Reviewer comments

Author responses

RC2

This manuscript presents long-term water quality time series data from the Elbe River in Europe. The authors use alkalinity and pH measurements to estimate dissolved CO2 concentrations, which they use to estimate CO2 emissions from the river and tributaries from 1984 to 2018. They then compare the temporal changes in CO2 emissions with the temporal changes in DIC, DOC and POC loads at the watershed's outlet, along with other water quality parameters. The authors show a decrease in CO2 emissions, which they relate to an improvement in water quality, particularly a decrease in DOC.

The paper suffers from several shortcomings in methodology, a poor presentation of results, and considerable issues with the English language. I must admit this comes as a surprise considering the list of authors, some of whom are widely recognized and respected in the scientific community. I think there is potential to improve this paper substantially, because the dataset holds significant value—but much more guidance will need to be provided by the co-authors. In the following I will elaborate on the three main concerns I have.

Reply:

Thank you for your thorough review. We apologize for shortcomings in the scientific quality. We will do our best to revise the text, keeping in mind also language issues.

In the methodology section, we will include an analysis of the uncertainties associated with pCO_2 and provided more details about the load calculations.

In the results section, we will incorporate a time series analysis of pCO_2 and biomass. The results of the Mann-Kendall test will also be included.

Methodological limitations

One limitation is that the entire paper is based on the use of two indirect methods to estimate CO2 emissions. First, pCO2 estimates are indirectly calculated from pH and alkalinity measurements. While this is a common undertaking, the authors must at least provide a quantification of uncertainties. Their plot comparing pCO2 estimates based on two different packages (PHREEQC and CO2SYS) raises concerns as it shows large differences between the two sets of estimates. Second, the CO2 emission estimates lack actual measurements. The authors use an empirical model which was primarily developed for smaller streams and might not be suitable to large rivers. The model results are not evaluated against actual measurements. Again, this needs to be justified (i.e. why was this particular model chosen and not another one?), and an assessment of uncertainties should be presented.

Reply:
Thank you very much for your suggestions.

We selected CO2SYS (Lewis & Wallace, 1998) over Phreeqc (Parkhurst & Appelo, 2013) for the calculations as not all datapoints provided the anions and cations required for a reliable calculation in Phreeqc. About 60% of the sample points have major ion data. The comparison of both calculations shows that there is an offset between measurements, resulting in about 16% higher pCO_2 values (Figure S4), when calculated with CO2SYS. To keep results consistent and comparable, in the study we calculated all data with CO2SYS, accepting the potential error.

To evaluate the uncertainty in pCO_2 estimates, we will focus on the calculation methods, as direct pCO_2 or F_{CO2} measurements are unavailable. CO2SYS (Humphreys et al., 2022) provides an approach to calculate the error propagation. The errors included in the propagation are:

1) pH: General precision of standard commercial pH probes is typically between ±0.01 to ±0.1,so we will assume ±0.05.
2) TA: General precision of TA measurement by titration methods ranging from ±10 to ±50 μmol L^{-1}, so we will assume ±20 μmol L^{-1}.
3) Temperature: assumed as ±0.1 ℃.

Finally, this approach leads to an estimated uncertainty of around ±12% by CO2SYS. Additionally, we will re-estimate the propagation errors in CO_2 efflux calculations using the Monte Carlo method.

For the width estimation model by flow discharge from Raymond et al. (2012), which is designed to the estimate of smaller rivers. For analysis the potential errors caused by this equation. We compare our results from different Strahler orders:

Most of the Elbe River's flow, categorized with Strahler orders from 1 to 6, matches the flow discharge range used to create the equation by Raymond et al. (2012) (Figure R1).

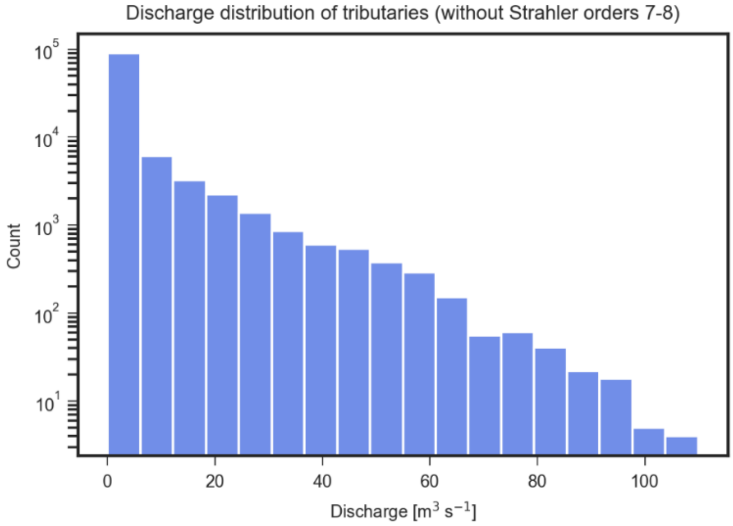

Figure R1. Flow discharge distribution of tributaries of the Elbe River. Flow discharge data obtained and resampled from GRADES (The Global Reach-scale A priori Discharge Estimates for SWOT) (Lin et al., 2019; Yang et al., 2019).

For the larger segments of the river, classified as Strahler orders 7 and 8, primarily the mainstem, we compared our estimated river widths with the research of Mallast et al. (2020). Their measurements were derived from satellite imagery. The average river width we estimated showed good agreement with their findings (this research: 177 m for Strahler order 7&8 (Figure R2), versus Mallast et al. (2020): 183 m, with an area of 107 km^2 divided by a length of 594 km).

Therefore, we believe the error introduced by our method in this research should be minor. An additional discussion of uncertainties will be added.

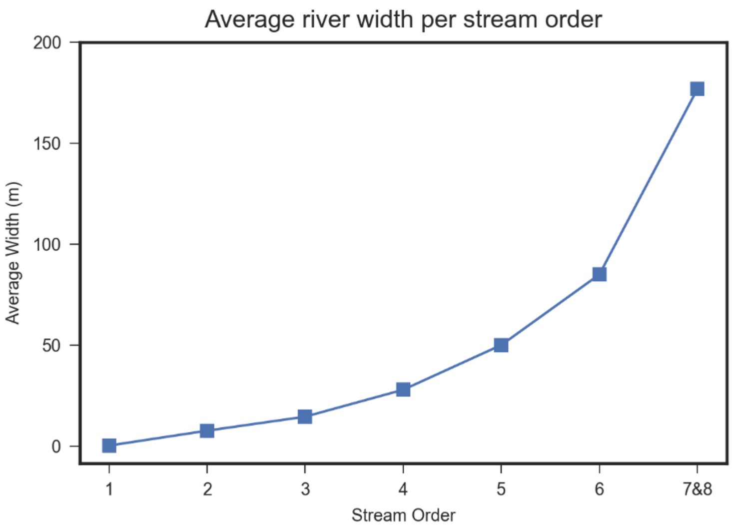

Figure R2: Estimated River width across different Strahler orders.

Another critical issue is with the use of discharge values for k600 estimates. From what I gather, the authors have used only one discharge value for each river location. This approach is problematic because k600 is highly influenced by discharge fluctuations, and failing to account for discharge fluctuations will result in erroneous CO2 emission flux estimates. This issue becomes evident in Figure 3f, where the relationship between FCO2 and pCO2 is almost perfectly linear—either suggesting that k600 has no influence on FCO2, or that k600 remains constant across space and time, both of which are improbable.

Reply:
In our study, the flow data used for calculations were extracted from the GRADES database (Lin et al., 2019; Yang et al., 2019), specifically selected to correspond with the dates of hydro-chemical data sampling. This database offers daily records of flow discharge, inherently accounting for the influence of flow variations on the seasonal k_{600} values. We will also provide the correlation analysis between variations in k_{600} and F_{CO2}.

Additionally, a comparative analysis between data from the GRADES database and actual measurements provided by hydrological stations will be conducted. A short discussion in long-term changes in discharge and the impact on F_{CO2} will be included.

A third issue is with the DOC data. It appears that two methods are used for the DOC flux estimation, yet only one is presented in the Results section. Furthermore, the first method does not present a way to calculate loads, but simply provides a framework for classifying C-Q patterns, which is rather confusing.

Reply:
Two calculation methods are described in the text, both founded on the principle of fitting the concentration to a model that utilizes the flow discharge to adjust the concentration. These approaches result in final errors that stem from the differences between the measured values and the values derived from model fitting.

Detailed explanations of this calculation process and uncertainties analysis will be provided in the methods section and in the supplementary.

Furthermore, upon comparison, the results from two methods show little differences. Therefore, we have applied the average of the two as the result.

Presentation of results
The results of statistical tests are not consistently reported throughout the paper. For example, Mann-Kendall trend test results are not presented for pCO2 and FCO2 (L231-261) as well as for DIC, DOC and POC (L276-291), making it challenging to assess the significance of the observed trends. Furthermore, there are no reported step change test results, despite the mention of these tests in the Methods section.

Reply:

The results of the Mann-Kendall test and the step change test for parameters such as pCO_2, F_{CO2}, DIC, DOC, and POC, etc., will be added.

I also noted some inconsistent statements between the results and discussion: while on L281 the authors state that “POC, DOC and DIC loads did not show significant trends”, this contradicts the following statement that the DOC load “showed relatively robust decreasing trend” (L310-311).

Reply:

According to the Mann-Kendall test, DOC load exhibit significant decreasing trends ($p<0.01$).

We will address this correction and thoroughly review the entire manuscript to avoid such mistakes.

Lastly, several figures are missing. For instance, the pCO2 time-series data are not shown despite these data being arguably one of the most critical data of the paper.

Reply:

A set of more comprehensive analyses, including additional time-series plots with pCO_2 and other parameters, will be added to the manuscript.

English language

The paper is very challenging to understand, and clearly the more senior authors (some of whom are well-published) have not provided the necessary feedback. It seems like only the abstract and the first few paragraphs of the introduction have been edited. The language used is awkward at best, and completely incoherent at worst. As a reviewer, I am not willing to invest one or two days correcting grammar and editing the entire paper. I strongly recommend that the senior authors fulfil their responsibilities of reviewing and editing this paper.

Reply:

We will fully revise the manuscript from all the co-authors.

References

Humphreys, M. P., Lewis, E. R., Sharp, J. D., & Pierrot, D. (2022). PyCO2SYS v1.8: marine carbonate system calculations in Python. *Geosci. Model Dev.*, *15*(1), 15-43. https://doi.org/10.5194/gmd-15-15-2022

Lewis, E., & Wallace, D. (1998). *Program developed for CO2 system calculations*.

Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., David, C. H., Durand, M., Pavelsky, T. M., Allen, G. H., Gleason, C. J., & Wood, E. F. (2019). Global Reconstruction of Naturalized River Flows at 2.94 Million Reaches. *Water Resour Res*, *55*(8), 6499-6516. https://doi.org/10.1029/2019WR025287

Mallast, U., Staniek, M., & Koschorreck, M. (2020). Spatial upscaling of CO2 emissions from exposed river sediments of the Elbe River during an extreme drought. *Ecohydrology*, *13*(6). https://doi.org/10.1002/eco.2216

Parkhurst, D. L., & Appelo, C. (2013). Description of input and examples for PHREEQC version 3—a computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations. *US geological survey techniques and methods*, *6*(A43), 497.

Raymond, P. A., Zappa, C. J., Butman, D., Bott, T. L., Potter, J., Mulholland, P., Laursen, A. E., McDowell, W. H., & Newbold, D. (2012). Scaling the gas transfer velocity and hydraulic geometry in streams and small rivers. *Limnology and Oceanography: Fluids and Environments*, *2*(1), 41-53. https://doi.org/10.1215/21573689-1597669

Yang, Y., Pan, M., Beck, H. E., Fisher, C. K., Beighley, R. E., Kao, S. C., Hong, Y., & Wood, E. F. (2019). In Quest of Calibration Density and Consistency in Hydrologic Modeling: Distributed Parameter Calibration against Streamflow Characteristics. *Water Resources Research*, *55*(9), 7784-7803. https://doi.org/10.1029/2018wr024178