSFH was historically regarded as the major point source for TP in Big Platte Lake with a peak phosphorus loading of approximately 2000 kg/yr in the mid 1970s (Canale et al. 2004). Since the late 1970s, phosphorus loading from the PRSFH has declined steadily, with a loading of between 300 and 400 kg/yr in the late 1980s/early 1990s to around 80 kg/yr starting in the late 1990s through to the present (Canale et al. 2004, 2010). The year effect predicted from the generalized additive mixed model fit to the Big Platte Lake TP monitoring program predicted a consistent decline in TP across all sampling depths from 1989 to the late 1990s, mirroring the decrease in phosphorus loading from the hatchery. However, the increase in the predicted TP year effect from the late
1990s to the mid 2000s suggests that whereas phosphorus discharge from the PRSFH was reduced, loading from other sources increased. Except for the PRSFH, no other major point source of phosphorus has been identified in the Platte River watershed (Canale et al. 2010), which points to increased phosphorus input from nonpoint sources, internal loading, or some other source for the increase in TP concentrations. Canale et al. (2004) similarly noted that volume-weighted averaged TP concentrations had declined by approximately 35% from the mid 1970s to the early 2000s despite an approximate 95% reduction in point source phosphorus loading. Canale et al. (2010) attributed the lack of greater reductions in TP concentrations in Platte Lake to increases in non-point source loading and evaluated some remedial actions that might help to further reduce TP concentrations in the lake, although they acknowledged that predicting internal loading of phosphorus can be difficult. In other systems, internal loading of phosphorus has been implicated as a major reason why water quality does not immediately improve post-implementation of management actions (Søndergaard et al. 2003). Fluctuations in internal loading of phosphorus can result from changes in water chemistry, degree of external loading of organic material, chemical concentrations of surface water run-off, and changes in fish and invertebrate community composition (Søndergaard et al. 2003; Orihel et al. 2017). While Big Platte Lake is presently classified as oligotrophic, based on past litigation history it seems clear that residents near the lake have ongoing concerns about TP concentrations and consequently efforts to identify the phosphorus sources should be undertaken.

Monthly variations in TP concentrations can vary considerably across systems, with peak TP concentrations in some systems occurring in the summer while in other systems peak concentrations occur in late fall or early winter, or there is very little monthly variation in concentrations (Johengen et al. 1994; Nicholls et al. 2001). In Big Platte Lake, peak TP
concentrations at most depths occurred in June and in October; an additional peak occurred in February at the surface and deepest sampling depths. The June peak is likely caused by high precipitation or runoff from melting snow during this time of year, which leads to excessive runoff from surrounding watersheds or atmospheric deposition. Fall peaks of TP concentrations in other lakes have been attributed to lake turnover, which results in increases in TP concentrations due to release from sediments (Stewart and Markello 1974). Big Platte Lake does stratify every summer with the deeper (≥18.29 m) portions of the lake turning anoxic, which spurs internal loading of phosphorus (Orihel et al. 2017). Based on the monthly smoothing component estimated from the generalized additive mixed model, a peak concentration of TP at the deepest sampling depths occurred approximately one month earlier than at the other sampling depths, which perhaps is suggestive of phosphorus release from the sediments around this time of year. Additional research into factors causing seasonal variations in TP concentrations would be beneficial.

Based on breakpoint analyses of the model residuals, there is evidence to suggest that the laboratories that have been responsible for determining the TP concentrations from the collected samples have varied in their performance. As previously indicated, TP concentrations in Big Platte Lake are near the detection limits for laboratory operations. When the first laboratory change occurred in 2002, the new lab switched from using a spectrophotometer with a light path of 1 cm to one with a light path of 10 cm, which provided more accurate measures of absorption and thus more accurate measurements of TP concentrations (G. Whelan, personal observation). The laboratory change that occurred in 2012 was primarily to improve the timeliness with which measurements of TP concentrations from the Big Platte Lake monitoring program could be obtained. The timeliness of obtaining TP concentration measurement is important as quicker
results allows for faster adjustments in PRSFH operations, which facilitates the hatchery’s ability
to meet the guidelines agreed upon in the settlement agreement. Methodologies between the
send and third laboratories are believed to be consistent, including the use of 10-cm
spectrophotometer light path for measuring TP concentrations. Nevertheless, the proximity of
the identified breakpoints across all sampling depths to when the laboratory change occurred in
2012 suggests some possible methodological change that is contributing to greater variation in
TP concentration measurements, although we cannot entirely rule out that the increased variation
is environmentally caused. With long-term monitoring programs, shifts in laboratories or
laboratory methods are likely unavoidable. Consequently, developing contingency plans for how
to deal with these type of changes, such as instituting a time-period where samples are processed
by both laboratories or methodologies so results can be compared and contrasted, should be a
specified component of a monitoring program framework so that consequences of these changes
can be conclusively determined.

Whitfield (1988) recommended that water-monitoring programs initially be very
conservative and collect samples frequently, with the aim of modifying the program after initial
evaluation of collected data. Similarly, Lindenmayer and Likens (2009) suggested that long-
term monitoring programs switch to an adaptive framework that allows sampling methodology,
as well as underlying questions and analytic approaches, to evolve over time, while
simultaneously ensuring the integrity of the long-term data record is maintained. Based on our
modeling results, if cost-saving measures were to be implemented to the Big Platte Lake water-
quality monitoring program, perhaps the best option would be to reduce the number of
subsamples collected at each sampling depth. We would recommend reducing the number of
subsamples rather than reducing the sampling frequency given the differences in sizes of the
standard deviations for the sampling date random effect and the residual component of the
model. With respect to other modifications to the sampling program, maintaining monthly
sampling would be prudent given the degree of variation observed across months. Given
qualitatively similar temporal trends observed at some depths, an additional cost-saving measure
that might be warranted would be to reduce the number of sampling depths to a subset of what is
currently sampled. For example, results at the 22.86 and 27.43 m depths for both year and
monthly effects were sufficiently similar that it may not be necessary to continue sampling both
depths. Similarly, results at the 4.57 to 13.72 m depths may also be sufficiently similar that it is
not necessary to continue sampling each of these depths.

The intent of this study was to assess temporal variation in TP concentrations from the
long-term monitoring that has been conducted on Big Platte Lake to inform possible changes to
the lake’s sampling program and facilitate program design for other lakes in the region.
Regional monitoring of TP concentrations in inland lakes can be beneficial for understanding
broad-scale eutrophication fluctuation stemming from land-use changes in an area, but also can
be used as a basis for understanding for assessing aquatic communities of monitored systems
(Paukert and Willis 2003; Bachmann et al. 2012; Gorman et al. 2014). While the Big Platte Lake
monitoring program provides a wealth of information pertaining to temporal variability in TP
concentration, the dataset cannot be used to assess other important aspects of water quality
monitoring programs, such as spatial variation or explorations of factors that might have given
rise to the temporal variation in TP concentrations that we observed. Previous research
conducted on large inland lakes in North America such as Lakes Champlain, Huron, Erie, and
Ontario have shown that trends in TP concentrations can vary considerable across regions within
a system (Nicholls et al. 2001; Smeltzer et al. 2012). Although Big Platte Lake is considerably
smaller than the aforementioned systems, how TP concentrations may spatially vary across the
system and how any spatial variation might compare to temporal variation is not clear. Some
additional water quality attributes are collected as part of the Big Platte Lake monitoring
program; however, these data were not collected across the entire time series, which limited our
ability to conduct analyses to explain some of the observed variation in TP concentrations.
Future monitoring programs on either Big Platte Lake or other inland lakes should consider the
costs and benefits of expanding sampling coverage to more than one region and collecting
information on possible explanatory variables for the water quality attribute under study to
strengthen the forecasting quality of constructed models. Additionally, according to Franklin
(1989) and Lindenmayer et al. (2010), long-term studies benefit when they are able to encompass
elements of experimentation so that responses tied to experimental alteration can be explicitly
measured. These changes will undoubtedly elevate costs of monitoring programs, but would
also increase the chances of novel scientific discoveries from the programs (Lindenmayer et al.
2010).

Conclusions

Modeling revealed nonlinear year and month trends in TP concentrations from Big Platte Lake,
MI based on measurements collected from the multidecadal monitoring program. Additionally,
there was a high degree of daily variation in TP concentrations, with considerably lower
variation associated with conducting triplicate measurements at each sampling depth. Overall
temporal trends in TP concentrations were different among some of the sampling depths, with
none of the trends aligning well with phosphorus loading reductions that have occurred due to
operational changes at the PRSFH, which is the only major point source for phosphorus to the lake. This mismatch between TP trends and PRSFH phosphorus loading suggests that reduced loading from the hatchery has been offset by increases in other sources. Follow-up analyses of model residuals suggest laboratories that have processed Big Platte Lakes water samples have possibly differed in their ability to obtain precise measurements. To lower monitoring program costs, reducing the number of readings at each sampling depth or reducing the number of sampled depths would be the best option based on modeling results. Given widespread concerns about socio-economic and human health consequences of eutrophication, we anticipate TP monitoring of aquatic systems will continue to be a routine part of water quality management; the degree of temporal variation observed in this study suggest that sporadic or haphazard collections will unlikely yield an accurate picture of TP levels in the monitored system. When designing long-term water quality monitoring programs, procedures for dealing with laboratory or methodological changes should be included in designs to ensure consistency in the time series.

Acknowledgements Funding for this research was provided by contributing partners of the Quantitative Fisheries Center, which includes Michigan State University, Great Lakes Fisheries Commission, Michigan Department of Natural Resources – Fisheries Division, and other Council of Lake Committee fishery management agencies. The authors thank Wilfred Swiecki and other members of the Platte Lake Improvement Association for their efforts in maintaining the Big Platte Lake total phosphorus monitoring program and database. This work was supported in part by Michigan State University through computational resources provided by the
Institute for Cyber-Enabled Research. This is publication 20XX-XX of the Quantitative Fisheries Center at Michigan State University.

References


Table 1. Depth-specific intercepts and standard errors from generalized additive mixed model fit to the $\log_e$ TP concentrations from Big Platte Lake, Michigan.

<table>
<thead>
<tr>
<th>Depth (m)</th>
<th>Coefficient Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>1.960</td>
<td>0.010</td>
</tr>
<tr>
<td>2.29</td>
<td>1.995</td>
<td>0.022</td>
</tr>
<tr>
<td>4.57</td>
<td>1.993</td>
<td>0.007</td>
</tr>
<tr>
<td>9.14</td>
<td>2.004</td>
<td>0.008</td>
</tr>
<tr>
<td>13.72</td>
<td>2.008</td>
<td>0.007</td>
</tr>
<tr>
<td>18.29</td>
<td>2.025</td>
<td>0.008</td>
</tr>
<tr>
<td>22.86</td>
<td>2.183</td>
<td>0.009</td>
</tr>
<tr>
<td>27.43</td>
<td>2.321</td>
<td>0.010</td>
</tr>
</tbody>
</table>
Table 2. Standard deviation estimates and 95% confidence intervals (in parentheses) for the smoothing components, sampling date random effects, and residual component from the generalized additive mixed model fit to the log$_e$ TP concentrations from Big Platte Lake, Michigan. Standard deviations exist for smoothing components because the mgcv package estimates degree of smoothness as a random effect. Two standard deviations exist for the year×month interaction because of how the interaction is parameterized. The standard deviation estimate for the residual effect represents remaining variation in TP concentrations and includes variation across subsamples.

<table>
<thead>
<tr>
<th>Model Effect</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year (Depth 0.00 m)</td>
<td>0.083 (0.043 – 0.159)</td>
</tr>
<tr>
<td>Year (Depth 2.29 m)</td>
<td>0.061 (0.031 – 0.119)</td>
</tr>
<tr>
<td>Year (Depth 4.57 m)</td>
<td>0.13 (0.069 – 0.245)</td>
</tr>
<tr>
<td>Year (Depth 9.14 m)</td>
<td>0.076 (0.031 – 0.187)</td>
</tr>
<tr>
<td>Year (Depth 13.72 m)</td>
<td>0.047 (0.024 – 0.091)</td>
</tr>
<tr>
<td>Year (Depth 18.29 m)</td>
<td>0.056 (0.027 – 0.118)</td>
</tr>
<tr>
<td>Year (Depth 22.86 m)</td>
<td>0.05 (0.024 – 0.106)</td>
</tr>
<tr>
<td>Year (Depth 27.43 m)</td>
<td>0.065 (0.034 – 0.125)</td>
</tr>
<tr>
<td>Month (Depth 0.00 m)</td>
<td>0.095 (0.049 – 0.186)</td>
</tr>
<tr>
<td>Month (Depth 2.29 m)</td>
<td>0.000 (0.000 – N.E.)</td>
</tr>
<tr>
<td>Month (Depth 4.57 m)</td>
<td>0.106 (0.058 – 0.193)</td>
</tr>
<tr>
<td>Month (Depth 9.14 m)</td>
<td>0.111 (0.059 – 0.206)</td>
</tr>
<tr>
<td>Month (Depth 13.72 m)</td>
<td>0.058 (0.022 – 0.153)</td>
</tr>
<tr>
<td>Month (Depth 18.29 m)</td>
<td>0.107 (0.056 – 0.205)</td>
</tr>
<tr>
<td>Parameter</td>
<td>Value</td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Month (Depth 22.86 m)</td>
<td>0.215 (0.129 – 0.357)</td>
</tr>
<tr>
<td>Month (Depth 27.43 m)</td>
<td>0.268 (0.160 – 0.450)</td>
</tr>
<tr>
<td>Year×Month (Depth 0.00 m)</td>
<td>0.009 (0.002 – 0.03); 0.032 (0.016 – 0.067)</td>
</tr>
<tr>
<td>Year×Month (Depth 2.29 m)</td>
<td>0.016 (0.008 – 0.033); 0.043 (0.021 – 0.086)</td>
</tr>
<tr>
<td>Year×Month (Depth 4.57 m)</td>
<td>0.015 (0.006 – 0.038); 0.038 (0.022 – 0.064)</td>
</tr>
<tr>
<td>Year×Month (Depth 9.14 m)</td>
<td>0.014 (0.005 – 0.036); 0.034 (0.019 – 0.061)</td>
</tr>
<tr>
<td>Year×Month (Depth 13.72 m)</td>
<td>0.052 (0.024 – 0.111); 0.020 (0.011 – 0.037)</td>
</tr>
<tr>
<td>Year×Month (Depth 18.29 m)</td>
<td>0.037 (0.018 – 0.077); 0.020 (0.011 – 0.036)</td>
</tr>
<tr>
<td>Year×Month (Depth 22.86 m)</td>
<td>0.027 (0.009 – 0.08); 0.024 (0.012 – 0.050)</td>
</tr>
<tr>
<td>Year×Month (Depth 27.43 m)</td>
<td>0.093 (0.039 – 0.222); 0.018 (0.009 – 0.034)</td>
</tr>
<tr>
<td>Sampling Date (Depth 0.00 m)</td>
<td>0.218 (0.204 – 0.234)</td>
</tr>
<tr>
<td>Sampling Date (Depth 2.29 m)</td>
<td>0.185 (0.171 – 0.200)</td>
</tr>
<tr>
<td>Sampling Date (Depth 4.57 m)</td>
<td>0.155 (0.144 – 0.168)</td>
</tr>
<tr>
<td>Sampling Date (Depth 9.14 m)</td>
<td>0.172 (0.160 – 0.186)</td>
</tr>
<tr>
<td>Sampling Date (Depth 13.72 m)</td>
<td>0.159 (0.148 – 0.172)</td>
</tr>
<tr>
<td>Sampling Date (Depth 18.29 m)</td>
<td>0.174 (0.162 – 0.188)</td>
</tr>
<tr>
<td>Sampling Date (Depth 22.86 m)</td>
<td>0.187 (0.174 – 0.201)</td>
</tr>
<tr>
<td>Sampling Date (Depth 27.43 m)</td>
<td>0.225 (0.209 – 0.242)</td>
</tr>
<tr>
<td>Residual</td>
<td>0.090 (0.088 – 0.091)</td>
</tr>
</tbody>
</table>
Figure Captions

Fig. 1. Platte River watershed and location of Big Platte Lake and Platte River State Fish Hatchery. Land use/land cover in the watershed is also shown and is based on a 2001 land cover dataset derived from classification of Landsat Thematic Mapper imagery (Michigan Geographic Data Library; https://www.mcgi.state.mi.us/mgdl/?rel=thext&action=thmname&cid=5&cat=Land+Cover+2001). The inset shows the location of the Platte River watershed in the state of Michigan.

Fig. 2. Total phosphorus in μg/L by sampling depth from Big Platte Lake, Michigan. The vertical lines identify when laboratories that analyzed collected water samples changed. The horizontal lines indicate the mean total phosphorus concentration across the entire time series at each depth.

Fig. 3. Loge transformed total phosphorus in μg/L by sampling depth from Big Platte Lake, Michigan. The vertical lines identify when laboratories that analyzed collected water samples changed. The horizontal lines indicate the mean total phosphorus concentration across the entire time series at each depth.

Fig. 4. Depth-specific partial predictions (i.e., additive effects) (± 1 SE) of loge total phosphorus in μg/L from Big Platte Lake, Michigan as a function of year based on the fitted generalized additive mixed model. The vertical lines identify when laboratories that analyzed collected water samples changed.

Fig. 5. Depth-specific partial predictions (i.e., additive effects) (± 1 SE) of loge total phosphorus in μg/L from Big Platte Lake, Michigan as a function of month based on the fitted generalized additive mixed model.
Fig. 6. Year-by-month partial predictions (i.e., additive effects) by sampling depth from the generalized additive mixed model fit to loge total phosphorus concentration from Big Platte Lake, Michigan. A positive value indicates year and month combinations where predicted loge total phosphorus is greater than the additive main effects of year (Fig. 4) and month (Fig. 5), whereas a negative effect indicates year and month combinations where predicted loge total phosphorus is smaller than the additive main effects of year (Fig. 4) and month (Fig. 5).

Fig. 7. Observed (circles) and predicted (line) total phosphorus in μg/L by sampling depth from the generalized additive mixed model fit to the loge total phosphorus concentrations from Big Platte Lake, Michigan. The generalized additive mixed model predictions do not include the random effect term for sampling depth meaning the predictions just describe the large-scale temporal trends in total phosphorus. The vertical lines identify when laboratories that analyzed collected water samples changed.

Fig. 8. Observed (circles) and predicted (line) total phosphorus in μg/L by sampling depth from the generalized additive mixed model fit to the loge total phosphorus concentrations from Big Platte Lake, Michigan. Unlike Fig. 7, the generalized additive mixed model predictions include the random effect term for sampling date. The vertical lines identify when laboratories that analyzed collected water samples changed.

Fig. 9. Pearson residuals (black circles) by sampling depth from the generalized additive mixed model fit to the loge total phosphorus concentrations from Big Platte Lake, Michigan. The black ×s overlaying the residuals indicate the location of breakpoints in the mean or variance of the residuals identified by the cross-entropy method (Priyadarshana and
Sofronov 2015). The vertical lines identify when laboratories that analyzed collected water samples changed.
Partial Prediction of Logₐ Total Phosphorous µg/L by Month

Sampling Month

- 22.86 m
- 27.43 m
- 9.14 m
- 13.72 m
- 18.29 m
- 0 m
- 2.29 m
- 4.57 m