Dear Drs. Hessen and Cloern,

Thank you very much for the opportunity to revise our manuscript (LOL2-21-0001 "A unified dataset of co-located sewage pollution, periphyton, and benthic macroinvertebrate community and food web structure from Lake Baikal (Siberia)"). We believe that these revisions have improved the clarity of our data article as well as the reproducibility and transparency of our data products.

Where applicable, we have also incorporated revisions to the data product itself. Those edits have been published on the EDI repository as a newly versioned data product (https://doi.org/10.6073/pasta/9554b7f19ddd4a614e854f18be978dca) and on the project's associated Open Science Framework portal (DOI 10.17605/OSF.IO/9TA8Z). We have likewise changed the EDI citation within the manuscript to reflect the new version of the dataset.

Additionally, we received comments on the companion L&O manuscript, which we believe also clarify the macroinvertebrate collection methods in this manuscript. We have added clarifying text on LN 1026-1040 to further describe the macroinvertebrate diversity detailed in this dataset, especially in the context of Lake Baikal's exceptional biodiversity.

Below, we have included the reviewers' unabridged comments with our response below each comment in bold. All line numbers in our responses correspond to line numbers in the cleaned, revised version. For the reviewers' reference, we have also provided a "tracked changes" version that includes all edits with the exception of text reordering (as suggested by reviewer 1). This change has been made to the "clean" version, but we felt that the inclusion of those tracked revisions could make the document difficult to understand.

Michael

AE Comments to Authors:
Associate Editor
Comments to Authors:
Dear Dr. Meyer,

I regret the long handling time of your ms, due to problems with finding competent and willing reviewers (a well know problem in these days...). Hence to avoid further delays I have decided to base my recommendations on the very thorough referee report that we obtained, and my own, independent judgement which is pasted below. As you will see we both agree that this is a very valuable dataset that fits the scope of LOL data articles, and it is a very elaborate and accessible set of data from a unique system. Still there are a number of aspects that need to be considered and elaborated on for clarity, notably on the method description.

Sincerely

Dag Hessen (HE)

This ms offer a set of benthic water quality properties for a 40 km long shoreline of Lake Baikal. The data is clearly of value given the uniqueness of the lake, and the authors also argue that their study design could be of value for other sites as well. I find the metadata well organized and easily accessible. My two major concerns with the data are their representativity for Lake Baikal as such given that this is based on one sampling event (late August 2015) over a specific part of the shore, the other is the lack of rigor in the sampling procedures, i.e., there are missing information on depths, replicates etc. At least this latter concern is more of technical nature, however, and can be solved in a revision. The identification and definition of microplastics may be provided, but I did not find it, and what about macroplastics? Also it is unclear to me if some or all of potential PPCPs were included. More specific comments are given below with reference to line numbers:

Line61: Most voluminous? According to my record it is number 7.

Lake Baikal is the largest freshwater lake in the world by volume and the seventh largest lake in the world by surface area. We have modified the text by highlighting Baikal's position among the world's freshwater lakes.

L64: Since all data are based on one single sampling in late august 2015, it is worth discussion how representative this is for other seasons/years. Likewise the rationale for picking this part of the shoreline and how it represent the rest.

We have added a "Site Information" section to the methods on LN 835-860, so as (1) to offer justification for choosing this area of the shoreline and time for sampling and (2) to contextualize this area of the Baikal's shoreline relative to the entire lake.

Additionally, we have included text in the "Data Use and Recommendations for Reuse" section to emphasize that much of the previous work on sewage pollution in Baikal has centered around the same area that we sampled.

L88: Melting permafrost could affect DOC, but likely gives a negligible contribution of nutrients relative to sewage and agriculture.

We agree that DOC would likely be among the most concentrated constituents in melting permafrost; previous works focusing on nutrient sources in Lake Baikal's watershed have demonstrated the potential for melting permafrost to contribute significant nutrients and potentially cause nearshore eutrophication. To highlight this point, we have added references to Anisimov and Reneva (2006) and Moore et al (2009).

L143: The information may be provided somewhere, but the actual distance from the shore to the pelagic stations is hard to judge. Also, which parameters that were samples on these three sites (chla and nutrients?). It is hard to extract this information from the Data description (L 261 and on). Apparently all listed species are littoral.

We agree that the actual distance to shore for the pelagic sampling locations may be difficult to intuit from the map of sampling locations (Figure 1). To aid readers in estimating distances, we included scale bar to Figure 1.

With respect to the potential confusion about which samples were collected at which locations, we have added clarifying text on LN 855-856 detailing that macroinvertebrate and algal collections occurred only at littoral sites, whereas water column samples occurred at both littoral and pelagic sites. Additionally, brief introductions underneath each filename details that periphyton and macroinvertebrate samples were collected only at littoral sites.

L931 and 949: Were fatty acids extracted from the same samples as used for species identification and quantity assessment?

No, separate collections were used for species abundance estimates, stable isotopes, and fatty acids. We have modified the text on LN 1042-1047 to reflect that separate collections were gathered for each variable. Additionally, we have clarified that only samples collected for taxonomic abundance were preserved in ethanol prior to processing. Samples collected for stable isotopes and fatty acids were frozen prior to analysis and were not preserved in ethanol.

L972: Done for all species?

Due to some samples not being well preserved upon arrival to the United States, we prioritized processing samples that were completely frozen upon arrival. Therefore, only certain taxa have paired abundance, stable isotope, and fatty acid data. We have added information about these potential limitations as well as an additional table (Table 2), which explicitly states which species are included for each measurement. We have also added clarifying text explaining the issues with transporting some samples back to the U.S. on LN 1049-1056.

****

Reviewer(s)' Comments to Author:
Reviewer: 1

The manuscript makes a convincing argument that the data sets, combined here, are of high value with a strong likelihood of reuse if advertised and explained in a data paper format. However, there are a number of minor comments and concerns with the data presentation that should be addressed prior to publication (and updated in the current EDI submission as well).

My comments are generally directed at maximizing the potential for data re-use by pointing out instances where details are lacking in the data table or methods descriptions.

As there are multiple copies of the methods text between the manuscript text and EDI metadata file and the line numbers start over again on page 66 in the combined PDF for review, the review process was complicated. As such, all line numbers have been combined with page numbers in my comments below. The notation for comments below is page#.line# (e.g. page 3, line 143 is 3.143). NOTE: If text was duplicated in the document, I did not note every place where the comment applied. My comments are organized as general and then by data table to ease discussion.

General
20.821 – Unless there is a specific reason not to, please list the methods in the same order that the data tables are presented earlier in the text.

We made the change to the "clean" manuscript document. In the "track changes" version, we did not make this change so as to make the document more readable.

24.1001 – Unless specified to do otherwise in the template, it would be beneficial to list the QA/QC procedures (i.e., method detection limits, blanks, standards, duplication rate, etc) with each data table methods description instead of as a separate section.

We agree that some readers could expect this information in the methods as opposed to the technical validation section. However, we do believe that some of this information is pertinent to the validation section, so we edited the text such that the information can be found in both the "Methods" and "Technical Validation" sections.

Chlorophyll
7.172 – specify that this is chlorophyll a concentrations in suspended particles so as not to be confused with any potential measures from periphyton

We have clarified this point by explicitly stating chlorophyll concentrations were measured within the water column.

8.187 – The samples were analyzed with a spectrophotometer according to the methods (23.872), but the description of the data and column headers indicate fluorometric analysis. Please rectify. Note, this is also noted in the EDI metadata file (page 51) and the EDI submission that is currently live:
https://urldefense.com/v3/__https://portal.edirepository.org/nis/metadataviewer?packageid=edi.677.1__;!!JmPEgBY0HMszNaDT!9BWH3vJe6lhUiqDoS7P_RJOr9XR0DbKRQP0-xmETxsXBsJ13rbLooccVUIAMPFgid-TwIQ$

Thank you for noting this discrepancy. This comment prompted us to contact the service lab that processed our chlorophyll samples, and the methods that we originally received were not those that were actually performed. Instead of using a spectrophotometer, the service lab used a fluorometer, which required a different calibration formula than the one noted in the manuscript.

We have corrected the methods both in the manuscript on LN 929-937 and also in the EDI metadata.

22.864 – If known, please state the type of plastic for the sample collection bottle

Unfortunately, we do not know the exact type of plastic bottle we used to collect chlorophyll samples. Due to field limitations, we used cleaned, 1.5-L beverage bottles, which were likely made of polyethylene terephthalate. We have added this detail to the text on LN 923-924.

22.865 – Were the water samples held in the dark prior to extraction or exposed to the light?

We have added clarifying text that after filters were placed in 35-mm petri dishes, the petri dishes were wrapped in aluminum foil to prevent exposure to light. This detail has been added on LN 926-927.

22.876 – Given the oligotrophic conditions, it is not surprising that the chlorophyll a concentrations were quite low. As such, it is imperative to include the method detection limit for the instrument used and additional information on quality control procedures (e.g. were blanks and standards used for each analysis batch?).

We have added a note on LN 937 that the MDL for this analysis was approximately 0.02 mg/L. Additionally, we have included information about fluorometric readings for blank samples as well as notes on the calibration procedure on LN 933-935.

IDW Population
8.198 – please state the timeframe for the population estimate, e.g. “current population” or “population at the time of sampling”

We modified the text (LN 240 and 244-246) to reflect that population estimates are based on census data and do not reflect swings in population due to the tourism season.

21.841 – again, please note what the time frame is for the population size estimate

As with the above comment, we modified the text (LN 885-887) to reflect that population estimates are based on census data and do not reflect swings in population due to the tourism season.

Fatty Acids
12.406 – Are there other potential quality control flags that should be considered? For example, were none of the samples below the limit of detection for some of the fatty acids? (I am not familiar with FA analysis, but every instrument has detection limits, so I'm just hypothesizing...)

With respect to a detection limit, we estimate that our detection limits were approximately 1 ng/mL. In any event, we quantified and identified every lipid compound that showed up in the chromatogram. Each sample contained peaks that were associated with known fatty acids, and among the 59 fatty acids contained in our dataset, few fatty acids were completely absent from a sample. In practice, future data users would likely filter for fatty acids that are above a given threshold (e.g., greater than 0.5% total fatty acids), much like community abundance data, so as to reduce the influence of rare fatty acids on a particular analysis and interpretation.

Other than this detection and peak quantification caveat, we could not think of other quality control flags for the fatty acid data.

We have included text regarding fatty acid peak identification and quantification in the methods on LN 1095-1100.

Invertebrates
No comments

"Metadata"
15.529 – I personally found it confusing to refer to this file and data table as "metadata" as it is not the metadata for the whole data submission, rather it is site and sampling information. I would recommend naming this file/data table "site_information" or something similar to reduce confusion.

We agree that the "metadata" data table and file could be confusing. As suggested, we have renamed this file as "site_information". The only place where the "metadata" filename is maintained is in the original, raw data on the Open Science Framework ("original_data/baikal_nearshore_metadata_201508.csv"), as this naming system stems from how we originally collected the data in August 2015. Script "00_disaggregated_data_cleaning.R" contains a comment about this point, and why the output file is now called "site_information.csv". All successive scripts have been updated to reflect this filename change.

Additionally, all variables named "metadata" within successive scripts (i.e., variables that contain measurements from the "site_information.csv" file) have been renamed to either

“site_information” or “site_info”. For example, the variable “ppcp_meta_dist” has been renamed to “ppcp_site_info_dist”.

15.540 – please note the data format, if appropriate to do so here (julian date? MM-DD-YYYY? Something else?)

We have clarified the text to read that “Day sampling occurred” refers to the day of the month in which sampling occurred. We have also clarified this text throughout the manuscript for all files containing a “day” column.

15.561 – depth of what? Sampling? Maximum depth? Please specify

We have clarified the text to read “Maximum depth at sampling location in meters”.

15.571 – for the mid_temp_celsius it would be helpful to explictly state the formula by which data users can calculate the depth of sampling (e.g. “depth_m”/2?) or add another column that specifies the mid_temp_celsius measurement depth.

As suggested, we have added clarifying text with the specific formula “depth_m/2” to explicitly define the mid_temp_celsius variable’s sampling depth.

Microplastics
15.594 – consider indicating that these are suspended microplastics and not samples associated with sediments or other potential pools of microplastics?

We have clarified that microplastics are indeed “suspended” microplastics on LN 394.

22.900 – what depth were samples collected at?

We have clarified that samples were collected at a depth of approximately 0.75 m on LN 962.

Nutrients
16.619 – indicate the depth at which nutrient samples were taken

We have clarified that samples were collected at a depth of approximately 0.75 m on LN 260 in the data description section as well as the methods on LN 893.

22.856 – are nitrate and ammonium reported “as N” or not? Might be helpful to specify.

Nitrate and ammonium are not reported “as N” but rather as ion concentrations. We have clarified this detail on LN 913.

22.857-860 – I could not find the texts referenced in the citations (also, please include an author for these citations in addition to the year). Please provide more methods information to increase reuse of these data. Also, perhaps consider relating the methods used to similar analysis methods in Standard Methods or other freely available analytical methods sources (e.g. USGS, US EPA, European standards)?

We uploaded pdf copies of these methods to the Open Science Framework portal within the directory "Nearshore_sampling/methods". We also amended the citation to these methods in the text. The methods cited are intergovernmental, standardized methods, approved for monitoring in Russia, Kyrgyzstan, Kazakhstan, and Belarus, and as such, are written in Russian, which may not be most accessible to the international scientific community. To help compare these methods with international, standardized methods, we have included references to International Standards Organization (ISO) protocols, specifically ISO 6878:2004 and ISO 6777:1984, both of which were used to develop the Russian-language methods. We have clarified this point on LN 915-918.

22.857-860 – please include information on instrument-specific method detection limits, standards and blanks, and holding times used for quality control in the analysis

We added detail about instrument make and model, detection limits, standards, and holding times to the nutrient methods section on LN 899-911. We have also included detail about nutrient minimal detection limits in the "nutrient.csv" portion of the "Data Description" section on LN 262-263.

Periphyton
17.643 – please add that the periphyton was sampled from rocks at each location to the description

We added detail that periphyton was sampled from rocks at each littoral location on LN 421-422.

23.938 – was the total volume of the sample recorded as well? This information is useful for standardizing the cell counts across samples (e.g. 10 mL aliquot from 50 mL sample is 20% is different than a 10 mL aliquot with ~300 cells from a 500 mL sample).

Total volume was not recorded for each sample, but each sample contained approximately 10-15 mL of preserved periphyton sample. We have added this information on LN 1002.

PPCP
17.685 – were these water samples? I'm assuming so, but it would be helpful to explicitly state that.

Yes, the PPCP samples were water samples. We have clarified this point on LN 312-313.

17.686 – thank you for including detection limit information here

We are thankful for the reviewer bringing up this inconsistency among various method descriptions. We believe this additional information is very useful for data's potential for reuse.

19.747-754 – do these dates align with the information in the "metadata" file (site and sampling information)?

Yes, the dates align with the information in the "metadata" (now "site_information") file. We kept those dates in this file as that information could be useful for users wanting to know when a sample was collected vs. when the sample was processed completely.

We have noted that these dates should be the same as those in the "site_information" file, and have likewise adjusted the collection_day and analysis_day descriptions to read as "Day of the month" that the sample was either collected or analyzed.

22.888 – given the potential for contamination, were blank samples also filtered for quality control? Or is this practice unnecessary? Please elaborate.

We agree that there is potential for contamination. Due to the complexities of sampling in Siberia and the cost of sample processing, trip and field blanks were not included in this study.

Stable Isotopes
24.968 – How were the tissues preserved and water removed? Freeze drying? Please specify

We have added detail about how samples were preserved and freeze-dried on LN 1060.

Total Lipids
It appears that a methods description is completely missing for this data file

We accidentally omitted these methods in the submitted manuscript. We apologize for this oversight and have included methods describing how total lipids were calculated on LN 1102-1107.

Additionally, we noticed that the datafile itself was missing from the EDI portal, although the data are included in the EDI metadata and in our original data submission to EDI. This appears to have been an error when the data were processed for publication. We have added the new data file in the newly versioned data product.

DATA USE AND RECOMMENDATIONS FOR REUSE

I appreciated the thoughtful discussion for how the data may be used by others. No suggestions for improvement.