Response to Comments on EF-biodiversity paper.

Comments on EF violation indicators

1) **Reviewer comment:** "The flow data in this study are based on simulated outputs (GCM-GHM linkage) with no reported validation of the flow estimates, even for contemporary time periods where streamflow gauge data are available. Absent such validation, there is unspecified uncertainty in the accuracy of the driver variables and, accordingly, uncertainty in the precision of conclusions based on these simulated variables"

Response: 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). We thus think that performing a global-scale validation of discharge is not required again, and beyond the scope of an application study like this. We do agree, however, that the validation of the used discharge data was not communicated properly in the manuscript and this is now added to the revised manuscript (see lines 148 - 153).

2) **Reviewer comment:** No evidence for reduction of uncertainty with the use of multiple GHM, GCM and EF methods

Response: We do agree with the reviewer that our statement was not fully accurate. The use of multiple GHMs, GCMs and EF methods didn't directly reduce the uncertainty in the study but provided a good understanding of the uncertainties involved in an analysis of this scale. The revised manuscript describes the use of multiple GCMs, GHMs and EF methods as a way of understanding uncertainty (see line 209 - 214), and states that the study is based on a robust ensemble methodology and the uncertainty behind the violation estimates is well accounted for.

3) **Reviewer comment:** No discussion of the well known "fact" that GCMs are notoriously poor at estimating precipitation outputs and "corrections" are needed before their application (e.g., see Woldemeskel et al., 2014, "A framework to quantify GCM uncertainties for use in impact assessment studies," J.Hydrol. 519(B), 1453-1465).

Response: The authors agree that raw GCMs outputs often exhibit biases and inter-model differences when it comes to precipitation in particular. However, This study does not use the GCM outputs directly, but instead uses prepared outputs from ISIMIP-2b that provide better estimates for water-related variables (Frieler et al., 2017). As discussed in response 1, the GCM forcings used in the ISIMIP are bias corrected with EWEMBI dataset (0.5 deg; daily; 1979-2013) using a first-order conservative remapping scheme. Moreover, the GCM outputs also come from CMIP5, which is also a well-established intercomparison project.

4) **Reviewer comment:** Not describe how the “five different EF calculation models” (lines 218-221) were used to extract the EFEs

Response: Additional descriptions on the EF methods are added in the supplementary information (see Table S3 in supplementary information).

5) **Reviewer comment**: Eliminating the “upper bound” of a flow “envelope” presents a fundamental problem in adequately characterizing the effects of hydrologic alteration. Philosophically, the use of the word “envelope” for a metric that has only one boundary is not logical! It would be better to call it the “environmental flow lower limit” or similar.

Response: The initial rationale of excluding the upper boundary violation was that it occurs only in few aquatic biodiversity hotspots. However, all the results are now revised by including both upper and lower boundary violations to capture the effect of very high flows on fish and amphibians.

6) **Reviewer comment**: Many of the river basins used in this analysis will have dams but the authors have not reported on the occurrence or density of dams in the level 5 HydroBASINS, so we cannot exclude the likelihood the estimated EFEs are inadequate to appropriately capture the hydrologic alteration at basin outlets.

Response: The dams are indeed a large contributing factor to the results uncertainty. Dams are represented in the ISIMIP data used here, albeit somewhat differently between the different hydrological models (see e.g.(Masaki et al., 2017)). The large dams (~844 dams around the world from GRanD dataset) are considered as a human influence driver in the ISIMIP historical simulation. However, in order to explicitly isolate the effects of dams in this analysis from other drivers, the information on dam operation schemes for each sub-basin would be necessary and this would require a paper on its own. Therefore, the effects of the dams are incorporated in this study but are not explicitly analyzed separately from other drivers. The lack of explicit consideration of dams is discussed as a limitation of the work and we suggest it as a potential future improvement of this study and when the required data is available (see lines 466 - 472).

Comments on Biodiversity

7) **Reviewer comment**: "The method of richness aggregation imposes a homogenizing effect of richness values across sub basins within a level 5 HydroBASIN catchment, and it ignores the very likely variation in flow alteration associated with the sub-basins."

"heterogeneous habitat types and samples are simply combined to create an aggregate species list at the basin level. It is overly simplistic to construct a hypothesis that posits that this aggregated diversity number can be explained by a single environmental factor (such as EFE indicators) derived for the whole of the basin (basin outlet) that is homogeneously applied to all the local diversity sampling locations. "

Response: The fish dataset used in this study are reported at a Hydro Basin scale and all reported biodiversity datasets had undergone rigorous quality evaluation including but not limited

to local or regional checks (Su et al., 2021; Tedesco et al., 2017). Moreover, the authors of this paper had also used different methods to match the scales of EF violation and biodiversity data (see SI, section S2). However, we agree that some amount of heterogeneity is lost with any aggregation. 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. Given the scale of this study, HydroBasins scale level 5 was selected because it can be rasterized into a 0.5-degree resolution grid without losing too much detail that is smaller than a grid cell. Therefore, the results from this study are less accurate for smaller subbasins, for which we recommend to use observed data with high detail as opposed to global data simulated on a coarse grid.

8) **Reviewer comment**: The expectation that amphibian richness will be a function of flow alteration is not justified. Very few amphibian species are obligate stream/river dwellers and are thus unlikely to be vulnerable to flow alteration. Indeed, the amphibian data for this analysis were not derived from river surveys; rather, they are simply extracted from a global terrestrial coverage that is “cut out” to fit the basins

Response: The initial rationale in including the amphibians in this study was to increase the biotic representation in this study. However, considering the suggestion from the reviewer, the study is received with only freshwater fish data.

9) **Reviewer comment**: "fish richness (FiR) varies NATURALLY across latitude and climate regions and historical isolation – i.e., freshwater fish biogeography. The paper does not adjust for this natural variation (e.g., the whole of Europe comprises fewer fish species than single sub-basins in the Amazon region)." "there is a fundamental scaling problem with assembling the fish richness data at the whole basin scale to test the overarching EF-biodiversity hypothesis. This limitation is a fundamental challenge to any testing of “global” scale hydrologic alteration and global scale diversity data, which are comprised of aggregated point values"

Response: The paper does account for some of the natural/spatial variability by considering the main categories of Freshwater ecoregions (G200). However, the authors agree that the study does not capture all the variability in freshwater fish across climate and latitude. In Section 3.2, along with the global analysis, the relation between EF-biodiversity was also evaluated at seven G200 freshwater ecoregions. The G200 freshwater ecoregions are classified to encompass relatively large units of land/water with similar characteristics and coincide with ecologically relevant areas level (Olson and Dinerstein, 2002).

10) **Reviewer comment**: In the present study, the authors do not provide any justification as to why one would expect a correlation between fish diversity and EFE

Response: Several studies show that all principal threats to freshwater ecosystems (overexploitation, water pollution, fragmentation, destruction or degradation of habitat and invasion by non-native species) are linked to flow alteration (Arthington et al., 2010). In addition, numerous hydro-ecological principles illustrate the myriad ecological functions of flow (Arthington et al., 2010; Pinay et al., 2002; Lytle and Poff, 2004). However, these principles and paradigms

should be accompanied by methods and models capable of accurately linking flows and biodiversity and ecological processes. Therefore this study is testing whether the widely used EF estimation methods actually captures this relationship. The motivation for conducting such an exploratory analysis was due to the fact that several large scale EF estimation methods are derived based on this assumption but none have tested whether the correlation holds at the global scale.

Comments on Hypothesis/Methodology

11) **Reviewer comment**: Scaling: there is a fundamental scaling problem with assembling the fish richness data at the whole basin scale to test the overarching EF-biodiversity hypothesis

Response: We agree with the reviewer that a lot of variability is lost due to this step; given the scale of this study, HydroBasins scale level 5 was selected because it can be rasterized into a 0.5-degree resolution grid without losing too much detail that is smaller than a grid cell. Therefore, the results from this study are less accurate for smaller subbasins, for which we recommend to use observed data with high detail as opposed to global data simulated on a coarse grid. We added this as a limitation in the discussion section.

12) **Reviewer comment**: The masking of local scale hydrologic factors can affect diversity. It is overly simplistic to construct a hypothesis that posits that this aggregated diversity number can be explained by a single environmental factor (such as EFE indicators) derived for the whole of the basin (basin outlet) that is homogeneously applied to all the local diversity sampling locations.

Response: Again, we agree to this point. However, as discussed in the responses of previous comments, this paper is neither constructing nor testing a hypothesis on the relationship between flow violation and biodiversity, but tests a widely used but rarely verified assumption on the relationship at global and ecoregion scale. Moreover we have also included the aspect of other confounding environmental factors that might strongly influence the result into the discussion/limitation section.

13) **Reviewer comment**: The implication seems to be that the approach taken in this paper is some kind of definitive test of the EF-biodiversity hypothesis at a global scale. Surely, there are other approaches that could also result in negative results! This single example of negative results should not be considered the final say on the matter.

Response: The paper is not intended to be a definitive test to disprove the relationship between EF and aquatic biodiversity. It is intended to be an exploratory analysis to identify the validity of the relation. The primary motivation for conducting such a study is due to the fact that, majority of the methods used to estimate EF operate at a coarser scale with an underlying assumption that the proportion of flow allocated directly impacts the ecosystem health. However, what this study is doing is reevaluating this assumption. The authors of this paper in no way intend to disregard the importance of flow, but instead were aiming to estimate the usability of large scale generalized EF estimation methods by evaluating its relationship to aquatic biodiversity indicators. The single

negative result is not a final say but it is a call for conducting more study on existing generalized and well applied methods

14) **Reviewer comment**: It seems that this is the crux of the paper and puts the lack of relationship between EF deviation and biodiversity into proper context, yet it receives very limited discussion. There remains a risk that a reader only draws the conclusion that EF violations have no effect on global freshwater biodiversity, therefore EF violations need not be considered important.

Response: Authors acknowledge the risk of reporting a non correlation between EF and biodiversity. However, as described in the response for comment 13, the aim of this study is to test the EF biodiversity relationship at a larger scale. In order to avoid the risk of misjudgement by the readers, we have strengthened the discussion that our findings are only applicable at global or ecoregion scale and with currently available data. At a scale smaller than this, several studies have already proved the importance of flow for maintaining ecosystem services.

Reference

Arthington, Á. H., Naiman, R. J., Mcclain, M. E., and Nilsson, C.: Preserving the biodiversity and ecological services of rivers: new challenges and research opportunities, 55, 1–16, 2010.

Frieler, K., Lange, S., Piontek, F., Reyer, C. P., Schewe, J., Warszawski, L., Zhao, F., Chini, L., Denvil, S., and Emanuel, K.: Assessing the impacts of 1.5 C global warming–simulation protocol of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP2b), 10, 4321–4345, 2017.

Gädeke, A., Krysanova, V., Aryal, A., Chang, J., Grillakis, M., Hanasaki, N., Koutroulis, A., Pokhrel, Y., Satoh, Y., and Schaphoff, S.: Performance evaluation of global hydrological models in six large Pan-Arctic watersheds, 163, 1329–1351, 2020.

Lytle, D. A. and Poff, N. L.: Adaptation to natural flow regimes, 19, 94–100, 2004.

Masaki, Y., Hanasaki, N., Biemans, H., Schmied, H. M., Tang, Q., Wada, Y., Gosling, S. N., Takahashi, K., and Hijioka, Y.: Intercomparison of global river discharge simulations focusing on dam operation—multiple models analysis in two case-study river basins, Missouri–Mississippi and Green–Colorado, 12, 055002, 2017.

Olson, D. M. and Dinerstein, E.: The Global 200: Priority ecoregions for global conservation, 199–224, 2002.

Pinay, G., Clément, J. C., and Naiman, R. J.: Basic principles and ecological consequences of changing water regimes on nitrogen cycling in fluvial systems, 30, 481–491, 2002.

Su, G., Logez, M., Xu, J., Tao, S., Villéger, S., and Brosse, S.: Human impacts on global freshwater fish biodiversity, 371, 835–838, 2021.

Tedesco, P. A., Beauchard, O., Bigorne, R., Blanchet, S., Buisson, L., Conti, L., Cornu, J.-F., Dias, M. S., Grenouillet, G., and Hugueny, B.: A global database on freshwater fish species occurrence in drainage basins, 4, 1–6, 2017.

Zaherpour, J., Gosling, S. N., Mount, N., Schmied, H. M., Veldkamp, T. I., Dankers, R., Eisner, S., Gerten, D., Gudmundsson, L., and Haddeland, I.: Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts, 13,

065015, 2018.