Comment on hess-2022-87

Anonymous Referee #1

Referee 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-RC1, 2022

Review of "Poor correlation between large-scale environmental flow violations and freshwater biodiversity: implications for water resource management and water planetary boundary" by Moher et al, for possible publication in HESS.

Summary

The authors use results from global-scale hydrological modeling to calculate changes in environmental flow (EF) metrics between an un-altered, pre-industrial (1800's) period and an altered contemporary (1976 - 2005) period, then attempt to correlate changes in EF metrics to changes in biodiversity metrics over the same time periods. In most instances, the authors do not find any significant detrimental effect of EF violation on biodiversity.

General comments

The authors were very thorough in the layout and logical flow of this topically important contribution. As I was reading, a number of "red flag" issues came up in my mind, but these were then addressed either through the development of an appropriate method, or were satisfactorily caveated and discussed. For example, shouldn't a relative biodiversity change metric be used? What are the different ways to deal with scale matching the flow and biodiversity datasets? How do we deal with low flow watersheds given the challenge of hydrologically modeling low flows? The authors recognize all of these issues, and address them.

That being said, the lack of resolution of some of these issues is a bit disappointing, especially given the null result of the paper: 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.

Specific comments

Abstract: 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.

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.

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.

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?

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.

Many references in the bibliography do not have a journal.