Referee #1

Referee general comment:

Thank you very much for the opportunity to review this manuscript describing the internal (physical and biological) and external factors leading to inter-annual variability in the extent anoxia in Lake Mendota. This study uses a combination of three very different types of models to evaluate these various factors. I found this paper very interesting, very well written, and may be very useful to the scientific community. I applaud the authors in using this multi-model approach. However, I think two of the three models have serious flaws that need to be addressed prior to publication.

My main concern is that one of the main takeaways from this paper (internal productivity has very limited effect on interannual differences in anoxia) may not be true. It may be true that physical mixing drives the overall extent of anoxia (baseline), but I think it is too early to say interannual variability in productivity has little affect. I think two of the models need to be reevaluated prior to making those conclusions:

Referee comment:

GLM-AED2. GLM-AED2 simulated the annual progression of anoxia very well, and simulated the importance of stratification driving not only the average changes in DO depletion but also much of the interannual variability in DO associated with stratification. But the model did not capture the interannual variability in surface productivity that may drive the other interannual variability in DO. It clearly could not reproduce the interannual variability in AF. This model had an R2 of only 0.08 and a negative NSE. Part of the problem may be that the model is trying to simulate two very different lakes (one without spiny water fleas and one with them) - all with one set of coefficients (that may not even represent the lake in the first place). Without simulating the big biological change, I am not sure you can get there with this model.

Referee suggestion: Use GLM-AED2 to only simulate one of the periods, either prior to or after the change in biology. If this does not improve the overall ability to predict AF, then the phytoplankton parameters may have to be adjusted. Without being able to predict most of the variability in AF, I really don't see its use in this paper.

Author response:

We are very thankful for this comment by the reviewer, which gives us the chance to discuss the GLM-AED2 performance and hopefully improve the overall manuscript. We make two over-arching points here. The first is that a visible shift in AF occurred in 2010 (Fig. 10b), and this may be explained by changes in the foodweb that affect primary production and organic matter cycling. We have no conclusive evidence of the cause, but the shift is coincident with the invasion of the predacious zooplankton, *Bythotrephes*. We discuss this and have added it to the abstract. The second is that our model reproduces well the ecosystem dynamics prior to 2010, and as the reviewer suggests, the lake is likely in different states, separated by the shift that occurs in 2010. Further, we also acknowledge that the GPP is actually an important driver of the variability in summer anoxia (rel. importance 15 %). The main text was revised accordingly in the abstract and in the discussion:

> L27-32: The summer heat budget, the timing of thermal stratification, and the gross primary production in the epilimnion were the most important predictors of the spatial and temporal extent of summer anoxia periods in Lake Mendota. Inter-annual variability in anoxia was largely driven by physical factors: earlier onset of thermal stratification in combination with a higher vertical stability strongly affected the duration and spatial extent of summer anoxia. A step change upward in summer anoxia in 2010 was unexplained by the GLM-AED2 model. Although the cause remains unknown, possible factors include invasion by the predacious zooplankton, *Bythothrephes longiman*us.

L441-443: We also acknowledge that a step change in the Anoxic Factor occurred in 2010 and was unexplained by our model. Although the cause remains unknown, the timing was coincident with large increases in the invasive zooplankton, *Bythotrephes* (Walsh et al., 2017).

To the point about the model not capturing variability in surface productivity, we added a new figure to the Supplement: Figure A8 which shows the time-series comparison between observed and modeled DOC concentrations. Here, you can see that the model replicated the overall dynamics of DOC concentrations in three different depths over time, which highlights its ability to replicate net aquatic production in the surface layer and its contribution to dissolved organic matter. Fig. 9a of the main text showed that the model overestimates surface dissolved oxygen concentration. This overestimation must have a concomitant increase in organic matter as a consequence of photosynthesis, and in this case is particulate organic matter (POM). Considering our proxy for phytoplankton biomass is well predicted (Fig. 5), this suggests our over-estimate of primary production results in increase in POM that is exported from the epilimnion to the hypolimnion. Unfortunately, we do not have observed POM to calibrate this part of the model, but we feel it is likely that our model has overestimated the contribution of primary production to hypolimnetic organic matter and subsequent oxygen depletion. This underlies our conclusion that primary production may be less important to inter-annual variability than physical factors. We added these sentences to the main text to state this:

L481-487: Although the model replicated well the long-term DOC dynamics (Supplement Figure A8), it also overestimated surface layer dissolved oxygen concentrations compared to the observed data. This overestimation must have a concomitant increase in organic matter as a consequence of photosynthesis, and in this case in POC. Considering our proxy for the dynamics of phytoplankton biomass is reasonably well predicted (Fig. 5), this suggests our over-estimate of primary production results in increase in POC that is exported from the epilimnion to the hypolimnion. Unfortunately, we do not have observed POC to calibrate this part of the model, but we feel it is likely that our model has overestimated the contribution of primary production to hypolimnetic organic matter and subsequent oxygen depletion.

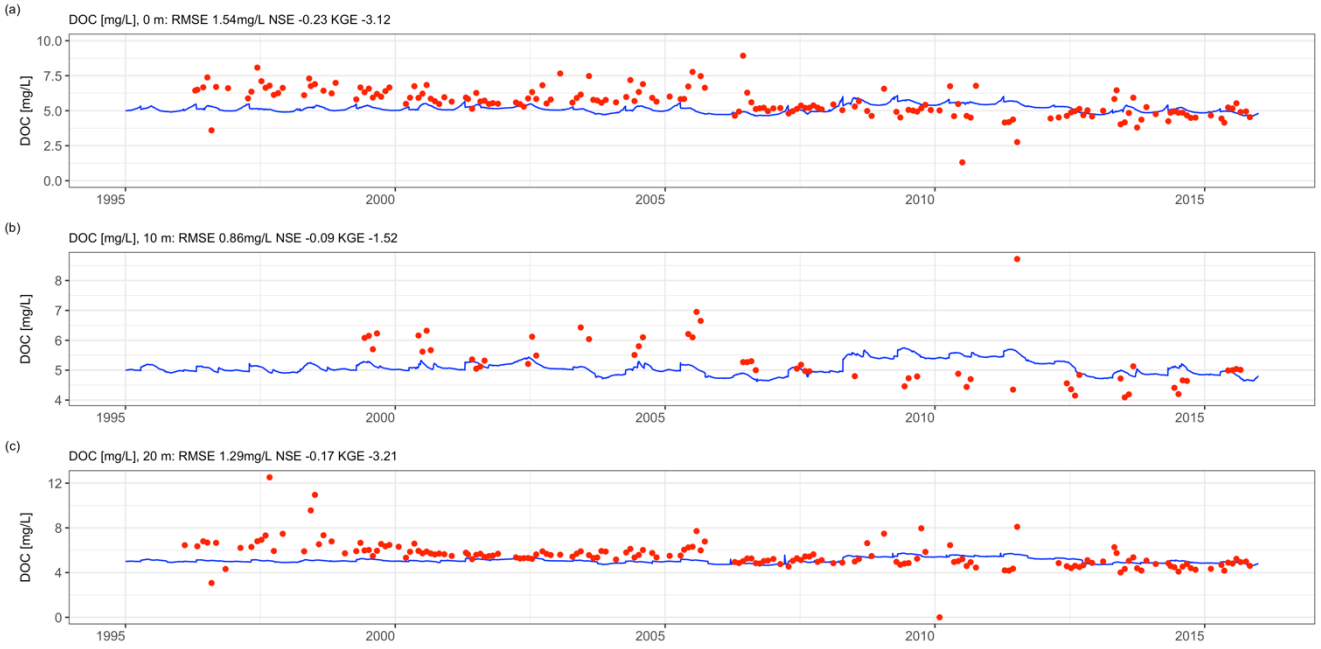


Figure A8 Time-series comparison between observed (red dots) and modeled dissolved organic carbon concentrations (blue lines). The fit criteria root-mean square error (RMSE), Nash-Sutcliffe coefficient of efficiency (NSE) and Kling-Gupta coefficient of efficiency (KGE).

Regarding the capture of interannual anoxia dynamics: Yes, it seems there was a shift in the ecosystem happening beginning in 2010 with higher annual Anoxic Factors. We changed Figure 10 to also show the comparison between simulated Anoxic Factor and the observed data for the periods pre-2010 and post-2010. We also recalculated goodness of fit separately for the two time periods. For the total time period (Fig 10b, 1992-2015 when observed data was available) the model achieved an RMSE of 7.12 d, NSE of -0.22, KGE of 0.26 and r of 0.28 showing that on average it was a week off in replicating the Anoxic Factor, but the KGE and r values proved that the general dynamics and interannual variability could be replicated. When comparing with the pre-2010 period (Fig 10c), the model achieved an RMSE of 6.79 d, an NSE of -0.25, an KGE of 0.44 and r of 0.45, which highlights the model's ability to replicate anoxia dynamics in this period (please note that the model was calibrated for the period 2005-2015 which proves, at least in our opinion, the success of the calibration if there indeed was an ecosystem shift). When comparing with the post-2010 period (Fig 10d), the model achieved an RMSE of 8.04 d, an NSE of -31.99, an KGE of 0.21 and r of 0.62. Here, the model is biased as the observed Anoxic Factor is higher in all years except 2013. Still, the interannual variability expressed by the correlation coefficient r was captured very well by the model. The p-value for the pre-2010 period of the correlation coefficient was p=0.0591. For the post-2010 period, the p-value = 0.19, reducing our confidence in the model for this shorter time period. The visual inspection of these plots (10c and 10d) highlights that they represent different ecosystem states, as there is step-change in the Anoxic Factor starting in 2010.

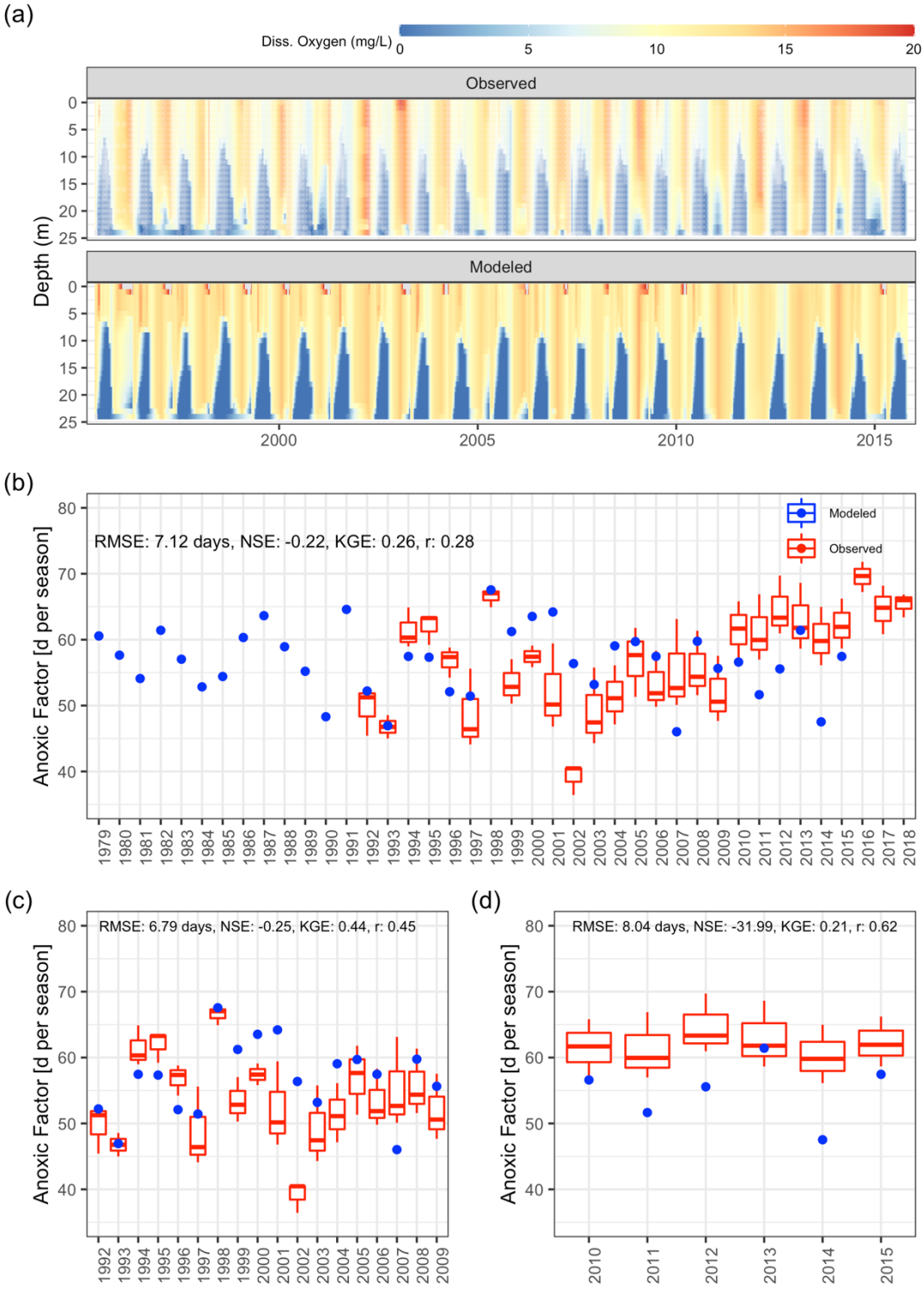

(a)
Diss. Oxygen (mg/L)
0
5
10
15
20
Observed
Modeled
Depth (m)
0
5
10
15
20
25
2000
2005
2010
2015
(b)
Anoxic Factor [d per season]
RMSE: 7.12 days, NSE: -0.22, KGE: 0.26, r: 0.28
Modeled
Observed
80
70
60
50
40
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
(c)
Anoxic Factor [d per season]
RMSE: 6.79 days, NSE: -0.25, KGE: 0.44, r: 0.45
80
70
60
50
40
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
(d)
Anoxic Factor [d per season]
RMSE: 8.04 days, NSE: -31.99, KGE: 0.21, r: 0.62
80
70
60
50
40
2010
2011
2012
2013
2014
2015

Figure 1 Comparison of observed to modeled dissolved oxygen concentrations and ecosystem response. A **Contour plot of observed (upper figure, white dots mark sample events) and simulated dissolved oxygen concentrations.** B **Comparison of simulated Anoxic Factor (red dots) against interpolated range of Anoxic Factor derived from observed data (box-whisker plots) over the period 1979 to 2018.** C **Comparison of simulated Anoxic Factor (red dots) against interpolated range of Anoxic Factor derived from observed data (box-whisker plots) over the period 1992 to 2009.** B **Comparison of simulated Anoxic Factor (red dots) against interpolated range of Anoxic Factor derived from observed data (box-whisker plots) over the period 2010 to 2015.**

Therefore, we focused our regression analysis on the pre-2010 period. First, we inspected if the distributions of the observed and modeled Anoxic Factors were similar by investigating the null hypothesis that they are identical populations as determined by the Wilcoxon test (see attached figure below). This test achieved a non-significant p-value of 0.13, indicating strong overlap in populations, and therefore are comparable. We added this figure as Figure A9 to the Supplement A of the manuscript. Also, a similar comparison of the Anoxic Factors for the post-2010 period revealed that observed and modeled distributions were significantly different with a p-value of 0.032. This effectively highlights that we can talk about "two different lakes here". We added these sentences to the main text in the results and in the discussion:

L412-421: The simulated Anoxic Factor over the total time period averaged 56.7 ± 5.2 days with an RMSE of 7 days, an NSE of -0.22, and an KGE of 0.26 (correlation coefficient r = 0.28). The model's underestimation of the recent positive trend of Anoxic Factors starting in 2010 was investigated by quantifying the fits during two periods: 1992-2009 (Figure 10C) and 2010-2005 (Figure 10D). In the pre-2010 period (1992-2009), the model achieved an RMSE of 6.79 days, an NSE of -0.25, an KGE of 0.44 and r of 0.45 for Anoxic Factor predictions. In the post-2010 period (2010-2015), the model achieved an RMSE of 8.04 days, an NSE of -31.99, an KGE of 0.21 and r of 0.62. A subsequent Wilcoxon signed-rank test highlighted, that the observed average and modelled Anoxic Factors from the pre-2010 period showed no significant differences between the two distributions, suggesting they belong to the same population (p-value = 0.13, Supplement Figure A9), whereas the distributions of observed mean Anoxic Factors and modeled ones after 2010 were significantly different (p-value = 0.032, Supplement Figure A9).

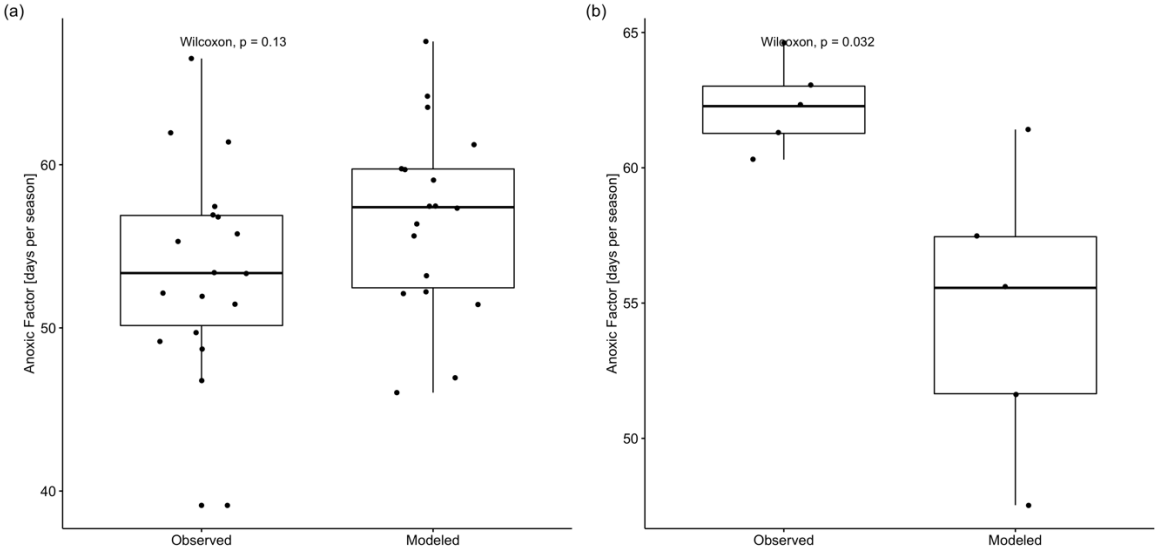


Figure A9 Box-whisker plots of observed to modeled Anoxic Factor for (a) the period 1992-2009 and (b) for the period 2010-2015.

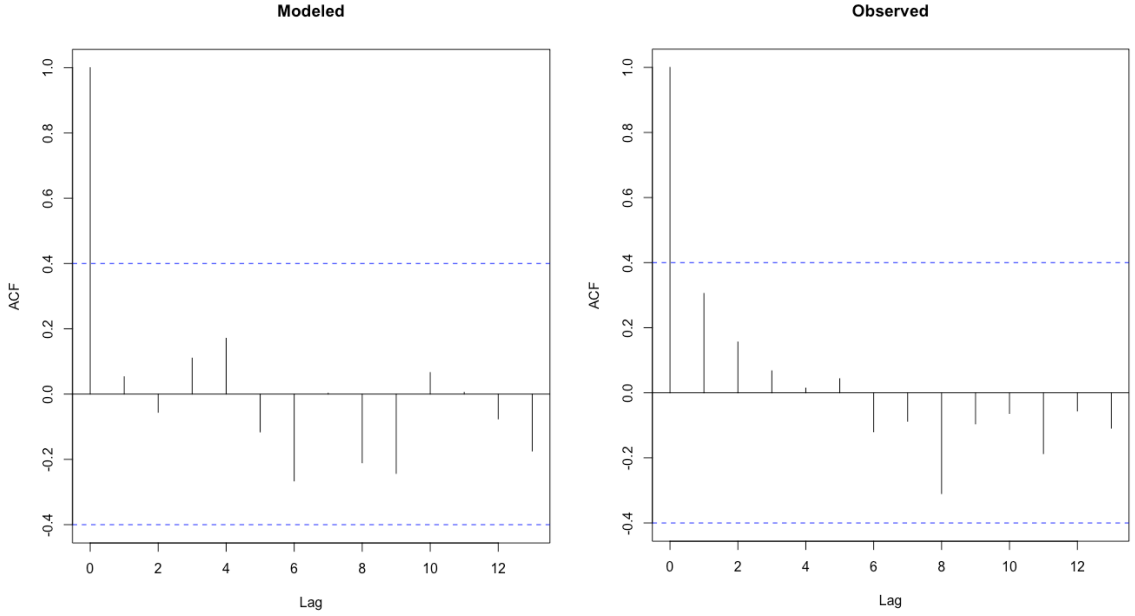
We discussed novel insights into these two distinct periods by expanding this paragraph

L498-518: The model replicated the maximum anoxia event in 1998 but struggled to replicate the minimum in 2002. The discrepancies of 5-10 days between the simulated and observed range of the Anoxic Factor beginning in 2010 are related to an increased spatial as well as temporal extent of summer anoxia (Supplement Figure A10), which was not captured by the model. This was highlighted by the statistical analysis of the pre-2010 (1992-2009)

and post-2010 (2010-2015) Anoxic Factors. Prior to 2010, there were no significant differences between observed and modeled distributions (p=0.13); whereas, after 2010, the observed distribution was significantly higher than the modeled distribution (p=0.032) (Supplement Figure A9). For simplicity and due to limitations in Lake Mendota monitoring data post-2010, we focused the regression analysis of the Anoxic Factor in this study only on the pre-2010 period.

The change in Anoxic Factor post-2010 may be due to an ecosystem shift in Lake Mendota that began in 2009, when the invasive spiny water flea (*Bythothrephes longimanus*) was detected in surprisingly high densities in the lake (Walsh et al., 2016b, 2018). Spiny water flea effectively became the dominant *Daphnia* grazer, causing historically low *Daphnia* biomass in 2010, 2014 and 2015 (Walsh et al., 2016a) and reducing water clarity. The spiny water flea may have increased organic matter supply to the hypolimnion by grazing down certain phytoplankton. Mendota's Daphnia population historically consisted of *Daphnia pulicaria* and the smaller-bodied *Daphnia galeata mendotae,* who compete differently with spiny water flea. *D. mendotae* biomass increased in spring after the spiny water flea invasion (Walsh et al., 2017), grazing on phytoplankton and probably accelerating organic matter mineralization before stratification onset. This could be one potential cause that contributed to the increase in hypolimnetic oxygen depletion after 2010. Our GLM-AED2 model could not replicate this food web change, and subsequent shift in anoxia dynamics, due to limitations of the numerical model, i.e., GLM-AED2 had constant ecological parameters over the entire modeling period and did not have zooplankton dynamics instantiated. We envision future monitoring and modeling studies that focus entirely on ecosystem differences and shifts between the pre-2010 and post-2010 periods of Lake Mendota.

Further, by analyzing the autocorrelation function (ACF) of the observed mean Anoxic Factors and the modeled ones (see figure below), we concluded that there is no autocorrelation between annual Anoxic Factors. It may be the case that the interannual variation in the Anoxic Factor (investigated by ACF) is effectively random, which does not mean that the Anoxic Factor is necessarily random, but that the variation in external drivers may be random. Still, our model's simulated Anoxic Factors are from the same distribution as the observed mean values highlighting the model's ability to capture the overall distribution of anoxia. Further, the fit metrics (highlighted in revised Figure 10) show that the model can capture inter-annual variability significantly prior to 2010, even if the average value is off by about a week.



We therefore followed the reviewer's suggestion to only include the pre-2010 period for the regression analysis, which is discussed in the next comment and response block:

L314-317: Only model output and model driver data from the period 1980-2009 were used in the regression analysis. The first year, 1979, was dropped from the investigations due to a lack of prior winter information. The years 2010-2015 were dropped due to an apparent ecosystem shift (see Section '3.4 Oxygen Dynamics').

Referee comment:

Regression model. I think there are four flaws in the approach used here: 1. Not including loading and in-lake variables that would potentially describe interannual variability in productivity. 2) Including modeling results in a regression analysis. Given that the model does not simulate AF, it appears that using modeling results in the regression may just add noise to the regression or reinforce parameters that are in the model. 3) Using one correlation and one regression to simulate two very different types of lakes, and 4) Using way too many variables in a single multiple regression equation. Even though it appears based on stepwise regression all of the variables are significant, I think it is way over parameterized. Several studies have shown that with regressions using very few observations, many variables can look significant – with each variable coming in to describe one or a few unique observations. A good rule of thumb is to keep only 1 variable in a multiple regression for each 8-10 observations. So for this regression with 37 (and actually only 28 monitored years) observations, there should only be maybe 3 independent variables.

Referee suggestion: 1) include variables like actual loading rather than concentrations, include variables that describe inlake productivity (total phosphorus, chlorophyll, Secchi). I am not sure what GPP actually represents. If GPP does describe the changes in chlorophyll, it should be stated. I also do not think it is a good idea to include things describing DO (like maximum height of anoxia) when you are trying to predict AF (this can get to circular reasoning) 2) Only use the 28 actual observations in the correlations and regressions. 3) Look at the correlations for each part of the record (different biological conditions) separately. 4) Stick to correlations and not use regressions. Or if you do look at regressions start simple and add variables only significant when you consider the change in AIC.

Author response:

Thank you for your very thoughtful explanation of the regression analysis' flaws and your very helpful suggestions how to overcome these.

1) We changed the inflow variables, total phosphorus inflow concentration and total nitrogen inflow concentration (both in g per m2), to total phosphorus inflow loading and total nitrogen inflow loading (both now in g per d per m2 of lake area). These loading variables were included in the model to assess the importance of external hydrological drivers for the extent of anoxia. To capture in-lake productivity variables, our regression included the cumulative gross primary production in the surface and bottom lake that represent the total sum of photosynthesis, hence expressed as carbon uptake, of each functional phytoplankton group, and scales directly with in-lake Chl-a concentrations. Further, our regression also includes the temporal change of dissolved as well as particulate organic carbon in the bottom layer from stratification onset to fall mixing onset. To make it clearer what GPP represents, we added this sentence to the main text:

L300-308: Here, GPP represents the sum of all functional phytoplankton group's photosynthesis rates parameterized as the total carbon uptake:

$$f_{uptake}^{PHY_C} = R_{growth}^{PHY}(1 - k_{pr}^{PHY})\,\phi_{temp}^{PHY}(T)\,\phi_{stress}^{PHY}(X)\,min\{\phi_{light}^{PHY}(I)\,\phi_{N}^{PHY}(NO_3, NH_4 PHY_N)\,\phi_{P}^{PHY}(PO_4, PHY_P)\,\phi_{Si}^{PHY}(Rsi)\}[PHY] \quad (7)$$

where the carbon uptake f_{uptake}^{PHYc} of an individual group *PHY* depends on the growth rate R_{growth}^{PHY}, the photorespiratory loss $(1 - k_{pr}^{PHY})$, temperature scaling $\phi_{temp}^{PHY}(T)$, metabolic stress $\phi_{stress}^{PHY}(X)$, and a minimum function taking into account limitations by light $\phi_{light}^{PHY}(I)$, nitrogen $\phi_{N}^{PHY}(NO_3, NH_4 PHY_N)$, phosphorus $\phi_{P}^{PHY}(PO_4, PHY_P)$ and silica $\phi_{Si}^{PHY}(Rsi)\}$ (Hipsey et al., 2017; adapted from Hipsey and Hamilton, 2008). As the GPP is the main model output variable for phytoplankton dynamics, it scales directly with biomass and Chl-a concentrations.

Following the reviewer's suggestion, we removed the maximum height of anoxia in the regression analysis. The variable was removed from all paragraphs (2.3.4 Regression Model, Table 1)

2) For the regression we only used modeled results and no actual observed data. This was done to identify internal connections in the numerical model and its mathematical equations. Similar analyses of modeled output and model driver data were done in Farrell et al., 2020; Snortheim et al., 2017; Ward et al., 2020. We added these sentences to the Methods section:

   L279-282: All candidate predictors were either modeled output or boundary data for the model. This enabled the regression analysis to identify internal connections in the numerical model itself (similar analyses of modeled output and driver data were done in Snortheim et al., 2017; Ward et al., 2020; Weng et al., 2020).

3) Following our reasoning in the first comment and response section, and the suggestions by the reviewer we revised our regression analysis by only using model data from 1980-2009. This excludes the first year as warm-up period and the post-2010 period due to different ecosystem conditions (probably spiny water flea invasion). We added additional discussions regarding the ecosystem shift (see response to first comment).

4) Following the suggestion of the reviewer, we re-did the regression analysis with 21 candidate predictors using model output and model drivers from 1980-2009 (we removed the anoxia height from the sediment) using the Boruta algorithm (random forest classifier). This analysis identified 10 variables as important. Subsequently, we did a step-wise analysis of the AIC of each model. This resulted in the identification of seven predictors: HBR ratio during spring, HBR ratio during summer, Birgean Work in spring, epilimnetic GPP, Schmidt Stability in summer, Birgean Work in summer, and onset date of stratification. The AICs of each model with any of these variables removed did not result in significant changes (this table was added to the manuscript as Table A3):

Table A3 Step-wise model-selection by removing predictors of the multiple linear regression model using seven predictors.

Predictor	AIC
HBR ratio during spring *(Spring.HBR)*	-61.820
HBR ratio during summer *(Summer.HBR)*	-60.529
Birgean Work during spring *(Spring.Birgean)*	-60.189
Gross primary production in the epilimnion (*Epi.GPP*)	-58.952
Schmidt Stability during summer *(Summer.St)*	-51.829
Birgean Work during summer *(Summer.B)*	-50.848
Onset of stratification *(Onset.Strat)*	-42.900

We reduced the final model to only three predictors (as the reviewer suggested) including onset date of stratification, Schmidt Stability in summer (as the AIC was similar to Birgean but the concept of Schmidt Stability is more generally known) and epilimnetic GPP. The text in "2.3.4 Regression Model" was accordingly changed to:

L321-330: This multiple linear regression model to predict Anoxic Factor included seven variables: HBR ratio during spring, HBR ratio during summer, Birgean Work in spring, epilimnetic GPP, Schmidt Stability in summer, Birgean Work in summer, and onset date of stratification. We reduced the complexity of the final multiple linear regression model to only three predictors of Anoxic Factor: onset date of stratification, Schmidt Stability in summer, and epilimnetic GPP. Schmidt Stability was included instead of Birgean Work as the resulting AIC of both models were similar, but the concept of Schmidt Stability is more commonly used in the limnological research community (Supplement Table A3). The final multiple linear regression model was configured as (scaled predictors, adjusted R^2 = 0.84, $p < 0.001$ Supplement Table A4).

$$\hat{y} = 0.24 Epi.GPP + 0.54 Summer.St - 0.46 Onset.Strat - 5.44 * 10^{-17} + \hat{\epsilon}, \quad (8)$$

where $\hat{\epsilon}\, N(0,38^2)$.

The results text in "3.5 Regression Model" was changed to:

L423-430: We included in total 3 predictors in our final multiple linear regression which were deemed important by the Boruta algorithm and stepwise linear model investigations using AIC for the period 1980-2009: Schmidt Stability during summer, the onset date of stratification, and gross primary production in the epilimnion (Supplement Table A4).

The linear model showed a good agreement between simulated and predicted Anoxic Factor (Figure 11 A, Supplement Table A4). The Anoxic Factor was positively correlated to the summer Schmidt Stability (r = 0.72, Figure 11 B) and the gross primary production in the epilimnion (r = 0.48). It was negatively correlated to the onset of stratification (r = -0.78, Figure 11 B).

We changed Supplement Table A3 (formerly A2) and Figure 11 accordingly:

Table A3 Most parsimonious multiple linear regression model (adjusted R^2 = 0.84, $p < 0.001$) explaining the summer Anoxic Factor.

	Estimate	Std. Error	t value	Pr(>|t|)	Rel. importance [%]
Intercept	-1.04e-15	5.70e-2	0.00	1.00	
Schmidt Stability during summer *(Summer.St)*	5.386e-1	7.920e-2	6.800	3.23e-7	43
Onset of stratification *(Onset.Strat)*	-4.581-1	9.006e-2	-5.086	2.68e-5	42
Gross primary production in the epilimnion (*Epi.GPP*)	2.436e-1	8.327e-2	2.926	0.00704	15

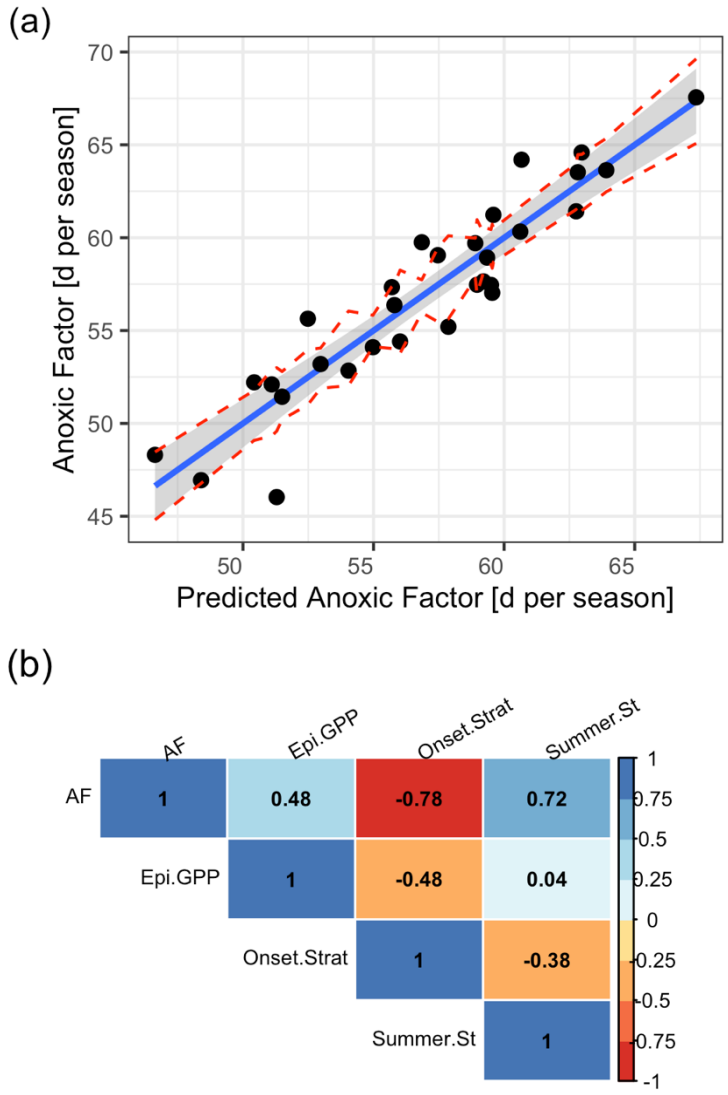


Figure 2 Predicted against simulated summer Anoxic Factor. A Linear model with a prediction which was done using a multiple linear regression model of the form: $\hat{y} = 0.24Epi.GPP + 0.54Summer.St - 0.46Onset.Strat - 5.44 * 10^{-17} + \hat{\epsilon}$, where $\hat{\epsilon} \sim N(0, 38^2)$. The red lines represent confidence intervals. B Correlogram of the input data using Pearson correlation coefficients

We changed the following sentences in the main text to reflect these changes:

L434-438: The Schmidt Stability during summer (rel. importance of 43 %) as well as the timing of stratification (rel. importance of 42 %) all influence Anoxic Factor, and are all driven mainly by atmospheric drivers and heat convection

throughout the water column. The most important predictor of Anoxic Factor directly related to biological processes is gross primary production in the epilimnion (rel. importance of 15 %), Supplement Table A4).

L596-599: Physical metrics – summer Schmidt Stability and onset date of stratification – were the most important predictors driving the summer Anoxic Factor. Although the gross primary production was still influential in affecting year-to-year variability of hypolimnetic anoxia, biological control over the Anoxic Factor was limited in our study period.

Referee comment:

My other main concern is that the deductive model seems to say that it is the inlake productivity that is driving the interannual variability in AF, and the other models seem to be saying it is driven by physics and sediment oxygen demand. Maybe with further analysis the models will come to more similar conclusions. If I am wrong with this interpretation, it should be explained better.

Author response:

The deductive model itself can only determine between two sources of depletion, either a volumetric one or an area sink. It cannot distinguish between biological or physical drivers of these depletion causes. Although the deductive model states that the volumetric sink is higher than the area sink, this is only of importance for the in-lake biological drivers (as the area sink depends on in-situ biogeochemical conditions). In the manuscript we state that the anoxia variability over a summer season is mainly driven by changes in the physical drivers, whereas we acknowledge that oxygen depletion itself (as shown in the regression model) is a function of biological and chemical activity. The deductive model itself does not consider any physical drivers, even diffusion is neglected. We added these lines to the main text to clarify our message:

L529-532: We note that the simple deductive model itself can only differentiate between two sources of depletion and neglects any physical transport drivers of oxygen, e.g., diffusion. Therefore, the results of the deductive model only add direct information to the actual depletion process of dissolved oxygen, but not of the dominant drivers.

Referee comment:

1. Line-125. Very little information is given on the actual loading. Can these estimates be compared with others?

Author response:

Thank you. We compared our loadings with literature values, especially regarding phosphorus. Previous estimates range from about 15-67 t of total phosphorus (TP) per year (Kara 2011). Our estimates are at the higher end of this range. There is a concern that previous estimates did not fully account for loads of adsorbed phosphorus (hence, phosphate bound on sediment), because of the importance of extreme storm events on particulate loads (Carpenter 2017). To accommodate for a potential underestimation of TP loads, we added to the inflow boundary condition the adsorbed phosphate variables, which was set roughly equal in magnitude to non-adsorbed phosphorus. This puts our estimates of total P load near the upper range of previous estimates. Bennett (1999) estimated the long-term average annual TP input with 34 t P. Our Yahara inflow had an average annual TP load of about 25.3 t/y and ranged between 2.69 to 73.09 t/y over the period 1979-2015. Due to the use of a hydrological model, our inflows accounted for a closed water balance and included near-lake groundwater/spring inflows. Our average annual load of 25.3 t/y is slightly higher than the loadings by of Lathrop (2009). We added these lines to the main text:

L136-144: To provide information regarding adsorbed soluble reactive phosphate, we doubled measured total phosphorus (TP) concentrations and applied specific ratios to individual phosphorus forms (Farrell et al., 2020; Snortheim et al., 2017; Weng et al., 2020). This put our estimates of TP near the upper range of previous load

estimates. Bennett et al., (1999) estimated the long-term average annual TP load to be about 34 t, whereas our average annual TP load (with adsorbed phosphate) was about 50.6 t and ranged between 5.3 to 146.1 t (1979-2015). Our average annual TP load (without adsorbed phosphate) was about 25.3 t and ranged between 2.7 to 73.1 t (1979-2015), which is similar to previous estimates between 15 to 67 t (Kara et al., 2012). By doubling our TP by adding adsorbed phosphate, we accommodate a potential TP load underestimation due to the importance of extreme storm events on particulate loads (Carpenter et al., 2018).

Further, we checked our derived annual TP loadings using the Vollenweider model by assuming winter TP concentrations, TP_{lake}, of 140 ug/L, a residence time, RT, of 4 years, P retention, σ, of 0.7, and a mean depth, z_{mean}, of 12.8 m:

$$TP_{lake} = \frac{L}{z_{mean}\left(\frac{1}{RT} + \sigma\right)}$$

$$L = 0.14\ g/m3\ (12.8\ m\ (0.25\ y^{-1}\ +\ 0.7\ y^{-1})) = 1.70\ g/m2/y$$

By multiplying L with the lake area of Lake Mendota (approx. 39.61 km2), the Vollenweider model quantifies the annual load for steady-state conditions with 67 t/y, which is slightly above our average annual TP load (with adsorbed phosphate) of 50.6 t/y.

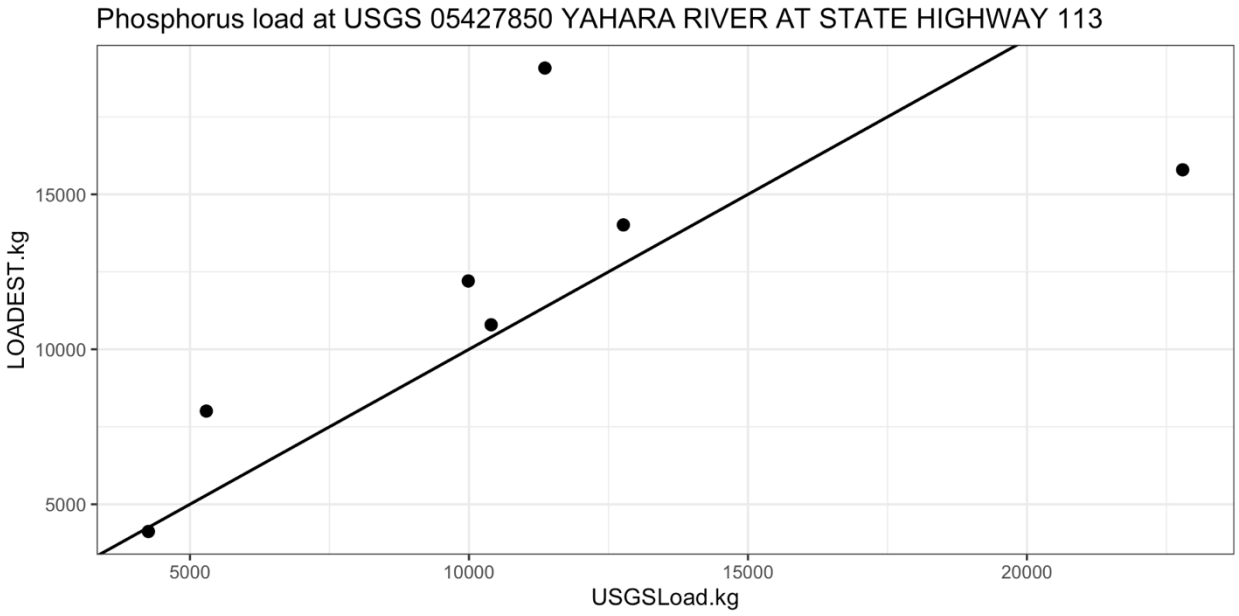
Referee comment:

2. Line 128 – It says here to look at Weng et al. 2020 for a description of the loading regression, but when I look at that paper, I don't see any more than they used a regression, with no statistics either for the monitored sites or the watershed modeling.

Author response:

Thank you for pointing this out. Yes, there are no previous publications describing the regression fit analysis. We described in the previous response that our TP loads were near the upper range of previous estimates due to our addition of adsorbed phosphate (due to extreme storm events and land erosion). For the regression analysis between discharges and nutrient concentrations, we used the state-of-the-art loadflex R-package (https://github.com/USGS-R/loadflex, Appling et al., 2015). As monitored TP estimates are rare, a comprehensive statistical analysis is challenging. The attached figure visualizes the fit for 8 years (2008-2015), which was satisfactory for most years. USGS monitoring began Oct 2008, so a comparison cannot be made for the entire year. Overall, our model tended to overestimate nutrient loads into the lake.

Year	LOADEST.kg	USGSLoad.kg
1996	10848	NA
1997	8827	NA
1998	14289	NA
1999	12268	NA
2000	15558	NA
2001	14384	NA
2002	10571	NA
2003	6703	NA
2004	15039	NA
2005	5139	NA
2006	9515	NA
2007	12412	NA
2008	25341	1478
2009	15793	22790
2010	19075	11361
2011	10792	10398
2012	4122	4255
2013	14013	12767
2014	12205	9988
2015	8007	5293



Referee comment:

3. Line 136 – You mention other data earlier years, who collected that?

Author response:

The additional data points were measured by Patricia Soranno for her thesis. We added that information in the sentence:

L149: The dissolved oxygen data set was complemented with historical measured dissolved oxygen data from 1992 to 1994.

And we acknowledged her in the Acknowledgement section "We are thankful for supplementary dissolved oxygen field data from 1992-1994 by Patricia Soranno." Her data does not officially belong to the NTL-LTER monitoring data set of Lake Mendota, but it gave us valuable early spring-summer information regarding oxygen dynamics.

Referee comment:

4. Line 159 – See comments above about mixing real observations with modeled data. 5. Line 190 – There are lots and lots of parameters in AED, how did you narrow it down to the ones to start with, you need to start somewhere?

Author response:

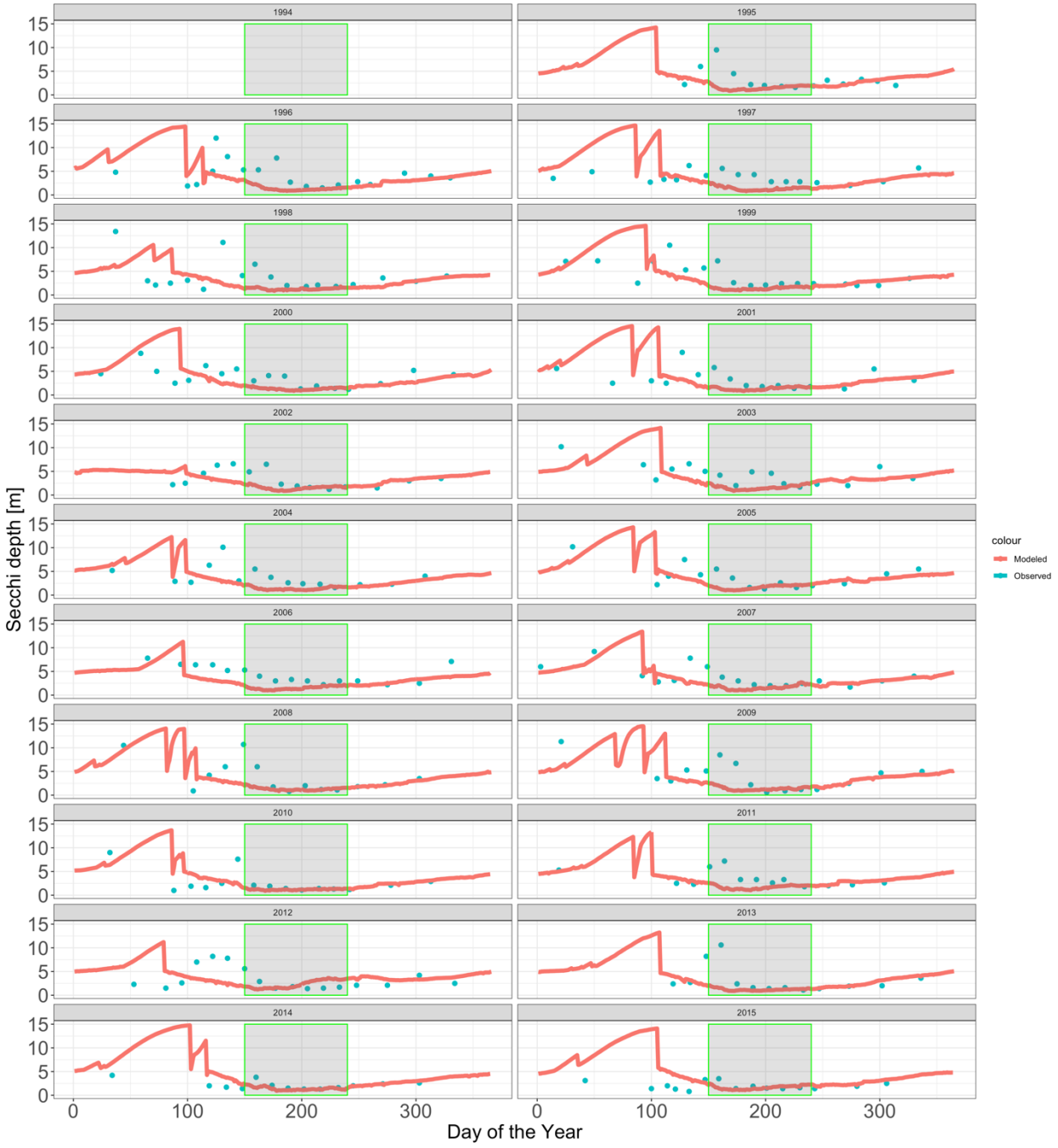
We used the Morris Sensitivity Method to identify crucial parameters for the calibration. For this analysis we included the main model parameters regarding sediment flux and in-water biogeochemical reactions, mainly, of the main nutrient modules: oxygen, carbon, silica, nitrogen and phosphate. For the initial values, we chose starting values either from the AED2 webpage (https://aed.see.uwa.edu.au/research/models/aed/modules.html, default values, or values inside the typical range) or from previous modeling work on Lake Mendota (see Snortheim et al. 2017). We added this sentence to the main text for clarification:

L231: Initial model parameter values were taken from default parameter values and ranges, as well as literature values (Hipsey et al., 2017; Snortheim et al., 2017).

Referee comment:

6. Line 215-Can you expect to capture interannual variability in productivity without having the phytoplankton simulate things specific to Lake Mendota?

Author response:
This is a good point, thank you for raising this. Although we did not calibrate the functional phytoplankton groups specifically to Lake Mendota, we still checked simulated Chl-a and Secchi depth values, as well as timings of phytoplankton bloom peaks. In general, the model did replicate the seasonal succession well. We've attached the following figure that compares the observed to modeled Secchi depths for the reviewer to inspect. The summer Secchi depths from the model are similar to the observed ones, highlighting that the ecosystem dynamics during anoxia are similar. The gray boxes highlight the time period from day of the year 150 to day of the year 180 (June to end of August): for the majority of years the model can replicate the Secchi depth dynamics of the June period., whereas generally it underestimates the initial summer Secchi depth.



Referee comment:

7. Line 260 – My bet is that anoxia does occur under the ice, but you can't get that from one measurement during the winter.

Author response:

Yes, measurements and the monitoring have shown that there is anoxia under the ice in Lake Mendota, but it varies a lot. We agree that we cannot determine the full anoxia extent under ice with only one or two measurements per season. We changed sentence the sentence accordingly to:

L276: We quantified the seasonal Anoxic Factor only for the summer season.

Referee comment:

8. Line 267 – Loads would be better than concentrations. Concentrations generally do not vary much from year to year. If you did really use loads, you should state that. But you should describe this better.

Author response:

We changed the inflow variables, total phosphorus inflow concentration and total nitrogen inflow concentration (both in g per m2), to total phosphorus inflow loading and total nitrogen inflow loading (both now in g per d per m2). These loading variables were included in the model to assess the importance of external hydrological drivers for the extent of anoxia. We changed the information in Table 1 accordingly:

Total phosphorus inflow loading	Winter/spring/summer of year n-1 and n	Extracted from driver data	g P per day per m^2
Total nitrogen inflow loading	Winter/spring/summer of year n-1 and n	Extracted from driver data	g N per day per m^2

Referee comment:

9. Line 273 – See comments above.

Author response:

Please see our response above.

Referee comment:

10. Line 278- Since Gross primary productivity (GPP) is your only in-lake productivity term, you should describe this in more detail. If this is directly related to chlorophyll, maybe this addresses some of my concerns.

Author response:

Thank for raising this point. GPP (gross primary productivity) in the lake is the cumulative photosynthesis, represented by cumulative carbon uptake per time step, of all functional phytoplankton groups. Therefore, it scales directly with the simulated Chl-a output. We clarified this in the main text:

L300-308: Here, GPP represents the sum of all functional phytoplankton group's photosynthesis rates parameterized as the total carbon uptake:

$$f_{uptake}^{PHY_C} = R_{growth}^{PHY}(1 - k_{pr}^{PHY})\, \phi_{temp}^{PHY}(T)\, \phi_{stress}^{PHY}(X)\, min\{\phi_{light}^{PHY}(I)\, \phi_N^{PHY}(NO_3, NH_4 PHY_N)\, \phi_P^{PHY}(PO_4, PHY_P)\, \phi_{Si}^{PHY}(Rsi)\}[PHY] \quad (7)$$

where the carbon uptake f_{uptake}^{PHYC} of an individual group *PHY* depends on the growth rate R_{growth}^{PHY}, the photorespiratory loss $(1 - k_{pr}^{PHY})$, temperature scaling $\phi_{temp}^{PHY}(T)$, metabolic stress $\phi_{stress}^{PHY}(X)$, and a minimum function taking into account limitations by light $\phi_{light}^{PHY}(I)$, nitrogen $\phi_{N}^{PHY}(NO_3, NH_4 PHY_N)$, phosphorus $\phi_{P}^{PHY}(PO_4, PHY_P)$ and silica $\phi_{Si}^{PHY}(Rsi)$} (Hipsey et al., 2017; adapted from Hipsey and Hamilton, 2008). As the GPP is the main model output variable for phytoplankton dynamics, it scales directly with biomass and Chl-a concentrations.

Referee comment:

11. Line 281 – Consider dropping this whole paragraph.

Author response:
Although we understand the reasoning behind dropping this paragraph as the same results could probably be achieved by either starting with a simple linear regression model and extending it, or by step-wise analysis of AIC, we decided to keep the Boruta algorithm and analysis in the manuscript. This method allows us to analyze 21 potential predictors in a comprehensive framework before reducing the final number of important predictors by step-wise analysis.

Referee comment:

12. Line 306 – The major conclusion of the deductive model says that water column respiration controls oxygen depletion, yet everything else seems to point to physics. Am I missing something here?? Is water column respiration the cause and physics drives the variability in this? More explanation is needed.

Author response:
Please see our response above regarding the limitations of the deductive models and its incapability to acknowledge physical drivers.

Referee comment:

13. Line 322 – Please give the stats for DO. This is really what matters in this paper, especially in the part that varies from year to year.

Author response:
We agree, thank you. We added these sentences to the main text:

L356-360: The simulated dissolved oxygen concentrations in the whole water column achieved an RMSE of 3.22 mg L-1, an NSE of 0.56, and an KGE of 0.77. Here, the average fits were better in the surface layer (RMSE of 2.77 mg L-1) compared to the bottom layers (RMSE of 3.31 mg L-1), whereas the temporal dynamics (as expressed in NSE and KGE) were slightly better in the bottom layer (an NSE of 0.64, KGE of 0.81) compared to the surface layer (NSE of -0.36, KGE of 0.47).

Further, the discussion of the oxygen fit has its own subparagraph in "3.4 Oxygen Dynamics" where we state that:

L405-409: Dissolved oxygen dynamics, including the spatial extent of oxygen depletion in the water column, and the timing of summer anoxia periods, were replicated by the GLM-AED model (Figure 9A-B); although the model overestimated spring and summer time surface oxygen concentrations due to a higher net ecosystem production. The depth-averaged fit criteria of dissolved oxygen concentrations were similar to a recent study from Farrell et al. (2020) in which the RMSE were 1.88 mg/L and 2.49 mg/L in the epilimnion and hypolimnion, respectively, of a GLM-AED model calibrated for Lake Mendota.

Referee comment:

14. Line 333 – Reorder this paragraph to put the peaks later when you talk about summer.

Author response:
We agree. We moved the sentence to a later paragraph and combined it with the description of the annual course of Schmidt Stability:

L395: Schmidt Stability peaked on average in August at approx. 720 J m-2 (Figure 6), followed by a peak in the Birgean Work at approx. 1250 J m-2.

Referee comment:

15. Line 345 – This paragraph could probably be deleted.

Author response:
As the main take-away message of our manuscript is related to physical drivers, we decided to keep this short paragraph describing the deep-water stagnancy in the manuscript. By comparing the additional energy demands of Lake Mendota with other similar sized lakes, the reader gets valuable information regarding the lake's energy budget, and potential conclusions to the anoxia drivers of similar lake systems.

Referee comment:

16. Line 370 – It says the model captured annual anoxia events. Yes it described the annual development, but right now it does not seem to have any interannual capabilities??

Author response:
We quantified the correlation coefficient for the Anoxic Factor with r = 0.28 (total period), r = 0.45 (pre-2010) and r = 0.62 (post-2010), see also Figure 10. Especially for the pre-2010 period the p-value of the correlation coefficient was p=0.0591, which was slightly above the significance level. Overall, this highlights the model's overall ability to predict interannual changes and dynamics.

Referee comment:

17. Line 374 – See above.

Author response:
See response above.

Referee comment:

18. Discussion – Need to tie all three model results together better. Right now two say physics and one says productivity.

Author response:
We disagree that two models point to physical drivers and one to biological ones. The deductive model distinguished the main oxygen consumption as either being a volumetric or an area sink term. This information was used to set up the sediment oxygen demand in the GLM-AED2 model. The results of the calibrated GLM-AED2 model were then used in a regression analysis to identify internal connections of the numerical model and its mathematical equations. This confirmed that in the process-based GLM-AED2 model three variables were important predictors of anoxia and its interannual variability. The deductive model itself does not consider any physical drivers (see responses above please).

Referee comment:

19. Line 394 – Although I completely agree with you, I am not sure where this comes from given the model results.

Author response:
The statement regarding that […] Biology matters but its interannual dynamics are not that influential […]" originates from the regression analysis. This analysis highlighted GPP as one of the most influential terms in projecting the variability of anoxia in Lake Mendota, but not as influential as physical variables (GPP only explained 15 % of the interannual variance of the Anoxic Factor).

Referee comment:

20. Line 420 – Again I agree with you, but other than one variable in seven in the regression, I don't know where this comes from. Need to describe this variables importance.

Author response:
As GPP is an ecosystem-scale metric that represents phytoplankton carbon uptake, net aquatic primary production as well as ecosystem respiration it surely highlights the biological control over Anoxic Factor, even if the regression deemed physical variables as more important.

Referee comment:

21. Line 425 – Maybe the lack of relations is due to using loading concentrations rather than actual loads. This is what I think the methods say.

Author response:
We changed the inflow parameters of total phosphorus and total nitrogen from concentrations to loadings and still the effect of anoxia is low. This is probably due to Lake Mendota's long water residence time of approx. 4 years.

Referee comment:

22. Line 433 – Is it loads or concentrations. If it is concentrations, that wouldn't surprise me at all. It is not the annual variations in concentrations that drive things, it is the difference in loads.

Author response:
See point above.

Referee comment:

23. Line 440 - This could be an important point, maybe there is so much oxygen consumption in the bottom, that it dwarfs any water column consumption. But this disagrees with findings of the other models.

Author response:
We discussed the sediment oxygen demand in the main text:

> L519-529: The simple deductive model established that the volumetric oxygen sink (i.e. water column oxygen demand) is consistently higher (on average about four times higher) than the sediment oxygen sink. The volumetric sink in lakes has been found to be strongly dependent on the trophic state of the lake, whereas the sediment sink is not (Rippey and McSorley, 2009). Eutrophic lakes tend to have high volume sinks that reach maxima of about 0.23 g m-3 d-1 (Rippey and McSorley, 2009) similar to the average volume sink of 0.16 g m-3 d-1 quantified by the deductive model for Lake Mendota. This finding is confirmed by the works of Conway (1972) who found that the high hypolimnetic oxygen demand of Lake Mendota was driven by algae decomposition, originating from the surface layer. Although eutrophic lakes tend to have a high sediment oxygen demand, the specific values can range from 0.3 g m-2 d-1 (Romero et al., 2004; Steinsberger et al., 2019) to extreme values of 80 g m-2 d-1 (Cross and Summerfelt, 1987), most studies measured or applied a value between 1 to 4 g m-2 d-1 (Mi et al., 2020; Veenstra

and Nolen, 1991). The sediment oxygen demand calculated by our deductive model of 0.04 g m-2 d-1 was closer to the average value of approx. 0.08 g m-2 d-1 measured by Rippey and McSorley (2009) on 32 lakes.

In our numerical model the sediment oxygen demand (SOD) is replicating the volumetric and area sink as explained in the "Methods" section. Also, the model SOD is represented over the whole vertical axis (sediment area per volume for each grid cell) instead of a stagnant bottom layer only near lake's bottom. The results of the deductive model did not confirm a very high SOD compared to other eutrophic lakes, see extreme values in Cross and Summerfelt (1987) of up to 80 g per m2 per d.

Referee comment:

24. Line 445. The apparent changes caused by the Spiny water flea may be totally confounding any correlations, regressions, and your GLM-AED2 modeling. You may have to stick to one of the periods to really describe the effects of physics vs internal. Or have two different models.

Author response:

See our initial response please. We described the Anoxic Factors for the pre-2010 and post-2010 periods in more details and focused our regression analysis only on the pre-2010 period.

Referee comment:

25. Line 472. Rather than implementing a different type of dynamic model, maybe better capturing change in productivity and clarity, will help in describing the physics.

Author response:

We agree that a better replication of changes in ecosystem-scale metrics like productivity or even water clarity would improve the simulations a lot. Still, water quality models are generally way overparameterized and have problems regarding equifinality. The occurrence of tiny water flea has proven that ecosystem changes will have strong effects on other ecosystem characteristics like anoxia. Therefore, even the best calibrated fixed water quality model will have problems replicating a dynamic ecosystem. Further, our monitoring campaigns do not capture important water quality variables on a high temporal scale, e.g. daily, which generates further uncertainty. Therefore, in our opinion an improvement of the hydrodynamic calculations for example by using a state-of-the-art turbulence closure scheme is the most applicable approach to improve the simulations in the near future.

Referee comment:

26. Line 481 – you didn't calibrate the biological parameters, so this should be rewrit-ten.

Author response:

We calibrated physical as well as chemical parameters in GLM-AED2 but did not modify the biological parameters of the functional phytoplankton blooms. As these functional variables represent multiple phytoplankton species, a direct calibration would potentially result in an over-calibration of the model for specific time periods, which we tried to avoid. We changed the sentence accordingly to:

L551: Our GLM-AED2 model overestimated spring phytoplankton biomass, which resulted in an overestimation of surface dissolved oxygen concentrations.

Referee comment:

27. Line 497 – Rather than thinking the deductive model is biased, maybe it is the only approach capturing the effects of the biology.

Author response:

555 Please see statement above regarding the limitations of the deductive model, and the lines that were revised to better formulate this in the main text.

Referee #2

Referee general comment:
This manuscript describes a one-dimensional model (i.e. GLM-AED2) study for Lake Mendota which analyzed its long-term changes of anoxia and the driving factors. As a major result, the model showed good performance in reproducing oxygen dynamics, especially the low oxygen concentration in the hypolimnion,in the lake and based on the statistical analysis, it suggested that the physical structure (e.g. Schmidt Stability, onset of stratification, water temperature in the hypolimnion) had a big influence on the spatial and temporal development of anoxia.
This is an interesting and important study, which could be considered for publication after a minor revision. Although there are quite a few studies analyzing hypolimnetic anoxia for inland waters, most of them draw their conclusion based on the short-term measurements and there is still a need to comprehensively illustrate this phenomenon and mechanisms behind its formation based on long-term database. Based on this prospective, this research fills in a research gap. In my opinion, this paper is well organized and its content, especially the discussion part will improve our understanding about anoxia and its future development under climate warming. Detailed comments are shown below.

Referee comment:
2.1 Study Site: It is better to show a topographic map of this lake, as well as the location for the water quality measurements.
Author response:
Thank you for this suggestion. We have added a new figure to the manuscript that shows the location and landuse overview of Lake Mendota, as well as the location of the measurement stations.

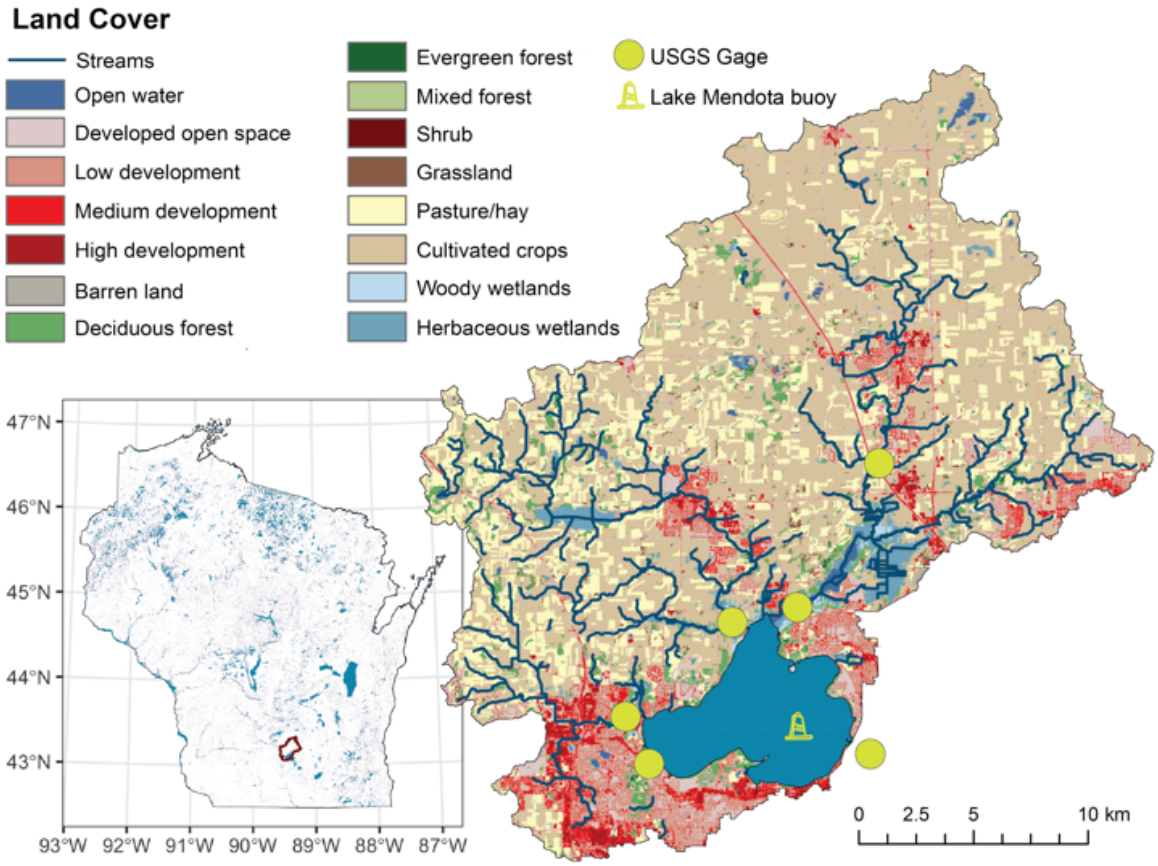


Figure 3 Location and overview map of Lake Mendota, Wisconsin, which is located in the Yahara River catchment in southern Wisconsin, USA. USGS gage stations for the PIHM-Lake model and the location of the Lake Mendota monitoring buoy are placed in the map. Land cover was obtained from the US National Land Cover database.

Referee comment:

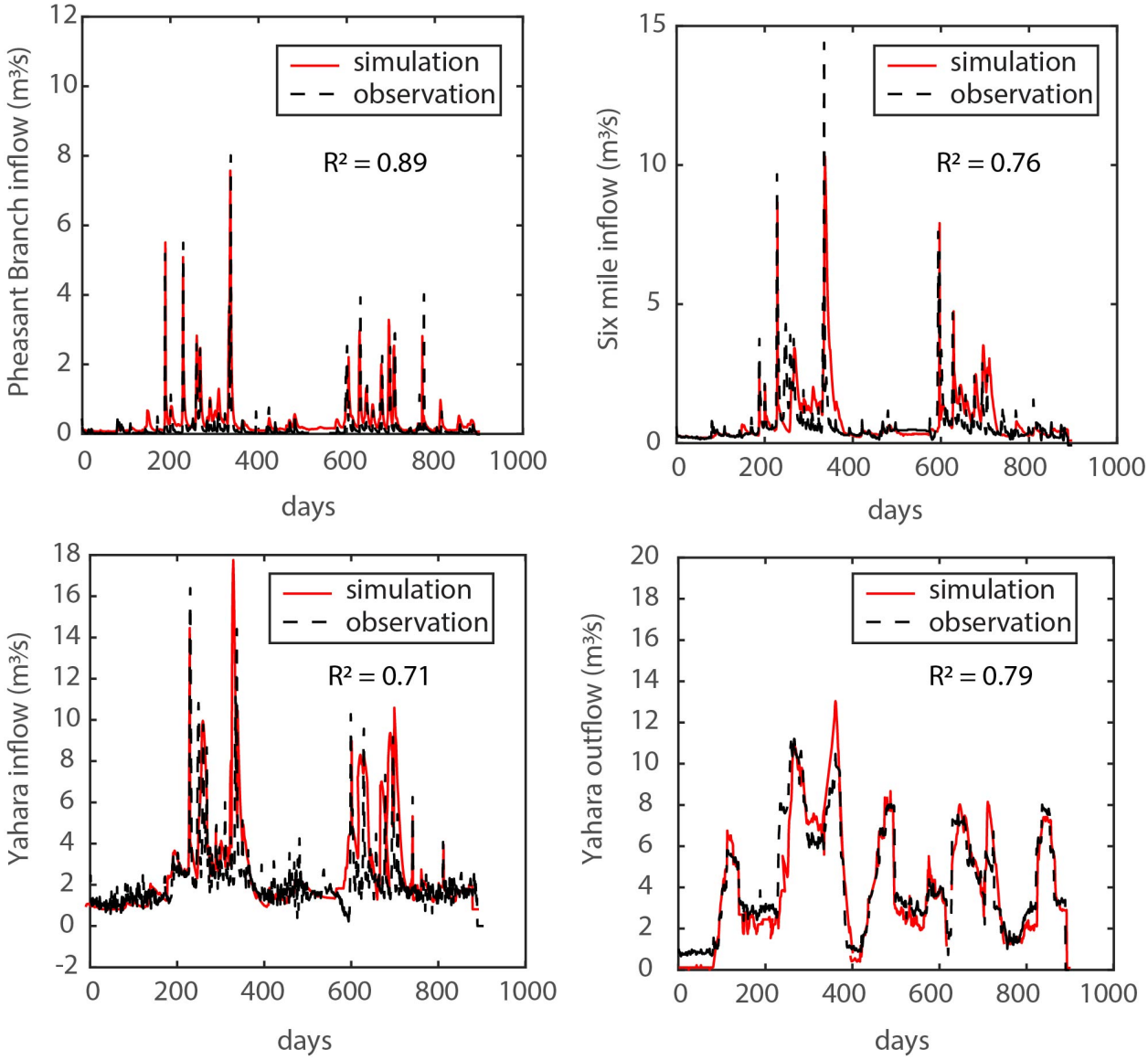
L 115: 1.How you calibrated the hydrological model?

Author response:

The hydrological PIHM-Lake model was calibrated to measured stream inflows to the lake and outflow discharges from the lake to the catchments. The model was calibrated by using the observations from 2009 to 2011 and validated by using the measurements from 2012 to 2014, within which all stream flow observations are available. To state this clearer in the main text, we slightly modified this sentence in the manuscript:

L125: The PIHM-Lake simulation covers a 37-year period (from 1979 to 2015), and its parameters were calibrated and validated with *in-situ* measured stream inflow and lake outflow discharges from the US Geological Survey.

We also attached the following figure to this reply here, which shows the fit between observed discharges of three Lake Mendota inflows (Pheasant branch, Six Mile, Yahara) and the outflow from Lake Mendota to the simulated discharges by PIHM-Lake (calibrated).



595

Referee comment:

2.From I know for the historical simulation, the inflow discharge is always drawn from the real measurements, instead of hydrological models. Do you have the measured inflow discharge for Lake Mendota?

Author response:

600 Yes, we have used measured inflow discharges for Lake Mendota at 4 inflows gages, see Fig. 1, but these monitoring stations only present an incomplete water balance as all groundwater inflow and surface overland flow to Lake Mendota are not observed, which could also contribute to the water balance of Lake Mendota. Therefore, we additionally used a calibrated hydrological PIHM-Lake model (using monitored flow data and lake surface water level fluctuations) to create two general inflows terms that close the overall lake water balance. To clarify this, we have added a sentence to the main
605 text:

L127: The application of the PIHM-Lake model for quantifying the lake inflows helped closing the water balance of Lake Mendota as groundwater inflow and surface overland flow were not measured, and the model simulations provided these inflows.

Referee comment:
L 125: How many types of nutrients were included here as the inflow boundary conditions? It is better clarify it here.
Author response:
Thank you for pointing this out. We included a sentence in the main text:

L133-136: We included the following nutrients in the inflow boundary conditions soluble reactive phosphate, adsorbed soluble reactive phosphate, dissolved organic phosphorus, particulate organic phosphorus, dissolved organic nitrogen, ammonia, nitrate, refractory dissolved organic carbon, dissolved inorganic carbon, and reactive silica.

Referee comment:
L133: I am not sure whether it is appropriate to define the inflow loading as the mean values from the water column. It means that there is no seasonal changes of DIC and silica, which is unrealistic. Could you explain why you set the inflow DIC and silica in this way?
Author response:
After a long internal discussion, we set DIC and silica to an average value as these variables are not part of the routine measurement program. As the average in-lake value is quite high, we did not expect any sensitivity of these values on the model results. Further, in-lake measurements have shown that the average concentration in the lake does not fluctuate much.

Referee comment:
2.3 Modelling Framework: Just a recommendation, it may be better to combine 2.3 to 2.7 into one part, since all of such content be longs to the model description.
Author response:
In accordance with the reviewer's suggestion, we changed the levels of sectioning of these paragraphs, e.g. "Deductive Model", "GLM-AED2", "Post-Processing of GLM-AED2 Output" and "Regression Model" are now all sub-paragraphs of "2.3 Modeling Framework".

Referee comment:
L 198: For water temperature simulation, I supposed the most important parameters should be wind factor and light extinction coefficient. How you defined these two in the model?
Author response:
For identification of calibration parameters, we used the Morris Sensitivity method, which declared the short-wave solar radiation factor, the long-wave radiation factor, the bulk aerodynamic sensible heat transfer coefficient, and the sediment temperatures as the most sensitive model parameters. Therefore, we did not calibrate the wind factor and left it a 1.0, e.g., we used the measured wind data from a close airport. The light extinction coefficient was set to a low water background value of 0.1, because the value was dynamically changing in the water quality model AED2, which backfed any changes in light extinction due to abundance of dissolved substances to the hydrodynamic model. We checked the dynamic modeled light extinction values with measured Secchi depth data, and the seasonal dynamics were replicated by the model.

Referee comment:
L 293: How you calculated GPP? It is better to clarify it here.
Author response:
GPP (in mmol C per m3 per d) was internally calculated by the AED2 model as the daily total carbon uptake of all functional phytoplankton groups. We clarified this in the main text:

L300-308: Here, GPP represents the sum of all functional phytoplankton group's photosynthesis rates parameterized as the total carbon uptake:

$$f_{uptake}^{PHY_C} = R_{growth}^{PHY}(1 - k_{pr}^{PHY})\ \phi_{temp}^{PHY}(T)\ \phi_{stress}^{PHY}(X)\ min\{\phi_{light}^{PHY}(I)\ \phi_{N}^{PHY}(NO_3, NH_4PHY_N)\ \phi_{P}^{PHY}(PO_4, PHY_P)\ \phi_{Si}^{PHY}(Rsi)\}[PHY] \quad (7)$$

where the carbon uptake $f_{uptake}^{PHY_C}$ of an individual group *PHY* depends on the growth rate R_{growth}^{PHY}, the photorespiratory loss $(1 - k_{pr}^{PHY})$, temperature scaling $\phi_{temp}^{PHY}(T)$, metabolic stress $\phi_{stress}^{PHY}(X)$, and a minimum function taking into account limitations by light $\phi_{light}^{PHY}(I)$, nitrogen $\phi_{N}^{PHY}(NO_3, NH_4PHY_N)$, phosphorus $\phi_{P}^{PHY}(PO_4, PHY_P)$ and silica $\phi_{Si}^{PHY}(Rsi)\}$ (Hipsey et al., 2017; adapted from Hipsey and Hamilton, 2008). As the GPP is the main model output variable for phytoplankton dynamics, it scales directly with biomass and Chl-a concentrations.

Referee comment:

L 333: There existed some negative values for Birgean Work in Figure 5, what is the reason for that?

Author response:

As the Birgean Work is

$$B = \int_0^{z_m} A_z(1 - \rho_z)z dz$$

a negative value can occur when a dominant part of the water column has water densities that are above 1,000 kg per m3. Hypothetically speaking, a negative Birgean value would mean that no energy is needed (or negative energy would be needed) to achieve the current stratification from a completely mixed state, which means that the current state is probably also completely mixed. We decided against discussing this in the manuscript as we focused on oxygen dynamics over time.

Referee comment:

L 371: In Figure 9B, why was the simulated AF represented by dots, instead of box plots as the measured one?

Author response:

The simulated AF is represented by dots as we calculated it from the GLM-AED2 output and were therefore able to quantify it using daily data. On the other hand, the observed data were only available every two weeks, therefore we used different interpolation techniques to get daily data. These uncertainties were captured in a box-plot.

Lake thermal structure drives inter-annual variability in summer anoxia dynamics in a eutrophic lake over 37 years

Robert Ladwig[1], Paul C. Hanson[1], Hilary A. Dugan[1], Cayelan C. Carey[2], Yu Zhang[3], Lele Shu[4], Christopher J. Duffy[5], Kelly M. Cobourn[6]

[1]Center for Limnology, University of Wisconsin-Madison, Madison, WI, USA
[2]Department of Biological Sciences, Virginia Tech, Blacksburg, VA, USA
[3]Earth and Environmental Sciences, Los Alamos National Laboratory, Los Alamos, NM, USA
[4]Department of Land, Air and Water Resources, University of California, Davis, Davis, CA, USA
[5]Department of Civil & Environmental Engineering, The Pennsylvania State University, State College, PA, USA
[6]Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, VA, USA

Correspondence to: Robert Ladwig (rladwig2@wisc.edu)

Abstract. The concentration of oxygen is fundamental to lake water quality and ecosystem functioning through its control over habitat availability for organisms, redox reactions, and recycling of organic material. In many eutrophic lakes, oxygen depletion in the bottom layer (hypolimnion) occurs annually during summer stratification. The temporal and spatial extent of summer hypolimnetic anoxia is determined by interactions between the lake and its external drivers (e.g., catchment characteristics/nutrient loads, meteorology), as well as internal feedback mechanisms (e.g., organic matter recycling, phytoplankton blooms). How these drivers interact to control the evolution of lake anoxia over decadal time scales will determine, in part, the future lake water quality. In this study, we used a vertical one-dimensional hydrodynamic-ecological model (GLM-AED2) coupled with a calibrated hydrological catchment model (PIHM-Lake) to simulate the thermal and water quality dynamics of the eutrophic Lake Mendota (USA) over a 37-year period. The calibration and validation of the lake model consisted of a global sensitivity evaluation as well as the application of an evolutionary optimization algorithm to improve the fit between observed and simulated data. ~~By quantifying~~We calculated stability indices (Schmidt Stability, Birgean Work, stored internal heat), ~~we~~ identified spring mixing and summer stratification periods, and quantified the energy required for stratification and mixing. To qualify which external and internal factors were most important in driving the inter-annual variation in summer anoxia, we applied a random-forest classifier and multiple linear regression to modeled ecosystem variables (e.g., stratification onset and offset, ice duration, gross primary production.) Lake Mendota exhibited prolonged hypolimnetic anoxia each summer, lasting between 50-60 days. The summer heat budget, ~~as well as~~ the timing of thermal stratification, and the gross primary production in the epilimnion were the most important predictors of the spatial and temporal extent of summer anoxia periods in Lake Mendota. ~~An~~Inter-annual variability in anoxia was largely driven by physical factors: earlier onset of thermal stratification in combination with a higher vertical stability strongly affected the duration and spatial extent of summer anoxia. A step change upward in summer anoxia in 2010 was unexplained by the GLM-AED2 model. Although the cause remains unknown, possible factors include invasion by the predacious zooplankton, *Bythothrephes longimanus*. As the heat budget depended primarily on external meteorological conditions, the spatial and temporal extent of summer anoxia in Lake Mendota is likely to increase in the near future as a result of projected climate change in the region.

1 Introduction

The availability of dissolved oxygen in lakes governs ecological habitats and niches, the rates of redox reactions, and the processing of organic matter throughout the water column (Cole and Weihe, 2016). When a lake is thermally stratified, metabolism in the surface layer (epilimnion) can act as a net source or sink of oxygen, depending on the balance of gross primary production and ecosystem respiration, and deviations of dissolved oxygen from saturation values are modulated by atmospheric exchange (Odum, 1956). Additionally, entraining inflows can also act as an important oxygen sink or source depending on the lake morphometry, the inlet discharge, and the carrying capacity for dissolved oxygen (Burns, 1995). Below the thermocline, dissolved oxygen is depleted in the bottom layer (hypolimnion) by organic matter mineralization in the water column and the sediment oxygen demand (Livingstone and Imboden, 1996). These oxygen depletion processes can be quantified as either a volumetric sink (e.g., due to organic matter mineralization in the water column) or as an areal sink (e.g. oxygen demand in the sediments). The depletion rates of oxygen depend on the organic matter pool (Müller et al., 2012, 2019), the trophic status of the lake (Rhodes et al., 2017; Rippey and McSorley, 2009), the area to volume relationship over depth (Livingstone and Imboden, 1996), and the chemical demand of the water column and sediments (Yin et al., 2011).

While the biogeochemistry of lake oxygen is well studied, there is much to be learned about decadal-scale controls over ecosystem patterns in oxygen and the interactions of external drivers with internal processes that control those patterns. Oxygen depletion in the hypolimnion that results in hypoxia (dissolved oxygen < 2 mg L^{-1}; Diaz and Rosenberg, 2008) and anoxia (dissolved oxygen < 1 mg L^{-1}; Nürnberg, 1995b) is a product of interacting external drivers (e.g., climate, land use practices in the catchment) that control mass fluxes (Jenny et al., 2016b), morphometric characteristics, and productivity that influences vertical transport and water column stability (Meding and Jackson, 2003). Unprecedented changes to the climate and catchment land use are likely to have nonlinear consequences on aquatic water bodies and will potentially intensify hypolimnetic anoxia (Jenny et al., 2014, 2016a; Sánchez-España et al., 2017).

The influence of physical controls on lake anoxia is of particular interest because it provides clues to how lakes might respond to long-term changes in exogenous drivers. The timing of anoxia has been found to be strongly related to the onset and offset of stratification, as well as sediment oxygen demand in small eutrophic lakes (Biddanda et al., 2018; Foley et al., 2012; Nürnberg, 2004). A reduction of winter/spring mixing and increase in stratification can be major drivers of deep-water oxygen depletion (North et al., 2014). For Lake Mendota (USA), Snortheim et al. (2017) concluded that changes in air temperature ~~was~~were the main driver of the spatio-temporal extent of summer hypolimnetic anoxia~~.~~, but was unable to disentangle the direct effect of air temperature (i.e., warmer water temperatures) vs. its indirect effect (i.e., stronger thermal stratification) on oxygen dynamics. The question thus remains open~~, even for single ecosystems~~: Under what circumstances and at what time scales does thermal stratification strength act as the dominant driver of hypolimnetic anoxia versus biogeochemical processes?

Studying decadal~~-~~-scale lake anoxia ~~would benefit from~~requires an ecosystem-scale metric of lake anoxia and an analytical framework for tying that metric to physical and biological processes. Several metrics of oxygen availability have

been developed by previous studies, such as mean areal hypolimnetic oxygen depletion (AHOD, in mg O_2 m^{-2} d^{-1}, Cornett and Rigler, 1979; Hutchinson, 1938), volumetric rate of oxygen consumption (VOD, in mg O_2 m^{-3} d^{-1}, Cornett, 1989), areal hypolimnetic mineralization (AHM, in g O_2 m^{-2} d^{-1}, Matzinger et al., 2010), and the Anoxic Factor (AF, in days, Nürnberg, 1995a, 1995b, 2004). Compared to the Anoxic Factor, most metrics calculate an oxygen depletion rate, whereas the Anoxic Factor provides an integrated metric that includes the spatial as well as temporal dimensions of anoxia per season. As the Anoxic Factor sums up the product of anoxia duration with the corresponding area, it is therefore a useful metric to evaluate long-term dynamics of hypolimnetic anoxia, and to compare the intensity of anoxia between years and different study sites. The Anoxic Factor and its derivative the Hypoxic Factor (the difference being the threshold of dissolved oxygen; (Nürnberg, 2004) have been used in several studies, and observations range from 0 to 83 days per summer for different lake ecosystems (Nürnberg, 1995b).

~~Hydrodynamic~~Coupled hydrodynamic-water quality models are an established approach to studying lake physical and biological responses to external drivers (Hipsey et al. 2019). An advantage of a lake ecosystem model calibrated to observed long-term data is that it can reproduce finer temporal and spatial resolution than observational data permits for most ecosystems (Stanley et al., 2019) allowing for the investigation of complex ecosystem dynamics (Ward et al., 2020). By applying an ecosystem model driven by sub-daily meteorological and daily hydrological inflow data, physical processes relevant to hypolimnetic oxygen depletion (such as the onset and seasonal evolution of thermal stratification and gas transfer velocities) can be resolved at an hourly resolution, and can subsequently be incorporated into stochastic models to gain an understanding about the relationships between drivers and their respective impacts on hypolimnetic anoxia (Snortheim et al., 2017). Results from deterministic lakes models can be analysed using statistical models to derive general empirical relationships. Results can also be compared with alternative, deductive approaches, which tend to be simpler models meant to reproduce gross ecosystem properties. An example relevant to lake anoxia is the simple deductive hypolimnetic oxygen depletion model by Livingstone and Imboden (1996), which established that ~~already~~ minor year-to-year meteorological variations during spring can cause an expansion of the thickness of the summer anoxia layer.

This study aims to determine the extent to which physical, chemical, and biological (internal and external) factors control the inter-annual variability of the summer Anoxic Factor over 37~~-~~ years in the eutrophic Lake Mendota. We first use a lake hydrodynamic-water quality model to generate fine scale ecosystem states and fluxes based on observational data ~~from the North Temperate Lakes Long Term Ecological Research program. Further~~. Second, we use a deductive lake anoxia model and data-driven empirical models to evaluate observed and simulated data and to determine broad-scale control over lake anoxia. We answer three questions with our modeling framework: (1) Overall, do internal biogeochemical processes or external loadings control year-to-year variability of the Anoxic Factor? (2) What are essential in-lake physical and biological controls over the long-term variability in anoxia? (3) As the timing of thermal stratification governs hypolimnetic oxygen depletion, what is the year-to-year variability of Lake Mendota's head budget? Answers to these questions will further our understanding of lake ecosystem responses to climate ~~or~~and landscape changes ~~and~~to support ~~future~~water quality management ~~decisions~~.

2 Materials and Methods

2.1 Study Site

Lake Mendota is a 39.61 km^2, 25 m deep, eutrophic lake in southern Wisconsin ~~(USA).~~, USA (Figure 4). The lake has a mean water residence time of 4.3 years (McDonald and Lathrop, 2017). Lake Mendota's mixing regime is characterized by a summer stratification period from late April through October and an inverse winter stratification period under ice (Brock, 1985). Lake Mendota's air temperature ranges from -39 to 40 °C with a mean annual value of 8 °C, and an annual precipitation ranging from 540 to 990 mm with an average of 780 mm (Lathrop, 1992). The 604 km^2 watershed is dominated by agricultural land (67%) and developed urban land (22%) (Duffy et al., 2018). Since 1995, physical, chemical, and biological characteristics have been sampled biweekly to monthly by the North Temperate Lakes Long Term Ecological Research Program (Magnuson et al., 2006). ~~-~~We note that Lake Mendota is a "hard water" lake with pH > 7 and exhibits consistently high dissolved inorganic carbon concentrations, with speciation dominated by bicarbonate and carbonate~~.~~ (Hart et al., 2020).

2.2 Driver Data Acquisition

Meteorological forcing data were obtained from the second phase of the North American Land Data Assimilation System (NLDAS-2, Xia et al., 2012). The data from the grid cell were centered at -89.4375, 43.0625. The NLDAS-2 grid cells are 1/8th-degree spacing and data are at an hourly resolution from January 1, 1979 to present (Mitchell, 2004). Meteorological parameters used in this study included wind speed, air temperature, specific humidity, surface pressure, surface downward short- and longwave radiation, and total precipitation, which were used primarily as boundary data for GLM-AED2. Relative humidity was calculated *post hoc* as a function of specific humidity, air temperature, and surface pressure.

To quantify the water budget in Lake Mendota, we simulated the water inflow from the catchment (through stream flow, overland flow, and groundwater flow) to the lake, and water outflow from the lake to the catchments using a physically based distributed hydrologic model, PIHM-Lake (Penn State Integrated Hydrologic Model, Qu and Duffy, 2007). PIHM-Lake integrates hydrologic processes in a lake-catchment coupled system simulating the surface and subsurface hydrologic ~~interaction~~interactions within the catchment and between the catchment and the lake. Hydrologic interactions within the catchment are modeled in three dimensions, while the lake is represented in PIHM-Lake as a simplified one-dimensional bucket model assuming a uniform lake surface area and depth. PIHM-Lake tracks the change in water storage from the watershed's vegetation canopy, ground surface, unsaturated soil zone, saturated soil zone, and lake by using the semi-discrete finite volume method and a triangular irregular network (TIN). The PIHM-Lake simulation covers a 37-year period (from 1979 to 2015), and ~~was~~its parameters were calibrated and validated ~~using~~with *in-situ* measured stream ~~flow, groundwater table level, lake surface water level,~~inflow and lake outflow discharges from the US Geological Survey. The application of the PIHM-Lake model for quantifying the lake inflows helped closing the water balance of Lake Mendota as groundwater inflow and surface overland flow were not measured, and the model simulations provided these inflows.

Surface nutrient loadings from the Yahara River and Pheasant Branch inflows into Lake Mendota were estimated by regression models using discharge and nutrient concentration data from USGS gages~~.~~ (Appling et al., 2015; data is available at Ladwig et al. (2020) via the Environmental Data Initiative). Combined with the simulated discharge time series from PIHM-Lake, these regressions were ~~using~~used to compute daily loading data. We included the following nutrients in the inflow boundary conditions soluble reactive phosphate, adsorbed soluble reactive phosphate, dissolved organic phosphorus, particulate organic phosphorus, dissolved organic nitrogen, ammonia, nitrate, refractory dissolved organic carbon, dissolved inorganic carbon, and reactive silica. For a complete description of the inflow loading ~~regression~~regressions, see Weng et al. (2020). To provide information regarding adsorbed soluble reactive phosphate, we doubled measured total phosphorus (TP) concentrations and applied specific ratios to individual phosphorus forms (Farrell et al., 2020; Snortheim et al., 2017; Weng et al., 2020). This put our estimates of TP near the upper range of previous load estimates. Bennett et al., (1999) estimated the long-term average annual TP load to be about 34 t, whereas our average annual TP load (with adsorbed phosphate) was about 50.6 t and ranged between 5.3 to 146.1 t (1979-2015). Our average annual TP load (without adsorbed phosphate) was about 25.3 t and ranged between 2.7 to 73.1 t (1979-2015), which is similar to previous estimates between 15 to 67 t (Kara et al., 2012). By doubling our TP by adding adsorbed phosphate, we accommodate a potential TP load underestimation due to the importance of extreme storm events on particulate loads (Carpenter et al., 2018). As direct measurements of inlet loadings of refractory organic matter, dissolved inorganic carbon (DIC) and silica were not available, we ~~assumed~~used constant average values for the inflow loadings similar to the long-term mean values of the water column.

Monitored NTL-LTER data from 1995 to 2015 were used for model ~~verification.~~calibration and validation. Data included water temperature and dissolved oxygen concentrations (Magnuson et al., 2020b) with a vertical spatial resolution of 1 m from the surface to 24 m. Data were measured biweekly during summer, monthly during fall and once per winter. The dissolved oxygen data set was complemented with historical measured dissolved oxygen data from 1992 to 1994. NTL-LTER data also included pH, dissolved inorganic carbon, dissolved organic carbon, nitrate, ammonia, soluble reactive phosphate and silica sampled at the depths 0, 3, 8, 10, 12, 14, 16, 18, 20 and 22 m (Magnuson et al., 2020a). Surface integrated~~-~~ samples of epilimnetic chlorophyll-a and Secchi depth were used to evaluate GLM-AED2's predictions of phytoplankton biomass and light extinction (Magnuson et al., 2020c, 2020d).

2.3 Modeling Framework

Our modeling framework to investigate drivers of hypolimnetic anoxia consisted of three components (Figure 5):

(1) *Deductive model:* A deductive model formulated by Livingstone and Imboden (1996) was run on the monitored field data to characterize the empirical relationships between observed dissolved oxygen data and oxygen depletion processes and to quantify the contributions of water column and sediments to hypolimnetic oxygen demand (Figure 5). The deductive model furthered our ecosystem-scale understanding of the partitioning between volumetric and areal oxygen depletion sinks in Lake Mendota. Therefore, this approach was used independently of the other modeling approaches as a "check" on the sediment oxygen demand rates of Lake Mendota used in GLM-AED2.

(2a) *GLM-AED2*: To gain a more mechanistic understanding of how processes driving oxygen depletion lead to ecosystem-scale oxygen dynamics, we used the vertical one-dimensional hydrodynamic water quality model, GLM-AED2 (Hipsey et al., 2019b). GLM-AED2 uses meteorological, hydrological, and nutrient load data as inputs and predicts lake physical, chemical, and biological dynamics, including those of dissolved oxygen. The advantage of using GLM-AED2 is that it quantifies and tracks processes relevant to oxygen cycling using well-accepted physical and biogeochemical interactions that otherwise are difficult to infer from observational data alone ~~(~~(see section '2.3.2 GLM-AED2'; Figure 5). Although GLM-AED2 is a deterministic model, hypolimnetic anoxia is an emergent ecosystem property that derives from a complex suite of interactions within the model (Snortheim et al., 2017). Therefore, we used GLM-AED2 to simulate and track states and fluxes of modelled variables.

(2b) *Regression model*: To derive generalized relationships between the interannual variation in hypolimnetic anoxia and the driving data, as well as the output from GLM-AED2, we used statistical models on our combined dataset of monitored and modeled data. Because the number of potential candidate predictors is high, we used a machine learning approach to determine the most significant predictors of seasonal hypolimnetic anoxia at the inter-annual scale (Figure 5). These predictors were then used in a multiple linear regression to rank their influence on hypolimnetic anoxia.

2.43.1 Deductive Model

Using temporal and spatial linearly interpolated observed dissolved oxygen data, we applied the simple deductive oxygen depletion model according to Livingstone and Imboden (1996) in which the oxygen depletion rate *J(z)* at depth *z* is

$$J(z) = J_v(z) + J_A(z)\alpha(z), \quad (1)$$

where J_V is the volume sink (mass per volume per time) representing organic matter mineralization processes, e.g. microbial respiration in the water column, J_A is the area sink (mass per area per time) representing sediment oxygen demand, and α is a function for the ratio of sediment area to water volume (Bossard and Gächter, 1981; Livingstone and Imboden, 1996):

$$\alpha(z) = -\frac{1}{A(z)}\frac{dA(z)}{dz}. \quad (2)$$

We used observed dissolved oxygen data from 1992 to 2015 (measured biweekly after ice offset) to calculate the specific oxygen depletion over depth for each year individually from the date of spring mixing offset to the date when oxygen concentrations were below 2 mg L-1 (criterium for hypoxia). Only dissolved oxygen data below a depth of 15 m were used. The derivatives of area to depth were approximated by using forward and backward differencing. The terms J_V and J_A were assumed to be constant for every year (assuming the hypolimnion to be homothermic) and were determined by using weighted linear regression.

2.53.2 GLM-AED2

For simulating Lake Mendota, we used the coupled 1D vertical hydrodynamic-ecological model GLM-AED2 (GLM: v.3.1.0a1, AED2: 1.3.4, developed by University of Western Australia, Hipsey et al. 2019). The hydrodynamic model GLM

incorporates a flexible Lagrangian grid with each layer's thickness dynamically changing in response to the respective water density (Hipsey et al., 2019b). Surface mixing processes are computed via an energy balance approach that compares the available (turbulent) kinetic energy to the internal potential energy of the water column (Hipsey et al., 2019b).

The water quality module, AED2, was configured to simulate the dynamics of dissolved oxygen, silica, inorganic carbon, organic matter (refractory, particulate and dissolved C, N and P), inorganic matter (refractory, particulate and dissolved C, N and P) as well as PO_4, NO_3, and NH_4, and two functional phytoplankton groups (representing diatoms and cyanobacteria, ~~Appendix~~Supplement 1). The model was run on an hourly time step and output data were saved at a daily timestep on noon**.** The thickness of each model layer (set to a maximum of 75 layers) could vary between 0.15 and 1.5 m with a minimum layer volume of 0.1 m^3. The source code of the model's version, configuration files, input and output data are stored and accessible at Ladwig et al. (2020) via the Environmental Data Initiative.

A global sensitivity analysis (Morris Method after Morris 1991) was conducted to identify the most influential parameters for the predictions of water temperature, dissolved oxygen, dissolved inorganic carbon, silica, nitrate and phosphate, respectively. Using the Morris Method with 10 iterative runs, the distributions of the absolute elementary effects (the model change quantified by a fit function, here the root-mean squared error (RMSE) between observed and simulated data) of each parameter were calculated. According to Morris (1991) and Saltelli et al. (2004), the mean of the absolute elementary effects represents the overall sensitivity of the model outcome to each parameter, and the standard deviations are a metric of the interactions between different parameters. All parameters with a normalized mean elementary effect over 0.1 were declared sensitive and were used for the calibration.

According to our calculated absolute elementary effects, we included the following ~~10~~ parameters in the calibration, listed according to their respective state parameter (~~Appendix~~Supplement Figure A1). ~~Water~~For calibrating the six state variables, these parameters were: 1) water temperature: bulk aerodynamic coefficient for sensible heat transfer (ch), long-wave radiation factor (lw_factor), mean sediment temperature (sed_temp_mean), shortwave radiation factor (sw_factor~~).~~); 2) Dissolved oxygen: Sediment flux (Fsed_oxy), mineralization rate of dissolved organic matter (Rdom), temperature multiplier for sediment flux (theta_sed_oxy~~).~~); 3) Dissolved inorganic carbon: Sediment flux (Fsed_dic), half-saturation constant for oxygen dependence on sediment flux (Ksed_dic), temperature multiplier for sediment flux (theta_sed_dic~~).~~); 4) Silica: Sediment flux (Fsed_rsi), half-saturation constant for oxygen dependence on sediment flux (Ksed_rsi), temperature multiplier for sediment flux (theta_sed_rsi~~).~~); 5) Nitrate: Sediment flux (Fsed_nit), half-saturation constant for oxygen dependence on denitrification (Kdenit), half-saturation constant for oxygen dependence on sediment flux (Ksed_nit), maximum reaction rate of denitrification at 20 °C (Rdenit~~).~~); and 6) Phosphate: Sediment flux (Fsed_frp), half-saturation constant for oxygen dependence on sediment flux (Ksed_frp), temperature multiplier for sediment flux (theta_sed_frp).

We applied a combination of an automatic calibration technique and manual calibration for the calibration period from 2005 to 2015. First, the derivative-free, evolutionary optimization algorithm (CMA-ES, Hansen (2016)) was used to minimize the RMSE between observed and simulated data (data were split into a calibration, 2005-2015, and a validation period, 1995-2004). We used a time period prior to the calibration period for validation to stress-test the model by applying it

a time period with potential different ecological characteristics. The model parameters were calibrated iteratively (and fixed for the next calibration step) in the following order: water temperature, dissolved oxygen, dissolved inorganic carbon, silica, nitrate and, last, phosphate. We did not calibrate for phytoplankton functional group biomass because it was out of scope for this analysis, but the model qualitatively recreated observed seasonal succession. Initial model parameter values were taken from default parameter values and ranges, as well as literature values (Hipsey et al., 2017; Snortheim et al., 2017). Calibration of water temperature and dissolved oxygen concentrations were run for 300 iterations, the other variables for 200 iterations. The fit criteria were root-mean square error (RMSE), Nash-Sutcliffe Coefficient of Efficiency (NSE) and Kling-Gupta Efficiency (KGE) (Gupta et al., 2009) for the calibration period, the validation period and the total time period. The advantage of combining an automatic approach and a manual post-calibration for an overparameterized model such as GLM was that CMA-ES first limited the possible parameter space of each parameter, then in a second calibration step, parameters could be manually changed to improve overall dynamics and behavior without relying on a fixed objective function. The manual calibration was done to ensure that the model was not overoptimized with unrealistic parameter combinations of the biological parameters. This calibration approach was done in accordance with other aquatic ecosystem modeling studies (Fenocchi et al., 2019; Mi et al., 2020), that did not apply computational optimization to water quality models.

2.~~6~~3.3 Post-Processing of GLM-AED2 Output

We quantified two heat budget metrics from simulated water temperature data, the Schmidt Stability (Idso, 1973; Read et al., 2011; Schmidt, 1928) and the Birgean Work (Birge, 1916; Idso, 1973). Schmidt Stability (*St*) is a stability index that expresses the amount of energy needed to mix the entire water column to uniform temperatures without affecting the amount of internal energy, whereas Birgean Work (*B*) is a stability index that quantifies the amount of external energy that is theoretically needed to build up the current stratification from a hypothetical completely mixed state. The sum of both terms, the total work *G*, gives an estimate of the energy needed to keep a lake isothermal during stratified conditions:

$$G = St + B, \tag{3}$$

$$G = \frac{g}{A_s}\int_0^{z_m} A_z(1-\rho_z)(z_v - z)dz + \frac{g}{A_s}\int_0^{z_m} A_z(1-\rho_z)z dz\,, \tag{4}$$

where g is gravity, A_s is the surface area (m^2), z_m is the maximum depth (m), A_z is the respective area at the depth z, ρ_z is the respective density at the depth z (kg m^{-3}), z_v is the depth of the center of volume ($z_v = \frac{1}{V}\int_0^{z_m} A_z z dz$), and V is the volume (m^3). The stagnancy of deep water can be quantified by calculating a heat budget ratio (*HBR*):

$$HBR = \frac{G}{B}, \tag{5}$$

which compares the amount of energy needed to maintain isothermal conditions to the amount of available external energy (Kjensmo, 1994). Here, an increased stagnancy of deep waters results in a reduced exchange of fluxes between the surface mixed and the bottom layer. Therefore, HBR values > 1 indicate the isolation of the bottom water layers from surface fluxes in a lake.

Internal energy - as the stored thermal energy in the water column - was quantified using the R package rLakeAnalyzer (Winslow et al., 2019) as:

$$E_{internal} = \frac{1}{A_s}\int_0^{z_m} T_z * c_w * m_z dz\,, \quad (6)$$

where T_z is the water temperature at depth z (°C), c_w is the specific heat of water ($J\ kg^{-1}\ K^{-1}$), and m_z is the mass of water at depth z (kg).

The thermocline depth was defined as a planar separation between the surface mixed and the bottom stagnant layer. The specific depth of this planar thermocline was quantified as the depth of the maximum density difference over the vertical axis where the minimum water temperature was above 4 °C and the density difference between surface and bottom layer was above 0.1 kg m^{-3}, signaling stratified conditions.

The temporal and spatial extent of anoxia during the summer season was quantified using the Anoxic Factor:

$$AF = \sum_{i=1}^{n} \frac{t_i A_i}{A_s}, \quad (7)$$

which sums the product of the anoxia duration t (days) with the corresponding area A (m^2) to the total surface area A_S when the in-water dissolved oxygen concentrations were below a threshold of 1 mg L^{-1} (Nürnberg, 1995b). As the Anoxic and Hypoxic Factor use the same equation with different thresholds relating essentially all anoxia information also to hypoxia, we focused on only quantifying the Anoxic Factor in this study. Anoxic Factor was calculated using the simulated dissolved oxygen concentrations as well as the approximately bi-weekly monitored field data, in which case data were temporally and spatially interpolated using an ensemble of approaches (linear, constant and spline interpolation between neighboring data points). We quantified the seasonal Anoxic Factor only ~~during~~for the summer ~~as Lake Mendota is not experiencing winter hypolimnetic anoxia under ice~~season.

2.~~7~~3.4 Regression Model

We evaluated ~~22~~21 candidate predictors on their relative importance in predicting the simulated summer Anoxic Factor of the respective year n (see Table 1 for an overview and further explanation). All candidate predictors were either modeled output or boundary data for the model. This enabled the regression analysis to identify internal connections in the numerical model itself (similar analyses of modeled output and driver data were done in Snortheim et al., 2017; Ward et al., 2020; Weng et al., 2020). For the calculation of certain candidate predictors, the water column was separated into an upper layer (from surface to a depth of 10 m) and a lower layer (from 10 m to maximum depth). Although this is a rough approximation, this depth roughly represents the thermocline depth and further separates the water column into a zone without light limitation and one with light limitation.

To represent external ~~control~~forcing processes, we included the seasonal total phosphorus inflow and seasonal total nitrogen inflow ~~concentrations~~loadings for the pre-summer period (winter, spring, summer) of each respective year. Further, we included the Birgean Work for spring and summer of each year as the Birgean Work represents the amount of external

energy (mostly by wind shear stress) that is needed to build up the current thermal structure. In addition to Birgean Work, we also included Schmidt Stability, the HBR ratio, the onset, end and duration of spring mixing, the onset, end and duration of summer stratification, the mean hypolimnetic water temperature at the onset of stratification, as well as the end and duration of the ice period prior to summer to investigate the effects of physical control on ~~hypolimentic~~hypolimnetic anoxia.

In-lake biogeochemical processes were represented by ~~the maximum height of anoxia during summer,~~ the dissolved oxygen concentration differences between spring mixing onset and offset in the hypolimnion (Livingstone and Imboden (1996) suggested that in eutrophic lakes dissolved oxygen reductions during the mixing phase can have profound effects on the summer anoxia), organic carbon (both dissolved and particulate, respectively) concentration differences between spring mixing and stratification in the hypolimnion, as well as cumulative gross primary production in the epilimnion and hypolimnion. Organic matter gradients were investigated because dissolved organic carbon can be used as a proxy for allochthonous organic matter contributions to bacterial mineralization rates (Hanson et al., 2003). Gross primary production (GPP) was included as an example organic matter source that can fuel bacterial mineralization (Yuan and Jones, 2019). Here, GPP represents the sum of all functional phytoplankton group's photosynthesis rates parameterized as the total carbon uptake:

$$f_{uptake}^{PHY_C} = R_{growth}^{PHY}(1 - k_{pr}^{PHY})\,\phi_{temp}^{PHY}(T)\,\phi_{stress}^{PHY}(X)\,min\{\phi_{light}^{PHY}(I)\,\phi_{N}^{PHY}(NO_3, NH_4 PHY_N)\,\phi_{P}^{PHY}(PO_4, PHY_P)\,\phi_{Si}^{PHY}(Rsi)\}[PHY] \quad (7)$$

where the carbon uptake $f_{uptake}^{PHY_C}$ of an individual group *PHY* depends on the growth rate R_{growth}^{PHY}, the photorespiratory loss $(1 - k_{pr}^{PHY})$, temperature scaling $\phi_{temp}^{PHY}(T)$, metabolic stress $\phi_{stress}^{PHY}(X)$, and a minimum function taking into account limitations by light $\phi_{light}^{PHY}(I)$, nitrogen $\phi_{N}^{PHY}(NO_3, NH_4 PHY_N)$, phosphorus $\phi_{P}^{PHY}(PO_4, PHY_P)$ and silica $\phi_{Si}^{PHY}(Rsi)\}$ (Hipsey et al., 2017; adapted from Hipsey and Hamilton, 2008). As the GPP is the main model output variable for phytoplankton dynamics, it scales directly with biomass and Chl-a concentrations.

To determine the relative importance of these candidate predictors that may influence the duration and extent of anoxia, we applied the Boruta R package (Kursa and Rudnicki, 2010) to identify the relevant predictors by using a wrapper built around a random forest classifier. The Boruta feature selection duplicates predictor values, which are then randomly shuffled to create so-called shadow attributes. If the variable predictor values (here the averaged accuracy loss normalized by the standard deviation, obtained from multiple random forest classifier runs) of the original values are significantly greater than the shadow predictor values, these variables are deemed relevant (Kursa and Rudnicki, 2010). Only model output and model driver data from the period 1980-2009 were used in the regression analysis. The first year, 1979, was dropped from the investigations due to a lack of prior winter information. The years 2010-2015 were dropped due to an apparent ecosystem shift (see Section '3.4 Oxygen Dynamics'). Meteorological quarterly divisions (DJF, MAM, JJA, SON) of the year were used to define seasons. After selecting important predictors driving the inter-annual variability in Anoxic Factor using the random forest method, we applied the remaining ~~10~~7 selected predictors in a multiple linear regression model to quantify their respective importance on predicting the Anoxic Factor. ~~Further, stepwise~~Stepwise model-selection iteratively removed predictors to improve the regression model's AIC. ~~Our final~~This multiple linear regression model to predict Anoxic Factor

included seven variables: HBR ratio during ~~summer, maximum height of anoxia above sediment, Hypolimnetic water temperature at stratification onset, hypolimnetic~~spring, HBR ratio during summer, Birgean Work in spring, epilimnetic GPP, Schmidt Stability in summer, Birgean Work in summer, and onset date of stratification. We reduced the complexity of the final multiple linear regression model to only three predictors of Anoxic Factor: onset date of stratification, Schmidt Stability in summer, and epilimnetic GPP. Schmidt Stability was included instead of Birgean Work as the resulting AIC of both models were similar, but the concept of Schmidt Stability is more commonly used in the limnological research community (Supplement Table A3). The final multiple linear regression model was configured as (scaled predictors, adjusted R2 = 0.~~88~~84, $p < 0.001$ Supplement Table A4).

$$\hat{y} = 0.47Summer.HBR - 0.15Intensity + 0.29Wtemp.Strat + 0.23Epi24Epi.GPP + 0.53Summer54Summer.St - 0.64Summer.B - 0.66Onset46Onset.Strat - 1.04 * 10^{-15}5.44 * 10^{-17} + \hat{\epsilon}, \quad (8)$$

where $\hat{\epsilon} \sim N(0,34^2)(0,38^2)$.

Relative importance of model fit was calculated as the R^2 contribution averaged over ordering among regressors (relaimpo package, Grömping 2006).

3 Results

3.1 Oxygen Depletion Rates

The derived annual oxygen depletion rates by the deductive model confirmed Lake Mendota's hypolimnetic anoxia as primarily driven by mineralization of organic matter. Observed oxygen depletion rates and area-volume ratios were positively correlated for all years except 1993, 1997 and 2007 (Figure 6). For years with a positive relationship, the average ~~volume~~volumetric sink J_V as 0.16 g m^{-3} d^{-1} and the average ~~area~~areal sink J_A with 0.04 g m^{-2} d^{-1} (adjusted $R^2 = 0.13$, $p < 0.001$). Lake Mendota's hypolimnetic oxygen depletion ~~is~~was mainly driven by water column respiration processes over sediment oxygen demand. The annual volumetric depletions rate followed a normal distribution with an increase in the volumetric sink in recent years. The areal depletion rate distribution was positively skewed. An inspection of the residuals from the model fits indicates that the linear regression model may not be appropriate for some years, especially for values of the sediment to area volume ratio $\alpha(z)$ near 0.5 m^2 m^{-3}.

Averaging this total oxygen depletion rate (volume and area ~~sink~~sinks) over the hypolimnion~~,~~ gave a potential total oxygen depletion of ~~approx. ~~~1 g m^{-3}~~2~~ d^{-1}~~.~~ (~ 32 mmol [O_2] m^{-2} d^{-1}). To conceptualize ~~such~~this a depletion rate in our deterministic GLM-AED2 model, we used a maximum sediment oxygen demand (SOD) of 100 mmol [O_2] m^{-2} d^{-1}~~, which~~. This rate represented the total sum of volumetric and areal oxygen sinks indirectly as ~~we aimed to represent~~ internal fluxes of organic carbon from the sediment back into the water column ~~that~~ would ~~additionally~~ drive additional oxygen depletion. This high value of SOD was scaled by the water temperature using an Arrhenius multiplier, effectively reducing it to a value between 1 to 1.5 g m^{-3} ~~d^{-1}~~2 d^{-1} (32 to 47 mmol [O_2] m^{-2} d^{-1}) of maximum oxygen depletion by the sediment sink in the hypolimnion

during summer stratification. A recent modeling study investigating the formation of metalimnetic oxygen minima in a drinking water reservoir by Mi et al. (2020) confirmed that such high maximum SOD values are typical for many lakes.

3.2 GLM-AED2 Calibration and Validation

The thermal characteristics of Lake Mendota were replicated well, especially water temperatures in the surface layers ~~(~~(Table 2, Figure 7, ~~Appendix~~Supplement Figure A2), with an RMSE of 1.~~3~~30 °C, ~~and~~ an NSE and ~~a~~an KGE of 0.97, which is within the range of previous modeling studies (Bruce et al., 2018; Read et al., 2014). The simulated dissolved oxygen concentrations in the whole water column achieved an RMSE of 3.22 mg L^{-1}, an NSE of 0.56, and an KGE of 0.77. Here, the average fits were better in the surface layer (RMSE of 2.77 mg L^{-1}) compared to the bottom layer (RMSE of 3.31 mg L^{-1}), whereas the temporal dynamics (as expressed in NSE and KGE) were slightly better in the bottom layer (an NSE of 0.64, KGE of 0.81) compared to the surface layer (NSE of -0.36, KGE of 0.47).

In contrast, the water quality model reproduced concentrations of the biogeochemical variables better at depth than at the surface, as evidenced by higher NSE values (Table 2, ~~Appendix~~Supplement Figure A3 – ~~A7~~A8). The density distributions of residuals (observed minus simulated data) are in agreement (Figure 8) for water temperature, dissolved oxygen concentrations, nitrate, phosphate and ~~ammonia,~~ammonium (we did not use ammonium data during calibration, but included it in the visual inspection to check general nitrogen dynamics replicated by the GLM-AED2 model), whereas the model overestimated dissolved inorganic carbon concentrations and chlorophyll-a concentrations, and underestimated silica concentrations. ~~It should be noted that~~As described above, the ~~inlet~~inflow concentrations of DIC and silica were assumed to be constant over the simulation period, ~~probably~~likely causing the discrepancies between model results and observed data. ~~An overview of the used and calibrated parameter values is given in Appendix Table A2.~~

3.3 Heat Budget Dynamics

Lake Mendota's annual stratification dynamics were characterized by a short spring mixing period followed by a very stable summer stratification period, which further promotes hypolimnetic oxygen depletion. ~~On average, Schmidt Stability peaked in July at approx. 720 J m^{-2} (Figure 5), followed by a peak in the Birgean Work at approx. 1250 J m^{-2}.~~ A low Schmidt Stability value in spring close to zero was representative of the overturn period (period I, Figure 9). During this time, the Birgean Work, as well as stored internal energy, increased rapidly, and the water column remained well oxygenated. The spring overturn period (period I) was characterized by low HBR values (ratio of St+B to B) with an average of 0.85 (MMA in Figure 10A). Low HBR denoted very unstable regimes due to an abundance of external energy compared to the required energy to keep the lake mixed. The start of the spring overturn period coincided with ice melt and open-water conditions, although in some years, the thermal structure of the lake was well mixed prior to ice off and spring overturn. May was the earliest month ~~where~~when the average HBR was above 1 (Figure 10B), which indicated that the water layers below the thermocline became isolated from the surface layers. Following period I, Schmidt Stability increased in conjunction with the spatial extent of anoxia in the lake water column (Figure 9). The heat budgets, as well as the anoxic area, peaked during this second phase (period II) and declined,

although the peak of anoxic area lagged behind the heat budget peaks. As the Schmidt Stability decreased to near zero in fall, mixing is initiated causing the water column to become oxygenated.

The stratification phase (period 2) had an average HBR value of 1.45, which indicated that an additional energy input of 45% would be needed to keep Lake Mendota isothermal during stratified summer conditions (Figure 10A). Lake Mendota's mean summer HBR value was similar to Lake Steinsfjord, Norway (max. depth 22 m, Kjensmo, (1994)) and was larger than the HBR values of the unstable lake systems Lake Marion, USA (max. depth 4.5 m) and Lake Wingra, USA (max. depth 6.1 m) (Kjensmo, 1994). The oxygenation of the water column lagged behind the stratification period, and even when the Schmidt Stability values at the end of the 2nd period were close to zero, a certain amount of the lake's area can remain anoxic.

Heat storage in Lake Mendota began after ice off and increased rapidly between the end of the mixing period and the onset of stratification (Figure 11A). The amount of internal energy stored at the beginning of stratification correlated with the maximum available amount of internal energy that will be stored during the stratified period~~.~~, despite year-to-year fluctuations in internal energy. Over the course of each year, the amount of stored internal energy was positively correlated with Schmidt Stability (Figure 11~~B). Nonetheless, the stored internal energy fluctuated year-to-year.~~B). Birgean Work was also positively correlated with both Schmidt Stability and Internal Energy. The relationship between Schmidt Stability and the spatial anoxia extent exhibited a clockwise hysteresis (Figure 11C). Beginning in June, Schmidt Stability increased as stratified conditions established in the water column. Schmidt Stability peaked on average in August~~.~~ at ~720 J m^{-2} (Figure 9), followed by a peak in the Birgean Work at ~1250 J m^{-2}. Simultaneously~~ the~~, the depth of anoxia in the water column (anoxia height) followed the progression of Schmidt Stability, ~~but peaked~~peaking in September. In Lake Mendota, the anoxia height was limited by the thermocline depth, as the low vertical turbulent diffusivity of the thermocline acted as a barrier for an encroachment of anoxic conditions into the surface mixed layer. Anoxia height decreased after September ~~with decreasing~~as Schmidt Stability ~~values~~decreased. Thermocline depth and anoxia height declined in parallel until Schmidt Stability reached zero. In Lake Mendota, as in most lakes, the surface layer was the region of significant heat storage (Figure 11D). Once stratified, heat storage in deeper water layers was limited, whereas heat in the upper 5 m of the lake increased throughout the summer, and accounted for up to 40% of the total internal energy stored during summer.

3.4 Oxygen Dynamics

Dissolved oxygen dynamics, including the spatial extent of oxygen depletion in the water column, and the timing of summer anoxia periods, were replicated by the GLM-AED model (Figure 12A-B); although the model overestimated spring and summer time surface oxygen concentrations due to a higher net ecosystem production. The depth-averaged fit criteria of dissolved oxygen concentrations were similar to a recent study from Farrell et al. (2020) in which the RMSE were 1.88 mg/L and 2.49 mg/L in the epilimnion and hypolimnion~~ were about 1.88 mg/L and 2.49 mg/L~~, respectively~~. The~~, of a GLM-AED model calibrated for Lake Mendota. Our model captured annual anoxia events in the hypolimnion (Figure 13A), and the range of the simulated Anoxic Factor was similar to the derived Anoxic Factor from observed data (Figure 13B). The model failed to replicate extreme events (~~for instance~~e.g., the very low Anoxic Factor in 2002) and did not capture a recent positive trend

of Anoxic Factor since 2010. ~~The simulated Anoxic Factor averaged 56.7 ± 5.2 days with an RMSE of 7 days (correlation coefficient r = 0.28).~~ The simulated Anoxic Factor over the total time period averaged 56.7 ± 5.2 days with an RMSE of 7 days, an NSE of -0.22, and an KGE of 0.26 (correlation coefficient r = 0.28). The model's underestimation of the recent positive trend of Anoxic Factors starting in 2010 was investigated by quantifying the fits during two periods: 1992-2009 (Figure 10C) and 2010-2005 (Figure 10D). In the pre-2010 period (1992-2009), the model achieved an RMSE of 6.79 days, an NSE of -0.25, an KGE of 0.44 and r of 0.45 for Anoxic Factor predictions. In the post-2010 period (2010-2015), the model achieved an RMSE of 8.04 days, an NSE of -31.99, an KGE of 0.21 and r of 0.62. A subsequent Wilcoxon signed-rank test highlighted, that the observed average and modelled Anoxic Factors from the pre-2010 period showed no significant differences between the two distributions, suggesting they belong to the same population (p-value = 0.13, Supplement Figure A9), whereas the distributions of observed mean Anoxic Factors and modeled ones after 2010 were significantly different (p-value = 0.032, Supplement Figure A9).

3.5 Regression Model

We included in total ~~7~~three predictors in our final multiple linear regression which were deemed important by the Boruta algorithm and ~~a~~ stepwise linear model ~~investigation~~investigations using AIC for the period 1980-2009: Schmidt Stability ~~and Birgean Work~~ during summer~~, the intensity~~ (rel. importance of ~~anoxia (maximum depth from the surface of anoxia in the water column),~~43 %), the onset date of stratification~~,~~ (rel. importance of 42 %), and gross primary production in the epilimnion~~, the water temperature at stratification onset in the hypolimnion, as well as the HBR ratio in summer (Appendix~~ (rel. importance of 15 %) (Supplement ~~Table A3~~4).

The linear model showed a good agreement between simulated and predicted Anoxic Factor (Figure 14 A, ~~Appendix~~Supplement ~~Table A3).~~4). The Anoxic Factor was positively correlated to the summer Schmidt Stability~~, the summer HBR ratio~~ (r = 0.72, Figure 14 B) and the gross primary production in the epilimnion~~.~~ (r = 0.48). It was negatively correlated to the ~~summer Birgean Work, the water temperature at~~ onset of stratification ~~onset, the maximum height of anoxia and the onset of stratification (Figure 10~~(r = -0.78, Figure 14 B).

4 Discussion

4.1 Controls of Inter-annual variability on Hypolimnetic Anoxia

Inter-annual variability in the Anoxic Factor for Lake Mendota is influenced primarily by physical processes that regulate thermal and stratification dynamics, and less so by processes that influence organic matter. The Schmidt Stability during summer (rel. importance of 43 %) as well as the timing of stratification~~, summer Schmidt Stability, and the summer ratio HBR~~ (rel. importance of 42 %) all influence ~~anoxic factor~~Anoxic Factor, and are all driven mainly by atmospheric drivers and heat convection throughout the water column. The ~~only significant~~most important predictor of ~~anoxic factor~~ Anoxic Factor directly related to biological processes is gross primary production in the epilimnion~~. Together, these variables explain 65%~~ (rel.

importance of ~~the total relative variability in Anoxic Factor (Appendix Table A3~~15 %), Supplement Table A4). For eutrophic lakes, this suggests two critical points. First, climate has direct control over lake phenology. Climate drives the timing of stratification onset and stratification strength, and that controls the year-to-year variability in Anoxic Factor. Second, biology matters, but its interannual dynamics are not that influential, at least for this eutrophic lake with a residence time greater than one year. We also acknowledge that a step change in the Anoxic Factor occurred in 2010 and was unexplained by our model. Although the cause remains unknown, the timing was coincident with large increases in the invasive zooplankton, *Bythotrephes* (Walsh et al., 2017).

4.2 Physical Control over Anoxic Factor

~~Oxygen~~Our work demonstrates that oxygen dynamics in Lake Mendota are strongly governed by the stratification strength in the water column. Snortheim et al. (2017) came to a similar conclusion~~, analyzing~~ in an analysis of Lake Mendota during a shorter time period (2007-2010), arguing that changes in the atmospheric boundary conditions - air temperature, wind speed and relative humidity - are driving changes in the hypolimnetic anoxia development of Lake Mendota. Here, we link these atmospheric drivers to changes in the water column's stratification (as quantified by Schmidt Stability and Birgean Work). Over our 37-year simulation, anoxia onset occurred in the days following stratification onset. During stratification, the establishment of a strong density gradient between the upper and the lower layers in the water column reduces vertical turbulent diffusivities and limits the downward flux of dissolved oxygen. Without any additional oxygen source (e.g., atmospheric fluxes or primary production), dissolved oxygen concentrations below the thermocline are rapidly consumed by bacterial mineralization of organic matter in the water column and sediment.

In Lake Mendota, the temporal and spatial extent of anoxia is limited by the length of the summer stratification period (e.g. onset and offset of stratification, heat storage in water column prior to stratification, see Figure 9), the stratification strength and thermocline depth (e.g. Schmidt Stability, wind shear stress, see Figure 11), respectively. The number of days between the onset of spring mixing, which begins immediately following ice off, and summer stratification, determines the maximum amount of internal energy stored in summer. An early spring overturn and a slightly later stratification start would lead to increased anoxia height in the water column, though not necessarily a higher Anoxic Factor as the duration of anoxia could be shorter. The mixing period is essentially a turning point in the year for the gradient of the internal heat accumulation, which increases rapidly following mixing. Still, as most energy is stored in a thin surface layer, short-duration extreme wind events or cold weather periods can deplete that additional stored heat before summer stratified conditions are reached. The storage of heat simultaneously increases Birgean Work, and later Schmidt Stability, increasing the resistance of the water column to mixing and limiting vertical fluxes from the epilimnion to the hypolimnion and vice versa. In summer, a higher amount of stored internal energy is also related to a higher Schmidt Stability, further increasing the spatial extent of anoxia. Ultimately, the spatial extent of anoxia is limited by the thermocline depth, as in all simulated years the anoxia height reaches a maximum during late summer when the thermocline depth was already deepening.

4.3 Biological Control over Anoxic Factor

Gross primary production (GPP) in the epilimnion prior to summer stratification is a secondary, but still important, predictor of anoxia. GPP fuels the sinking of particulate organic carbon (POC) into deeper layers before the establishment of a thermocline. In the hypolimnion, POC is readily decomposed into DOC and mineralized by bacteria in the numerical model, and reflects the dissolved oxygen volume sink. Unexpectedly, factors controlling year-to-year variation in GPP, such as external loadings of nutrients (specifically nitrogen and phosphorus), are not evident in the anoxia patterns in Lake Mendota. This is likely due to the historically high autochthony of the eutrophic lake~~,~~ (Hart, 2017), with phytoplankton blooms ~~document~~documented back to the early 1900s (Lathrop, 2007~~), and importance of autochthony over allochthony (Hart, 2017~~), thereby minimizing the need for external nutrient loads to stimulate phytoplankton production. While biological contributions to volumetric and sediment oxygen demands are well-described for a broad range of lakes (Gelda and Auer, 1996; Matzinger et al., 2010; Müller et al., 2012; Rippey and McSorley, 2009; Yuan and Jones, 2019), for eutrophic lakes, the control over available organic substrate for hypolimnetic oxygen demand may depend more on internal processing (autochthony) than external subsidies (allochthony).

Although the model replicated well the long-term DOC dynamics (Supplement Figure A8), it also overestimated surface layer dissolved oxygen concentrations compared to the observed data. This overestimation must have a concomitant increase in organic matter as a consequence of photosynthesis, and in this case in POC. Considering our proxy for the dynamics of phytoplankton biomass is reasonably well predicted (Fig. 5), this suggests our over-estimate of primary production results in increase in POC that is exported from the epilimnion to the hypolimnion. Unfortunately, we do not have observed POC to calibrate this part of the model, but we feel it is likely that our model has overestimated the contribution of primary production to hypolimnetic organic matter and subsequent oxygen depletion.

The regression models showed that variables related to load dynamics were not significant predictors of Anoxic Factor over nearly four decades. The total phosphorus and nitrogen loads, the change in dissolved oxygen during spring overturn, the temporal change in organic carbon pools, and ice duration were not found important based on the random-forest classifier. Phosphorus cycling in Lake Mendota is complex, so it may not be surprising that load dynamics in any one year are, to a certain extent, uncoupled from the hypolimnetic oxygen demand (Hanson et al., 2020). The relatively long water residence time of Lake Mendota (approx. 4 years, McDonald and Lathrop, 2017), along with the high internal phosphorus loading rate, means that external phosphorus loads represent only about 1/3 of the available phosphorus in the epilimnion (Soranno et al., 1997). Furthermore, high primary production rates that exceed the total lake mineralization, along with external loads of organic carbon, lead to a high storage of organic matter in the sediments that can likely carry over from one year to the next (Hart, 2017). In a more nutrient-poor system, the nutrient and carbon availability would likely be more important predictors.

The model replicated the ~~extreme~~ maximum anoxia event in 1998 but struggled to replicate the minimum in 2002. The discrepancies of 5-10 days between the simulated and observed range of the Anoxic Factor beginning in 2010 are related to an increased spatial as well as temporal extent of summer anoxia (~~Appendix Figure A8), which was not captured by the~~

~~model.~~ Supplement Figure A10), which was not captured by the model. This was highlighted by the statistical analysis of the pre-2010 (1992-2009) and post-2010 (2010-2015) Anoxic Factors. Prior to 2010, there were no significant differences between observed and modeled distributions (p=0.13); whereas, after 2010, the observed distribution was significantly higher than the modeled distribution (p=0.032) (Supplement Figure A9). For simplicity and due to limitations in Lake Mendota monitoring data post-2010, we focused the regression analysis of the Anoxic Factor in this study only on the pre-2010 period.

The change in Anoxic Factor post-2010 may be due to an ecosystem shift in Lake Mendota that began in 2009, when the invasive spiny water flea (*Bythothrephes longimanus*) was detected in surprisingly high densities in the lake (Walsh et al., 2016b, 2018). Spiny water flea effectively became the dominant ~~daphnia~~*Daphnia* grazer, causing historically low ~~daphnia~~*Daphnia* biomass in 2010, 2014 and 2015 (Walsh et al., 2016a) and reducing water clarity. The ~~impact of~~ spiny water flea ~~on specific phytoplankton groups~~ may have increased organic matter supply to the hypolimnion by grazing down certain phytoplankton. Mendota's *Daphnia* population historically consisted of *Daphnia pulicaria* and the smaller-bodied *Daphnia galeata mendotae*, who compete differently with spiny water flea. *D. mendotae* biomass increased in spring after the spiny water flea invasion (Walsh et al., 2017), grazing on phytoplankton and probably accelerating organic matter mineralization before stratification onset. This could be one potential cause that contributed to the increase in hypolimnetic oxygen depletion after 2010. Our GLM-AED2 model could not replicate this food web change, and subsequent shift in anoxia dynamics, due to limitations of the numerical model, i.e., GLM-AED2 had constant ecological parameters over the entire modeling period and did not have zooplankton dynamics instantiated. -We envision future monitoring and modeling studies that focus entirely on ecosystem differences and shifts between the pre-2010 and post-2010 periods of Lake Mendota.

The simple deductive model established that the volumetric oxygen sink (i.e. water column oxygen demand) is consistently higher (on average about four times higher) than the sediment oxygen sink. The volumetric sink in lakes has been found to be strongly dependent on the trophic state of the lake, whereas the sediment sink is not (Rippey and McSorley, 2009). Eutrophic lakes tend to have high ~~volume~~volumetric sinks that reach maxima of about 0.23 g m^{-3} d^{-1} (Rippey and McSorley, 2009) similar to the average ~~volume~~volumetric sink of 0.16 g m^{-3} d^{-1} quantified by the deductive model for Lake Mendota. This finding is confirmed by the works of Conway (1972~~)~~), who found that the high hypolimnetic oxygen demand of Lake Mendota was driven by algae decomposition, originating from the surface layer. Although eutrophic lakes tend to have a high sediment oxygen demand, the specific values can range from 0.3 g m^{-2} d^{-1} (Romero et al., 2004; Steinsberger et al., 2019) to extreme values of 80 g m^{-2} d^{-1} (Cross and Summerfelt, 1987), most studies measured or applied a value between 1 to 4 g m^{-2} d^{-1} (Mi et al., 2020; Veenstra and Nolen, 1991). The sediment oxygen demand calculated by our deductive model of 0.04 g m^{-2} d^{-1} was closer to the average value of approx. 0.08 g m^{-2} d^{-1} measured by Rippey and McSorley (2009) on 32 lakes. We note that the simple deductive model itself can only differentiate between two sources of depletion and neglects any physical transport drivers of oxygen, e.g., diffusion. Therefore, the results of the deductive model only add direct information to the actual depletion process of dissolved oxygen, but not of the dominant drivers.

4.4 Improving the Modeling Framework

The coupled GLM-AED2 model was able to generally replicate the thermal dynamics and biogeochemistry of Lake Mendota. In contrast to the calibration of Lake Mendota by Bruce et al. (2018) using an earlier version of GLM (v. 2.2.0), our model reproduced the water temperatures in the surface layer better than the bottom layer dynamics (RMSE for epilimnion and hypolimnion water temperatures, respectively, from Bruce et al., 2018: 1.94 °C and 1.42 °C). This is probably due to the close proximity of the atmospheric forcing boundary condition to the surface layers, whereas the energy balance approach used by GLM potentially underestimates vertical mixing, and hence overpredicts bottom layer water temperatures. In contrast, the model achieved better fits of the biogeochemical variables in the bottom layer. Better fits in the hypolimnion were likely achieved through directed calibration of sediment fluxes during the calibration-validation approach. The implementation and testing of alternative vertical mixing schemes for the Lake Mendota model (e.g. vertical mixing using a k- ε turbulence model) could potentially improve vertical transport and water temperature dynamics in deep layers. Further, using transient sediment boundary conditions with dynamic parameters over time could improve the model fit with the observed data, and could replicate potential ecosystem shifts. As the spatial extent of hypolimnetic anoxia is fundamentally three-dimensional (Biddanda et al., 2018), fully resolving anoxia in space and time likely requires a 3D model (Bocaniov and Scavia, 2016). Still, such a model has higher computational needs for long-term calibration-validation analysis, and current monitoring is inadequate to validate the results as most measurements are only made at the deepest point of the lake. Therefore, additional monitoring sites would need to be established. Improved spatial monitoring would be useful in validating our 1D approach and setting up higher dimensional numerical models.

Our ~~calibrated~~ GLM-AED2 model overestimated spring phytoplankton biomass, which resulted in an overestimation of surface dissolved oxygen concentrations. This primary overproduction is a potential source of uncertainty for the anoxia timing below the thermocline as the model's anoxia dynamics lag behind the observed ones. The time difference between the simulated and observed dissolved oxygen decline below the thermocline during stratification could be explained by an underestimation of sinking simulated organic material into the hypolimnion. Discrepancies between simulated and observed Anoxic Factors, therefore, could be rooted in our simplifications of the phytoplankton dynamics and the related organic matter fluxes, and highlight the importance of improving the representation of phytoplankton and zooplankton dynamics in numerical models. Simulating a magnitude of individual species rather than functional phytoplankton groups has been shown to improve numerical water quality and ecosystem predictions (Hellweger, 2017), though it is unclear if it could improve spring bloom predictions in Lake Mendota. Further, better numerical representations of phytoplankton life cycles (Hense, 2010; Shimoda and Arhonditsis, 2016), and/or allometric scaling (Shimoda et al., 2016) could significantly improve numerical phytoplankton predictions. It is noteworthy that biweekly monitored data of Lake Mendota required interpolation of the observed data in order to calculate the observed Anoxic Factors. This adds uncertainty to the observed Anoxic Factor, as monitoring likely missed important daily (or even sub-biweekly) fluctuations in dissolved oxygen.

It should be noted that as a statistical approach, the deductive regression model does not account for important mechanisms that may explain nonlinearities in the hypothetical linear relationships between oxygen depletion rate and the sediment to volume rate. Thus, the deductive regression model may be biased for Lake Mendota. As the model still advanced our broader system understanding by quantifying the range of the sediment oxygen demand, it was still helpful to investigate observed dissolved oxygen concentration data.

4.5 Implications for landscape and climate change

The strong relationship between anoxia and water column stability suggests that a changing climate might increase Anoxic Factor. Future climate in the region is expected to warm (Veloz et al., 2012), which may amplify and prolong water column stratification through increasing air temperatures (O'Reilly et al., 2015; Winslow et al., 2017). Shorter ice duration or even the total loss of ice (Sharma et al., 2019) could promote earlier heat storage in Lake Mendota, which could potentially increase summer Schmidt Stability, as demonstrated by ~~(~~Farrell et al~~.,~~. (2020~~)).~~). Earlier heat accumulation would cause a stability increase and an earlier onset of stratification, thereby extending the duration of anoxia. Earlier onset of anoxia may cause the anoxia height to be spatially limited by an earlier, and therefore lower, thermocline depth. Therefore, warming air temperatures will likely increase the Anoxic Factor of Lake Mendota through prolonged temporal and increased spatial extent of anoxia. It is worth noting that our initial regression quantified the correlation between Anoxic Factor and the water temperature in the hypolimnion at stratification onset ~~was~~as weakly negative. Higher water temperatures in a mixed water column prior to stratification onset are related to less stable stratified summer conditions. This feedback, potentially enhanced by shorter ice periods and warmer spring overturn periods, could shorten the extent of summer anoxia~~.~~ (similar findings were reported in Flaim et al., 2020).

Although our model evaluation supported the claim that external phosphorus loads are not important predictors of inter-annual variability in anoxia, future changes in the landscape (Motew et al., 2019), e.g. reduced agricultural application of phosphorus, less direct run-off pathways from the catchment to the lake, or more urbanization, may change these relationships. Lakes with nutrient concentrations lower than Lake Mendota would almost certainly experience higher primary production with elevated nutrient loads, and higher primary production would likely fuel higher hypolimnetic respiration (Rippey and McSorley, 2009). Thus, the link between catchment processes and lake anoxia, which was not detectable in this study, would likely be important in lakes with meso- or oligotrophic states (~~Farrell~~Ward et al., 2020). For Lake Mendota, the only reasonable management approach to reducing anoxia is to lower external nutrient loads, especially given that anoxia duration in Lake Mendota is related to thermal stratification, which is predicted to increase with future warmer air temperatures.

5 Conclusions

We presented a novel modeling framework combining three complementary approaches (deductive model, numerical GLM-AED2 model, and regression model) to conceptually identify the important drivers of year-to-year variability in the spatial and

temporal summer hypolimnetic anoxia extent of eutrophic Lake Mendota over a period spanning nearly four decades. Physical metrics – ~~onset date of stratification,~~ summer Schmidt Stability~~,~~ and ~~the ratio~~onset date of ~~St+B to B~~stratification – were the most important predictors driving the summer Anoxic Factor. Although the gross primary production was still influential in affecting year-to-year variability of hypolimnetic anoxia, biological control over the Anoxic Factor was limited in our study period. As climate change is positively correlated to lake stratification characteristics (earlier, longer and more intense summer stratification), we expect an increase in the Anoxic Factor of Lake Mendota in coming decades. The only local management option to mitigate future hypolimnetic anoxia in Lake Mendota is a reduction of external nutrient loads, which aims at shifting the lake towards oligotrophic conditions. ~~As managers and decision makers undertake forward planning to guard against a decline in lake water quality as a result of climate change, decision support tools that support an understanding of lake dynamics over the long term are essential.~~ The modeling framework developed here can be extended by an advanced sediment diagenesis model and an uncertainty analysis, e.g. Bayesian analysis, to develop greater insight into effective strategies to mitigate environmental degradation. Consequently, as managers and decision makers work to prevent a decline in lake water quality as a result of climate change, decision support tools that support an understanding of lake dynamics over the long term are essential.

Code and data availability

The data to run the 37-year simulation, including configuration files, driver data and model source code, as well as the simulation output used in this study are available at the Environmental Data Initiative (https://portal-s.edirepository.org/nis/mapbrowse?scope=knb-lter-ntl&identifier=389&revision=1, Ladwig et al., 2020).

Author Contributions

PCH, HAD, CCC, KMC and RL designed the conceptual model study and aim. HAD pre-processed the input data. RL and PCH performed the lake model simulations. YZ, LS and CD performed the catchment model simulations. RL analysed the data and prepared the manuscript with contributions from all co-authors.

Acknowledgments

The project was funded through an NSF ABI development grant (#DBI 1759865), an NSF CNH grant (#1517823), as well as NSF grants DEB 1753639 and DEB 1753657. Lake Mendota data were obtained from the North Temperate Lakes Long Term Ecological Research program (#DEB-1440297). We are thankful for supplementary dissolved oxygen field data from 1992-1994 by Patricia Soranno and insights from the CNH-Lakes team.

Competing Interests

The authors declare that they have no conflict of interest.

References

Appling, A. P., Leon, M. C. and McDowell, W. H.: Reducing bias and quantifying uncertainty in watershed flux estimates: the R package loadflex, Ecosphere, 6(12), art269, doi:10.1890/ES14-00517.1, 2015.

Bennett, E. M., Reed-Andersen, T., Houser, J. N., Gabriel, J. R. and Carpenter, S. R.: A Phosphorus Budget for the Lake Mendota Watershed, Ecosystems, 2(1), 69–75, doi:10.1007/s100219900059, 1999.

Biddanda, B. A., Weinke, A. D., Kendall, S. T., Gereaux, L. C., Holcomb, T. M., Snider, M. J., Dila, D. K., Long, S. A., VandenBerg, C., Knapp, K., Koopmans, D. J., Thompson, K., Vail, J. H., Ogdahl, M. E., Liu, Q., Johengen, T. H., Anderson, E. J. and Ruberg, S. A.: Chronicles of hypoxia: Time-series buoy observations reveal annually recurring seasonal basin-wide hypoxia in Muskegon Lake – A Great Lakes estuary, J. Gt. Lakes Res., 44(2), 219–229, doi:10.1016/j.jglr.2017.12.008, 2018.

Birge, E. A.: The work of wind in warming a lake, Transl. Wis. Acad Sciens Arts Lett., 18 Pt II, 341, 1916.

Bocaniov, S. A. and Scavia, D.: Temporal and spatial dynamics of large lake hypoxia: Integrating statistical and three-dimensional dynamic models to enhance lake management criteria: LAKE HYPOXIA: INTEGRATING MODELS TO ENHANCE LAKE MANAGEMENT, Water Resour. Res., 52(6), 4247–4263, doi:10.1002/2015WR018170, 2016.

Bossard, P. and Gächter, R.: Methan- und Sauerstoffhaushalt im mesotrophen Lungernsee, Schweiz. Z. Für Hydrol., 43(2), 219–252, doi:10.1007/BF02502135, 1981.

Brock, T. D.: A Eutrophic Lake: Lake Mendota, Wisconsin, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo., 1985.

Bruce, L. C., Frassl, M. A., Arhonditsis, G. B., Gal, G., Hamilton, D. P., Hanson, P. C., Hetherington, A. L., Melack, J. M., Read, J. S., Rinke, K., Rigosi, A., Trolle, D., Winslow, L., Adrian, R., Ayala, A. I., Bocaniov, S. A., Boehrer, B., Boon, C., Brookes, J. D., Bueche, T., Busch, B. D., Copetti, D., Cortés, A., de Eyto, E., Elliott, J. A., Gallina, N., Gilboa, Y., Guyennon, N., Huang, L., Kerimoglu, O., Lenters, J. D., MacIntyre, S., Makler-Pick, V., McBride, C. G., Moreira, S., Özkundakci, D., Pilotti, M., Rueda, F. J., Rusak, J. A., Samal, N. R., Schmid, M., Shatwell, T., Snorthheim, C., Soulignac, F., Valerio, G., van der Linden, L., Vetter, M., Vinçon-Leite, B., Wang, J., Weber, M., Wickramaratne, C., Woolway, R. I., Yao, H. and Hipsey, M. R.: A multi-lake comparative analysis of the General Lake Model (GLM): Stress-testing across a global observatory network, Environ. Model. Softw., 102, 274–291, doi:10.1016/j.envsoft.2017.11.016, 2018.

Burns, N. M.: Using hypolimnetic dissolved oxygen depletion rates for monitoring lakes, N. Z. J. Mar. Freshw. Res., 29(1), 1–11, doi:10.1080/00288330.1995.9516634, 1995.

Carpenter, S. R., Booth, E. G. and Kucharik, C. J.: Extreme precipitation and phosphorus loads from two agricultural watersheds: Extreme precipitation and phosphorus load, Limnol. Oceanogr., 63(3), 1221–1233, doi:10.1002/lno.10767, 2018.

Cole, G. and Weihe, P.: Textbook of Limnology, 5., Waveland Press, Inc., 2016.

Conway, C. J.: Oxygen depletion in the hypolimnion, M.Sc. thesis, UW-Madison, Madison., 1972.

Cornett, R. J.: Predicting changes in hypolimnetic oxygen concentrations with phosphorus retention, temperature, and morphometry, Limnol. Oceanogr., 34(7), 1359–1366, doi:10.4319/lo.1989.34.7.1359, 1989.

Cornett, R. J. and Rigler, F. H.: Hypolinimetic Oxygen Deficits: Their Prediction and Interpretation, Science, 205(4406), 580–581, doi:10.1126/science.205.4406.580, 1979.

Cross, T. K. and Summerfelt, R. C.: OXYGEN DEMAND OF LAKES: SEDIMENT AND WATER COLUMN BOD, Lake Reserv. Manag., 3(1), 109–116, doi:10.1080/07438148709354766, 1987.

Diaz, R. J. and Rosenberg, R.: Spreading Dead Zones and Consequences for Marine Ecosystems, Science, 321(5891), 926–929, doi:10.1126/science.1156401, 2008.

Duffy, C. J., Dugan, H. A. and Hanson, P. C.: The age of water and carbon in lake-catchments: A simple dynamical model: Age of water and carbon in lake-catchments, Limnol. Oceanogr. Lett., 3(3), 236–245, doi:10.1002/lol2.10070, 2018.

Farrell, K. J., Ward, N. K., Krinos, A. I., Hanson, P. C., Daneshmand, V., Figueiredo, R. J. and Carey, C. C.: Ecosystem-scale nutrient cycling responses to increasing air temperatures vary with lake trophic state, Ecol. Model., 430, 109134, doi:10.1016/j.ecolmodel.2020.109134, 2020.

Fenocchi, A., Rogora, M., Morabito, G., Marchetto, A., Sibilla, S. and Dresti, C.: Applicability of a one-dimensional coupled ecological-hydrodynamic numerical model to future projections in a very deep large lake (Lake Maggiore, Northern Italy/Southern Switzerland), Ecol. Model., 392, 38–51, doi:10.1016/j.ecolmodel.2018.11.005, 2019.

Flaim, G., Andreis, D., Piccolroaz, S. and Obertegger, U.: Ice Cover and Extreme Events Determine Dissolved Oxygen in a Placid Mountain Lake, Water Resour. Res., 56(9), doi:10.1029/2020WR027321, 2020.

Foley, B., Jones, I. D., Maberly, S. C. and Rippey, B.: Long-term changes in oxygen depletion in a small temperate lake: effects of climate change and eutrophication, Freshw. Biol., 57(2), 278–289, doi:10.1111/j.1365-2427.2011.02662.x, 2012.

Gelda, R. K. and Auer, M. T.: Development and Testing of a Dissolved Oxygen Model for a Hypereutrophic Lake, Lake Reserv. Manag., 12(1), 165–179, doi:10.1080/07438149609354006, 1996.

Grömping, U.: Relative Importance for Linear Regression in *R*: The Package **relaimpo**, J. Stat. Softw., 17(1), doi:10.18637/jss.v017.i01, 2006.

Gupta, H. V., Kling, H., Yilmaz, K. K. and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377(1–2), 80–91, doi:10.1016/j.jhydrol.2009.08.003, 2009.

Hansen, N.: The CMA Evolution Strategy: A Tutorial, ArXiv160400772 Cs Stat [online] Available from: http://arxiv.org/abs/1604.00772 (Accessed 25 November 2019), 2016.

Hanson, P. C., Bade, D. L., Carpenter, S. R. and Kratz, T. K.: Lake metabolism: Relationships with dissolved organic carbon and phosphorus, Limnol. Oceanogr., 48(3), 1112–1119, doi:10.4319/lo.2003.48.3.1112, 2003.

Hanson, P. C., Stillman, A. B., Jia, X., Karpatne, A., Dugan, H. A., Carey, C. C., Stachelek, J., Ward, N. K., Zhang, Y., Read, J. S. and Kumar, V.: Predicting lake surface water phosphorus dynamics using process-guided machine learning, Ecol. Model., 430, 109136, doi:10.1016/j.ecolmodel.2020.109136, 2020.

Hart, J.: Greenhouse Gas Formation and Organic Carbon Dynamics in a Eutrophic Lake, Master Thesis, University of Wisconsin-Madison, Madison., 2017.

Hart, J., Dugan, H., Carey, C., Stanley, E. and Hanson, P.: Lake Mendota Carbon and Greenhouse Gas Measurements at North Temperate Lakes LTER 2016 ver 19, [online] Available from: https://doi.org/10.6073/pasta/170e5ba0ed09fe3d5837ef04c47e432e, 2020.

Hellweger, F.: 75 years since Monod: It is time to increase the complexity of our predictive ecosystem models (opinion), Ecol. Model., 346, 77–87, doi:10.1016/j.ecolmodel.2016.12.001, 2017.

Hense, I.: Approaches to model the life cycle of harmful algae, J. Mar. Syst., 83(3–4), 108–114, doi:10.1016/j.jmarsys.2010.02.014, 2010.

Hipsey, M., Huang, P., Paraska, D., Busch, B. and Boon, C.: AED2 Modules: Parameters and configuration, Aquat. Ecodynamics Model. Libr. AED2 [online] Available from: http://aed.see.uwa.edu.au/research/models/AED/index.html (Accessed 19 December 2019a), 2019.

Hipsey, M. R. and Hamilton, D. P.: The Computational Aquatic Ecosystem Dynamics Model (CAEDYM): v3 Science Manual, Centre for Water Research Technical Report, Centre for Water Research, Perth, Australia., 2008.

Hipsey, M. R., Huang, P., Paraska, D., Busch, B. and Boon, C.: Aquatic Ecodynamics (AED/AED2) model library, Aquat. Ecodynamics Model. Libr. AED2 [online] Available from: https://aed.see.uwa.edu.au/research/models/aed/ (Accessed 7 October 2020), 2017.

Hipsey, M. R., Bruce, L. C., Boon, C., Busch, B., Carey, C. C., Hamilton, D. P., Hanson, P. C., Read, J. S., Sousa, E. de, Weber, M. and Winslow, L. A.: A General Lake Model (GLM 3.0) for linking with high-frequency sensor data from the Global Lake Ecological Observatory Network (GLEON), Geosci. Model Dev., 12(1), 473–523, doi:https://doi.org/10.5194/gmd-12-473-2019, 2019b.

Hutchinson, G. E.: On the Relation between the Oxygen Deficit and the productivity and Typology of Lakes, Int. Rev. Gesamten Hydrobiol. Hydrogr., 36(2), 336–355, doi:10.1002/iroh.19380360205, 1938.

Idso, S. B.: On the concept of lake stability, Limnol. Oceanogr., 18(4), 681–683, doi:10.4319/lo.1973.18.4.0681, 1973.

Jenny, J.-P., Arnaud, F., Alric, B., Dorioz, J.-M., Sabatier, P., Meybeck, M. and Perga, M.-E.: Inherited hypoxia: A new challenge for reoligotrophicated lakes under global warming: Holocene hypoxia dynamics in large lakes, Glob. Biogeochem. Cycles, 28(12), 1413–1423, doi:10.1002/2014GB004932, 2014.

Jenny, J.-P., Francus, P., Normandeau, A., Lapointe, F., Perga, M.-E., Ojala, A., Schimmelmann, A. and Zolitschka, B.: Global spread of hypoxia in freshwater ecosystems during the last three centuries is caused by rising local human pressure, Glob. Change Biol., 22(4), 1481–1489, doi:10.1111/gcb.13193, 2016a.

Jenny, J.-P., Normandeau, A., Francus, P., Taranu, Z. E., Gregory-Eaves, I., Lapointe, F., Jautzy, J., Ojala, A. E. K., Dorioz, J.-M., Schimmelmann, A. and Zolitschka, B.: Urban point sources of nutrients were the leading cause for the historical spread of hypoxia across European lakes, Proc. Natl. Acad. Sci., 113(45), 12655–12660, doi:10.1073/pnas.1605480113, 2016b.

Kara, E. L., Heimerl, C., Killpack, T., Van de Bogert, M. C., Yoshida, H. and Carpenter, S. R.: Assessing a decade of phosphorus management in the Lake Mendota, Wisconsin watershed and scenarios for enhanced phosphorus management, Aquat. Sci., 74(2), 241–253, doi:10.1007/s00027-011-0215-6, 2012.

Kjensmo, J.: Internal energy, the work of the wind, and the thermal stability in Lake Tyrifjord, southeastern Norway, Hydrobiologia, 286(1), 53–59, doi:10.1007/BF00007280, 1994.

Kursa, M. B. and Rudnicki, W. R.: Feature Selection with the Boruta Package, J. Stat. Softw., 36(11), doi:10.18637/jss.v036.i11, 2010.

Ladwig, R., Hanson, P. C., Dugan, H. A., Carey, C. C., Zhang, Y., Shu, L., Duffy, C. and Cobourn, K. M.: Lake thermal structure drives inter-annual variability in summer anoxia dynamics in a eutrophic lake over 37 years ver 1, [online] Available from: https://portal-s.edirepository.org/nis/mapbrowse?scope=knb-lter-ntl&identifier=389&revision=1 (Accessed 3 July 2020), 2020.

Lathrop, R. C.: Lake Mendota and the Yahara River chain, in Food web management: A case study of Lake Mendota, Wisconsin, pp. 17–29, Springer-Verlag, N.Y., 1992.

Lathrop, R. C.: Perspectives on the eutrophication of the Yahara lakes, Lake Reserv. Manag., 23(4), 345–365, doi:10.1080/07438140709354023, 2007.

Livingstone, D. M. and Imboden, D. M.: The prediction of hypolimnetic oxygen profiles: a plea for a deductive approach, Can. J. Fish. Aquat. Sci., 53(4), 924–932, doi:10.1139/f95-230, 1996.

Magnuson, J., Carpenter, S. and Stanley, E.: North Temperate Lakes LTER: Chemical Limnology of Primary Study Lakes: Nutrients, pH and Carbon 1981 - current ver 52, [online] Available from: https://doi.org/10.6073/pasta/8359d27bbd91028f222d923a7936077d, 2020a.

Magnuson, J., Carpenter, S. and Stanley, E.: North Temperate Lakes LTER: Physical Limnology of Primary Study Lakes 1981 - current ver 27, [online] Available from: https://doi.org/10.6073/pasta/c120b223f80c63982457a2e1e76f6038, 2020b.

Magnuson, J., Carpenter, S. and Stanley, E.: North Temperate Lakes LTER: Phytoplankton - Madison Lakes Area 1995 - current ver 28, [online] Available from: https://doi.org/10.6073/pasta/13ea8f578654493155a660ab886f695e, 2020c.

Magnuson, J., Carpenter, S. and Stanley, E.: North Temperate Lakes LTER: Secchi Disk Depth; Other Auxiliary Base Crew Sample Data 1981 - current ver 29, [online] Available from: https://doi.org/10.6073/pasta/c0b0b52c4c41446b76e14662f9a9a0ce, 2020d.

Magnuson, J. J., Kratz, T. K. and Benson: Long-term dynamics of lakes in the landscape : long-term ecological research on North Temperate Lakes, Oxford; New York: Oxford University Press. [online] Available from: https://books.google.com/books?hl=en&lr=&id=CkFuLsV5tuMC&oi=fnd&pg=PR17&dq=Long-term+dynamics+of+lakes+in+the+landscape+:+long-term+ecological+research+on+North+Temperate+Lakes&ots=T8-QMy4uNX&sig=lhoqbzi6bitAPJIGrBy-2hslDjs#v=onepage&q=Long-term%20dynamic., 2006.

Matzinger, A., Müller, B., Niederhauser, P., Schmid, M. and Wüest, A.: Hypolimnetic oxygen consumption by sediment-based reduced substances in former eutrophic lakes, Limnol. Oceanogr., 55(5), 2073–2084, doi:10.4319/lo.2010.55.5.2073, 2010.

McDonald, C. P. and Lathrop, R. C.: Seasonal shifts in the relative importance of local versus upstream sources of phosphorus to individual lakes in a chain, Aquat. Sci., 79(2), 385–394, doi:10.1007/s00027-016-0504-1, 2017.

Meding, M. E. and Jackson, L. J.: Biotic, chemical, and morphometric factors contributing to winter anoxia in prairie lakes, Limnol. Oceanogr., 48(4), 1633–1642, doi:10.4319/lo.2003.48.4.1633, 2003.

Mi, C., Shatwell, T., Ma, J., Wentzky, V. C., Boehrer, B., Xu, Y. and Rinke, K.: The formation of a metalimnetic oxygen minimum exemplifies how ecosystem dynamics shape biogeochemical processes: A modelling study, Water Res., 175, 115701, doi:10.1016/j.watres.2020.115701, 2020.

Mitchell, K. E.: The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system, J. Geophys. Res., 109(D7), D07S90, doi:10.1029/2003JD003823, 2004.

Morris, M. D.: Factorial Sampling Plans for Preliminary Computational Experiments, Technometrics, 33(2), 161–174, doi:10.1080/00401706.1991.10484804, 1991.

Motew, M., Chen, X., Carpenter, S. R., Booth, E. G., Seifert, J., Qiu, J., Loheide, S. P., Turner, M. G., Zipper, S. C. and Kucharik, C. J.: Comparing the effects of climate and land use on surface water quality using future watershed scenarios, Sci. Total Environ., 693, 133484, doi:10.1016/j.scitotenv.2019.07.290, 2019.

Müller, B., Bryant, L. D., Matzinger, A. and Wüest, A.: Hypolimnetic Oxygen Depletion in Eutrophic Lakes, Environ. Sci. Technol., 120824131307006, doi:10.1021/es301422r, 2012.

Müller, B., Steinsberger, T., Schwefel, R., Gächter, R., Sturm, M. and Wüest, A.: Oxygen consumption in seasonally stratified lakes decreases only below a marginal phosphorus threshold, Sci. Rep., 9(1), 18054, doi:10.1038/s41598-019-54486-3, 2019.

North, R. P., North, R. L., Livingstone, D. M., Köster, O. and Kipfer, R.: Long-term changes in hypoxia and soluble reactive phosphorus in the hypolimnion of a large temperate lake: consequences of a climate regime shift, Glob. Change Biol., 20(3), 811–823, doi:10.1111/gcb.12371, 2014.

Nürnberg, G.: Quantifying anoxia in lakes, Limnol. Oceanogr., 40(6), 1100–1111, doi:https://doi.org/10.4319/lo.1995.40.6.1100, 1995a.

Nürnberg, G. K.: The Anoxic Factor, a Quantitative Measure of Anoxia and Fish Species Richness in Central Ontario Lakes, Trans. Am. Fish. Soc., 124(5), 677–686, doi:10.1577/1548-8659(1995)124<0677:TAFAQM>2.3.CO;2, 1995b.

Nürnberg, G. K.: Quantified hypoxia and anoxia in lakes and reservoirs, ScientificWorldJournal, 4, 42–54, doi:10.1100/tsw.2004.5, 2004.

Odum, H. T.: Primary Production in Flowing Waters, Limnol. Oceanogr., 1(2), 102–117, doi:10.4319/lo.1956.1.2.0102, 1956.

O'Reilly, C. M., Sharma, S., Gray, D. K., Hampton, S. E., Read, J. S., Rowley, R. J., Schneider, P., Lenters, J. D., McIntyre, P. B., Kraemer, B. M., Weyhenmeyer, G. A., Straile, D., Dong, B., Adrian, R., Allan, M. G., Anneville, O., Arvola, L., Austin, J., Bailey, J. L., Baron, J. S., Brookes, J. D., de Eyto, E., Dokulil, M. T., Hamilton, D. P., Havens, K., Hetherington, A. L., Higgins, S. N., Hook, S., Izmest'eva, L. R., Joehnk, K. D., Kangur, K., Kasprzak, P., Kumagai, M., Kuusisto, E., Leshkevich, G., Livingstone, D. M., MacIntyre, S., May, L., Melack, J. M., Mueller-Navarra, D. C., Naumenko, M., Noges, P., Noges, T., North, R. P., Plisnier, P.-D., Rigosi, A., Rimmer, A., Rogora, M., Rudstam, L. G., Rusak, J. A., Salmaso, N., Samal, N. R., Schindler, D. E., Schladow, S. G., Schmid, M., Schmidt, S. R., Silow, E., Soylu, M. E., Teubner, K., Verburg, P., Voutilainen, A., Watkinson, A., Williamson, C. E. and Zhang, G.: Rapid and highly variable warming of lake surface waters around the globe, Geophys. Res. Lett., 42(24), 10,773-10,781, doi:10.1002/2015GL066235, 2015.

Qu, Y. and Duffy, C. J.: A semidiscrete finite volume formulation for multiprocess watershed simulation: MULTIPROCESS WATERSHED SIMULATION, Water Resour. Res., 43(8), doi:10.1029/2006WR005752, 2007.

Read, J. S., Hamilton, D. P., Jones, I. D., Muraoka, K., Winslow, L. A., Kroiss, R., Wu, C. H. and Gaiser, E.: Derivation of lake mixing and stratification indices from high-resolution lake buoy data, Environ. Model. Softw., 26(11), 1325–1336, doi:10.1016/j.envsoft.2011.05.006, 2011.

Read, J. S., Winslow, L. A., Hansen, G. J. A., Van Den Hoek, J., Hanson, P. C., Bruce, L. C. and Markfort, C. D.: Simulating 2368 temperate lakes reveals weak coherence in stratification phenology, Ecol. Model., 291, 142–150, doi:10.1016/j.ecolmodel.2014.07.029, 2014.

Rhodes, J., Hetzenauer, H., Frassl, M. A., Rothhaupt, K.-O. and Rinke, K.: Long-term development of hypolimnetic oxygen depletion rates in the large Lake Constance, Ambio, 46(5), 554–565, doi:10.1007/s13280-017-0896-8, 2017.

Rippey, B. and McSorley, C.: Oxygen depletion in lake hypolimnia, Limnol. Oceanogr., 54(3), 905–916, doi:10.4319/lo.2009.54.3.0905, 2009.

Romero, J. R., Antenucci, J. P. and Imberger, J.: One- and three-dimensional biogeochemical simulations of two differing reservoirs, Ecol. Model., 174(1–2), 143–160, doi:10.1016/j.ecolmodel.2004.01.005, 2004.

Saltelli, A., Tarantola, S., Campolongo, F. and Ratto, M.: Sensitivity Analysis in Practice: A Guide to Assessing Scientific Methods, John Wiley and Sons, Ltd., 2004.

Sánchez-España, J., Mata, M. P., Vegas, J., Morellón, M., Rodríguez, J. A., Salazar, Á., Yusta, I., Chaos, A., Pérez-Martínez, C. and Navas, A.: Anthropogenic and climatic factors enhancing hypolimnetic anoxia in a temperate mountain lake, J. Hydrol., 555, 832–850, doi:10.1016/j.jhydrol.2017.10.049, 2017.

Schmidt, W.: Über die Temperatur- und Stabilitätsverhältnisse von Seen, Geogr. Ann., 10, 145, doi:10.2307/519789, 1928.

Sharma, S., Blagrave, K., Magnuson, J. J., O'Reilly, C. M., Oliver, S., Batt, R. D., Magee, M. R., Straile, D., Weyhenmeyer, G. A., Winslow, L. and Woolway, R. I.: Widespread loss of lake ice around the Northern Hemisphere in a warming world, Nat. Clim. Change, 9(3), 227–231, doi:10.1038/s41558-018-0393-5, 2019.

Shimoda, Y. and Arhonditsis, G. B.: Phytoplankton functional type modelling: Running before we can walk? A critical evaluation of the current state of knowledge, Ecol. Model., 320, 29–43, doi:10.1016/j.ecolmodel.2015.08.029, 2016.

Shimoda, Y., Rao, Y. R., Watson, S. and Arhonditsis, G. B.: Optimizing the complexity of phytoplankton functional group modeling: An allometric approach, Ecol. Inform., 31, 1–17, doi:10.1016/j.ecoinf.2015.11.001, 2016.

Snortheim, C. A., Hanson, P. C., McMahon, K. D., Read, J. S., Carey, C. C. and Dugan, H. A.: Meteorological drivers of hypolimnetic anoxia in a eutrophic, north temperate lake, Ecol. Model., 343, 39–53, doi:10.1016/j.ecolmodel.2016.10.014, 2017.

Soranno, P. A., Carpenter, S. R. and Lathrop, R. C.: Internal phosphorus loading in Lake Mendota: response to external loads and weather, Can. J. Fish. Aquat. Sci., 54(8), 1883–1893, doi:10.1139/f97-095, 1997.

Stanley, E. H., Collins, S. M., Lottig, N. R., Oliver, S. K., Webster, K. E., Cheruvelil, K. S. and Soranno, P. A.: Biases in lake water quality sampling and implications for macroscale research: Biases in studying and monitoring lakes, Limnol. Oceanogr., 64(4), 1572–1585, doi:10.1002/lno.11136, 2019.

Steinsberger, T., Müller, B., Gerber, C., Shafei, B. and Schmid, M.: Modeling sediment oxygen demand in a highly productive lake under various trophic scenarios, edited by W. Duan, PLOS ONE, 14(10), e0222318, doi:10.1371/journal.pone.0222318, 2019.

Veenstra, J. and Nolen, S.: In-Situ sediment oxygen demand in five Southwestern U.S. lakes, Water Res., 25(3), 351–354, doi:10.1016/0043-1354(91)90016-J, 1991.

Veloz, S., Williams, J. W., Lorenz, D., Notaro, M., Vavrus, S. and Vimont, D. J.: Identifying climatic analogs for Wisconsin under 21st-century climate-change scenarios, Clim. Change, 112(3–4), 1037–1058, doi:10.1007/s10584-011-0261-z, 2012.

Walsh, J. R., Carpenter, S. R. and Vander Zanden, M. J.: Invasive species triggers a massive loss of ecosystem services through a trophic cascade, Proc. Natl. Acad. Sci., 113(15), 4081–4085, doi:10.1073/pnas.1600366113, 2016a.

Walsh, J. R., Munoz, S. E. and Vander Zanden, M. J.: Outbreak of an undetected invasive species triggered by a climate anomaly, Ecosphere, 7(12), doi:10.1002/ecs2.1628, 2016b.

Walsh, J. R., Lathrop, R. C. and Vander Zanden, M. J.: Invasive invertebrate predator, Bythotrephes longimanus, reverses trophic cascade in a north-temperate lake: Invasive invertebrate reverses cascade, Limnol. Oceanogr., 62(6), 2498–2509, doi:10.1002/lno.10582, 2017.

Walsh, J. R., Lathrop, R. C. and Zanden, M. J. V.: Uncoupling indicators of water quality due to the invasive zooplankter, Bythotrephes longimanus, Limnol. Oceanogr., 63(3), 1313–1327, doi:10.1002/lno.10773, 2018.

Ward, N. K., Steele, B. G., Weathers, K. C., Cottingham, K. L., Ewing, H. A., Hanson, P. C. and Carey, C. C.: Differential Responses of Maximum Versus Median Chlorophyll-a to Air Temperature and Nutrient Loads in an Oligotrophic Lake Over 31 Years, Water Resour. Res., 56(7), doi:10.1029/2020WR027296, 2020.

Weng, W., Boyle, K. J., Farrell, K. J., Carey, C. C., Cobourn, K. M., Dugan, H. A., Hanson, P. C., Ward, N. K. and Weathers, K. C.: Coupling Natural and Human Models in the Context of a Lake Ecosystem: Lake Mendota, Wisconsin, USA, Ecol. Econ., 169, 106556, doi:10.1016/j.ecolecon.2019.106556, 2020.

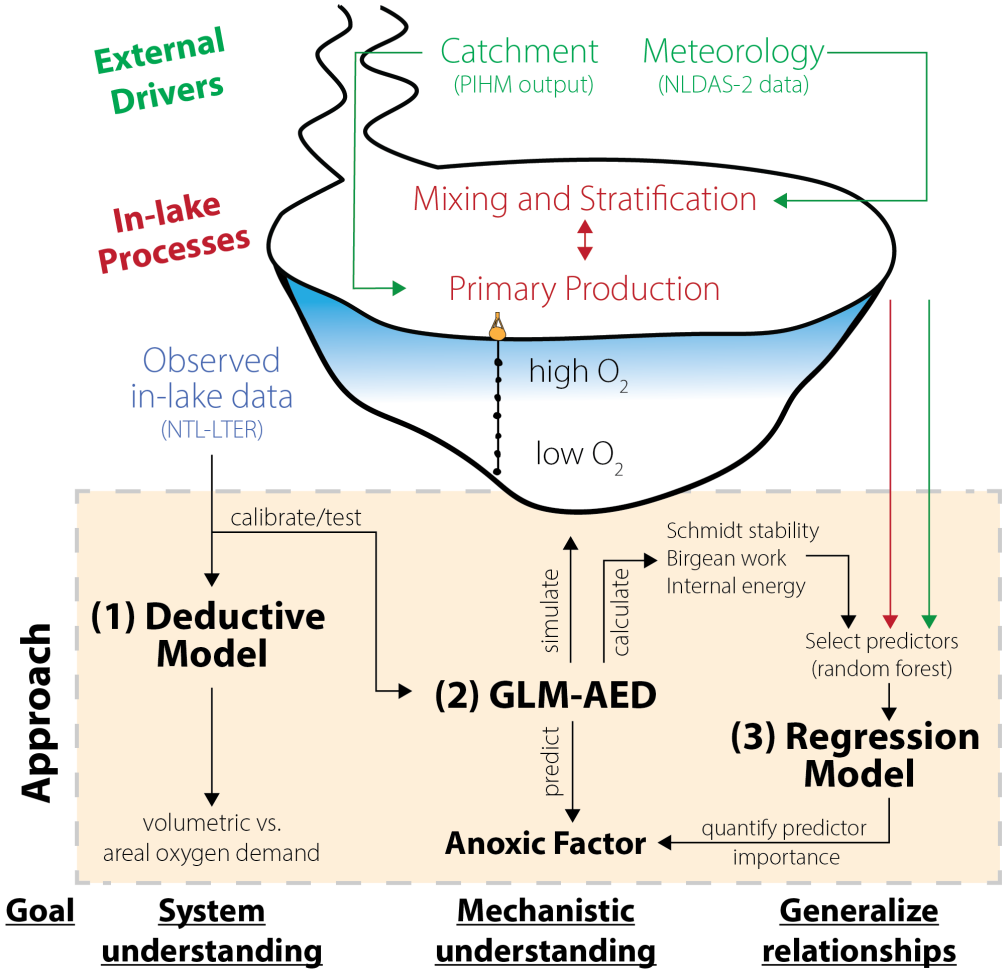
Winslow, L. A., Read, J. S., Hansen, G. J. A., Rose, K. C. and Robertson, D. M.: Seasonality of change: Summer warming rates do not fully represent effects of climate change on lake temperatures: Seasonal heterogeneity in lake warming, Limnol. Oceanogr., 62(5), 2168–2178, doi:10.1002/lno.10557, 2017.

Winslow, L. A., Read, J. S., Woolway, R. I., Brentrup, T., Leach, H., Zwart, J. A., Snortheim, C., Albers, S. and Collinge, D.: rLakeAnaluyzer: Lake physics tool. [online] Available from: doi:10.5281/zenodo.58411, 2019.

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D., Koren, V., Duan, Q., Mo, K., Fan, Y. and Mocko, D.: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products: WATER AND ENERGY FLUX ANALYSIS, J. Geophys. Res. Atmospheres, 117(D3), n/a-n/a, doi:10.1029/2011JD016048, 2012.

Yin, Y., Zhang, Y., Liu, X., Zhu, G., Qin, B., Shi, Z. and Feng, L.: Temporal and spatial variations of chemical oxygen demand in Lake Taihu, China, from 2005 to 2009, Hydrobiologia, 665(1), 129–141, doi:10.1007/s10750-011-0610-y, 2011.

Yuan, L. L. and Jones, J. R.: Modeling hypolimnetic dissolved oxygen depletion using monitoring data, Can. J. Fish. Aquat. Sci., 1–10, doi:10.1139/cjfas-2019-0294, 2019.


External Drivers
Catchment
(PIHM output)
Meteorology
(NLDAS-2 data)
In-lake Processes
Mixing and Stratification
Primary Production
Observed in-lake data
(NTL-LTER)
high O_2
low O_2
calibrate/test
Schmidt stability
Birgean work
Internal energy
simulate
calculate
Select predictors
(random forest)
(1) Deductive Model
(2) GLM-AED
(3) Regression Model
predict
volumetric vs.
areal oxygen demand
Anoxic Factor
quantify predictor
importance
Approach
Goal
System understanding
Mechanistic understanding
Generalize relationships

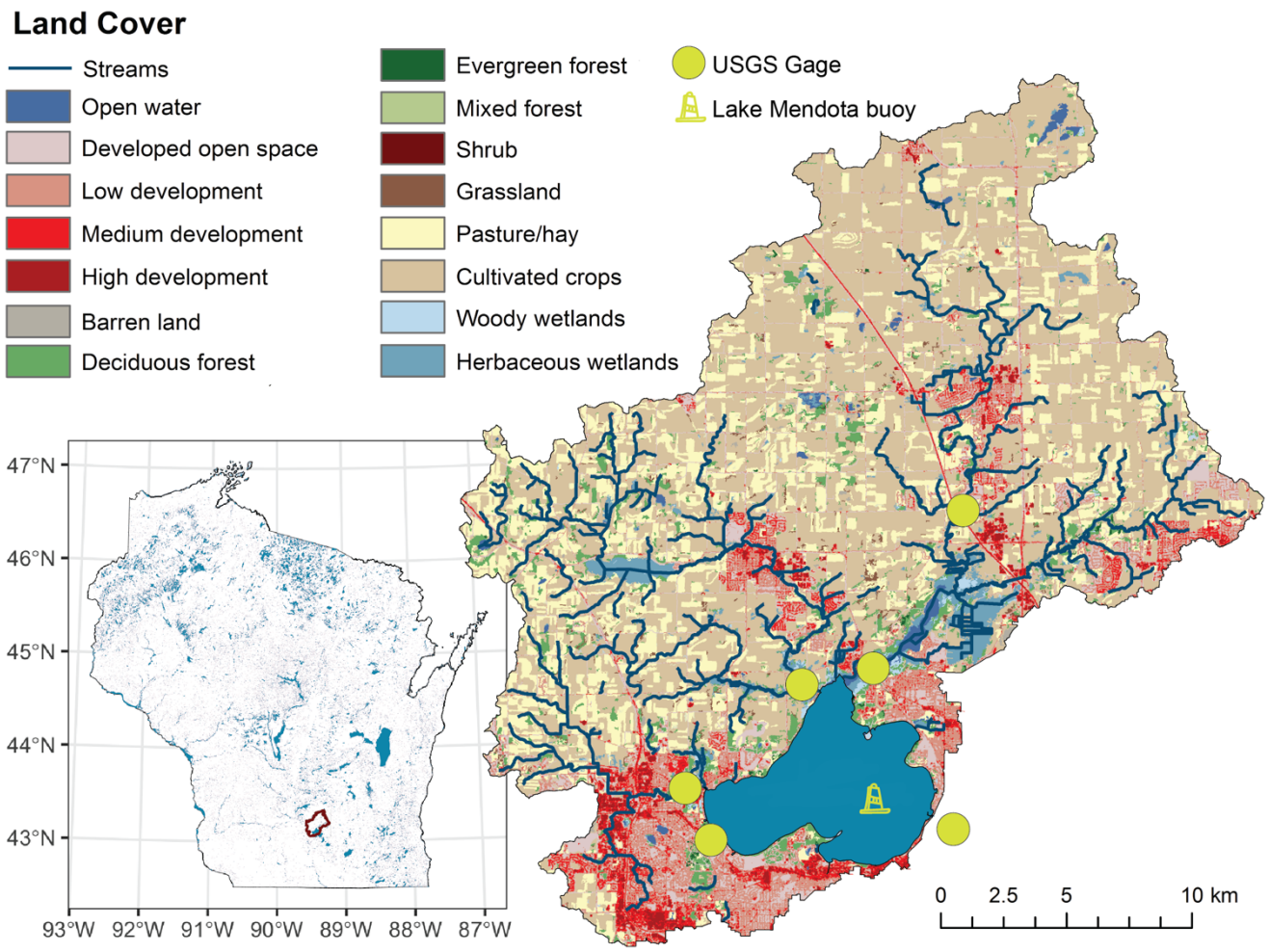


Figure 4 Location and overview map of Lake Mendota, Wisconsin, which is located in the Yahara River catchment in southern Wisconsin, USA. USGS gage stations for the PIHM-Lake model and the location of the Lake Mendota monitoring buoy are placed in the map. Land cover was obtained from the US National Land Cover database.

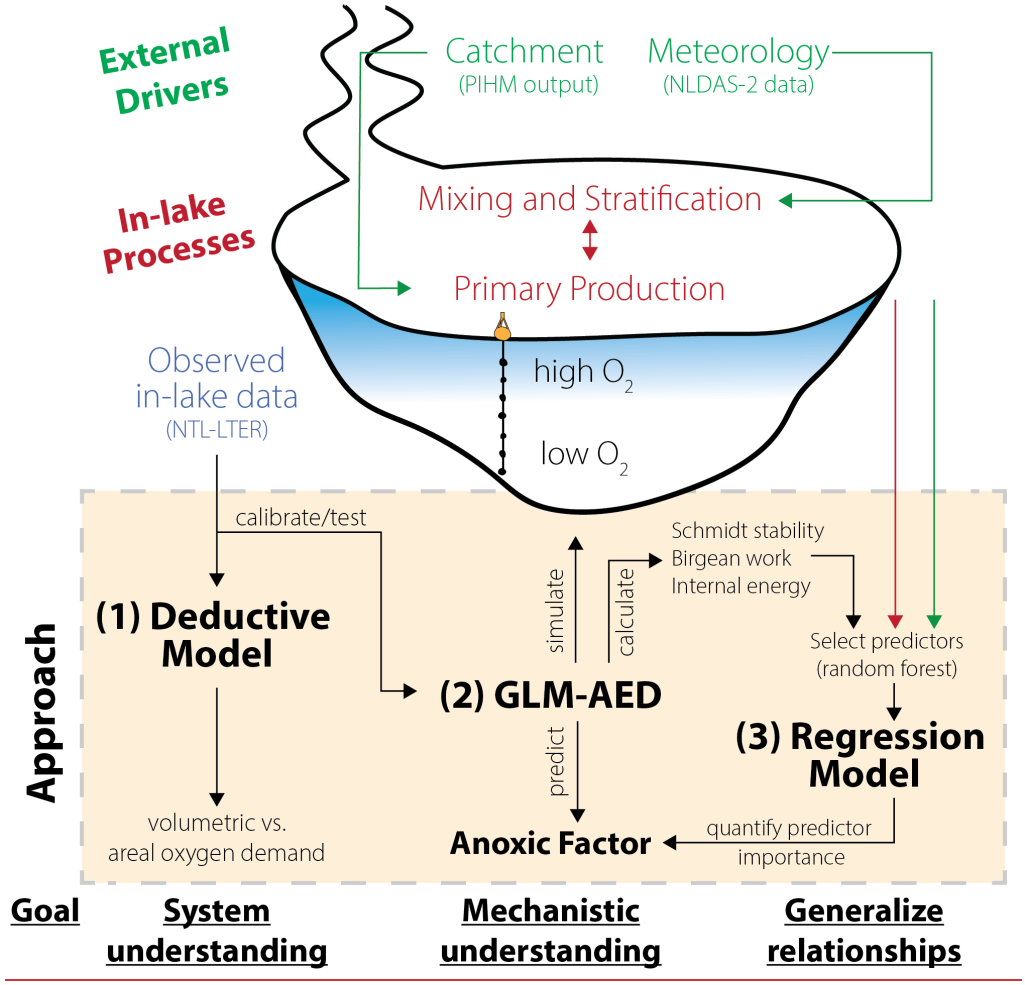


Figure 5 Schematic overview of the modeling workflow: application of a (1) deductive model to further our system understanding about oxygen sink terms; replication of Lake Mendota using (2) GLM-AED2 to investigate hydrodynamic and ecosystem mechanisms; and application of (3) regression models to quantify the importance of ecosystem predictors on the Anoxic Factor.

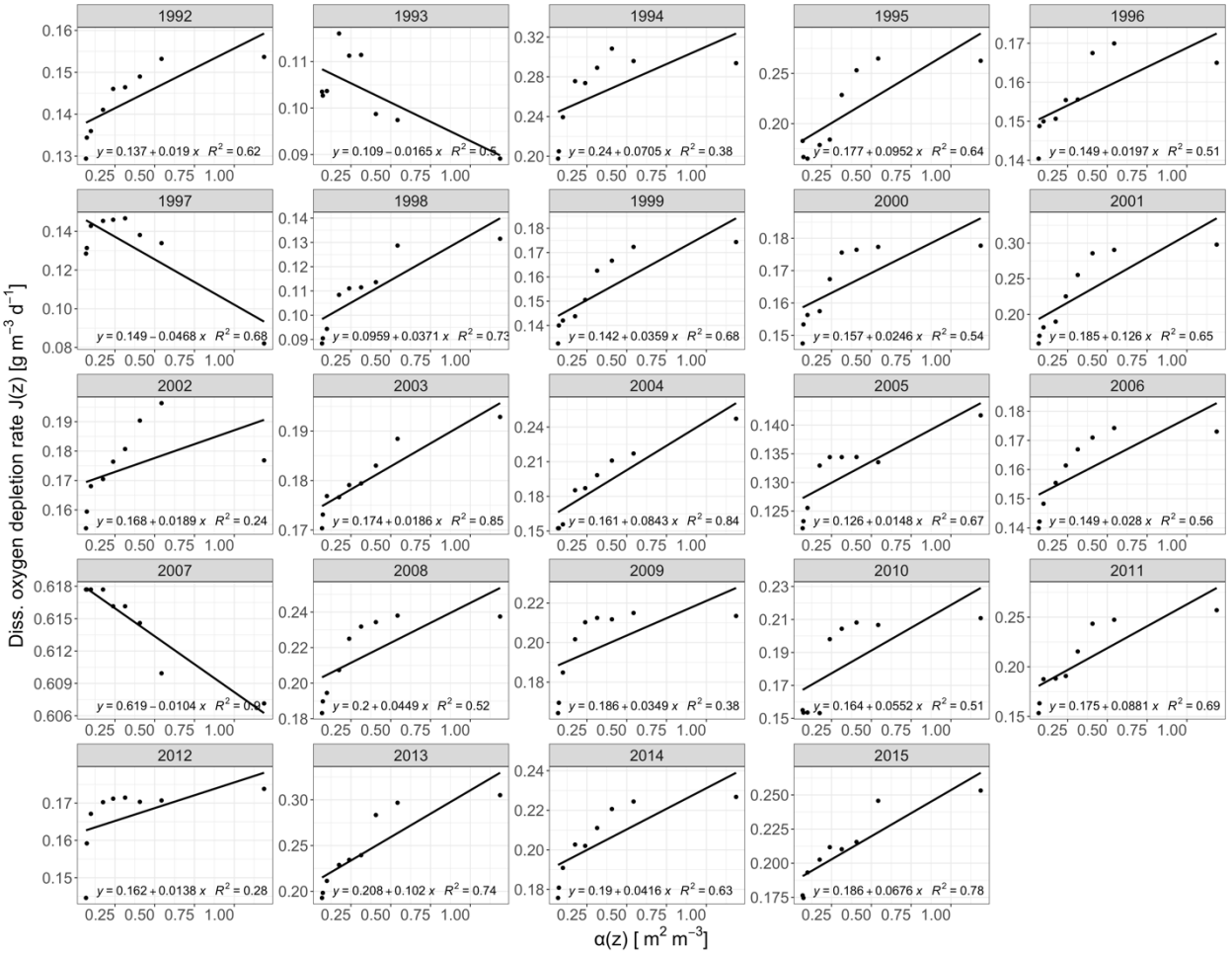


Figure 6 Regression plots of the morphometric function α(z) against oxygen depletion rates for the years 1992 to 2018, which were calculated from temporally linearly interpolated observed data. The respective equations represent weighted linear regressions.

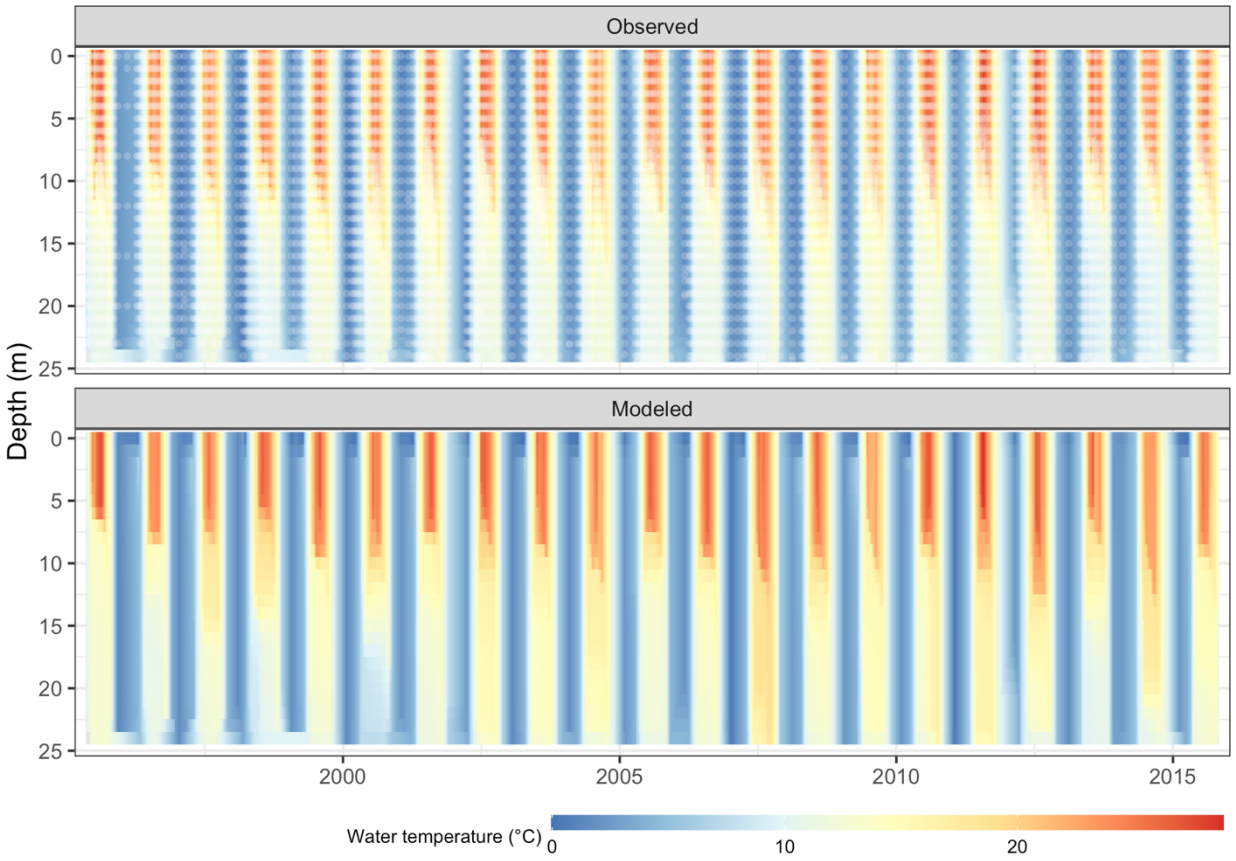
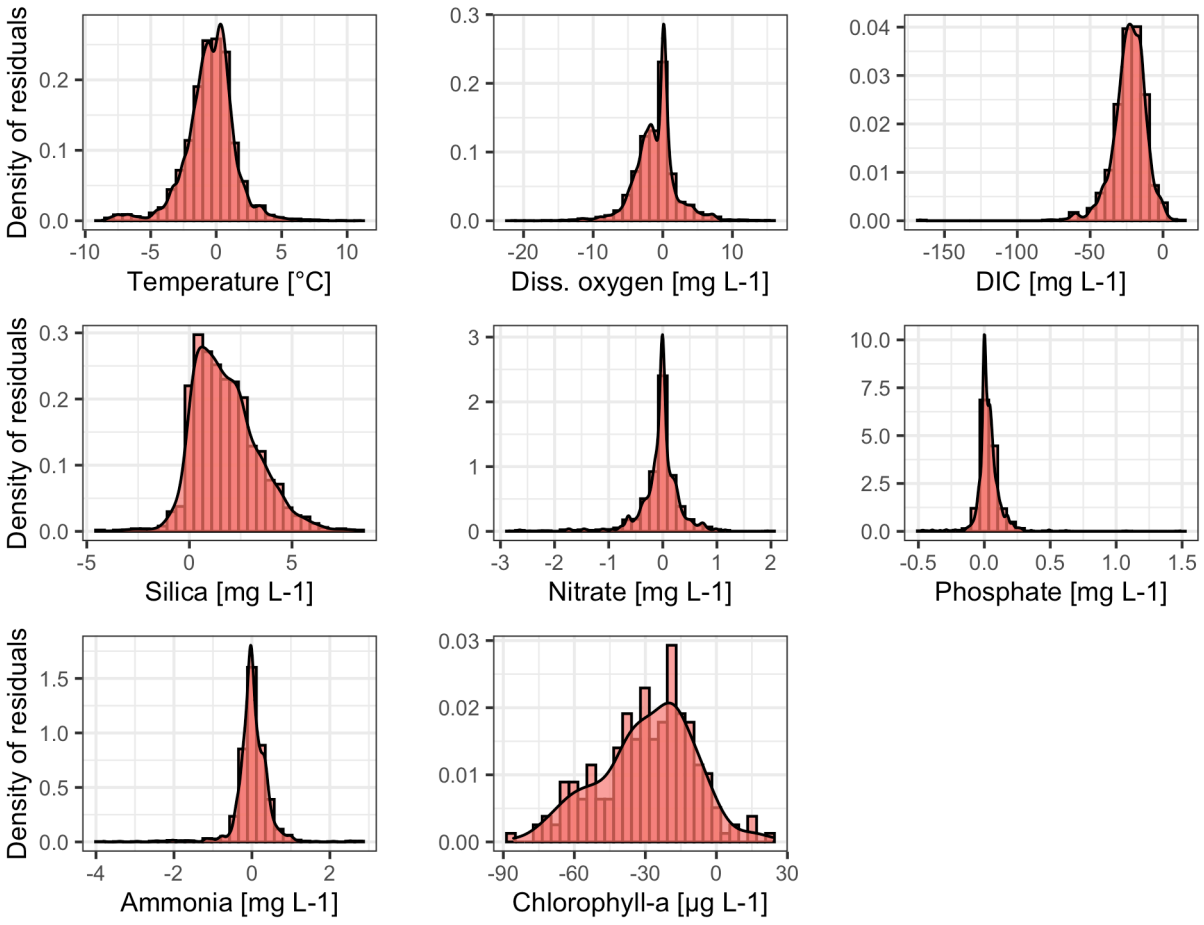


Figure 7~~ Water~~ Comparison of Lake Mendota water temperature from ~~observation~~observations (upper ~~figure~~panel, white dots mark ~~sample~~sampling events) and GLM-AED2 ~~simulation~~simulations (lower panel).


Density of residuals
Temperature [°C]
Diss. oxygen [mg L-1]
DIC [mg L-1]
Silica [mg L-1]
Nitrate [mg L-1]
Phosphate [mg L-1]
Ammonia [mg L-1]
Chlorophyll-a [µg L-1]

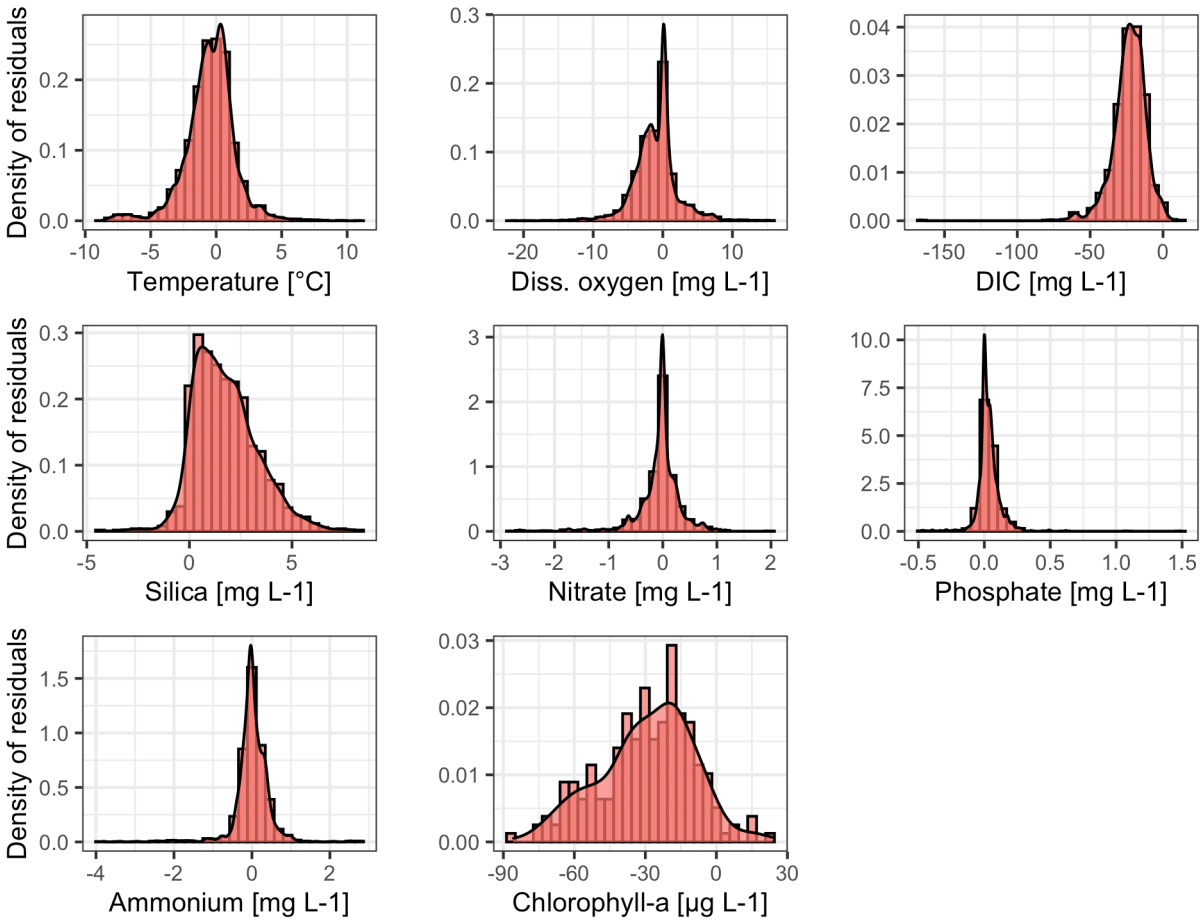


Figure 8 Density distributions of residuals (observed - modeled) for water temperature, dissolved oxygen, dissolved inorganic carbon (DIC), silica, nitrate, phosphate, ~~ammonia~~ammonium and phytoplankton. The density distributions include residuals over all data points (over each time-step over each depth), calculated from observations minus model predictions.

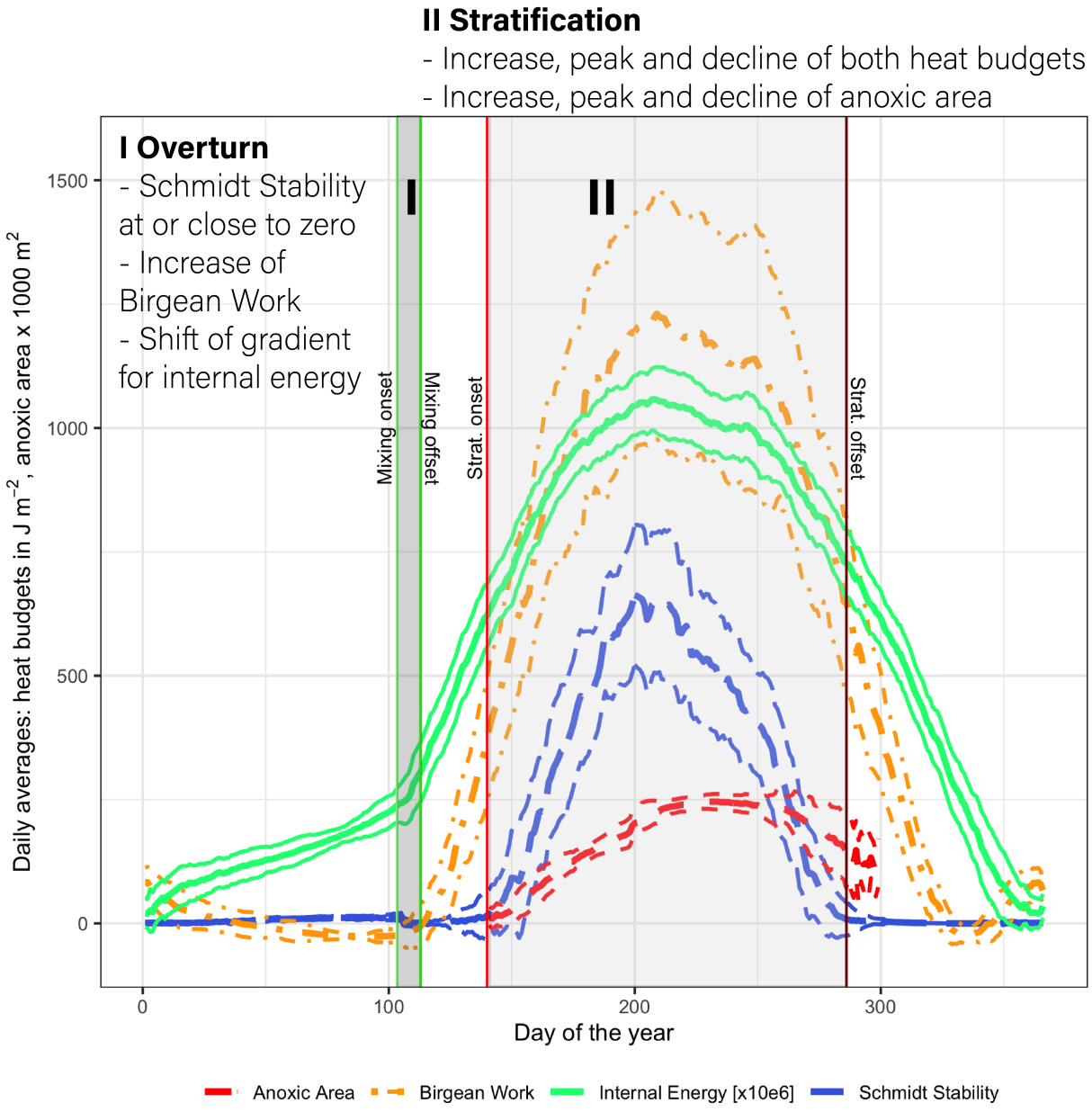


Figure 9 Daily average values of Schmidt Stability, Birgean Work, internal energy and anoxic area (below 1 mg L^{-1}) plus/minus the respective standard deviations (dashed lines) (internal energy is given in 10e6 J m^{-2}). [Anoxic area units were adjusted for display ~~purposes~~.]

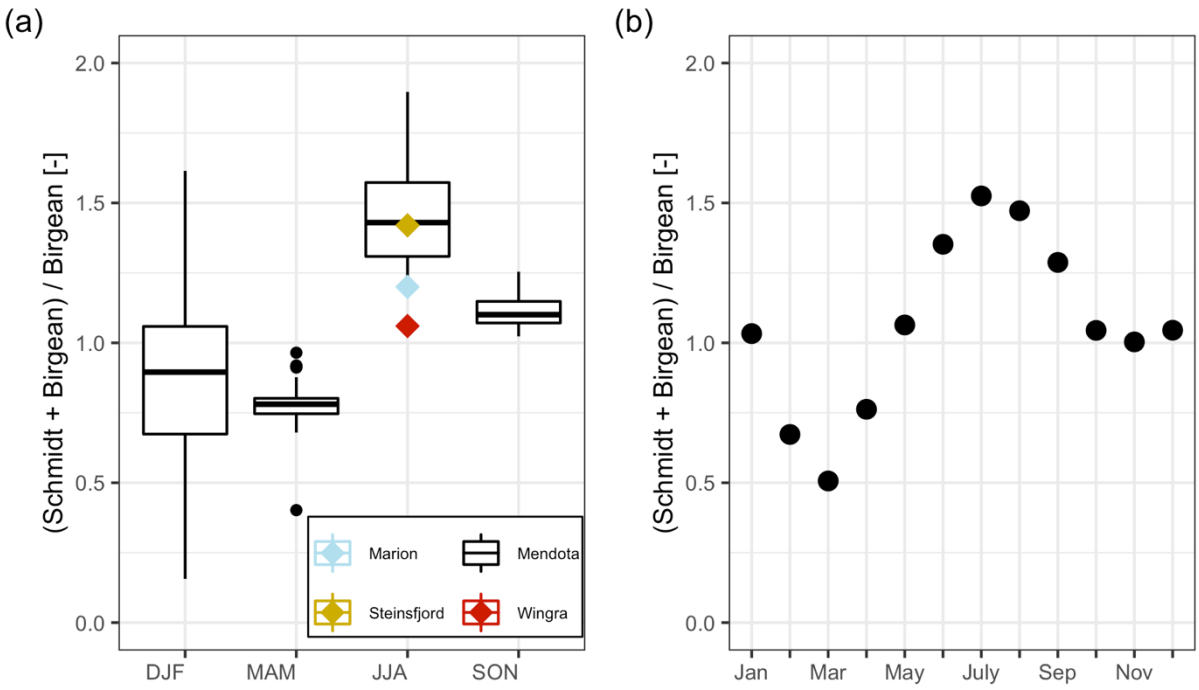
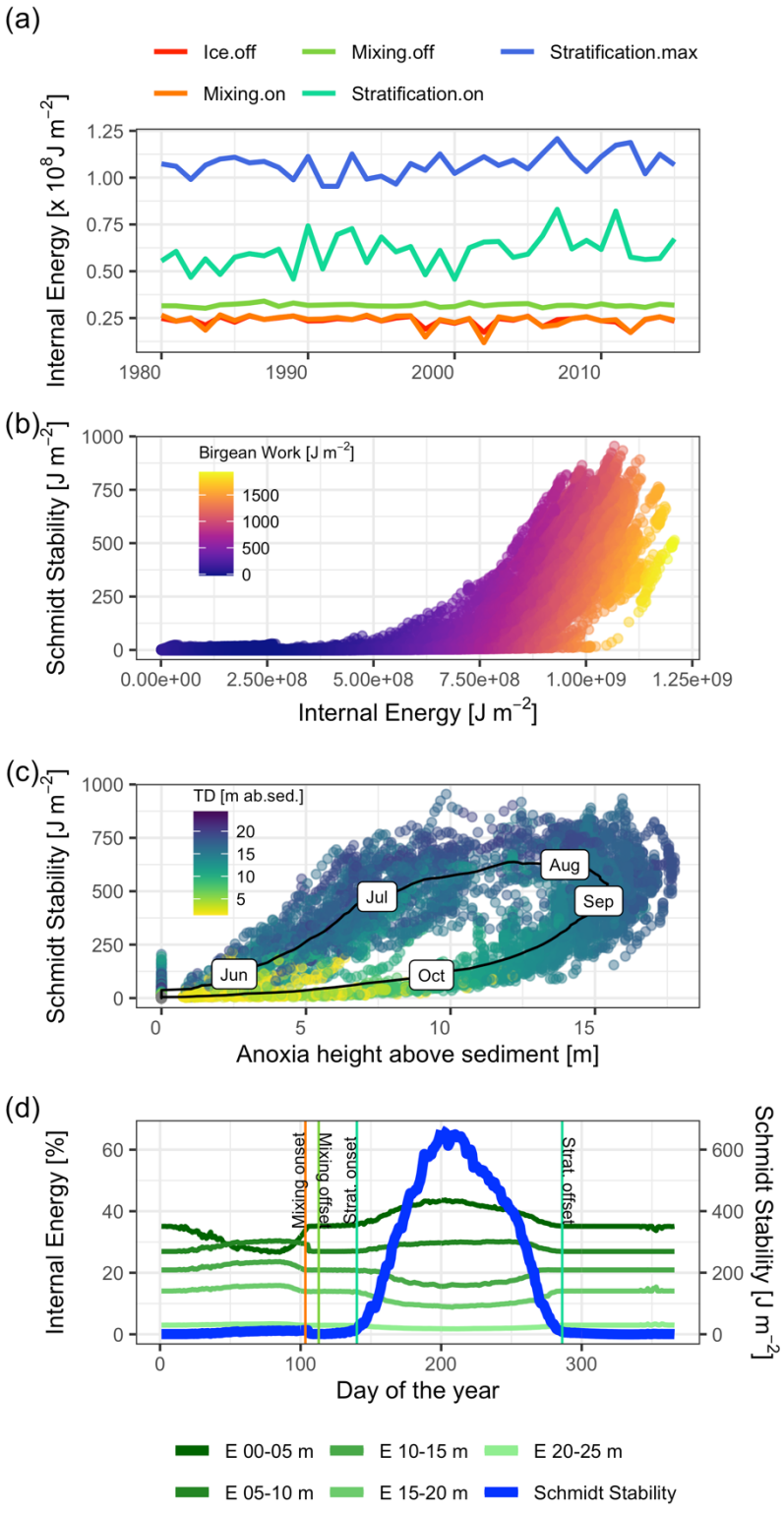


Figure 10 Dynamics of average HBR = (St+B)/B over time. A ~~box~~: Box-plots of HBR over the meteorological seasons, which represent seasonal quarters of the year, beginning in December. The summer HBR values for Lake Marion, Lake Steinsfjord, and Lake Wingra were taken from Kjensmo (1994). ~~B Scatter plot~~B: Scatterplot of average HBR values for each month


(a)
Ice.off
Mixing.off
Stratification.max
Mixing.on
Stratification.on
Internal Energy [x 10^8 J m^-2]
1.25
1.00
0.75
0.50
0.25
1980
1990
2000
2010
(b)
Schmidt Stability [J m^-2]
Birgean Work [J m^-2]
1500
1000
500
0
1000
750
500
250
0
0.00e+00
2.50e+08
5.00e+08
7.50e+08
1.00e+09
1.25e+09
Internal Energy [J m^-2]
(c)
Schmidt Stability [J m^-2]
TD [m ab.sed.]
20
15
10
5
Jun
Jul
Aug
Sep
Oct
1000
750
500
250
0
0
5
10
15
Anoxia height above sediment [m]
(d)
Internal Energy [%]
Schmidt Stability [J m^-2]
Mixing onset
Mixing offset
Strat. onset
Strat. offset
60
40
20
0
600
400
200
0
0
100
200
300
Day of the year
E 00-05 m
E 10-15 m
E 20-25 m
E 05-10 m
E 15-20 m
Schmidt Stability

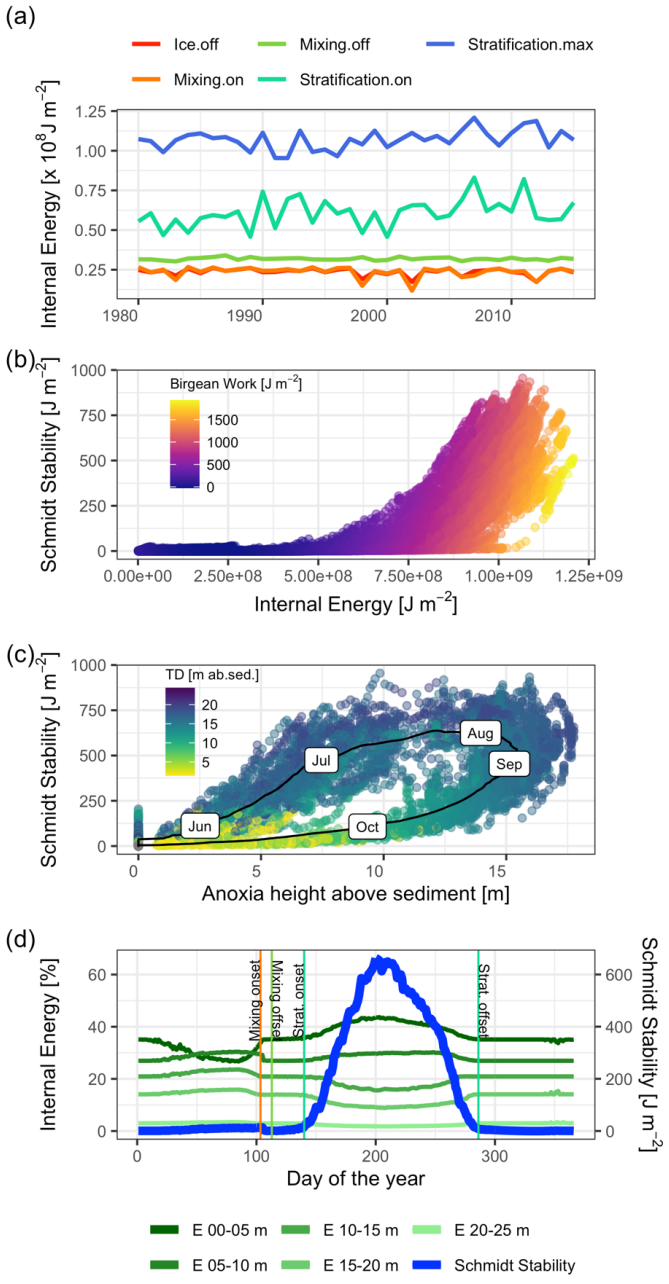


Figure 11 Stored heat dynamics and relationships to stratification strength, thermocline depth and anoxia height~~.~~ (hypsographic height of anoxia area in the lake above sediment). A: Time series of internal energy at the respective dates of ice off, mixing onset, mixing offset, stratification onset, and stratification offset. B: Scatter plot of internal energy against Schmidt Stability whereas the color represents the magnitude of Birgean Work. C: Scatter plot of anoxia height against Schmidt Stability. The black line represents the average dynamic over the course of a year with the respective months as labels. The color corresponds to the thermocline depth in meters above the sediment. D: Time series of daily averaged internal energies stored over different depths and Schmidt Stability). The main heat storage happens in a shallow surface layer effectively after ice off and the simultaneous onset of a mixing period.

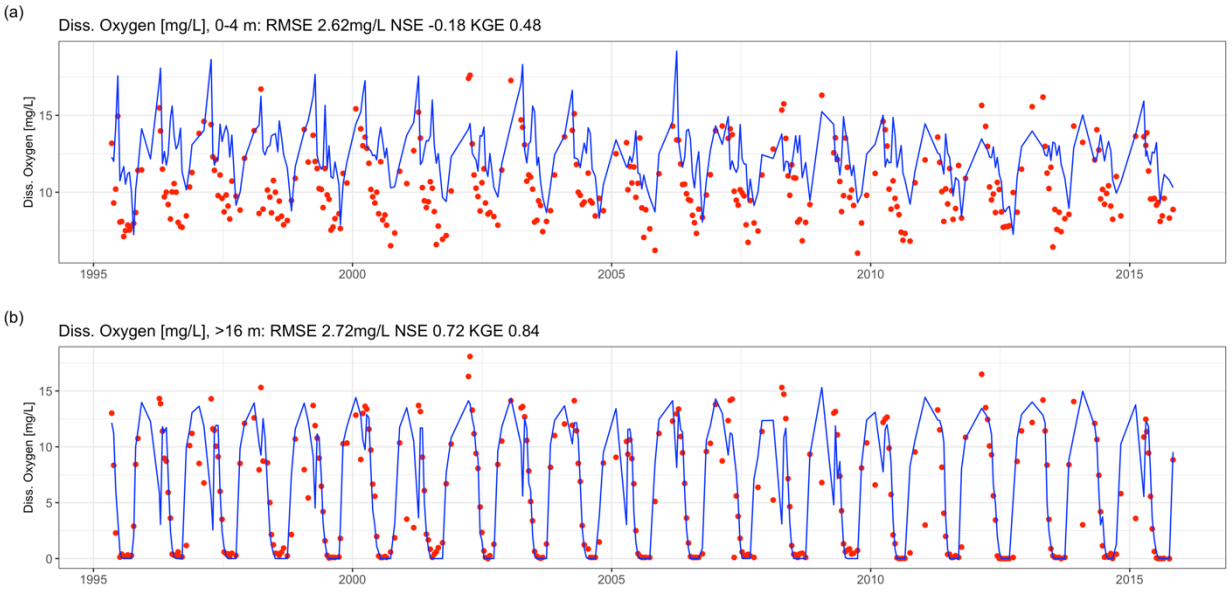
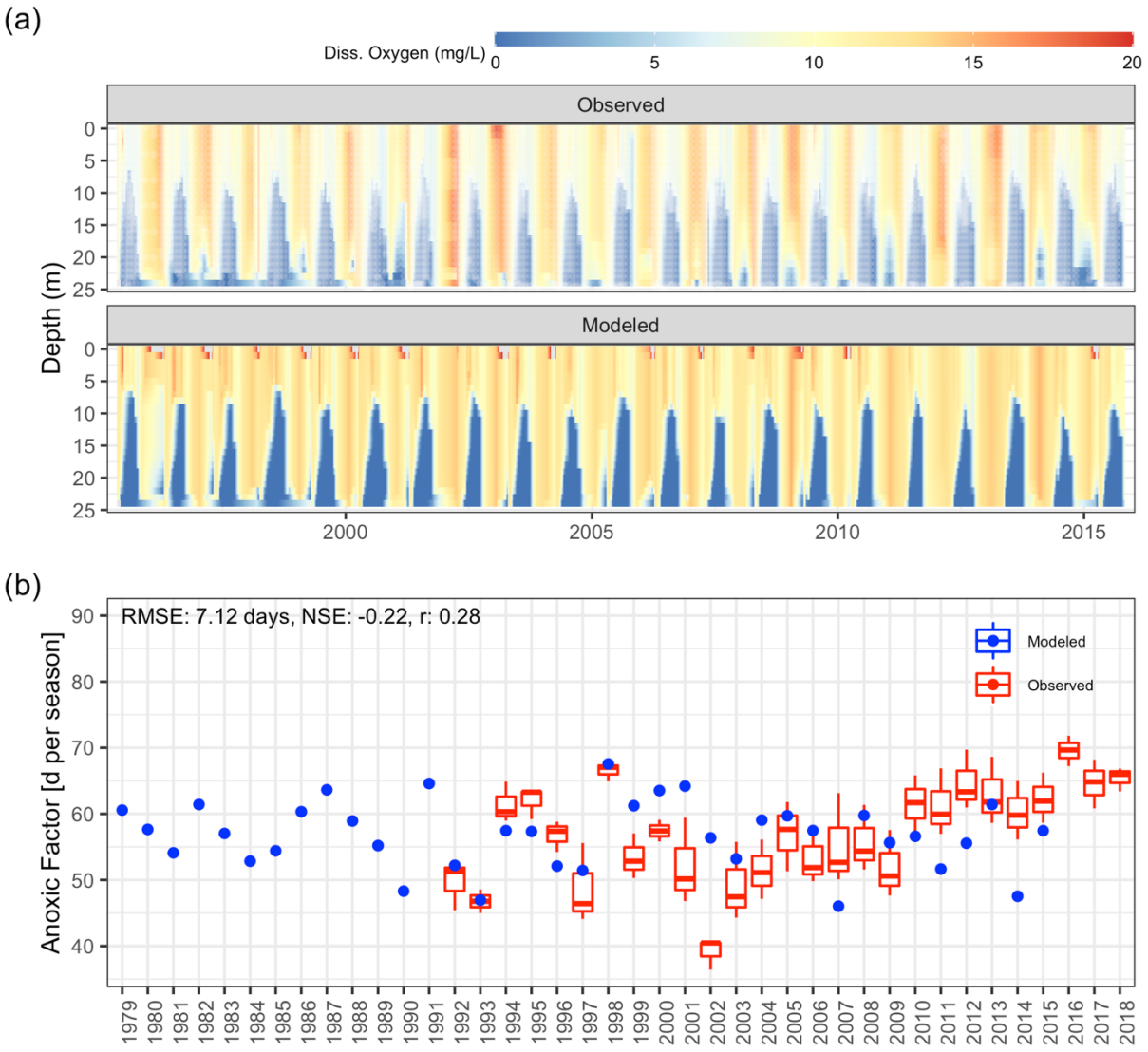


Figure 12 Time-series comparison between observed (red dots) and modeled dissolved oxygen concentrations (blue lines). The fit criteria root-mean square error (RMSE), Nash-Sutcliffe coefficient of efficiency (NSE) and Kling-Gupta coefficient of efficiency (KGE). A: averaged dissolved oxygen concentrations in the epilimnion (0-4 m). B: averaged dissolved oxygen concentrations in the hypolimnion (deeper than 16 m).


(a)
Diss. Oxygen (mg/L)
0
5
10
15
20
Observed
Modeled
Depth (m)
0
5
10
15
20
25
2000
2005
2010
2015
(b)
RMSE: 7.12 days, NSE: -0.22, r: 0.28
Modeled
Observed
Anoxic Factor [d per season]
90
80
70
60
50
40
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018

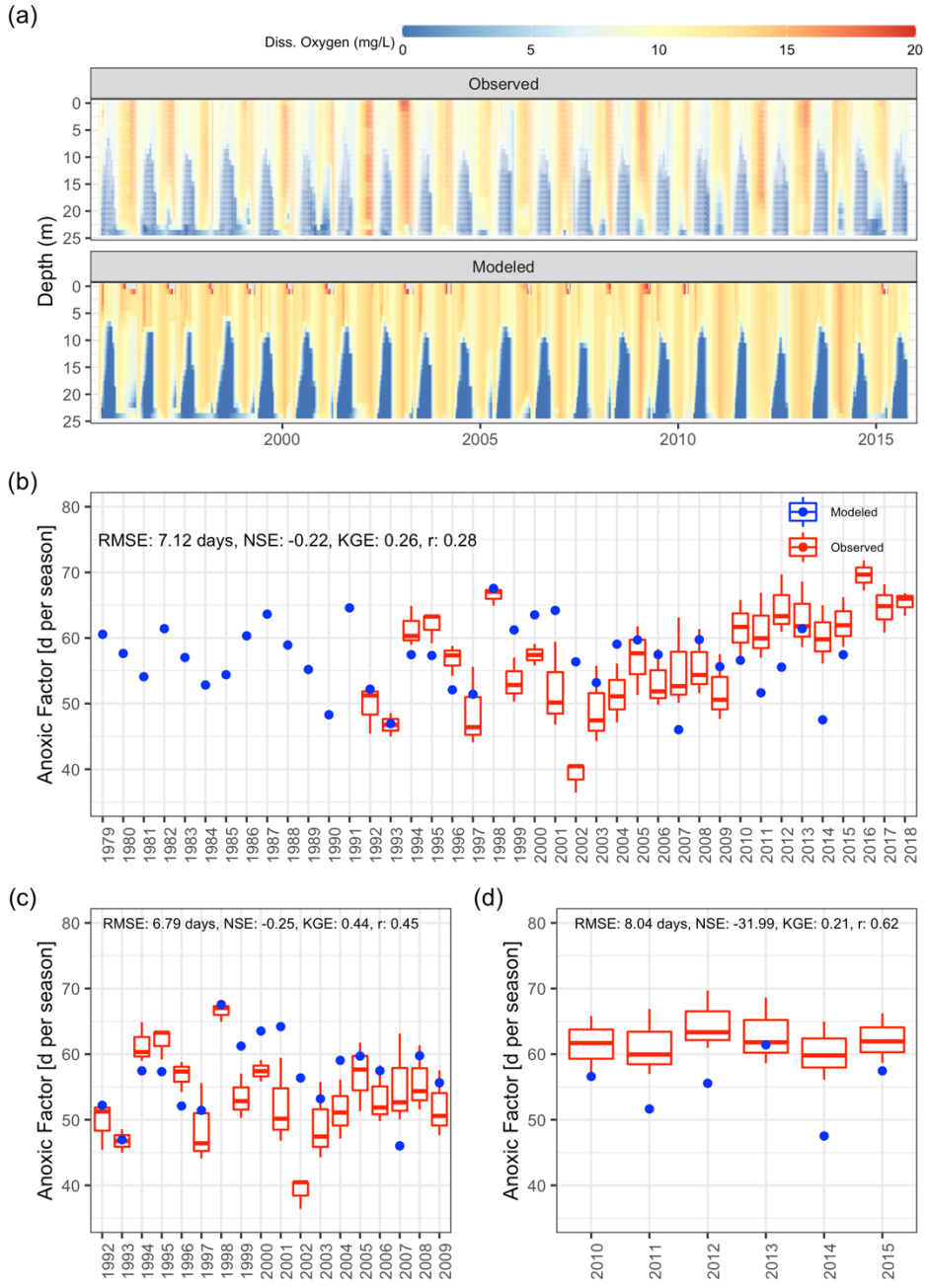
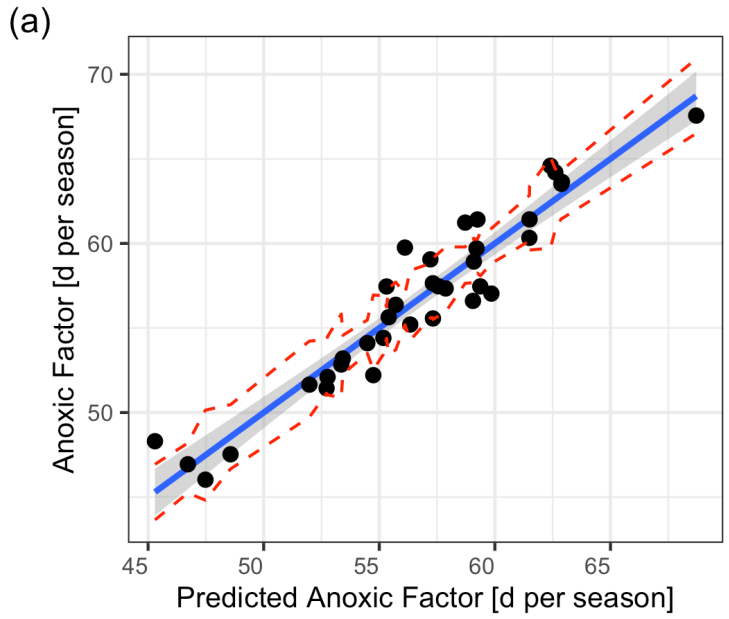
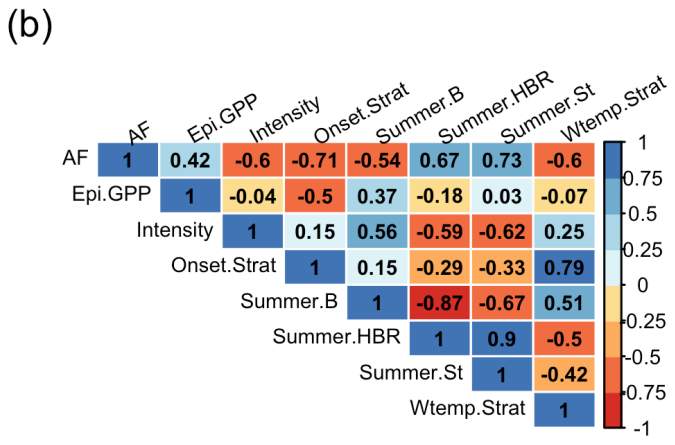
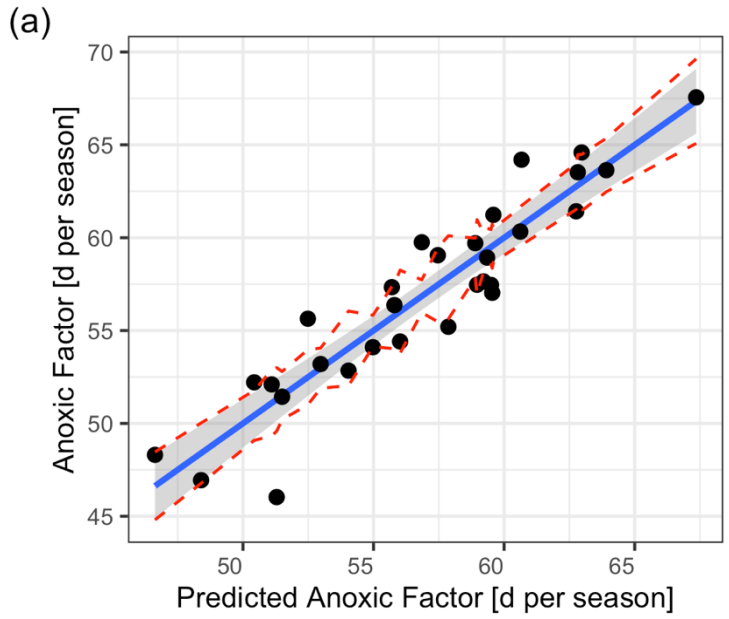


Figure 13 Comparison of ~~observed~~observations to GLM-AED modeled dissolved oxygen concentrations and ecosystem response. A: **Contour plot of observed (upper figure, white dots mark sample events) and simulated dissolved oxygen concentrations.** ~~B~~B: **Comparison of simulated Anoxic Factor (red dots) against interpolated range of Anoxic Factor derived from observed data (box-whisker plots) over the period 1979 to 2018.** C: **Comparison of simulated Anoxic Factor (red dots) against interpolated range of Anoxic Factor derived from observed data (box-whisker plots) over the period 1992 to 2009.** D: **Comparison of simulated Anoxic**

Factor (red dots) against interpolated range of Anoxic Factor derived from observed data (box-whisker plots) over the period 2010 to 2015.







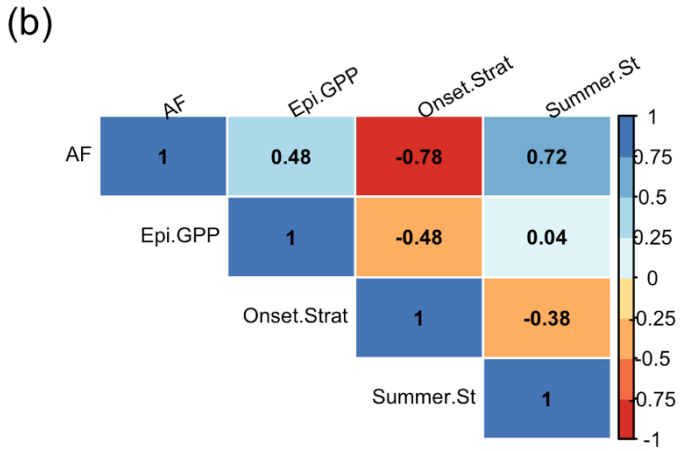


Figure 14 Predicted against simulated summer Anoxic Factor. A: Linear model with a prediction which was done using a multiple linear regression model of the form: $\hat{y} = -= 0.47Summer.HBR - 0.15Intensity + 0.29Wtemp.Strat + 0.23Epi24Epi.GPP + 0.53Summer54Summer.St - 0.64Summer.B - 0.66Onset46Onset.Strat - 1.04 * 10^{-15}5.44 * 10^{-17} + \epsilon$, where $\epsilon \sim N(0, 34^2)$. $(0, 38^2)$. The red lines represent confidence intervals. B: Correlogram of the input data using Pearson correlation coefficients.

Tables

Table 1 Overview of investigated predictors in a linear regression model on estimating the Anoxic Factor

Candidate predictor	Temporal period	Method	Unit
Schmidt Stability	Summer of year n	See section '2.3.3 Post-~~processing~~Processing of ~~model~~GLM-AED2 output'	J per m^2
Schmidt Stability	Spring of year n	See section '2.3.3 Post-~~processing~~Processing of ~~model~~GLM-AED2 output'	J per m^2
Birgean Work	Summer of year n	See section ~~'~~''2.3.3 Post-~~processing~~Processing of ~~model~~GLM-AED2 output'	J per m^2
Birgean Work	Spring of year n	See section '2.3.3 Post-~~processing~~Processing of ~~model~~GLM-AED2 output'	J per m^2
Ratio HBR	Summer of year n	See section '2.3.3 Post-~~processing~~Processing of ~~model~~GLM-AED2 output'	-
Ratio HBR	Spring of year n	See section '2.3.3 Post-~~processing~~Processing of ~~model~~GLM-AED2 output'	-
Onset, end and duration of spring mixing	Spring/summer of year n	Determined by Schmidt Stability values (close to or at zero, indicating mixed conditions)	Day of the year, or days
Onset, end and duration of summer stratification	Spring/summer/fall of year n	Stratification was defined when density difference between surface and bottom layer was above 0.1 kg m^{-3} and surface temperature was above 4 °C	Day of the year, or days
~~Maximum height above sediment of anoxia~~	~~Summer of year n~~	~~Extracted from simulation output~~	~~m above sediment~~
End and duration of ice period	Winter/spring of year n-1 and n	Extracted from simulation output	Day of the year, or days
Dissolved oxygen difference between mixing end and mixing onset in the lower layer	Spring/summer of year n	Extracted from simulation output	mmol O_2 per m^2 d

Dissolved organic carbon gradient between stratification onset and mixing onset in the lower layer	Spring/summer of year n	Extracted from simulation output	mmol C per m^2 d
Particulate organic carbon gradient between stratification onset and mixing onset in the lower layer	Spring/summer of year n	Extracted from simulation output	mmol C per m^2 d
Mean water temperature in the lower layer at the onset of stratification	Summer of year n	Extracted from simulation output	°C
Total phosphorus inflow ~~concentration~~loading	Winter/spring/summer of year n-1 and n	Extracted from driver data	g P per day per m^2
Total nitrogen inflow ~~concentration~~loading	Winter/spring/summer of year n-1 and n	Extracted from driver data	g N per day per m^2
Cumulative gross primary production in the upper layer	Winter/spring of year n-1 and n	Extracted from simulation output	mmol per m^2 per day
Cumulative gross primary production in the lower layer	Winter/spring of year n-1 and n	Extracted from simulation output	mmol per m^2 per day

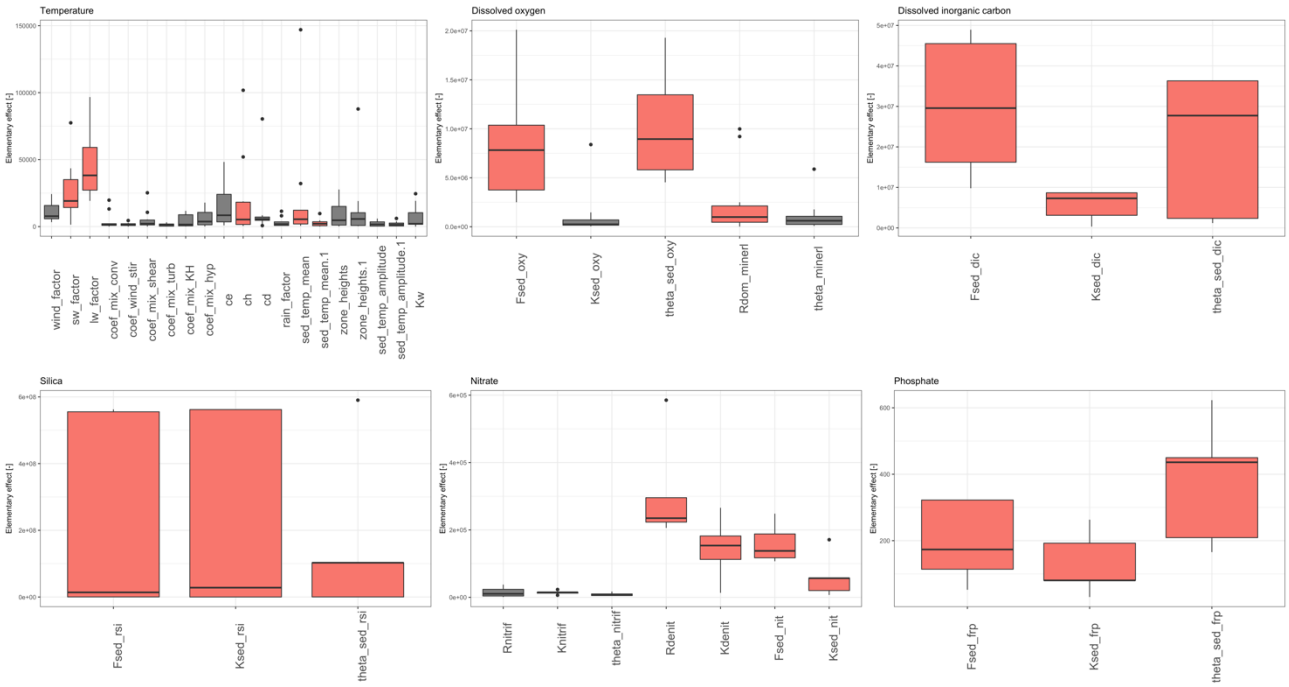
Table 2 Model performance for water temperature, dissolved oxygen, dissolved inorganic carbon, silica, nitrate, ~~ammonia~~ammonium, and phosphate. During calibration and validation, only the total fits over all depths and time-steps were calculated. Surface layers refers to a depth of 0 m below water table, and bottom layer to a depth of 20 m below water table. Fits for surface and bottom layer during calibration and validation are not shown as the fit over the whole water column and over time were used.

Variable	Calibration (2005 – 2015)			Validation (1995 – 2004)			Total period (1995 – 2015)		
	RMSE	NSE	KGE	RMSE	NSE	KGE	RMSE	NSE	KGE
Water temperature [°C]	2.26	0.87	0.92	1.78	0.92	0.96	1.96	0.91	0.94
surface layer							1.30	0.97	0.97
bottom layer							2.43	0.20	0.71
Dissolved oxygen [mg L^{-1}]	3.33	0.54	0.76	3.29	0.53	0.76	3.22	0.56	0.77
surface layer							2.77	-0.36	0.46
bottom layer							3.31	0.64	0.81
Dissolved inorganic carbon [mg L^{-1}]	25.87	-7.03	0.18	15.41	-8.65	0.20	25.92	-10.13	0.20
surface layer							19.71	-10.79	0.22
bottom layer							29.19	-12.84	0.15
Silica [mg L^{-1}]	2.83	-1.32	-4.55	1.42	-0.55	-1.77	2.33	-0.83	-3.10

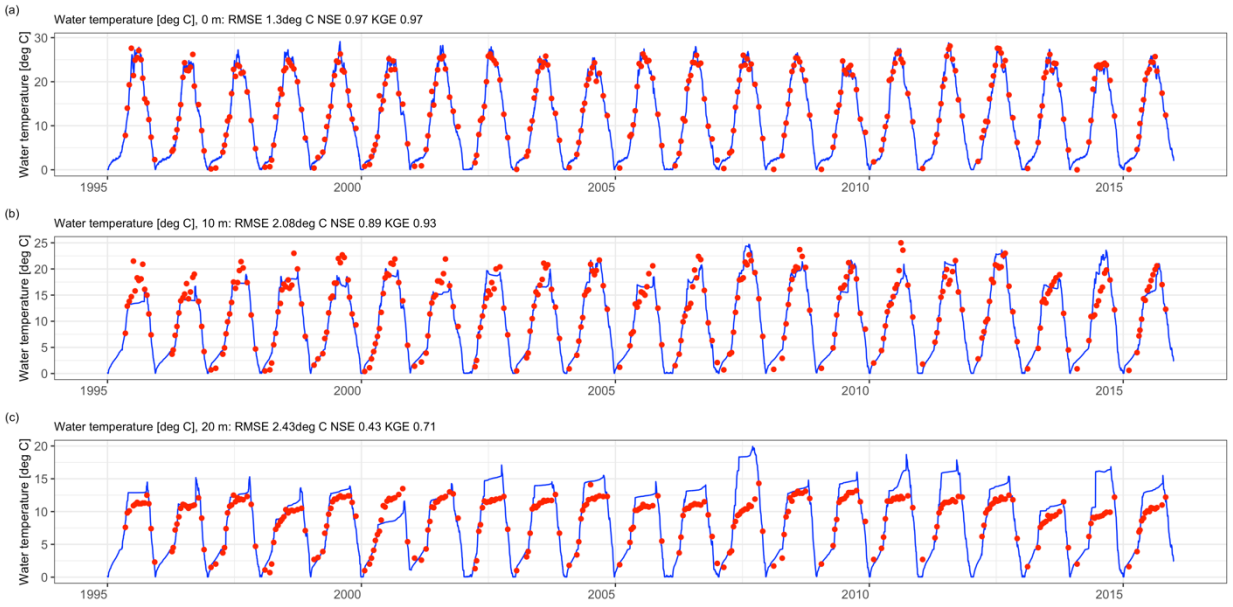
surface layer							1.61	-1.32	-6.53
bottom layer							2.78	-0.97	-0.90
Nitrate [mg L^{-1}]	0.36	-0.01	0.56	0.45	-1.44	0.26	0.40	-0.40	0.44
surface layer							0.35	-0.18	0.49
bottom layer							0.30	0.29	0.34
~~Ammonia~~Ammonium [mg L^{-1}]	0.60	-3.03	0.17	0.48	-3.28	0.15	0.56	-3.05	0.17
surface layer							0.25	-1.76	0.08
bottom layer							0.64	0.41	0.70
Phosphate [mg L^{-1}]	0.10	0.56	0.51	0.09	0.59	0.62	0.09	0.56	0.51s
surface layer							0.03	0.40	0.43
bottom layer							0.17	0.40	0.36

~~Appendix A~~

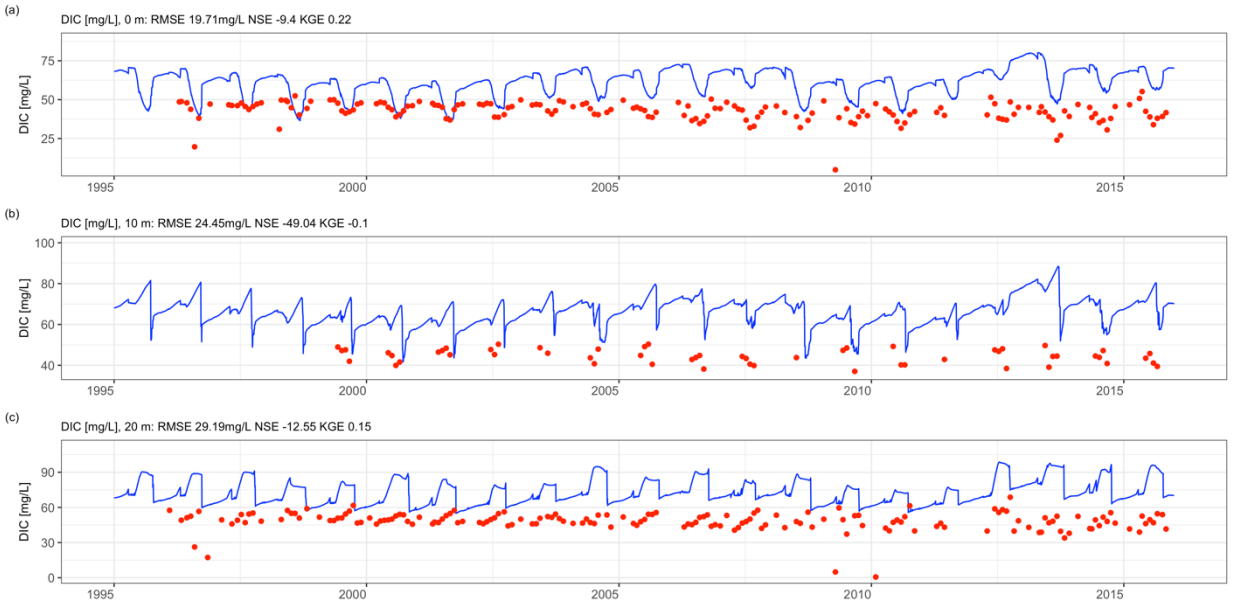
~~Correspondence to~~: ~~Robert Ladwig (rladwig2@wisc.edu)~~



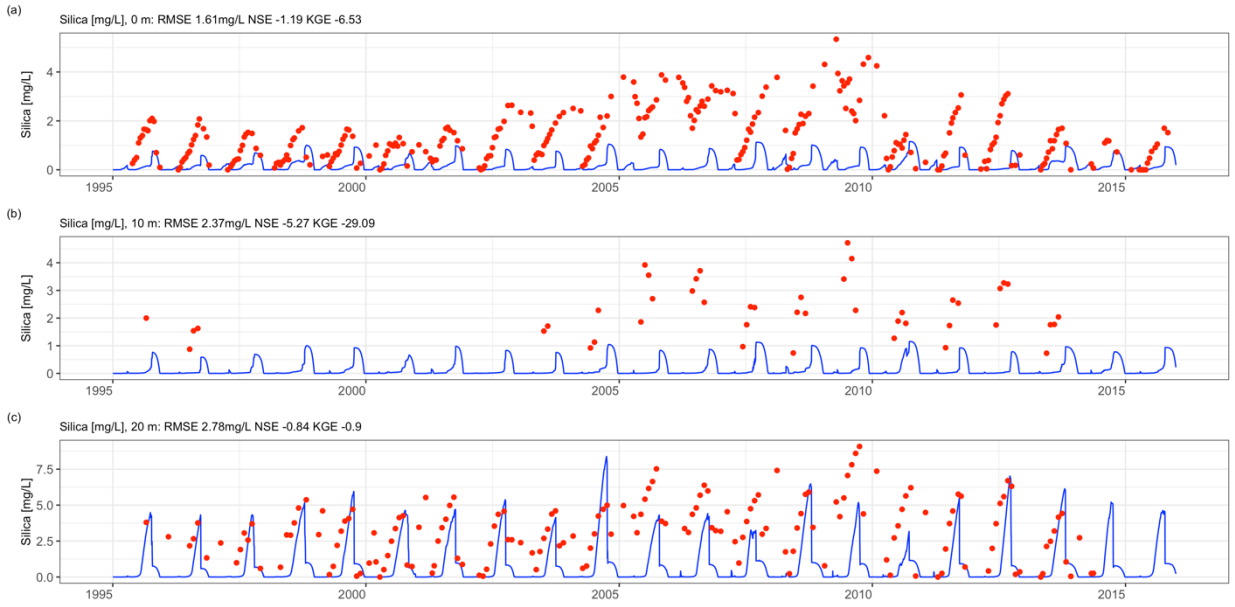
~~Figure A1 Mean and standard deviations of absolute elementary effects quantified by the Morris Method for water temperature, dissolved oxygen, dissolved inorganic carbon, silica, nitrate and phosphate. Colored bars are sensitive parameters that were used in the calibration.~~



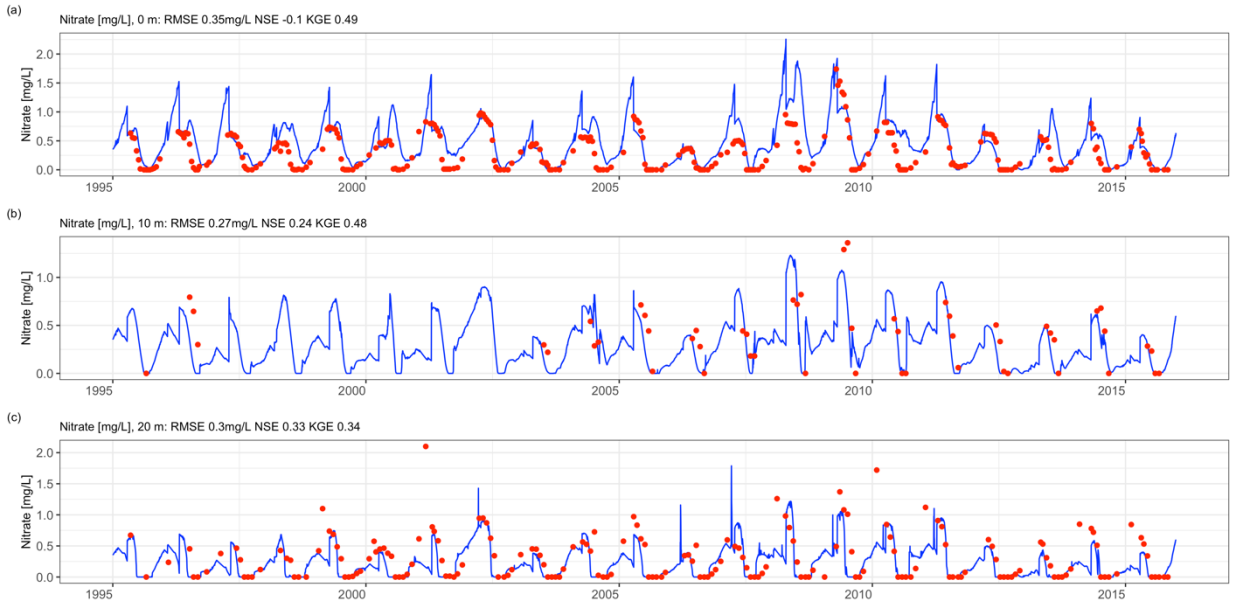
~~Figure A2 Time-series comparison between observed (red dots) and water temperatures (blue lines). The fit criteria root-mean square error (RMSE), Nash-Sutcliffe coefficient of efficiency (NSE) and Kling-Gupta coefficient of efficien~~



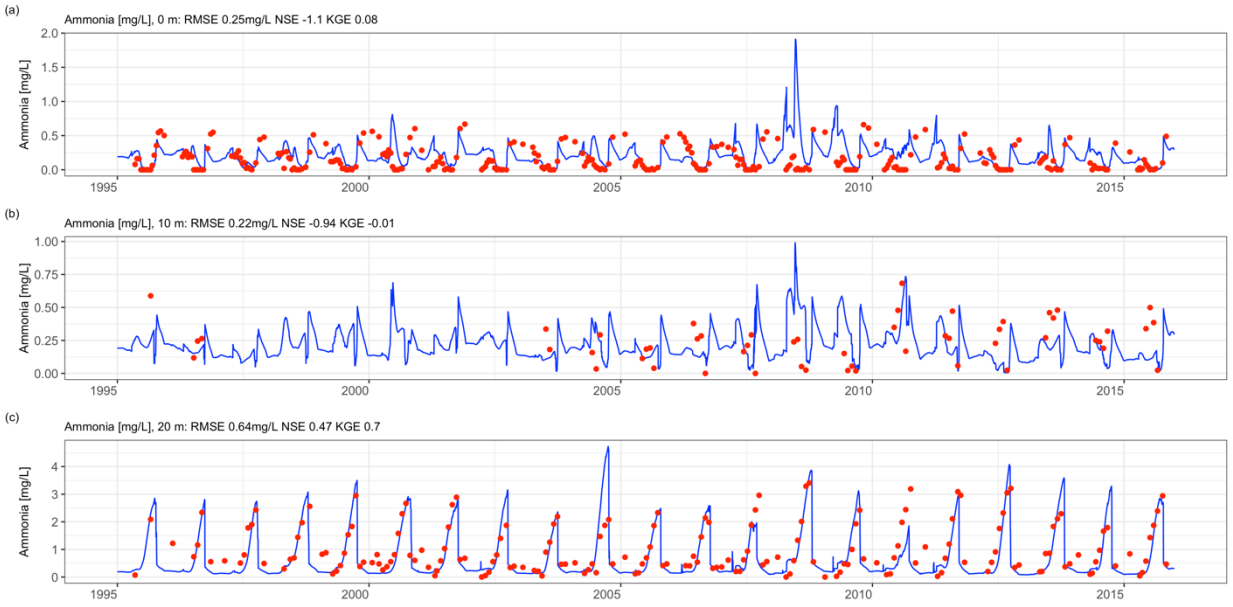
~~Figure A3 Time-series comparison between observed (red dots) and modeled dissolved inorganic carbon concentrations (blue lines). The fit criteria root-mean-square error (RMSE), Nash-Sutcliffe coefficient of efficiency (NSE) and Kling-Gupta coefficient of efficiency (KGE).~~



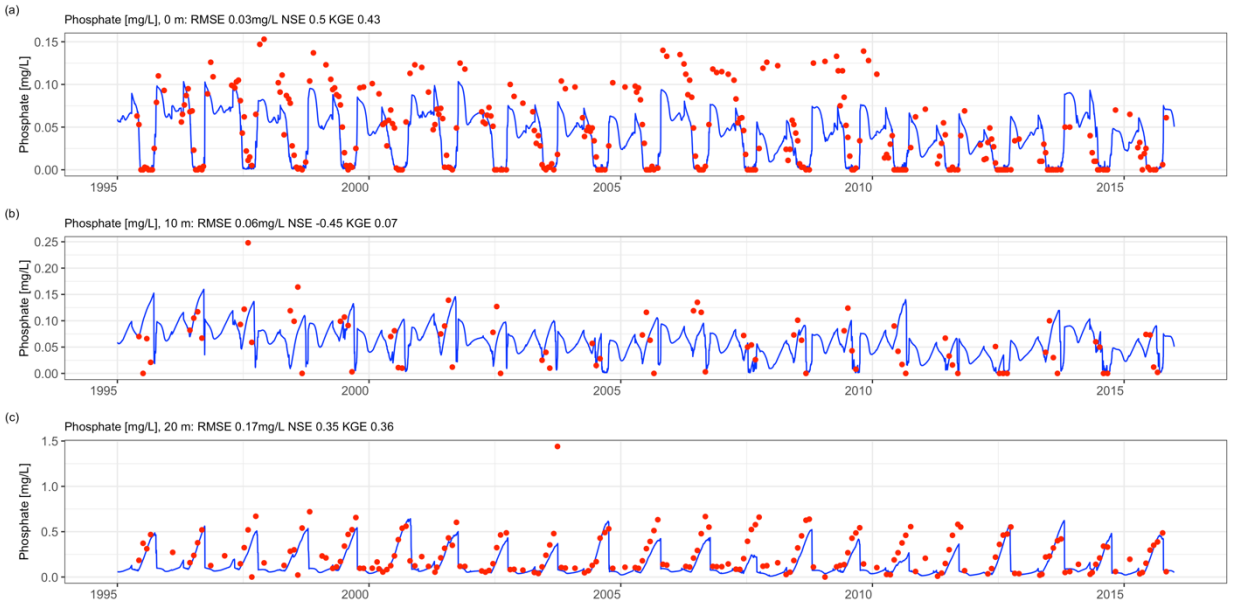
~~Figure A4 Time-series comparison between observed (red dots) and modeled silica concentrations (blue lines). The fit criteria root-mean square error (RMSE), Nash-Sutcliffe coefficient of efficiency (NSE) and Kling-Gupta coefficient of efficiency (KGE).~~



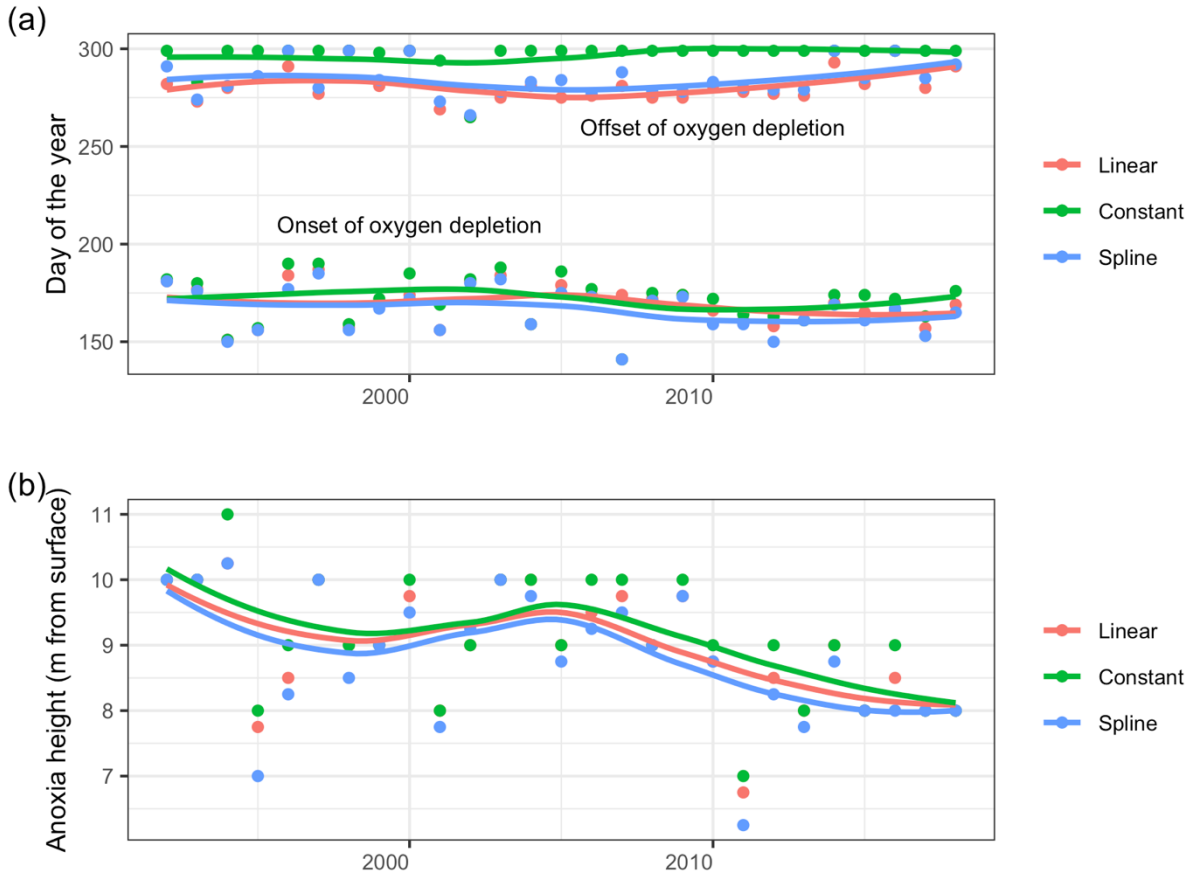
~~Figure A5 Time-series comparison between observed (red dots) and modeled nitrate concentrations (blue lines). The fit criteria root-mean square error (RMSE), Nash-Sutcliffe coefficient of efficiency (NSE) and Kling-Gupta coefficient of efficiency (KGE).~~



~~Figure A6Time-series comparison between observed (red dots) and modeled ammonia concentrations (blue lines). The fit criteria root-mean square error (RMSE), Nash-Sutcliffe coefficient of efficiency (NSE) and Kling-Gupta coefficient of efficiency (KGE).~~



~~Figure A7 Time-series comparison between observed (red dots) and modeled phosphate concentrations (blue lines). The fit criteria root-mean square error (RMSE), Nash-Sutcliffe coefficient of efficiency (NSE) and Kling-Gupta coefficient of efficiency (KGE).~~



~~Figure A8 Observed anoxia onset, offset (a) and height (b) dynamics. The colored lines refer to the interpolation method.~~

~~Table A1 Model parameters for functional phytoplankton groups~~

~~Parameter~~	~~Description~~	~~Cyanobacteria~~	~~Diatoms~~
~~P_initial~~	~~Initial concentration of phytoplankton (mmol C/m3)~~	~~10~~	~~8.4~~
~~P0~~	~~Minimum concentration of phytoplankton (mmol C/m3)~~	~~0.03~~	~~0.03~~
~~W_p~~	~~sedimentation rate (m/d)~~	~~0~~	~~-0.05~~
~~Xcc~~	~~carbon to chlorophyll ratio (mg C/mg chla)~~	~~50~~	~~50~~
~~R_growth~~	~~Phyto max growth rate @20C (/day)~~	~~0.8~~	~~2.8~~
~~fT_Method~~	~~Temperature limitation function of growth~~	~~CAEDYM style~~	~~CAEDYM style~~

Theta_growth	Arrhenius temperature scaling for growth function (-)	1.06	1.06
T_std	Standard temperature (deg C)	20	15
T_opt	Optimum temperature (deg C)	28	20
T_max	Maximum temperature (deg C	35	32
lightModel	Type of light response function	no photoinhibition	no photoinhibition
I_K	Half saturation constant for light limitation of growth (microE/m^2/s)	25	10
KePHY	Specific attenuation coefficient ((mmol C m^3^-1)^1 m^-1)	0.005	0.001
F_pr	Fraction of primary production lost to exudation (-)	0.005	0.002
R_resp	Phytoplankton respiration/metabolic loss rate @ 20 (deg C)	0.08	0.12
Theta_resp	Arrhenius temperature scaling factor for respiration (-)	1.05	1.07
K_fres	Fraction of metabolic loss that is true respiration (-)	0.6	0.6
K_fdom	Fraction of metabolic loss that is DOM (-)	0.05	0.05
simDINUptake	Simulate DIN uptake	True	True
simINDynamics	Simulate internal N	Fixed C:N	Dynamic C:N
N_0	Nitrogen concentration below which uptake is 0 (mmol N/m^3)	0	0
K_N	Half saturation concentration of nitrogen (mmol N/m^3)	1	3.5
X_ncon	Constant internal nitrogen concentration (mmol N/ mmol C)	0.035	0.035
X_nmin	minimum internal nitrogen concentration (mmol N/ mmol C)	0.06	0.077
X_nmax	maximum internal nitrogen concentration (mmol N/ mmol C)	0.206	0.129
R_nuptake	maximum nitrogen uptake rate (mmol N/m^3/d)	0.068	0.13
R_nfix	nitrogen fixation rate (mmol N/mmol C/day)	0.13	0
simDIPUptake	Simulate DIP uptake	True	True
simIPDynamics	Simulate internal phosphorus dynamics	Dynamic C:P	Dynamic C:P

P_0	Phosphorus concentration below which uptake is 0 (mmol P/m^3)	0	0
K_P	Half saturation concentration of phosphorus (mmol P/m^3)	0.5	0.7
X_pmin	Minimum internal phosphorus concentration (mmol P/mmol C)	0.0019	0.0081
X_pmax	Maximum internal phosphorus concentration (mmol P/mmol C)	0.0089	0.033
R_puptake	Maximum phosphorus uptake rate (mmol P/m^3/d)	0.0039	0.007
simSIUptake	Simulate Si uptake	False	True
Si_0	Silica concentration below which uptake is 0 (mmol Si/m^3	-	0
K_Si	Half saturation concentration of silica (mmol Si /m3)	-	2.5
X_sicon	Constant internal silica concentration (mmol Si/mmol C)	-	0.04

Table A2 Calibrated model parameters

Parameter	Description	Unit	Default value (Hipsey et al., 2019a, 2019b)	Model value
f_{sw}	Solar radiation scaling factor	-	1.0	0.84
f_{lw}	Long-wave radiation scaling factor	-	1.0	0.99
C_H	Bulk aerodynamic coefficient for sensible heat transfer	-	0.0013	0.0014
$T_{z=1,mean}$	Annual mean temperature of the upper sediment zone	°C	-	5.07
$T_{z=z,mean}$	Annual mean temperature of the lower sediment zone	°C	-	13.47
F_{max}^{oxy}	Max. sediment flux for dissolved oxygen	$\frac{mmol}{m^2 d^2}$	-100.0	-100.0
K_{sed}^{oxy}	Half saturation concentration controlling oxygen sediment flux	$\frac{mmol}{m^3}$	50.0	15.0

θ_{sed}^{oxy}	Temperature multiplier for oxygen sediment flux	-	1.0	1.08
$R_{mineral}^{dom}$	Maximum rate of aerobic mineralisation of labile dissolved organic matter at 20 °C	d^{-1}	0.5	0.5
F_{max}^{dic}	Max. sediment flux for dissolved inorganic carbon (DIC)	$\frac{mmol}{m^2 d^2}$	4.0	250.0
K_{sed}^{dic}	Half saturation concentration controlling DIC sediment flux	$\frac{mmol}{m^3}$	30.0	7.0
θ_{sed}^{dic}	Arrhenius temperature multiplier for DIC sediment flux	-	1.0	1.08
F_{max}^{rsi}	Max. sediment flux for reactive silica	$\frac{mmol}{m^2 d^2}$	-	16.42
K_{sed}^{rsi}	Half saturation concentration controlling silica sediment flux	$\frac{mmol}{m^3}$	50.0	1.90
θ_{sed}^{rsi}	Arrhenius temperature multiplier for silica sediment flux	-	1.0	1.08
R_{nitrif}	Maximum rate of nitrification at 20 °C	d^{-1}	0.1	0.03
R_{aenit}	Maximum rate of denitrification at 20 °C	d^{-1}	0.3	2.0
K_{aenit}	Half saturation concentration for denitrification	$\frac{mmol}{m^3}$	2.0	3.0
F_{max}^{nit}	Max. sediment flux for nitrate		-5.0	-9.55
K_{sed}^{nit}	Half saturation concentration controlling nitrate sediment flux	$\frac{mmol}{m^3}$	100.0	173.13
F_{max}^{frp}	Max. sediment flux for phosphate	$\frac{mmol}{m^2 d^2}$	-	0.49

K_{sed}^{frp}	Half saturation concentration controlling phosphate sediment flux	$\frac{mmol}{m^3}$	-	200.0
θ_{sed}^{frp}	Arrhenius temperature multiplier for phosphate sediment flux	-	1.0	1.0

Table A3 Most parsimonious multiple linear regression model (adjusted $R^2 = 0.88$, $p < 0.001$) explaining the Anoxic Factor during summer.

	Estimate	Std. Error	t value	Pr(>|t|)	Rel. importance [%]
Intercept	-1.04e-15	5.70e-2	0.00	1.00	
Onset of stratification *(Onset.Strat)*	-6.65e-1	1.49e-1	-4.46	1.2e-4	22
Schmidt Stability during summer *(Summer.St)*	5.38e-1	1.79e-1	2.99	5.0e-3	17
HBR ratio during summer *(Summer.HBR)*	-4.76e-1	2.64e-1	-1.80	0.08	13
Gross primary production in the epilimnion *(Epi.GPP)*	2.34e-1	9.35e-2	2.50	0.01	13
Maximum depth of anoxia during summer below surface *(Intensity)*	-1.53e-1	7.90e-2	-1.93	0.06	11
Birgean Work during summer *(Summer.B)*	-6.44e-1	1.72e01	-3.72	0.8e-4	11
Water temperature in the hypolimnion at the onset of stratification (*Wtemp.Strat*)	2.98e-1	1.39e-1	2.13	0.04	10