Review of: Modeling of the large-scale nutrient biogeochemical cycles in Lake Onego

I previously reviewed this paper when it was submitted to Limnology & Oceanography. Of interest to me was to understand if the authors had responded to my earlier comments, the comments of another reviewer and the handling editor. These comments should have translated to changes throughout the paper. In the first instance, I noted the Abstracts of the two papers (L&O and Biogeosciences) were identical. I have gone through the reviewers' comments from the earlier review (in blue) and added new text (in black) that reflects whether I consider the earlier comments are adequately dealt with in the current submission to Biogeosciences.

Reviewer: 1

This paper presents a largely theoretical physical-biogeochemical simulation of Lake Onego for a period of 40 years. The paper seeks to use a modelling approach to collate some of the relatively sparse and disparate sources of information available on the lake. While this approach is commendable, it needs to be well supported with a sound underpinning modelling framework. Such a framework would involve:

- Being sure to collate the available and relevant sources of information available on the lake. This was not done adequately in my opinion and several of the papers that were part of a special issue on Lake Onego (Inland Waters Vol 9, Issue 2), and contained relevant information, were not cited, while at the same time the authors stated that "there is almost no empirical information on the major biogeochemical variables and fluxes".

The authors have partially addressed this point – noting the inclusion of the Efremova et al. 2019 paper (as cited by the authors).

- A sound modelling process involves calibration and validation against concentrations and/or biogeochemical fluxes. The comparison of temperature made by the authors: "Taking this expected bias into consideration, the simulated water temperature (11.85 ± 3.92 °C; median 13.15 °C) matches the observations (13.05 ± 4.82 °C, median 14.40 °C) well" is inadequate and direct comparisons (observations vs. simulation output) should have been made together with the relevant underpinning errors statistics (e.g., R2, RMSE, PBIAS, etc.).

The authors have not in my opinion adequately addressed this point. They suggest that "the simulated water temperature (11.85 ± 3.92 °C; median 13.15 °C) matches the observations (13.05 ± 4.82 °C, median 14.40 °C) rather well" but the reader is not told what the error statistic (±) relates to and it is not clear why model output was not aligned with comparable location of stations in claiming that there was bias due to 70% of measurements coming from coastal areas and bays. It looks from Fig. 5 like a comparison was done for three specific stations, but the R values are rather poor for temperature and a 1:1 line should also be shown to examine if there was systematic bias. I note that the authors have included a figure (Fig. 4) showing observed and simulated ice cover but give no statistics for goodness of fit and no detail about the way in which ice cover was simulated. Figure 6 caption should explain the black and blue lines and give the temperature for the 'biological summer'. Most importantly, the reader has no information if vertical thermal stratification is captured in the model; are there really no vertical profiles of temperature?

I appreciate that data may have been sparse but there are methods to counter this, e.g., use of remote sensing to provide optically active surface water constituents that can be compared with model output. Sensitivity analysis is also another useful approach to develop confidence in the model simulations and be able to define a range of output as part of a model error analysis. Bootstrapping approaches are also useful for sparse data. Without this the model simulations become a largely theoretical exercise because we do not know the accuracy of the simulation output.

I didn't see a lot of additional effort to address suggestions around sensitivity analysis or the use of remote sensing data to demonstrate spatial variability in the model.

On line 215-216, it is stated that "Presented concentrations of inorganic and total nutrients were averaged over the whole water body, from the surface to the bottom". I understand why this might be done for a winter mixed period, but not for summer. I'm unclear on the following sentence as well, and it seems to be that distributing DIP through the water column in summer would lead to some large inaccuracies. Indeed, earlier it is stated (line 211) that "…variables averaged over the entire Lake Onego"; was this a volumetric average or was it a vertical average for the water column? The authors seem to add doubt between observations and model assumptions with the statement that: "Consequently [because the authors chose to average nutrient concentrations over the whole water body], the dissolved inorganic phosphorus (DIP), comprising summer phosphorus accumulation in the hypolimnion, was never fully depleted'.

- Almost no information is given on the parameters that go into the model. Parameters like sediment nutrient release rates, deoxygenation rates, algal growth rates, etc., need to be provided in any modelling exercise; they serve as a basis for future work and refinement (e.g., in experimental work) and they should generally fit within literature ranges. A hint of parameterization is given in the sediment N and P content values given in Table 1 but it is notable that the model sediment N content is mostly greater than the range given for the measured values – and this is mostly the case for sediment P also. This raises some major question marks about the N and P mass balances that are given for the lake.

Information on parameters is still not given. There is a great deal of uncertainty for the reader relating to the way in which the biogeochemical model was calibrated. The paragraph in the Ecosystem variables section (lines 211 to 222) did not alleviate these concerns.

- The reader is given no overview of measured forcing data inputs into the model. For example, the 'estimated' river runoff and N and P loads (Fig. 3) should have had the observations included also. Further, the meteorological variable inputs to the model should have been clearly specified. There were quite a few typographical errors through the paper (e.g., spelling mistakes in Fig. 1) that would need to be corrected in future iterations of this paper.

The authors now provide a brief description of inflow, outflow and meteorological forcing data for the model input.

I would suggest not mixing the results and discussion into a single section, which would help to add clarity.

Not done. This paper could easily be improved by separation of Results and Discussion sections.

In other instances, elements of Introduction should not appear in the Methods. The Introduction would benefit from a clearly defined scientific hypothesis or test being put forward (in the last paragraph of the Introduction), i.e., this would form the basis of the model testing or scenario generation.

No clear scientific test or hypothesis presented. What was the purpose or objective of this paper?

What efforts were made to validate the discharge and nutrient concentrations measurements? It is not clear why the reviewer cannot see the interpolated load values (Fig. 3) plotted against the actual measurements (as points), so that the reader can see the frequency of measurements. Were the so-called "upward tendency" and "distinctly decreased" changes in loads with time actually significant? How were concentrations of dissolved nutrients in the inflows determined?

No formal calibration and validation of the model is carried out by the authors, raising doubts about the predictive capabilities of the model and more effort (e.g., remote sensing) should have gone into supporting this process.

It is not clear what the authors are pointing out in the following: "Taking into account all the uncertainties of such comparisons, starting with measurements being made over years from the mosaic distribution of real sediment types, and ending up with simplified and spatially invariable sediment parameterizations disregarding sediment types, the simulated areal concentrations could be considered plausible". It is not clear what would be plausible or implausible.