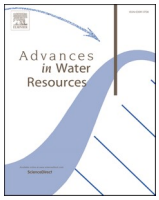
Advances in Water Resources 165 (2022) 104212

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres



# Multi-model evaluation of catchment- and global-scale hydrological model simulations of drought characteristics across eight large river catchments

Amit Kumar [a,*], Simon N. Gosling [a], Matthew F. Johnson [a], Matthew D. Jones [a], Jamal Zaherpour [a], Rohini Kumar [b], Guoyong Leng [c], Hannes Müller Schmied [d,e], Jenny Kupzig [f], Lutz Breuer [g,h], Naota Hanasaki [i], Qiuhong Tang [c], Sebastian Ostberg [j], Tobias Stacke [k], Yadu Pokhrel [l], Yoshihide Wada [m], Yoshimitsu Masaki [i,n]

[a] School of Geography, University of Nottingham, Nottingham, UK
[b] Department of Computational Hydrosystems, Helmholtz Centre for Environmental Research – UFZ, Leipzig, Germany
[c] Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China
[d] Institute of Physical Geography, Goethe University, Frankfurt, Germany
[e] Senckenberg Leibniz Biodiversity and Climate Research Centre (SBiK-F), Frankfurt am Main, Germany
[f] Institute of Engineering Hydrology and Water Resources Management, Ruhr-University, Bochum, Germany
[g] Institute for Landscape Ecology and Resources Management (ILR), Research Centre for Biosystems, Land Use and Nutrition (iFZ), Justus Liebig University, Giessen, Germany
[h] Centre for International Development and Environmental Research (ZEU), Justus Liebig University, Giessen, Germany
[i] National Institute for Environmental Studies, Tsukuba, Japan
[j] Earth System Analysis, Potsdam Institute for Climate Impact Research, Potsdam, Germany
[k] Regional Land and Atmosphere Modelling, Institute of Coastal Research, Helmholtz-Zentrum, Geesthacht, Germany
[l] Department of Civil and Environmental Engineering, Michigan State University, Michigan, USA
[m] International Institute for Applied Systems Analysis (IIASA), Laxenburg, Asutria
[n] Ibaraki University, Japan

ARTICLE INFO

*Keywords:*
Global hydrological models
Catchment hydrological models
Hydrological droughts
Model evaluation
Model validation
ISIMIP

ABSTRACT

Although global- and catchment-scale hydrological models are often shown to accurately simulate long-term runoff time-series, far less is known about their suitability for capturing hydrological extremes, such as droughts. Here we evaluated simulations of hydrological droughts from nine catchment scale hydrological models (CHMs) and eight global scale hydrological models (GHMs) for eight large catchments: Upper Amazon, Lena, Upper Mississippi, Upper Niger, Rhine, Tagus, Upper Yangtze and Upper Yellow. The simulations were conducted within the framework of phase 2a of the Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP2a). We evaluated the ability of the CHMs, GHMs and their respective ensemble means (Ens-CHM and Ens-GHM) to simulate observed hydrological droughts of at least one month duration, over 31 years (1971–2001). Hydrological drought events were identified from runoff-deficits and the Standardised Runoff Index (SRI). In all catchments, the CHMs performed relatively better than the GHMs, for simulating monthly runoff-deficits. The number of drought events identified under different drought categories (i.e. SRI values of -1 to -1.49, -1.5 to -1.99, and $\leq$-2) varied significantly between models. All the models, as well as the two ensemble means, have limited abilities to accurately simulate drought events in all eight catchments, in terms of their occurrence and magnitude. Overall, there are opportunities to improve both CHMs and GHMs for better characterisation of hydrological droughts.

## 1. Introduction

A drought is an event where water availability is lower than normal, resulting in a failure to fulfil the water demands of different natural systems and socioeconomic sectors (WMO, 1986). From 1991 to 2005, 950 million people were affected by droughts worldwide and economic

* Corresponding author.
*E-mail address:* amit1792kumar@gmail.com (A. Kumar).

https://doi.org/10.1016/j.advwatres.2022.104212
Received 14 October 2021; Received in revised form 13 April 2022; Accepted 18 April 2022
Available online 21 April 2022

damage of 100 billion US dollars was reported UN and UNISDR, UNDP, I., (2009). Droughts are usually the consequence of a prolonged period of below normal precipitation that also affects many other environmental, climate and socio-economic variables (Lloyd-Hughes, 2014; Van Loon, 2015). Drought can be difficult to identify in space and time, which makes it one of the most complex natural hazard (Wilhite, 1993; Wilhite et al., 2000). Researchers, managers and policy makers quantify drought events using drought indices based on climate data (reviewed in Heim, 2002; Keyantash and Dracup, 2002; Mishra and Singh, 2010). Whilst precipitation is a key input to calculate these indices, other climate and environmental variables that affect water storage and availability, are also significant.

Droughts are complex natural disasters as their onset and magnitude are related to the interaction between many hydrological and climatological processes. Droughts can be classified into different types namely meteorological, hydrological, agricultural and socioeconomic droughts. Meteorological droughts represent below normal precipitation and are mainly presented by precipitation driven indices such as the Standardised Precipitation Index (SPI; McKee et al., 1993), Regional Drought Area Index (RDAI; Fleig et al., 2011) and Effective drought index (EDI; Byun and Wilhite, 1999). In contrast, hydrological droughts define effects on freshwater storage, which are represented by indices that use stream flows, reservoir levels, groundwater levels or other similar variables. Hydrological droughts are often closely related to meteorological droughts and can also be exacerbated by environmental changes, anthropogenic activities, and mismanagement of water resources (Tallaksen et al., 2004).

Studies on hydrological droughts at global or continental scales increasingly use Land Surface Models (LSMs), Global Hydrological Models (GHMs) and Catchment Scale Hydrological Models (CHMs) to quantify and predict drought events (Gosling, Zaherpour, et al., 2017; Hattermann et al., 2017). GHMs, LSMs and CHMs have been widely used to model flood hazards and risk (Arnell and Gosling 2016), climate change mitigation (Irvine et al., 2017), forecasting at shorter time scales (Emerton et al., 2016) and food security (Elliott et al., 2014). The use of these models to study droughts is also relatively common (Van Huijgevoort, et al., 2013; Prudhomme et al., 2014) but there is relatively less information on the performance of the models for simulating drought. Given the societal significance of drought prediction under climate change scenarios (Pokhrel et al., 2021) using these tools, and the influence of results on climate change adaptation and mitigation decision-making, it is critical to understand their strengths and limitations when specifically focusing on drought assessment and prediction.

Whilst some studies have assessed the ability of multiple CHMs (Huang et al., 2017) and GHMs (Zaherpour et al., 2018) to simulate historical low flows, there has to date been no study that compares the performance of CHMs and GHMs with a focus on drought characteristics. Thus the main novelty of this study is that it is the first cross-scale model evaluation of drought event *frequency, intensity/severity* and *duration*, together referred to as 'droughts' hereafter. This was achieved through the computation of observed and simulated runoff-deficits and the Standardised Runoff Index (SRI).

The main objective of this study was to systematically evaluate the performance of several global scale and catchment scale hydrological models to simulate droughts. We also discuss the opportunities for improving the simulation of drought events by both GHMs and CHMs.

## 2. Data

### 2.1. Study catchments

Eight large (> 65,000 $km^2$) catchments were selected to cover diverse climate zones and hydrological systems around the globe. These were, the Upper Amazon, Lena, Upper Mississippi, Upper Niger, Rhine, Tagus, Upper Yangtze and Upper Yellow (Fig. 1). The same eight catchments used in previous GHM-CHM comparisons (Gosling et al., 2017; Hattermann et al., 2017). For the analysis, only the upper part of the Amazon, Mississippi, Niger, Yangtze and Yellow were modelled due to their complicated geomorphological structure and human alterations further downstream (Krysanova and Hattermann, 2017). Catchment boundaries were defined according to Drainage Direction Maps at 30′ (DDM30; Döll and Lehner, 2002) for the GHMs and CHMs, and according to the Global Runoff Data Centre (GRDC; http://grdc.bafg.de) for the observed data.

### 2.2. Models and input data

We used CHM and GHM simulations from ISIMIP2a (Gosling et al., 2017) to identify historical drought events from their respective monthly runoff simulations, and evaluated how these droughts compared to the observed record. The simulations are the most up-to-date opportunity that exists to robustly and consistently compare

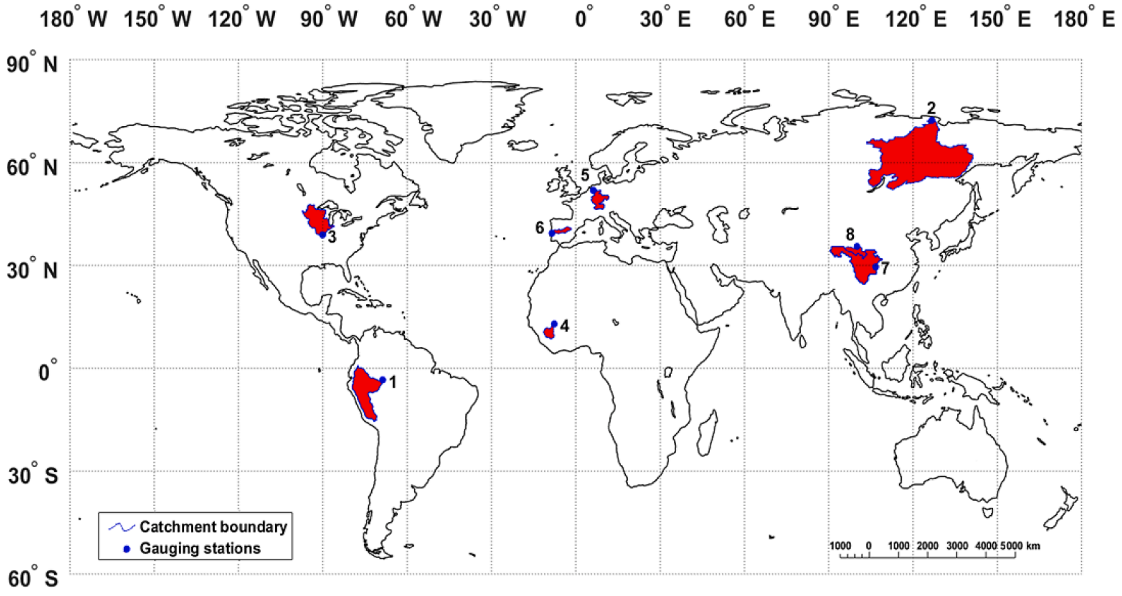



**Fig. 1.** Location of the eight study catchments labelled as (1) Upper Amazon, (2) Lena, (3) Upper Mississippi, (4) Upper Niger, (5) Rhine, (6) Tagus, (7) Upper Yangtze, and (8) Upper Yellow.

the performance of several CHMs and GHMs.

Simulated runoff data from nine CHMs and eight GHMs along with corresponding observed runoff data from the GRDC, for 1971–2001, were used in this study. Observed monthly runoff data was acquired from the most downstream gauge in the GRDC catalogue for each catchment. All GHMs and CHMs were run with daily input climate data from WATCH ERA-40 (Weedon et al., 2011) for the period 1971–2001, which we aggregated to monthly temporal resolution. Output from the models is openly available from the Earth System Grid Federation (ESGF; https://esg.pik-potsdam.de/search/isimip) for the GHMs (Gosling et al., 2017) and the CHMs (Krysanova et al., 2017). All the GHMs provided outputs for all catchments; however, the number of CHMs with simulated runoff varied by catchment (Table 1). Following the method described by Haddeland et al. (2011), monthly observed and simulated runoff data was converted to catchment-mean monthly runoff by using the area upstream of the gauge according to the DDM30 river network. Thus, an area correction factor was applied to the GRDC runoff data to account for the fact that the river network, which is at 0.5° spatial resolution, may not perfectly overlap with the GRDC river catchment boundaries (Table 1).

Both CHMs and GHMs simulate the full hydrological cycle with predominantly daily precipitation and temperature as input data. All the GHMs simulated hydrological processes at a spatial resolution of 0.5° x 0.5° across the global land surface. In contrast, CHMs operated using various approaches; three CHMs run on a grid (mHM, VIC, WaterGAP3), four by splitting the catchment into sub-catchments and smaller hydrological response units (HBV, HYPE, SWAT, SWIM) and one by considering the whole catchment as a single entity (HYMOD). The GHMs were not calibrated to catchment specific conditions, except WaterGAP2 (which was calibrated against long-term average annual streamflow for 1319 gauges worldwide) while the CHMs were calibrated and the performance of the calibration evaluated in a separate validation period using the WATCH ERA-40 climate forcing data (Huang et al., 2017).

In addition to the individual model results, we calculated the corresponding ensemble mean (using monthly runoff values) for the GHMs (denoted Ens-GHM) and CHMs (Ens-CHM) for every catchment, and included them in the analysis (**SM 1**). All the calculations of runoff-deficits and SRI for Ens-GHM and Ens-CHM are based on ensemble monthly runoff values. For the purposes of this study, all the models were treated as independent even though some of the models employ similar model parameterisations for some hydrological processes. No hydrological model was excluded or weighted based on their ability to simulate runoff.

## 3. Method

### 3.1. Runoff-deficit

We calculated runoff-deficit values based on a variable threshold and compared the simulated runoff-deficit with observed values. The threshold was defined as the 80th exceedance percentile runoff for each month separately, a value between the 70th and 95th percentile, commonly used for perennial rivers (Hisdal et al., 2001; Andreadis et al., 2005; Fleig et al., 2006; Tallaksen et al., 2009; van Loon, 2015). A positive runoff deficit indicates dry conditions i.e. when the runoff falls below the threshold value.

### 3.2. Standardised runoff index (SRI)

Several standardised indices have been developed for identifying hydrological droughts, such as the Palmer Hydrological Drought Index (PDHI; Jacobi et al., 2013), Water Supply Index (WSI; Garen, 1993) and

**Table 1**
Catchments, gauging station, upstream area of gauging station, and the GHMs and CHMs that comprise ensemble (marked with x).

| Catchment | | Upper Amazon | Lena | Upper Mississippi | Upper Niger | Rhine | Tagus | Upper Yangtze | Upper Yellow |
|---|---|---|---|---|---|---|---|---|---|
| **GRDC number** | | 3,623,100 | 2,903,430 | 4,119,800 | 1,134,100 | 6,435,060 | 6,113,050 | — | — |
| **Gauging station** | | Sao Paulo de Olivenca | Stolb | Alton | Koulikoro | Lobith | Almourol | Cuntan | Tangnaihai |
| **Upstream drainage area (km$^2$) - GRDC** | | 990,781 | 2460,000 | 444,185 | 120,000 | 160,800 | 67,490 | 804,859 | 121,000 |
| **Upstream drainage area (km$^2$) – DDM30** | | 994,469 | 2456,513 | 448,575 | 121,058 | 162,092 | 71,007 | 851,303 | 117,543 |
| **Difference between GRDC and DDM30 areas (%)** | | −0.4 | 0.1 | −1.0 | −0.9 | −0.8 | −5.2 | −5.8 | 2.9 |
| **GHM ensemble (Ens-GHM)** | CLM (Oleson et al., 2010) | x | x | x | x | x | x | x | x |
| | DBH (Tang et al., 2007) | x | x | x | x | x | x | x | x |
| | H08 (Hanasaki et al., 2008) | x | x | x | x | x | x | x | x |
| | MATSIRO (Pokhrel et al., 2015) | x | x | x | x | x | x | x | x |
| | MPI-HM (Hagemann and Dümenil, 1998) | x | x | x | x | x | x | x | x |
| | PCR-GLOBWB (Wada et al., 2014) | x | x | x | x | x | x | x | x |
| | WaterGAP2 (Muller Schmied et al., 2016) | x | x | x | x | x | x | x | x |
| | LPJmL (Bondeau et al., 2007) | x | x | x | x | x | x | x | x |
| **Number of GHM simulations** | | **8** | **8** | **8** | **8** | **8** | **8** | **8** | **8** |
| **CHM ensemble (Ens-CHM)** | ECOMAG (Motovilov et al., 1999) | | x | | | | | | |
| | HBV (Bergstrom and Forsman, 1973) | x | | x | x | x | x | x | x |
| | HYMOD (Boyle, 2001) | x | | x | x | x | | | x |
| | HYPE (Lindström et al., 2010) | | x | | | x | x | | |
| | mHM (Samaniego et al., 2010) | x | | x | x | x | | | x |
| | SWAT (Arnold et al., 1993) | x | | x | x | x | | x | |
| | SWIM (Krysanova et al., 1998) | x | x | x | x | x | | x | x |
| | VIC (Liang et al., 1994) | x | x | x | x | x | x | x | x |
| | WaterGAP3 (Verzano 2009) | x | x | x | x | x | x | | x |
| **Number of CHM simulations** | | **7** | **5** | **7** | **7** | **8** | **4** | **4** | **6** |

Standardised Streamflow/Runoff Index (SSI/SRI; Vicente-Serrano et al., 2012). In comparison with PHDI and WSI, SSI is more commonly accepted because it is simple to calculate, can be used on various time scales and requires fewer inputs. SSI is extensively used in many studies (Shukla and Wood, 2008; Vicente-Serrano et al., 2012; Wu et al., 2018; Liu et al., 2019). Calculation of SSI/SRI is similar to that of Standardised Precipitation Index (SPI) proposed by McKee et al. (1993b) but considering runoff instead of precipitation. The SRI values are determined based on long-term runoff records (preferably >30 years) by aggregating the monthly runoff over an accumulation period (1, 3, 6, 12, or 24 months). The new series formed after accumulation (1 month for this study) is then fitted to a probability distribution that subsequently is transformed to a normal distribution such that the mean SRI is zero. We tested three different probability distributions, Gamma, Normal and Weibull, to see which had the best fit to observed runoff. The Gamma distribution was the best fit for all eight catchments (**SM 2**) and was used for all the SRI calculations.

The "SPEI" package in R (Beguería and Vicente-Serrano, 2017) was used for the calculations of SRI. "SPEI" i.e. Standardized Precipitation-Evapotranspiration Index package, facilitates computation of SPI and other variants of SPI (SRI in our study) by providing defined functions that can be used directly in the R working environment. Any positive SRI values indicate runoff values greater than the mean monthly runoff and vice versa. SRI values less than −1 were considered to indicate drought conditions. These drought conditions were categorised based on SRI values into three drought categories, *moderate, severe* and *extreme* droughts. SRI values from −1 to −1.49 indicate *moderate* drought, from −1.5 to −1.99 *severe* drought, and all the SRI values less than −2 indicate *extreme* drought (McKee et al., 1993a; Lloyd-Hughes and Saunders, 2002; Mishra et al., 2007; Bloomfield and Marchant, 2013; Wasko et al., 2021).

### 3.3. Drought characteristics and performance evaluation

SRI drought events (drought events based on SRI values) were defined as temporally continuous SRI values <−1 lasting until the SRI is above −1 again. Runoff-deficit drought events (drought events based on runoff-deficit values) were defined as temporally continuous positive runoff-deficit values lasting until it is zero or less.

We calculated three drought characteristics for both runoff-deficit and SRI drought events from the observed and simulated data: drought *intensity* (or *severity* in case of runoff-deficit drought events), drought *duration* and *frequency* of drought events. Run theory (Yevjevich and Ica Yevjevich, 1967) was used for the extraction of drought characteristics from the SRI and runoff deficit time series. The Run theory method is based on a threshold level approach which considers the statistical properties of runs (here, length of negative or positive deviations in SRI or runoff-deficit values respectively) above or below a given threshold level. The drought *duration* (in months) was taken as the period in months for which the SRI remained less than a threshold value of −1 or runoff-deficit was greater than a threshold value of 0. The minimum SRI value in the drought period was used as the drought *intensity* for SRI drought events, while the collective sum of all positive runoff-deficits over the drought period was taken as drought *severity* for runoff-deficit drought events.

We used the coefficient of determination i.e. the square of Pearson product moment correlation coefficient ($R^2$), and the Nash-Sutcliffe Efficiency coefficient (NSE; Nash and Sutcliffe, 1970) to evaluate the goodness of fit between observed and simulated runoff-deficits and SRI values, for each CHM, GHM, and the Ens-CHM and Ens-GHM. $R^2$ indicated the strength of relationship between the simulated and observed runoff, and NSE indicated the models' efficiency (range -∞–1). NSE values approaching 1 indicate a perfect match of simulated and observed runoff values while negative values show that the observed mean is a better predictor than the model.

## 4. Results

### 4.1. Comparison of observed and simulated runoff deficits

Table 2 shows the $R^2$ and NSE values for all the CHMs and GHMs, as well as the Ens-CHM and Ens-GHM. An $R^2$ value >0.7 was interpreted as satisfactory model performance, as was an NSE >0.70, following Moriasi et al. (2007), (2015). For all eight catchments, the CHMs were better than GHMs for the estimation of monthly runoff-deficits because the $R^2$ values for Ens-CHM were higher than those for Ens-GHM. Both ensembles performed well in four catchments according to $R^2$ because the Ens-CHM and Ens-GHM $R^2$ values were greater than 0.7 (Moriasi et al., 2007, 2015): the Upper Mississippi, Rhine, Tagus and Upper Yellow. NSE values for Ens-CHM were greater than 0.7 for five catchments, and closer to 1 for the Rhine and Tagus catchments, while for Ens-GHM only two catchments (Upper Mississippi and Tagus) had NSE values greater than 0.7.

All individual CHMs that simulated runoff for the Rhine, Tagus and Upper Yangtze catchments had $R^2$ and NSE values >0.70 except WaterGAP3 for the Tagus catchment achieving 0.77 ($R^2$) and 0.60 (NSE). Both the $R^2$ and NSE values were <0.70 for all other catchments, the only exception being mHM for the Upper Mississippi and Upper Yellow catchments.

None of the individual CHMs or GHMs had consistently satisfactory $R^2$ and NSE values across all catchments. The Rhine and Tagus catchments had individual GHMs with $R^2$ values >0.70 for over half the models but with unsatisfactory NSE values. In addition, the Upper Amazon, Lena and Upper Niger catchments had no individual GHMs with $R^2$ or NSE values >0.70. Individual GHMs showed good performance in the Tagus catchment, where both indicators were >0.70.

### 4.2. Comparison of observed and simulated SRI

$R^2$ and NSE values for the SRI series are displayed in Table 3. Both $R^2$ and NSE vary widely across all catchments for both CHMs and GHMs. The $R^2$ values of the ensemble series varied between 0.18 (Ens-GHM, Upper Niger) and 0.94 (Ens-CHM, Rhine), while the NSE ranged from −0.14 (Ens-GHM, Upper Niger) to 0.95 (Ens-CHM, Rhine). For only three catchments, Ens-CHM had $R^2$ and NSE values >0.70 and Ens-GHM for only two.

The catchments where most of the models (both CHMs and GHMs) performed well were the Upper Mississippi and Rhine, both with $R^2$ and NSE values for Ens-CHM and Ens-GHM >0.85. For both these catchments, individual models always achieved $R^2$ >0.50 and NSE values >0.40. No individual CHM or GHM achieved $R^2$ and NSE values >0.7 for the other six catchments, with the exception of HBV for the Tagus catchment ($R^2 = 0.75$ and NSE = 0.74). The catchment where all the models performed the worst was the Upper Niger, where no individual GHM or CHM (except SWIM) achieved an $R^2$ value >0.35 and NSE values were always close to, or below zero.

### 4.3. Evaluation of the frequency of drought events

Table 4 presents the number of runoff deficit and SRI drought events, identified from the observed and simulated runoff (Ens-CHM and Ens-GHM) for all eight catchments. The total number of individual drought events identified by Ens-GHM and Ens-CHM were comparable in most catchments. For all eight catchments, 136 individual SRI drought events were identified from the observed data, and the ensemble models successfully simulated 133 and 123 drought events for Ens-GHM and Ens-CHM, respectively. Similarly, Ens-GHM and Ens-CHM successfully simulated 233 and 220 individual drought events respectively from 244 individual runoff-deficit drought events, identified from observed records.

Ens-CHM and Ens-GHM performed similarly in their estimation of the number of drought events for most catchments, with five catchments

**Table 2**
$R^2$ and NSE (in parentheses) values for simulated (individual GHMs and CHMs including both ensembles) versus observed monthly runoff-deficit across all eight catchments. Cells with bold text denotes where both $R^2$ and NSE values are above 0.7. Cells marked with x denote that the particular model was not run for the specific catchment.

| | | Upper Amazon | Lena | Upper Mississippi | Upper Niger | Rhine | Tagus | Upper Yangtze | Upper Yellow |
|---|---|---|---|---|---|---|---|---|---|
| **GHMs** | **CLM** | 0.36 (−0.13) | 0.05 (−0.04) | 0.58 (0.09) | 0.49 (−1.33) | 0.47 (−1.13) | 0.59 (0.53) | 0.19 (0.08) | 0.42 (0.40) |
| | **DBH** | 0.38 (0.32) | 0.07 (−0.16) | 0.69 (0.03) | 0.41 (−2.37) | 0.73 (0.01) | 0.82 (−0.08) | 0.54 (0.5) | 0.64 (0.62) |
| | **H08** | 0.37 (0.20) | 0.01 (−0.17) | 0.60 (−0.01) | 0.40 (−3.67) | 0.64 (−0.47) | 0.85 (−1.26) | 0.44 (0.38) | 0.45 (0.39) |
| | **MATSIRO** | 0.30 (0.24) | 0.31 (0.26) | 0.69 (0.49) | 0.63 (−0.87) | 0.55 (0.51) | **0.85** **(0.79)** | 0.58 (0.56) | 0.58 (0.35) |
| | **MPI-HM** | 0.36 (0.32) | 0.10 (−0.05) | 0.74 (0.50) | 0.56 (−0.08) | 0.67 (0.36) | 0.76 (0.55) | 0.39 (0.37) | 0.50 (0.42) |
| | **PCR-GLOBWB** | 0.38 (0.34) | 0.41 (0.36) | 0.74 (0.67) | 0.31 (0.15) | 0.81 (0.67) | **0.88** **(0.87)** | **0.77** **(0.70)** | **0.70** **(0.70)** |
| | **WaterGAP2** | 0.46 (0.39) | 0.39 (0.35) | **0.75** **(0.70)** | 0.44 (0.42) | **0.86** **(0.86)** | **0.86** **(0.85)** | 0.62 (0.61) | 0.65 (0.56) |
| | **LPJmL** | 0.37 (0.23) | 0.23 (−1.73) | 0.45 (−0.33) | 0.41 (−4.05) | 0.84 (0.12) | 0.86 (−0.89) | 0.66 (0.66) | 0.54 (0.53) |
| **CHMs** | **ECOMAG** | x | 0.67 (0.58) | x | x | x | x | x | x |
| | **HBV** | 0.52 (0.51) | x | 0.65 (0.63) | 0.55 (0.41) | **0.86** **(0.86)** | **0.94** **(0.88)** | **0.80** **(0.79)** | 0.69 (0.67) |
| | **HYMOD** | 0.43 (0.39) | x | 0.54 (0.48) | 0.56 (0.46) | **0.86** **(0.84)** | x | x | 0.61 (0.59) |
| | **HYPE** | x | 0.72 (0.62) | x | x | **0.91** **(0.88)** | **0.94** **(0.92)** | x | x |
| | **mHM** | 0.50 (0.48) | x | **0.85** **(0.83)** | 0.54 (0.51) | **0.90** **(0.90)** | x | x | **0.76** **(0.73)** |
| | **SWAT** | 0.49 (0.43) | x | 0.77 (0.68) | 0.49 (0.47) | **0.94** **(0.92)** | x | **0.79** **(0.78)** | x |
| | **SWIM** | 0.55 (0.51) | 0.58 (0.47) | 0.70 (0.65) | 0.67 (0.59) | **0.89** **(0.89)** | x | **0.79** **(0.72)** | 0.71 (0.68) |
| | **VIC** | 0.49 (0.45) | 0.53 (0.52) | 0.65 (0.56) | 0.59 (0.49) | **0.91** **(0.90)** | **0.91** **(0.87)** | **0.83** **(0.82)** | 0.59 (0.46) |
| | **WaterGAP3** | 0.17 (−0.35) | 0.37 (0.31) | 0.67 (0.65) | 0.35 (0.22) | **0.85** **(0.85)** | 0.77 (0.60) | x | 0.59 (0.56) |
| **Ens-GHM** | | 0.49 (0.47) | 0.37 (0.30) | **0.82** **(0.80)** | 0.53 (−0.35) | 0.82 (0.65) | **0.9** **(0.71)** | 0.63 (0.62) | 0.72 (0.62) |
| **Ens-CHM** | | 0.50 (0.48) | **0.73** **(0.73)** | **0.83** **(0.82)** | 0.61 (0.59) | **0.96** **(0.95)** | **0.96** **(0.95)** | **0.85** **(0.85)** | 0.73 (0.66) |

differing by 6 or less runoff-deficit drought events and only by 2 or less SRI drought events. For half of the catchments, the number of SRI drought events estimated by the Ens-GHM was closer to the observed data than the Ens-CHM simulations, while for observed runoff-deficit drought events estimation by the Ens-CHM was better in the majority of catchments.

Despite the comparable estimates of the total number of SRI and runoff-deficit drought events across most catchments, both Ens-CHM and Ens-GHM struggled in estimating the *frequency* of droughts within SRI drought categories. Both, Ens-CHM and Ens-GHM underestimated the number of *extreme* SRI droughts ($\leq -2$) in the Upper Amazon, Upper Niger and Tagus catchments. While in the Upper Yellow catchment, Ens-CHM and Ens-GHM both overestimated the number of *extreme* SRI droughts, showing 3 and 4 *extreme* events respectively when no such SRI drought events were identified in the observed data.

### 4.4. Evaluation of drought intensity

Fig. 2 and **SM 3** (supplementary materials) show *extreme* SRI and *severe* runoff deficit drought events respectively, and the performance of individual models in representing these events with regards to drought *intensity* or *severity* along with drought *duration*. Substantial variation was seen in the ability of individual CHMs and GHMs to simulate drought *intensity* or *severity*. The over- or under-estimation of drought *intensity* in Fig. 2 (drought *severity* in **SM 3**) shown as colour coded cells from shades of green to brown is based on the difference in estimated and observed drought *intensity* or *severity*. There are marked differences across catchments in whether the models over- or under-estimated the *intensity* (or *severity*) of observed droughts, and in the ability to simulate the very occurrence of an observed drought itself. We observed cases where ensemble models simulated a drought that never occurred in the observed record, and cases where a model fails to simulate a drought that occurred in the observed record, in most catchments for *extreme* SRI and *severe* runoff-deficit drought events.

Out of the 27 observed *extreme* SRI drought events across all catchments, Ens-CHM and Ens-GHM failed to identify 1 and 3 drought events, respectively. In total 14 *extreme* SRI drought events were identified by the models which were not observed (shaded grey under 'Observed' in Fig. 2), and 5 of these events were identified by both Ens-CHM and Ens-GHM. For the Upper Amazon, individual CHMs and GHMs performed similarly, with most models under-estimating drought *intensity* for 6 of the observed *extreme* SRI drought events. For one event in the Upper Amazon, an observed SRI drought event went undetected by 10 out of 17 models. All the individual models did not accurately simulate hydrological conditions in the Upper Yellow catchment, with 4 *extreme* drought events simulated that were not observed in reality. Model performance is more nuanced in the other catchments, with some observed droughts within a catchment under-estimated and other events over-estimated (Upper Niger, Rhine, and Upper Yangtze).

Ens-CHM and Ens-GHM each failed to identify 12 drought events across all eight catchments from 70 observed *severe* runoff-deficit droughts (**SM 3**, supplementary material). In addition to 70 observed

**Table 3**
$R^2$ and NSE (in parentheses) values for simulated (individual GHMs and CHMs including both ensembles) versus observed SRI across all eight catchments. Cells with bold text denotes where both $R^2$ and NSE values are >0.70 and marked with x denotes that the particular model was not run for the specific catchment.

| | | Upper Amazon | Lena | Upper Mississippi | Upper Niger | Rhine | Tagus | Upper Yangtze | Upper Yellow |
|---|---|---|---|---|---|---|---|---|---|
| **GHMs** | **CLM** | 0.42 (0.30) | 0.17 (−0.29) | 0.64 (0.6) | 0.15 (−0.20) | 0.50 (0.43) | 0.38 (0.24) | 0.37 (0.22) | 0.4 (0.27) |
| | **DBH** | 0.44 (0.34) | 0.12 (−0.43) | 0.67 (0.64) | 0.11 (−0.32) | **0.76 (0.75)** | 0.32 (0.15) | 0.45 (0.35) | 0.27 (0.05) |
| | **H08** | 0.44 (0.34) | 0.14 (−0.35) | 0.66 (0.63) | 0.11 (−0.32) | 0.63 (0.59) | 0.31 (0.13) | 0.34 (0.18) | 0.23 (−0.03) |
| | **MATSIRO** | 0.32 (0.14) | 0.30 (0.01) | **0.71 (0.7)** | 0.24 (−0.01) | 0.55 (0.49) | 0.42 (0.31) | 0.44 (0.33) | 0.34 (0.18) |
| | **MPI-HM** | 0.39 (0.26) | 0.19 (−0.24) | **0.75 (0.74)** | 0.27 (0.05) | **0.71 (0.7)** | 0.65 (0.62) | 0.39 (0.25) | 0.43 (0.32) |
| | **PCR-GLOBWB** | 0.48 (0.39) | 0.33 (0.06) | **0.75 (0.74)** | 0.01 (−0.75) | **0.80 (0.79)** | 0.56 (0.50) | 0.52 (0.45) | 0.45 (0.35) |
| | **WaterGAP2** | 0.53 (0.47) | 0.28 (−0.02) | **0.78 (0.77)** | 0.24 (−0.01) | **0.87 (0.88)** | 0.65 (0.62) | 0.54 (0.48) | 0.5 (0.43) |
| | **LPJmL** | 0.50 (0.43) | 0.12 (−0.44) | 0.53 (0.46) | 0.12 (−0.28) | **0.80 (0.79)** | 0.21 (−0.07) | 0.47 (0.37) | 0.29 (0.09) |
| **CHMs** | **ECOMAG** | x | 0.45 (0.28) | x | x | x | x | x | x |
| | **HBV** | 0.55 (0.49) | x | 0.68 (0.65) | 0.25 (0.01) | **0.88 (0.88)** | **0.75 (0.74)** | 0.59 (0.54) | 0.25 (0.01) |
| | **HYMOD** | 0.50 (0.42) | x | 0.59 (0.54) | 0.28 (0.07) | **0.78 (0.77)** | x | x | 0.41 (0.29) |
| | **HYPE** | x | 0.44 (0.27) | x | x | **0.89 (0.89)** | 0.64 (0.61) | x | x |
| | **mHM** | 0.56 (0.50) | x | **0.87 (0.87)** | 0.27 (0.05) | **0.88 (0.89)** | x | x | 0.54 (0.47) |
| | **SWAT** | 0.54 (0.47) | x | **0.72 (0.70)** | 0.31 (0.13) | **0.93 (0.93)** | x | 0.56 (0.5) | x |
| | **SWIM** | 0.62 (0.58) | x | **0.73 (0.72)** | 0.49 (0.4) | **0.88 (0.88)** | x | 0.51 (0.43) | 0.51 (0.44) |
| | **VIC** | 0.52 (0.45) | 0.30 (0.01) | 0.64 (0.60) | 0.27 (0.05) | **0.91 (0.91)** | 0.34 (0.18) | 0.55 (0.49) | 0.32 (0.13) |
| | **WaterGAP3** | 0.23 (−0.02) | 0.29 (0) | **0.77 (0.76)** | 0.16 (−0.17) | **0.86 (0.86)** | 0.57 (0.51) | x | 0.49 (0.41) |
| **Ens-GHM** | | 0.56 (0.50) | 0.35 (0.10) | **0.86 (0.86)** | 0.18 (−0.14) | **0.85 (0.85)** | 0.47 (0.38) | 0.57 (0.52) | 0.56 (0.46) |
| **Ens-CHM** | | 0.56 (0.50) | 0.46 (0.30) | **0.88 (0.88)** | 0.33 (0.15) | **0.94 (0.95)** | **0.73 (0.71)** | 0.65 (0.62) | 0.53 (0.46) |

*severe* runoff-deficit drought events, 19 additional runoff-deficit drought events were identified by the models which were not observed and 9 of these additional events were identified as *severe* runoff-deficit drought events by both Ens-GHM and Ens-CHM. Performance of CHMs and GHMs was analogous in estimating drought *severity* across all catchments, however there were fewer cases in CHMs (than GHMs) where observed *severe* runoff-deficit drought events were identified only by one individual model. Two runoff-deficit drought events went undetected by all individual models in the Upper Yellow catchment.

### 4.5. Evaluation of drought duration

For all the observed runoff-deficit and SRI drought events (independently under each drought category i.e. *moderate, severe* and *extreme* for SRI droughts events), we calculated the absolute error in drought *duration* for each model. The mean of all the absolute errors (under each drought category for SRI droughts events) for a model yielded the mean absolute error (MAE) for that model. Figs. 3 and 4 shows the MAE values for SRI (under the three drought categories) and runoff-deficit drought events respectively from the Ens-CHM and Ens-GHM, for every catchment along with their respective mean observed drought *duration*. **SM 4** (supplementary materials) presents MAE values for the individual models averaged across all catchments for both SRI and runoff-deficit drought events.

For SRI droughts, the largest total MAE (i.e. the MAE summed across the three drought categories) was seen for Upper Niger for both ensembles, followed by Upper Mississippi for Ens-CHM and Upper Amazon for Ens-GHM (Fig. 3). Both ensembles display MAE's smaller than all individual models (**SM 4**), for *extreme* SRI droughts across all three drought categories. The models simulate drought *duration* for *extreme* SRI droughts better than for lower *intensity* SRI droughts (Fig. 3). However for SRI drought events, overall, both the CHMs and GHMs struggle to accurately model drought duration, with MAE consistently >1 month (**SM 4**). For runoff-deficit droughts, the MAE was similar between Ens-CHM and Ens-GHM, except for the Lena with a difference of 1 month between the two ensembles (Fig. 4). MAEs were generally lower for runoff-deficit droughts than for SRI droughts.

## 5. Discussion

The aim of this study was to assess the performance of CHMs and GHMs in simulating observed drought events using runoff-deficit and SRI as indicators of hydrological drought. No two drought events are the same and, as such, drought events cannot be judged based on a single characteristic. Here we used three characteristics; the *intensity* or *severity, duration* and *frequency* of drought events. We used $R^2$ and NSE to judge the ability of the models and their ensemble means to replicate observed monthly runoff-deficits and SRI.

### 5.1. Comparison of CHM and GHM performance

Different thresholds defining runoff-deficit droughts for each catchment meant there were not standard drought classes, which limited the distinctive analysis of drought occurrences. *Frequency* of runoff-deficit

**Table 4**
Number of runoff-deficit and SRI drought events identified from observed and simulated runoff (Ens-GHM and Ens-CHM) for all eight catchments, SRI drought events are classified into *moderate, severe* and *extreme* drought events based on SRI values.

| | | SRI drought events | | | | Runoff-deficit drought events |
|---|---|---|---|---|---|---|
| Catchments(↓) | Drought category (→) | Moderate(−1 to −1.49) | Severe(−1.5 to −1.99) | Extreme(−2 or Below) | Total individualdrought events(−1 or Below) | Total individualdrought events(with positive drought deficit) |
| **Upper Amazon** | **Ens-GHM** | 3 | 5 | 4 | 12 | 24 |
| | **Ens-CHM** | 6 | 5 | 4 | 15 | 29 |
| | **Obs** | **6** | **4** | **6** | **16** | **27** |
| **Lena** | **Ens-GHM** | 7 | 4 | 3 | 14 | 19 |
| | **Ens-CHM** | 7 | 5 | 3 | 15 | 25 |
| | **Obs** | **11** | **4** | **3** | **18** | **36** |
| **Upper Mississippi** | **Ens-GHM** | 8 | 4 | 2 | 14 | 28 |
| | **Ens-CHM** | 6 | 3 | 3 | 12 | 23 |
| | **Obs** | **9** | **3** | **2** | **14** | **27** |
| **Upper Niger** | **Ens-GHM** | 10 | 7 | 2 | 19 | 40 |
| | **Ens-CHM** | 9 | 4 | 2 | 15 | 31 |
| | **Obs** | **11** | **6** | **5** | **22** | **30** |
| **Rhine** | **Ens-GHM** | 7 | 7 | 4 | 18 | 32 |
| | **Ens-CHM** | 8 | 6 | 3 | 17 | 30 |
| | **Obs** | **10** | **6** | **3** | **19** | **31** |
| **Tagus** | **Ens-GHM** | 11 | 6 | 0 | 17 | 38 |
| | **Ens-CHM** | 2 | 5 | 1 | 8 | 16 |
| | **Obs** | **5** | **3** | **2** | **10** | **27** |
| **Upper Yangtze** | **Ens-GHM** | 12 | 5 | 5 | 22 | 26 |
| | **Ens-CHM** | 12 | 5 | 7 | 24 | 38 |
| | **Obs** | **10** | **5** | **6** | **21** | **34** |
| **Upper Yellow** | **Ens-GHM** | 6 | 7 | 4 | 17 | 26 |
| | **Ens-CHM** | 9 | 5 | 3 | 17 | 28 |
| | **Obs** | **7** | **9** | **0** | **16** | **32** |
| **Total individual drought events across all catchments** | **Ens-GHM** | 64 | 45 | 24 | 133 | 233 |
| | **Ens-CHM** | 59 | 38 | 26 | 123 | 220 |
| | **Obs** | **69** | **40** | **27** | **136** | **244** |

drought events was higher than that of SRI drought events (Table 4), which resulted in almost consistent mean observed drought *duration* and very less variation was seen in the MAE between Ens-CHM and Ens-GHM across all catchments.

There was a marked difference in performance of CHMs and GHMs when estimating drought *intensity* and *duration* with the SRI. This is owing to the data used for identifying simulated drought events, which here is simulated monthly runoff data. The accuracy of the SRI computation is directly proportional to the quality of data used for its calculation (Hayes et al., 1999). Huang et al. (2017) reported that CHMs accurately reproduced monthly runoff, seasonal dynamics, moderate or high-flows but simulations of low-flows were problematic in most catchments. Zaherpour et al. (2018) found that the majority of GHMs overestimated low-flows considerably more than they overestimated high-flows and that GHMs overestimated minimum flow return periods. The majority of the GHMs showed a tendency for overestimating monthly runoff with a wider magnitude range (Veldkamp et al., 2018). Previous studies highlight that this wider spread around ensembles in every catchment is due to the structure of GHMs (Haddeland et al., 2011; Gudmundsson et al., 2012). Physical processes such as transmission losses, having less presence in the GHMs is one main reason for some of the differences between simulated and observed runoff (Gosling and Arnell, 2011). In addition, evapotranspiration simulation has been reported to vary widely among the GHMs (Wartenburger et al., 2018).

### 5.2. Performance of the ensemble means

Although the $R^2$ and NSE values for the SRI series are less than satisfactory for many individual GHMs and CHMs across catchments, the ensemble models were better at estimating drought *frequency* for most of the catchments. Moreover, both the ensembles display, overall, a better performance compared to the individual models that make up each respective ensemble, and showed comparable outputs (runoff-deficit and SRI values) despite the GHMs having a wider spread across the ensemble. Although the GHMs used the same climate forcing, they used different formulations to compute potential evapotranspiration (Telteu et al., 2021), which contributes to differences in simulated runoff between the GHMs (Beck et al., 2017). For monthly runoff-deficits, the Ens-GHM outperformed individual GHMs because the individual models over- or under-estimated low-flow conditions. The computation of an ensemble mean for the GHMs essentially balances out the over- and under-estimation of runoff-deficits by individual models. However the use of a large number of models does not necessarily ensure better performance of the ensemble mean. For the Upper Amazon, Upper Mississippi and Rhine where the number of models used for the ensemble calculation was higher, the performance of the Ens-CHM was better at estimating drought *frequency* but showed higher MAE for drought *duration* with exception of the Rhine.

### 5.3. Implications for model calibration

The performance of WaterGAP2 for runoff-deficit estimation was similar to the other GHMs. Despite WaterGAP2 being the only GHM to be calibrated with long-term annual river discharge, it did not perform noticeably better than other GHMs, and independent of calibration, the identification of hydrological extremes was not satisfactory across all models. This suggests that it is important to calibrate models with high temporal resolution data that allows extreme events to be accounted for in the calibration process. The use of multiple criteria during model calibration, specifically for low- or no-flow could be helpful (Krysanova et al., 2018).

### 5.4. Representation of processes and quality of observed data

Our results indicate relatively better performances of GHMs for runoff-deficits and SRI patterns in the Upper Mississippi and Rhine compared to other catchments. Furthermore, none of the GHMs missed any observed SRI drought events, while only few GHMs failed to simulate observed runoff-deficit drought events. However, tendency of GHMs (across all catchments) towards simulating drought events that were not

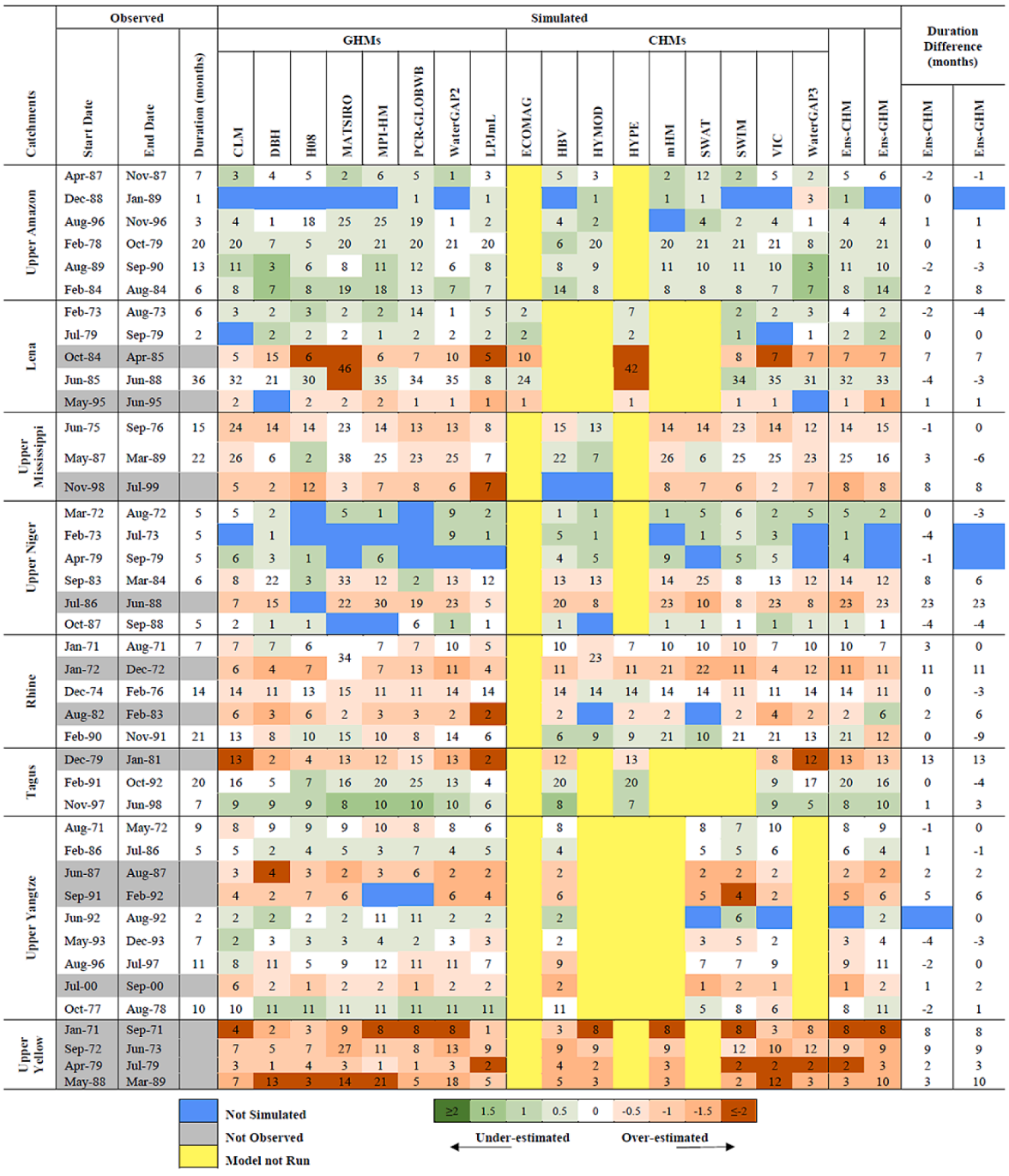



**Fig. 2.** Observed and simulated extreme SRI drought events (SRI $\leq -2$). Displayed is each observed drought's start and end date, with the numbers indicating the observed and simulated drought duration in months. Drought events are colour coded based on the difference in simulated and observed drought intensity (from 2 to $-2$), marked from shades of green to brown (under- and over-estimated respectively). Blue cells are observed droughts not present in simulated records. Grey cells are simulated droughts (either by Ens-CHM or Ens-GHM or both) not present in the observed record. Yellow cells denote that the particular model was not run for the specific catchment.

observed, can be attributed to a dry bias introduced by the choice of potential evapotranspiration formulation for individual models. For example, PCR-GLOBWB consistently appeared near the dry end of Ens-GHM, perhaps because it includes a temperature based evaporation formulation (Hamon) that has been shown to induce a large bias when applied outside its calibration range (Milly and Dunne, 2017). In general, for GHMs it is difficult to estimate a drought event at the right time because multiple errors propagate from the inputs (meteorological parameters) and some GHMs struggle to capture the magnitude and timing of processes like abstraction losses and snowmelt accurately, which is likely to have an impact on drought timing (initiation and *duration* of drought events).

For CHMs, large biases have been reported in simulating low-flow conditions across majority of the catchments we studied, especially the Upper Yangtze (Huang et al., 2017). Inaccuracies of low-flow observations may be a factor affecting the estimation of observed drought conditions, as might river ice in some catchments, while the inability of individual CHMs to replicate low-flow or no-flow may be due to the choice of objective functions for calibration of the CHMs (Huang et al., 2017). Similarities in performance of both sets of individual models in

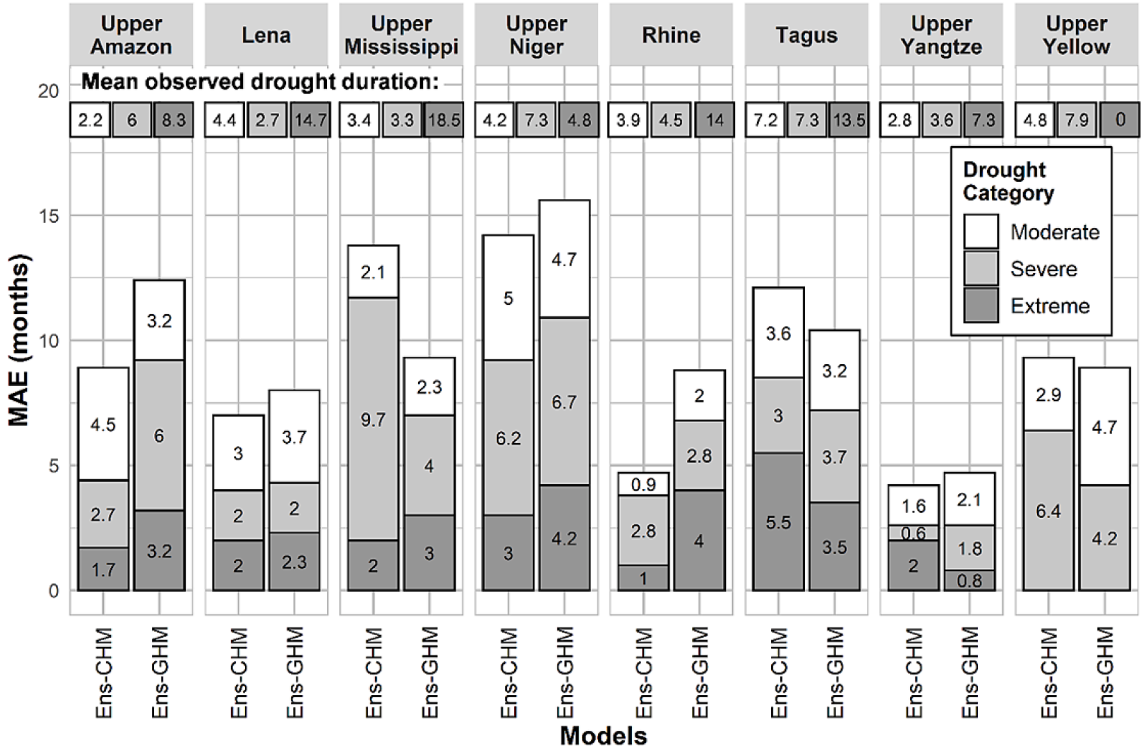



**Fig. 3.** MAE for SRI drought *duration* under each drought category for Ens-GHM and Ens-CHM, for all eight catchments, along with respective mean observed drought *duration*.

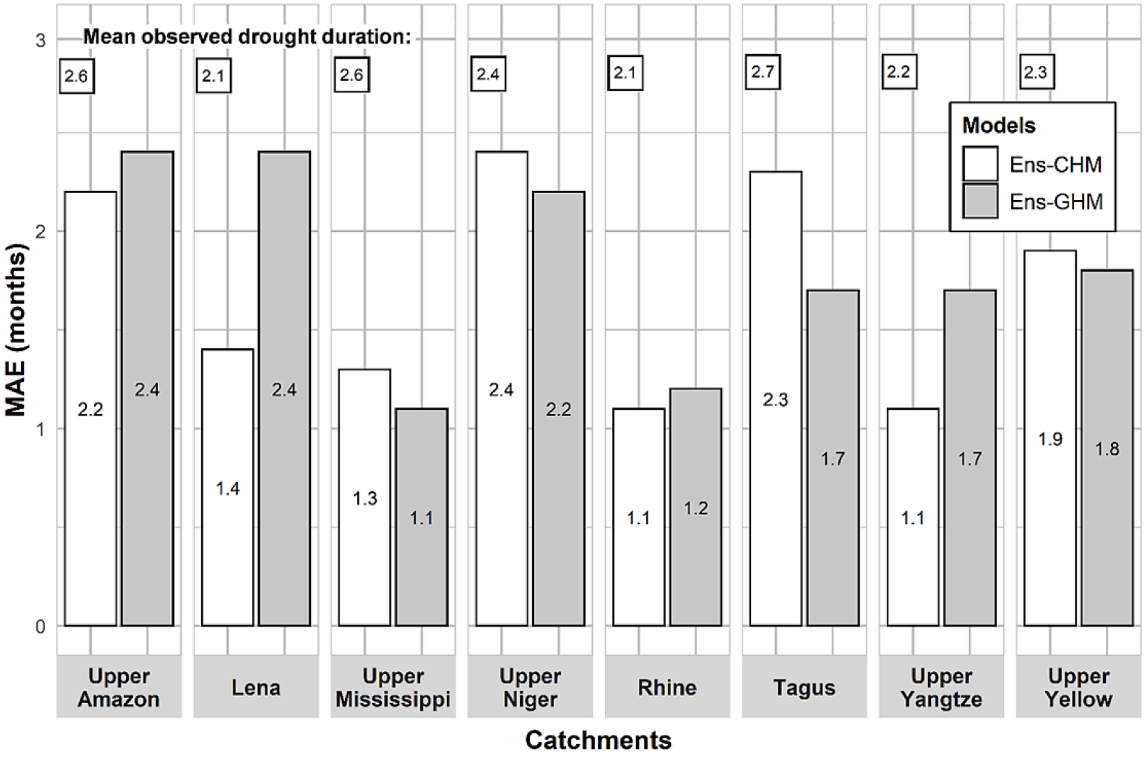



**Fig. 4.** MAE for runoff-deficit drought *duration* for Ens-GHM and Ens-CHM, for all eight catchments, along with respective mean observed drought *duration*.

the Upper Amazon, Lena, Tagus, and Upper Yangtze catchments for drought events in and SM 3 (supplementary materials) can likely be attributed towards the quality of the meteorological data used for simulations. Some studies have reported WATCH ERA-40 data to be unreliable due to inaccuracies in observed precipitation records, caused by fog/mist (Strauch et al., 2017). WaterGAP3 among all CHMs comparatively showed the weakest performance, which may be attributed to fewer parameters used for model calibration compared to other

CHMs.

## 6. Conclusion

Our study focused upon the effectiveness of catchment- and global-scale hydrological models to estimate drought conditions at the catchment scale, for 8 large catchments. We found comparably lower performance by most GHMs in simulating monthly runoff-deficits, while CHMs and GHMs were similar in estimating SRI. Both sets of models show limited ability to simulate the finer, more granular and detailed characteristics (*intensity* or *severity, duration* and *frequency*) of observed droughts but the ensembles performed better compared to the individual models that make up each respective ensemble. Whilst the Ens-CHM and Ens-GHM simulated drought *frequency* well for runoff-deficit and SRI drought events, marked differences were observed in ability to simulate the occurrence of observed and simulated drought events (not simulating several observed drought events and simulating drought events out of observed records). For both the ensembles, the error is, overall, smallest for *duration* of extreme SRI droughts across all three drought categories. However, it can also be concluded that both CHMs and GHMs struggled in accurately modelling drought *duration* for moderate and severe SRI drought categories. We believe that there is still room for improvement in runoff simulations to facilitate drought identification and accurate estimation of drought characteristics.

## CRediT authorship contribution statement

**Amit Kumar:** Investigation, Data curation, Writing – original draft. **Simon N. Gosling:** Methodology, Data curation, Writing – original draft, Writing – review & editing. **Matthew F. Johnson:** Methodology, Data curation, Writing – original draft, Writing – review & editing. **Matthew D. Jones:** Methodology, Writing – original draft, Writing – review & editing. **Jamal Zaherpour:** Data curation, Writing – original draft, Writing – review & editing. **Rohini Kumar:** Writing – original draft, Writing – review & editing. **Guoyong Leng:** Writing – original draft, Writing – review & editing. **Hannes Müller Schmied:** Writing – original draft, Writing – review & editing. **Jenny Kupzig:** Writing – original draft, Writing – review & editing. **Lutz Breuer:** Writing – original draft, Writing – review & editing. **Naota Hanasaki:** Writing – original draft, Writing – review & editing. **Qiuhong Tang:** Writing – original draft, Writing – review & editing. **Sebastian Ostberg:** Writing – original draft, Writing – review & editing. **Tobias Stacke:** Writing – original draft, Writing – review & editing. **Yadu Pokhrel:** Writing – original draft, Writing – review & editing. **Yoshihide Wada:** Writing – original draft, Writing – review & editing. **Yoshimitsu Masaki:** Writing – original draft, Writing – review & editing.

## Declaration of Competing Interest

The authors declare no conflict of interest.

*Funding*

The research was funded by the University of Nottingham Vice Chancellor's Scholarship for Research Excellence (International).

*Availability of data and material*

Output from the models used for this study are openly available from the Earth System Grid Federation (ESGF; https://esg.pik-potsdam.de/search/isimip) for the GHMs (Gosling et al., 2017) and the CHMs (Krysanova et al., 2017).

*Code availability*

Not applicable.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.advwatres.2022.104212.

## References

Andreadis, K.M., et al., 2005. Twentieth-century drought in the conterminous United States. J. Hydrometeorol. 6 (6), 985–1001.
Arnell, N.W., Gosling, S.N., 2016. The impacts of climate change on river flood risk at the global scale. Clim. Chang. https://doi.org/10.1007/s10584-014-1084-5 [Preprint].
Arnold, J.G., Allen, P.M., Bernhardt, G., 1993. A comprehensive surface-groundwater flow model. J. Hydrol. https://doi.org/10.1016/0022-1694(93)90004-S [Preprint].
Beck, H.E., et al., 2017. Global evaluation of runoff from 10 state-of-the-art hydrological models. Hydrol. Earth Syst. Sci. https://doi.org/10.5194/hess-21-2881-2017 [Preprint].
Beguería, S. and Vicente-Serrano, S.M. (2017) "Calculation of the standardised precipitation-evapotranspiration Index (SPEI)".
Bergstrom, S., Forsman, A., 1973. Development of a conceptual deterministic rainfall-runoff model. Nord. Hydrol. https://doi.org/10.2166/nh.1973.0012 [Preprint].
Bloomfield, J.P., Marchant, B.P., 2013. Analysis of groundwater drought building on the standardised precipitation index approach. Hydrol. Earth Syst. Sci. 17 (12) https://doi.org/10.5194/hess-17-4769-2013.
Bondeau, A., et al., 2007. Modelling the role of agriculture for the 20th century global terrestrial carbon balance,. Glob. Chang. Biol. https://doi.org/10.1111/j.1365-2486.2006.01305.x [Preprint].
Boyle, D., 2001. Multicriteria Calibration of Hydrologic Models. The University of Arizona.
Byun, H.R., Wilhite, D.A., 1999. Objective quantification of drought severity and duration. J. Clim. https://doi.org/10.1175/1520-0442(1999)012<2747:OQODSA>2.0.CO;2 [Preprint].
Döll, P., Lehner, B., 2002. Validation of a new global 30-min drainage direction map. J. Hydrol. 258 (1–4) https://doi.org/10.1016/S0022-1694(01)00565-0.
Elliott, J., et al., 2014. Constraints and potentials of future irrigation water availability on agricultural production under climate change. Proc. Natl. Acad. Sci. U. S. A. https://doi.org/10.1073/pnas.1222474110 [Preprint].
Emerton, R.E., et al., 2016. Continental and global scale flood forecasting systems. WIREs Water. https://doi.org/10.1002/wat2.1137 [Preprint].
Fleig, A.K., et al., 2006. A global evaluation of streamflow drought characteristics. Hydrol. Earth Syst. Sci. 10 (4) https://doi.org/10.5194/hess-10-535-2006.
Fleig, A.K., et al., 2011. Regional hydrological drought in north-western Europe: linking a new regional drought area index with weather types. Hydrol. Process. https://doi.org/10.1002/hyp.7644 [Preprint].
Garen, D.C., 1993. Revised surface-water supply index for Western United States. J. Water Resour. Plan. Manag. https://doi.org/10.1061/(asce)0733-9496(1993)119:4(437 [Preprint].
Gosling, S., Zaherpour, J., 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. Chang. https://doi.org/10.1007/s10584-016-1773-3 [Preprint].
Gosling, S., Müller Schmied, H., et al. (2017) ISIMIP2a simulation data from water (global) sector. doi:10.5880/PIK.2017.010.
Gosling, S.N., Arnell, N.W., 2011. Simulating current global river runoff with a global hydrological model: model revisions, validation, and sensitivity analysis. Hydrol. Process. https://doi.org/10.1002/hyp.7727 [Preprint].
Gudmundsson, L., et al., 2012. Comparing large-scale hydrological model simulations to observed runoff percentiles in Europe. J. Hydrometeorol. https://doi.org/10.1175/JHM-D-11-083.1 [Preprint].
Haddeland, I., et al., 2011. Multimodel estimate of the global terrestrial water balance: setup and first results. J. Hydrometeorol. https://doi.org/10.1175/2011JHM1324.1 [Preprint].
Hagemann, S., Dümenil, L., 1998. A parametrization of the lateral waterflow for the global scale. Clim. Dyn. https://doi.org/10.1007/s003820050205 [Preprint].
Hanasaki, N., et al., 2008. An integrated model for the assessment of global water resources - part 1: model description and input meteorological forcing. Hydrol. Earth Syst. Sci. https://doi.org/10.5194/hess-12-1007-2008 [Preprint].
Hattermann, F.F., et al., 2017. Cross-scale intercomparison of climate change impacts simulated by regional and global hydrological models in eleven large river basins. Clim. Chang. 141 (3), 561–576. https://doi.org/10.1007/s10584-016-1829-4.
Hayes, M.J., et al., 1999. Monitoring the 1996 drought using the standardized precipitation index. Bull. Am. Meteorol. Soc. https://doi.org/10.1175/1520-0477(1999)080<0429:MTDUTS>2.0.CO;2 [Preprint].
Heim, R.R., 2002. A review of twentieth-century drought indices used in the United States. Bull. Am. Meteorol. Soc. 83 (8), 1149–1166. https://doi.org/10.1175/1520-0477-83.8.1149.
Hisdal, H., et al., 2001. Have streamflow droughts in Europe become more severe or frequent? Int. J. Climatol. 21 (3) https://doi.org/10.1002/joc.619.
Huang, S., et al., 2017. Evaluation of an ensemble of regional hydrological models in 12 large-scale river basins worldwide. Clim. Chang. https://doi.org/10.1007/s10584-016-1841-8 [Preprint].
Van Huijgevoort, M.H.J., et al., 2013. Global multimodel analysis of drought in runofffor the second half of the twentieth century. J. Hydrometeorol. https://doi.org/10.1175/JHM-D-12-0186.1 [Preprint].

Irvine, P.J., et al., 2017. Towards a comprehensive climate impacts assessment of solar geoengineering. Earth's Future. https://doi.org/10.1002/2016EF000389 [Preprint].
Jacobi, J., et al., 2013. A tool for calculating the palmer drought indices. Water Resour. Res. 49 (9), 6086–6089. https://doi.org/10.1002/wrcr.20342.
Keyantash, J., Dracup, J.A., 2002. The quantification of drought: an evaluation of drought indices. Bull. Am. Meteorol. Soc. https://doi.org/10.1175/1520-0477-83.8.1167 [Preprint].
Krysanova, V. et al. (2017) "ISIMIP2a simulation data from water (regional) sector." doi:10.5880/PIK.2018.007.
Krysanova, V., et al., 2018. How the performance of hydrological models relates to credibility of projections under climate change. Hydrol. Sci. J. 63 (5) https://doi.org/10.1080/02626667.2018.1446214.
Krysanova, V., Hattermann, F.F., 2017. Intercomparison of climate change impacts in 12 large river basins: overview of methods and summary of results. Clim. Chang. 141 (3) https://doi.org/10.1007/s10584-017-1919-y.
Krysanova, V., Müller-Wohlfeil, D.I., Becker, A., 1998. Development and test of a spatially distributed hydrological/water quality model for mesoscale watersheds. Ecol. Model. 106 (2–3), 261–289. https://doi.org/10.1016/S0304-3800(97)00204-4.
Liang, X., et al., 1994. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. J. Geophys. Res. https://doi.org/10.1029/94jd00483 [Preprint].
Lindström, G., et al., 2010. Development and testing of the HYPE (hydrological predictions for the environment) water quality model for different spatial scales. Hydrol. Res. https://doi.org/10.2166/nh.2010.007 [Preprint].
Liu, Y., et al., 2019. Understanding the spatiotemporal links between meteorological and hydrological droughts from a three-dimensional perspective. J. Geophys. Res. Atmos. https://doi.org/10.1029/2018JD028947 [Preprint].
Lloyd-Hughes, B., 2014. The impracticality of a universal drought definition. Theor. Appl. Climatol. https://doi.org/10.1007/s00704-013-1025-7 [Preprint].
Lloyd-Hughes, B., Saunders, M.A., 2002. A drought climatology for Europe. Int. J. Climatol. 22 (13) https://doi.org/10.1002/joc.846.
van Loon, A.F., 2015. Hydrological drought explained. Wiley Interdiscip. Rev. Water. https://doi.org/10.1002/wat2.1085 [Preprint].
McKee, T.B., Doesken, N.J., Kleist, J., 1993a. The relationship of drought frequency and duration to time scales. In: Proceedings of the 8th Conference on Applied Climatology. Boston, pp. 179–183.
McKee, T.B., Doesken, N.J., Kliest, J., 1993b. The relationship of drought frequency and duration to time scales. In: Proceedings of the 8th Conference of Applied Climatology. Anaheim, CA, 17-22 JanuaryPreprints, Eighth Conf. on Applied Climatology, Amer. Meteor, Soc.
Milly, P.C.D., Dunne, K.A., 2017. A hydrologic drying bias in water-resource impact analyses of anthropogenic climate change. J. Am. Water Resour. Assoc. https://doi.org/10.1111/1752-1688.12538 [Preprint].
Mishra, A.K., Desai, V.R., Singh, V.P., 2007. Drought forecasting using a hybrid stochastic and neural network model. J. Hydrol. Eng. 12 (6) https://doi.org/10.1061/(asce)1084-0699(2007)12:6(626.
Mishra, A.K., Singh, V.P., 2010. A review of drought concepts. J. Hydrol. https://doi.org/10.1016/j.jhydrol.2010.07.012 [Preprint].
Moriasi, D.N., et al., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50 (3). https://doi.org/10.13031/2013.23153.
Moriasi, D.N., et al., 2015. Hydrologic and water quality models: performance measures and evaluation criteria. Trans. ASABE 58 (6). https://doi.org/10.13031/trans.58.10715.
Motovilov, Y.G., et al., 1999. Validation of a distributed hydrological model against spatial observations. Agric. For. Meteorol. https://doi.org/10.1016/S0168-1923(99)00102-1 [Preprint].
Muller 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 (7) https://doi.org/10.5194/hess-20-2877-2016.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I-a discussion of principles. J. Hydrol. [Preprint].
Oleson, K.W. et al. (2010) Technical description of version 4.0 of the community land model (CLM). doi:10.1.1.172.7769.
Pokhrel, Y., et al., 2021. Global terrestrial water storage and drought severity under climate change. Nat. Clim. Chang. https://doi.org/10.1038/s41558-020-00972-w [Preprint].
Pokhrel, Y.N., et al., 2015. Incorporation of groundwater pumping in a global land surface model with the representation of human impacts. Water Resour. Res. 51 (1) https://doi.org/10.1002/2014WR015602.
Prudhomme, C., et al., 2014. Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment. Proc. Natl. Acad. Sci. U. S. A. https://doi.org/10.1073/pnas.1222473110 [Preprint].
Samaniego, L., Kumar, R., Attinger, S., 2010. Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. Water Resour. Res. 46 (5) https://doi.org/10.1029/2008WR007327.
Shukla, S., Wood, A.W., 2008. Use of a standardized runoff index for characterizing hydrologic drought. Geophys. Res. Lett. https://doi.org/10.1029/2007GL032487 [Preprint].
Strauch, M., et al., 2017. Adjustment of global precipitation data for enhanced hydrologic modeling of tropical Andean watersheds. Clim. Chang. https://doi.org/10.1007/s10584-016-1706-1 [Preprint].
Tallaksen, L.M., Hisdal, H., Lanen, H.A.J.V., 2009. Space-time modelling of catchment scale drought characteristics. J. Hydrol. 375 (3–4) https://doi.org/10.1016/j.jhydrol.2009.06.032.
Tallaksen, L.M.L.M., et al., 2004. Hydrological Drought: Processes and Estimation Methods For Streamflow and Groundwater Developments in Water Science. Elsevier [Preprint].
Tang, Q., et al., 2007. The influence of precipitation variability and partial irrigation within grid cells on a hydrological simulation. J. Hydrometeorol. https://doi.org/10.1175/JHM589.1 [Preprint].
Telteu, C.E., et al., 2021. Understanding each other's models an introduction and a standard representation of 16 global water models to support intercomparison, improvement, and communication. Geosci. Model Dev. https://doi.org/10.5194/gmd-14-3843-2021.
UN and UNISDR, UNDP, I., 2009. Making Disaster Risk Reduction Gender Sensitive: Policy and Practical Guidelines. UNDP.
Veldkamp, T.I.E., et al., 2018. Human impact parameterizations in global hydrological models improve estimates of monthly discharges and hydrological extremes: a multi-model validation study. Environ. Res. Lett. https://doi.org/10.1088/1748-9326/aab96f.
Verzano, K., 2009. Climate change impacts on flood related hydrological processes: further development and application of a global scale hydrological model. Rep. Earth Syst. Sci. [Preprint].
Vicente-Serrano, S.M., et al., 2012. Accurate Computation of a streamflow drought index. J. Hydrol. Eng. https://doi.org/10.1061/(asce)he.1943-5584.0000433 [Preprint].
Wada, Y., Wisser, D., Bierkens, M.F.P., 2014. Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources. Earth Syst. Dyn. https://doi.org/10.5194/esd-5-15-2014 [Preprint].
Wartenburger, R., et al., 2018. Evapotranspiration simulations in ISIMIP2a-Evaluation of spatio-temporal characteristics with a comprehensive ensemble of independent datasets. Environ. Res. Lett. 13 (7) https://doi.org/10.1088/1748-9326/aac4bb.
Wasko, C., et al., 2021. Understanding trends in hydrologic extremes across Australia. J. Hydrol. 593. https://doi.org/10.1016/j.jhydrol.2020.125877.
Weedon, G.P., et al., 2011. Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century. J. Hydrometeorol. 12 (5) https://doi.org/10.1175/2011JHM1369.1.
Wilhite, D.A., 1993. The enigma of drought. Drought Assessment, Management, and Planning: Theory and Case Studies. https://doi.org/10.1007/978-1-4615-3224-8_1.
Wilhite, D.A., Hayes, M.J. and Svoboda, M.D. (2000) "Drought monitoring and assessment: status and trends in the United States," in. doi:10.1007/978-94-015-9472-1_11.
WMO, 1986. Report on Drought and Countries Affected by Drought During 1974–1985. World Meteorological Organization.
Wu, J., et al., 2018. Impacts of reservoir operations on multi-scale correlations between hydrological drought and meteorological drought. J. Hydrol. https://doi.org/10.1016/j.jhydrol.2018.06.053 [Preprint].
Yevjevich, V., Ica Yevjevich, V., 1967. An objective approach to definitions and investigations of continental hydrologic droughts an objective approach to definitions and investigations of continental hydrologic droughts. Hydrol. Pap. [Preprint].
Zaherpour, J., et al., 2018. Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts. Environ. Res. Lett. https://doi.org/10.1088/1748-9326/aac547 [Preprint].