We would like to thank the reviewers for their useful and detailed reviews of our manuscript. Please see below for our responses to each reviewer comment.

Reviewer 2

The authors made a good work while revising the paper.
I appreciated that they restructured the introduction, but I think some work is still required to limit overstatements.
I have a few specific comments below, but my main criticism, which is an easy fix, is that the authors justify their work by stating that "the relative importance of autochthonous and allochthonous OC in depleting oxygen across diverse lake systems remains uncertain" (Abstract). Similarly, they present their main finding, that hypolimnetic oxygen consumption co mes mostly from the respiration of autochthonous OC, as surprising (l598-599).
Although I find that the study is providing a nice mechanistic understanding of the links between the C and O cyclings in lakes, the authors must tone down on the "lack of knowledge" in the intro and the "surprise" of the conclusion. The link between nutrient enrichment and oxygen depletion at lake bottoms is certainly not novel, so the fact that excess PP fuels oxygen depletion in hypolimnion is undoubtedly not a novelty. Already Odum in 1956 was on this line. I suggest the authors return to Nürnberg, Gertrud K., (1995), Quantifying anoxia in lakes, Limnology and Oceanography, 40, doi: 10.4319/lo.1995.40.6.1100. to justify their study while not undermining the basic limnological knowledge. Muller et al, 2012 also is a good reference on which they can build (already in the ref list).

Response: In the Abstract we replaced “relative importance” with “relative contribution” and rewrote the sentence. Similarly, in the 4th paragraph of the Introduction, we replaced “importance” with “contributions.” We have also included the following sentence in the final paragraph of the introduction:

The importance of excess primary production to anoxia has been established (Nürnberg et al., 1995; Müller et al., 2012). We build upon this research by quantifying the timing and magnitude of OC contributions to hypolimnetic anoxia.

Specific comments

R1: L255-257
"We choose to develop our own model instead of GLM or Simstrat to … metabolic… " . Both Simstrat and GLM are physical models. Besides, just below, you mention that the GLM was used to simulate temperatures. So my guess is that you did not use available water quality models such as AED 2 for the C/O compartments

Response: We apologize for the confusion. We meant to say that we developed our own “water quality” model, and that we used the output of GLM for physical dynamics. The GLM output we used for temperature profiles was previously published (Read et al. 2021). We have clarified this in the manuscript.

R2: Table 3: duplicated line for the Respiration rate of POCR

Response: Thank you for pointing this out. We have fixed this error in the manuscript.

R3: Fig 6 and l 537: how much can you say from the lake metabolism given the uncertainties around the model, which are usually bigger than the difference between GPP and R? I think some care is required here.

Response: We have added the following text to section 3.2 of the manuscript:

“It is worth noting that our interpretation of metabolism dynamics in the results are based on the median NPP and Respiration flux values produced by the model. Because of the high uncertainty associated with these fluxes, we should be cautious about asserting inferences about long term changes in trophic state.”

R4: L557 "Water column respiration contributes more than sediment respiration to total hypolimnetic respiration in the southern lakes compared to the northern lakes, with the exception of TR, where cumulative water column respiration is much larger than cumulative sediment respiration"
This sentence is quite circumvoluted. Maybe"Water column respiration contributes more than sediment respiration to total hypolimnetic respiration in the deepest lakes.

Response: We agree with this suggestion and have reflected this change in the manuscript.

R5: L 604 "Autochthonous OC pools have higher turnover rates than allochthonous OC pools (Dordoni et al., 2022) and often are lower in concentration than the more recalcitrant allochthonous pools (Wilkinson et al. 2013)." In the study lakes, autochthonous DOC dominates by far.

Response: We have added further text to the Discussion section (copied below), to help clarify our point. We also note that the lakes in our study (specifically, AL, BM, SP, and TR) were in the Wilkinson et al (2013) study.

New text added to section 4.1 of the manuscript:
"Similar to what was found by Wilkinson et al (2013), the standing stock of DOC in the water column of lakes in our study was from predominantly allochthonous sources. However, we emphasize in our study that autochthonous OC pools have higher turnover rates than allochthonous OC pools (Dordoni et al., 2022) and often are lower in concentration than the more recalcitrant allochthonous pools (Wilkinson et al. 2013)."

R6: I need help understanding Figure 8. Is the total average annual hypolimnetic respiration per lake different from the annual respiration flux presented in Figure 5? If it does not, why is there an order of magnitude of difference between fluxes (approx 40-150 g/m2/yr in Fig 5 and 400-1300 g/m2/yr in Fig 8)? If it is different, can you clarify?

Response: Fig 5 shows the annual whole-lake respiration flux (epi and hypo), and describes the flux over the entire lake surface area. Fig 8 shows the total annual lake respiration for just the hypolimnion, described over just the area of the hypolimnion (lake sediments). We estimated the area of the hypolimnion from the area of the thermocline. This difference of respiration pool (gC) and area (m2) results in different flux magnitudes (gC/m2).

Reviewer 3

The revised manuscript improves significantly on the earlier draft. However, there is still no information regarding the basis of the model linking NPP to TP. Furthermore, the author's response to my original comment does not mention a citation, nor does it really address my comment regarding the basis of this model formulation. The citations that are mentioned in the manuscript describe NPP models based on chl a (McCullough et al. 2018) and based on nitrate and phosphate concentrations (Ladwig et al. 2021, Hippsey et al. 2022). Both of these parameters provide a better basis for estimating NPP than TP, which is used in the current study. Total phosphorus includes contributions from sediment-bound P, so there is likely to be variability in TP that is not associated with changes NPP. I suspect that data limitations dictated the use of TP in the NPP model, and if so, the manuscript should make this clear. Furthermore, lacking a citation for a TP-NPP model, the authors should provide some justification for why this model formulation is valid.

Response: The reviewer highlights an important issue, which is the relevance of total phosphorus to primary production in the water column. We have added the following text into section 2.3.4 or the manuscript:

"For our most eutrophic lake, Lake Mendota, most of the TP in Lake Mendota's water column is from internal loading (Soranno et al.1997) as a consequence of mineralization of organic forms of phosphorus (principally from phytoplankton) or released in mineral form from sediments (Hoffman et al. 2013). As shown by Read et al. (2014), most of the total phosphorus in the water column is in reactive forms – either dissolved reactive or as part of phytoplankton biomass. Considering phosphorus can cycle rapidly among forms, and considering the high correlation between TP and reactive forms of phosphorus (Read et al., 2014), we consider total phosphorus to be a suitable index of available nutrients for primary production. This simplification obviates the need for additional processes in the model for cycling phosphorus among various forms, which reduces model data requirements and the need for more complex calibration. Lakes in our study have relatively low sediment loading, so we assume tightly bound phosphorus in the water column (i.e., phosphorus generally unavailable to biotic processes) is a small fraction of the total phosphorus. In lakes with high sediment loads, the assumption that TP is a reasonable index for NPP may not hold true."

R1: Line 208: How were these discharge discrepancies quantified?

Response: We quantified these discrepancies by comparing the modeled discharge estimates to published water residence time values for our study lakes. We have added the following text to help clarify this point:

"After comparing the modeled discharge data with published water residence times for our study lakes (Table 1), we found that the derived discharge data for ME, TR, AL, and SP was approximately 20-50% higher than previously reported values (Hunt et al. 2013, Webster et al. 1996), depending on the lake, while hydrology in BM was approximately 25% too low (Hunt et al. 2013)"

R2: Line 254: Should this be Ladwig (2021) or is there a 2022 publication not listed in the reference list?

Response: Thank you for pointing out this oversight on our part. We have fixed this issue in the manuscript

R3: Table 3: Respiration rate for POCR listed twice.

Response: Thank you for pointing this out. We have fixed this error in the manuscript.

R4: Table 3: Why do POCL and POCR have different settling rates? Is there a reason to expect this difference?

Response: For POCR, we note that the settling rate [1.2 m/d] equates approximately to a first order decay of 0.1 1/d, when divided by the mean depth of the lake (e.g., 1.2 m/d / 12.8m ~= 0.1 m/d). This rate is roughly 2 orders of magnitude greater than the decay rate for POCR. Thus, the model is not particularly sensitive to the POCR settling rate. The POCR represents leaves, pollen, detritus, etc. from the catchment. POCL represents the community of phytoplankton in our model, which may include fast-sinking diatoms and blue-greens that sometimes have buoyancy. We cite a diversity of sinking rates in the manuscript. It seems reasonable that POCL would sink more slowly than POCR.

R5: Line 346: should this read: "…in a layer is taken from…"

Response: Yes, we agree and have reflected these changes in the manuscript.

R6: Line 418: Move descriptions of "constant", "manually calibrated", and "parameters calibrated…" earlier in the manuscript so that the reader understands these categories when looking at Table 3.

Response: Thank you for this suggestion. We agree that Table 3 should be closer to the model parameter group descriptions. Rather than move the descriptions further up in the manuscript, we have moved Table 3 down further closer to the Model Sensitivity and Parameter Calibration section.

R7: Line 424: Which observations were used for the model fit? DO? DOC? POC? Secchi? Were all groups of observations equally weighted so that the parameters were estimated to maximize model fit to all measurements? If so, wouldn't it make more sense to more heavily weight predictions of DOC and POC, given the focus of this analysis?

Response: While the reviewer makes an interesting point about weighting DOC and POC in model fits, we note that DOC data are somewhat sparse and not especially dynamic, and that

Secchi is our index for POC. We have high confidence in values of DO, and DO is critical to calibrating metabolic rates. Thus, we do not feel a sufficient justification for weighting more heavily any one observational variable.

The following text was included in the original manuscript submission but was removed:

During the model fitting, errors in modeled DO, DOC, and Secchi depth are weighted equally in the southern lakes. In the northern lakes, fitting Secchi depth was challenging. Initial model fits revealed that patterns in observed Secchi did not show regular seasonality and were highly stochastic. Therefore, we use a moving average on observational data and predictions of Secchi depth and calculate the residuals as the difference between the two averaged time series. This is done to remove stochasticity from the observational data and fit the model predictions to the average observed Secchi value. We use a moving average window of 15 observations because we want to capture the average annual Secchi depth trend, and there are roughly 15 observations per year.

We have updated section 2.4 to include some of this original text (copied below):

During the model fitting, errors in modeled DO, DOC, and Secchi depth are weighted equally in the southern lakes. Secchi depths in the northern lakes were highly stochastic, and therefore we use a moving average on observational data and predictions of Secchi depth and calculate the residuals as the difference between the two averaged time series. We use a moving average window of 15 observations because we want to capture the average annual Secchi depth trend, and there are roughly 15 observations per year.

R8: Figure 2: Hypolimnetic DO depletion rate seems notably off in SP. Any reason why?

Response: The hypolimnetic DO depletion rate for SP is higher than observed in some years and lower in others, indicating more uncertainty. We can only speculate about the cause. The highly uncertain DO depletion rate may have to do with the morphometry of SP, which is more of a flat-bowl-shape than the rest of the northern study lakes and therefore may have a higher proportional sediment area. Although lake hypsometry, along with thermal profile, controls the volume of hypolimnion in contact with sediments in our model, there may be other factors related to morphometry (e.g., sediment focusing) that remain unaccounted.

We have added additional text to section 4.2 to help clarify this point.

R9: Figure 3: It seems like the amplitude of variation in DOC (difference between the peak of each seasonal cycle and the trough) is overestimated in all lakes except for BM. I would guess that the majority of this seasonal variation in DOC is due to autochthonous sources, so could this difference in amplitude indicate that the predicted autochthonous contributions to DOC are too high? If so, this would affect the relative importance of OC sources presented in later figures (e.g., Line 512). I think the conclusions regarding the importance of OC sources would be stronger if the amplitudes of variations of observed vs. predicted were more similar (see comment for Line 424 regarding model fitting).

Response: As the reviewer points out, most of the DOC dynamics are from autochthony. In our model, autochthony is split between dissolved and particulate fractions. Lowering the annual amplitude of modeled DOC would perhaps raise the annual amplitude of modeled POC, but the OC would still be autochthonous. Most autochthony is respired in the water column in our lakes, regardless of its form (Fig. 5). We also note that BM, pointed out by the reviewer as the lake with the lowest DOC annual amplitude, does not have OC fates that are remarkably different from other similar lakes in our study (Fig. 5). Although recalibrating the model always changes numbers, we believe the differences would be minor and would not alter our conclusions.

R10: Line 581: Isn't the smaller proportion of total respiration attributed to DOC a direct consequence of the fact that selected values for r for particulates is set to 5x that of dissolved values? So, doesn't this finding simply reflect model assumptions?

Response: Relative rates do, of course, have an effect on fluxes and standing stocks. However, there is no way to accurately predict a priori what will happen at the ecosystem scale. The total respiratory flux is the product of the respective labile and recalcitrant pool sizes and their decay rates, the available oxygen, and the ambient temperature. The oxygen demand, therefore, depends very much on the OC pool sizes, which vary by seasons and differ between thermal strata. Based on ecosystem observations alone, the pool size of the recalcitrant fraction of OC is higher in many lakes, suggesting that ecosystem respiration scales with allochthony. Without quantifying primary production and its fate, we have no way to account for its contribution to total respiration. How OC moves through the lake matters, as well. For example, it is conceivable that primary production in the epilimnion could have been offset by epilimnetic respiration, thus eliminating the autochthonous pool before it had a chance to contribute to hypolimnetic respiration. However, through modeling, we were able to quantify the autochthonous OC production and its export from the epilimnion to the hypolimnion and its total contribution to the organic carbon cycle, including the effects on oxygen consumption and the seasonality of those dynamics.

R11: Line 597: I think this approach has merit, and this statement may well be true, but I also think it makes sense to place this statement in the context of the different model assumptions.

Response: The criticism raised here is of a similar nature to previous criticisms about being careful not to inflate the importance of the findings. We agree that the text could be “toned down” and tightened and edited the first Discussion paragraph accordingly.

R12: Line 609: Are there DOC datasets available that are collected more frequently? It would be interesting to validate against more intense data.

Response: We wish there were more DOC observational data available for this study, and we agree that would make for an interesting comparison with our results. Although it is outside of the scope of this manuscript, we agree that a field campaign focusing on alternative DOC pools and their sources in lakes would be a great topic for future research.