Environmental flow envelopes: quantifying global, ecosystem–threatening streamflow alterations

Vili Virkki1*#, Elina Alanärä1#, Miina Porkka1,2, Lauri Ahopelto1,3, Tom Gleeson4,5, Chinchu Mohan4,6, Lan Wang-Erlandsson7,8, Martina Flörke9, Dieter Gerten10,11, Simon N Gosling12, Naota Hanasaki13, Hannes Müller Schmied14,15, Matti Kummu1*

1 Water and Development Research Group, Aalto University, Espoo, Finland
2 Global Economic Dynamics and the Biosphere, Royal Swedish Academy of Sciences, Stockholm, Sweden
3 Finnish Environment Institute, Helsinki, Finland
4 Department of Civil Engineering, University of Victoria, Victoria, British Columbia, Canada
5 School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia, Canada
6 Global Institute for Water Security, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
7 Stockholm Resilience Centre, Stockholm University, Stockholm, Sweden
8 Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden
9 Institute of Engineering Hydrology and Water Resources Management, Ruhr-University Bochum, 44801, Bochum, Germany
10 Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Potsdam, Germany
11 Humboldt-Universität zu Berlin, Geography Department, Berlin, Germany
12 School of Geography, University of Nottingham, Nottingham, NG7 2RD, United Kingdom
13 National Institute for Environmental Studies, Tsukuba, Japan
14 Institute of Physical Geography, Goethe University Frankfurt, Frankfurt am Main, Germany
15 Senckenberg Leibniz Biodiversity and Climate Research Centre (SBiK-F), Frankfurt am Main, Germany

# equal contribution to the article
* Correspondence to: Vili Virkki (vili.virkki@aalto.fi), Matti Kummu (matti.kummu@aalto.fi)

Abstract. Human actions and climate change have drastically altered river flows across the world, resulting in adverse effects on riverine ecosystems. Environmental flows (EFs) have emerged as a prominent tool for safeguarding riverine ecosystems. However, at the global scale, the assessment of EFs is associated with significant uncertainty. Here, we present a novel method to determine EFs by Environmental Flow Envelopes (EFE), which is an envelope of variability bounded by discharge limits within which riverine ecosystems are not seriously compromised. The EFE is defined globally in approximately 4,400 sub–basins at monthly time resolution, considering also the methodological uncertainties related with global EF studies. In addition to a lower bound of discharge, the EFE introduces an upper bound of discharge, identifying areas where streamflow has increased substantially. Further, instead of only showing whether EFs are violated, as commonly done, we quantify, for the first time, the frequency, severity, and trends of EFE violations, which can be considered as potential threats to riverine ecosystems.

We use pre–industrial (1801–1860) quasi-natural discharge and a suite of hydrological EFR methods and global hydrological models to estimate EFE, applying data from the ISIMIP 2b ensemble. We then compare the EFEs to recent past (1976–2005)
discharge to assess the violations of the EFE. We found that the EFE violations most commonly manifest themselves by insufficient streamflow during the low flow season, with less violations during intermediate flow season, and only few violations during high flow season. These violations are widespread: discharge in half of the sub–basins of the world has violated the EFE during more than 5% of the months between 1976 and 2005. The trends in EFE violations have mainly been increasing during the past decades and will likely remain problematic with projected increases in anthropogenic water use and hydroclimatic changes. Indications of excessive streamflow through EFE upper bound violations are relatively scarce and spatially distributed, although signs of increasing trends can be identified and potentially attributed to climate change. While the EFE provides a quick and globally robust way of determining environmental flow allocations at the sub–basin scale, local fine–tuning is necessary for practical applications and further research on the coupling between quantitative discharge and riverine ecosystem responses is required.

1 Introduction

The human exploitation of rivers is a sensitive balance between benefits gained from water use and adverse Earth system responses. While also enabling the development of societies, rivers upkeep two major regulatory Earth system functions: maintaining the hydrological cycle and providing habitat for freshwater ecosystems (Gleeson et al., 2020). Regardless and because of their importance, rivers are under increasing anthropogenic pressure due to direct human actions, such as flow regulation and damming, excessive water withdrawals, pollution, and land use change (Best, 2019; Kummu et al., 2016). Moreover, human–induced climate change can increase or decrease the seasonal streamflow at different spatial scales (Arnell and Gosling, 2013; Asadieh and Krakauer, 2017; Gudmundsson et al., 2021; Moragoda and Cohen, 2020; van Vliet et al., 2013). The pressure on freshwater ecosystems is only expected to increase in the future due to population growth and projected climate change (Best, 2019; Graham et al., 2020).

Freshwater ecosystems consisting of rivers, lakes, wetlands, and other freshwater features contain nearly 6% of all known species concentrated in 0.8% of Earth’s surface (Dudgeon et al., 2006). The riverine parts of freshwater ecosystems have been seriously compromised by human actions: rivers containing 65% of the global discharge are classified under moderate to high threat in terms of biodiversity, 53% of the global rivers have experienced marked changes in fish biodiversity, and 48% of global river reaches are impaired by diminished connectivity (Grill et al., 2019; Su et al., 2021; Vörösmarty et al., 2010). One of the root causes behind this degradation is the anthropogenic alteration of the natural flow regime of a river, i.e. the magnitude, frequency, duration, timing, and rate of change in flow (Poff et al., 1997). These natural streamflow dynamics have already changed in major rivers across the globe (Grill et al., 2015). Due to its profound effect on the physical habitat of the river, the flow regime is one of the key factors in defining the well–being of riverine ecosystems and maintaining their ecological status (Bunn and Arthington, 2002; Poff and Zimmerman, 2010).
To safeguard riverine ecosystems, the concept of environmental flows (hereafter EF; often used interchangeably with ecological flows) has emerged during the past decades (Poff and Matthews, 2013). While multiple definitions of EF exist, the most comprehensive recent definition comes from The Brisbane Declaration 2018 (Arthington et al., 2018), which states that “Environmental flows describe the quantity, timing, and quality of freshwater flows and levels necessary to sustain aquatic ecosystems which, in turn, support human cultures, economies, sustainable livelihoods, and well-being.” The concept of EFs is often quantified by computing environmental flow requirements (EFRs, sometimes also environmental flow needs), which refer to the minimum discharge required to sustain healthy and functional riverine ecosystems (Pastor et al., 2014). Hence, the EFR corresponds to a boundary not to be transgressed. Beyond simple EFRs, more nuanced quantification of anthropogenic impacts on discharge based on a multitude of different metrics include e.g. the Indicators of Hydrological Alteration (IHA; Richter et al., 1997, 1996). To date, EF assessments have become well-established parts of conserving and restoring riverine ecosystems and are implemented in the legislation of many countries (Acreman et al., 2014; Arthington et al., 2018; Tickner et al., 2020).

Ideally, EFs would incorporate in situ data and local expert knowledge to determine EFRs consistent with actual ecosystem water needs of each river, however, this data is unavailable at the global scale. Thereby, global studies accommodating EFs rather use hydrological EFR methods that express the EFR as a share of discharge on a specific timescale, considering it as a viable proxy for riverine ecosystem well-being (e.g. Gerten et al., 2020, 2013; Hanasaki et al., 2008; Hoekstra and Mekonnen, 2011; Hogeboom et al., 2020; Jägermeyr et al., 2017; Pastor et al., 2019, 2014; Steffen et al., 2015). However, the underlying discharge data based on which global studies often determine EFRs is uncertain: runoff and discharge estimated by Global Hydrological Models (GHMs) that are forced with modelled climate from General Circulation Models (GCMs) tend to be highly dispersed between different models (Dirmeyer et al., 2016; Gädeke et al., 2020; Hattermann et al., 2018; Müller Schmied et al., 2016; Schewe et al., 2014; Veldkamp et al., 2018; Zaherpour et al., 2019). As the GHM outputs are generally uncertain, determining EFRs based on GHMs and hydrological EFR methods is equally uncertain. Moreover, hydrological EFR methods often set only a minimum discharge boundary, disregarding the potentially adverse effects of flows increasing significantly above natural levels especially in floodplain ecosystems (Hayes et al., 2018; Junk et al., 1989; Schneider et al., 2017; Talbot et al., 2018). Although reviews of EFs have recognised this threat of excessive flows (Acreman et al., 2014; Poff and Zimmerman, 2010; Richter, 2010), no global scale methodology exists yet to quantify it.

In addition to the methodological uncertainties, existing global studies are also limited in their EF violation assessment. Commonly, EFs are treated as simple limits that are either violated or not, lacking in quantifying either how frequently or how severely these violations manifest themselves (Gerten et al., 2020; Pastor et al., 2019; Steffen et al., 2015). Some of the more detailed studies incorporate additional factors, such as the magnitude with which EFs are violated, but lack in accounting for the seasonality of streamflow (Hogeboom et al., 2020; Jägermeyr et al., 2017). Given that particularly low flows are often the most impacted by anthropogenic actions, such as water withdrawals and flow regulation by damming (Döll et al., 2009;...
Schneider et al., 2017), EF assessments should be able to separate violations during different flow seasons. Finally, while recent studies have shown that river flows have changed considerably due to direct human actions (Graham et al., 2020; Müller Schmied et al., 2016) and climate change (Gudmundsson et al., 2021; Moragoda and Cohen, 2020) during the past decades, no study has yet assessed the past trends in EF violations. Therefore, new knowledge is required to compose a combined and comprehensive outlook on these three aspects of EF violation.

Here, we show a significant advance in global EF assessment by introducing and applying a robust, global–scale methodology of Environmental Flow Envelopes (EFEs). Defined at the sub–basin scale in monthly time resolution, the EFE is an envelope of safe discharge variability that addresses the pitfalls of existing global studies. First, to reduce uncertainties in global EF assessments, the EFE is composed of a number of hydrological EFR methods applied to an ensemble of GHM outputs simulated using multiple GCMs. In addition, we newly suggest to include an upper bound of the EFE, aiding in identifying areas where streamflow has increased above the EFE. Second, we present a novel quantification of the seasonal frequency, severity, and trends of EFE violations by comparing recent, anthropogenically influenced discharge to pristine state EFEs. For the first time, this pristine state is estimated by pre–industrial (1801–1860) discharge.

2 Methods and data

Estimating EFE violations was divided into three parts, which are outlined in Fig. 1 and detailed in the following sections. First, we obtained ISIMIP 2b simulated discharge data from four global hydrological models (GHMs; H08, LPJmL, PCR-GLOBWB, and WaterGAP2). The GHMs model the global terrestrial hydrological cycle through mechanistic equations. Each of the four GHMs is parameterised with modelled climate from four different general circulation models (GCMs), thereby providing us with 16 data sets of gridded daily discharge. First, for each distinct combination of GHMs and GCMs, we transformed the gridded daily discharge to monthly discharge at the sub–basin scale according to HydroBASINS sub–basin division, both for the pre–industrial (1801–1860) and the recent past (1976–2005) period. Second, we estimated the EFEs for each GHM using pre–industrial discharge and five hydrological EFR methods for all GHM–GCM combinations separately. Finally, we compared the recent past discharge to the EFEs to estimate the frequency, severity, and trends of EFE violations, again for each GHM separately.
1. Pre-processing of discharge data

4 GHMs, each with 4 GCMs \( n_{\text{ensemble}} = 16 \)

- Daily discharge (1801–1860)
- Aggregate daily data to monthly
- Aggregate grid scale data to sub-basin scale

Daily gridded discharge to monthly discharge at the sub-basin scale

4 GHMs, each with 4 GCMs \( n_{\text{ensemble}} = 16 \)

- Aggregate daily data to monthly
- Aggregate grid scale data to sub-basin scale

Monthly discharge (1801–1860)

\( n_{\text{ensemble}} = 16 \)


\( n_{\text{ensemble}} = 4 \)

2. Defining EFESs based on pre-industrial (1801–1860) discharge

- Monthly discharge (1801–1860) 
- Estimate EFRs in sub-basins using 5 EFR methods, \( n_{\text{GHM}} = 20 \) (4 GCMs and 5 EFR methods for a GHM)
- Take 50th percentile of the EFR distribution = lower EFE boundary

- Take 95th percentile of discharge for each GHM = upper EFE boundary

- Monthly EFE for each GHM separately

3. Comparing recent past (1976–2005) discharge to EFESs

- Take median over GCMs

- Compare discharge to EFE

- Monthly EFE for each GHM separately

<table>
<thead>
<tr>
<th>Legend</th>
</tr>
</thead>
<tbody>
<tr>
<td>EFE upper bound</td>
</tr>
<tr>
<td>EFE lower bound</td>
</tr>
<tr>
<td>Monthly discharge</td>
</tr>
<tr>
<td>Individual EFR</td>
</tr>
<tr>
<td>EFE violation</td>
</tr>
</tbody>
</table>

**Figure 1.** The methodological outline of this study: defining environmental flow envelopes (EFESs) and estimating the frequency and severity of EFE violations. GHM stands for global hydrological model, GCM for general circulation model, and EFR for environmental flow requirement.
2.1 Data

We used the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) simulation round 2b outputs of global daily discharge (Frieler et al., 2017; available at https://esg.pik-potsdam.de). ISIMIP is a community–driven climate–impacts modelling initiative that collects and harmonises global model outputs (The Inter-Sectoral Impact Model Intercomparison Project, 2021). To decrease the uncertainties related to using single GHMs with single or few GCMs, we chose to use discharge estimates from four different GHMs (H08 (Hanasaki et al., 2018), LPJmL (Sitch et al., 2003), PCR-GLOBWB (Sutanudjaja et al., 2018), and WaterGAP2 (Müller Schmied et al., 2016)), each forced with modelled climate from four GCMs (GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR, MIROC5). Adopting this kind of an ensemble decreases uncertainty stemming from two separate sources: 1) using more than one GCM within one GHM decreases the GHM parameterisation uncertainty, and 2) using a number of GHMs in an analysis decreases the uncertainty of modelling the hydrological cycle within a single GHM (Döll et al., 2016; Schewe et al., 2014; Sood and Smakhtin, 2015). Simple metrics, such as the ensemble mean or median, often provide globally decent estimates when compared to observed discharge (see e.g. Arsenault et al., (2015) and Huang et al., (2017)), although individual members of the ensemble may outperform the ensemble result at the catchment scale (Zaherpour et al., 2018).

The discharge data (over both periods 1801–1860 and 1976–2005) were first temporally aggregated from daily to monthly discharge by taking the mean of daily values and then spatially aggregated at the sub-basin scale according to the HydroBASINS level 5. HydroBASINS is a global polygon layer series, which divides the world into consistently sized and hierarchically nested sub-basins at different scales (Lehner and Grill, 2013). We selected the level 5 since it is the highest level of detail that can be rasterized into a 0.5–degree resolution grid without an excessive loss of sub-basins that are smaller than a grid cell. In total, 352 out of 4,734 sub-basins were excluded due to their small size, while the average size of the remaining sub-basins was 30,700 km² and median size 19,600 km². Minor additional exclusions of five to six sub-basins per GHM were caused by non-overlapping discharge data grids. To aggregate the discharge at the sub-basin scale, we selected the maximum discharge cell value within the borders of each sub-basin, assuming that the sub-basin drains out from that cell. Hence, we consider this cell – and any violation in it – as representative of the whole sub-basin, though the situation may vary in different parts of the sub-basin.

2.2 Defining EFEs

We defined the EFEs based on the pre-industrial (1801–1860) time period, which in this study represents the natural flow regime and therefore relatively intact riverine ecosystems in the absence of significant anthropogenic flow alteration. Following Pastor et al. (2014), we selected five different EFR methods to accommodate for the differences in the methods’ definitions of ecosystem water needs. The selected EFR methods include Smakhtin's method (Smakhtin et al., 2004), Tennant's method (Tennant, 1976), Tessmann's method (Tessmann, 1980), the Q90-Q50-method (Pastor et al., 2014), and the variable
monthly flow (VMF) -method (Pastor et al., 2014). These methods are based on simple flow metrics, such as mean annual or monthly flow, determining EFRs according to hydrological seasons. All methods separate between low–flow and high–flow months while the Tessmann and VMF methods supplement this with a third class for intermediate–flow months. The equations to compute EFRs according to the selected EFR methods are presented in Table 1.

Table 1. Descriptions of methods used to calculate environmental flow requirements (EFRs) in this study (adapted from Pastor et al. (2014)). MMF refers to mean monthly flow of each month, MAF to mean annual flow (the mean monthly flow of all months within a year), Q50 and Q90 to flow exceeding 50% and 90% of the flows during the period of interest respectively, and coefHF to high-flow coefficient used in Smakhtin’s method.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Low–flow month definition</td>
<td>MMF ≤ MAF</td>
<td>MMF ≤ MAF</td>
<td>MMF ≤ MAF</td>
<td>MMF ≤ 0.4 x MAF</td>
<td>MMF ≤ 0.4 x MAF</td>
</tr>
<tr>
<td>EFR of low–flow month</td>
<td>(Q90)</td>
<td>0.2 x MAF</td>
<td>(Q90)</td>
<td>MMF</td>
<td>0.6 x MMF</td>
</tr>
<tr>
<td>High–flow month definition</td>
<td>MMF &gt; MAF</td>
<td>MMF &gt; MAF</td>
<td>MMF &gt; MAF</td>
<td>MMF &gt; 0.4 x MAF and 0.4 x MMF &gt; 0.4 x MAF</td>
<td>MMF &gt; 0.8 x MAF</td>
</tr>
<tr>
<td>EFR of high–flow month</td>
<td>coefHF x MAF(^{(a)})</td>
<td>0.4 x MAF</td>
<td>(Q50)</td>
<td>0.4 x MMF</td>
<td>0.3 x MMF</td>
</tr>
<tr>
<td>Intermediate–flow month definition</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>MMF &gt; 0.4 x MAF and 0.4 x MMF ≤ 0.4 x MAF</td>
<td>MMF &gt; 0.4 x MAF and MMF ≤ 0.8 x MAF</td>
</tr>
<tr>
<td>EFR of intermediate–flow month</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.4 x MAF</td>
<td>0.45 x MMF</td>
</tr>
</tbody>
</table>

\(^{(a)}\) If \(Q90 > 0.3 x MAF\), coefHF = 0; if \(0.2 x MAF < Q90 \leq 0.3 x MAF\), coefHF = 0.07; if \(0.1 x MAF < Q90 \leq 0.2 x MAF\), coefHF = 0.15; if \(Q90 \leq 0.1 x MAF\), coefHF = 0.2.

For each GHM, we applied the selected five EFR methods to four discharge data sets simulated using modelled climate from four GCMs, resulting in a monthly distribution of 20 independent EFR estimates per GHM. Before computing EFRs, we removed monthly outlier discharge further than three standard deviations away from mean monthly discharge. Similarly for the resulting EFR distribution, EFRs further than three standard deviations away from mean EFR were removed. This way, we avoided skewing the EFR distribution with extreme outliers in pre–industrial data. As the EFE lower bound, we selected the median of the EFR distribution. Selecting the midway EFR estimate excludes the tails of the EFR distribution that potentially consist of unrealistically low or high EFR estimates, caused by either highly deviant discharge provided by certain GCMs or distinctively different representation of ecosystem water needs in the EFR method.
As the EFE upper bound, for each GHM, we selected the 95th percentile of pre–industrial monthly discharge over all GCMs. While minor flooding can still be beneficial for riverine ecosystems, extreme floods often result in adverse effects (Talbot et al., 2018) and especially floodplain ecosystems require a distinctive dry period (Hayes et al., 2018; Junk et al., 1989; Schneider et al., 2017). This dry period can be compromised by increased dry season flows, for example due to hydropower operation. Other factors that potentially cause increases in flows across all flow seasons include natural variability of climate, anthropocentric climate change, inter–basin water transfers, and land use change, for example. Exceeding the 95th percentile of pre–industrial monthly discharge – including all GCMs – can thereby be considered as a significant signal of increased flows, although the underlying drivers vary. For illustration, a conceptual definition of the EFE is presented in Fig. A1, a comparison between monthly pre–industrial discharge and the EFE lower bound is presented in Fig. A2, and a comparison between EFEs and recent past discharge in sub–basins in variable flow regimes across the world is presented in Fig. A3.

2.3 Evaluating EFE violations

Finally, we compared the recent past discharge to the EFEs at the sub–basin scale. We considered the recent past discharge to cover years 1976–2005, the end date being limited by the ISIMIP 2b simulation period. For each GHM, we calculated a monthly violation ratio between the median discharge over four GCMs and the GHM–specific EFE (Table 2). The violation ratio yields a value between 0 and 100 if the discharge is within the EFE, a negative value if the discharge is below the EFE lower bound, and a value over 100 if the discharge is above the EFE upper bound. In the few cases where the EFE was unavailable due to no recorded flow in the pre–industrial time series, we considered the violation ratio to be zero, i.e. no violation.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Equation for violation ratio</th>
<th>Violation ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q &lt; EFE_{lower}$</td>
<td>$\frac{Q - EFE_{lower}}{EFE_{lower}} \times 100$</td>
<td>$&lt; 0$</td>
</tr>
<tr>
<td>$EFE_{lower} \leq Q \leq EFE_{upper}$</td>
<td>$\frac{Q - EFE_{lower}}{EFE_{upper} - EFE_{lower}} \times 100$</td>
<td>$0$ to $100$ (no EFE violation)</td>
</tr>
<tr>
<td>$Q &gt; EFE_{upper}$</td>
<td>$\left(\frac{Q - EFE_{upper}}{EFE_{upper}} + 1\right) \times 100$</td>
<td>$&gt; 100$</td>
</tr>
</tbody>
</table>

Throughout the analysis, we excluded time periods during which the EFE is violated for less than three consecutive months. This emphasises long–term flow alterations that are likely to threaten the riverine ecosystems beyond individual species (Biggs et al., 2005). Simultaneously, potential one–month outliers in recent past discharge are eliminated and do not therefore cause bias to violation metrics. In addition to results presented in the following section with a minimum three–month sequence of
violations, we repeated the analysis with other minimum lengths of the violation streak. The results of this sensitivity analysis are presented in the supplementary material (Fig. S1–S3). As often done in global studies (e.g. Gerten et al., 2020; Steffen et al., 2015), we excluded sub–basins with extremely low flow from our analysis; a sub–basin was excluded if at least three out of four GHMs estimated mean annual flow (the mean monthly flow of all months; MAF) less than 10 m$^3$ s$^{-1}$.

We analysed the EFE violations from two perspectives: the frequency and the severity of violations. Using equations in Table 2, we determined the violation ratio in each sub–basin for each month in 1976–2005. Considering the four GHMs, this resulted in a total of 1,440 violation ratios for each sub–basin (4 GHMs x 30 years x 12 months). We treated the violation ratios from different GHMs as independent observations of violation since the EFE was defined and evaluated strictly GHM–wise. The results for individual GHMs are presented in the supplementary material (Fig. S4–S11). We then defined two metrics: 1) violation frequency = fraction of violated months out of all 1,440 months in the time series, and 2) violation severity = mean violation ratio during those violated months. These metrics were computed separately for the lower and upper EFE bounds. A numerical example is provided in Fig. A1.

Elaborating the EFE violation patterns further, we analysed the violations with respect to flow seasons. For this, we classified each month of record into low (Q < 0.4MAF), intermediate (0.4MAF ≤ Q ≤ MAF), and high (Q > MAF) flow classes. For each GHM, we computed the flow season of each month from median discharge over all GCMs. MAF was computed from the respective year of each month, so that individual months could be classified to different seasons during different years, accommodating for drier and wetter years. Further, we conducted a seasonal trend analysis on the EFE violation frequency and severity. Here, we computed the frequency and severity of violations according to definitions above, but instead of all years (1976–2005), we applied five–year moving windows starting from the first window 1976–1980 and ending in the last window 2001–2005. Each of the moving window time series computed over four GHMs consisted of 240 violation ratios (4 GHMs x 5 years x 12 months). Then, for each sub–basin and separately for frequency and severity, we computed the Kendall rank correlation coefficient and fitted a linear regression model into the moving window time series (n = 26). We eliminated any statistically non–significant (p > 0.05) trends using the Kendall rank correlation test and the linear regression slope t–test. Finally, we combined the EFE violation frequency and severity throughout the recent past time series with the linear violation trend slopes and performed a fuzzy c–means clustering (Bezdek, 1981) to each flow season separately.

3 Results

Our findings show that EFE violations are widespread around the world, concentrating on lower bound violations in the arid and dry temperate climate zones (Fig. 2). In addition, notable EFE violation patterns emerge also in areas with high anthropogenic pressure, such as the Middle East, India, Eastern Asia, and Central America. The median discharge over GCMs violates the EFE in 49.8% of the total 3,860 sub–basins during more than 5.0% of the total 1,440 months of record across all

https://doi.org/10.5194/hess-2021-260
Preprint. Discussion started: 11 May 2021
© Author(s) 2021. CC BY 4.0 License.
GHMs (Fig. 2a). Discharge in 43.2% of sub–basins violates the EFE lower bound during more than 5.0% of all months (Fig. 2b) whereas the respective figure for the EFE upper bound is only 9.6% (Fig. 2c). Therefore, the EFE is rather violated by insufficient than excessive discharge, and regional patterns are more clearly visible in EFE lower bound violations whereas EFE upper bound violations are more dispersed into individual sub–basins.

Figure 2: Frequency of environmental flow envelope (EFE) violations between 1976 and 2005; for both upper and lower bounds (a), lower bound only (b), and upper bound only (c), all computed across four global hydrological models (GHMs). Sub–basins with mean annual flow (MAF) less than 10 m$^3$ s$^{-1}$ are excluded. Case examples on how the recent past discharge compares to the EFE in different flow regimes are shown in Fig. A3.

3.1 Characterisation of EFE violations

The low flow season is clearly the most impacted in terms of EFE lower bound violations, while the violations decrease gradually from low to intermediate and intermediate to high flow seasons (Fig. 3a–c). The distinction between flow seasons is stronger for the frequency than the severity of violations. Between 1976 and 2005, discharge violates the EFE in 83.4%, 59.0%,...
and 28.6% of sub–basins during low, intermediate, and high flow seasons for at least one three-month streak (frequency > 0). The medians of violation severities for low, intermediate, and high flow seasons are -37.1%, -19.0%, and -24.7%, respectively. These figures mean that the typical EFE lower bound violation is caused by discharge falling 19–37% below the EFE lower bound. Although the severity of violations appears to be less dependent on flow season than the frequency of violations, the low flow season remains the most impacted overall. This is also supported by the spatial coverage of sub–basins in the class of the most frequent (> 25%) and the most severe (Q < 0.5EFElower) violations, which reaches over all continents during low flow season (Fig. 3c) and decreases in prevalence during intermediate and high flow seasons (Fig. 3a–b).

The EFE upper bound violations are less dependent on flow season and exhibit less consistent spatial patterns of frequency and severity than EFE lower bound violations (Fig. 3d–f). The shares of sub–basins in which discharge violates the EFE upper bound for at least one three-month streak between 1976 and 2005 are 15.5%, 24.6%, and 18.9% for low, intermediate, and high flow seasons, respectively. The medians of violation severities during low, intermediate, and high flow seasons are 153%, 121%, and 123%. Although the summarised statistics would suggest typical EFE upper bound violations to be caused by discharge exceeding the EFE upper bound by 21–53%, many of the sub–basins experiencing EFE upper bound violations fall into the high–severity categories in which discharge exceeds the EFE upper bound at least twofold (Fig. 3d–f). These extremes occur in relatively small areas, covering a small number of sub–basins (Fig. 3e–f; e.g. Tigris–Euphrates, Northern China, Niger river), whereas larger–scale patterns covering more sub–basins show less frequent and less severe EFE upper bound violations (Fig. 3d–e; e.g. northeastern Europe, Central Asia). Hence, while EFE lower bound violations pertain similar characteristics over wide regions with strong dependence on flow season, EFE upper bound violations are more case–specific with the more serious violations covering only a small number of sub–basins in a region at a time.
Figure 3: Seasonal frequency and severity of environmental flow envelope (EFE) violations of the lower bound (a-c) and the upper bound (d-f). Q stands for monthly discharge and MAF for mean annual flow. For each sub-basin in each season, violation frequency corresponds to the fraction of violated months out of all months in that season, and violation frequency to the mean violation ratio during those violated months.
3.2 Past trends in EFE violations

Between 1976 and 2005, the frequency and severity of EFE violations of both lower and upper bounds have mainly co–developed in the same direction with more sub–basins experiencing amplifying rather than attenuating trends. For the EFE lower bound violations, a statistically significant violation trend is observed for 51.9%, 31.1%, and 15.0% of all sub–basins during low, intermediate, and high flow season, respectively (Fig. 4a–c). Of these detected trends, 41.0%, 54.3%, and 64.8% consist of a frequency and a severity trend in the same direction. Respectively, for the EFE upper bound and 10.3%, 16.6%, and 11.0% of all sub–basins showing statistically significant violation trends, 69.2%, 68.4%, and 72.1% of trends consist of changes in the same direction (Fig. 4d–f). Across both bounds and all three flow seasons, shares of trends consisting of an increase in one variable and a decrease in the other range from 0.5% to 5.4%, leaving the remaining 28–59% of trends to consist of a trend in one variable and no trend in the other. This highlights that the trends in EFE violation frequency and severity rather co–develop than conflict. Since increasing violation frequency combined with increasing violation severity is the single most common trend for both EFE lower bound and upper bound violations (28.7% and 53.0% of all detected trends across all flow seasons), the general trend of EFE violations has been towards intensifying direction during the past decades.

In most of the world, the trends of EFE lower and upper bound violations are independent, but signs of EFE violation trends shifting from the lower bound to the upper bound can be identified especially in the Northern Hemisphere and the Pan–Arctic areas. Trends in which the EFE lower bound violation frequency and severity are decreasing prevail in e.g. parts of Russia and Northern Canada (Fig. 4c), but the same regions show increasing trends in EFE upper bound violations (Fig. 4e). Therefore, increasing discharge alleviating EFE lower bound violations may turn out to be amplifying for EFE upper bound violations in some regions and downplay the positive indications of decreasing EFE lower bound violation trends. For most of the world, however, this shifting of violations is not visible, and trends – as well as the violations overall – concentrate on one boundary of the envelope only.
Figure 4: Trends of frequency and severity of environmental flow envelope (EFE) violations of the lower bound (a–c) and the upper bound (d–f). Q stands for monthly discharge and MAF for mean annual flow. The trends are computed using five–year moving windows from 1976 to 2005; only statistically significant trends are shown. The steepness of the trend slope is not considered here but the trends are classified only by increasing or decreasing direction.
3.3 Categorisation of sub–basins by EFE lower bound violations and trends

The arid mid-latitudes along with parts of tropical South America and subtropical Africa and Asia emerge as the most impacted regions in terms of EFE lower bound violations when the frequency, severity, and trends associated with both are combined together in a cluster analysis. In the relative paucity of sub–basins experiencing EFE upper bound violations, we performed the cluster analysis for the EFE lower bound violations only. In Fig. 5, the presented six clusters are ordered according to increasing prevalence of EFE lower bound violations: the blue and turquoise clusters A–B show areas with relatively little violations, the areas in purple and yellow clusters C–D begin to show increasing violation patterns, and the orange and red clusters E–F correspond to the areas with the absolute highest frequency and severity of EFE lower bound violations. The aforementioned highly impacted regions are mainly covered by clusters D–F, which contain the overall highest frequency and severity of violations through all flow seasons. Further, sub–basins in the yellow cluster D are currently experiencing moderate violations but showing the steepest increasing trends in both violation severity and frequency during low flow season (Fig. 4c). These sub–basins consisting of, for example, highly populous areas in Asia as well as regions in South America with rich riverine ecosystems, can be considered to be under the most significant threat of intensifying EFE violations if the past trends continue.
Figure 5: Clustering of sub–basins based on environmental flow envelope (EFE) lower bound violations during high flow season (a), intermediate flow season (b) and low flow season (c). Q stands for monthly discharge and MAF for mean annual flow. The four clustering variables constitute violation frequency, violation severity, and linear trend slopes associated with both variables. Sub–basins with mean annual flow (MAF) less than 10 m$^3$s$^{-1}$ are excluded.

4 Discussion

In this work, we show that recent past discharge in nearly half of the sub–basins of the world violates the EFE – a safe envelope of discharge variability – for extensive and recurrent periods between 1976 and 2005 (Fig. 2a). The emerging EFE lower bound violation patterns are strongly seasonal with low flow season being the most affected by both frequent and severe violations,
whereas the EFE upper bound violation patterns are more dispersed and harder to characterise (Fig. 3). Further, trends in both EFE lower and upper bound violations have rather been amplifying than attenuating during the past decades, showing increases in both violation frequency and severity in many areas (Fig. 4). Our results show that many sub–basins in the most populous and ecologically diverse areas, such as East Asia, South Asia, and parts of South America, are already experiencing considerable EFE lower bound violations, which can be expected to intensify based on the past trends (Fig. 5). To date, our study is the first to quantitatively address these three aspects of frequency, severity, and trends combined.

Parts of the most affected areas in terms of EFE violations, such as the arid mid–latitudes, India, Eastern Asia, and the west coast of North America, compare well with other global scale estimates of EF violations (Gerten et al., 2020; Jägermeyr et al., 2017; Steffen et al., 2015). These regions contain some of the most fragmented and regulated rivers globally, indicating drastic anthropogenic flow alteration (Grill et al., 2019, 2015). On the other hand, EF violations reported by the aforementioned global studies are not as widespread in large parts of Australia, South America, and Southern Africa as the EFE violations shown in our work (Fig. 2–3). Our results show that parts of Europe and parts of North America are among the areas where EFE violations are the least prevalent (Fig. 2–3), although rivers in these regions are highly fragmented, regulated and threatened in terms of biodiversity (Grill et al., 2019, 2015; Vörösmarty et al., 2010). Since these areas show relatively little EFE violations, it can be inferred that even though the quantitative discharge would be within the EFE, the anthropogenic flow alteration can still be major. Regarding the degree to which the EFs are undermined, Jägermeyr et al. (2017) report mainly discharge deficits under 10% whereas our results suggest substantially higher violation severities (Fig. 3a–c). However, the baselines between these studies differ since Jägermeyr et al. (2017) determine EFRs based on pristine discharge simulation between 1980 and 2009 and report annual averages whereas our EFEs are pre–industrial and we conduct the analysis per flow season.

4.1 Key drivers of EFE violations

Three key drivers for the prevalence and change in EFE violations can be identified from previous research: the two main direct anthropogenic impacts of increasing water use and flow regulation, especially by dam operation (Döll et al., 2009; Graham et al., 2020; Müller Schmied et al., 2016; Schneider et al., 2017), and the indirect impact of climate change on streamflow (Arnell and Gosling, 2013; Asadieh and Krakauer, 2017; Gudmundsson et al., 2021; Moragoda and Cohen, 2020; van Vliet et al., 2013; Wanders et al., 2015). The frequent and severe EFE violations in the densely populated mid-latitudes can largely be attributed to anthropogenic impact dominating the long–term streamflow alterations (Müller Schmied et al., 2016), which is also reflected in the projected increase of water stress (use to–availability ratio) that is driven primarily by increasing water use (Graham et al., 2020). The net anthropogenic flow alteration within a sub–basin can further be affected by water use and land use change beyond the sub–basin scale, either in upstream sub–basins or in remotely teleconnected regions (Munia et al., 2020; Wang-Erlandsson et al., 2018).
In the subtropical Southern Hemisphere, the EFE lower bound violations can be expected to follow the projected trends of increasing droughts as both are driven by abnormally low amounts of water in a system (Asadieh and Krakauer, 2017; Wanders et al., 2015). On the other hand, especially the decreasing trends of EFE lower bound violations (Fig. 4b–c) and the increasing trend of EFE upper bound violations in high-latitude Europe and Siberia (Fig. 4d–e) can at least partially be attributed to the past and projected increase in discharge due to climate change (Arnell and Gosling, 2013; Asadieh and Krakauer, 2017; Gudmundsson et al., 2021). However, dam operation alters flow regimes even in these sparsely populated regions and can potentially increase especially low season flows, resulting in EFE upper bound violations (Döll et al., 2009; Poff et al., 2007; see also Fig. A3b). While the three main drivers of flow alteration can either attenuate or amplify the net effect on EFE violations depending on the region, limiting anthropogenic flow alteration with special attention to low flow season would still be the key practical measure to decrease EFE violations in the most affected areas.

4.2 Relationship between EFE and riverine ecosystem well–being

The key assumption behind our results is that violating the EFE, either by insufficient or excessive streamflow, is a potential threat to riverine ecosystems. The simple correlation between a discharge proxy variable and ecosystem well–being is, however, a view that has been challenged in the past (Poff and Zimmerman, 2010; Richter, 2010), and the practical allocation of EFs based on insufficient methods has even been argued to potentially cause further degradation of riverine ecosystems (Arthington et al., 2006; Shenton et al., 2012). This is because of the multifaceted biodiversity response to altered flow regimes including variation across spatial scales and distinct parts of the riverine ecosystem, as well as the adaptation of species to flow regime changes over long timespans (Biggs et al., 2005; Poff et al., 1997; Rolls et al., 2018). Moreover, a recent study on global fish biodiversity has shown that several other factors, such as water quality and the presence of invasive species, may be more important in maintaining riverine ecosystems than quantitative flow (Su et al., 2021). Despite their flaws, hydrological EFR methods have remained as the primary option for global scale studies since direct assessments of riverine ecosystem well–being or more advanced EFR methods require in situ data, ancillary variables, and local expert knowledge (Tharme, 2003). In addition, the hydrological EFR methods applied in this study have been validated by Jägermeyr et al. (2017) and Pastor et al. (2014) with comparisons to locally defined EFRs that better portray the case–specific dependence of quantitative flow and riverine ecosystem well–being. Therefore, while the EFE may not be able to provide a globally generalised relationship between quantitative discharge and riverine ecosystem well–being, it is still a viable tool in illustrating the impacts of anthropogenic flow alteration at the sub–basin scale. However, local studies with more case–dependent knowledge and the incorporation of factors beyond quantitative flow will be required for practical implications.

By selecting the pre–industrial discharge as the baseline for defining EFEs, this study adheres to the paradigm of natural flows (Poff et al., 1997). This paradigm states that serious deviation from a natural baseline state is detrimental for the riverine ecosystem, and its globally equal absoluteness therefore suits the study well. Comparing the Anthropocene to previous, Holocene–like baseline conditions is also one of the leading rationales behind the Planetary Boundaries, which has emerged...
as a highly influential framework on quantifying anthropogenic impacts on the Earth system (Rockström et al., 2009; Steffen et al., 2015). However, regarding EFs, the pre–industrial natural flow baseline can be or has already been rendered unreachable as anthropogenic climate change continuously alters flow regimes even in pristine basins (Poff and Matthews, 2013). Moreover, in practical terms, returning to a natural flow state is an impossibility in many regions due to profound anthropogenic modification of rivers, such as large–scale damming and inter–basin transfer schemes. At the time of completion, this modification from natural into designed flows has been deemed to yield social and economic benefits beyond ecosystems and it has been accepted to partially compromise the natural flows (Acreman et al., 2014). Hence, while the EFEs based on the natural flow regime provide a valuable reference point, the policy targets based on them should more comprehensively consider the dynamics and contexts of local scale social–ecological systems, as well as the practical limits of flow restoration, in order to yield maximal co–benefits for all.

4.3 Methodological discussion and limitations

Our rationale based on which we define EFEs is strongly associated with ensemble thinking. Even though in some sub–basins, individual members of the ensemble (here, one EFR out of 20, see Fig. 1) would be the best fit locally, the ensemble median is deemed globally feasible as shown in studies regarding ensemble runoff and discharge (Arsenault et al., 2015; Huang et al., 2017; Zaherpour et al., 2018). Therefore, our results based on the ensemble could be assumed to be relatively robust compared to single–model or single–method studies (e.g. Gerten et al., 2020; Hoekstra and Mekonnen, 2011; Pastor et al., 2019; Steffen et al., 2015). Hogeboom et al. (2020) present an ensemble EFR similar to our EFE lower bound, although constructed from fewer ensemble members. Their comparison between annual EFR and runoff largely agrees with our comparison between EFE lower bound and monthly pre–industrial discharge (Fig. A2), although our method sets the EFE lower bound high in areas where Hogeboom et al. (2020) set it low, such as Australia and many other arid areas. As the spread between individual EFR methods applied over different GHMs is substantial (Hogeboom et al., 2020; Jägermeyr et al., 2017; Pastor et al., 2014), the uncertainty on how well the EFE bounds would correspond to EFRs determined from observed discharge remains, although adopting the ensemble decreases it.

Regarding the EFE upper bound, our selection of the 95th percentile of pre–industrial discharge is only a first step towards a more informed choice. On one hand, the link between the EFE upper bound and ecosystem responses remains weak in some areas, but on the other hand, it has shown to be a very important dry season factor in e.g. monsoon flood pulse systems in which floodplain ecosystems require distinct dry and wet periods (Hayes et al., 2018; Junk et al., 1989; Schneider et al., 2017). Further, the EFE upper bound is intentionally set to be very high by including all GCMs within a GHM – potentially containing very high discharge estimates. Hence, violations of the EFE upper bound are strong signals of excessive flows, although it cannot be inferred from this study whether these are detrimental to the riverine ecosystems outside the monsoon area.
The underlying hydrological data partially restricts the conclusions that can be made based on our results. First of all, determining EFEs based on monthly data aggregated from daily data is a substantial simplification and incurs a loss of temporal detail especially regarding extreme high and low flows. Moreover, we consider the sub–basin outlet cells as representative for the whole upstream area, although local EFE violations may vary within the sub–basin. While constructing the hydrological ensemble could be advanced by incorporating more sophisticated methods based on e.g. weighting by model performance in different regions (Arsenault et al., 2015; Beck et al., 2017; Zaherpour et al., 2019), the global modelling efforts, such as ISIMIP, remain the primary raw data source for global hydrological studies. Stemming from the model structural differences and varying parameterisation, GHMs are always uncertain to an extent (Telteu et al., 2021, in review). Especially, regarding the Pan–Arctic areas, GHMs have recently been shown to perform relatively poorly (Gädeke et al., 2020). In addition, while the data from ISIMIP 2b should be representative of historical land use and other human influences including dams and reservoirs (Frieler et al., 2017), the inclusion and parameterisation of different human impacts in GHMs plays a significant role in the results, particularly in terms of flooding and dam operation (Masaki et al., 2017; Veldkamp et al., 2018). The between–GHM uncertainty is illustrated in our sensitivity analysis which replicates the main results using individual GHMs (Fig. S4–S11).

4.4 Way forward

In the future, developing and applying the EFE methodology presented in this study should concentrate on validating the correspondence between the estimated EFEs and riverine ecosystem responses. Although derived from a robust ensemble, the EFE is still based on rule–of–thumb style EFR methods, which must be augmented with local knowledge for practical applications. Furthermore, quantification of the riverine ecosystem responses to prolonged and excessive flows through case studies would benefit the development of the EFE upper bound. While anthropogenic water use, river regulation and climate change are recognised as the leading drivers of flow alteration causing EFE violations, a more systematic and independent analysis on the couplings between these three drivers and EFE violations would provide more insights into our results. Despite the needs for further research and the limited direct applicability, the EFEs can already be used in global analysis for identifying sub–basins where anthropogenic flow alteration could potentially be considered to threaten riverine ecosystems. In its current state, the EFE methodology is lightly parameterised and applicable with open global data sets, availability and quality of which is constantly increasing. While methodological fine–tuning remains to be required for local contexts, the EFEs provide a quick and globally robust way of assessing the threats to riverine ecosystems posed by flow alteration, and allocating streamflow to the environment and anthropogenic uses.

5 Conclusion

Direct and indirect anthropogenic flow alterations are threatening the integrity of riverine ecosystems across the world. In this study, we have developed and applied a novel methodology of Environmental Flow Envelopes (EFEs) to quantify both the frequency and severity of these threats. Comparing recent past discharge with the EFEs based on pre–industrial conditions
shows that a significant part of global sub–basins is experiencing long-standing flow alteration. These EFE violations most commonly manifest themselves as insufficient flow during the low flow season, although in individual sub–basins, excessive flows can also be identified. With widespread increasing trends in both violation frequency and severity, the EFE violations can be expected to be amplified in response to projected future increases in human water use, building of new dams and climate change. On one hand, our results highlight the need to consider environmental flows in global research and policies on water resources management, while on the other hand, operationalising our results at the basin scale requires assimilation of cross-scale information and interdisciplinary knowledge.
Appendix A. EFE conceptualisation and assessment

**Figure A1:** Case example on the conceptual definition of the environmental flow envelope (EFE) and the assessment of EFE violations. The example sub-basin is a part of the Rio Paraguay basin: the observation point is located a little upstream from Asunción, Paraguay. For simplicity, we show discharge and assess EFE violations only for the lower bound and year 2000. In addition, we do not enforce the 3-month violation streak rule (see Sect. 2.3) in this example but count all individual violated months. If the 3-month rule was enforced, violations from H08 model only would be counted. For each global hydrological model (GHM; H08, LPJmL, PCR-GLOBWB, and WaterGAP2), the discharge is the median estimate over four general circulation models (GCMs). The EFE violation frequency and severity are computed according to definitions in Sect. 2.3.
Figure A2: Comparison between the environmental flow envelope (EFE) lower bound and pre–industrial discharge. Q stands for monthly discharge and MAF for mean annual flow. Here, for each global hydrological model (GHM) and month, we took the pre–industrial median discharge over all general circulation models (GCMs) and divided the EFE lower bound with it, yielding a total of 2,880 ratios for each sub–basin (4 GHMs x 60 years x 12 months). Outlier discharge was removed from monthly discharge before taking the median as outlined in Sect. 2.2. Then, for each season and across all GHMs, we took the median of the resulting EFElower / monthly discharge ratios (a–c) and computed the median absolute deviation around this median value (d–f). Some EFE lower bound estimates exceed the median low flow season discharge due to high variation in pre–industrial discharge affecting the distribution of environmental flow requirements (EFRs) from which the EFE lower bound is drawn (see Fig. 1). Moreover, the spread of ratios between EFE lower bound and low flow season monthly discharge is relatively high, further indicating high variability in low flow season discharge modelled by GHMs in the pre–industrial time series.
Figure A3. Case examples of environmental flow envelopes (EFEs) and mean monthly discharge in variable flow regimes. For the sake of illustration, we show both EFE lower and upper bounds as mean values over four global hydrological models (GHMs). Accordingly, the discharge presented here is the mean monthly discharge between 1976 and 2005, computed over four discharge data sets from four GHMs. Further, for each GHM, the discharge is the median over four general circulation models (GCMs) as outlined in Sect. 2.3. The anthropogenic modification of flow regimes is clearly visible in some of these sub-basins: for example, the spring peak flow in Fig. A3b has decreased whereas summer flows have substantially increased compared to pre-industrial EFEs.

Code and data availability

The code and data used in producing the results shown in this research article will be released in an open repository upon publication.

Author contribution

MK, EA, and VV conceptualised the study with input from MP, LA, TG, CM, LWE, and DG. DG, MF, NH, and HMS performed the ISIMIP simulations, which were coordinated by SNG and HMS. EA processed the raw data, wrote the implementation of the EFE methodology, and conceived the initial analysis with help from VV, MP, LA, and MK. VV revised and performed the final analysis and produced the results and visualisation shown in the study, discussing together with MK, MP, LA, TG, CM, and LWE. VV wrote the manuscript based on EA’s work with contributions from all authors.
Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

We would like to thank the ISIMIP team and all participating modelling teams for making the outputs available.

VV was funded by the Aalto University School of Engineering Doctoral Programme. EA and LA were funded by Maa- ja vesitekniikan tuki ry. MK received funding from Academy of Finland funded project WATVUL (grant no. 317320) and European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 819202). MP was funded by European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 819202).

References


