Reply on RC1

Chinchu Mohan et al.

Author comment on "Poor correlation between large-scale environmental flow violations and freshwater biodiversity: implications for water resource management and the freshwater planetary boundary" by Chinchu Mohan et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2022-87-AC1, 2022

Authors sincerely thank Anonymous Referee #1 for their constructive comments which have helped to improve the article. We address each comment in turn below.

Note: All line numbers in responses correspond to the revised manuscript.

Comment 1.1: EF violation does not explain changes in biodiversity at the scales of consideration. It would seem that the authors have the tools and the datasets to address some of the major challenges they discuss. In particular, the authors recognize scale and scale matching as issues, and even discuss a solution for dealing with it (Line 266).

"*Aggregation of any scale will lead to some level of homogenization of the data. A reach-by-reach evaluation will be an ideal solution to capture all the heterogeneity*. However, this is not very practical for a global study due to data and computational limitations."

I would ask the authors: Why is this analysis impractical? None of the metrics being calculated in the manuscript are computationally complex, and the most complex statistical technique is to regress change in biodiversity onto change in EFmetric. Importantly, the authors list scale-matching as a potential explanation for their null finding (Line 445). Of all the limitations listed, this one seems the most straightforward to address in the present manuscript without needing to find new datasets or formulate a more complex model (e.g. reasons 1, 3, 4, 5 in Section 4.3).

There are likely other simple checks the authors could perform to explore whether the results might change if the scale-matching is performed differently. For example, how strongly do EF violations correlate within a watershed? Alternatively, if EF violation is observed at the outlet, do we see that an elevated fraction of sub-watersheds also exhibit EF violation (maybe test using some form of logistic regression)? Or, can the authors show that the results are more robust in the watersheds where scales ARE matched (e.g. Figure S4b)? These may help us to answer whether we might expect the results to change if the analysis was performed at a different (e.g. reach) scale, or if scales were more appropriately matched.

Response 1.1: As discussed in the revised manuscript (Line 294 -295), a reach by reach or a finer resolution comparison of the EF violation and biodiversity indicators might be an ideal way to capture the heterogeneity of the aquatic ecosystems. However, to our understanding, there is no fine resolution gridded datasets available for biodiversity except the ones derived based on streamflow deviations (Mean Species Abundance from GLOBIO-Aquatic, [*Janse et al.*, 2015]). Comparing two model-derived values rooted in streamflow deviations would not satisfy the overarching aim of this study, especially when both are based on the same underlying process assumptions, would not add insight into biodiversity responses to EF violations, and hence is not included in the manuscript.

Additionally, to partially address the heterogeneity limitation, in addition to our global study, we used the RivFishTIME dataset by Comte et al (2021) – compiled from long-term riverine fish surveys from 46 regional and national monitoring programmes and from individual academic research efforts – and repeated the analysis with the proposed EF violation indicators.

The results were consistent with the findings of the main study and there was no significant correlation between EF violation indicators and fish abundance data over time (see results for five selected fish species based on data completeness and geographical distribution in Supplementary Information section S8). The details of the RivFishTIME dataset and the findings using this dataset are also included in the revised manuscript (see Table 1, Line 180 - 186 and Line 359 – 368)

Line 180 -186: In addition to FiR, we used the RivFishTIME dataset by Comte et al (2021) – compiled from long-term riverine fish surveys from 46 regional and national monitoring programmes and from individual academic research efforts. Though the RivFishTIME dataset is highly spatially skewed towards the already data-rich regions of Europe, North America (particularly the United States of America), and Australia and temporally discontinuous, it is the only species-specific fish abundance time-series data available and is useful to have independent verification of the findings using FiR and relative biodiversity indicators.

Line 359-368: The increase in the fish assemblage over time was verified using an independent dataset RivFishTIME (see SI; Fig. S8, Fig. S9) (Comte et al., 2021). The increase in the fish richness facets primarily stems from the introduction of alien species introduced into streams for commercial purposes (Su et al., 2021). The invasion of alien species can tamper with the existing natural ecosystem equilibrium resulting in further degradation of the overall ecosystem health. The results using RivFishTIME data sets were also consistent with the findings using FiR and 6 relative biodiversity indicators and there was no significant correlation between EF violation indicators and fish abundance data over time (see results for five selected fish species based on data completeness and geographical distribution in Supplementary Information section S8; Fig. S8).

With respect to the scale matching, analysis similar to the one shown in the main manuscript are carried out using different aggregation/scale matching techniques and the results are included in the Supplementary Information (see Fig. S5 and S6). Additionally, the manuscript is revised to reflect the results from different scale matching techniques (see Line 295-299 and Line 379-380)

Additionally, as suggested by the Referee, when revising the manuscript we estimated the variance of EF violation indicators within the catchment boundary (consistent with Su et al.'s facets). The results were added to the Supplementary material (See section S10, Fig. S12 in Supplementary Information).

Fig. S12 EF violation indicators' coefficient of variance within fish facets data catchment boundary (Su et al., 2021): EF violation (a) frequency and (b) severity

Comment 1.2: There may be some confusion for "uninitiated" readers regarding terminology in the abstract. I suggest defining important terms, like "EF violation", "a planetary boundary for freshwater", etc. I was unfamiliar with some of these terms

Response 1.2: Definitions of uncommon terms are provided in footnotes along with the abstract.

Comment 1.3: Line 214: The authors exclude catchments with MAF < 10 cms. However, many low flows are seasonally observed, such that MAF may be quite large due to elevated wet season flows, with very low flows during a dry season. This is definitely the case in California. Even though many Coastal CA Level 5 watersheds have MAF > 10 cms, low flows during the dry season can be very small, and difficult to model (e.g. the Eel River, Level 5 basin 7050014040). Yet, many of the most important EF metrics are based on low flows, as these represent the period of the year when water is most limiting for ecosystems. In general, I'd like to hear a bit more about the success of the ISIMIP flow model in these seasonally dry low flow watersheds, given how finicky many low flow EF metrics can be. Section 2.1: In general, given the challenges associated with hydrological models, have the authors considered using the gage data used to cal/val ISIMIP? Understandably there will be no pre-industrial record, but presumably, trends in EF metrics could be calculated in "early" (e.g. 1980) versus "late" (e.g. 2015) periods, and any associated trend in biodiversity metric could be explored? This would circumvent any issues with low flow modeling.

Response 1.3: All the GHM outputs used in this study are extensively validated and evaluated in several previous studies (e.g. Gädeke et al., 2020; Zaherpour et al., 2018). Moreover, as part of the ISIMIP impact model intercomparison activity, all the GCM climate input data were bias-corrected using compiled reference datasets covering the entire globe at 0.5 deg resolution (Frieler et al., 2017). Additionally, the GHM outputs are also validated using historical data to better fit reality (Frieler et al., 2017). Though the seasonal performance of ISIMIP data is not conducted at the global scale, there are several studies that evaluate the seasonal performance of GHMs at large basin scales (Huang et al., 2016; Gädeke et al., 2020; Zaherpour et al., 2018). All these studies report reasonable performance in capturing the seasonal dynamics of the GHMs. We thus think that performing a global-scale validation of discharge is not required again, and is beyond the scope of an application study like this (see lines 151 - 156). The authors, however, agree that the current analysis was carried out at an annual time step which overlooks the seasonal variations in the EF-biodiversity relationships. In the revised manuscript, we will take note of this special case where the intra-annual variability in discharge is very high and implicate that higher detail - both in sub-basin catchment boundaries and sub-annual discharge data - would be required for practical evaluation uncertainty in these cases (see lines 230-234)

Comment 1.4: Line 191 Are the biogeographical realms just the base spatial units of the biodiversity datasets? E.g. the gray shapes in Figure S4b?

Related to the previous question, I'm having some trouble understanding some of the biodiversity metrics, and how they relate to scale. This is probably just some confusion on terminology on my part; I'm a bit new to this particular topic.

So, for example, can the authors more clearly define "dissimilarity" (Line 181)? It is stated that it, "...accounts for the difference between each pair of fish assemblage in one biogeographical realm." It would be really helpful to have some basic equations here, and some explanation of how the calculations correspond to the different scales of aggregation discussed in the flow section and the aggregation section.

Response 1.4: Biogeographical realms (ecoregions) are the spatial units used in this study for classifying the results into hydro-ecologically similar groups. The grey shapes in Fig. S4b are the spatial scale of relative freshwater fish facets (TR, FR, PR, TD, FD, PD) obtained from Su et al. 2021. The freshwater fish richness data (FiR), from Tedesco et al., 2017, however, is at 30 arc second and is restricted to 3119 drainage basins. The spatial and temporal scale of individual data is included in Table S1 in Supplementary Information.

In order to better visually explain the concept of richness and dissimilarity fish facets, a reference to figure 1 in Su et al., 2021 is added in the manuscript (Lines 201-205). Additionally, the following sentence explaining the how the calculations correspond to the different scales of aggregation

Line 201-205: "The scale at which the fish facets are estimated does not necessarily align with the scale at which the EF violations are estimated in all cases. The basin scale facet estimates were then matched with corresponding EF violation indices using different aggregation/data matching methods (see Section 2.4 for more details)."

Comment 1.5: line 231 - First it is stated this is calculated as the absolute mean of the deviation magnitude, but then it is normalized? Should this be interpreted as a percent change in the mean? Is this how the other metrics (e.g. F) are also normalized?

Response 1.5: Annual violation severity (S) was calculated by taking the mean of the magnitude of monthly deviation beyond upper or lower EFE boundary. The magnitude of violation is based on the violation ratio proposed by Virkki et al. 2022.

Table 2 from Virkki et al., 2022. Computing the EFE violation ratio. Q stands for monthly discharge between 1976 and 2005; EFE_{lower} for the EFE lower bound, and EFE_{upper} for the EFE upper bound

According to this definition, the lower bound violations will have negative magnitude while the upper bound will have positive magnitude. Therefore, the absolute values were taken in this study to avoid mutual cancellation of the upper and lower bound violations. Additionally, in order to make the different EF violation indices comparable, the values of violation indicators (F and S) were scaled (or normalized) between 0 to 1 using the following formula

where, X_{normal} = normalised value; X = actual value; X_{min}= minimum value in the dataset; X_{max} = maximum value in the dataset

Comment 1.6: Line 236: On the probability of a shift from nonviolated to violated. Is total years in the denominator incorrect? I would think it would be conditional on the occurrence of a nonviolated state, as you can't shift from nonviolated to violated if you're currently in a violated state.

Response 1.6: The logic behind using the total number of years in the denominator was to estimate the probability to shift given the entire period of time. This enabled easy and logical comparison between different regions. Authors, however, agree with the Referee that using nonviolated years in the denominator is also an alternate way of looking at this shift.

Please also note the supplement to this comment:
https://hess.copernicus.org/preprints/hess-2022-87/hess-2022-87-AC1-supplement.pdf