Earth Syst. Sci. Data Discuss., referee comment RC1
https://doi.org/10.5194/essd-2021-43-RC1, 2021

Comment on essd-2021-43

Anonymous Referee #1

Referee comment on "The Surface Water Chemistry (SWatCh) database: a standardized global database of water chemistry to facilitate large-sample hydrological research" by Lobke Rotteveel et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2021-43-RC1, 2021

The manuscript by Rotteveel and Sterling presents the global surface water chemistry (SWatCh) database, which contains data for 17 variables (Al, Fe, major ions, nutients, organic C, pH, etc.) from 9 million samples collected between 1960 and 2019. This database has the specific purpose to support research on surface water acidification. To create this database, the authors used data from 6 exiting hydrochemical databases/dataset, which they put in a uniform format, and then removed samples that were flagged as problematic and duplicates that exist as some of the databases used have culled data from the other databases.

I was able to download and use SWatCh without any problem. The download process is straight forward, and the database is easy to use.

While it is a very important task to assemble available data into such a large, publically available database, I feel that the authors have done a very poor job with regard to quality checks. They only discarded data that was already flagged as problematic, or which had very clearly unrealistic values, which was limited to negative values for concentrations. I think a much more robust quality check would be required to publish this dataset with an article in ESSD. Further, I feel that the authors did a rather poor job at analysing and presenting the data. For these two points, please see my major comments further below. Finally, I would like to highlight that this database does not contain any data on alkalinity, acid neutralisation capacity (ANC), DIC or HCO3- concentrations. This information can be found in at least a few of the databases from which data was taken for SWatCh. More importantly, these parameters represent the buffering capacity of a surface water body against acidification, and would thus be of huge importance for the study of surface water acidification and recovery. It is completely incomprehensible for me why these parameters were not included in SWatCh.

I suggest that major revision are necessary before this study can be considered for

publication in ESSD. Please, see my major and general comments below:

Major comment #1: Quality checks of database

You should check all parameter values if they are reasonable, even if they are not flagged. You should check for instance for unreasonable high values, which can be due to mistakes made with the units (in particular for a database like GloRiCh, where data was assembled from lots of different dataset). If for instance mg and ug (or mM and uM) have been mixed up at some point (that could already be a mistake in the dataset you are taking data from), this might lead to errors of three orders of magnitude. I would suggest to first define for each parameter a realistic value range. Then, for all values lying outside of that range, you should first check is that concerns only one value in a time series, or the whole time series of a sampling site, or all values of a certain data source (note that for instance GloRiCh gives references of the data sources it used, and already in GloRich such mistakes might be present). If extreme values concern one specific sampling location, it might be worth investigating if that might be due to an exceptional site. For instance extremely high F- concentrations might be due to hydrothermal influence. Extreme PO4- concentrations can be due to phosphate deposits, like in the Peace River catchment, Florida. Sediments from dried out lakes might yield high concentrations in NaSO4. Etc.

For each sampling location you should look at the time-series and try to identify potential outliers within the time series. For each outlier, you might want to check if other parameters are also affected, which could mean that either something exceptional has happened or that data from another sampling location has been wrongly attributed. Anyway, you should flag those values. You cannot assume that all suspicious data has already been flagged accordingly, in particular as data comes from very different sources, and some of them, like GloRiCh, are again assembled from different sources with different degree of quality checks.

Major comment #2: Presentation of database

Your results section is very short, and your discussion section doesn’t make many links to your own results. Figure 2 is a good beginning to represent the available data, but it would be more interesting if the spatial coverage was represented separately for different types of inland water bodies. It is not clear at all from your manuscript how well lakes vs. reservoirs vs. rivers are represented in tat database.

It would also be interesting to know the numbers of samples per water body type that have measurements for a specific combination of parameters that are interesting with regard to acidification, like: How many samples are there with all major ions and pH? Here you should maybe start with an overview of which combinations of parameters are usually used to study acidification. I guess samples where only sodium or phosphate was measure are not that interesting. Maybe you can make a ranking of parameter combinations that allow you to study acidification with a different degree of conclusiveness and certitude. And then list the number of samples that have measurements for these parameter combinations, and do that separately for different kinds of water bodies (lakes, reservoirs, canals, ditches, rivers, etc.) and different world regions (at least continents, or major biomes/climate zones). You should also give an overview about which time-periods are covered in different parts of the world. That would very important if you want to investigate temporal trends in acidification recovery.

You should also think about presenting data density (number of sites, number of samples per site, average length and frequency of time-series, etc.) for different types of inland waters as a map. You could take inspiration from figure 2a in Regnier et al. 2013 (Nature Geoscience, DOI: 10.1038/ngeo1830), that created a density index for pCO2 values.

You should also take into account global geodatasets that allow for regional classification of water bodies, like for instance the HydroAtlas (Linke et al. 2019, https://doi.org/10.1038/s41597-019-0300-6). Like this you could make more qualified statement about which kind of river or lake is underrepresented in your dataset. You state you cannot link your chemistry data to catchment properties, but with HydroAtlas you could get a good idea what kind of river-catchment systems are well represented and what kind underrepresented.

General comments:

L74-76: How did you perform those checks? Where can I see the results? I think a quality assessment of this kind is very important.

L85-86: I wonder how you identified these “untreated” water bodies. I know that in GloRiCh this information is not given. Here, some analysis of the water chemistry data

itself could have been useful to spot suspicious cases, for which some investigation could have been performed based on the location information.

L88: By “phosphorus”, do you mean “total phosphorus”?

L91-92: Error message instead of reference.

Section 4.2: When discussing these effect of methodological changes on time-series data, you should combine that indeed with an analysis of at least the longest time-series you have in your database.

Section 4.4: Here you mention that you often do not have the discharge data associated to the water chemistry data. Did you try to match the river water sampling locations with stream gauges from the Global Runoff Data Centre (GRDC, https://www.bafg.de/GRDC/EN/Home/homepage_node.html)?