LETTER • OPEN ACCESS

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

View the article online for updates and enhancements.

Related content

- The critical role of the routing scheme in simulating peak river discharge in global hydrological models
  Fang Zhao, Ted I E Veldkamp, Katja Frieler et al.

- Intercomparison of global river discharge simulations focusing on dam operation—multiple models analysis in two case-study river basins, Missouri–Mississippi and Green–Colorado
  Yoshimitsu Masaki, Naota Hanasaki, Hester Biemans et al.

- Towards a global water scarcity risk assessment framework: incorporation of probability distributions and hydro-climatic variability
  T I E Veldkamp, Y Wada, J C J H Aerts et al.

Recent citations

- Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts
  Jamal Zaherpour *et al*

Environmental Research Letters

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

T I E Veldkamp[1,9,13] , F Zhao[2] , P J Ward[1] , H de Moel[1], J C J H Aerts[1,3], H Müller Schmied[4,5] , F T Portmann[4], Y Masaki[6], Y Pokhrel[7], X Liu[8] , Y Satoh[9], D Gerten[2,10], S N Gosling[11] , J Zaherpour[11] and Y Wada[9,12]

1   Institute for Environmental Studies (IVM), VU Amsterdam, the Netherlands
2   Potsdam Institute for Climate Impact Research, Potsdam, Germany
3   Department of Geography, University of California, Santa Barbara, Santa Barbara, United States of America
4   Institute of Physical Geography, Goethe-University Frankfurt, Frankfurt, Germany
5   Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Frankfurt, Germany
6   National Institute for Environmental Studies, Tsukuba, Japan
7   Department of Civil and Environmental Engineering, Michigan State University, Michigan, United States of America
8   Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, People's Republic of China
9   International Institute for Applied Systems Analysis, Laxenburg, Austria
10 Department of Geography, Humboldt-Universität zu Berlin, Berlin, Germany
11 School of Geography, University of Nottingham, Nottingham, United Kingdom
12 Department of Physical Geography, Utrecht University, the Netherlands
13 Author to whom any correspondence should be addressed.

E-mail: ted.veldkamp@vu.nl

Abstract

Human activity has a profound influence on river discharges, 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 mean, high- and low-flows. The analysis is performed for 471 gauging stations across the globe for the period 1971–2010. We find that the inclusion of HIP improves the performance of the 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 the GHMs, although the level of improvement and the reasons for it vary greatly. The inclusion of HIP leads to a significant decrease in the bias of the long-term mean monthly discharge in 36%–73% of the studied catchments, and an improvement in the modeled 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 the simulated high-flows, it can lead to either increases or decreases in the low-flows. This is due to the relative importance of the timing of return flows and reservoir operations as well as their associated uncertainties. Even with the inclusion of HIP, we find that the model performance is still not optimal. This highlights the need for further research linking human management and hydrological domains, especially in those areas in which human impacts are dominant. 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 activity has a profound influence on river discharges, 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 activity in global hydrological models (hereafter GHMs; a full list of abbreviations is presented in supplementary table 2 available at stacks.iop.org/ERL/13/055008/mmedia) (Bierkens 2015, Pokhrel *et al* 2016). These model parameterizations include the incorporation of dam and reservoir operations, the representation of human water use and return flows, and representations of land use, land management and land cover change (Pokhrel *et al* 2016, Wada *et al* 2016a, 2017).

GHMs are widely used in scientific studies. For example, they have been used to assess the historical and future impacts of socioeconomic developments and/or hydro-climatic variability and change on freshwater resources, droughts and water scarcity (Biemans *et al* 2011, Döll *et al* 2009, Döll and Müller Schmied 2012, Fujimori *et al* 2017, Gosling *et al* 2017, Haddeland *et al* 2006, 2007, 2014, Hanasaki *et al* 2013, Van Huijgevoort *et al* 2013, Kummu *et al* 2016, Müller Schmied *et al* 2016, Munia *et al* 2016, Rost *et al* 2008, Veldkamp *et al* 2015a, 2015b, 2016, 2017, Wada *et al* 2011, 2013a, 2013b, 2014a, Wanders *et al* 2015). They are also increasingly used in practice. Global institutions are relying on GHMs more and more to conduct first-order assessments of water-related hazards, because data, time or resources are in short-supply for setting-up and executing multiple in-depth local studies. For example, GHMs have provided input into a multitude of high-level policy documents, such as the UN World Water Development Reports (e.g. Alcamo and Gallopin 2009), Global Environmental Outlooks (UNEP 2007), the World Bank series on climate change and development (Hallegatte *et al* 2016, 2017 and IPCC assessment reports (IPCC 2007, 2013).

As GHMs continue to improve in terms of detail, granularity and speed, their importance for global, regional and local applications is likely to increase further (Bierkens 2015). Therefore, it is essential to have a thorough understanding of how well these GHMs represent real-world hydrological conditions. However, most GHM validation studies are limited to near-natural river catchments and make use of naturalized discharge data (Beck *et al* 2016, Gudmundsson *et al* 2011, 2012). Studies that have validated GHM simulations where human activity is included have either focused on a single GHM and/or a few selected river catchments (Biemans *et al* 2011, Döll *et al* 2003, 2009, De Graaf *et al* 2014, Haddeland *et al* 2006, Masaki *et al* 2017, Müller Schmied *et al* 2014, Pokhrel *et al* 2012, Wada *et al* 2011, 2013a, 2014a).

To date, a comprehensive validation of the ability of multiple GHMs to represent the influence of human activity on discharge and hydrological extremes in near-natural and managed catchments is missing. As a result, there is a limited understanding of whether (and where) the parameterization of human activity in GHMs leads to an increase (or decrease) in model performance. To address this issue, the main objectives of this study are: (a) to evaluate the performance of five state-of-the-art GHMs that include the parameterization of human activity in their modeling scheme; and (b) to compare the performance of these GHMs when run with and without human impact parameterizations.

2. Data and methods

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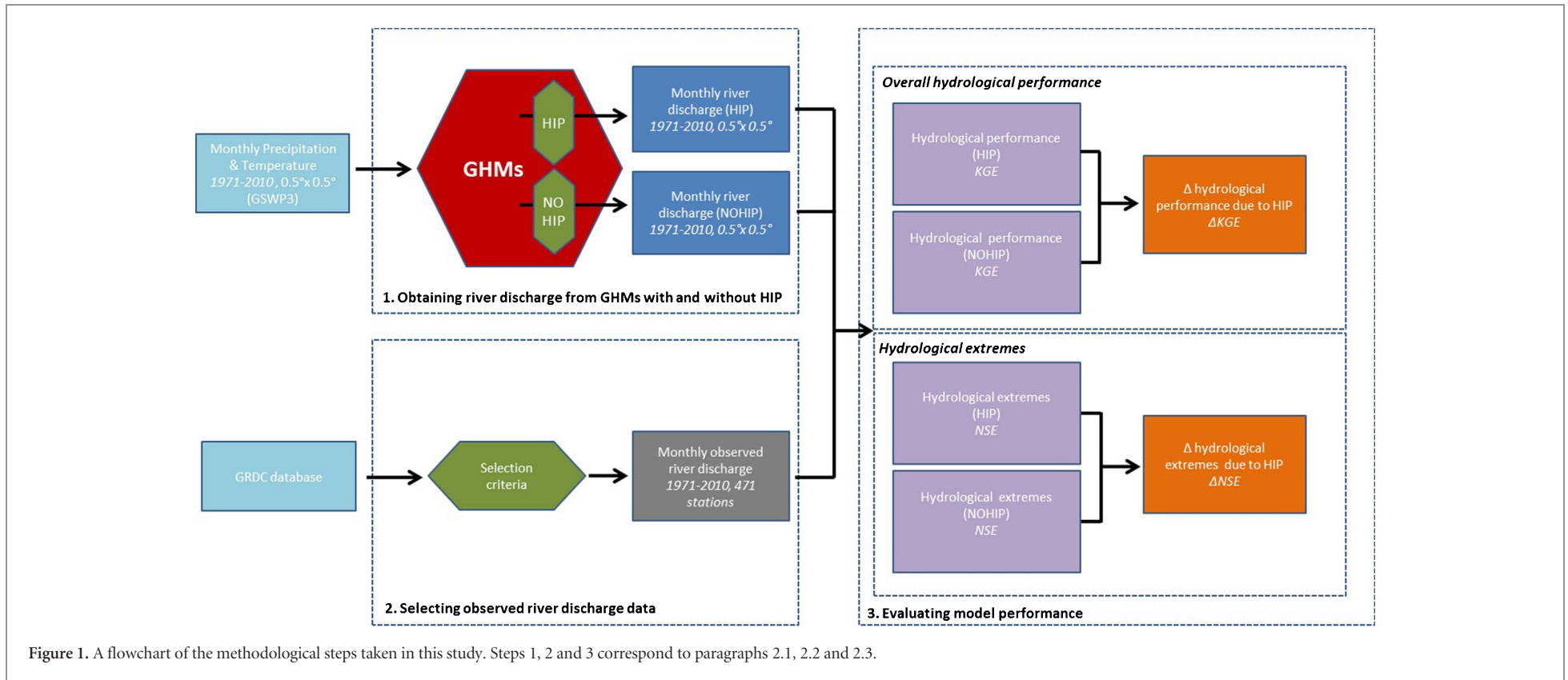
2.1. Obtaining modeled river discharge from GHMs with and without HIP

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

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

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 dataset (http://hydro.iis.u-tokyo.ac.jp/GSWP3). The GHMs applied in this study differ

Figure 1. A flowchart of the methodological steps taken in this study. Steps 1, 2 and 3 correspond to paragraphs 2.1, 2.2 and 2.3.

in hydrological representation and parameterization (supplementary table 1A). H08 and MATSIRO model the energy balance explicitly and use the bulk formula in the evaporation scheme (Hanasaki *et al* 2008a, 2008b, 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).

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 modeling 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–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.

2.1.2. Parameterization of human impact

All GHMs use a combination of socioeconomic and hydro-climatological parameters to estimate sectoral water demands (Hanasaki *et al* 2008a, 2008b, 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 table 1B) are estimated by combining historical gridded livestock density maps with their species-specific water demands. Domestic water demands (supplementary table 1C) 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 1D) are based on historical country-scale estimates from the WWDR-II dataset (Shiklomanov 1997, Vörösmarty *et al* 2005, WRI 1998) and the FAO-AQUASTAT database (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 1E) by multiplying the area equipped for irrigation with its utilization intensity, the total crop-specific water requirements—determined by the hydro-climatic conditions (temperature, precipitation, potential evapotranspiration, soil moisture,

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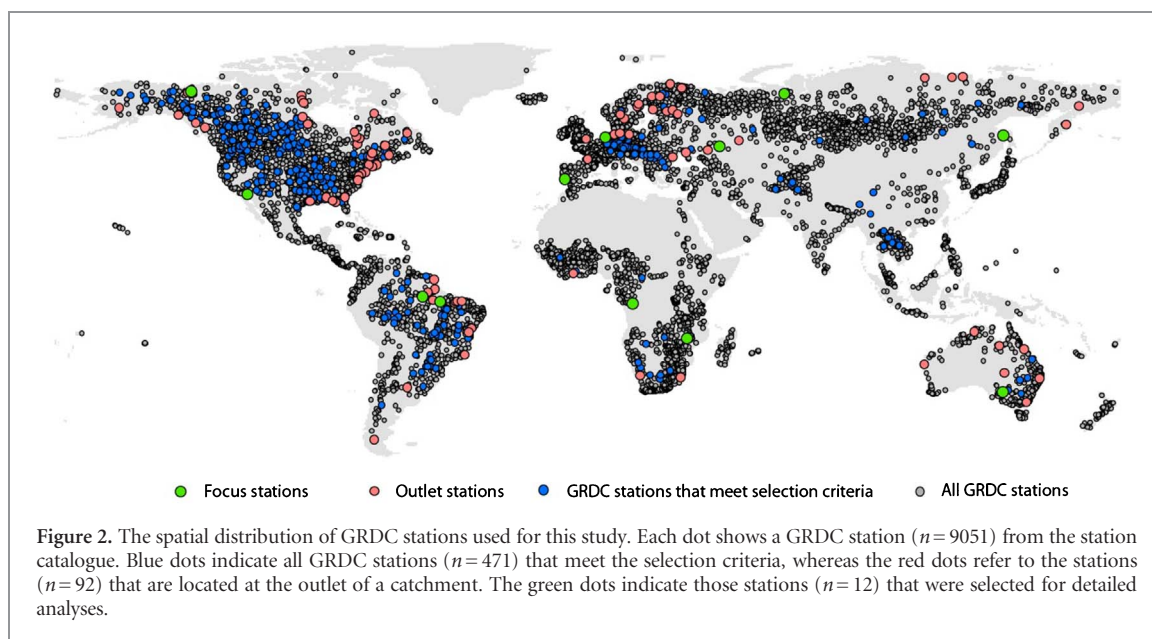
LPJmL, H08 and MATSIRO use surface water (first) to accommodate the sectoral water needs (supplementary table 1F). WaterGAP2 uses the groundwater to fulfill water demands, and surface water is only used if enough is available. PCR-GLOBWB applies a share of readily available groundwater reserves, based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

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

2.2. Selecting observed river discharge data

Observed monthly river discharge data was taken from the Global Runoff Data Centre (GRDC, 56068 Koblenz, Germany). From the 9051 gauging stations in the GRDC database, we selected stations that meet the following criteria: (1) a minimum of 5 years' coverage (not necessarily consecutive) during the period 1971–2010 with a completeness of observations of $\geq$95%; and (2) a minimum catchment area of 9000 km^2, to omit catchments whose hydrological processes cannot be adequately represented by models operating at $0.5° \times 0.5°$ (Hunger and Döll 2008). We discarded the stations for which the difference in the catchment area in the GRDC database and as estimated by using the DDM30 river routing network (Döll and Lehner 2002) is >25%.

We then made a distinction between near-natural and managed catchments. Following Beck *et al* (2016), a catchment is classified as near-natural if the share of land-area subject to irrigation is <2% and the

Figure 2. The spatial distribution of GRDC stations used for this study. Each dot shows a GRDC station ($n = 9051$) 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.

total reservoir capacity is <10% of its long-term mean annual discharge. If these conditions were not met the catchment was classified as managed. The classification was based on the HYDE 3/MIRCA land cover dataset (Fader *et al* 2010, Klein Goldewijk and Van Drecht 2006, Portmann *et al* 2010, Ramankutty *et al* 2008) together with the Global Reservoir and Dam database (Lehner *et al* 2011). Two stations shifted from near-natural to human-impacted conditions between 1971 and 2010 and were discarded from further analysis.

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

Figure 2 shows that the majority of selected stations (blue) are located in northern and Latin America, Europe, southern Africa and Australia. The number of stations in northern and central Africa and Asia is relatively small. We selected 12 stations in river basins located in different geographic regions (green circles in figure 2: Amazonas, Amur, Colorado, Congo, Guadiana, Mackenzie, Murray, Ob, Rhine, Tocantins, Volga and the Zambezi) for which a detailed analysis is provided in the supplementary results section (supplementary).

2.3. Evaluating model performance

To evaluate the GHM simulation of monthly discharge and hydrological extremes under HIP and NOHIP conditions, we compared the modeled results with observed river discharge data using several evaluation

metrics described below. To ensure a consistent comparison between the modeled and observed data, we only used modeled data for the same years for which observations were available. We also corrected modeled discharges for potential over- and underestimations caused by the difference in catchment size between the model and the GRDC. To do this, we used a multiplier that represented the difference in the upstream area as reported by the GRDC and as estimated from the DDM30 network.

First, we applied the modified Kling–Gupta efficiency index (KGE) with its sub-components: the linear correlation coefficient (rKGE), the bias ratio (βKGE) and the variability ratio (γKGE) (Gupta *et al* 2009, Kling *et al* 2012). The KGE is a widely applied indicator for the validation of hydrological performance in modeling studies at the global and regional scale and provides a good representation of the 'closeness' of simulated discharges to observations (Huang *et al* 2017, Kuentz *et al* 2013, Nicolle *et al* 2014, Revilla-Romero *et al* 2015, Thiemig *et al* 2013, 2015, Thirel *et al* 2015, Wöhling *et al* 2013). Moreover, use of its three sub-components enables the identification of reasons for sub-optimal model performance (Gupta *et al* 2009, Kling *et al* 2012, Thiemig *et al* 2013). This was achieved by estimating for each sub-parameter its distance to optimal performance, and by subsequently comparing these distances across the different sub-parameters. The statistical significance of the change in KGE outcomes due to the inclusion of HIP was tested by means of regular bootstrapping ($n = 1000$, $p \le 0.05$ (two-tailed)), following the method of Livezey and Chen (1982) and Wilks (2006).

Second, we applied the Nash–Sutcliffe efficiency test (NSE, Nash and Sutcliffe 1970) to evaluate the representation of Q_1 (high-flow) and Q_{99} (low-flow) conditions (e.g. Beck *et al* 2017a, Blösch *et al* 2013, Hejazi and Moglen 2008, Mohamoud 2008), obtained

under fixed threshold level settings (van Loon 2015). By means of a two-sample Kolmogorov–Smirnov (KS) test (Massey 1951, $p \leq 0.05$) we tested how often HIP leads to significant changes in the fit of the full modeled exceedance probability curve for hydrological extremes compared to the full observed exceedance probability curve.

3. Results

3.1. Validation and influence of human impact parameterization on overall model performance

Including the parameterization of human impacts in the GHMs leads to a large improvement in overall model performance. The 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 3(a)). 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 (figures 3(a)–(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 3(b)). For the variability ratio, improvements were found for 31.4%–74.4% of the land area for all catchments (48.9%–92.6% for the managed, 23.0%–73.2% for the near-natural) (figure 3(c)). The lowest improvements are found for the correlation coefficient, with improvements for 15.9%–58.1% of the total land area for all catchments (22.1%–75.1% for the managed, 13.9%–61.4% for the near-natural) (figure 3(d)).

The 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 the northwest US. 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).

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

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

Spatially explicit results vary per GHM and are shown in supplementary figure 3. The distribution of dominant contributors to the sub-optimal overall hydrological performance is similar for H08, LPJmL, and PCR-GLOBWB. For these GHMs, we find dominant contributions from the bias ratio in southern Africa, Australia and inland US, and dominant contributions of the variability ratio and the correlation coefficient in Latin America as well as at higher latitude and altitude regions. For Europe, the dominant contributions for H08, LPJmL and PCR-GLOBWB are the variability ratio, the correlation coefficient, and the bias ratio respectively. The dominant contributors that cause sub-optimal overall hydrological performance for MATSIRO and WaterGAP2 are more equally distributed across the globe. While all sub-components contribute to sub-optimal overall hydrological model performance for MATSIRO, it is predominantly the correlation coefficient and the variability ratio that determines the sub-optimal performance in WaterGAP2.

3.2. Validation and influence of human impact parameterizations on the simulation of hydrological extremes

The inclusion of HIP in the simulations affects the ability of GHMs to estimate hydrological extremes cor-
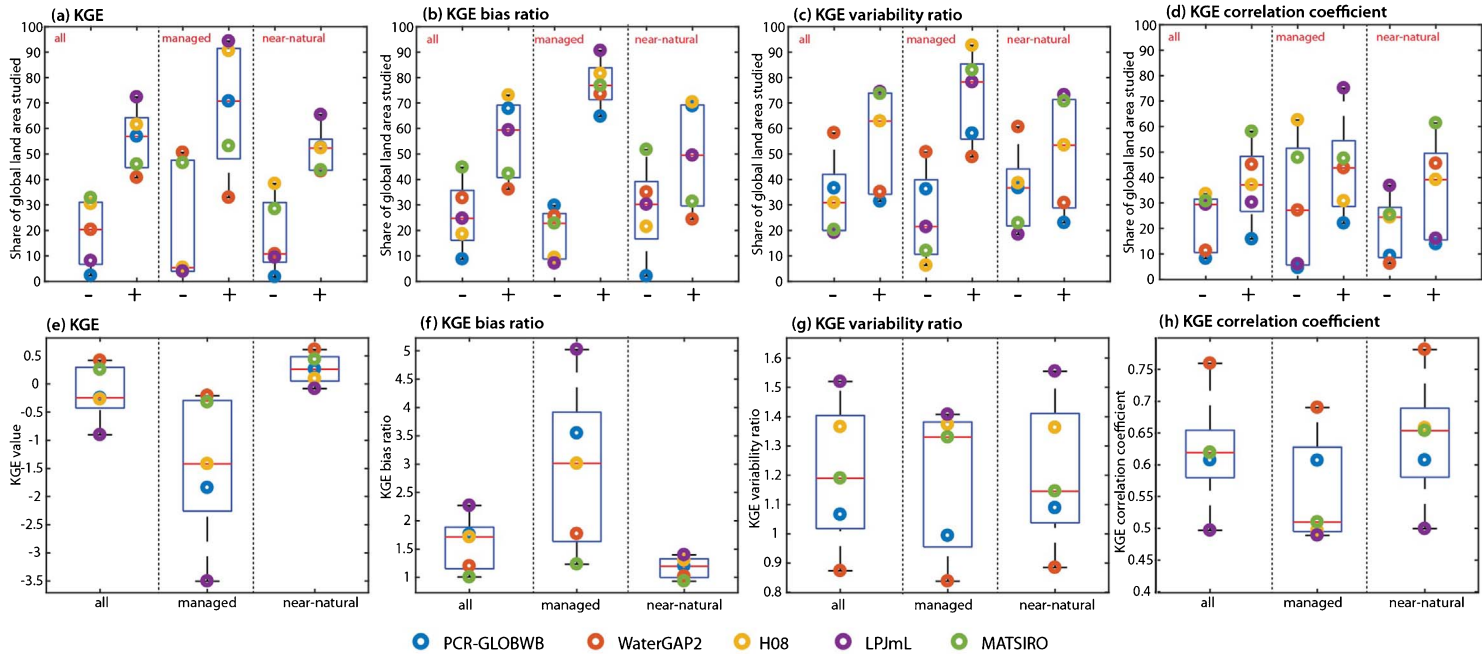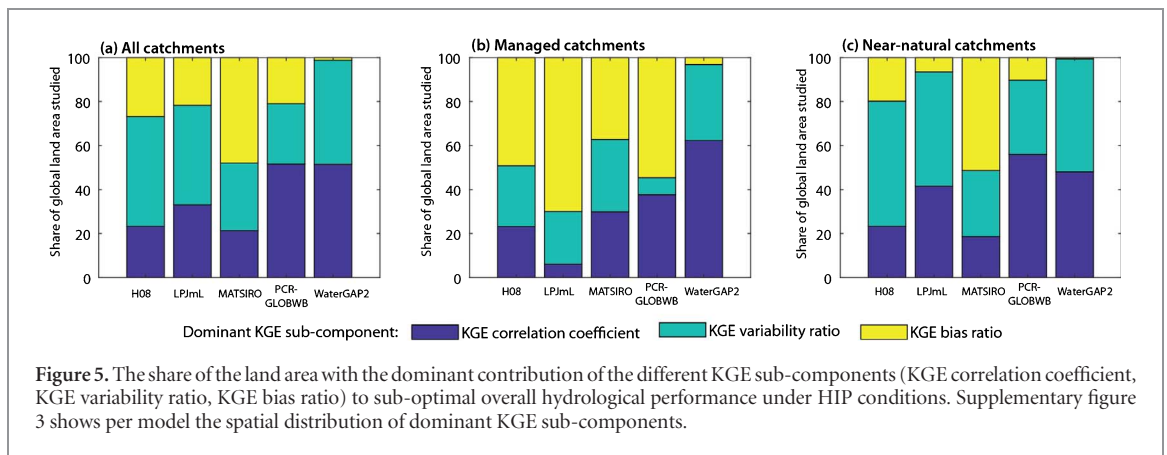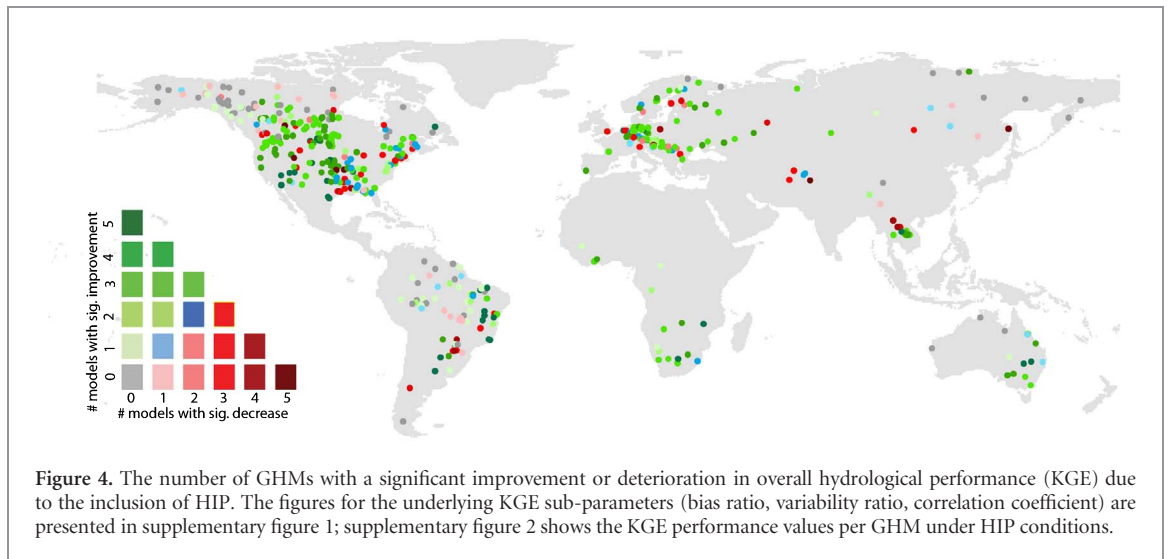
Figure 3. Global weighted-mean (improvement ('+') or deterioration ('−')) in the representation of hydrological performance due to HIP for all catchments, managed catchments and near-natural catchments. Figures 3(a)–(d) visualize for each GHM the share of land area with a significant change in overall hydrological performance due to the inclusion of HIP. Figures 3(e)–(h) indicate the globally weighted-mean hydrological performance after the inclusion of HIP. On each box, the red mark indicates the median. The bottom and top edges of the box indicate the 25th and 75th percentiles of the model ensemble, respectively.

Figure 4. The number of GHMs with a significant improvement or deterioration in overall hydrological performance (KGE) due to the inclusion of HIP. The figures for the underlying KGE sub-parameters (bias ratio, variability ratio, correlation coefficient) are presented in supplementary figure 1; supplementary figure 2 shows the KGE performance values per GHM under HIP conditions.

Figure 5. The share of the land area with the dominant contribution of the different KGE sub-components (KGE correlation coefficient, KGE variability ratio, KGE bias ratio) to sub-optimal overall hydrological performance under HIP conditions. Supplementary figure 3 shows per model the spatial distribution of dominant KGE sub-components.

rectly in the majority of the land area studied (figure 6). The inclusion of HIP leads to better model performance for all GHMs, across a substantial share of the land area studied (figures 6(*a*)–(*b*)). For high-flows, HIP improves model performance significantly across 34.6%–77.0% of the land area for all catchments (36.4%–94.7% for managed, 24.1%–79.2% for near-natural). For low-flows, HIP improves model performance significantly across 39.4%–80.4% of the land area for all catchments (29.3%–81.8% for managed, 42.7%–90.3% for near-natural). The KS-test results (supplementary figure 4) show that HIP only leads to significant changes in the representation of the exceedance probability curve in a limited number of cases for H08 and LPJmL (up to 14.1% of the land area studied), predominantly in managed catchments.

Overall, hydrological extremes are represented reasonably well under HIP conditions, with globally weighted mean NSE values ranging between 0.8–0.98 for high-flows, and 0.84–0.98 for low-flows (figures 6(*c*)–(*d*)). However, there is a significant difference in the ability of the GHMs to represent hydrological extremes between managed and near-natural catchments.

Figure 7 indicates that for the majority of stations, the inclusion of HIP leads to an improvement in the representation of hydrological extremes, for most GHMs. A deterioration in the representation of hydrological extremes across the majority of GHMs as a result of the inclusion of HIP was only found in selected areas, for example at higher latitudes and along the east coast of the US. When comparing the results for Q_1 high-flows with Q_{99} low-flows, no large differences in the spatial distribution of the number of GHMs are found with a significant improvement or deterioration.

The effects of HIP on the magnitude of extreme discharge differ for low-flows and high-flows (supplementary figure 5). Whilst the magnitude of high-flows mostly decreases with the inclusion of HIP, the effects on the magnitude of low-flows are both positive and negative. The convergence of results towards higher observed discharges, in both high- and low-flow estimates (as identified for all models in supplementary figure 5), indicates that HIP becomes less important for the correct representation of hydrological extremes with increasing discharge volumes.

Table 1. The performance metrics used in this study and their calculation procedure. Here, s_i and o_i are the simulated and observed monthly discharge at station i; μ_s and μ_o are the simulated and observed mean monthly discharge at station i; σ_s and σ_o are the standard deviation of the simulated and observed 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	$\mathrm{KGE} = 1 - \sqrt{(r\mathrm{KGE}^* - 1)^2 + (\beta\mathrm{KGE}^* - 1)^2 + (\gamma\mathrm{KGE}^* - 1)^2}$	$-\infty - 1$ (ideal value: 1)
rKGE	KGE correlation coefficient (Pearson)	$r\mathrm{KGE} = \dfrac{\sum_{i=1}^{n}(s_i - \mu_{s,i})(o_i - \mu_{o,i})}{\sqrt{\sum_{i=1}^{n}(s_i - \mu_{s,i})^2}\sqrt{\sum_{i=1}^{n}(o_i - \mu_{o,i})^2}}$	$-1 - 1$ (ideal value: 1)
βKGE	KGE bias ratio	$\beta\mathrm{KGE} = \mu_{s,i}/\mu_{o,i}$	$0 - \infty$ (ideal value: 1)
γKGE	KGE variability ratio	$\gamma\mathrm{KGE} = \dfrac{\sigma_{s,i}/\mu_{s,i}}{\sigma_{o,i}/\mu_{o,i}}$	$0 - \infty$ (ideal value: 1)
NSE	Nash–Sutcliffe model efficiency	$\mathrm{NSE} = 1 - \dfrac{\sum(Q_s - Q_o)^2}{\sum(Q_o - \overline{Q_o})^2}$	$-\infty - 1$ (ideal value: 1)
Q_1	High-flow indicator	Monthly discharge (m s^{-3}) that is exceeded on average in 1 out of 100 months	
Q_{99}	Low-flow indicator	Monthly discharge (m s^{-3}) that is exceeded on average in 99 out of 100 months	
KS	Two-sample Kolmogorov–Smirnov test	$[h, p] = \mathrm{kstest2}\,(\mathrm{cdf}\,(Q_s), \mathrm{cdf}\,(Q_o), \text{'Alpha'}, 0.05)^{\mathrm{a}}$	For $p > 0.05$, H_0 (the two cdfs coming from the same distribution) is not rejected.

[a] The calculation procedure for the two-sample Kolmogorov–Smirnov test presented in the table is the Matlab function for the KS-test.
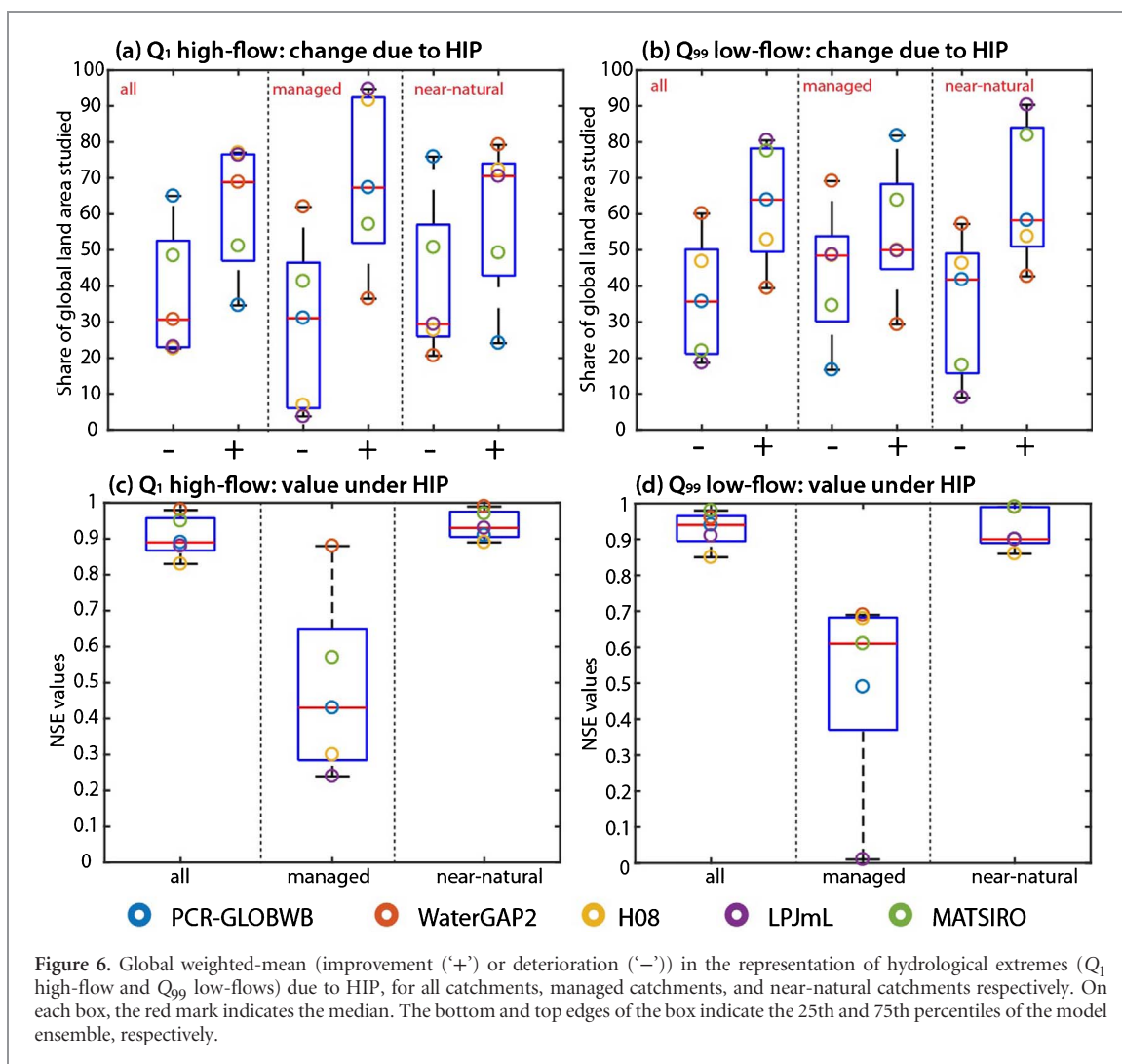

Figure 6. Global weighted-mean (improvement ('+') or deterioration ('−')) in the representation of hydrological extremes (Q_1 high-flow and Q_{99} low-flows) due to HIP, for all catchments, managed catchments, and near-natural catchments respectively. On each box, the red mark indicates the median. The bottom and top edges of the box indicate the 25th and 75th percentiles of the model ensemble, respectively.

Figure 7. The number of GHMs with a significant improvement or deterioration in the representation of hydrological extremes due to the inclusion of HIP.
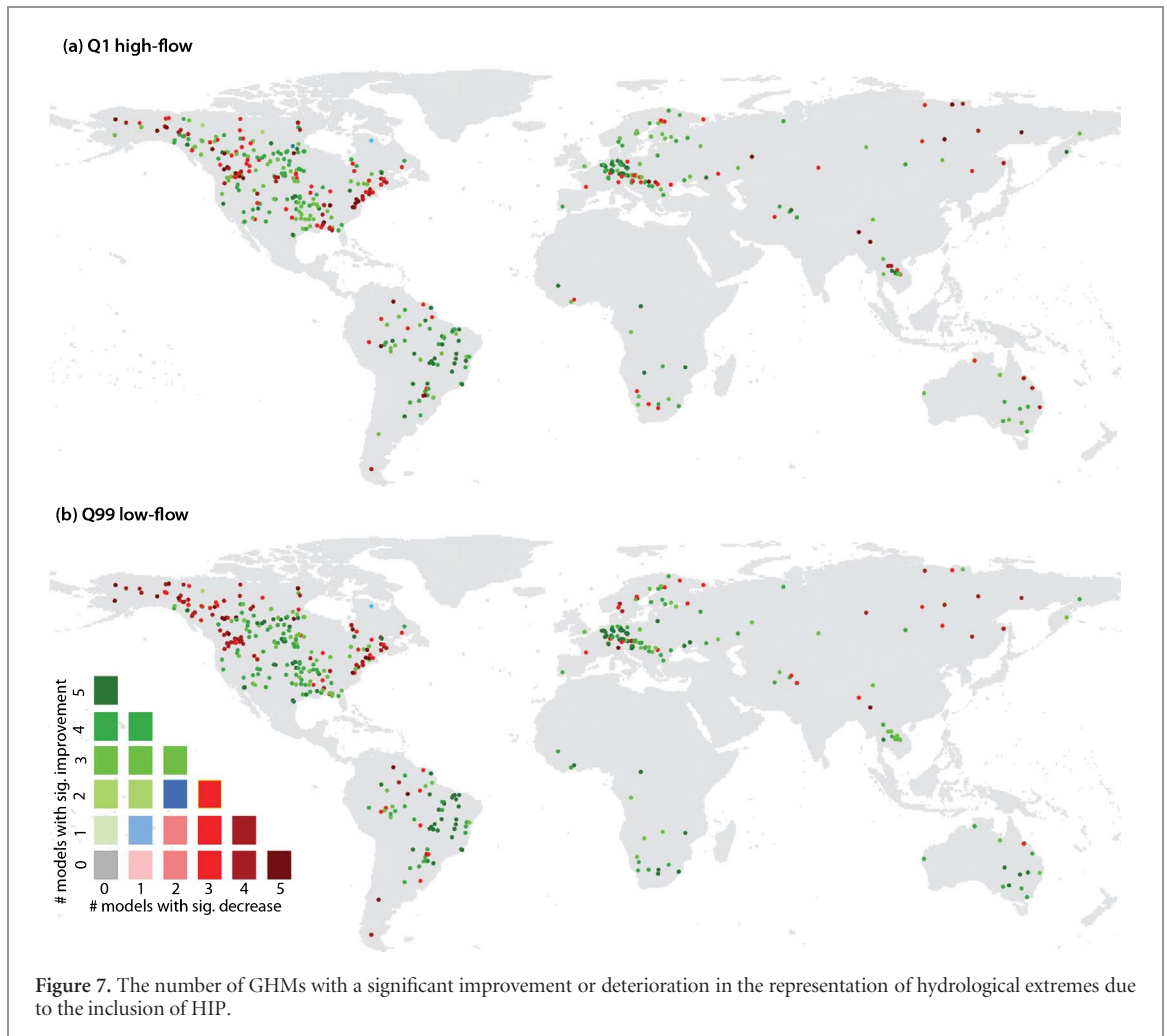
4. Discussion

Our results show that including HIP in GHMs generally improves the overall hydrological performance of the GHMs, as well as their representation of hydrological extremes. However, we also show that further improvements are needed. In this section, we discuss: (1) possible reasons for the improved model performance due to HIP; (2) the main limitations of the current modeling frameworks and their representation of HIP, as well as potential ways to improve them; and (3) we reflect on the general limitations in the current study design and provide suggestions for further research.

4.1. Improvements in model performance due to HIP and challenges ahead

Whilst the inclusion of HIP predominantly leads to the largest improvements in simulated discharge in the managed catchments, simulated discharge is also improved in a large share of the near-natural catchments. Improvements in model performance associated with the inclusion of HIP can be attributed to improvements in the different KGE sub-components, and in turn to different model components

parameterizing hydrological and human processes. In addition, insights into those factors bounding the optimal hydrological model performance under HIP conditions may help to identify priorities for further model improvement.

4.1.1. Representation of long-term mean discharges (bias ratio)

Our study shows that the representation of long-term mean discharges significantly improved with the inclusion of HIP, especially in the managed catchments. The inclusion of HIP generally results in lower simulated discharges. As most GHMs systematically overestimate river discharges in the NOHIP simulation, this results in an improved performance. When HIP is included, we only find a deterioration in the bias ratio in selected higher latitude/altitude regions, where discharges are underestimated; this finding is in line with the outcomes of single-model studies performed by Döll *et al* (2009), De Graaf *et al* (2014) and Haddeland *et al* (2006). Improvements in the bias ratios due to the inclusion of HIP can be attributed to the inclusion of water abstractions and return flows (supplementary table 1B–G), and the incorporation of irrigated areas and irrigation rules, which influence

evapotranspiration rates and the generation of runoff (supplementary table 1E).

However, despite improvement in the bias ratio with the inclusion of HIP, this KGE sub-indicator contributes most to sub-optimal performance in managed catchments for H08, LPJmL, MATSIRO and PCR-GLOBWB under HIP conditions. As the GHMs continue to overestimate long-term mean discharges in most cases under HIP conditions, future model improvements should target the correction of this bias in these locations. This may be achieved by critically revisiting the methods used to represent evapotranspiration rates (supplementary table 1A), runoff generation processes (supplementary table 1A) and the level of water abstractions in managed catchments (supplementary table 1B–E). The relatively good performance of WaterGAP2, in which biases in long-term mean annual discharge are adjusted using a parameter that determines the portion of effective precipitation that becomes surface runoff (Müller Schmied *et al* 2014), highlights the potential importance of including a calibration routine (supplementary table 1I). Calibration is also performed for H08, but this calibration routine aims to minimize runoff bias by modifying two parameters of subsurface flow for four climatic groups (Hanasaki *et al* 2008a, 2008b); it is therefore less effective at minimizing the bias ratio under HIP conditions.

4.1.2. Representation of hydrological variability (variability ratio)

The inclusion of HIP leads to mixed results regarding the representation of hydrological variability. Whilst HIP improved the representation of variability in some catchments and for some GHMs, it deteriorated the representation of variability for others. For example, it led to improvements in the west coast US, southern Africa and Australia, but a deterioration for most GHMs in Europe and inland US. Similar results were found by Biemans *et al* (2011), De Graaf *et al* (2014) and Masaki *et al* (2017) for a selection of catchments. Changes in the variability ratio due to the inclusion of HIP are predominantly driven by the timing of water abstractions and return flows, as well as by reservoir operation rules (supplementary table 1F–H). These human activities influence the relative size of high- and low-flows compared to their long-term mean discharge values.

The variability ratio is the KGE sub-parameter that contributes most to the sub-optimal performance in near-natural catchments with the inclusion of HIP, for H08, LPJmL and WaterGAP2. These GHMs significantly overestimate hydrological variability in near-natural catchments (except WaterGAP2, which underestimates variability in managed and near-natural catchments), and model improvement should therefore focus on better representing the speed of hydrological response, e.g. through an improved representation of the soil moisture storage capacity or the ratio between surface and sub-surface runoff (supplementary table 1A). In those cases where the variability ratio is also the KGE sub-parameter that contributes most to sub-optimal performance in managed catchments, model improvement should target the timing of water abstractions, return flows and reservoir management (supplementary table 1F–H).

4.1.3. Representation of the goodness-of-fit (correlation coefficient)

The inclusion of HIP only led to improved correlation coefficients in limited cases, and often resulted in a deterioration, even in managed catchments. Correlation coefficients between observed and modeled discharges, which are predominantly determined by the hydro-meteorological forcing data (Döll *et al* 2016, Beck *et al* 2016), were found to be generally high under both HIP and NOHIP conditions. Perturbations of the hydrological cycle due to human activity leading to changes in the timing of discharges and in the shape of the hydrograph, like return flows and reservoir operations, explain the observed decrease in the correlation coefficient in a substantial share of catchments and models globally (supplementary table 1F–H).

Under HIP conditions, the correlation coefficient is the KGE sub-parameter that contributes most to sub-optimal performance only in PCR-GLOBWB for near-natural catchments and WaterGAP2 for managed catchments. It should be acknowledged, though, that correlation coefficients for PCR-GLOBWB and WaterGAP2 are relatively high, especially compared to the other GHMs. The relatively low correlation coefficients in near-natural catchments found at higher latitudes in all models may be addressed by critically reviewing the snow accumulation and melt processes in the GHMs (supplementary table 1A). Higher correlation coefficients in the managed catchments may be established by improving the timing and quantification of return flow estimates and the representativeness of reservoir operations (supplementary table 1F–H).

4.1.4. Representation of hydrological extremes

The inclusion of HIP also led to significant changes in the ability of most GHMs to represent hydrological extremes (both high- and low-flows), although the strength of this change is very much dependent on the location and GHM in question. Whilst the magnitude of high-flow estimates mainly decreased due to the inclusion of HIP, low-flow estimates showed mixed results. This is because the impact of human activity tends to be greater for lower discharges, as the relative 'size' of human perturbations (such as water abstractions, return flows or delayed releases of water via reservoir operations) is higher as a percentage of the overall discharge when flows are low. Both De Graaf *et al* (2014) and Wada *et al* (2013a) found similar results when investigating hydro-climatic extremes. However, even with the inclusion of

HIP, the representation of hydrological extremes is sub-optimal. Future model improvements should aim to better characterize these extremes and to improve the representation of human activity during extreme hydrological conditions.

4.2. Limitations and further research

As the GHMs have very different parameterizations of hydrological and human processes, the current study does not allow a systematic assessment of specific cause–effect relations between HIP and the observed improvements in performance (Döll *et al* 2016, Haddeland *et al* 2014, Hagemann *et al* 2013, Schewe *et al* 2014, Beck *et al* 2016). To do this, a substantial Monte Carlo analysis would be required, whereby individual parameters and combinations of parameters are systematically modified for all GHMs (Döll *et al* 2016). Undertaking such an analysis in parallel to the different GHMs incorporated is computationally expensive and requires a strict modeling protocol. It may, however, provide additional information on how to adapt and improve the individual GHMs and would be a valuable addition to the results presented in this study.

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

The spatial resolution of the GHMs applied in this study is $0.5° \times 0.5°$ ($\sim50\,km \times 50\,km$ at the equator), dictated by the resolution of the GSWP3 input dataset. At a $0.5° \times 0.5°$ spatial resolution, hydrological processes are often represented by GHMs in a simplified or generalized form which are not fit for local applications (Bierkens 2015). To account for this, we applied a minimum catchment size of $9000\,km^2$, thereby omitting catchments that are too small to be adequately represented by GHMs (Hunger and Döll 2008). Newer versions of several of the GHMs now operate at higher resolutions; for example WaterGAP and PCR-GLOBWB have recently been published in 5 m/6 m versions respectively (Verzano *et al* 2012, Wada *et al* 2016b). Future research can investigate whether the inclusion of these high-resolution model-runs improves the representation of discharges and hydrological extremes in the selected catchments and whether these high-resolution runs also allow for the inclusion of smaller catchments.

In this study, a relatively simple distinction was made between managed and near-natural catchments using two parameters: irrigated agriculture and reservoirs. These parameters were chosen as they have been reported to be the most significant human param-

eters on river hydrology (Beck *et al* 2016, 2017a). However, to make a more detailed distinction between catchments that are impacted by human activity and those that are not, future studies could consider incorporating additional criteria, such as the share of sectoral water abstractions and return flows, and the share of built-up land area. Additional catchment descriptors (Eisner 2016), like climate conditions and the physiographic properties of the drainage area, could also be applied to further assess the important controls on modeled discharges.

When evaluating the impact of HIP on hydrological extremes, we only incorporated results for the Q_1 high-flow and Q_{99} low-flow. In this study, we did not consider other ranges of the extreme value distribution explicitly. Although the inclusion of HIP influences these hydrological extremes substantially, we found very few instances for which this led to a significant change in the full exceedance probability curve. Future research should therefore also incorporate other ranges of the probability exceedance curve in order to do a full assessment of the influence of HIP on high- and low-flow extremes.

Next to the parameterizations and representation of hydrological processes and human impacts, other sources contribute to uncertainty in the modeling of discharges and hydrological extremes. These include the quality of the data, the uncertainties in the input data and observation datasets and the calibration/validation strategy (Döll *et al* 2016, Sood and Smakhtin 2015). The quality of the selected forcing data, for example, may limit the representation of monthly discharges and hydrological extremes significantly (Döll *et al* 2016, Beck *et al* 2016), although this not been evaluated explicitly here. However, climate forcing uncertainty is probably a dominant driver for model outputs (Müller Schmied *et al* 2014, 2016). Benchmarking of the GSWP3 dataset against historical observations of precipitation and temperature, or against other forcing datasets (e.g. similar to Beck *et al* 2017b, Sun *et al* 2018), may therefore be of added value.

Differences in the quality and trustworthiness of the historical discharge observations (e.g. due to sampling, measurement and interpretation errors), may potentially result in artificial biases in the validation results (Renard *et al* 2010). The spatial representativeness of our results is limited by the availability of consistent publicly available *in situ* observations of sufficient quality. Future research should therefore consider extending the GRDC data-points to the regional repositories of observed discharges, as recently attempted by Beck *et al* (2016), Do *et al* (2017) and Gudmundsson *et al* (2017). However, increasing spatial representation comes at the cost of consistency, and special attention should be paid to the harmonization of these different databases. The use of remotely sensed data could also provide a valuable way of carrying out calibration and validation in ungauged

regions (Döll *et al* 2014a, 2014b, Scanlon *et al* 2018). Remotely sensed data can also be of added value in the assessment of water consumed by agricultural irrigation (Peña-Arancibia *et al* 2016), operational drought monitoring and early warning (Ahmadalipour *et al* 2017), as well as the estimation of terrestrial water budgets (Zhang *et al* 2017). Moreover, clear potential exists for assimilating remotely sensed data into the models (Eicker *et al* 2014).

Calibration and validation are essential to compensate for factors such as the impossibility of measuring all required model parameters at the applied scale, the lack of process understanding, the simplistic process representation in GHMs and errors in forcing data (Beck *et al* 2016, Bierkens 2015, Döll *et al* 2016, Liu *et al* 2017). Hence, calibration/validation is key for obtaining realistic model performance. It should be acknowledged, though, that the representation of hydrological and/or human processes is artificially altered by means of calibration/validation processes and that limited calibration may introduce uncertainties to the model output (Sood and Smakhtin 2015). Before using any calibrated/validated model data, one should therefore critically reflect on whether the calibration/validation procedure executed—together with their optimization objectives—are fit for the specific application in mind.

5. Summary and conclusions

This study shows that the inclusion of human activity in GHMs can significantly improve the simulation of monthly discharges and hydrological extremes, for the majority of catchments studied. The finding is robust across both managed and near-natural catchments. The global and spatially distributed results presented in this study indicate that the inclusion of human impact parameterizations (HIP) is associated with improvements in the bias ratio and the variability ratio. Whilst the biases in long-term mean monthly discharge decrease significantly in 36.1%–73.0% of the studied catchments due to the inclusion of HIP, the modeling of hydrological variability improves significantly in 31.4%–74.4% of the catchments. Estimates of hydrological extremes are also significantly influenced by the inclusion of HIP, although the influence is highly dependent on the location and GHM in question. While HIP generally leads to a decrease (and thus improvement) in the absolute magnitude of simulated high-flows, its impact on low-flows is mixed.

Even when human activity is included in GHMs, their performance is still limited; this is particularly the case in managed catchments. Moreover, the systematic misrepresentation of hydrological extremes across all GHMs calls for a careful interpretation of risk assessments based on their results, and further study into the overarching research theme

of water resources, hydrological extremes, human interventions, and feedback linkages. The large variation in performance between GHMs, regions and performance indicators, highlights the importance of carefully selecting models, model components and evaluation metrics in future model applications. For example, for a study of droughts it is essential to correctly represent hydrological variability, whilst to study water scarcity it is crucial to minimize biases.

Sub-KGE results, which were presented in this study for each GHM, allow for the attribution of different hydrological and human impact model-components limiting optimal hydrological performance. In most GHMs, model performance is limited due to the overestimation of long-term mean discharges. The correlation coefficient is the limiting factor for optimal model performance for Water-GAP2, despite the high correlation coefficients that were found for this model relative to the other GHMs studied. A better understanding of these factors, as provided by this study, may assist in the identification of priorities for further model improvement.

ORCID iDs

T I E Veldkamp ⓘ https://orcid.org/0000-0002-2295-8135
F Zhao ⓘ https://orcid.org/0000-0002-4819-3724
P J Ward ⓘ https://orcid.org/0000-0001-7702-7859
H Müller Schmied ⓘ https://orcid.org/0000-0001-5330-9923
X Liu ⓘ https://orcid.org/0000-0001-5726-7353
S N Gosling ⓘ https://orcid.org/0000-0001-5973-6862

References

Adam J C, Haddeland I, Su F and Lettenmaier D P 2007 Simulation of reservoir influences on annual and seasonal streamflow changes for the Lena, Yenisei, and Ob' rivers *J. Geophys. Res. Atmos.* **112** 1–22

Ahmadalipour A, Moradkhani H, Hongxiang Y and Zarekarizi M 2017 Remote sensing of drought: vegetation, soil moisture, and data assimilation *Remote Sensing of Hydrological Extremes* ed V Lakshmi (Switzerland: Springer) pp 121–49

Alcamo J and Gallopín G 2009 Building a 2nd generation of world water scenarios *Side Publications Series—Insights* (Paris: United Nations World Water Assessment Programme)

Beck H E, Van Dijk A I J, De Roo A, Miralles D G, McVicar T M, Schellekens J and Bruijnzeel L A 2016 Global-scale regionalization of hydrologic model parameters *Water Resour. Res.* **52** 3599–622

Beck H E, Van Dijk A I J, De Roo A, Dutra E, Fink G, Orth R and Schellekens J 2017a Global evaluation of runoff from 10 state-of-the-art hydrological models *Hydrol. Earth Syst. Sci.* **21** 2881–903

Beck H E, Vergopolan N, Pan M, Levizzani V, Van Dijk A I J, Weedom G P, Brocca L, Pappenberger F, Huffman G J and Wood E F 2017b Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modelling *Hydrol. Earth Syst. Sci.* **21** 6201–17

van Beek L P H, Wada Y and Bierkens M F P 2011 Global monthly water stress: I. Water balance and water availability *Water Resour. Res.* **47** W07517

Biemans H, Haddeland I, Kabat P, Ludwig F, Hutjes R W A, Heinke J, Von Bloh W and Gerten D 2011 Impact of reservoirs on river discharge and irrigation water supply during the 20th century *Water Resour. Res.* **47** W03509

Bierkens M F P 2015 Global hydrology 2015: state, trends, and directions *Water Resour. Res.* **51** 4923–47

Blösch G, Sivapalan M, Wagener T, Viglione A and Savenije H 2013 *Runoff Predictions in Ungauged Basins: Synthesis Across Processes, Places and Scales* (Cambridge: Cambridge University Press)

Bondeau A *et al* 2007 Modelling the role of agriculture for the 20th century global terrestrial carbon balance *Glob. Change Biol.* **13** 679–706

Do H X, Gudmundsson L, Leonard M, Westra S and Seneviratne S I 2017 The global streamflow indices and metadata archive (GSIM)—Part 1: the production of daily streamflow archive and metadata *Earth Syst. Sci. Data Discuss.* in review

Döll P and Lehner B 2002 Validation of a new global 30 m in drainage direction map *J. Hydrol.* **258** 214–31

Döll P and Siebert S 2002 Global modelling of irrigation water requirements *Water Resour. Res.* **38** W1037

Döll P, Kaspar F and Lehner B 2003 A global hydrological model for deriving water availability indicators: model tuning and validation *J. Hydrol.* **270** 105–34

Döll P, Fiedler K and Zhang J 2009 Global-scale analysis of river flow alterations due to water withdrawals and reservoirs *Hydrol. Earth Syst. Sci.* **13** 2413–32

Döll P and Müller Schmied H 2012 How is the impact of climate change on river flow regimes related to the impact on mean annual runoff? A global-scale analysis *Environ. Res. Lett.* **7** 014037

Döll P, Fritsche M, Eicker A and Müller Schmied H 2014a Seasonal water storage variations as impacted by water abstractions: comparing the output of a global hydrological model with GRACE and GPS observations *Surv. Geophys.* **35** 1311–31

Döll P, Müller Schmied H, Schuh C, Portmann F T and Eicker A 2014b Global-scale assessment of groundwater depletion and related groundwater abstractions: combining hydrological modeling with information from well observations and GRACE satellites *Water Resour. Res.* **50** 5698–720

Döll P, Douville H, Güntner A, Müller Schmied H and Wada Y 2016 Modelling freshwater resources at the global scale: challenges and prospects *Surv. Geophys.* **37** 195–221

Eicker A, Schumacher M, Kusche J and Döll P and Müller Schmied H 2014 Calibration/data assimilation approach for integrating GRACE data into the WaterGAP global hydrology model (WGHM) using an ensemble Kalman filter: first results *Surv. Geophys.* **35** 1285–309

Eisner S 2016 Comprehensive evaluation of the WaterGAP3 Model across climatic, physiographic, and anthropogenic gradients *PhD Thesis* University of Kassel, Germany (http://nbn-resolving.de/urnnbndehebis34-2016031450014)

Fader M, Rost S, Muller C, Bondeau A and Gerten D 2010 Virtual water content of temperate cereals and maize: Present and potential future patterns *J. Hydrol.* **384** 218–31

Flörke M, Kynast E, Bärlund I, Eisner S, Wimmer F and Alcamo J 2013 Domestic and industrial water uses of the past 60 years as a mirror of socio-economic development: A global simulation study *Glob. Environ. Change* **23** 144–56

Fujimori S, Hanasaki N and Masui T 2017 Projections of industrial water withdrawal under shared socioeconomic pathways and climate mitigation scenarios *Sustain. Sci.* **12** 275–92

Gosling S N *et al* 2017 A comparison of changes in river runoff from multiple global and catchment-scale hydrological models under global warming scenarios of 1 °C, 2 °C, and 3 °C *Clim. Change* **141** 577–95

Gupta H V, Kling H, Yilmaz K K and Martinez G F 2009 Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling *J. Hydrol.* **377** 80–91

de Graaf I E M, van Beek L P H, Wada Y and Bierkens M F P 2014 Dynamic attribution of global water demand to surface water and groundwater resources: effects of abstractions and return flows on river discharges *Adv. Water Resour.* **64** 21–33

Gudmundsson L *et al* 2011 Comparing large-scale hydrological model simulations to observed runoff percentiles in Europe *J. Hydrometeor.* **13** 604–20

Gudmundsson L, Wagener T, Tallaksen L M and Engeland K 2012 Evaluation of nine large-scale hydrological models with respect to the seasonal runoff climatology in Europe *Wat. Resour. Res.* **48** W11504

Gudmundsson L, Do H X, Leonard M, Westra S and Seneviratne S I 2017 The global streamflow indices and metadata archive (GSIM)—Part 2: quality control, time-series indices and homogeneity assessment *Earth Syst. Sci. Data Discuss.* in review

Haddeland I, Skaugen T and Lettenmaier D P 2006 Anthropogenic impacts on continental surface water fluxes *Geophys. Res. Lett.* **33** L08406

Haddeland I, Skaugen T and Lettenmaier D P 2007 Hydrologic effects of land and water management in North America and Asia: 1700–1992 *Hydrol. Earth Syst. Sci.* **11** 1035–45

Haddeland I *et al* 2014 Global water resources affected by human interventions and climate change *Proc. Natl Acad. Sci. USA* **111** 3251–6

Hagemann S *et al* 2013 Climate change impact on available water resources obtained using multiple global climate and hydrology models *Earth Syst. Dyn.* **4** 129–44

Hallegatte S, Bangalore M, Bonzanigo L, Fay M, Kane T, Narloch U, Rozenberg J, Treguer D and Vogt-Schilb A 2016 Shock waves: managing the impacts of climate change on poverty *Climate Change and Development* (Washington, DC: World Bank)

Hallegatte S, Vogt-Schilb A, Bangalore M and Rozenberg J 2017 Unbreakable: building the resilience of the poor in the face of natural disasters *Climate Change Development* (Washington, DC: World Bank)

Hanasaki N, Kanaeo S and Oki T 2006 A reservoir operation scheme for global river routing models *J. Hydrol.* **327** 22–41

Hanasaki N, Kanae S, Oki T, Masuda K, Motoya K, Shirakawa N, Shen Y and Tanaka K 2008a An integrated model for the assessment of global water resources—Part 1: model description and input meteorological forcing *Hydrol. Earth Syst. Sci.* **12** 1007–25

Hanasaki N, Kanae S, Oki T, Masuda K, Motoya K, Shirakawa N, Shen Y and Tanaka K 2008b An integrated model for the

assessment of global water resources—Part 2: applications and assessments *Hydrol. Earth Syst. Sci.* **12** 1027–37

Hanasaki N *et al* 2013 A global water scarcity assessment under shared socio-economic pathways—Part 2: water availability and scarcity *Hydrol. Earth Syst. Sci.* **17** 2393–413

Hejazi M I and Moglen G E 2008 The effect of climate and land use change on flow duration in the Maryland Piedmont region *Hydrol. Proc.* **22** 4710–22

Huang S *et al* 2017 Evaluation of an ensemble of regional hydrological models in 12 large-scale river basins worldwide *Clim. Change* **141** 381–97

Hunger M and Döll P 2008 Value of river discharge data for global-scale hydrological modelling *Hydrol. Earth Syst. Sci.* **12** 841–61

van Huijgevoort M H J *et al* 2013 Global multimodel analysis of drought in runoff for the second half of the twentieth century *J. Hydrometeor.* **14** 1535–52

IPCC 2007 Climate change 2014: synthesis report *Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Geneva: IPCC)

IPCC 2013 *Climate Change 2013: The Physical Science Basis* (Geneva: IPCC)

Klein Goldewijk K and Van Drecht G 2006 HYDE 3: Current and historical population and land cover. MNP 2006 *Integrated Modelling of Global Environmental Change. An Overview of IMAGE 2.4* ed A F Bouwman, T Kram and K K Goldewijk (Bilthoven: Netherlands Environmental Assessment Agency (MNP))

Kling H, Fuchs M and Paulin M 2012 Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios *J. Hydrol.* **424** 264–77

Kuentz A, Mathevet T, Gailhard J, Perret C and Andreassian V 2013 Over 100 years of climatic and hydrologic variability of a Mediterranean and mountainous watershed: the Durance river Cold and Mountain region hydrological systems under climate change: towards improved projections proceedings of H02, *IAHS-IAPSO-IASPEI Assembly* (Gothenburg: IAHS) p 360

Kummu M, Guillaume J H A, De Moel H, Eisner S, Flörke M, Porkka M, Siebert S, Veldkamp T I E and Ward P J 2016 The world's road to water scarcity: shortage and stress in the 20th century and pathways towards sustainability *Sci. Rep.* **6** 38495

Lehner B *et al* 2011 High resolution mapping of the world's reservoirs and dams for sustainable river flow management *Front. Ecol. Environ.* **9** 494–502

van Loon A F 2015 Hydrological drought explained *WIREs Water* **2** 359–92

van Loon A F *et al* 2016 Drought in the anthropocene *Nat. Geosci.* **9** 89–91

Livezey R E and Chen W Y 1982 Statistical field significance and its determination by Monte Carlo techniques *Mon. Weather Rev.* **111** 46–59

Liu X, Tang Q, Cui H, Mu M, Gerten D, Gosling S, Masaki Y, Wada Y and Satoh Y 2017 Multimodel uncertainty changes in simulated river flows induced by human impact parameterizations *Environ. Res. Lett.* **12** 025009

Masaki *et al* 2017 Intercomparison of regulated river discharge among multiple global hydrological models under multiple forcings—Part II: multiple models in two case-study river basins, Missouri-Mississippi and Green-Colorado *Environ. Res. Lett.* **12** 055002

Massey F J 1951 The Kolmogorov-Smirnov test for goodness of fit *J. Am. Stat. Assoc.* **46** 68–78

Mohamoud Y M 2008 Prediction of daily flow duration curves and streamflow for ungauged catchments using regional flow duration curves *Hydrol. Sci. J.* **53** 706–24

Müller Schmied H, Eisner S, Franz D, Wattenbach M, Portmann F T, Flörke M and Döll P 2014 Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration *Hydrol. Earth. Syst. Sci.* **18** 3511–38

Müller Schmied H *et al* 2016 Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use *Hydrol. Earth Syst. Sci.* **20** 2877–98

Munia H, Guillaume J H A, Mirumachi N, Porkka M, Wada Y and Kummu M 2016 Water stress in global transboundary river basins: significance of upstream water use on downstream stress *Environ. Res. Lett.* **11** 014002

Nash J E and Sutcliffe J V 1970 River flow forecasting through conceptual models part I—A discussion of principles *J. Hydrol.* **10** 282–90

Nicolle P *et al* 2014 Benchmarking hydrological models for low-flow simulation and forecasting on French catchments *Hydrol. Earth Syst. Sci.* **18** 2829–57

Padowski J C, Gorelick S M, Thompson B H, Rozelle S and Fendorf S 2015 Assessment of human-natural system characteristics influencing global freshwater supply vulnerability *Environ. Res. Lett.* **10** 104014

Peña-Arancibia J L, Mainuddin M, Kirby J M, Chiew F H S, McVicar T R and Vaze J 2016 Assessing irrigated agriculture's surface water and groundwater consumption by combining satellite remote sensing and hydrologic modelling *Sci. Total Environ.* **542** 372–82

Pokhrel Y, Hanasaki N, Koirala S, Cho J, Yeh P J-F, Kim H, Kanae S and Oki T 2012 Incorporating anthropogenic water regulation modules into a land surface model *J. Hydrometeorol.* **13** 255–69

Pokhrel Y N, Koirala S, Yeh P J-F, Hanasaki N, Longuevergne L, Kanae S and Oki T 2015 Incorporation of groundwater pumping in a global land surface model with the representation of human impacts *Water Resour. Res.* **51** 7896

Pokhrel Y N, Hanasaki N, Wada Y and Kim H 2016 Recent progresses in incorporating human land-water management into global land surface models toward their integration into Earth system models *Wiley Interdiscip. Rev. Water* **3** 548–74

Portmann F T, Siebert S and Döll. P 2010 MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: a new high-resolution data set for agricultural and hydrological modelling *Glob. Biogeochem. Cycles* **24** GB1011

Ramankutty N, Evan A T, Monfreda C and Foley J A 2008 Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000 *Glob. Biogeochem. Cycles* **22** GB1003

Renard B, Kavetski D, Kuczera G, Thyer M and Franks S W 2010 Understanding predictive uncertainty in hydrologic modeling: the challenge of identifying input and structural errors *Water Resour. Res.* **46** W05521

Revilla-Romero B, Beck H E, Burek P, Salamon P, De Roo A and Thielen J 2015 Filling the gaps: calibrating a rainfall-runoff model using satellite-derived surface water extent *Remote Sens. Environ.* **171** 118–31

Rost S, Gerten D, Bondeau A, Lucht W, Rohwer J and Schaphoff S 2008 Agricultural green and blue water consumption and its influence on the global water system *Water Resour. Res.* **44** W09405

Scanlon B R *et al* 2018 Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data *Proc. Natl Acad. Sci.* **2017** 04665

Schaphoff S, Heyder U, Ostberg S, Gerten D, Heinke J and Lucht W 2013 Contribution of permafrost soils to the global carbon budget *Environ. Res. Lett.* **8** 014026

Schewe J *et al* 2014 Multimodel assessment of water scarcity under climate change *Proc. Natl Acad. Sci. USA* **111** 3245–50

Shiklomanov I A 1997 *Assessment of Water Resources and Water Availability in the World, Comprehensive Assessment of the Freshwater Resources of the World* (Stockholm: Stockholm Environment Institute)

Sood A and Smakhtin V 2015 Global hydrological models: a review *Hydrol. Sci. J.* **60** 549–65

Sun Q, Miao C, Duan Q, Ashouri H, Sorooshian S and Hsu K-L 2018 A review of global precipitation data sets: data sources, estimation, and intercomparisons *Rev. Geophys.* **56**

Takata K, Emori S and Watanabe T 2003 Development of minimal advanced treatments of surface interaction and runoff *Glob. Planet. Change* **38** 209–22

Thiemig V, Bisselink B, Pappenberg F and Thielen J 2015 A pan-African medium-range ensemble flood forecast system *Hydr. Earth. Syst. Sci.* **19** 3365–85

Thiemig V, Rojas R, Zambrano-Bigiarini M and De Roo A 2013 Hydrological evaluation of satellite-based rainfall estimates over the Volta and Baro-Akobo basin *J. Hydrol.* **499** 324–38

Thirel G *et al* 2015 Hydrology under change: an evaluation protocol to investigate how hydrological models deal with changing catchments *Hydrol. Sci. J.* **60** 1184–99

UNEP 2007 *Global Environment Outlook 4: environment for development* (Valletta: Progress Press) (http://drustage.unep.org/geo/assessments/global-assessments/global-environment-outlook-4)

Veldkamp T I E, Eisner S, Wada Y, Aerts J C J H and Ward P J 2015a Sensitivity of water scarcity events to ENSO-driven climate variability at the global scale *Hydrol. Earth Syst. Sci.* **19** 4081–98

Veldkamp T I E *et al* 2017 Water scarcity hotspots travel downstream due to human interventions in the 20th and 21st century *Nat. Comm.* **8** 15697

Veldkamp T I E, Wada Y, Aerts J C J H and Ward P J 2016 Towards a global water scarcity risk assessment framework: incorporation of probability distributions and hydro-climatic variability *Environ. Res. Lett.* **11** 024006

Veldkamp T I E, Wada Y, de Moel H, Kummu M, Eisner S, Aerts J C J H and Ward P J 2015b Changing mechanisms of global water scarcity events: Impacts of socioeconomic changes and inter-annual hydro-climatic variability *Glob. Environ. Change* **32** 18–29

Verzano K, Barlund I, Flörke M, Lehner B, Kynast E and Voß F 2012 Modeling variable river flow velocity on continental scale: current situation and climate change impacts in Europe *J. Hydrol.* **424** 238–51

Vörösmarty C J, Leveque C and Revenga C 2005 Freshwater ecosystems *Millennium Ecosystem Assessment Volume 1: Conditions and Trends* (Washington, DC: Island Press) pp 165–207

Wada Y, Van Beek L P H and Bierkens M F P 2011 Modelling global water stress of the recent past: on the relative importance of trends in water demand and climate variability *Hydrol. Earth Syst. Sci.* **15** 3785–808

Wada Y, Van Beek L P H, Wanders N and Bierkens M F P 2013a Human water consumption intensifies hydrological drought worldwide *Environ. Res. Lett.* **8** 034036

Wada Y *et al* 2013b Multimodel projections and uncertainties of irrigation water demand under climate change *Geophys. Res. Lett.* **40** 4626–32

Wada Y, Gleeson T and Esnault L 2014a Wedge approach to water stress *Nat. Geosc.* **7** 615–7

Wada Y, Wisser D and Bierkens M F P 2014b Global modelling of withdrawal, allocation and consumptive use of surface water and groundwater resources *Earth Syst. Dyn.* **5** 15–40

Wada Y *et al* 2016a Modeling global water use for the 21st century: the water futures and solutions (WFaS) initiative and its approaches *Geosci. Model Dev.* **9** 175–222

Wada Y, De Graaf I E M and Van Beek L P H 2016b High-resolution modeling of human and climate impacts on global water resources *J. Adv. Model. Earth Syst.* **8** 735–63

Wada Y *et al* 2017 Human–water interface in hydrological modelling: current status and future directions *Hydrol. Earth Syst. Sci.* **21** 4169–93

Wanders N, Wada Y and van Lanen H A J 2015 Global hydrological droughts in the 21st century under a changing hydrological regime *Earth Syst. Dyn.* **6** 1–15

Wilks D S 2006 On 'field significance' and the false discovery rate *J. Appl. Meteorol. Climatol.* **45** 1181–9

Winsemius H C *et al* 2016 Global drivers of future flood risk *Nat. Clim. Change* **6** 381–5

Wöhling T, Samaniego L and Kumar R 2013 Evaluating multiple performance criteria to calibrate the distributed hydrological model of the upper Neckar catchment *Env. Earth. Sci.* **59** 453–68

WRI, UNEP, UND, and Bank World 1998 *World Resources 1998–99: A Guide to the Global Environment—Environmental Change and Human Health* (Washington, DC)

Yoshikawa S, Cho J, Yamada H G, Hansaki N and Kanae S 2014 An assessment of global net irrigation water requirements from various water supply sources to sustain irrigation: rivers and reservoirs 1960–2050 *Hydrol. Earth Syst. Sci.* **18** 4289–310

Zhang L, Dobslaw H, Stacke T, Güntner A, Dill R and Thomas M 2017 Validation of terrestrial water storage variations as simulated by different global numerical models with GRACE satellite observations *Hydrol. Earth Syst. Sci.* **21** 821–37

Zhao F *et al* 2017 The critical role of the routing scheme in simulating peak river discharge in global hydrological models *Environ. Res. Lett.* **12** 075003