

ACCEPTED MANUSCRIPT • OPEN ACCESS

## Human impact parameterizations in global hydrological models improves estimates of monthly discharges and hydrological extremes: a multi-model validation study

To cite this article before publication: Ted Isis Elize Veldkamp *et al* 2018 *Environ. Res. Lett.* in press <https://doi.org/10.1088/1748-9326/aab96f>

### Manuscript version: Accepted Manuscript

Accepted Manuscript is “the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an ‘Accepted Manuscript’ watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors”

This Accepted Manuscript is ©2017 IOP Publishing Ltd.

As the Version of Record of this article is going to be / has been published on a gold open access basis under a CC BY 3.0 licence, this Accepted Manuscript is available for reuse under a CC BY 3.0 licence immediately.

Everyone is permitted to use all or part of the original content in this article, provided that they adhere to all the terms of the licence <https://creativecommons.org/licenses/by/3.0>

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required. All third party content is fully copyright protected and is not published on a gold open access basis under a CC BY licence, unless that is specifically stated in the figure caption in the Version of Record.

View the [article online](#) for updates and enhancements.

# Human impact parameterizations in global hydrological models improve estimates of monthly discharges and hydrological extremes: a multi-model validation study

T.I.E. Veldkamp<sup>1,9</sup>, F. Zhao<sup>2</sup>, P.J. Ward<sup>1</sup>, H. de Moel<sup>1</sup>, J.C.J.H., Aerts<sup>1,3</sup>, H. Müller Schmied<sup>4,5</sup>, F.T. Portmann<sup>4</sup>, Y. Masaki<sup>6</sup>, Y. Pokhrel<sup>7</sup>, X. Liu<sup>8</sup>, Y. Satoh<sup>9</sup>, D. Gerten<sup>2,10</sup>, S.N. Gosling<sup>11</sup>, J. Zaherpour<sup>11</sup>, Y. Wada<sup>9,12</sup>

<sup>1</sup> Institute for Environmental Studies (IVM), VU Amsterdam, the Netherlands

<sup>2</sup> Potsdam Institute for Climate Impact Research, Potsdam, Germany

<sup>3</sup> Department of Geography, University of California, Santa Barbara, Santa Barbara, USA

<sup>4</sup> Institute of Physical Geography, Goethe-University Frankfurt, Frankfurt, Germany

<sup>5</sup> Senckenberg Biodiversity and Climate Research Centre (BiK-F), Frankfurt, Germany

<sup>6</sup> National Institute for Environmental Studies, Tsukuba, Japan

<sup>7</sup> Department of Civil and Environmental Engineering, Michigan State University, Michigan, USA

<sup>8</sup> Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China

<sup>9</sup> International Institute for Applied Systems Analysis, Laxenburg, Austria

<sup>10</sup> Department of Geography, Humboldt-Universität zu Berlin, Berlin, Germany.

<sup>11</sup> School of Geography, University of Nottingham, Nottingham, United Kingdom

<sup>12</sup> Department of Physical Geography, Utrecht University, Netherlands

E-mail: ted.veldkamp@vu.nl

## Abstract

Human activities have a profound influence on river discharge, hydrological extremes, and water-related hazards. In this study, we compare the results of five state-of-the-art global hydrological models (GHMs) with observations to examine the role of human impact parameterizations (HIP) in the simulation of the mean, high, and low flows. The analysis is performed for 471 gauging stations across the globe and for the period 1971-2010. We find that the inclusion of HIP improves the performance of GHMs, both in managed and near-natural catchments. For near-natural catchments, the improvement in performance results from improvements in incoming discharges from upstream managed catchments. This finding is robust across GHMs, although the level of improvement and reasons for improvement vary greatly by GHM. The inclusion of HIP leads to a significant decrease in the bias of long-term mean monthly discharge in 36-73% of the studied catchments, and an improvement in modelled hydrological variability in 31-74% of the studied catchments. Including HIP in the GHMs also leads to an improvement in the simulation of hydrological extremes, compared to when HIP is excluded. Whilst the inclusion of HIP leads to decreases in simulated high-flows, it can lead to either increases or decreases in low-flows. This is due to the relative importance of the timing of return flows and reservoir operations and their associated uncertainties. Even with the inclusion of HIP, we find that model performance still not optimal. This highlights the need for further research linking the human management and hydrological domains, especially in those areas with a dominant human impact. The large variation in performance between GHMs, regions, and performance indicators, calls for a careful selection of GHMs, model components, and evaluation metrics in future model applications.

## 1. Introduction

Human activities have a profound influence on river discharge, hydrological extremes, and water-related hazards, like flooding, droughts, water scarcity, and water quality issues (Van Loon et al., 2016; Liu et al., 2017; Padowski et al., 2015; Veldkamp et al., 2017; Wada et al., 2011; Winsemius et al., 2016). As a result, research efforts have been made to parameterize human activities in global hydrological models (hereafter: GHMs, a full list of abbreviations is presented in **supplementary table 2**) (Bierkens, 2015; Pokhrel et al., 2016). These model parameterizations include: the

1  
2  
3 52 incorporation of dam and reservoir operations; the representation of human water use and return  
4 53 flows; and the representations of land use, land management, and land cover change (Pokhrel et al.,  
5 54 2016; Wada et al., 2016a, 2017).  
6  
7

8 55  
9 56 GHMs are widely used in scientific studies. For example, they have been used to assess the historical  
10 57 and future impacts of socioeconomic developments and/or hydro-climatic variability and change, on  
11 58 freshwater resources, droughts, and water scarcity (Biemans et al., 2011; Döll et al., 2009; Döll and  
12 59 Müller Schmied, 2012; Fujimori et al., 2017; Gosling et al., 2017; Haddeland et al., 2006, 2007, 2014;  
13 60 Hanasaki et al., 2013; Van Huijgevoort et al., 2013; Kummu et al., 2016; Müller Schmied et al., 2016;  
14 61 Munia et al., 2016; Rost et al., 2008; Veldkamp et al., 2015a,b, 2016, 2017, Wada et al., 2011,  
15 62 2013a,b, 2014a, Wanders et al., 2015). They are also increasingly used in practice. Global institutions  
16 63 increasingly rely on GHMs to conduct first-order assessments of water-related hazards because data,  
17 64 time, or resources are in short-supply for setting-up and executing multiple in-depth local studies. For  
18 65 example, GHMs have provided input into a multitude of high-level policy documents, such as: UN  
19 66 World Water Development Reports (e.g. Alcamo and Gallopin, 2009); Global Environmental  
20 67 Outlooks (UNEP, 2007); World Bank series on climate change and development (Hallegatte et al.,  
21 68 2016, 2017); and IPCC assessment reports (IPCC, 2007, 2013).  
22  
23  
24  
25  
26  
27  
28  
29

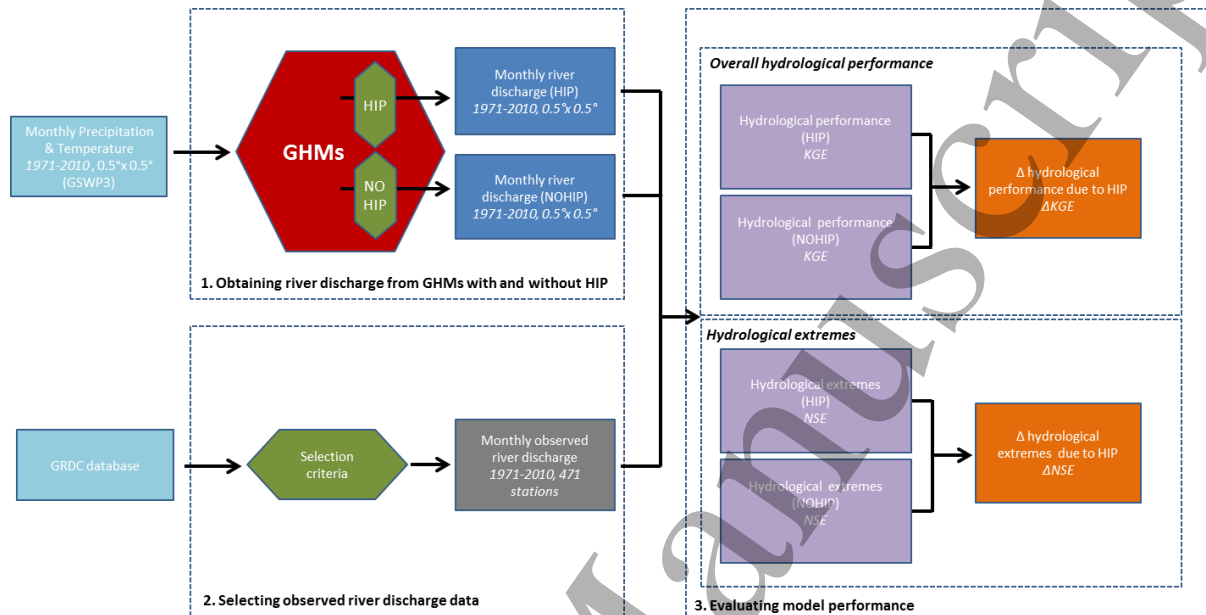
30 69  
31 70 As GHMs continue to improve in terms of detail, granularity, and speed, their importance for global,  
32 71 regional, and local applications is likely to increase further (Bierkens, 2015). Therefore, it is essential  
33 72 to have a thorough understanding of how well these GHMs represent real-world hydrological  
34 73 conditions. However, most GHM validation studies are limited to near-natural river catchments and  
35 74 make use of naturalized discharge data (Beck et al., 2016; Gudmundsson et al., 2011, 2012). Studies  
36 75 that have validated GHM simulations where human activities included have either focused on a single  
37 76 GHM and/or few selected river catchments (Biemans et al., 2011; Döll et al., 2003; 2009; De Graaf et  
38 77 al., 2014; Haddeland et al., 2006; Masaki et al., 2017; Müller Schmied et al., 2014; Pokhrel et al.  
39 78 2012; Wada et al., 2011, 2013a, 2014a).  
40  
41  
42  
43  
44  
45

46 79  
47 80 To date, a comprehensive validation of the ability of multiple GHMs to represent the influence of  
48 81 human activities on discharge and hydrological extremes in near-natural and managed catchments is  
49 82 missing. As a result, there is a limited understanding of whether (and where) the parameterizations of  
50 83 human activities in GHMs leads to an increase (or decrease) in model performance. To address this  
51 84 issue, the main objectives of this study are: (a) to evaluate the performance of five state-of-the-art  
52 85 GHMs that include the parameterizations of human activities in their modelling scheme; and (b) to  
53 86 compare the performance of these GHMs when run with and without human impact  
54 87 parameterizations.  
55  
56  
57  
58  
59

60 88

## 89 2. Data and Methods

90 The overall methodological framework used in this study is shown in **figure 1**. In brief, the method  
 91 involves three main steps: (1) obtaining river discharge from GHMs with human impact  
 92 parameterizations (HIP) and without human impact parameterizations (NOHIP); (2) selecting  
 93 observed river discharge data; and (3) evaluating model performance. Each of these steps is explained  
 94 in the following subsections.



95

96 **Figure 1: Flowchart of the methodological steps taken in this study.** Steps 1, 2, and 3 correspond to  
 97 paragraphs 2.1, 2.2 and 2.3.

98

### 99 2.1 Obtaining river discharge from GHMs with and without HIP

100 We used modelled monthly discharge (0.5° x 0.5° spatial resolution) for the period 1971–2010 from  
 101 five GHMs: H08 (Hanasaki et al., 2008a,b), LPJmL (Bondeau et al., 2007; Rost et al., 2008;  
 102 Schaphoff, et al., 2013), MATSIRO (Pokhrel, et al., 2012, 2015; Takata et al., 2003), PCR-GLOBWB  
 103 (van Beek et al., 2011; Wada et al., 2011, 2014b), and WaterGAP2 (Müller Schmied et al., 2016). All  
 104 simulations were carried out under the modelling framework of phase 2a of the Inter-Sectoral Impact  
 105 Model Intercomparison Project (ISIMIP2a: <https://www.isimip.org/protocol/#isimip2a>). For each  
 106 GHM, we used two simulations: (1) HIP: a model run including time-varying land use and land cover  
 107 change, historical dam construction and operation, irrigation, and upstream consumptive water  
 108 abstractions; and (2) NOHIP: a ‘naturalized’ model run without HIP.

109

110 An overview of the model characteristics of each of the GHMs, and the methods used to parameterize  
 111 hydrological processes and human impacts, can be found in **supplementary table 1**, and details on  
 112 each GHM can be found in the individual model references provided therein. In the following  
 113 subsections, we briefly outline the most important characteristics of the hydrological and human  
 114 impacts parameterizations.

115

### 2.1.1 Parameterizations of hydrological processes

Each GHM in this study is forced with daily (MATSIRO: three-hourly) inputs from the GSWP3 historical climate data-set (<http://hydro.iis.u-tokyo.ac.jp/GSWP3>). The GHMs applied in this study differ in hydrological representation and parameterizations (**supplementary table 1.A**). H08 and MATSIRO model the energy balance explicitly and use the bulk formula in the evaporation scheme (Hanasaki et al., 2008a,b; Pokhrel, et al., 2012, 2015; Takata et al., 2003). LPJmL, PCR-GLOBWB, and WaterGAP2 do not include the energy balance explicitly and use the Priestley-Taylor and Hammon formulas in their evapotranspiration schemes (van Beek et al., 2011; Bondeau et al., 2007; Müller Schmied et al., 2014,2016; Schaphoff et al., 2013; Verzano et al., 2012; Wada et al., 2011).

125

To generate runoff, all GHMs use a saturation excess formula, although the formula is integrated differently in the various GHMs. Snow accumulation and melt are integrated in the modelling framework via the energy balance (H08, MATSIRO) or by means of a degree-day calculation method (LPJmL, PCR-GLOBWB, WaterGAP2). All GHMs use a linear reservoir method in their routing scheme. Whilst H08, LPJmL, and MATSIRO route with a constant flow velocity (based on Manning's Strickler), PCR-GLOBWB and WaterGAP2 use variable flow velocities. The number of soil layers and their depths vary significantly between GHMs, from one layer with varying depth (e.g. WaterGAP2, H08) to 12 fully resolved layers.

134

### 2.1.2 Parameterizations of human impacts

All GHMs use a combination of socioeconomic and hydro-climatological parameters to estimate sectoral water demands (Hanasaki et al., 2008a,b; Müller Schmied et al., 2016; Pokhrel, et al., 2015; Rost et al., 2008; Schaphoff, et al., 2013; Takata et al., 2003; Van Beek et al., 2011; Wada et al., 2014b). Livestock water needs (**supplementary 1.B**) are estimated by combining historical gridded livestock density maps with their species-specific water demands. Domestic water demands (**supplementary table 1.C**) are derived by applying a time-series regression at the country-scale, accounting for drivers like population and per capita GDP, and in some cases (PCR-GLOBWB) total electricity production, energy consumption, and temperature. Industrial water demands (**supplementary table 1.D**) are based on historical country-scale estimates from the WWDR-II dataset (Shiklomanov, 1997; Vorosmarty et al., 2005; WRI, 1998) and the FAO-AQUASTAT database (<http://www.fao.org/nr/water/aquastat/dbase/index.stm>), for PCR-GLOBWB and H08 respectively. WaterGAP2 simulates global thermoelectric water use using spatially explicit information on the location of power plants. Manufacturing water demand is simulated in WaterGAP2 for each country using its yearly Gross Value Added (GVA), and factors representing technological change and water use intensity. The models estimate irrigation water use (**supplementary table 1.E**) by multiplying the area equipped for irrigation with its utilization intensity, the total crop-specific

152 water requirements – determined by the hydro-climatic conditions (temperature, precipitation,  
153 potential evapotranspiration, soil moisture, crop-growth curves, length and timing of the crop-growth  
154 season), and a parameter that accounts for the irrigation water use efficiency.

155

156 LPJmL, H08, and MATSIRO use surface water (first) to accommodate the sectoral water needs  
157 (**supplementary table 1.F**). WaterGAP2 uses the groundwater to fulfil water demands, and surface  
158 water is only used if enough is available. PCR-GLOBWB applies a share of readily available  
159 groundwater reserves, based on the ratio between simulated daily base-flow and long-term mean river  
160 discharge, to be used for consumptive water needs. The remainder of the water needs are fulfilled in  
161 PCR-GLOBWB by means of surface water. Whilst all GHMs deal consistently with return flows  
162 (**supplementary table 1.G**) for industry (surface water, same day), domestic (surface water, same  
163 day), and livestock (no return flow), returns from irrigation water use are incorporated differently.  
164 PCR-GLOBWB and H08 allow excess irrigation water return to the soil and groundwater layers by  
165 means of infiltration and additional recharge. LPJmL and MATSIRO return directly to the rivers, for  
166 which LPJmL uses a fixed ratio of 50%. Excess irrigation water in WaterGAP2 is returned to the  
167 surface waters using a cell-specific artificial drainage fraction, while the rest of the excess water is  
168 returned to groundwater.

169

170 All GHMs include either irrigation and/or non-irrigation purposes in their reservoirs schemes  
171 (**supplementary table 1.H**), and PCR-GLOBWB also includes flood control and navigation. The  
172 retrospective operation schemes of Hanasaki et al. (2006), Biemans et al. (2011), and Haddeland et al.  
173 (2006) form the basis of the reservoir operation schemes in most models. PCR-GLOBWB uses a  
174 prospective reservoir operation scheme that integrates efforts of Haddeland et al. (2006) and Adam et  
175 al. (2007). H08 is the only model that does not account for increased evapotranspiration over  
176 reservoirs.

177

## 178 ***2.2 Selecting observed river discharge data***

179 Observed monthly river discharge data were taken from the Global Runoff Data Centre (GRDC,  
180 56068 Koblenz, Germany). From the 9,051 gauging stations in the GRDC database, we selected  
181 stations that meet the following criteria: (1) a minimum of 5-year coverage (not necessarily  
182 consecutive) during the period 1971–2010 with a completeness of observations of  $\geq 95\%$ ; and (2) a  
183 minimum catchment area of 9,000 km<sup>2</sup>, to omit catchments whose hydrological processes cannot be  
184 adequately represented by models operating at 0.5° x 0.5° (Hunger and Döll, 2008). Finally, we  
185 discarded the stations for which the difference in catchment area in GRDC database and that  
186 estimated by using the DDM30 river routing network (Döll and Lehner, 2002) is  $>25\%$ .

187

1  
2  
3 188 We then made a distinction between near-natural and managed catchments. Following Beck et al.  
4 189 (2016), a catchment is classified as near-natural if the share of land-area subject to irrigation is <2%  
5 190 and the total reservoir capacity is <10% of its long-term mean annual discharge. If these conditions  
6 191 are not met the catchment was classified as managed. The classification was based on the HYDE  
7 192 3/MIRCA land cover dataset (Fader et al., 2010; Klein Goldewijk and Van Drecht, 2006; Portmann et  
8 193 al., 2010; Ramankutty et al., 2008) together with the Global Reservoir and Dam database (Lehner et  
9 194 al., 2011). Two stations shifted from near-natural to human impacted conditions between 1971 and  
10 195 2010, and were discarded from further analysis.

196

17 197 The aforementioned steps resulted in 471 stations with a total catchment area covering 19.8% of the  
18 198 global land (**figure 2**), of which 92 are located at the outlet of a catchment area. The mean length of  
19 199 observations is 32.8 years for all stations. Of all stations, 226 are located in managed catchments and  
20 200 245 in near-natural catchments. Of the stations located at the outlet of a catchment, 45 are managed  
21 201 (4.8% of the global land area), and 47 are near-natural (15.1% of the global land area).

202

27 203 **Figure 2** shows that the majority of selected stations (blue) are located in Northern and Latin-  
28 204 America, Europe, Southern Africa, and Australia. The number of stations in Northern and Central  
29 205 Africa and Asia is relatively small. We selected 12 stations in river basins located in different  
30 206 geographic regions (green circles in **figure 2**: Amazonas, Amur, Colorado, Congo, Guadiana,  
31 207 Mackenzie, Murray, Ob, Rhine, Tocantins, Volga, and the Zambezi) for which a detailed analysis is  
32 208 provided in the **Supplementary results** section (**Supplementary**).

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

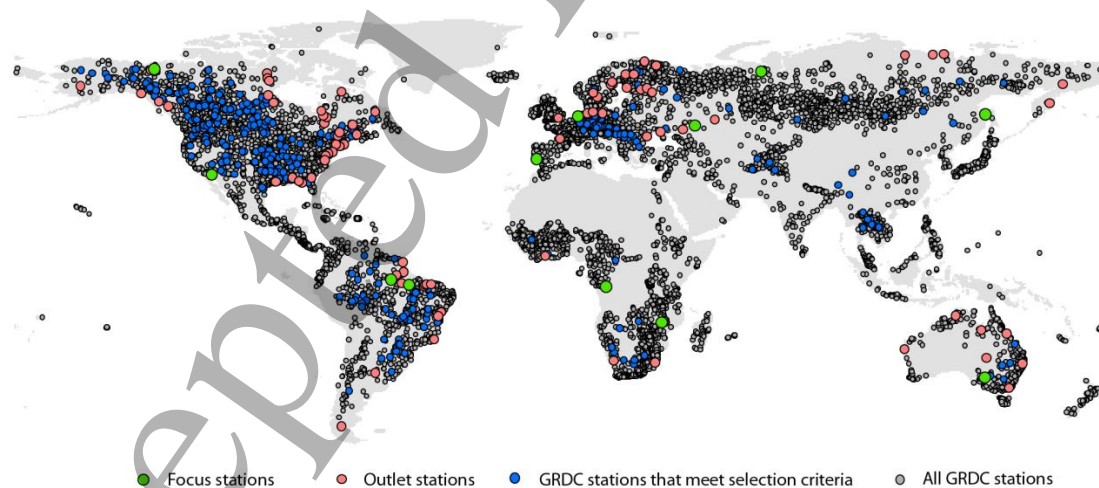
56

57

58

59

60



209

210

211

212

213

214

215

**Figure 2: Spatial distribution of GRDC stations used for this study.**

Each dot shows a GRDC station (n = 9,051) from the station catalogue. Blue dots indicate all GRDC stations (n = 471) that meet the selection criteria, whereas the red dots refer to the stations (n = 92) that are located at the outlet of a catchment. The green dots indicate those stations (n = 12) that were selected for detailed analyses.

### 2.3 Evaluating model performance

216 To evaluate the GHMs' simulation of monthly discharge and hydrological extremes under HIP and  
 217 NOHIP conditions, we compared modelled results with observed river discharge data using several  
 218 evaluation metrics described below. To ensure a consistent comparison between modelled and  
 219 observed data, we only used modelled data for the same years for which observations were available.  
 220 We also corrected modelled discharges for potential over-/underestimations caused by the difference  
 221 in catchment size between model and GRDC. To do this, we used a multiplier that represents the  
 222 difference in upstream area as reported by the GRDC and as estimated from the DDM30 network.

223  
 224 First, we applied the modified Kling-Gupta Efficiency index (KGE) with its sub-components: the  
 225 linear correlation coefficient (rKGE); the bias ratio ( $\beta$ KGE); and the variability ratio ( $\gamma$ KGE) (Gupta  
 226 et al., 2009; Kling et al., 2012). The KGE is a widely applied indicator for the validation of  
 227 hydrological performance in modelling studies at the global and regional scale and provides a good  
 228 representation of the "closeness" of simulated discharges to observations (Huang et al. 2017, Kuentz  
 229 et al., 2013; Nicolle et al., 2014; Revilla-Romero et al., 2015; Thiemiig et al., 2013, 2015; Thirel et al.,  
 230 2015; Wöhling et al., 2013). Moreover, use of its three sub-components enables the identification of  
 231 reasons for sub-optimal model performance (Gupta et al., 2009; Kling et al., 2012; Thiemiig et al.,  
 232 2013). This was achieved by estimating for each sub-parameter its distance to optimal performance,  
 233 and by subsequently comparing these distances across the different sub-parameters. Statistical  
 234 significance of the change in KGE outcomes due to the inclusion of HIP was tested by means of  
 235 regular bootstrapping ( $n = 1,000$ ,  $p \leq 0.05$  (two-tailed)), following the method of Livezey and Chen  
 236 (1982) and Wilks (2006).

237  
 238 Second, we applied the Nash-Sutcliffe Efficiency test (NSE, Nash and Sutcliffe, 1970) to evaluate the  
 239 representation of  $Q_1$  (high-flow) and  $Q_{99}$  (low-flow) conditions (e.g. Beck et al., 2017a; Blösch et al.,  
 240 2013; Hejazi and Moglen, 2008; Mohamoud, 2008), obtained under fixed threshold level settings (van  
 241 Loon, 2015). By means of a two-sample Kolmogorov-Smirnov (KS) test (Massey, 1951;  $p \leq 0.05$ ) we  
 242 tested how often HIP leads to significant changes in the fit of the full modelled exceedance  
 243 probability curve for hydrological extremes compared to the full observed exceedance probability  
 244 curve.

245  
 246 **Table 1: The performance metrics used in this study and their calculation procedure.**

247 Here,  $s_i$  and  $o_i$  are simulated and observed monthly discharge at station  $i$ ;  $\mu_s$  and  $\mu_o$  are simulated and observed  
 248 mean monthly discharge at station  $i$ ;  $\sigma_s$  and  $\sigma_o$  are the standard deviation of the simulated and observed  
 249 discharge at station  $i$ , respectively;  $Q_s$  and  $Q_o$  are the simulated and observed hydrological extremes.

Abbreviation	Name	Calculation procedure	Range and ideal value
KGE	Modified Kling-Gupta Efficiency Index	$KGE = 1 - \sqrt{(rKGE^* - 1)^2 + (\beta KGE^* - 1)^2 + (\gamma KGE^* - 1)^2}$	$-\infty - 1$ (ideal value: 1)
rKGE	KGE correlation coefficient	$rKGE = \frac{\sum_{i=1}^n (s_i - \mu_{s,i})(o_i - \mu_{o,i})}{\sqrt{\sum_{i=1}^n (s_i - \mu_{s,i})^2} \sqrt{\sum_{i=1}^n (o_i - \mu_{o,i})^2}}$	$-1 - 1$ (ideal value: 1)



	(Pearson)		
$\beta KGE$	KGE bias ratio	$\beta KGE = \mu_{s,i} / \mu_{o,i}$	0 - $\infty$ (ideal value: 1)
$\gamma KGE$	KGE variability ratio	$\gamma KGE = \frac{\sigma_{s,i} / \mu_{s,i}}{\sigma_{o,i} / \mu_{o,i}}$	0 - $\infty$ (ideal value: 1)
$NSE$	Nash-Sutcliffe Model Efficiency	$NSE = 1 - \frac{\sum(Q_s - Q_o)^2}{\sum(Q_o - \bar{Q}_o)^2}$	$-\infty - 1$ (ideal value: 1)
$Q_1$	High-flow indicator	Monthly discharge (m <sup>3</sup> /s) that is exceeded on average in 1 out of 100 months	
$Q_{99}$	Low-flow indicator	Monthly discharge (m <sup>3</sup> /s) that is exceeded on average in 99 out of 100 months	
$KS$	Two sample Kolmogorov-Smirnov test	[h, p] = kstest2(cdf(Q <sub>s</sub> ), cdf(Q <sub>o</sub> ), 'Alpha', 0.05)*	For p > 0.05 H <sub>0</sub> (the two cdfs come from the same distribution) is not rejected.

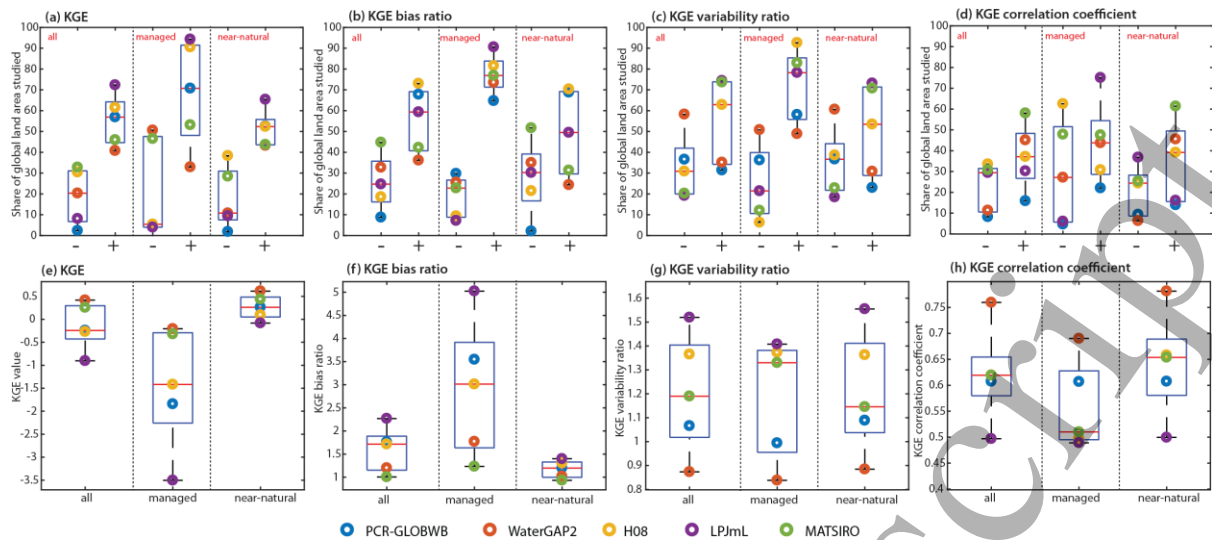
\* Calculation procedure for the two-sample Kolmogorov-Smirnov test presented in the table is the Matlab function for the KS-test.

### 3. Results

#### 3.1 Validation and influence of human impact parameterizations on overall model performance

Including the parameterizations of human impacts in the GHMs leads to a large improvement in overall model performance. Hydrological performance under the HIP simulations shows a significant improvement compared to the NOHIP simulations for between 40.8% and 72.3% of the land area studied, depending on the GHM (**figure 3a**). For most GHMs, the positive effects of including HIP in the simulations outweigh the negative effects. This is the case for both near-natural and managed catchments, although the positive effects are more pronounced for the managed catchments (**figure 3a-d**). Near-natural catchments are only indirectly impacted by HIP, for example by receiving improved or altered water simulations from upstream managed catchments. The KGE sub-components show significant improvement in performance in large shares of the land area studied, especially for the bias and variability ratio. The bias ratio improves significantly for 36.1-73.0% of the total land area for all catchments, compared to 64.8-90.6% and 24.3-70.4% in managed and near-natural catchments respectively (**figure 3b**). For the variability ratio, improvements were found for 31.4-74.4% of land area for all catchments (48.9-92.6% for managed / 23.0-73.2% for near-natural) (**figure 3c**). The lowest improvements are found for the correlation coefficient, with improvements for 15.9-58.1% of total land area for all catchments (22.1-75.1% for managed / 13.9-61.4% for near-natural) (**figure 3d**).

Results are shown for each station in **figure 4** for the overall model performance (KGE), and in **supplementary figure 1** for the KGE sub-parameters. The results show particularly strong improvements in overall performance in Latin America, Southern Africa, and Northwest U.S.. There are only a limited number of stations for which the inclusion of HIP leads to a significant decrease in overall hydrological performance for the majority of GHMs or where no to limited changes occur, for example in near-natural areas (e.g. the Amazonas).

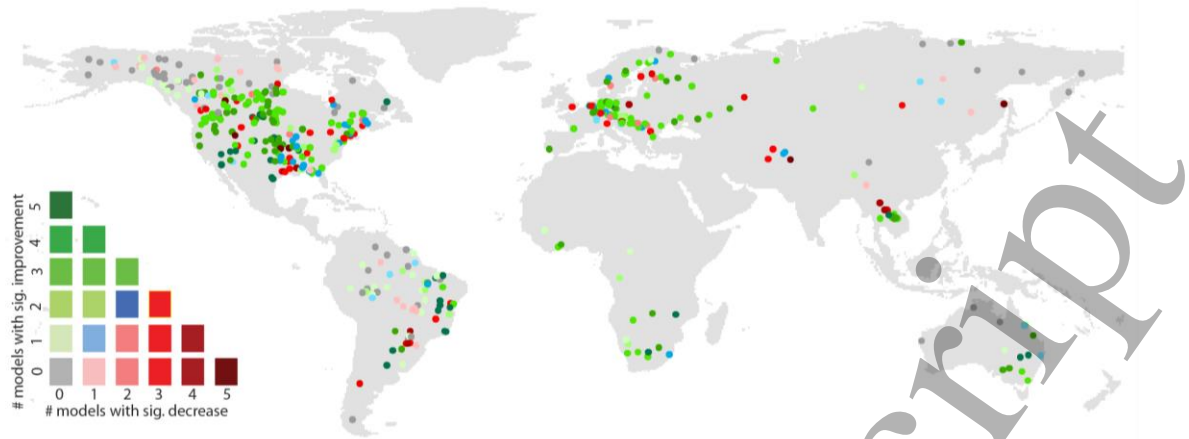


279

280 **Figure 3: Global weighted-mean (improvement ('+') or deterioration ('-') in the) representation of**  
 281 **hydrological performance due to HIP for all catchments, managed catchments, and near-natural**  
 282 **catchments.**

283 Figures 3a-d visualize for each GHM the share of land area with a significant change in overall hydrological  
 284 performance due to the inclusion of HIP. Figures 3e-h indicate the globally weighted-mean hydrological  
 285 performance after inclusion of HIP. On each box, the red mark indicates the median. The bottom and top edges  
 286 of the box indicate the 25th and 75th percentiles of the model ensemble, respectively

287  
 288 When considering overall hydrological performance for each GHM under HIP conditions (**figure 3e**),  
 289 WaterGAP2 and MATSIRO show the best performance globally. Even though the simulations with  
 290 HIP include human impact parameterizations by definition, all GHMs still show better performance in  
 291 near-natural catchments than in managed catchments (**figure 3e-h**). The KGE bias ratio values >1  
 292 indicate that all models systematically overestimate long-term mean monthly discharge (**figure 3f**), up  
 293 to 5-fold for LPJmL in managed catchments. For the variability ratio (**figure 3g**), WaterGAP2 is the  
 294 only GHM that tends to slightly underestimate variability (variability ratio <1) in monthly discharge,  
 295 in both the managed and near-natural catchments. All other GHMs show overestimations, up to 1.55-  
 296 fold for LPJmL for near-natural catchments. All GHMs show a reasonable correlation with observed  
 297 monthly discharge estimates (**figure 3h**), with values ranging between 0.49 to 0.69 in the managed  
 298 catchments and 0.50 to 0.79 in the near-natural catchments. The highest correlation coefficients  
 299 including HIP are found for WaterGAP2, with a global mean value across all catchments of 0.76 (0.69  
 300 for managed catchments / 0.78 for near-natural catchments).



301

302 **Figure 4: Number of GHMs with a significant improvement or deterioration in overall hydrological**  
 303 **performance (KGE) due to inclusion of HIP.**

304 Figures for the underlying KGE sub-parameters (bias ratio, variability ratio, correlation coefficient) are  
 305 presented in supplementary figure 1. Supplementary figure 2 shows the KGE performance values per GHM  
 306 under HIP conditions.

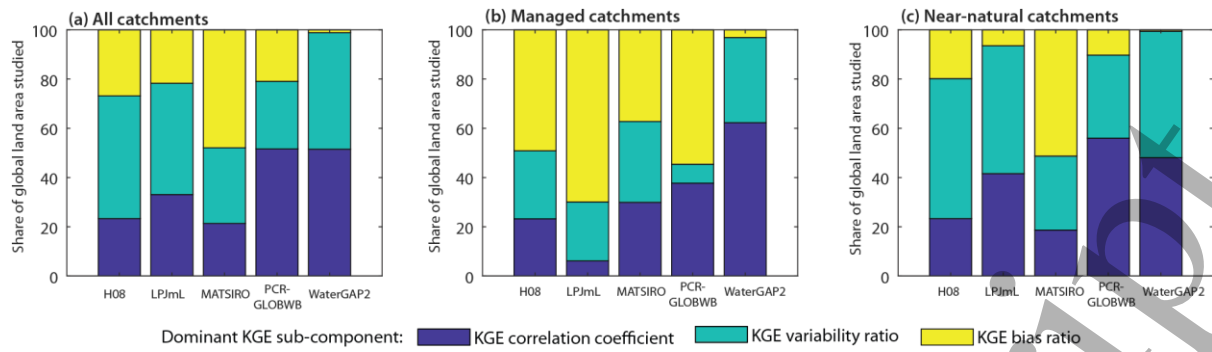
307

308 For each catchment (and therefore its associated land area), it is possible to distinguish which of the  
 309 KGE sub-parameters contributes most to sub-optimal performance. These results are summarised in  
 310 **figure 5**. The results show that under HIP conditions, the bias ratio contributes most to sub-optimal  
 311 performance in managed catchments for most GHMs, except WaterGAP2 (for which the correlation  
 312 coefficient contributes most). For near-natural catchments, sub-optimal performance is most often  
 313 caused by the variability ratio for H08, LPJmL and WaterGAP2, by the bias ratio for MATSIRO, and  
 314 by the correlation coefficient for PCR-GLOBWB.

315

316 Spatially explicit results vary per GHM and are shown in **supplementary figure 3**. The distribution  
 317 of dominant contributors to the sub-optimal overall hydrological performance is similar for H08,  
 318 LPJmL, and PCR-GLOBWB. For these GHMs, we find a dominant contribution of the bias ratio in  
 319 Southern Africa, Australia, and inland U.S. Dominant contributions of the variability ratio and the  
 320 correlation coefficient for these GHMs are found in Latin America, and at higher latitude and altitude  
 321 regions. For Europe, the dominant contributions for H08, LPJmL, and PCR-GLOBWB are the  
 322 variability ratio, the correlation coefficient, and the bias ratio respectively. The dominant contributors  
 323 that cause sub-optimal overall hydrological performance for MATSIRO and WaterGAP2 are more  
 324 equally distributed across the globe. While sub-components contribute to sub-optimal overall  
 325 hydrological model performance for MATSIRO, it is predominantly the correlation coefficient and  
 326 the variability ratio that determines the sub-optimal performance in WaterGAP2.

327



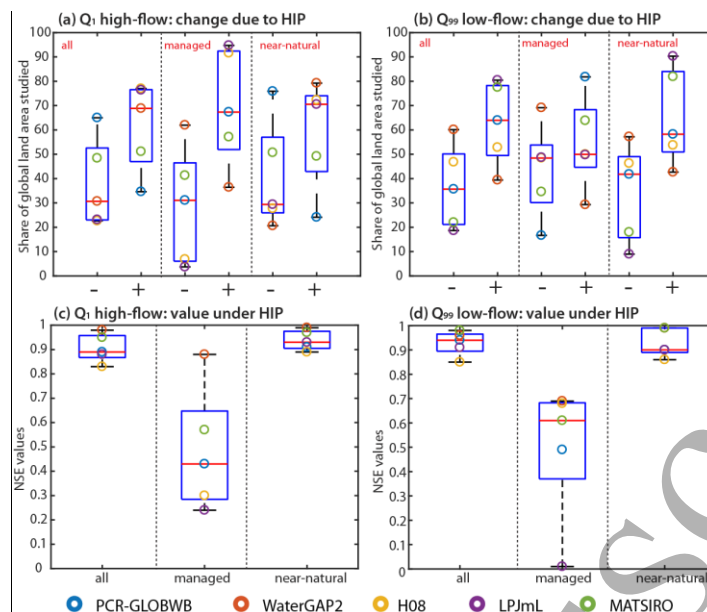
328 **Figure 5: Share of land area with dominant contribution of the different KGE sub-components (KGE**  
 329 **correlation coefficient, KGE variability ratio, KGE bias ratio) to sub-optimal overall hydrological**  
 330 **performance under HIP conditions.**

331 Supplementary figure 3 shows per model the spatial distribution of dominant KGE sub-components.  
 332  
 333

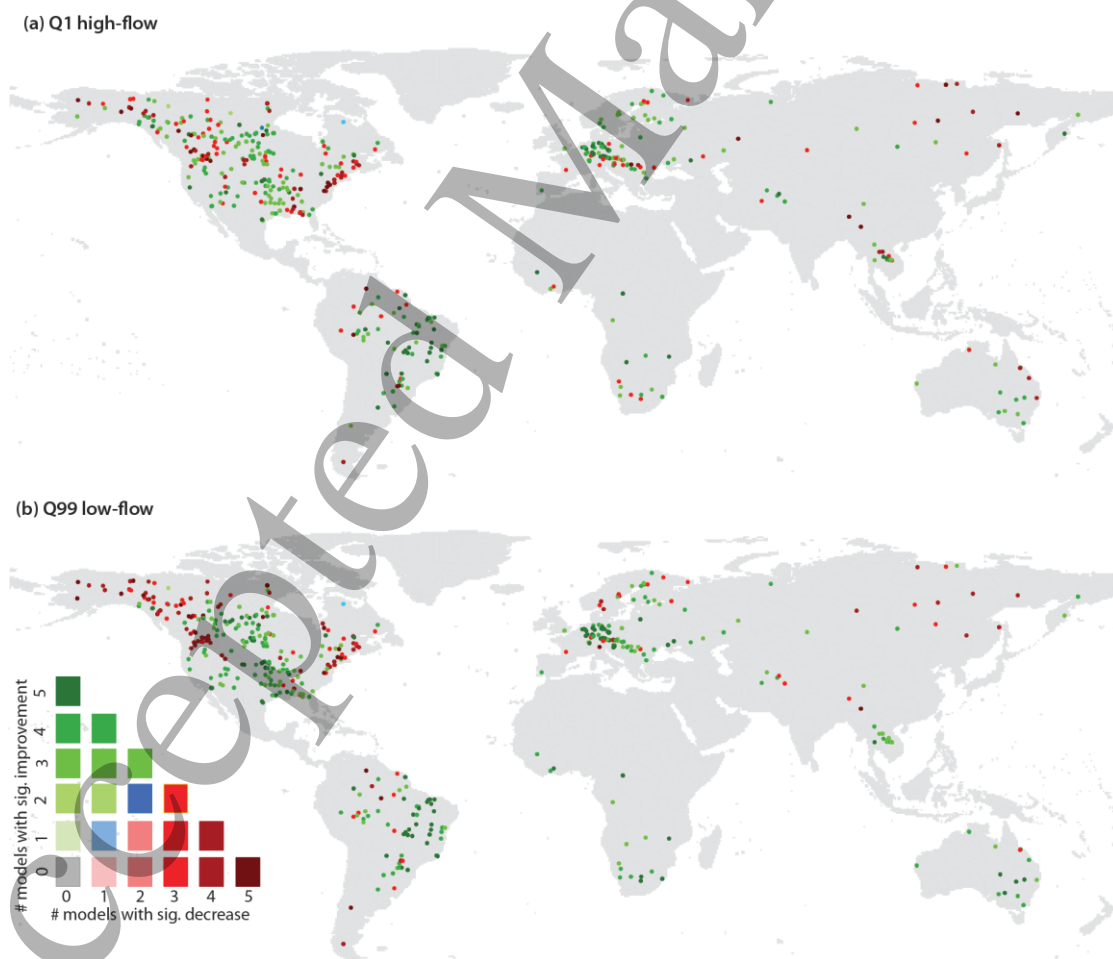
### 334 3.2 Validation and influence of human impact parameterizations on the simulation of hydrological 335 extremes

336 The inclusion of HIP in the simulations affects the ability of GHMs to estimate hydrological extremes  
 337 correctly in the majority of the land area studied (**figure 6**). The inclusion of HIP leads to better model  
 338 performance for all GHMs, across a substantial share of the land area studied (**figure 6a-b**). For high-  
 339 flows, HIP improves model performance significantly across 34.6-77.0% of the land area for all  
 340 catchments (36.4-94.7% for managed / 24.1-79.2% for near-natural). For low-flows, HIP improves  
 341 model performance significantly across 39.4-80.4% of the land area for all catchments (29.3-81.8%  
 342 for managed / 42.7-90.3% for near-natural). The KS-test results (**supplementary figure 4**) show that  
 343 HIP only leads to significant changes in the representation of the exceedance probability curve in a  
 344 limited number of cases for H08 and LPJmL (up to 14.1% of the land area studied), predominantly in  
 345 managed catchments.

346 Overall, hydrological extremes are represented reasonably well under HIP conditions, with globally  
 347 weighted-mean NSE values ranging between 0.80-0.98 for high-flows, and 0.84-0.98 for low-flows  
 348 (**figure 6c-d**). However, there is a significant difference in the ability of the GHMs to represent  
 349 hydrological extremes between managed and near-natural catchments.



350  
 351 **Figure 6: Global weighted-mean (improvement ('+') or deterioration ('-') in the) representation of**  
 352 **hydrological extremes (Q<sub>1</sub> high-flow and Q<sub>99</sub> low-flows) due to HIP, for all catchments, managed**  
 353 **catchments, and near-natural catchments respectively.**  
 354 On each box, the red mark indicates the median. The bottom and top edges of the box indicate the 25th and 75th  
 355 percentiles of the model ensemble, respectively  
 356  
 357



358  
 359 **Figure 7: Number of GHMs with a significant improvement or deterioration in representation of**  
 360 **hydrological extremes due to inclusion of HIP.**

1  
2  
3 361  
4 362 **Figure 7** indicates that for the majority of stations, the inclusion of HIP leads to an improvement in  
5  
6 363 the representation of hydrological extremes, for most GHMs. A deterioration in the representation of  
7  
8 364 hydrological extremes across the majority of GHMs as a result of the inclusion of HIP was only found  
9  
10 365 in selected areas, for example at higher latitudes and along the east-coast of the U.S.. When  
11  
12 366 comparing the results for the  $Q_1$  high-flows with the  $Q_{99}$  low-flows, no large differences in the spatial  
13  
14 367 distribution of the number of GHMs with a significant improvement or deterioration are found.

15  
16 369 The effects of HIP on the magnitude of extreme discharge differ for low-flows and high-flows  
17  
18 370 (**supplementary figure 5**). Whilst the magnitude of high-flows mostly decreases with the inclusion of  
19  
20 371 HIP, the effects on the magnitude of low-flows are both positive and negative. The convergence of  
21  
22 372 results towards higher observed discharges, in both high- and low-flow estimates (as identified for all  
23  
24 373 models in **supplementary figure 5**), indicates that HIP becomes less important for the correct  
25  
26 374 representation of hydrological extremes with increasing discharge volumes.

#### 27 375 28 376 **4. Discussion**

29 377 Our results show that including HIP in GHMs generally improves the overall hydrological  
30  
31 378 performance of the models, as well as their representation of hydrological extremes. However, we  
32  
33 379 also show that further improvements are needed. In this section, we discuss: (1) possible reasons for  
34  
35 380 the improved model performance due to HIP; (2) the main limitations of the current modelling  
36  
37 381 frameworks and their representation of HIP, and potential ways to improve them; and we reflect on  
38  
39 382 (3) general limitations in the current study design and provide suggestions for further research.

#### 40 383 41 384 **4.1 Improvements in model performance due to HIP and challenges ahead**

42 385 Whilst the inclusion of HIP predominantly leads to the largest improvements in simulated discharge  
43  
44 386 in the managed catchments, simulated discharge is also improved in a large share of the near-natural  
45  
46 387 catchments. Improvements in model performance associated with the inclusion of HIP can be  
47  
48 388 attributed to improvements in the different KGE sub-components, and in turn to different model  
49  
50 389 components parameterizing the hydrological and human processes. In addition, insights into those  
51  
52 390 factors bounding the optimal hydrological model performance under HIP conditions may help to  
53  
54 391 identify priorities for further model improvement.

##### 55 392 **4.1.1 Representation of long-term mean discharges (bias ratio)**

56 393 Our study shows that the representation of long-term mean discharges significantly improved with the  
57  
58 394 inclusion of HIP, especially in managed catchments. Inclusion of HIP generally results in lower  
59  
60 395 simulated discharges. As most GHMs systematically overestimate river discharges in the NOHIP  
396 simulation, this results in an improved performance. When HIP is included, we only find a

1  
2  
3 397 deterioration in the bias ratio in selected higher latitude/altitude regions, where discharges are  
4 398 underestimated; this finding is in line with outcomes of single-model studies performed by Döll et al.  
5 399 (2009), De Graaf et al. (2014), and Haddeland et al. (2006). Improvements in bias ratios due to the  
6 400 inclusion of HIP can be attributed to the inclusion of water abstractions and return flows  
7 401 (**supplementary table 1.B-G**), and the incorporation of irrigated areas and irrigation rules, which  
8 402 influence evapotranspiration rates and the generation of runoff (**supplementary table 1.E**).

9  
10  
11 403 However, despite improvement in the bias ratio with the inclusion of HIP, this KGE sub-indicator  
12 404 contributes most to sub-optimal performance in managed catchments for H08, LPJmL, MATSIRO,  
13 405 and PCR-GLOBWB under HIP conditions. As the GHMs continue to overestimate long-term mean  
14 406 discharges in most cases under HIP conditions, future model improvements should be targeted to  
15 407 correcting this bias in these locations. This may be achieved by critically revisiting the methods used  
16 408 to represent evapotranspiration rates (**supplementary table 1.A**), runoff generation processes  
17 409 (**supplementary table 1.A**) and the level of water abstractions in managed catchments  
18 410 (**supplementary table 1.B-E**). The relatively good performance of WaterGAP2, in which biases in  
19 411 long-term mean annual discharge are adjusted using a parameter that determines the portion of  
20 412 effective precipitation that becomes surface runoff (Müller Schmied et al., 2014), highlights the  
21 413 potential importance of including a calibration routine (**supplementary table 1.I**). Calibration is also  
22 414 performed for H08, but this calibration aims to minimize runoff bias by modifying two parameters of  
23 415 subsurface flow for four climatic groups (Hanasaki et al., 2008a,b); it is therefore less effective in  
24 416 minimizing the bias ratio under HIP conditions.

25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36 417

#### 37 418 *4.1.2 Representation of hydrological variability (variability ratio)*

38 419 The inclusion of HIP leads to mixed results regarding the representation of hydrological variability.  
39 420 Whilst HIP improved the representation of variability in some catchments and for some GHMs, it  
40 421 deteriorated the representation of variability for others. For example, it led to improvements in west-  
41 422 coast U.S., Southern Africa, and Australia, but a deterioration for most GHMs in Europe and inland  
42 423 U.S.. Similar results were found by Biemans et al. (2011), De Graaf et al. (2014), and Masaki et al.  
43 424 (2017) for a selection of catchments. Changes in the variability ratio due to the inclusion of HIP are  
44 425 predominantly driven by the timing of water abstractions and return flows, as well as by reservoir  
45 426 operation rules (**supplementary table 1.F-H**). These human activities influence the relative size of  
46 427 high- and low-flows compared to their long-term mean discharge values.

47  
48  
49  
50  
51  
52  
53 428 The variability ratio is the KGE sub-parameter that contributes most to the sub-optimal performance  
54 429 in near-natural catchments with the inclusion of HIP, for H08, LPJmL, and WaterGAP2. These  
55 430 GHMs significantly overestimate hydrological variability in near-natural catchments (except  
56 431 WaterGAP2, which underestimates variability in managed and near-natural catchments), and model  
57 432 improvement should therefore focus on better representing the speed of hydrological response, e.g.

1  
2  
3 433 through an improved representation of the soil moisture storage capacity or the ratio between surface  
4 434 and sub-surface runoff (**supplementary table 1.A**). In those cases where the variability ratio is also  
5  
6 435 the KGE sub-parameter that contributes most to sub-optimal performance in managed catchments,  
7  
8 436 model improvement should target the timing of water abstractions, return flows, and reservoir  
9  
10 437 management (**supplementary table 1.F-H**).

11 438

#### 12 439 *4.1.3 Representation of the goodness-of-fit (correlation coefficient)*

13  
14 440 The inclusion of HIP only led to improved correlation coefficients in limited cases, and often resulted  
15  
16 441 in a deterioration, even in managed catchments. Correlation coefficients between observed and  
17  
18 442 modelled discharges, which are predominantly determined by the hydro-meteorological forcing data  
19  
20 443 (Döll et al., 2016; Beck et al., 2016), were found to be generally high under both HIP and NOHIP  
21  
22 444 conditions. Perturbations of the hydrological cycle due to human activities leading to changes in the  
23  
24 445 timing of discharges and in the shape of the hydrograph, like return flows and reservoir operations,  
25  
26 446 explain the observed decrease in the correlation coefficient in a substantial share of catchments and  
27  
28 447 models globally (**supplementary table 1.F-H**).

29  
30 448 Under HIP conditions, the correlation coefficient is the KGE sub-parameter that contributes most to  
31  
32 449 sub-optimal performance only in PCR-GLOBWB for near-natural catchments and WaterGAP2 for  
33  
34 450 managed catchments. It should be acknowledged, though, that correlation coefficients for PCR-  
35  
36 451 GLOBWB and WaterGAP2 are relatively high, especially compared to the other GHMs. The  
37  
38 452 relatively low correlation coefficients in near-natural catchments found at higher latitudes in all  
39  
40 453 models may be addressed by critically reviewing the snow accumulation and melt processes in the  
41  
42 454 GHMs (**supplementary table 1.A**). Higher correlation coefficients in the managed catchments may  
43  
44 455 be established by improving the timing and quantification of return flow estimates and the  
45  
46 456 representativeness of reservoir operations (**supplementary table 1.F-H**).

#### 47 457 *4.1.4 Representation of hydrological extremes*

48  
49 458 The inclusion of HIP also led to significant changes in the ability of most GHMs to represent  
50  
51 459 hydrological extremes (both high- and low-flows), although the strength of this change is very much  
52  
53 460 dependent on the location and GHM in question. Whilst the magnitude of high-flow estimates mainly  
54  
55 461 decreased due to the inclusion of HIP, low-flow estimates showed mixed results. This is because the  
56  
57 462 impacts of human activities tend to be greater for lower discharges, as the relative 'size' of human  
58  
59 463 perturbations (such as water abstractions, return flows, or delayed releases of water via reservoir  
60  
464 operations) is higher as a percentage of overall discharge when flows are low. Both De Graaf et al.  
465 (2014) and Wada et al. (2013a) found similar results when investigating hydro-climatic extremes.  
466 However, even with inclusion of HIP, the representation of hydrological extremes is sub-optimal.  
467 Future model improvements should aim to better characterize these extremes and to improve the  
468 representation of human activities during extreme hydrological conditions.



1  
2  
3 4694 470 **4.2 Limitations and further research**

5 471 As the GHMs have very different parameterizations of hydrological and human processes, the current  
6 472 study does not allow a systematic assessment of specific cause-effect relations between HIP and the  
7 473 observed improvements in performance (Döll et al., 2016; Haddeland et al., 2014; Hagemann et al.,  
8 474 2013; Schewe et al., 2014; Beck et al., 2016). To do this, a substantial Monte-Carlo analysis would be  
9 475 required, whereby individual parameters and combinations of parameters are systematically modified  
10 476 for all GHMs (Döll et al., 2016). Undertaking such an analysis in parallel for the different GHMs  
11 477 incorporated is computationally expensive and requires a strict modelling-protocol. It may provide,  
12 478 however, additional information on how to adapt and improve the individual models and would be a  
13 479 valuable addition to the results presented in this study.

14 480

15 481 When interpreting the results of this study one must take into account that we only evaluated the  
16 482 GHMs with respect to monthly discharge. Whilst monthly discharge may be sufficient for the  
17 483 assessment and management of low-flows, droughts, and freshwater resource availability, flood risk  
18 484 assessment and management require information on daily peak discharge. Further research should  
19 485 therefore attempt to validate GHMs using daily peak discharge and assess how daily peak discharge is  
20 486 affected by the inclusion of HIP.

21 487

22 488 The spatial resolution of the GHMs applied in this study is  $0.5^\circ \times 0.5^\circ$  (~50 km x 50 km at the  
23 489 equator), dictated by the resolution of the GSWP3 input dataset. At a  $0.5^\circ$  spatial resolution  
24 490 hydrological processes are often represented by GHMs in a simplified or generalized form not fit for  
25 491 local applications (Bierkens, 2015). To account for this, we applied a minimum catchment size of  
26 492 9,000 km<sup>2</sup>, thereby omitting catchments too small to be adequately represented by GHMs (Hunger  
27 493 and Döll, 2008). Newer versions of several of the GHMs now operate at higher resolutions; for  
28 494 example WaterGAP and PCR-GLOBWB have recently published 5-min/6-min versions respectively  
29 495 (Verzano et al., 2012; Wada et al., 2016b). Future research could investigate whether the inclusion of  
30 496 these high-resolution model-runs improves the representation of discharges and hydrological extremes  
31 497 in the selected catchments and whether these high-resolution runs also allow for the inclusion of  
32 498 smaller catchments.

33 499

34 500 In this study, a relatively simple distinction was made between managed and near-natural catchments  
35 501 using two parameters: irrigated agriculture and reservoirs. These parameters were chosen as they have  
36 502 been reported to be the most significant human parameters on river hydrology (Beck et al., 2016,  
37 503 2017a). However, to make a more detailed distinction between catchments that are impacted by  
38 504 human activities and those that are not, future studies could consider incorporating additional criteria,  
39 505 such as the share of sectoral water abstractions and return flows, and the share of built-up land area.

1  
2  
3 506 Additional catchment descriptors (Eisner, 2016), like climate conditions and physiographic properties  
4 507 of the drainage area, could also be applied to further assess the important controls on modelled  
5  
6 508 discharges.  
7

8 509

9 510 When evaluating the impact of HIP on hydrological extremes we only incorporated results for the  $Q_1$   
10 511 high-flow and  $Q_{99}$  low-flow. In this study we did not consider other ranges of the extreme value  
11 512 distribution explicitly. Although the inclusion of HIP shows influences these hydrological extremes  
12 513 substantially, we found very few instances in which this led to a significant change in the full  
13 514 exceedance probability curve . Future research should therefore also incorporate other ranges of the  
14 515 probability exceedance curve in order to do a full assessment of the influence of HIP on high- and  
15 516 low-flow extremes.  
16  
17  
18  
19  
20

21 517 Next to the parameterizations and representation of hydrological processes and human impacts, other  
22 518 sources contribute to the uncertainty in the modelling of discharges and hydrological extremes. ,  
23 519 These include the quality of, and uncertainties in, input data and observation datasets, and the  
24 520 calibration/validation strategy (Döll et al., 2016; Sood and Smakhtin, 2015). The quality of the  
25 521 selected forcing data, for example, may limit the representation of monthly discharges and  
26 522 hydrological extremes significantly (Döll et al., 2016; Beck et al., 2016), but has not been evaluated  
27 523 explicitly in this study. However, climate forcing uncertainty is probably a dominant driver for model  
28 524 outputs (Müller Schmied et al 2014, 2016). A benchmarking of the GSWP3 dataset against historical  
29 525 observations of precipitation and temperature, or against other forcing datasets (e.g. similar to Beck et  
30 526 al., 2017b; Sun et al., 2017), may therefore be of added value.  
31  
32  
33  
34  
35  
36  
37  
38

37 527

39 528 Differences in the quality and trustworthiness of the historical discharge observations (e.g. due to  
40 529 sampling, measurement, and interpretation errors), may potentially result in artificial biases in the  
41 530 validation results (Renard et al., 2010). The spatial representativeness of our results is limited by the  
42 531 availability of consistent publicly available in situ observations of sufficient quality. Future research  
43 532 should therefore consider extending the GRDC data-points with regional repositories of observed  
44 533 discharges, such as recently attempted by Beck et al. (2016), Do et al. (2017), and Gudmundsson et al.  
45 534 (2017). However, increasing the spatial representation comes at the cost of consistency, and special  
46 535 attention should be paid to the harmonization of these different databases. The use of remotely sensed  
47 536 data could also provide a valuable way of carrying out calibration and validation in ungauged regions  
48 537 (Döll et al., 2014a,b; Scanlon, et al. 2018). Remotely sensed data can also be of added value in: the  
49 538 assessment of the water consumed by agricultural irrigation (Peña-Arancibia et al., 2016), operational  
50 539 drought monitoring and early warning (Ahmadalipour et al., 2017); and the estimation of terrestrial  
51 540 water budgets (Zhang et al., 2017). Moreover, a clear potential exists for the assimilation of remotely  
52 541 sensed data into models (Eicker et al., 2014).  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 542

4 543 Calibration and validation are essential for compensating for factors such as the impossibility to  
5 544 measure all required model parameters at the applied scale, the lack of process understanding, the  
6 545 simplistic process representation in GHMs, and errors in forcing data (Beck et al., 2016; Bierkens,  
7 546 2015; Döll et al., 2016; Liu et al., 2017). Hence, calibration/validation is key for realistic model  
8 547 performance. It should be acknowledged, though, that the representation of hydrological and/or  
9 548 human processes is artificially altered by means of calibration/validation processes and that a limited  
10 549 calibration may introduce uncertainties to the model output (Sood and Smakhtin, 2015). Before using  
11 550 any calibrated/validated model-data one should therefore critically reflect on whether the  
12 551 calibration/validation procedure executed, together with their optimization objectives, are fit for the  
13 552 specific application in-mind.

14 553

## 15 554 **5. Summary and conclusions**

16 555 This study shows that the inclusion of human activities in GHMs can significantly improve the  
17 556 simulation of monthly discharges and hydrological extremes, for the majority of catchments studied.  
18 557 The finding is robust across both managed and near-natural catchments. The global and spatially  
19 558 distributed results presented in this study indicate that the inclusion of human impact  
20 559 parameterizations is associated with improvements in the bias ratio and the variability ratio. Whilst  
21 560 the biases in long-term mean monthly discharge decrease significantly in 36.1-73.0% of the studied  
22 561 catchments due to the inclusion of HIP, the modelling of hydrological variability improves  
23 562 significantly in 31.4-74.4% of the catchments. Estimates of hydrological extremes are also  
24 563 significantly influenced by the inclusion of HIP, although the influence is highly dependent on the  
25 564 location and GHM in question. While HIP generally leads to a decrease (and thus improvement) in the  
26 565 absolute magnitude of simulated high-flows, its impact on low-flows is mixed.

27 566

28 567 Even when human activities are included in GHMs, their performance is still limited; this is  
29 568 particularly the case in managed catchments. Moreover, the systematic misrepresentation of  
30 569 hydrological extremes across all GHMs calls for a careful interpretation of risk assessments based on  
31 570 their results, and further study into the overarching research theme of water resources, hydrological  
32 571 extremes, human interventions, and feedback linkages. The large variation in performance between  
33 572 GHMs, regions, and performance indicators, highlights the importance of a careful selection of  
34 573 models, model components, and evaluation metrics in future model applications. For example, for a  
35 574 study of droughts it is essential to correctly represent hydrological variability, whilst to study water  
36 575 scarcity it is crucial to minimize biases.

37 576

38 577 Sub-KGE results, which were presented in this study for each GHM, allow for the attribution of  
39 578 different hydrological and human impact model-components limiting optimal hydrological

1  
2  
3 579 performance. In most GHMs model performance is limited due to the overestimation of long-term  
4 580 mean discharges. The correlation coefficient is the limiting factor for optimal model performance for  
5 581 WaterGAP2, despite the high correlation coefficients that were found for this model relative to the  
6 582 other GHMs studied. A better understanding of these factors, as provided by this study, may assist in  
7 583 the identification of priorities for further model improvement.  
8  
9  
10  
11 584

### 12 585 **Acknowledgements**

13 586 The Global Runoff Data Centre (GRDC, 56068 Koblenz, Germany) are thanked for providing the  
14 587 observed discharge data. This work has been conducted under the framework of phase 2 of the Inter-  
15 588 Sectoral Impact Model Intercomparison Project (ISIMIP2a: [www.isimip.org](http://www.isimip.org)) and the authors want to  
16 589 thank the coordination team responsible for bringing together the different global hydrological  
17 590 modelling groups and for coordinating the research agenda, which resulted in this manuscript. The  
18 591 research leading to this article is partly funded by the EU 7th Framework Programme through the  
19 592 project Earth2Observe (grant agreement no. 603608). JZ was funded by the Islamic Development  
20 593 Bank. PJW received additional funding from the Netherlands Organisation for Scientific Research  
21 594 (NWO) in the form of a Vidi grant (016.161.324). J.C.J.H.A. received funding from the Netherlands  
22 595 Organisation for Scientific Research (NWO) VICI (grant no. 453-14-006). YM was supported by the  
23 596 Environment Research and Technology Development Fund (S-10) of the Ministry of the  
24 597 Environment.  
25  
26 598

### 27 599 **References**

- 28 600 Adam, J. C., Haddeland, I., Su, F., & Lettenmaier, D. P. (2007). Simulation of reservoir influences on  
29 601 annual and seasonal streamflow changes for the Lena, Yenisei, and Ob' rivers. *Journal of*  
30 602 *Geophysical Research Atmospheres*, 112, 1–22.  
31  
32 603  
33 604 Ahmadalipour, A., Moradkhani, H., Hongxiang, Y., & Zarekarizi, M. (2017). Remote Sensing of  
34 605 Drought: Vegetation, Soil Moisture, and Data Assimilation. In V. Lakshmi (Ed.), *Remote Sensing of*  
35 606 *Hydrological Extremes* (pp. 121–149). Switzerland: Springer Remote Sensing/Photogrammetry.  
36  
37 607  
38 608 Alcamo, J. & Gallopin, G. (2009). The United Nations World Water Assessment Programme:  
39 609 Building a second generation of world water scenarios, Paris, France.  
40  
41 610  
42 611 Beck, H. E., Van Dijk, A. I. J., De Roo, A., Miralles, D. G., McVicar, T. M., Schellekens, J. &  
43 612 Bruijnzeel, L. A. (2016). Global-scale regionalization of hydrologic model parameters. *Water Resour.*  
44 613 *Res.* 52, 3599–3622.  
45  
46 614  
47 615 Beck, H. E., Van Dijk, A. I. J., De Roo, A., Dutra, E., Fink, G., Orth, R., Schellekens, J. (2017a).  
48 616 Global evaluation of runoff from 10 state-of-the-art hydrological models. *Hydrol. Earth Syst. Sci.*, 21,  
49 617 2881–2903.  
50  
51 618  
52 619 Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., Van Dijk, A. I. J., Weedon, G. P., Brocca, L.,  
53 620 Pappenberger, F., Huffman, G. J., Wood, E. F. (2017b). Global-scale evaluation of 22 precipitation  
54 621 datasets using gauge observations and hydrological modelling. *Hydrol. Earth Syst. Sci.*, 21, 6201-  
55 622 6217.  
56  
57 623  
58 624 van Beek, L.P.H., Wada, Y. & Bierkens, M.F.P. (2011). Global monthly water stress: I. Water  
59 625 balance and water availability. *Water Resour. Res.* 47, W07517.  
60

- 1  
2  
3 626  
4 627 Biemans, H., Haddeland, I., Kabat, P., Ludwig, F., Hutjes, R.W.A., Heinke, J., Von Bloh, W. &  
5 628 Gerten, D. (2011). Impact of reservoirs on river discharge and irrigation water supply during the 20th  
6 629 century. *Wat. Resour. Res.* 47, W03509.  
7  
8 630  
9 631 Bierkens, M.F.P. (2015). Global hydrology 2015: State, trends, and directions. *Water Resour. Res.*, 51,  
10 632 4923-4947.  
11 633  
12  
13 634 Blösch, G., Sivapalan, M., Wagener, T., Viglione, A., Savenije, H. (2013). Runoff predictions in  
14 635 ungauged basins: synthesis across processes, places and scales. Cambridge University Press, UK.  
15 636  
16 637 Bondeau, A. *et al.* (2007). Modelling the role of agriculture for the 20th century global terrestrial  
17 638 carbon balance. *Global Change Biology*, 13, 679–706.  
18 639  
19 640 Do, H. X., Gudmundsson, L., Leonard, M., Westra, S. & Seneviratne, S.I. (2017). The Global  
20 641 Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of daily streamflow  
21 642 archive and metadata. *Earth Syst. Sci. Data Discuss.*  
22 643  
23 644 Döll, P. & Lehner, B. (2002). Validation of a new global 30-min drainage direction map. *J. Hydrol.*,  
24 645 258, 214-231.  
25 646  
26 647 Döll, P. & Siebert, S. (2002). Global modelling of irrigation water requirements. *Water Resour. Res.*,  
27 648 38, W1037.  
28 649  
29 650 Döll, P., Kaspar, F. & Lehner, B. (2003). A global hydrological model for deriving water availability  
30 651 indicators: model tuning and validation. *J. Hydrol.*, 270, 105-134.  
31 652  
32 653 Döll, P., Fiedler, K. & Zhang, J. (2009). Global-scale analysis of river flow alterations due to water  
33 654 withdrawals and reservoirs. *Hydrol. Earth Syst. Sci.* 13, 2413-2432.  
34 655  
35 656 Döll, P., & Müller Schmied, H. (2012). How is the impact of climate change on river flow regimes  
36 657 related to the impact on mean annual runoff? A global-scale analysis. *Environ. Res. Lett.*, 7, 014037.  
37 658  
38 659 Döll, P., Fritsche, M., Eicker, A. & Müller Schmied, H. (2014a). Seasonal Water Storage Variations  
39 660 as Impacted by Water Abstractions: Comparing the Output of a Global Hydrological Model with  
40 661 GRACE and GPS Observations. *Surv. Geophys.*, 35(6), 1311-1331.  
41 662  
42 663 Döll, P., Müller Schmied, H., Schuh, C., Portmann, F.T. & Eicker, A. (2014b). Global-scale  
43 664 assessment of groundwater depletion and related groundwater abstractions: Combining hydrological  
44 665 modeling with information from well observations and GRACE satellites. *Water Resour. Res.*, 50,  
45 666 5698-5720.  
46 667  
47 668 Döll, P., Douville, H., Güntner, A., Müller Schmied, H. & Wada, Y. (2016). Modelling Freshwater  
48 669 Resources at the Global Scale: Challenges and Prospects. *Surv. Geophys.*, 37(2), 195-221.  
49 670  
50 671 Eicker, A., Schumacher, M., Kusche, J., Döll, P. & Müller Schmied, H. (2014). Calibration/data  
51 672 assimilation approach for integrating GRACE data into the WaterGAP Global Hydrology Model  
52 673 (WGHM) using an Ensemble Kalman Filter: First results. *Surveys in Geophysics* 35 (6), 1285-1309  
53 674

- 1  
2  
3 675 Eisner, S. (2016). Comprehensive evaluation of the WaterGAP3 Model across climatic,  
4 676 physiographic, and anthropogenic gradients. PhD thesis ([http://nbn-resolving.de/urn:nbn:de:hebis:34-](http://nbn-resolving.de/urn:nbn:de:hebis:34-2016031450014)  
5 677 2016031450014)  
6 678
- 7 679 Fader, M., Rost, S., Muller, C., Bondeau, A., Gerten, D. (2010). Virtual water content of temperate  
8 680 cereals and maize: Present and potential future patterns. *J. Hydrol.*, 384 (3-4), 218-231.  
9 681
- 10 682 Flörke, M., Kynast, E., Bärlund, I., Eisner, S., Wimmer, F. & Alcamo, J. (2013). Domestic and  
11 683 industrial water uses of the past 60 years as a mirror of socio-economic development: A global  
12 684 simulation study. *Global Environ. Change*, 23, 144–156.  
13 685
- 14 686 Fujimori, S., Hanasaki, N., Masui, T. (2017). Projections of industrial water withdrawal under shared  
15 687 socioeconomic pathways and climate mitigation scenarios. *Sust. Sci.*, 12, 275-292.  
16 688
- 17 689 Gosling, S.N. *et al.* (2017). A comparison of changes in river runoff from multiple global and  
18 690 catchment-scale hydrological models under global warming scenarios of 1 C, 2 C, and 3 C. *Clim.*  
19 691 *Change*, 141, 577-595.  
20 692  
21 693
- 22 694 Gupta, H.V., Kling, H., Yilmaz, K.K. & Martinez, G.F. (2009). Decomposition of the mean squared  
23 695 error and NSE performance criteria: implications for improving hydrological modelling. *J. Hydrol.*,  
24 696 377 (1–2), 80–91.  
25 697
- 26 698 de Graaf, I.E.M., Van Beek, L.P.H., Wada, Y. & Bierkens, M.F.P. (2014). Dynamic attribution of  
27 699 global water demand to surface water and groundwater resources: Effects of abstractions and return  
28 700 flows on river discharges. *Adv. Water Resour.*, 64, 21-33.  
29 701
- 30 702 Gudmundsson, L., *et al.* (2011). Comparing large-scale hydrological model simulations to observed  
31 703 runoff percentiles in Europe. *J. Hydrometeor.*, 13, 604 – 620.  
32 704
- 33 705 Gudmundsson, L., Wagener, T., Tallaksen, L.M. & Engeland, K. (2012). Evaluation of nine large-  
34 706 scale hydrological models with respect to the seasonal runoff climatology in Europe. *Wat. Resour.*  
35 707 *Res.*, 48, W11504.  
36 708
- 37 709 Gudmundsson, L., Do, H.X., Leonard, M., Westra, S. & Seneviratne, S. I. (2017). The Global  
38 710 Streamflow Indices and Metadata Archive (GSIM) – Part 2: Quality control, time-series indices and  
39 711 homogeneity assessment. *Earth Syst. Sci. Data Discuss.*  
40 712
- 41 713 Haddeland, I., Skaugen, T. & Lettenmaier, D. P. (2006). Anthropogenic impacts on continental  
42 714 surface water fluxes. *Geophys. Res. Lett.*, 33, L08406.  
43 715
- 44 716 Haddeland, I., Skaugen, T., & Lettenmaier, D. P. (2007). Hydrologic effects of land and water  
45 717 management in North America and Asia: 1700–1992. *Hydrol. Earth Syst. Sci.* 11, 1035–1045.  
46 718
- 47 719 Haddeland, I. *et al.* (2014). Global water resources affected by human interventions and climate  
48 720 change. *Proc. Natl. Acad. Sci. U. S. A.*, 111, 3251–3256.  
49 721
- 50 722 Hagemann, S. *et al.* (2013). Climate change impact on available water resources obtained using  
51 723 multiple global climate and hydrology models. *Earth Syst. Dynam.* 4, 129-144.  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 724  
725 Hallegatte, S., Bangalore, M., Bonzanigo, L., Fay, M., Kane, T., Narloch, U., Rozenberg, J., Treguer,  
726 D., Vogt-Schilb, A. (2016). Shock Waves: Managing the Impacts of Climate Change on Poverty.  
727 *Climate Change and Development Washington, DC: World Bank.*  
728 <https://openknowledge.worldbank.org/handle/10986/22787>  
729
- 730 Hallegatte, S., Vogt-Schilb, A., Bangalore, M. & Rozenberg, J. (2017). Unbreakable: Building the  
731 Resilience of the Poor in the Face of Natural Disasters. *Climate Change and*  
732 *Development; Washington, DC: World Bank.*  
733 <https://openknowledge.worldbank.org/handle/10986/25335>  
734  
735
- 736 Hanasaki, N., Kanae, S., Oki, T. (2006). A reservoir operation scheme for global river routing  
737 models. *J. Hydrol.*, 327, 1-2, 22-41.  
738
- 739 Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y. & Tanaka, K.  
740 (2008a). An integrated model for the assessment of global water resources – Part 1: Model description  
741 and input meteorological forcing. *Hydrol. Earth Syst. Sci.*, 12, 1007–1025.  
742
- 743 Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y. & Tanaka, K.  
744 (2008b). An integrated model for the assessment of global water resources – Part 2: Applications and  
745 assessments. *Hydrol. Earth Syst. Sci.*, 12, 1027–1037.  
746
- 747 Hanasaki, N. *et al.* (2013). A global water scarcity assessment under Shared Socio-economic  
748 Pathways – Part 2: Water availability and scarcity. *Hydrol. Earth Syst. Sci.*, 17, 2393-2413.  
749
- 750 Hejazi, M. I., Moglen, G.E. (2008). The effect of climate and land use change on flow duration in the  
751 Maryland Piedmont region. *Hydrol. Proc.*, 22, 4710-4722.  
752
- 753 Huang, S., Kumar, R., Flörke, M., Yang, T., Hundecha, Y., Kraft, P., Gao, C., Gelfan, A., Liersch, S.,  
754 Lobanova, A., Strauch, M., Van Ogtrop, F., Reinhardt, J., Haberlandt, U., Krysanova, V. (2017).  
755 Evaluation of an ensemble of regional hydrological models in 12 large-scale river basins worldwide.  
756 *Clim. Change*, 141, 381-397.  
757
- 758 Hunger, M. & Döll, P. (2008) Value of river discharge data for global-scale hydrological modelling.  
759 *Hydrol. Earth Syst. Sci.*, 12, 841-861.  
760
- 761 Van Huijgevoort, M. H. J., Hazenberg, P., Van Lanen, H. A. J., Teuling, A. J., Clark, D. B., Folwell,  
762 S., Gosling, S. N., Hanasaki, N., Heinke, J., Koirala, S., Stacke, T., Voss, F., Sheffield, J. &  
763 Uijlenhoet, R. (2013). Global multimodel analysis of drought in runoff for the second half of the  
764 twentieth century. *J. Hydrometeorol.*, 14, 1535-1552  
765
- 766 IPCC (2007). Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III  
767 to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC, Geneva,  
768 Switzerland.  
769
- 770 IPCC (2013). Climate change 2013: the physical science basis. IPCC, Geneva, Switzerland.  
771
- 772 Klein Goldewijk, K. & Van Drecht, G.(2006). HYDE 3: Current and historical population and land  
773 cover. MNP (2006) (Edited by A.F. Bouwman, T. Kram and K. Klein Goldewijk), Integrated  
774 modelling of global environmental change. An overview of IMAGE 2.4. Netherlands Environmental  
775 Assessment Agency (MNP), Bilthoven, The Netherlands.

- 1  
2  
3 776  
4 777 Kling, H., Fuchs, M. & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an  
5 778 ensemble of climate change scenarios. *J. Hydrol.*, 424–425, 264–277.  
6 779  
7 780 Kuentz, A., Mathevet, T., Gailhard, J., Perret, C., Andreassian, V. (2013). Over 100 years of climatic  
8 781 and hydrologic variability of a Mediterranean and mountainous watershed: the Durance river. In: *Cold*  
9 782 *and Mountain Region Hydrological Systems Under Climate Change: Towards Improved Projections*  
10 783 Proceedings of H02, IAHS-IAPSO-IASPEI Assembly, Gothenburg, Sweden. IAHS publication, 360.  
11 784  
12 785 Kummu, M., Guillaume, J. H. A., De Moel, H., Eisner, S., Flörke, M., Porkka, M., Siebert, S.,  
13 786 Veldkamp, T. I. E., Ward, P.J. (2016) The world's road to water scarcity: shortage and stress in the  
14 787 20th century and pathways towards sustainability. *Sci. Rep.*, 6, 38495.  
15 788  
16 789 Lehner, B., *et al.* (2011). High resolution mapping of the world's reservoirs and dams for sustainable  
17 790 river flow management. *Front. Ecol. Environ.*, 9, 494-502.  
18 791  
19 792 van Loon, A.F. (2015). Hydrological drought explained. *WIREs Water*, 2, 359-392.  
20 793  
21 794 van Loon, A.F., *et al.* (2016). Drought in the Anthropocene. *Nat. Geoscience*, 9, 89-91.  
22 795  
23 796 Livezey, R. E., & Chen, W. Y. (1982). Statistical field significance and its determination by monte  
24 797 carlo techniques. *Mon. Weather Rev.*, 111, 46–59.  
25 798  
26 799 Liu, X., Tang, Q., Cui, H., Mu, M., Gerten, D., Gosling, S., Masaki, Y., Wada, Y. & Satoh, Y. (2017).  
27 800 Multimodel uncertainty changes in simulated river flows induced by human impact parameterizations.  
28 801 *Environ. Res. Lett.*, 12, 025009.  
29 802  
30 803 Masaki, Y., *et al.* (2017). Intercomparison of regulated river discharge among multiple global  
31 804 hydrological models under multiple forcings – Part II: Multiple models in two case-study river basins,  
32 805 Missouri-Mississippi and Green-Colorado. *Environ. Res. Lett.*, 12, 055002.  
33 806  
34 807 Massey, F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American*  
35 808 *Statistical Association*, 46, 68–78.  
36 809  
37 810 Mohamoud, Y.M. (2008). Prediction of daily flow duration curves and streamflow for ungauged  
38 811 catchments using regional flow duration curves. *Hydrol. Sci. J.*, 53, 706-724.  
39 812  
40 813 Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F.T., Flörke, M. & Döll, P.  
41 814 (2014). Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological  
42 815 model structure, human water use and calibration. *Hydrol. Earth. Syst. Sci.*, 18, 3511-3538.  
43 816  
44 817 Müller Schmied, H. *et al.* (2016). Variations of global and continental water balance components as  
45 818 impacted by climate forcing uncertainty and human water use. *Hydrol. Earth Syst. Sci.* 20, 2877-2898.  
46 819  
47 820 Munia, H., Guillaume, J.H.A., Mirumachi, N., Porkka M., Wada, Y. & Kummu, M. (2016). Water  
48 821 stress in global transboundary river basins: significance of upstream water use on downstream stress.  
49 822 *Environ. Res. Lett.*, 11, 014002.  
50 823  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



- 1  
2  
3 824 Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A  
4 825 discussion of principles. *J. Hydrol.*,10(3), 282-290.  
5 826
- 6  
7 827 Nicolle, P., Pushpalatha, R., Perrin, C., François, D., Thiéry, D., Mathevet, T., Le Lay, M., Besson, F.,  
8 828 Soubeyroux, J.-M., Viel, C., Regimbeau, F., Andréassian, V., Maugis, P., Augeard, B., Morice, E.  
9 829 (2014). Benchmarking hydrological models for low-flow simulation and forecasting on French  
10 830 catchments. *Hydrol. Earth Syst. Sci.*, 18, 2829-2857.  
11 831
- 12  
13 832 Padowski, J.C., Gorelick, S.M., Thompson, B.H., Rozelle, S. & Fendorf, S. (2015). Assessment of  
14 833 human-natural system characteristics influencing global freshwater supply vulnerability. *Environ.*  
15 834 *Res. Lett.*, 10, 104014.  
16 835
- 17  
18 836 Peña-Arancibia, J. L., Mainuddin, M., Kirby, J. M., Chiew, F. H. S., McVicar, T. R., & Vaze, J.  
19 837 (2016). Assessing irrigated agriculture's surface water and groundwater consumption by combining  
20 838 satellite remote sensing and hydrologic modelling. *Sci. Total Environ.*, 542, 372–382.  
21 839
- 22  
23 840 Pokhrel, Y., Hanasaki, N., Koirala, S., Cho, J., Yeh, P. J.-F., Kim, H., Kanae, S. & Oki, T. (2012).  
24 841 Incorporating anthropogenic water regulation modules into a land surface model. *J. Hydrometeor.*,  
25 842 13(1), 255–269.  
26 843
- 27 844 Pokhrel, Y. N., Koirala, S., Yeh, P.J.-F., Hanasaki, N., Longuevergne, L., Kanae, S. & Oki, T. (2015).  
28 845 Incorporation of groundwater pumping in a global land surface model with the representation of  
29 846 human impacts, *Water Resour. Res.*, 51, 7896.  
30 847
- 31 848 Pokhrel, Y. N., Hanasaki, N., Wada, Y., & Kim, H. (2016). Recent progresses in incorporating human  
32 849 land-water management into global land surface models toward their integration into Earth system  
33 850 models. *Wiley Interdiscip. Rev. Water*, 3, 548-574.  
34 851
- 35 852 Portmann, F.T., Siebert, S. & Döll, P (2010). MIRCA2000 – Global monthly irrigated and rainfed  
36 853 crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological  
37 854 modelling. *Global Biogeochem. Cycles*, 24, GB1011.  
38 855
- 39 856 Ramankutty, N., Evan, A.T., Monfreda, C. & Foley, J.A. (2008). Farming the planet: 1. Geographic  
40 857 distribution of global agricultural lands in the year 2000. *Global Biogeochem. Cycles*, 22, GB1003  
41 858
- 42 859 Renard, B., Kavetski, D., Kuczera, G., Thyer, M., & Franks, S. W. (2010). Understanding predictive  
43 860 uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water*  
44 861 *Resources Research*,46, W05521.  
45 862
- 46 863 Revilla-Romero, B., Beck, H.E., Burek, P., Salamon, P., De Roo, A., Thielen, J. (2015). Filling the  
47 864 gaps: Calibrating a rainfall-runoff model using satellite-derived surface water extent. *Remote Sensing*  
48 865 *of Env.*, 171, 118-131.  
49 866
- 50 867 Rost, S., Gerten, D., Bondeau, A., Lucht, W., Rohwer, J. & Schaphoff, S. (2008). Agricultural green  
51 868 and blue water consumption and its influence on the global water system. *Water Resour. Res.* 44,  
52 869 W09405.  
53 870
- 54 871 Scanlon, B. R., et al. (2018). Global models underestimate large decadal declining and rising water  
55 872 storage trends relative to GRACE satellite data. *Proc. Nat. Acad. Sci.*, 201704665.  
56 873

- 874 Schaphoff, S., Heyder, U., Ostberg, S., Gerten, D., Heinke, J. & Lucht, W. (2013). Contribution of  
875 permafrost soils to the global carbon budget. *Env. Res. Lett.*, 8, 014026.  
876
- 877 Schewe, J., et al. (2014). Multimodel assessment of water scarcity under climate change. *Proc. Nat.  
878 Acad. Sci. U.S.A.*, 111, 3245-3250.  
879
- 880 Shiklomanov, I. A. (1997). Assessment of water resources and water availability in the world,  
881 Comprehensive assessment of the freshwater resources of the world. Stockholm, Sweden.  
882
- 883 Sood, A. & Smakhtin, V. (2015) Global hydrological models: a review. *Hydrol. Sci. J.*, 60, 549-565.  
884
- 885 Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S. & Hsu, K.-L. (2017). A review of global  
886 precipitation data sets: data sources, estimation, and intercomparisons, *Reviews of Geophysics*, 56.  
887
- 888 Takata, K., Emori, S. & Watanabe, T. (2003). Development of minimal advanced treatments of  
889 surface interaction and runoff. *Global Planet. Change*, 38, 209–222.  
890
- 891 Thiemig, V., Bisselink, B., Pappenberg, F., Thielen, J. (2015). A pan-African medium-range  
892 ensemble flood forecast system. *Hydr. Earth. Syst. Sci.*, 19, 3365-3385.  
893
- 894 Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., De Roo, A. (2013). Hydrological evaluation of  
895 satellite-based rainfall estimates over the Volta and Baro-Akobo basin. *J. of Hydrol.*, 499, 324-338.  
896
- 897 Thirel, G., Andréassian, V., Perrin, C., Audouy, J.-N., Berthet, L., Edwards, P, Folton, N., Furusho,  
898 C., Kuentz, A., Lerat, J., Lindström, G., Martin, E., Mathevet, T., Merz, R., Parajka, J., Ruelland, D.,  
899 Vaze, J. (2015). Hydrology under change: an evaluation protocol to investigate how hydrological  
900 models deal with changing catchments. *Hydrol. Sci. J.*, 60, 1184-1199.  
901
- 902 UNEP (2007). Global Environment Outlook 4: environment for development.  
903
- 904 Veldkamp, T. I. E., Eisner, S., Wada, Y., Aerts, J. C. J. H., & Ward, P. J. (2015). Sensitivity of water  
905 scarcity events to ENSO-driven climate variability at the global scale. *Hydrol. Earth Syst. Sci.*, 19,  
906 4081-4098.  
907
- 908 Veldkamp, T. I. E., Eisner, S., Wada, Y., Aerts, J. C. J. H., & Ward, P. J. (2015). Sensitivity of water  
909 scarcity events to ENSO-driven climate variability at the global scale. *Hydrol. Earth Syst. Sci.*, 19,  
910 4081-4098.  
911
- 912 Veldkamp, T. I. E., Wada, Y., Aerts, J. C. J. H., Döll, P., Gosling, S. N., Liu, J., Masaki, Y., Oki, T.,  
913 Ostberg, S., Pokhrel, Y., Satoh, Y., Kim, H., & Ward, P. J. (2017). Water scarcity hotspots travel  
914 downstream due to human interventions in the 20th and 21st century. *Nat. Comm.*, 8, 15697.  
915
- 916 Veldkamp, T. I. E., Wada, Y., Aerts, J. C. J. H., & Ward, P. J. (2016). Towards a global water  
917 scarcity risk assessment framework: incorporation of probability distributions and hydro-climatic  
918 variability. *Environ. Res. Lett.*, 11, 024006.  
919
- 920 Veldkamp, T. I. E., Wada, Y., de Moel, H., Kummu, M., Eisner, S., Aerts, J. C. J. H., & Ward, P. J.  
921 (2015b). Changing mechanisms of global water scarcity events: Impacts of socioeconomic changes  
922 and inter-annual hydro-climatic variability. *Glob. Env. Change*, 32, 18-29.  
923

- 1  
2  
3 924 Verzano, K., Barlund, I., Flörke, M., Lehner, B., Kynast, E. & Voß, F. (2012). Modeling variable  
4 925 river flow velocity on continental scale: Current situation and climate change impacts in Europe, *J.*  
5 926 *Hydrol.*, 424–425, 238–251.  
6 927
- 7 928 Vörösmarty, C. J., Leveque, C., & Revenga, C. (2005). Freshwater ecosystems. In *Millennium*  
8 929 *Ecosystem Assessment Volume 1: Conditions and Trends* (pp. 165–207).  
9 930
- 10 931 Wada, Y., Van Beek, L. P. H. & Bierkens, M. F. P. (2011). Modelling global water stress of the recent  
11 932 past: on the relative importance of trends in water demand and climate variability. *Hydrol. Earth Syst.*  
12 933 *Sci.*, 15, 3785–3808.  
13 934
- 14 935 Wada, Y., Van Beek, L. P. H., Wanders, N. & Bierkens, M. F. P. (2013a). Human water consumption  
15 936 intensifies hydrological drought worldwide. *Environ. Res. Lett.*, 8, 034036.  
16 937
- 17 938 Wada, Y., et al. (2013b). Multimodel projections and uncertainties of irrigation water demand under  
18 939 climate change. *Geophys. Res. Lett.*, 40, 4626-4632.  
19 940
- 20 941 Wada, Y., Gleeson, T. & Esnault, L. (2014a). Wedge approach to water stress. *Nat. Geosc.*, 7, 615-  
21 942 617.  
22 943
- 23 944 Wada, Y., Wisser, D. & Bierkens, M. F. P. (2014b). Global modelling of withdrawal, allocation and  
24 945 consumptive use of surface water and groundwater resources. *Earth Syst. Dyn.*, 5, 15–40.  
25 946
- 26 947 Wada, Y., et al. (2016a). Modeling global water use for the 21st century: the Water Futures and  
27 948 Solutions (WFaS) initiative and its approaches. *Geosci. Model Dev.*, 9, 175-222.  
28 949
- 29 950 Wada, Y., De Graaf, I. E. M. & Van Beek, L. P. H. (2016b) High-resolution modeling of human and  
30 951 climate impacts on global water resources. *J. Adv. Model. Earth Syst.*, 8, 735-763  
31 952
- 32 953 Wada, Y., et al. (2017). Human–water interface in hydrological modelling: current status and future  
33 954 directions, *Hydrol. Earth Syst. Sci.*, 21, 4169-4193  
34 955
- 35 956 Wanders, N., Wada, Y. & Van Lanen, H. A. J. (2015). Global hydrological droughts in the 21<sup>st</sup>  
36 957 century under a changing hydrological regime. *Earth Syst. Dynam.*, 6, 1-15.  
37 958
- 38 959 Wilks, D.S. (2006). On ‘Field significance’ and the false discovery rate. *J. Appl. Meteorol. Climatol.*,  
39 960 45, 1181-1189.  
40 961
- 41 962 Winsemius, H.C., et al. (2016). Global drivers of future flood risk. *Nat. Climate Change*, 6, 381-385.  
42 963
- 43 964 Wöhling, T., Samaniego, L., Kumar, R. (2013). Evaluating multiple performance criteria to calibrate  
44 965 the distributed hydrological model of the upper Neckar catchment. *Env. Earth. Sci.*, 59, 453-468.  
45 966
- 46 967 WRI, UNEP, UNDP, & World Bank (1998). World Resources 1998-99: A guide to the global  
47 968 environment -environmental change and human health. Washington D.C., U.S.A  
48 969
- 49 970 Yoshikawa, S., Cho, J., Yamada, H.G., Hansaki, N. & Kanae, S. (2014). An assessment of global net  
50 971 irrigation water requirements from various water supply sources to sustain irrigation: rivers and  
51 972 reservoirs (1960-2050). *Hydrol. Earth Syst. Sci.*, 18, 4289-4310.  
52 973

1  
2  
3 974 Zhang, L., Dobslaw, H., Stacke, T., Güntner, A., Dill, R., & Thomas, M. (2017). Validation of  
4 975 Terrestrial Water Storage Variations as Simulated by Different Global Numerical Models with  
5 976 GRACE Satellite Observations. *Hydrol. Earth Syst. Sci.*, 821–837.

6 977  
7 978 Zhao, F., *et al.* (2017). The critical role of the routing scheme in simulating peak river discharge in  
8 979 global hydrological models. *Environ. Res. Lett.*, 12, 075003.  
9 980

10  
11  
12 981  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Accepted Manuscript