

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

3
4 **Travis O. Brenden • René Reilly • Edward Eisch • Aaron Switzer • Gary E. Whelan**

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

11 Edward Eisch • Aaron Switzer • Gary E. Whelan
12 Fisheries Division, Michigan Department of Natural Resources, P.O. Box 30446, Lansing, MI
13 48909 USA

15 Corresponding author contact information: Travis O. Brenden; brenden@msu.edu; 517-930-
16 2156

17
18 **Abstract** Effective water quality management depends on enactment of appropriately-designed
19 monitoring programs to reveal current and forecasted conditions. Because water quality
20 conditions are influenced by numerous factors, commonly measured attributes such as total
21 phosphorus (TP) can be highly temporally varying. For highly varying processes, monitoring
22 programs should be long term and periodic quantitative analyses are needed so that temporal
23 trends can be distinguished from stochastic variation, which can yield insights into potential
24 modifications to the program. Using generalized additive mixed modeling, we assessed temporal
25 (yearly and monthly) trends and quantified other sources of variation (daily and subsampling) in
26 TP concentrations from a multidecadal depth-specific monitoring program on Big Platte Lake,
27 Michigan. Yearly TP concentrations decreased from the late 1980s to late 1990s before
28 rebounding through the early 2000s. At depths of 2.29 to 13.72 m, TP concentrations have
29 cycled around stationary points since the early 2000s, while at the surface and depths ≥ 18.29

30 concentrations have continued declining. Summer and fall peaks in TP concentrations were
31 observed at most depths, with the fall peak at deeper depths occurring one month earlier than
32 shallower depths. Daily sampling variation (i.e., variation within a given month and year) was
33 greatest at shallowest and deepest depths. Variation in subsamples collected from depth-specific
34 water samples constituted a small fraction of total variation. Based on model results, cost-saving
35 measures to consider for the monitoring program include reducing subsampling of depth-specific
36 concentrations and reducing the number of sampling depths given observed consistencies across
37 the program period.

38

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

41 **Introduction**

42

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

58 Water quality goals and standards vary greatly across systems and states, but usually
59 include protecting recreational uses of waters, ensuring consumable fish, protecting and restoring
60 aquatic ecosystems, and ensuring safe drinking water and public health. Likewise, goals of
61 water quality monitoring programs can be diverse, and include elements related to determination
62 of trends, compliance with water quality standards, and/or assessment of environmental impacts
63 (Whitfield 1988). Ideally, the sampling strategy associated with a particular monitoring program

64 is developed considering both water quality and monitoring goals in combination as they both
65 influence whether collected data can actually determine whether goals have been met (Moore et
66 al. 1976; Whitfield 1988). Critical design features of sampling strategies for water quality
67 monitoring include measurement time span, measurement frequency, and method of
68 measurement (Moore et al. 1976).

69 Water quality monitoring involves sampling a time-varying stochastic process (Loftis and
70 Ward 1980), and a range of factors can affect the measured attribute including anthropogenic
71 disturbance and/or management policies, climate, and instrumentation error/noise (Moore et al.
72 1976; Loftis and Ward 1980). Together, these factors can lead to a high degree of temporal
73 variability in water quality attributes. Ecological and environmental processes that are
74 characterized by high degree temporal variability require long-term monitoring programs so that
75 process patterns (i.e., trends) can be separated from noise, and that the relative importance of
76 different components of variation can be assessed (Hirsch et al. 1982; Franklin 1989; Pace and
77 Cole 1989; Dodds et al. 2012). When short-term monitoring programs are used to characterize a
78 process with high temporal variability, problems can arise because management decisions may
79 be made based on anomalous random results (Dodds et al. 2012).

80 For water-quality management to benefit fully from a long-term monitoring program,
81 periodic quantitative analysis of collected data also is necessary (Moore et al. 1976; Franklin
82 1989; Pace and Cole 1989). Ward et al. (1986) described water-quality monitoring as suffering
83 from a “data-rich but information-poor” syndrome because of what they believed were
84 inadequate attempts to extract meaningful information from collected data. This in turn can put
85 monitoring programs at risk of termination because benefits cannot be easily communicated to
86 members of the public, agency administrators, or government officials (Ward et al. 1986).

87 Quantitative analyses of the data resulting from long-term water-quality monitoring programs
88 can be used to assess short- (seasonal) and long-term (annual) temporal variation (i.e., trends) in
89 the attribute of interest, which can indicate whether management policies are having desired
90 effects or require modification (Whitfield 1988). Quantitative analyses can also be used to
91 assess other components of variation in the attribute of interest due to factors such as
92 instrumentation noise and spatial variability, which may provide beneficial information for
93 making improvements to the monitoring program (Moore et al. 1976; Beck 1987).

94 Big Platte Lake (44°41.48'N, 86°05.63'W) is a 1,020-ha lake located in the northwest
95 region of the state of Michigan's Lower Peninsula in Benzie County (Fig. 1). Since the late
96 1980s, total phosphorus (TP) in Big Platte Lake has been intensively monitored as part of
97 litigation involving phosphorus discharge from the state of Michigan's Platte River State Fish
98 Hatchery (PRSFH) located upstream from Big Platte Lake. As part of this monitoring, TP
99 concentrations have been measured at multiple depths from a single site approximately every
100 two weeks with triplicate readings taken at each depth. The long-term monitoring of Big Platte
101 Lake and the sampling strategy employed in the monitoring program (i.e., consistent
102 measurement techniques employed over a regular schedule) provide a rather unique opportunity
103 for assessing variation in TP from an inland lake (Hirsch et al. 1982). Prior research by Smith
104 and Canale (2015) assessed volume-weighted averaged TP concentrations from Big Platte Lake
105 using a subset (2005 to 2013) of data from the monitoring program for determining whether the
106 sampling program was appropriate for assessing compliance with a numerical standard (*see Site
107 description*). From this analysis, it was determined that the sampling program was more
108 intensive than needed based on recent measurements and that reducing the number of readings
109 per depth would still have high power for comparison against the numerical standard (Smith and

110 Canale 2015). An assessment of temporal trends in the depth-specific TP concentrations across
111 the entire period of the monitoring program has not previously been conducted. The goal for this
112 study was to quantify temporal (yearly, monthly) trends and assess other components of variation
113 (daily, subsampling) variation in depth-specific TP concentrations from the multidecadal, depth-
114 specific monitoring program from Big Platte Lake. A rigorous quantitative analysis
115 decomposing temporal trends of TP concentrations in Big Platte Lake and the variability in daily
116 and subsampling variations will offer insights into possible modifications to the lake's water
117 quality monitoring program and aid in the design of programs for other lakes in the region (Beck
118 1987). According to Pace and Cole (1989), dissemination of results on interannual variability in
119 monitored attributes from long-term studies is important because the findings can have broad
120 relevance.

121

122 **Materials and methods**

123

124 Site description

125

126 Big Platte Lake lies within the Platte River watershed, which has a total surface area of
127 49,840 ha (Fig. 1). Land use/cover in the watershed is predominantly upland and lowland forest
128 (61.1%), followed by upland openland (16.9%), agriculture (9.4%), water (7.5%), and urban
129 (2.7%) (Fig. 1). Mean and maximum depths of Big Platte Lake are 4.6 and 27.4 m, respectively
130 (Tonello 2010). Shoreline development of Big Platte Lake is heavy with many homes and
131 cottages located around the lake's perimeter with the exception of the southeast shoreline

132 (Tonello 2010). The lake is considered oligotrophic with algal growth limited by phosphorus
133 levels (Canale et al. 2004).

134 The PRSFH, which is operated by the Michigan Department of Natural Resources
135 Fisheries Division (DNR), is located approximately 13 km upstream from the upper end of Big
136 Platte Lake (Fig. 1). The PRSFH is the primary producer of Coho salmon (*Oncorhynchus*
137 *kisutch*) for stocking in Michigan, although Chinook salmon (*Oncorhynchus tshawytscha*),
138 Atlantic salmon (*Salmo salar*), and walleye (*Sander vitreus*) also are produced at the hatchery.
139 Historically, the PRSFH used surface water from the Platte River for fish production with the
140 water subsequently becoming enriched with phosphorus from fish egestion and unconsumed feed
141 prior to its being discharged back into the river. In the 1970s, phosphorus loading from the
142 PRSFH was estimated to be as high as 1,960 kg/yr (Canale et al. 2004), which prompted a
143 lawsuit in the 1980s by local residents of Big Platte Lake (Platte Lake Improvement Association)
144 against the DNR to reduce phosphorus discharge from the hatchery. In 2000, a settlement
145 agreement between the parties was reached whereby phosphorus discharge from the hatchery
146 after facility renovations would be reduced to a maximum of 79.5 kg/yr and no more than a total
147 of 34.0 kg in any 3-month period (Canale et al. 2004). The settlement agreement also stipulated
148 that volume-weighted averaged TP concentration of Big Platte Lake should be less than 8.0 µg/L
149 95% of the time (Canale et al. 2004). Facility renovations of the PRSFH were completed in
150 2004. Between 2000 and 2009, the PRSFH was occasionally out of compliance with the
151 settlement agreement. Since summer 2010, phosphorus discharge from the PRSFH has complied
152 with the settlement agreement.

153 Canale et al. (2010) constructed a phosphorus budget for Big Platte Lake using
154 monitoring data collected to the mid 2000s. According to their analysis, based on typical loads

155 and lake inflow rates, 86% of the baseline total phosphorus load to Big Platte Lake originated
156 from nonpoint sources (Canale et al. 2010). Other sources based on their analyses included
157 atmospheric deposition (4%), discharge from the PRSFH (3%), and internal loading from
158 sediment release (3.5%) (Canale et al. 2010).

159

160 Sampling methods

161

162 The description of the TP sampling in Big Platte Lake has previously been described in
163 Canale et al. (2004, 2010) and Smith and Canale (2015) and is only briefly summarized here. TP
164 concentrations have been measured at 8 depths (\approx 0.0, 2.29, 4.57, 9.14, 13.72, 18.29, 22.86, and
165 27.43 m below the surface) since 1989 from a single site located over the deepest portion of the
166 lake, although sampling at the 2.29 m depth did not begin until early 1993. For this study, we
167 used data collected from November 1989 to November 2014. Sampling has occurred
168 approximately every 2 weeks, weather permitting. Monitoring during the winter months is
169 sometimes difficult because it depends on ice conditions being suitable for safe sampling; the
170 longest time span between successive samples was 105 days during winter 2002. Early in the
171 monitoring period (pre 1999), sampling was sometimes conducted weekly. Across the entire
172 monitoring program period, water samples were collected on average every 16.8 days.

173 Water samples are collected by lowering a Kemmerer water sampler to the desired depth
174 and activating the sampler trip heads. A single water sample is collected at each depth, with
175 triplicate subsamples taken from each sample for TP analysis. TP concentrations are measured
176 using the acid persulfate digestion-ammonium molybdate method (Eaton et al. 2005).
177 Laboratories that have conducted the TP analyses changed in 2002 and 2012. Because TP

178 concentrations in Big Platte Lake are near the detection limits for laboratory operations, several
179 quality control measures are implemented to improve accuracy and precision of TP concentration
180 measurements (Smith and Canale 2015). Occasionally, TP concentrations from a subsample are
181 discarded because of presumed contamination. This was generally a rare occurrence as the
182 average number of subsamples available during the course of the study ranged from 2.95 to 2.97
183 for the various depths. The time series of measured TP concentrations by depth from Big Platte
184 Lake is shown in Fig. 2. In total, 12,488 TP concentration measurements were used for this
185 study.

186

187 Statistical analyses

188

189 For analyses, TP concentrations were \log_e transformed to help stabilize variation in
190 measurements across the time-series. A generalized additive mixed model was fit to the
191 transformed concentrations that included depth-specific intercepts, smoothing components for
192 sampling year, sampling month, and the tensor-product interaction (Wood 2017) between
193 sampling year and month, and a sampling date random effect term that was unique to each
194 measurement depth. The smoothing components for sampling year, sampling month, and the
195 interaction between sampling year and month were intended to describe the temporal trends in
196 TP while the sampling date random effect captured the short-term (i.e., daily) variation in TP
197 concentrations. With this model, the residual component accounted for the variation among
198 subsample concentrations across the sampling depths as well as other stochastic sources of
199 variation. Smoothing components were based on penalized regression splines with the degree of
200 smoothness estimated as part of the model fitting process. The number of knots for the spline

201 smoothing components was set at 24 knots for year, 12 knots for month, and 12 knots for the
202 year×month interaction. Models were fit by maximum likelihood in R version 3.3.2 (R Core
203 Team 2016) using the *bam* function from the *mcgv* library (Wood 2011). Because of the size of
204 the dataset and the complexity of the model, model estimation was performed on Michigan State
205 University high-performance compute clusters.

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

224

225 **Results**

226

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

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

240 The depth-specific intercepts for the generalized additive mixed model indicated that TP
241 concentrations increased with sampling depth with deeper areas having the largest differences
242 between sampling depths (Table 1). In other words, there was a larger difference in TP
243 concentrations between the 22.86 and 27.43 m sampling depths then between the 0.0 and 4.57 m
244 sampling depths. The smoothing components for year were largely consistent across the
245 different sampling depths and suggested generally declining TP concentrations from 1989 to the
246 late 1990s followed by increasing concentrations from the late 1990s to the early 2000s (Fig. 4).

247 At sampling depths of 2.29 m to 13.72 m, TP concentrations exhibited some cycling from the
248 early 2000s to 2014, whereas at the surface and sampling depths of 18.29 to 27.43 m TP
249 concentrations steadily declined with more rapid declines at the deeper depths (Fig. 4).

250 The smoothing components for month were consistent for sampling depths ranging from
251 4.57 to 18.29 m with peak concentrations occurring in June and around November (Fig. 5). At a
252 sampling depth of 0 m, a peak in transformed TP concentrations also occurred in November (Fig.
253 5), with a somewhat smaller peak around March. At the 22.86 and 27.43 m sampling depths,
254 transformed TP concentrations peaked in June with a smaller peak in September. At the deepest
255 sampling depth, there was another peak in concentrations around January (Fig. 5). For the 2.29
256 m sampling depth, the estimated smoothing component for month was linear and suggested
257 generally increasing concentrations during the course of a year (Fig. 5).

258 The smoothing components for the year×month interactions suggested that for each
259 sampling depth there were particular years where TP concentrations exhibited even greater
260 monthly fluctuations than what was suggested from the estimated monthly smoothing component
261 (Fig. 6). For example, across most sampling depths October to November was typically
262 associated with peak TP concentrations based on the estimated monthly smoothing component.
263 Based on the smoothing component for the interaction between year and month, in the early
264 years of the sampling program there was a negative effect (i.e., TP concentrations were lower
265 than what was predicted from the additive year and month effects) predicted from the year and
266 month interaction whereas in later years there was a positive effect (i.e., TP concentrations were
267 greater than what was predicted from the additive year and month effects) (Fig. 6). Conversely,
268 the opposite was true (positive effect predicted for early in the time series and negative effect
269 predicted for later in the time series) for the March and April sampling months (Fig. 6).

270 The standard deviation estimates for the sampling date random effect were the largest and
271 nearly equal at the shallowest (0 m; 0.218) and deepest (27.43 m: 0.225) sampling depths
272 meaning that these depths had the largest daily fluctuations in TP concentrations (Table 2). The
273 standard deviation estimates for the sampling date random effect for the other depths ranged
274 0.155 to 0.187 (Table 2). The standard deviation for the residual component of the generalized
275 additive mixed model, which accounts for all remaining unexplained variation in the data
276 including factors such as variation among subsamples, was 0.090 (Table 2).

277 Examination of model predictions based only on the smooth terms for year, month, and
278 year×month interactions (i.e., absent the predictions from the sampling date random effect),
279 supported the general pattern from the visual examination of the transformed TP concentrations
280 (i.e., initial decline early in the sampling period followed by somewhat of a rebound in the late
281 1990s and early 2000s) but also better revealed some of the seasonal trend in the concentrations
282 (Fig. 7). Including the random effect predictions in the model predictions clearly demonstrated
283 the extent of sampling date variation in the concentrations across the time series (Fig. 8).

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

290 No mean breakpoints were detected from the breakpoint analyses of the residuals from
291 the generalized additive mixed model at any of the sampling depths. When breakpoint analyses
292 were allowed to account for changes in mean or variances, some breakpoints were identified for

293 each sampling depth (Fig. 9). The number of estimated breakpoints for each of the sampling
294 depths ranged from two (0 m) to nine (27.43 m). All sampling depths except for the 27.43 m
295 sampling depth had four or fewer estimated breakpoints. At the 0.0 and 2.29 m sampling depths,
296 breakpoints were identified within a couple of months of when the first laboratory change
297 occurred (Fig 9). Across all sampling depths, breakpoints were identified within 8 months of
298 when the second laboratory occurred (Fig. 9).

299

300 **Discussion**

301

302 Maintaining water quality monitoring programs can be expensive and logistically
303 challenging (Dodds et al. 2012; La Beau et al. 2013); consequently, many monitoring programs
304 are characterized by short periods and irregular sampling (Whitfield 1988; Stow 1995).
305 Oftentimes, monitoring programs are initiated to evaluate the success of a particular restoration
306 project and consequently programs may have limited funding or have been instigated by a
307 political directive (Lindemayer and Likens 2009), which likely contributes to the paucity of long-
308 term monitoring programs. One proposed solution for dealing with limited funding to support
309 monitoring is to establish endowments and use the earned interest to support the program
310 (Steinman and Ogdahl 2004). Despite the associated challenges in maintaining long-term
311 monitoring programs, their importance is widely recognized among ecologists and natural
312 resource managers (Lindemayer and Likens 2009). Long-term monitoring programs are crucial
313 for separating pattern from noise, and increase the chances of finding ecological “surprises” (i.e.,
314 unexpected outcomes that lead to major paradigm shifts in thinking) in the measured attribute
315 (Lindenmayer et al. 2010; Dodds et al. 2012). As well, data arising from long-term monitoring

316 can prove useful for answering/testing future questions/hypotheses that were never foreseen
317 when monitoring was initiated (Burt et al. 2014).

318 What length of time constitutes “long term” for a monitoring program is admittedly
319 equivocal (Strayer et al. 1986). In a case-study review of long-term data sets, Dodds et al.
320 (2012) evaluated six monitoring programs that ranged in duration from 10 to 80 yrs. Similarly,
321 Lindenmayer et al. (2010) in a review of the types of ecological surprises that can result from
322 long-term studies considered monitored programs with durations of at least 25 years. For this
323 study, we analyzed a 25-year time series of TP concentrations from Big Platte Lake, which is in
324 the range of time spans of the case studies evaluated by Lindenmayer et al. (2010) and Dodds et
325 al. (2012). Dodds et al. (2012) noted that nearly every ecological study that involves some form
326 of active monitoring covers only a small fraction of time from a paleoecological perspective.
327 Nevertheless, they defined a long-term data set as one that is “measured through time using
328 standardized methods that allow for the elucidation of ecological system responses to drivers
329 (e.g., linear, lag, threshold, regime shift) to drivers, disturbances (e.g., presses or pulses)
330 recovery from disturbances, and relevant interactions for a given hypothesis” (Dodds et al.
331 2012). Major drivers of phosphorus levels in lakes include point sources, nonpoint sources, and
332 internal loading, with point sources tending to be temporally stable and nonpoint sources and
333 internal loading tending to be temporally variable due to linkages with seasonal agricultural
334 activities, irregular climate events, and anthropogenic activities (Carpenter et al. 1998; Orihel
335 2017). While a 25-year time span is perhaps not long enough to distinguish major land use/
336 cover changes in the surrounding watershed or rare climatic events, it should be of sufficient
337 duration for contrasting temporal variation at the scales of interest for this study (i.e., since major
338 changes in the PRSFH operations were implemented).

339 As noted earlier, for management to benefit fully from a monitoring program periodic
340 quantitative analysis of collected data is necessary (Moore et al. 1976; Franklin 1989; Pace and
341 Cole 1989). One of the recognized benefits from quantitative analysis of long-term monitoring
342 data is that it provides a framework for organizing information on the measured attributes
343 conditioned by the assumed process and underlying statistical model (Stow 2015). According to
344 Ward et al. (1986), water quality monitoring programs must have firmer scientific and systematic
345 bases if they are to provide useful information for water quality management. Analyses similar
346 to those conducted in this study can reveal the scale of variability in the attribute, which in turn
347 can yield important information for how a monitoring program can be modified. Expansion of
348 the type of continuous monitoring conducted at Big Platte Lake to more systems and watersheds
349 across the state or larger spatial areas (e.g., Laurentian Great Lakes) would likely prove
350 beneficial for providing key information on temporal and systematic changes in important water
351 quality attributes.

352 Monitoring of TP concentrations in Big Platte Lake was initiated because of concerns and
353 subsequent litigation regarding phosphorus discharge from the PRSFH. The PRSFH was
354 historically regarded as the major point source for TP in Big Platte Lake with a peak phosphorus
355 loading of approximately 2000 kg/yr in the mid 1970s (Canale et al. 2004). Since the late 1970s,
356 phosphorus loading from the PRSFH has declined steadily, with a loading of between 300 and
357 400 kg/yr in the late 1980s/early 1990s to around 80 kg/yr starting in the late 1990s through to
358 the present (Canale et al. 2004, 2010). The year effect predicted from the generalized additive
359 mixed model fit to the Big Platte Lake TP monitoring program predicted a consistent decline in
360 TP across all sampling depths from 1989 to the late 1990s, mirroring the decrease in phosphorus
361 loading from the hatchery. However, the increase in the predicted TP year effect from the late

362 1990s to the mid 2000s suggests that whereas phosphorus discharge from the PRSFH was
363 reduced, loading from other sources increased. Except for the PRSFH, no other major point
364 source of phosphorus has been identified in the Platte River watershed (Canale et al. 2010),
365 which points to increased phosphorus input from nonpoint sources, internal loading, or some
366 other source for the increase in TP concentrations. Canale et al. (2004) similarly noted that
367 volume-weighted averaged TP concentrations had declined by approximately 35% from the mid
368 1970s to the early 2000s despite an approximate 95% reduction in point source phosphorus
369 loading. Canale et al. (2010) attributed the lack of greater reductions in TP concentrations in
370 Platte Lake to increases in non-point source loading and evaluated some remedial actions that
371 might help to further reduce TP concentrations in the lake, although they acknowledged that
372 predicting internal loading of phosphorus can be difficult. In other systems, internal loading of
373 phosphorus has been implicated as a major reason why water quality does not immediately
374 improve post-implementation of management actions (Søndergaard et al. 2003). Fluctuations in
375 internal loading of phosphorus can result from changes in water chemistry, degree of external
376 loading of organic material, chemical concentrations of surface water run-off, and changes in
377 fish and invertebrate community composition (Søndergaard et al. 2003; Orihel et al. 2017).
378 While Big Platte Lake is presently classified as oligotrophic, based on past litigation history it
379 seems clear that residents near the lake have ongoing concerns about TP concentrations and
380 consequently efforts to identify the phosphorus sources should be undertaken.

381 Monthly variations in TP concentrations can vary considerably across systems, with peak
382 TP concentrations in some systems occurring in the summer while in other systems peak
383 concentrations occur in late fall or early winter, or there is very little monthly variation in
384 concentrations (Johengen et al. 1994; Nicholls et al. 2001). In Big Platte Lake, peak TP

385 concentrations at most depths occurred in June and in October; an additional peak occurred in
386 February at the surface and deepest sampling depths. The June peak is likely caused by high
387 precipitation or runoff from melting snow during this time of year, which leads to excessive
388 runoff from surrounding watersheds or atmospheric deposition. Fall peaks of TP concentrations
389 in other lakes have been attributed to lake turnover, which results in increases in TP
390 concentrations due to release from sediments (Stewart and Markello 1974). Big Platte Lake does
391 stratify every summer with the deeper (≥ 18.29 m) portions of the lake turning anoxic, which
392 spurs internal loading of phosphorus (Orihel et al. 2017). Based on the monthly smoothing
393 component estimated from the generalized additive mixed model, a peak concentration of TP at
394 the deepest sampling depths occurred approximately one month earlier than at the other sampling
395 depths, which perhaps is suggestive of phosphorus release from the sediments around this time
396 of year. Additional research into factors causing seasonal variations in TP concentrations would
397 be beneficial.

398 Based on breakpoint analyses of the model residuals, there is evidence to suggest that the
399 laboratories that have been responsible for determining the TP concentrations from the collected
400 samples have varied in their performance. As previously indicated, TP concentrations in Big
401 Platte Lake are near the detection limits for laboratory operations. When the first laboratory
402 change occurred in 2002, the new lab switched from using a spectrophotometer with a light path
403 of 1 cm to one with a light path of 10 cm, which provided more accurate measures of absorption
404 and thus more accurate measurements of TP concentrations (G. Whelan, *personal observation*).
405 The laboratory change that occurred in 2012 was primarily to improve the timeliness with which
406 measurements of TP concentrations from the Big Platte Lake monitoring program could be
407 obtained. The timeliness of obtaining TP concentration measurement is important as quicker

408 results allows for faster adjustments in PRSFH operations, which facilitates the hatchery's ability
409 to meet the guidelines agreed upon in the settlement agreement. Methodologies between the
410 send and third laboratories are believed to be consistent, including the use of 10-cm
411 spectrophotometer light path for measuring TP concentrations. Nevertheless, the proximity of
412 the identified breakpoints across all sampling depths to when the laboratory change occurred in
413 2012 suggests some possible methodological change that is contributing to greater variation in
414 TP concentration measurements, although we cannot entirely rule out that the increased variation
415 is environmentally caused. With long-term monitoring programs, shifts in laboratories or
416 laboratory methods are likely unavoidable. Consequently, developing contingency plans for how
417 to deal with these type of changes, such as instituting a time-period where samples are processed
418 by both laboratories or methodologies so results can be compared and contrasted, should be a
419 specified component of a monitoring program framework so that consequences of these changes
420 can be conclusively determined.

421 Whitfield (1988) recommended that water-monitoring programs initially be very
422 conservative and collect samples frequently, with the aim of modifying the program after initial
423 evaluation of collected data. Similarly, Lindenmayer and Likens (2009) suggested that long-
424 term monitoring programs switch to an adaptive framework that allows sampling methodology,
425 as well as underlying questions and analytic approaches, to evolve over time, while
426 simultaneously ensuring the integrity of the long-term data record is maintained. Based on our
427 modeling results, if cost-saving measures were to be implemented to the Big Platte Lake water-
428 quality monitoring program, perhaps the best option would be to reduce the number of
429 subsamples collected at each sampling depth. We would recommend reducing the number of
430 subsamples rather than reducing the sampling frequency given the differences in sizes of the

431 standard deviations for the sampling date random effect and the residual component of the
432 model. With respect to other modifications to the sampling program, maintaining monthly
433 sampling would be prudent given the degree of variation observed across months. Given
434 qualitatively similar temporal trends observed at some depths, an additional cost-saving measure
435 that might be warranted would be to reduce the number of sampling depths to a subset of what is
436 currently sampled. For example, results at the 22.86 and 27.43 m depths for both year and
437 monthly effects were sufficiently similar that it may not be necessary to continue sampling both
438 depths. Similarly, results at the 4.57 to 13.72 m depths may also be sufficiently similar that it is
439 not necessary to continue sampling each of these depths.

440 The intent of this study was to assess temporal variation in TP concentrations from the
441 long-term monitoring that has been conducted on Big Platte Lake to inform possible changes to
442 the lake's sampling program and facilitate program design for other lakes in the region.
443 Regional monitoring of TP concentrations in inland lakes can be beneficial for understanding
444 broad-scale eutrophication fluctuation stemming from land-use changes in an area, but also can
445 be used as a basis for understanding for assessing aquatic communities of monitored systems
446 (Paukert and Willis 2003; Bachmann et al. 2012; Gorman et al. 2014). While the Big Platte Lake
447 monitoring program provides a wealth of information pertaining to temporal variability in TP
448 concentration, the dataset cannot be used to assess other important aspects of water quality
449 monitoring programs, such as spatial variation or explorations of factors that might have given
450 rise to the temporal variation in TP concentrations that we observed. Previous research
451 conducted on large inland lakes in North America such as Lakes Champlain, Huron, Erie, and
452 Ontario have shown that trends in TP concentrations can vary considerable across regions within
453 a system (Nicholls et al. 2001; Smeltzer et al. 2012). Although Big Platte Lake is considerably

454 smaller than the aforementioned systems, how TP concentrations may spatially vary across the
455 system and how any spatial variation might compare to temporal variation is not clear. Some
456 additional water quality attributes are collected as part of the Big Platte Lake monitoring
457 program; however, these data were not collected across the entire time series, which limited our
458 ability to conduct analyses to explain some of the observed variation in TP concentrations.
459 Future monitoring programs on either Big Platte Lake or other inland lakes should consider the
460 costs and benefits of expanding sampling coverage to more than one region and collecting
461 information on possible explanatory variables for the water quality attribute under study to
462 strengthen the forecasting quality of constructed models. Additionally, according to Franklin
463 (1989) and Lindenmayer et al. (2010), long-term studies benefit when they are able to encompass
464 elements of experimentation so that responses tied to experimental alteration can be explicitly
465 measured. These changes will undoubtedly elevate costs of monitoring programs, but would
466 also increase the chances of novel scientific discoveries from the programs (Lindenmayer et al.
467 2010).

468

469 **Conclusions**

470

471 Modeling revealed nonlinear year and month trends in TP concentrations from Big Platte Lake,
472 MI based on measurements collected from the multidecadal monitoring program. Additionally,
473 there was a high degree of daily variation in TP concentrations, with considerably lower
474 variation associated with conducting triplicate measurements at each sampling depth. Overall
475 temporal trends in TP concentrations were different among some of the sampling depths, with
476 none of the trends aligning well with phosphorus loading reductions that have occurred due to

477 operational changes at the PRSFH, which is the only major point source for phosphorus to the
478 lake. This mismatch between TP trends and PRSFH phosphorus loading suggests that reduced
479 loading from the hatchery has been offset by increases in other sources. Follow-up analyses of
480 model residuals suggest laboratories that have processed Big Platte Lakes water samples have
481 possibly differed in their ability to obtain precise measurements. To lower monitoring program
482 costs, reducing the number of readings at each sampling depth or reducing the number of
483 sampled depths would be the best option based on modeling results. Given widespread concerns
484 about socio-economic and human health consequences of eutrophication, we anticipate TP
485 monitoring of aquatic systems will continue to be a routine part of water quality management;
486 the degree of temporal variation observed in this study suggest that sporadic or haphazard
487 collections will unlikely yield an accurate picture of TP levels in the monitored system. When
488 designing long-term water quality monitoring programs, procedures for dealing with laboratory
489 or methodological changes should be included in designs to ensure consistency in the time series.

490

491

492 **Acknowledgements** Funding for this research was provided by contributing partners of the
493 Quantitative Fisheries Center, which includes Michigan State University, Great Lakes Fisheries
494 Commission, Michigan Department of Natural Resources – Fisheries Division, and other
495 Council of Lake Committee fishery management agencies. The authors thank Wilfred Swiecki
496 and other members of the Platte Lake Improvement Association for their efforts in maintaining
497 the Big Platte Lake total phosphorus monitoring program and database. This work was
498 supported in part by Michigan State University through computational resources provided by the

499 Institute for Cyber-Enabled Research. This is publication 20XX-XX of the Quantitative Fisheries
500 Center at Michigan State University.

501

502 **References**

503

504 Adrian, R., O'Reilly, C., Zagarese, H., Baines, S. B., Hessen, D. O., Keller, W., Livingstone, D.
505 M., Sommaruga, R., et al. (2009). Lakes as sentinals of climate change. *Limnology and*
506 *Oceanography* 54:2283-2297. doi: 10.4319/lo.2009.54.6_part_2.2283.

507 Bachmann, R. W., Bigham, D. L., Hoyer, M. V., & Canfield, D. E., Jr. (2012). Phosphorus,
508 nitrogen, and the designated uses of Florida lakes. *Lake and Reservoir Management*
509 28:46-58. doi:10.1080/07438141.2011.650835.

510 Beck, M. B. (1987). Water quality modeling: a review of the analysis of uncertainty. *Water*
511 *Resources Research* 23:1393-1442. doi:10.1029/WR023i008p01393

512 Burt, T. P., Howden, N. J. K., & Worrall, F. (2014). On the importance of very long-term water
513 quality records. *WIREs Water* 1:41-48. doi: 10.1002/wat2.1001

514 Canale, R. P., Harrison, R., Moskus, P., Naperalala, T., Swiecki, W., & Whelan, G. (2004). Case
515 study: reduction of total phosphorus loads to Big Big Platte Lake, MI through point
516 source control and watershed management. *Proceedings of the Water Environment*
517 *Federation Watershed* 4:1060-1076. doi:10.2175/193864704790896829.

518 Canale, R. P., Redder, T., Swiecki, W., & Whelan, G. (2010). Phosphorus budget and
519 remediation plan for Big Big Platte Lake, Michigan. *Journal of Water Resources*
520 *Planning and Management* 136:576-586. doi:10.1061/(ASCE)WR.1943-5452.0000071.

521 Carpenter, S. R., Caraco, N. F., Correll, D. L., Howart, R. W., Sharpley, A. N., & Smith, V. H.
522 (1998). Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological*
523 *Applications* 8:559-568. doi:10.1890/1051-0761(1998)008[0559:NPOSWW]2.0.CO;2.

524 Dodds, W. K., Robinson, C. T., Gaiser, E. E., Hansen, G. J. A., Powell, H., Smith, J. M., Morse,
525 N. B., Johnson, S. L., et al. (2012). Surprises and insights from long-term aquatic data
526 sets and experiments. *BioScience* 62:709-721. doi:10.1525/bio.2012.62.8.4.

527 Eaton, A. D., Clesceri, L. S., Rice, E. W., Greenberg, A. E., & Franson, M. H. (2005). *Standard*
528 *methods for the examination of water and wastewater, 21st edition*. Washington D.C.:
529 American Public Health Association.

530 Fölster, J., Johnson, R. K., Futter, M. N., & Wilander, A. (2014). The Swedish monitoring of
531 surface waters: 50 years of adaptive monitoring. *Ambio* 43(Suppl. 1):3-18. doi:
532 10.1007/s13280-014-0558-z,

533 Franklin, J. F. (1989) Importance and justification of long-term studies in ecology, In. G. E.
534 Likens (Ed.), *Long-term studies in ecology* (pp. 3-19) New York: Springer.

535 Glaser, D., Rhea, J. R., Opdyke, D. R., Russell, K. T., Ziegler, C. K., Ku, W., Zheng, L., &
536 Mastriao, J. (2009). Model of zebra mussel growth and water quality impacts in the
537 Seneca River, New York. *Lake and Reservoir Management* 25:49-72. doi:
538 10.1080/07438140802714411.

539 Gorman, M. W., Zimmer, K. D., Herwig, B. R., Hanson, M. A., Wright, R. G., Vaughn, S. R., &
540 Younk, J. A. (2014). Relative importance of phosphorus, fish biomass, and watershed
541 land use as drivers of phytoplankton abundance in shallow lakes. *Science of the Total*
542 *Environment* 466-467:849-855. doi:10.1016/j.scitotenv.2013.07.106.

543 Hirsch, R. M., Slack, J. R., & Smith, R. A. (1982). Techniques of trend analysis for monthly
544 water quality data. *Water Resources Research* 18:107-121. doi:
545 10.1029/WR018i001p00107.

546 Johengen, T. H., Johannsson, O. E., Pernie, G. L., & Millard, E. S. (1994). Temporal and
547 seasonal trends in nutrient dynamics and biomass measures in Lakes Michigan and
548 Ontario in response to phosphorus control. *Canadian Journal of Fisheries and Aquatic
549 Sciences* 51:2570-2578. doi:10.1139/f94-257.

550 LaBeau, M. B., Gorman, H., Mayer, A., Dempsey, D., & Sherrin, A. (2013). Tributary
551 phosphorus monitoring in the U.S. portion of the Laurentian Great Lake Basin: Drivers
552 and challenges. *Journal of Great Lakes Research* 39: 569-577.
553 doi:10.1016/j.jglr.2013.09.014.

554 Lindenmayer, D. B., & Likens, G. E. (2009). Adaptive monitoring: a new paradigm for long-
555 term research and monitoring. *Trends in Ecology and Evolution* 24:482-486. doi:
556 10.1016/j.tree.2009.03.005.

557 Lindenmayer, D. B., Likens, G. E., Krebs, C. J., & Hobbs, R. J. (2010). Improved probability of
558 detection of ecological “surprises”. *Proceeding of the National Academy of Sciences of
559 the United States of America* 107:21957-21962. doi:10.1073/pnas.1015696107.

560 Loftis, J. C., & Ward, R. C. (1980). Water quality monitoring – some practical sampling
561 frequency considerations. *Environmental Management* 4:521-526. doi:
562 10.1007/BF01876889.

563 Moore, S. F., Dandy, G. C., & DeLucia, R. J. (1976). Describing variance with a simple water
564 quality model and hypothetical sampling programs. *Water Resources Research* 12:795-
565 804. doi:10.1029/WR012i004p00795.

566 Nicholls, K. H., Hopkins, G. J., Standke, S. J., & Nakamoto, L. (2001). Trends in total
567 phosphorus in Canadian near-shore waters of the Laurentian Great Lakes. *Journal of*
568 *Great Lakes Research* 27:402-422. doi:10.1016/S0380-1330(01)70656-9.

569 Orihel, D. M., Baulch, H. M., Casson, N. J., North, R. L., Parsons, C. T., Seckar, D. C. M., &
570 Venkiteswaran, J. J. (2017). Internal phosphorus loading in Canadian fresh waters: a
571 critical review and data analysis. *Canadian Journal of Fisheries and Aquatic Sciences*
572 74:2005-2029. doi: 10.1139/cjfas-2016-0500.

573 Pace, M. L., & Cole, J. J. (1989). What questions, systems, or phenomena warrant long-term
574 ecological study? In. G. E. Likens (Ed.), *Long-term studies in ecology* (pp. 183-185) New
575 York: Springer.

576 Paukert, C. P., & Willis, D. W. (2003). Aquatic invertebrate assemblages in shallow prairie lake:
577 fish and environmental influences. *Journal of Freshwater Ecology* 18:523-536.
578 doi:10.1080/02705060.2003.9663993.

579 Priyadarshana, W. J. R. M, & Sofronov, G. (2015). Multiple break-points detection in array CGH
580 data via the cross-entropy method. *IEEE/ACM Transactions on Computational Biology*
581 *and Bioinformatics* 12:487-498. doi:10.1109/TCBB.2014.2361639.

582 Priyadarshana, W. J. R. M, & Sofronov, G. (2016). *breakpoint: An R Package for multiple*
583 *break-point detection via the cross-entropy method*. R package version 1.2. (www.
584 CRAN.R-project.org/package=breakpoint).

585 R Core Team. (2016). *R: a language and environment for statistical computing*. Vienna, Austria:
586 R Foundation for Statistical Computing. (www.R-project.org).

587 Smeltzer, E., d. Shambaugh, A., & Stangel, P. (2012). Environmental change in Lake Champlain
588 revealed by long-term monitoring. *Journal of Great Lakes Research* 38(Suppl. 1):6-18.
589 doi: 10.1016/j.jglr.2012.01.002.

590 Smith, E. P., & Canale, R. P. (2015). An analysis of sampling programs to evaluate compliance
591 with numerical standards: total phosphorus in Big Platte Lake, MI. *Lake and Reservoir*
592 *Management* 31:190-201. doi:10.1080/10402381.2015.1061073.

593 Søndergaard, M., Jensen, J. P., & Jeppesen, E. (2003). Role of sediment and internal loading of
594 phosphorus in shallow lakes. *Hydrobiologia* 506-509:135-145. doi:
595 doi.org/10.1023/B:HYDR.0000008611.12704.dd.

596 Steinman, A. D., & Ogdahl, M. (2004). An innovative funding mechanism for the Muskegon
597 Lake AOC. *Journal of Great Lakes Research* 30:341-343. doi: 10.1016/S0380-
598 1330(04)70351-2.

599 Stewart, K. M., & Markello, S. J. (1974). Seasonal variation in concentrations of nitrate and total
600 phosphorus, and calculated nutrient loading for six lakes in western New York.
601 *Hydrobiologia* 44:61-89. doi:10.1007/BF00036157.

602 Stow, C. A. (2015). The need for sustained, long-term phosphorus modeling in the Great Lakes.
603 *Journal of Great Lakes Research* 41:315-316. doi:10.1016/j.jglr.2015.03.001.

604 Tonello, M. A. (2010). Big Platte Lake. Status of the Fishery Resource Report 2010-110.
605 Michigan Department of Natural Resources and Environment, Lansing.
606 (https://www.michigan.gov/documents/dnr/2010-110_351459_7.pdf)

607 Ward, R. C., Loftis, J. C., & McBride, G. B. (1986). The “data-rich but information-poor”
608 syndrome in water quality monitoring. *Environmental Management* 10:291-297. doi:
609 10.1007/BF01867251.

610 Whitfield, P. H. (1988). Goals and data collection designs for water quality monitoring. *Water*
611 *Resources Bulletin* 24:775-780. doi: 10.1111/j.1752-1688.1988.tb00928.x.

612 Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood
613 estimation of semiparametric generalized linear models. *Journal of the Royal Statistical*
614 *Society, B* 73:3-36. doi: 10.1111/j.1467-9868.2010.00749.x.

615 Wood, S. N. (2017). *Generalized additive models: an introduction with R, 2nd edition*. Boca
616 Raton, Florida: Chapman and Hall/CRC.

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

| Depth (m) | Coefficient Estimate | Standard Error |
|-----------|----------------------|----------------|
| 0.00 | 1.960 | 0.010 |
| 2.29 | 1.995 | 0.022 |
| 4.57 | 1.993 | 0.007 |
| 9.14 | 2.004 | 0.008 |
| 13.72 | 2.008 | 0.007 |
| 18.29 | 2.025 | 0.008 |
| 22.86 | 2.183 | 0.009 |
| 27.43 | 2.321 | 0.010 |

619

620

621 Table 2. Standard deviation estimates and 95% confidence intervals (in parentheses) for the
622 smoothing components, sampling date random effects, and residual component from the
623 generalized additive mixed model fit to the \log_e TP concentrations from Big Platte Lake,
624 Michigan. Standard deviations exist for smoothing components because the mgcv package
625 estimates degree of smoothness as a random effect. Two standard deviations exist for the
626 year×month interaction because of how the interaction is parameterized. The standard deviation
627 estimate for the residual effect represents remaining variation in TP concentrations and includes
628 variation across subsamples.

| Model Effect | Standard Deviation |
|-----------------------|-----------------------|
| Year (Depth 0.00 m) | 0.083 (0.043 – 0.159) |
| Year (Depth 2.29 m) | 0.061 (0.031 – 0.119) |
| Year (Depth 4.57 m) | 0.13 (0.069 – 0.245) |
| Year (Depth 9.14 m) | 0.076 (0.031 – 0.187) |
| Year (Depth 13.72 m) | 0.047 (0.024 – 0.091) |
| Year (Depth 18.29 m) | 0.056 (0.027 – 0.118) |
| Year (Depth 22.86 m) | 0.05 (0.024 – 0.106) |
| Year (Depth 27.43 m) | 0.065 (0.034 – 0.125) |
| Month (Depth 0.00 m) | 0.095 (0.049 – 0.186) |
| Month (Depth 2.29 m) | 0.000 (0.000 – N.E.) |
| Month (Depth 4.57 m) | 0.106 (0.058 – 0.193) |
| Month (Depth 9.14 m) | 0.111 (0.059 – 0.206) |
| Month (Depth 13.72 m) | 0.058 (0.022 – 0.153) |
| Month (Depth 18.29 m) | 0.107 (0.056 – 0.205) |

| | |
|-------------------------------|--|
| Month (Depth 22.86 m) | 0.215 (0.129 – 0.357) |
| Month (Depth 27.43 m) | 0.268 (0.160 – 0.450) |
| Year×Month (Depth 0.00 m) | 0.009 (0.002 – 0.03); 0.032 (0.016 – 0.067) |
| Year×Month (Depth 2.29 m) | 0.016 (0.008 – 0.033); 0.043 (0.021 – 0.086) |
| Year×Month (Depth 4.57 m) | 0.015 (0.006 – 0.038); 0.038 (0.022 – 0.064) |
| Year×Month (Depth 9.14 m) | 0.014 (0.005 – 0.036); 0.034 (0.019 – 0.061) |
| Year×Month (Depth 13.72 m) | 0.052 (0.024 – 0.111); 0.020 (0.011 – 0.037) |
| Year×Month (Depth 18.29 m) | 0.037 (0.018 – 0.077); 0.020 (0.011 – 0.036) |
| Year×Month (Depth 22.86 m) | 0.027 (0.009 – 0.08); 0.024 (0.012 – 0.050) |
| Year×Month (Depth 27.43 m) | 0.093 (0.039 – 0.222); 0.018 (0.009 – 0.034) |
| Sampling Date (Depth 0.00 m) | 0.218 (0.204 – 0.234) |
| Sampling Date (Depth 2.29 m) | 0.185 (0.171 – 0.200) |
| Sampling Date (Depth 4.57 m) | 0.155 (0.144 – 0.168) |
| Sampling Date (Depth 9.14 m) | 0.172 (0.160 – 0.186) |
| Sampling Date (Depth 13.72 m) | 0.159 (0.148 – 0.172) |
| Sampling Date (Depth 18.29 m) | 0.174 (0.162 – 0.188) |
| Sampling Date (Depth 22.86 m) | 0.187 (0.174 – 0.201) |
| Sampling Date (Depth 27.43 m) | 0.225 (0.209 – 0.242) |
| Residual | 0.090 (0.088 – 0.091) |

630 **Figure Captions**

631 Fig. 1. Platte River watershed and location of Big Platte Lake and Platte River State Fish

632 Hatchery. Land use/land cover in the watershed is also shown and is based on a 2001

633 land cover dataset derived from classification of Landsat Thematic Mapper imagery

634 (Michigan Geographic Data Library;

635 <https://www.mcgi.state.mi.us/mgdl/?rel=thext&action=thmname&cid=5&cat=Land+Cov>

636 [er+2001](https://www.mcgi.state.mi.us/mgdl/?rel=thext&action=thmname&cid=5&cat=Land+Cov)). The inset shows the location of the Platte River watershed in the state of

637 Michigan.

638 Fig. 2. Total phosphorus in $\mu\text{g/L}$ by sampling depth from Big Platte Lake, Michigan. The

639 vertical lines identify when laboratories that analyzed collected water samples changed.

640 The horizontal lines indicate the mean total phosphorus concentration across the entire

641 time series at each depth.

642 Fig. 3. \log_e transformed total phosphorus in $\mu\text{g/L}$ by sampling depth from Big Platte Lake,

643 Michigan. The vertical lines identify when laboratories that analyzed collected water

644 samples changed. The horizontal lines indicate the mean total phosphorus concentration

645 across the entire time series at each depth.

646 Fig. 4. Depth-specific partial predictions (i.e., additive effects) (± 1 SE) of \log_e total phosphorus

647 in $\mu\text{g/L}$ from Big Platte Lake, Michigan as a function of year based on the fitted

648 generalized additive mixed model. The vertical lines identify when laboratories that

649 analyzed collected water samples changed.

650 Fig. 5. Depth-specific partial predictions (i.e., additive effects) (± 1 SE) of \log_e total phosphorus

651 in $\mu\text{g/L}$ from Big Platte Lake, Michigan as a function of month based on the fitted

652 generalized additive mixed model.

653 Fig. 6. Year-by-month partial predictions (i.e., additive effects) by sampling depth from the
654 generalized additive mixed model fit to \log_e total phosphorus concentration from Big
655 Platte Lake, Michigan. A positive value indicates year and month combinations where
656 predicted \log_e total phosphorus is greater than the additive main effects of year (Fig. 4)
657 and month (Fig. 5), whereas a negative effect indicates year and month combinations
658 where predicted \log_e total phosphorus is smaller than the additive main effects of year
659 (Fig. 4) and month (Fig. 5).

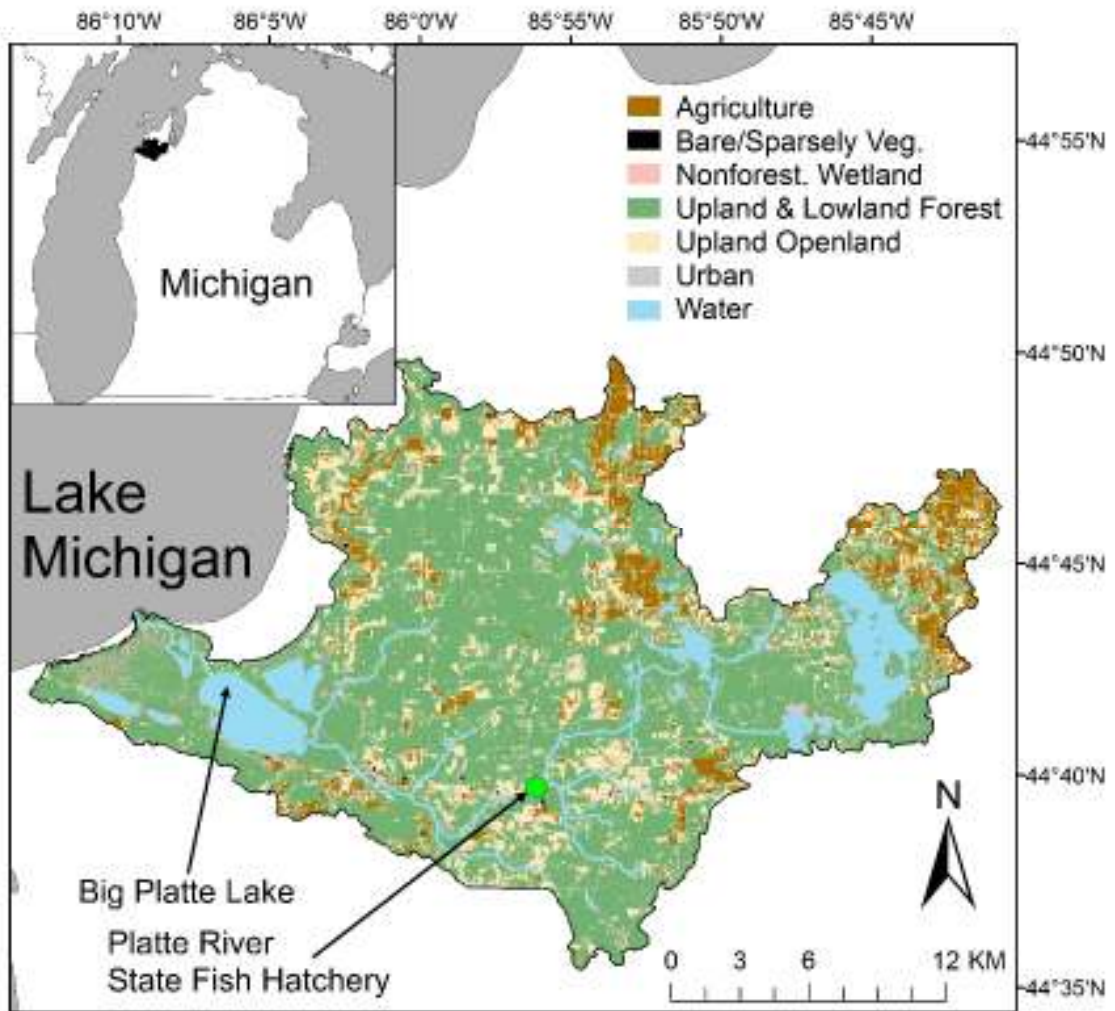
660 Fig. 7. Observed (circles) and predicted (line) total phosphorus in $\mu\text{g/L}$ by sampling depth from
661 the generalized additive mixed model fit to the \log_e total phosphorus concentrations from
662 Big Platte Lake, Michigan. The generalized additive mixed model predictions do not
663 include the random effect term for sampling depth meaning the predictions just describe
664 the large-scale temporal trends in total phosphorus. The vertical lines identify when
665 laboratories that analyzed collected water samples changed.

666 Fig. 8. Observed (circles) and predicted (line) total phosphorus in $\mu\text{g/L}$ by sampling depth from
667 the generalized additive mixed model fit to the \log_e total phosphorus concentrations from
668 Big Platte Lake, Michigan. Unlike Fig. 7, the generalized additive mixed model
669 predictions include the random effect term for sampling date. The vertical lines identify
670 when laboratories that analyzed collected water samples changed.

671 Fig. 9. Pearson residuals (black circles) by sampling depth from the generalized additive mixed
672 model fit to the \log_e total phosphorus concentrations from Big Platte Lake, Michigan. The
673 black \times s overlaying the residuals indicate the location of breakpoints in the mean or
674 variance of the residuals identified by the cross-entropy method (Priyadarshana and

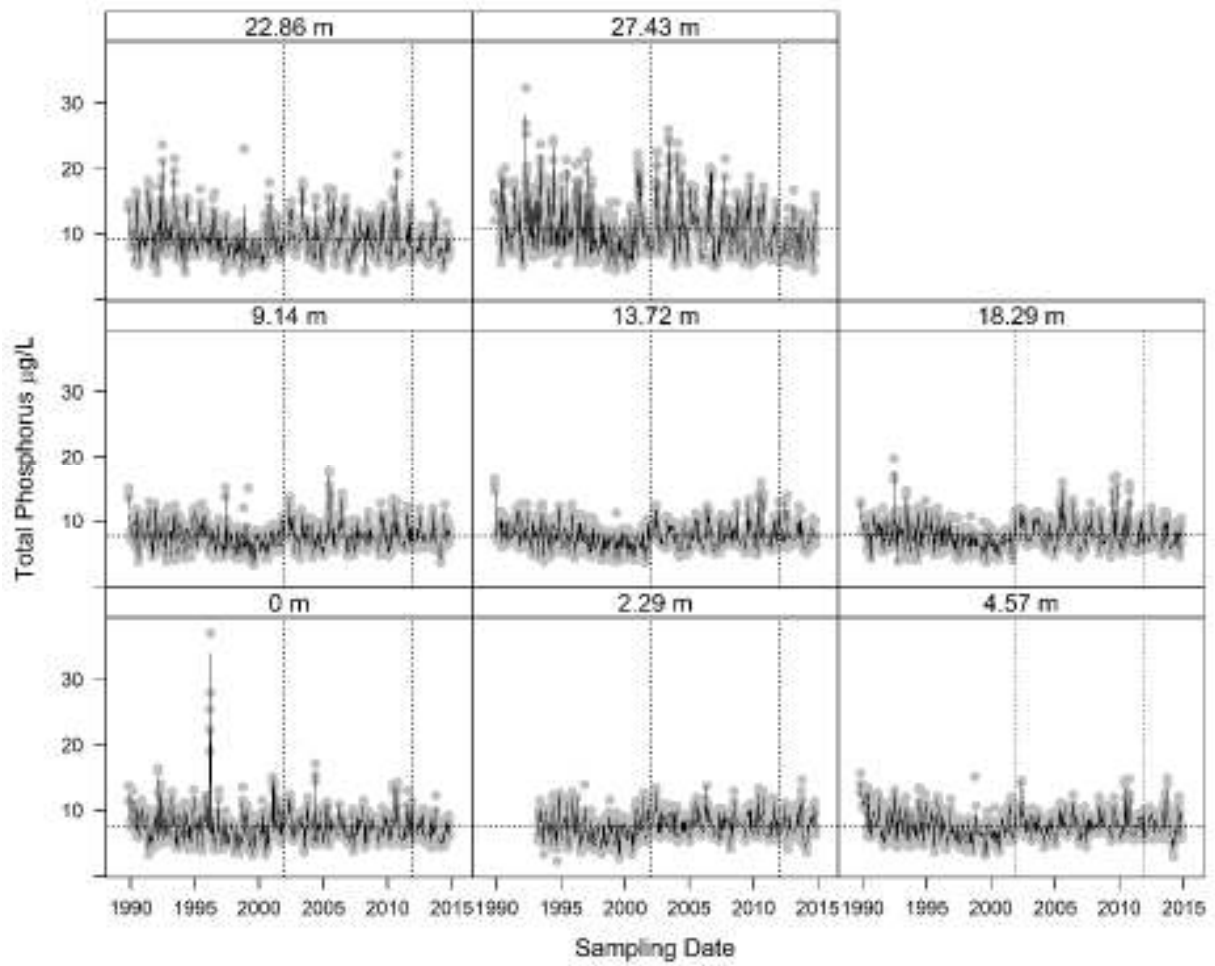
675 Sofronov 2015). The vertical lines identify when laboratories that analyzed collected
676 water samples changed.

677



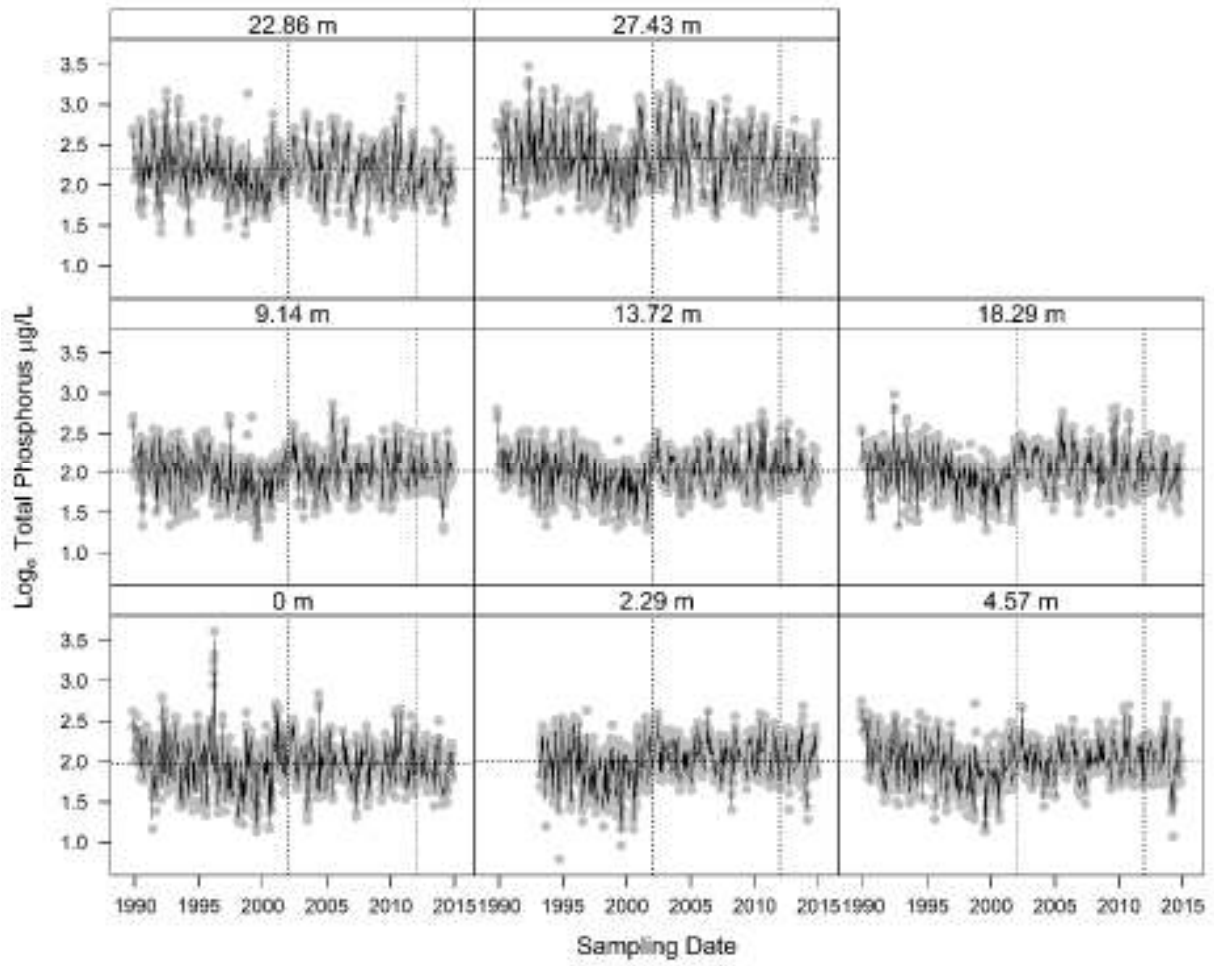
678

679

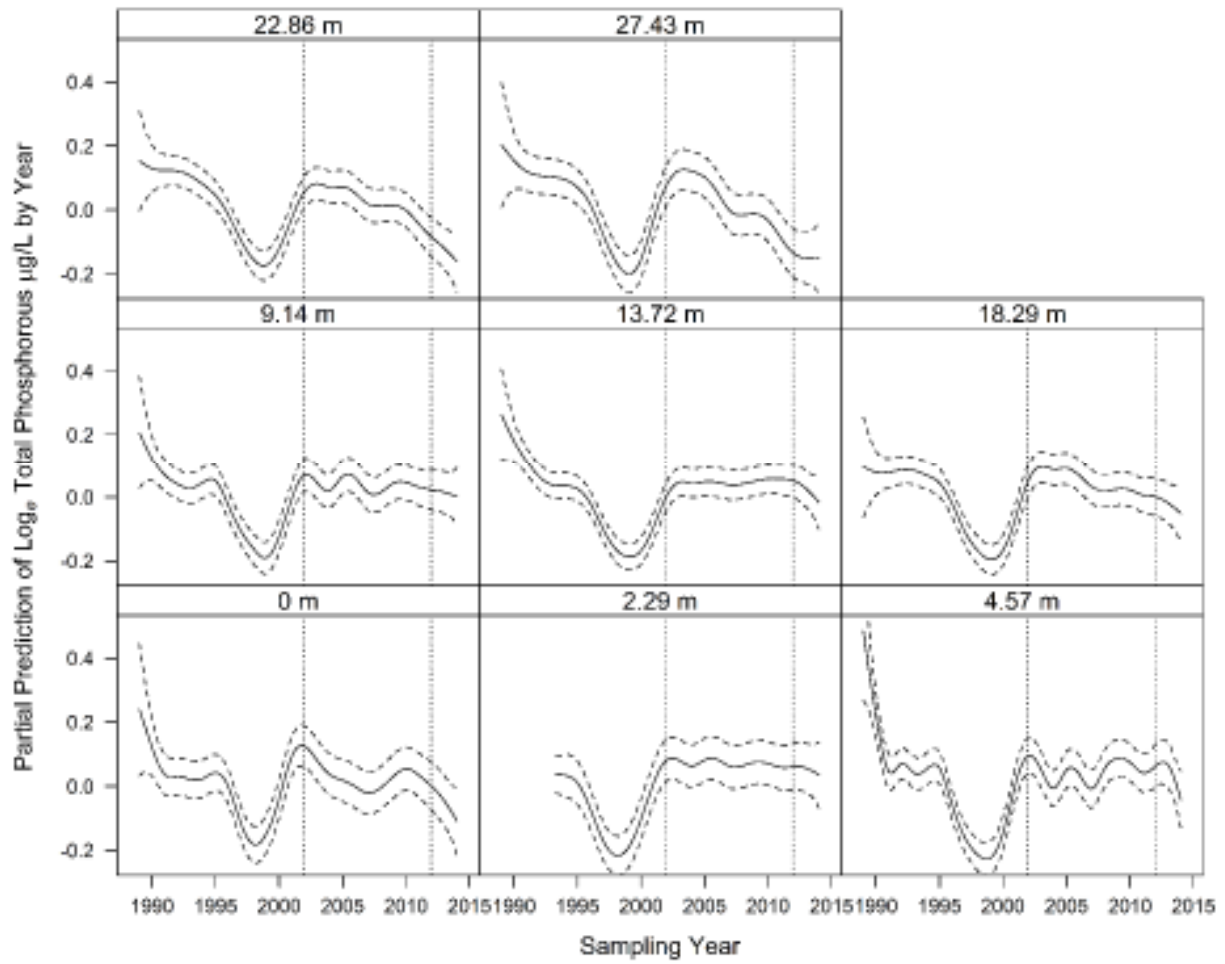


680

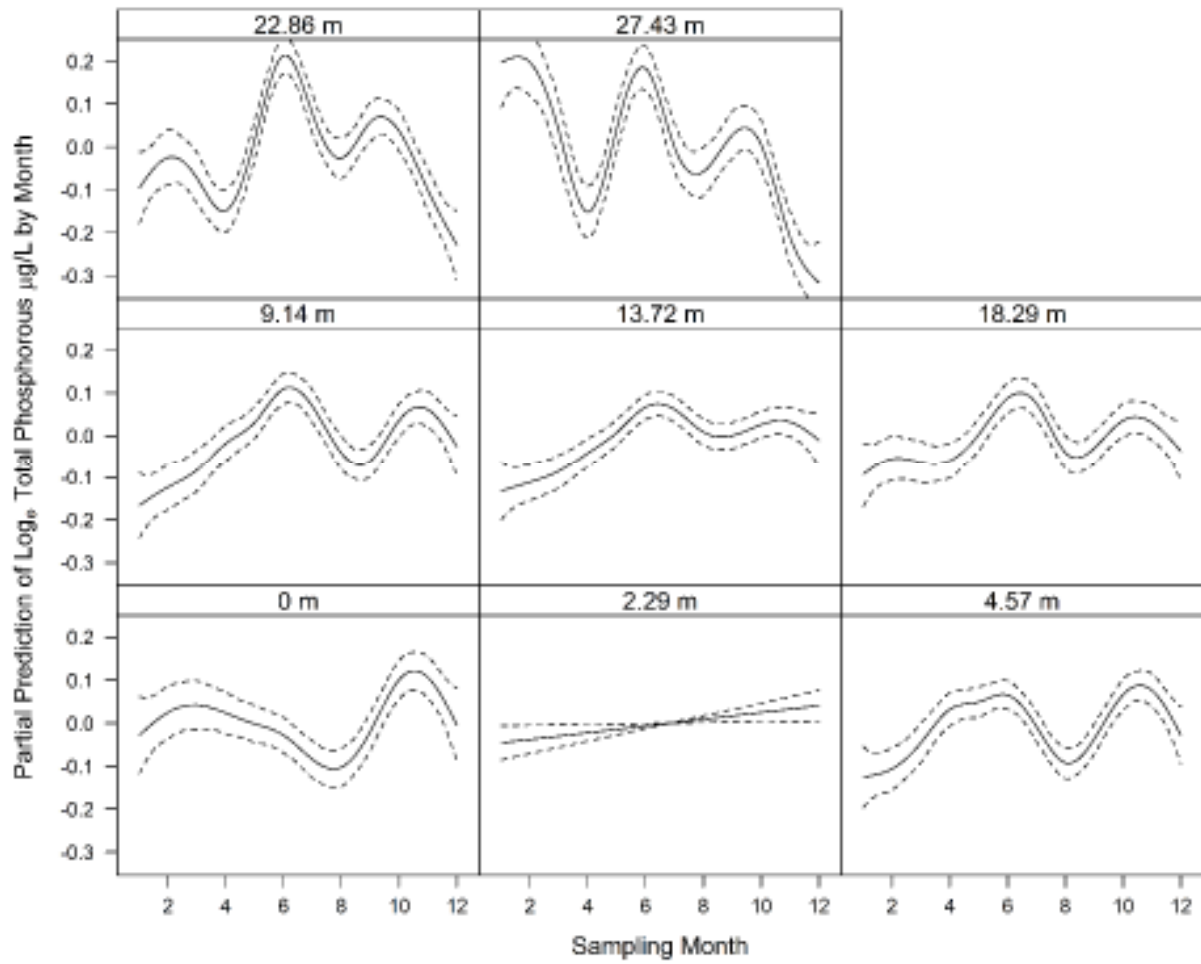
681



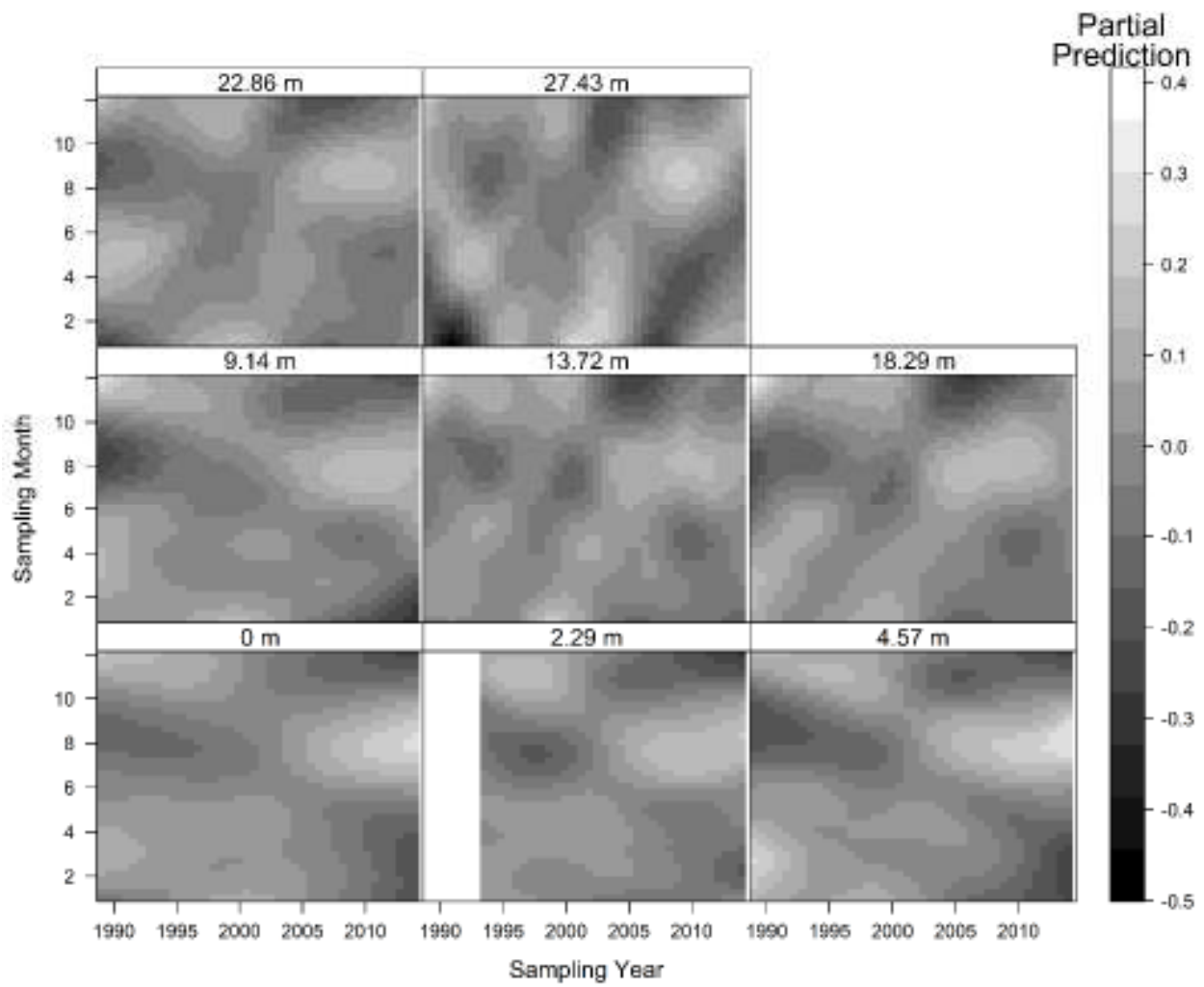
682



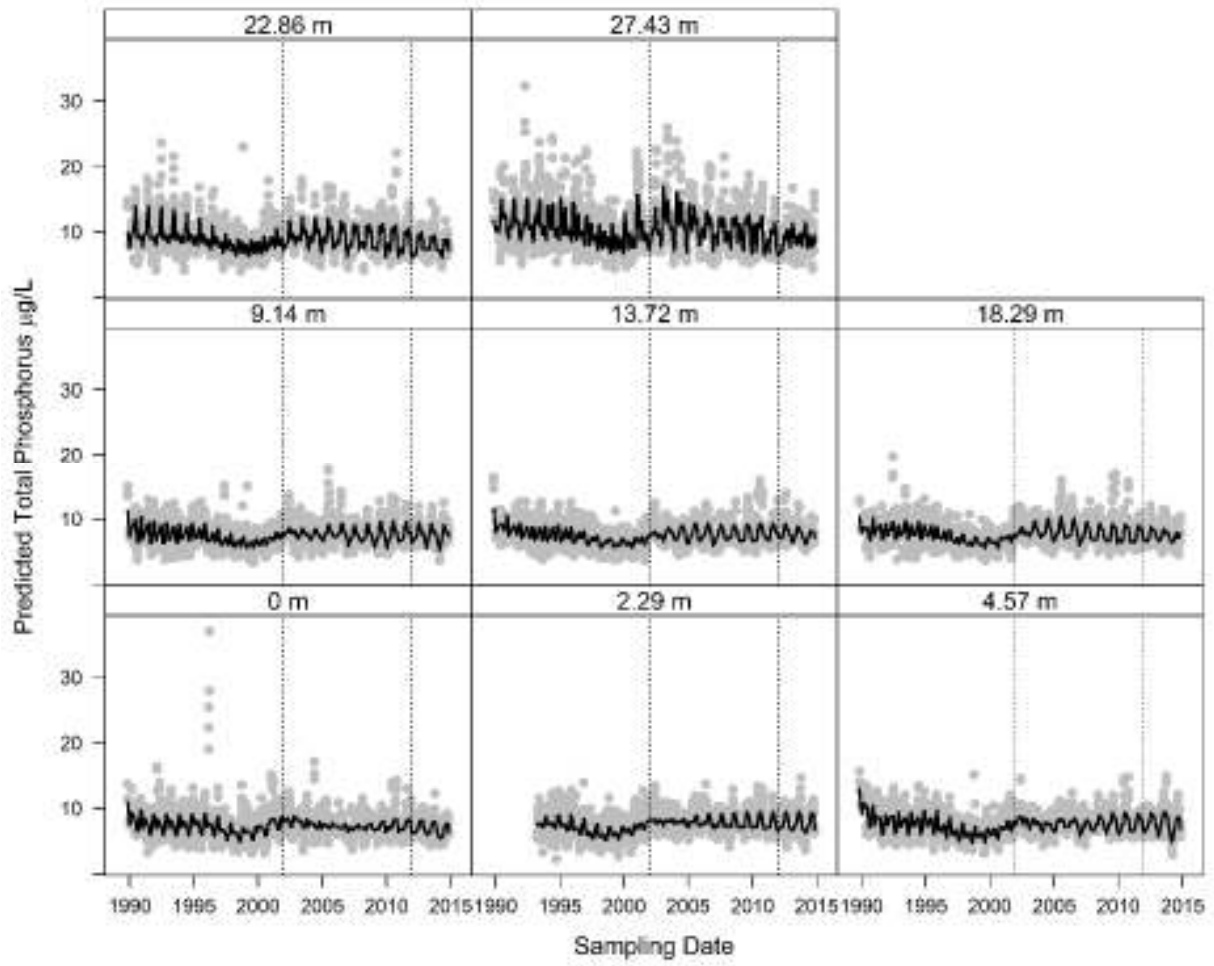
683



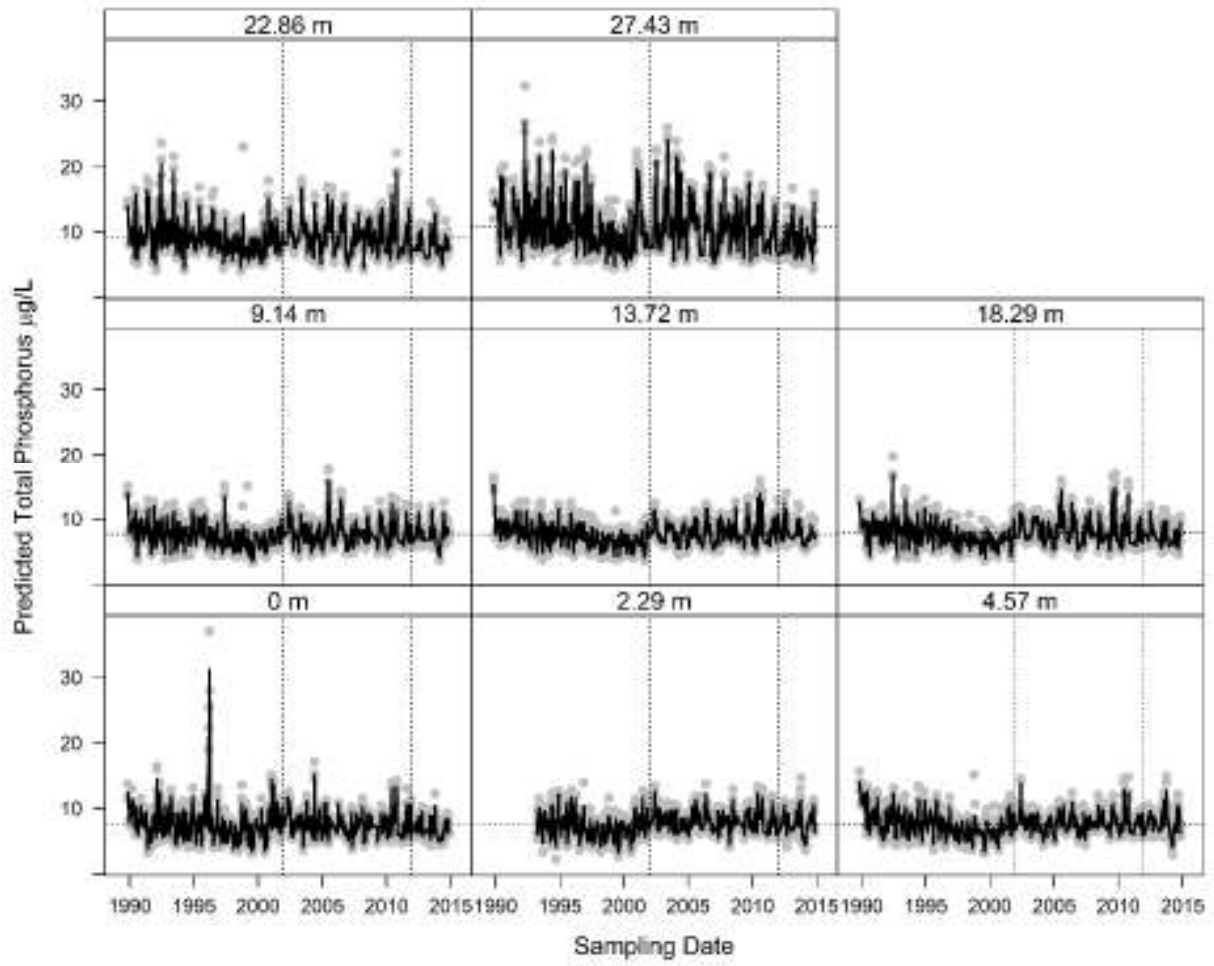
684



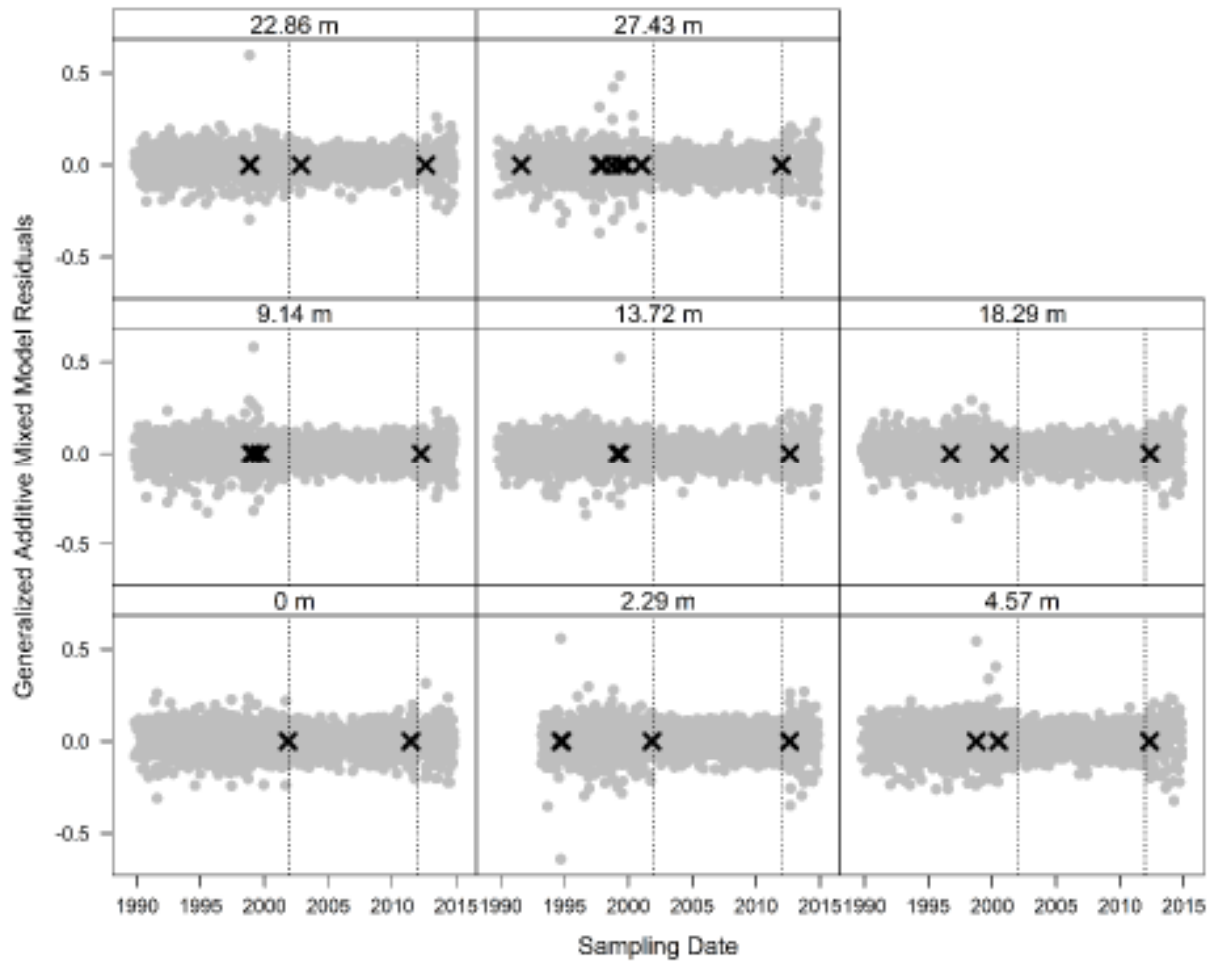
685



686



687



688