Environmental flow envelopes: quantifying global, ecosystem-threatenning streamflow alterations

General comments
The authors quantify the environmental flow violation for most worldwide sub-basins using environmental flow envelopes (EFE). EFE are based on previously established minimum environmental flow requirements and a newly established maximum environmental flow requirement. EFE quantification was based on pre-industrial simulations of several global hydrological models (GHM), under various global circulation models (GCM), and applied to historical simulations for each GHM. Results inform about the frequency and severity of EFE violation, and could thus indicate the state of riverine ecosystem security.

Environmental flow is a concept often used in hydrological modelling in order to simulate sustainable water use scenarios that do not interfere with riverine ecosystem security. Moreover, they can be a useful tool to determine sustainable dam and water management policies. The authors aim to reduce the uncertainty in the current (historical) violation of these requirements, and to expand the violation assessment to the worlds sub-basins, by using a GHM-GCM model ensemble. However, I have some concerns regarding the EFE quantification methods, the attribution of EFE violation, and some general comments

EFE quantification methods
The authors aim to advance environmental flow assessments in four ways: (1) improve the environmental flow concept to include an upper bound (line 95, line 114, line 186, line 458), (2) reduce the uncertainty in the actual quantification of environmental flows, in particular the large GHM uncertainties (line 87, line 92, line 112, line 409), (3) use the pre-industrial river streamflow as the environmental flow basis (line 116, line 160, line 460), and (4) present the quantification of environmental flows in terms of frequency, severity and trends (line 98, line 115, line 458).
1. While I understand the adverse effects of upper EFE violation, the selection of the upper bound seems arbitrary. What is the reason for the 95th percentile?

2. I feel this study does not properly address the GHM uncertainty. The ensemble strategy is only applied to the EFR methods and GCMs, not the GHMs. EFE violations are still determined for each model separately, without averaging, which can lead to additional (compounding) uncertainties. Moreover, if the aim of the authors is to provide accurate quantifications of EFE for each river, why did the study not use observational data or select the “best” performing GHM model for each basin?

3. What is the rationale behind using pre-industrial river streamflow? Should the simulated historical pristine (no human influence) river streamflow not be better to estimate EFE (as is done by other modelling studies)? It is briefly discussed that historical pristine river streamflow may already be violating EFE (line 399). Should this be the case?

4. I am assuming monthly pre-industrial EFE quantifications are combined into a multi-year monthly averages. However, the EFE are applied to historical monthly streamflow values. This could result in EFE violation during dry years (even if EFE were applied during the pre-industrial period). Is this correct?

**Attribution of EFE violation**

The authors describe the EFE violation in terms of frequency, severity and trends. To that end they use the gross total of all 1440 historical GHM-year-month combinations (line 232). Subsequently the key drivers are shortly explained in the discussion (line 353).

5. I assume the authors use multi-year monthly averaged pre-industrial simulations to quantify the EFE and subsequently apply these values to the historical period (see my comments above). Can the authors show the impact of this choice on your results? It would be useful to separate the effects of climatic changes (human induced or otherwise) from direct human alterations (e.g. abstractions and dam construction). Moreover, it would be useful to see the temporal variability in EFE violation, especially during dry years.

6. Results are given as the gross total of all GHM-year-month combination. However, the results do not show the variability between the GHMs. Could it be that a majority of the EFE violations originate from a single model?

7. The authors state that their findings show that EFE violations are widespread around the world (line 239). However, to me the results shown seem quite modest (about half the basins violate EFE for 5% of the time, not even a month per year). Even more so as dry years could contribute to this number. How do these results change for 10% or 25% of all months?
Specific comments
1. Line 110 and 658: “The significantly advanced global environmental flow assessments with the novel methodology of environmental flow envelopes” is overstated. The concept of EFE is heavily based on previous lower limit environmental flow requirements, while the newly introduced upper limit is unsubstantiated (in this study). Rather the application of the EFE methodology is new.

2. Line 177 and 178: Why are high and low flows omitted? Should they not be part of the natural flow regime (wet and dry periods)?

3. Line 241: Do these results still take into account the three consecutive month exceedance threshold (line 210)?

4. Line 284: Is it implied this trend is due to changes in climate? If so, a thirty year period to detect climate trends is rather short.

5. Line 296: It is interesting to see that the trend in upper bound EFE violation mostly increases in the northern regions, while the lower bound EFE violation mostly decreases in these regions. Do the authors know if this is a climatic or direct human alteration effect?

6. Figure 5: This figure is a little confusing to me. Firstly there is no y axis for the trend slope graph. Secondly there are clusters that have prevalent violations during all seasons, but subsequently do not display violations during for example the high flow months? Lastly, more information could be given about the clustering and why some areas could not be assigned to a cluster.

7. Line 329: The first section of the discussion could be incorporated in the conclusions.

8. Line 430: As the aggregated monthly sub-basin assessment restricts the conclusions that can be made, why did the authors not do a daily gridded assessment? The data is available.

9. Line 329 and 461: Where is it shown that the flow alterations reported here are recurrent or long-standing?

Technical corrections
1. Line 51: Should “because of their importance” be removed?