Multimodel assessment of flood characteristics in four large river basins at global warming of 1.5, 2.0 and 3.0 K above the pre-industrial level

Environmental Research Letters

LETTER

Multimodel assessment of flood characteristics in four large river basins at global warming of 1.5, 2.0 and 3.0 K above the pre-industrial level

Shaochun Huang[1,2,7], **Rohini Kumar**[3], **Oldrich Rakovec**[3,4], **Valentin Aich**[2,5], **Xiaoyan Wang**[6], **Luis Samaniego**[3], **Stefan Liersch**[2] and **Valentina Krysanova**[2]

[1] Norwegian Water Resources and Energy Directorate (NVE), Oslo, Norway
[2] Potsdam Institute for Climate Impact Research (PIK), Potsdam, Germany
[3] UFZ-Helmholtz Centre for Environmental Research, Leipzig, Germany
[4] Czech University of Life Sciences, Faculty of Environmental Sciences, Prague, Czechia
[5] World Meteorological Organization (WMO), Geneva, Switzerland
[6] State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Center for Global Change and Water Cycle, Hohai University, Nanjing, People's Republic of China
[7] Author to whom any correspondence should be addressed.

E-mail: shh@nve.no

Keywords: flood timing, 100 year floods, flood frequency, climate change, CMIP5-GCMs, multi-model ensemble

Supplementary material for this article is available online

OPEN ACCESS

RECEIVED
2 March 2018

REVISED
14 October 2018

ACCEPTED FOR PUBLICATION
18 October 2018

PUBLISHED
23 November 2018

Abstract

This study assesses the flood characteristics (timing, magnitude and frequency) in the pre-industrial and historical periods, and analyzes climate change impacts on floods at the warming levels of 1.5, 2.0 and 3.0 K above the pre-industrial level in four large river basins as required by the Paris agreement. Three well-established hydrological models (HMs) were forced with bias-corrected outputs from four global climate models (GCMs) for the pre-industrial, historical and future periods until 2100. The long pre-industrial and historical periods were subdivided into multiple 31-year subperiods to investigate the natural variability. The mean flood characteristics in the pre-industrial period were derived from the large ensemble based on all GCMs, HMs and 31-year subperiods, and compared to the ensemble means in the historical and future periods. In general, the variance of simulated flood characteristics is quite large in the pre-industrial and historical periods. Mostly GCMs and HMs contribute to the variance, especially for flood timing and magnitude, while the selection of 31-year subperiods is an important source of variance for flood frequency. The comparison between the ensemble means shows that there are already some changes in flood characteristics between the pre-industrial and historical periods. There is a clear shift towards earlier flooding for the Rhine (1.5 K scenario) and Upper Mississippi (3.0 K scenario). The flood magnitudes show a substantial increase in the Rhine and Upper Yellow only under the 3.0 K scenario. The floods are projected to occur more frequently in the Rhine under the 1.5 and 2.0 K scenarios, and less frequently in the Upper Mississippi under all scenarios.

Introduction

Floods are the most frequent natural disasters globally, causing many fatalities and losses with increasing trend despite the growing awareness (Kundzewicz *et al* 2014). For example, floods in 2016 caused an estimated total economic damage of around 56 billion USD on the globe (Guha-Sapir *et al* 2017). According to model projections, damages are expected to increase due to both economic growth and climate change (Hirabayashi *et al* 2013, Winsemius *et al* 2016, Alfieri *et al* 2017). However, there is a low confidence in the observed trends at the global scale that anthropogenic climate change has already affected the magnitude or frequency of floods due to limited period of instrumental records on floods at gauge stations, and interference with non-climatic drivers, such as land use change and engineering (Hall *et al* 2014, Intergovernmental Panel on Climate Change IPCC (2012)).

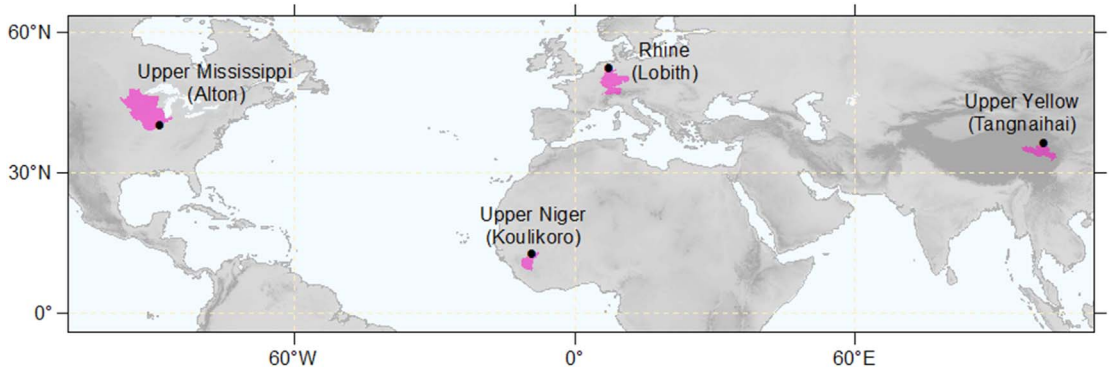

Figure 1. Location of the four case study river basins along with their gauging stations considered in the study. The grey shaded color indicates the terrain elevation with darker color being representative of higher altitude, and vice-versa.

In addition, existing climate impact studies are often unable to provide a consistent picture of projections on changes in flood magnitude in many regions due to selection of different warming scenarios, climate models, reference and future periods, bias-correction methods and hydrological models (HMs) (Kundzewicz *et al* 2017). Most recent impact studies select only one reference period (typically 30 years in the recent past period, e.g. 1971–2000) and several future periods under the Representative Concentration Pathways (RCP) scenarios to analyze changes in flood characteristics (Madsen *et al* 2014). However, such comparative analysis results should be cautiously interpreted due to long-term flood regime fluctuations (Hall *et al* 2014), which can be longer than 30 years. Moreover, the agreement of the Paris climate conference in December 2015 explicitly requests climate impact assessments of global warming at 1.5 and 2 K above the pre-industrial level (United Nations Framework Convention on Climate Change (UNFCCC) 2015) and not above the recent past period. Such assessments are rare in existing literature due to lack of knowledge about the pre-industrial level.

For the global projections on floods risks, global HMs are used (Arnell and Gosling, 2016, Alfieri *et al* 2018, Dottori *et al* 2018). However, such models often show larger bias in the simulated discharge than the calibrated regional-scale HMs at the basin scale (Ward *et al* 2015, Hattermann *et al* 2017, Krysanova *et al* 2018). Since the model performance in the historical period is important (Krysanova *et al* 2018), there is an increasing trend of using well calibrated and validated regional (or basin-scale) HMs at large river basins or continent scales for impact assessment (Aich *et al* 2014, Vetter *et al* 2015, Krysanova and Hattermann, 2017, Samaniego *et al* 2017, Vetter *et al* 2017, Thober *et al* 2018). Higher credibility of climate change impacts on average hydrological conditions and extremes can be expected using regional models at the basin scale (Krysanova *et al* 2018).

This study used two regional-scale and one large-scale HMs driven by climate projections from four global climate models (GCMs) applied by the regional water sector group in the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP, www.isimip.org) framework. All three HMs were calibrated and validated for average conditions and flood characteristics. Our aims are (1) to analyze flood characteristics, including timing, magnitude and frequency, in the pre-industrial and historical periods in terms of mean/median values and variance based on an ensemble of all GCMs, HMs and multiple subperiods and (2) to provide a first consistent view on changes in flood characteristics at the warming levels of 1.5, 2.0 and 3.0 K *above the pre-industrial level* for four large river basins located in different climate zones on four continents. The simulated impacts are further cross-checked using results of trend analysis, which also allow to qualitatively assess uncertainties related to GCMs and HMs.

Study Areas

Among the 12 large river basins used for the regional-scale water modeling in ISIMIP (Krysanova and Hattermann, 2017), we selected the study basins based on two criteria: (1) the basins should represent different geographic, physiographic, land cover, hydro-climatic characteristics as well as flood regimes, and (2) observed flood characteristics should be well reproduced by participating regional HMs—the selection in this case was based on a comprehensive model evaluation conducted by Huang *et al* (2017). As a result, four large-scale river basins (figure 1) located on different continents were selected: Upper Mississippi at Alton gauging station (drainage area 444.000 km^2), Rhine at Lobith (160.000 km^2), Upper Niger at Koulikoro (120.000 km^2), and Upper Yellow at Tangnaihai (121.000 km^2). The upper parts of three basins were chosen for the analysis because they are less influenced by water management or flood inundation processes (e.g. for the lower Niger basin (Aboubacar 2007)) than the lower parts. In the Rhine basin, the influence of land use and human water management

plays a minor role for the large-scale modeling (Bronstert *et al* 2007). Later in the text, the word 'Upper' will be omitted for better readability.

The selected basins differ in their physiographic, land cover and hydro-climatic characteristics (table S1 in supplement is available online at stacks.iop.org/ERL/13/124005/mmedia). The Yellow basin is situated in the Qinghai–Tibet Plateau with a mean elevation of approximately 4125 m a.s.l., whereas mean elevation of three other basins ranges from 300 to 500 m a.s.l. In terms of land cover, the cropland share is highest in the Mississippi (43%), followed by Rhine (38%) and Niger (24%). Climate of these basins also varies greatly. According to the Köppen climate classification scheme (Peel *et al* 2007), the Niger basin is mainly located in the tropical savanna climate region, the Mississippi in the warm summer continental climate, the Rhine in the oceanic climate and the Yellow in the alpine (montane) climate. Hence, long-term mean temperature/annual precipitation range between −2 °C/506 mm (Yellow) and 26.5 °C/1495 mm (Niger) (period 1971–2000).

The runoff coefficient representing the ratio of the long-term observed mean runoff to mean precipitation varies from 0.18 (Niger) to 0.44 (Rhine) across the study basins. The high flow seasons also vary in the basins (see fig. S1 in supplements). The Yellow river receives most of rainfall in summer with distinct high flows in this season. In the Niger basin, high flows occur mainly in late summer or early autumn. The rainfall-driven floods are dominant in these basins. The high flow season in the Mississippi basin is not so distinct and expands from spring to summer, resulting from both snowmelt and convective rainfall. In the Rhine basin, high flows mainly occur in winter and spring, but the seasonal fluctuation is less distinct compared to other three rivers. The last two basins have a mixture of rainfall- and snow-driven flood regimes. Readers may also refer to the introduction of the ISIMIP regional studies by Krysanova and Hattermann (2017) for further details on these basins.

Methods and Data

Observational and reanalysis datasets

Within the ISIMIP framework, setup of all impact models is done using the same land-surface and meteorological datasets. The Shuttle Radar Topography Mission Digital Elevation Model with 90 meter (3 arc seconds) horizontal resolution was used to delineate the river network, catchments, sub-basins, and slope parameters. Soil properties were derived from the Harmonized World Soil Database (FAO *et al* 2009) and information on land use/cover was taken from global land cover database for the year 2000 (Bartholome and Belward 2005). Both the soil and land use databases are at 30 arc-s resolution (ca. 926 m). Daily time-series of observed discharge data at the gauge stations for calibration/validation of HMs were obtained from the Global Runoff Data Centre (GRDC 2018) for the Rhine and Mississippi, from the Hydrology Bureau, the Yellow River Conservancy Commission (Yang *et al* 2014) for the Yellow and from the Niger-HYCOS monitoring network (Niger Basin Authority 2008) for the Niger. More detailed information on the model input data can be found in Krysanova and Hattermann (2017).

The EWEMBI climate dataset (EartH2Observe, **W**ATCH-Forcing-Data-ERA-Interim and **E**uropean Reanalysis (ERA)-Interim data **M**erged and Bias-corrected for **I**SIMIP) (Lange 2016) served as meteorological input to calibrate and validate three HMs. The EWEMBI dataset is based on ERA-Interim reanalysis data (Dee *et al* 2011), WATer and global CHange (WATCH) forcing data (Weedon *et al* 2014), EartH2Observe forcing data (Calton *et al* 2016) and National Aeronautics and Space Administration/Global Energy and Water cycle Experiment (NASA/GEWEX) Surface Radiation Budget data (Stackhouse *et al* 2011). It provides daily meteorological variables (mean, maximum and minimum temperatures, precipitation, solar radiation, relative humidity and wind) for the period 1979–2013 at a spatial resolution of 0.5° (ca. 55 km). This reanalysis data provides the consistent climate data for all basins with good quality.

Climate model datasets

Aligned with the pledges of the Paris agreement, the ISIMIP project proposed tailored climate simulations for the pre-industrial, historical and future periods (Frieler *et al* 2017).

They are:

(1) A multi-centennial pre-industrial reference simulation with fixed pre-industrial socio-economic conditions (*piControl,* 1661–1860);

(2) Historical simulations accounting for varying socio-economic conditions and climate change (*historical,* 1861–2005);

(3) and (4) future simulations with fixed year 2005 socio-economic conditions and climate change under the strong-mitigation scenario RCP2.6 closest to the global warming limits agreed on in Paris and non-mitigation scenario RCP6.0 (*RCP2.6 and RCP6.0*, 2006–2099).

Outputs of four GCMs from the fifth phase of the Coupled Model Intercomparison Project (CMIP5) (Taylor *et al* 2012) archive were used for this assessment, namely: Geophysical Fluid Dynamics Laboratory's Earth System Model 2M (GFDL-ESM2M) (Dunne *et al* 2012), Hadley Global Environment Model 2—Earth System (HadGEM2-ES) (Collins *et al* 2011), Institute Pierre Simon Laplace Climate Model 5A Low Resolution (IPSL-CM5A-LR) (Dufresne *et al* 2013) and

the Model for Interdisciplinary Research on Climate 5 (MIROC5) (Watanabe *et al* 2010). These GCMs were selected to represent the space of global mean temperature (GMT) change and relative precipitation changes of all GCMs from the CMIP5 as comprehensive as possible (Warszawski *et al* 2014) and they provide the full set of output variables for the planned multi-sectoral simulations to be used in the IPCC Special Report on the 1.5 K target.

The raw GCM data were interpolated to 0.5° horizontal resolution (ca. 55 km) using a first-order conservative remapping scheme (Jones, 1999), and linearly interpolated to the standard Gregorian calendar (365 d/year plus leap days) if necessary. Outputs were also bias-corrected to the EWEMBI dataset at daily time step using a trend preserving statistical bias-correction algorithm developed by Hempel *et al* (2013). More detailed information on the GCMs and their simulations of temperature and precipitation are given in Supplement.

The 1.5, 2.0 and 3.0 K climate scenarios were obtained from two future simulations by calculating 31 year running means of GMT minus GMT in piControl, 1661–1860, as suggested in the ISIMIP project (Frieler *et al* 2017). The middle years, in which the 31 year running mean of GMT crosses 1.5, 2.0 and 3.0 K are shown in table S3. The GFDL-ESM2M and MIROC5 models project slower GMT increase than other GCMs, so they reach the 1.5 K threshold in the middle of the century under the RCP2.6 and RCP6.0 scenarios. The GFDL-ESM2M run even fails to reach the 1.5 K threshold under the RCP2.6 scenario. The IPSL-CM5A-LR shows the fastest GMT increase among all GCMs, and reaches the 1.5 and 2 K thresholds before 2030 under both scenarios. The other three GCM runs project an increase by 2 K after 2050, and only the IPSL-CM5A-LR and HadGEM2-ES runs reach the 3 K threshold around 2070s under the RCP6.0 scenario. In total, seven 31-year periods representing the 1.5 K scenario, five periods for the 2.0 K scenario, and only two periods representing the 3.0 K scenario were found in future simulations (table S3). Since these 31-year scenario periods were taken from four GCMs and two RCPs, they were considered as ensembles of future projections (Mitchell *et al* 2016).

Since land use change and human influences were not considered in our hydrological modeling, comparison of simulation outputs between the chosen future periods and piControl allows analyzing the pure effect of climate warming by 1.5, 2.0 and 3.0 K above the pre-industrial level. Note that the piControl (200 years) and historical period (145 years) are much longer than the future simulation periods (31 years). In order to make the comparison and investigate the natural variability, we used a 31-year centered moving window for the pre-industrial and historical periods to calculate the flood indices. In total, there are 169 and 114 of 31-year subperiods for the piControl and historical simulations, respectively.

Hydrological Models

In this study, we used three spatially semi-distributed HMs for the assessment of changes in flood characteristics. These include two regional-scale models: the Soil and Water Integrated Model (SWIM) (Krysanova *et al* 1998) and the mesoscale Hydrologic Model (mHM) (Samaniego *et al* 2010, Kumar *et al* 2013), and one large-scale model: the Variable Infiltration Capacity model (VIC) (Liang *et al* 1994), all being a part of the community driven modeling effort in the ISIMIP regional water sector. All three HMs were forced by the EWEMBI reanalysis data and were applied at a regional-scale with parameters being calibrated against time-series of daily discharge in each catchment independently (see more details about these models and their calibration in supplement).

Flood Indices

In this study, we calculated three indices for flood timing, magnitude and frequency within each 31-year period: (1) the average flooding date, (2) 100 year flood discharge and (3) the total number of floods within the 31 years. Within each period, we extracted two variables from the daily discharge series: the annual maximum discharge (AMD) from hydrological years, and the independent peaks selected by the peak-over-threshold (POT) approach. The hydrological years start in October in the Rhine and Mississippi and in April in the Niger (Aich *et al* 2016, Petrow and Merz, 2009, US Geological Survey, 2016). There is no hydrological year defined in China, so the calendar year was used for the Yellow. The thresholds were selected from the piControl simulations so that two independent peaks per year are represented on average in the whole pre-industrial period. The same thresholds were then applied for the historical and scenario simulations to select peaks.

The dates of the AMD were averaged for each 31-year period to calculate the average flood date for each gauging station using circular statistics, following the approach by Bloschl *et al* (2017). The flood discharge related to the 100 year return period was estimated by fitting the generalized extreme value distribution (Coles 2001) to the AMD. The number of independent peaks over threshold in each 31-year period were considered as the total number of floods (Mallakpour and Villarini 2015) and compared between the piControl and other periods. More detailed information on the definition of flood indices and their calculation is given in supplement.

Analysis of variance in the pre-industrial period

We applied the Analysis of variance (ANOVA) method (Bosshard *et al* 2013, Vetter *et al* 2015) to partition the variance of simulated flood characteristics in the

Table 1. Validation results of hydrological models for the historical period driven by EWEMBI (EartH2Observe, WATCH-Forcing-Data-ERA-Interim and European Reanalysis (ERA)-Interim data Merged and Bias-corrected for ISIMIP) meteorological data with daily time step. The validation periods slightly differ between basins, as they were chosen based on hydrological data availability.

				Hydrological models		
Basin	Gauge	Period	Criteria	mHM	SWIM	VIC
Rhine	Lobith	1980–2010	NSE	0.9	0.86	0.37
			KGE	0.94	0.9	0.72
			PBIAS (%)	3	0	6
			AMD PBIAS (%)	8	−20	19
Yellow	Tangnaihai	1980–2009	NSE	0.8	0.71	0.48
			KGE	0.84	0.84	0.74
			PBIAS (%)	−5	−3	16
			AMD PBIAS (%)	−9	7	15
Mississippi	Alton	1980–2001	NSE	0.82	0.72	0.54
			KGE	0.91	0.8	0.78
			PBIAS (%)	0	4	2
			AMD PBIAS (%)	6	−4	26
Niger	Koulikoro	1980–2006	NSE	0.73	0.77	0.64
			KGE	0.86	0.84	0.78
			PBIAS (%)	1	−4	−9
			AMD PBIAS (%)	−1	−9	13

pre-industrial period into three contributing sources: the GCMs, HMs and the selection of 31-year periods (PE). The ANOVA method allows providing information about the contribution of these three major sources and their interaction terms (GCM × HM, GCM × PE, PE × HM and GCM × PE × HM) to the variance of flood characteristics. To avoid the bias caused by different sample sizes of the sources, the ANOVA was implemented for a number of subsamples, each of which includes three 31-year periods, three GCMs and three HMs. The obtained estimates were then averaged. For more explanation of the method and equations, please refer to Bosshard *et al* (2013).

Trend analysis

In addition to calculation of the flood indices in each 31-year period and analysis of impacts under different warming levels, the long-term evolution of indices from the historical period to 2099 was analyzed under the RCP2.6 and 6.0 scenarios. The period 1911–2099 covering equal lengths in the historical and scenario periods was chosen for that. The aim was to extend and crosscheck the analysis of 1.5, 2.0 and 3.0 K scenarios.

Similar to Bloschl *et al* (2017), the annual series (AMD, AMD dates and the number of floods over threshold per year) were subject to a 31-year moving average filter. The Mann–Kendall test (Hipel and McLeod, 1994) was used to detect the significance of trends of the averaged series at the 5% significance level, and the Sen's slope (Hipel and McLeod, 1994) was used to estimate the magnitude of the trend.

Results

Model validation

It is important to validate the HMs' performance against observations before they are used for climate impact assessments (Krysanova *et al* 2018). The model validation results for the historical periods are shown in table 1, where the model performance is presented in terms of Nash and Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970), Kling–Gupta efficiency (KGE) (Gupta *et al* 2009), percent bias of discharge (PBIAS), and percent bias in AMD (AMD PBIAS). For the AMD PBIAS, we consider the results within ±25% as satisfactory for the large-scale modeling according to Huang *et al* (2017), since none of the models was calibrated specifically against the flood peaks.

Based on the NSE, KGE and PBIAS criteria, the mHM and SWIM models have a good performance (NSE $>$ 0.7, KGE $>$ 0.7, PBIAS $<$ ±10%). The NSE for the VIC model is lower compared to other models but the KGE are above 0.7, indicating a reasonable model performance. In general, all models simulate the flood peaks satisfactorily (AMD PBIAS within ±25%), except one case: the VIC model for the Mississippi with 26%. Note that VIC in this study often overestimates the flood peaks, probably because the parameters related to snow processes were not used for calibration (Hurkmans *et al* 2008).

In addition to the statistical criteria, we also compared the observed and simulated AMD and their timing in figure 2. The specific discharge (mm/d) is used here to account for differences in basin areas. In general, the simulated AMD have a good agreement with the observed ones, as indicated by r^2 (coefficient of

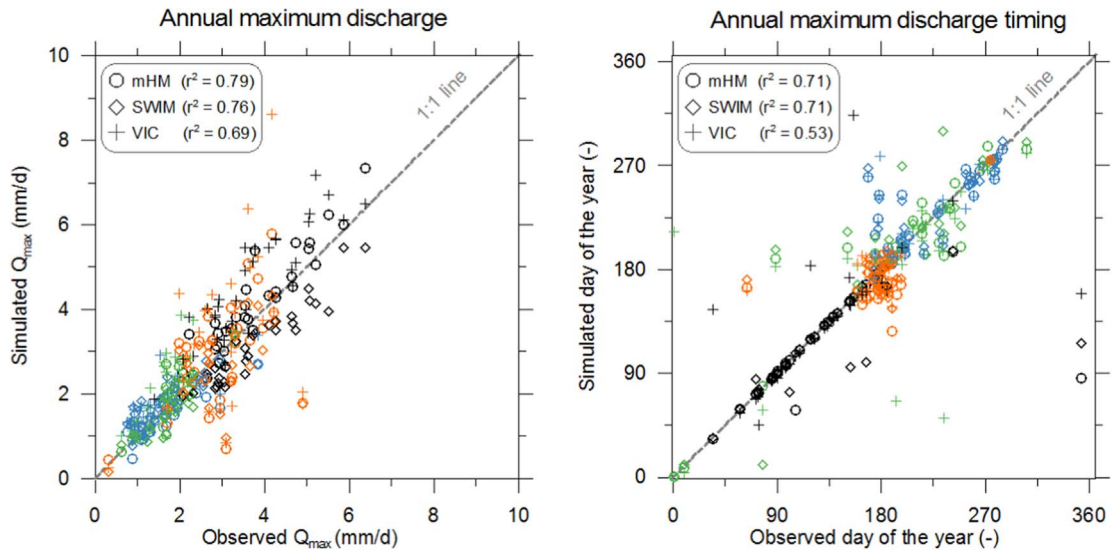

Figure 2. Evaluation of model simulations against observations for the annual maximum discharge (left) and annual maximum discharge timing (right) across the four basins (black: Rhine; blue: Yellow; green: Mississippi; orange: Niger). Discharge values shown here represent specific runoff in mm/d to account for the variable basin areas and to enable better comparison between basins. Note that the starting day of the hydrological year varies among basins (first day of the calendar month of October for the Rhine and Mississippi, April for the Niger, and January for the Yellow).

determination) $\geqslant$0.69 for all results. The simulation of timing is better for mHM and SWIM than for VIC ($r^2 = 0.53$).

Flood characteristics in the pre-industrial and historical periods: simulation variance

The first objective of this study was to analyze flood characteristics in the long pre-industrial and historical periods. The ensembles of flood characteristics based on combinations of four GCMs and three HMs are shown in figures S3–S8 (supplement). The spread of the boxplots illustrates the effect of multiple 31-year subperiods, i.e. the natural variability. The results clearly show substantial variance due to selection of GCMs, HMs as well as 31-year subperiods. Table 2 illustrates the contribution of three sources to the overall variance of the simulated flood characteristics in the pre-industrial periods calculated using the ANOVA method.

In general, the GCMs and HMs are the major sources of variance for flood timing and magnitude (explaining 60%–80% of the variance) whereas the selection of multiple 30-year periods (PE) and an interaction term GCM $\times$ PE are important for flood frequency. The GCMs explain more than 40% of the variance of the flood timing in the Rhine and Niger while the HMs contribute more than 35% in the Yellow and Mississippi. HMs are the dominant source for three of the four basins for flood magnitude, mainly arising from substantial variations of simulated flood discharges among different HMs. The large variations can be partly attributed to the simulation biases noticed in the validation period among different HMs; and partly due to different model responses under changed climate conditions. The GCMs and PE as well as their interaction term contribute more than 65% to the total ensemble variance of flood frequency in three of the four basins. In the Upper Niger, HMs are the dominant source for flood frequency.

Flood characteristics in the pre-industrial and historical periods: comparison of means, medians and spreads

Due to the large variance of flood characteristics in the pre-industrial period shown above, we focus on the ensemble results based on all GCMs, HMs and 31-year subperiods (grey and green boxplots in figure 3). Among the four basins, the Rhine has the widest spread of flood timing ranging from December to August in the pre-industrial period. The two-sample *t*-test (Snedecor and Cochran, 1989) was used to determine if the ensemble means in the pre-industrial and historical periods are equal. We observe that there is no substantial shift of flood timing between the pre-industrial and historical periods in the Rhine and Niger. In contrast, there is a clear shift towards later flood timing in the Mississippi and Yellow in the historical period suggested by the median/mean days at the 0.05 significance level (see the black points with stars in figure 3). The ranges of flood timing in the historical period are generally within the ranges of piControl for the Niger and Yellow, but it is larger for the Mississippi and smaller for the Rhine.

There is a large variance of flood magnitude simulated in both pre-industrial and historical periods in four basins. The spreads of the boxplots in the

Table 2. Variance decomposition of the simulated flood characteristics (timing, magnitude and frequency) for the four studied river basins in the pre-industrial period considering three contributing sources: general circulation models (GCM), hydrological models (HM) and selection of the 31-year subperiod (PE).

		Contributing sources of variance						
					Interaction terms			
		General circulation model (GCM)	Hydrological model (HM)	Selection of subperiods (PE)	GCM × HM × PE	GCM × PE	HM × PE	GCM × HM
Timing	Rhine	0.43	0.17	0.06	0.07	0.08	0.03	0.16
	Yellow	0.22	0.44	0.04	0.08	0.06	0.03	0.14
	Mississippi	0.28	0.36	0.07	0.05	0.16	0.03	0.05
	Niger	0.62	0.14	0.02	0.05	0.04	0.02	0.11
Magnitude	Rhine	0.11	0.66	0.03	0.04	0.08	0.02	0.06
	Yellow	0.63	0.17	0.03	0.03	0.06	0.02	0.04
	Mississippi	0.11	0.62	0.07	0.03	0.11	0.02	0.03
	Niger	0.33	0.43	0.02	0.02	0.05	0.01	0.15
Frequency	Rhine	0.22	0.03	0.17	0.10	0.37	0.05	0.06
	Yellow	0.17	0.05	0.16	0.14	0.32	0.09	0.07
	Mississippi	0.23	0.02	0.17	0.06	0.44	0.04	0.04
	Niger	0.13	0.76	0.01	0.02	0.02	0.01	0.06

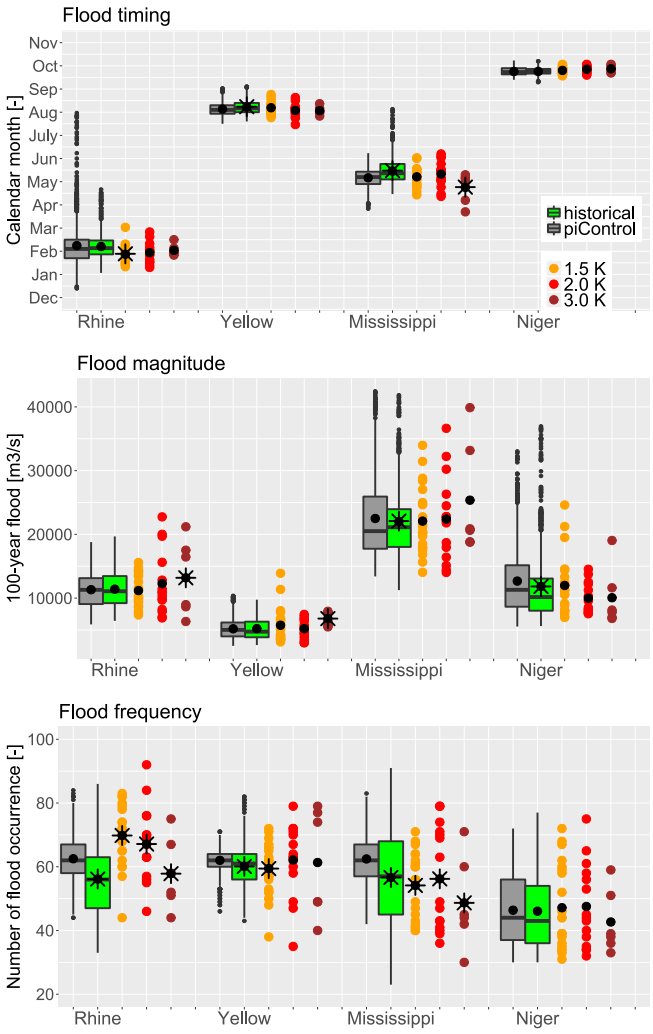

Figure 3. The multi-model ensemble results on the average flood date, 100 year flood discharge and total number of flood occurrences within 31 years in the pre-industrial and historical periods, as well as in future scenario periods at the global warming levels of 1.5, 2.0 and 3.0 K for four study basins. The box plots summarize ranges, median values and the 25 and 75 percentiles of the ensemble outputs based on simulation results using four global climate models, three hydrological models and multiple 31-year subperiods. The black points show the mean values of the ensembles and the points with stars show that the means are different at the 0.05 significance level compared to the pre-industrial means.

pre-industrial and historical periods are comparable. The median floods show a small increase in the Mississippi and a decrease in the Yellow and Niger. However, the ensemble means of flood magnitude are well comparable except for the Mississippi and Niger showing a minor decrease of −2% and −8%, respectively. Despite the large variability, there are few extreme flood events in the historical period in the Rhine and Niger that overshot the highest flood levels in the pre-industrial period.

Since the threshold was determined to select two independent flood events per year on average in the pre-industrial period, the median values of the total number of floods within 31-year subperiods at the pre-industrial level are around 62. Note that it was not possible to select two independent flood events for the Niger because there is only one flood peak in most years, especially as simulated by the SWIM and mHM models. The ensemble mean values suggest less frequent floods in the historical period than in piControl for the Rhine, Yellow and Mississippi (by 3%–10%), and no changes in the Niger. The 25–75 percentile ranges of the green boxes are wider than the grey ones for three basins, except the Niger, indicating that the flood frequency patterns changed in the historical period compared to the pre-industrial level in these basins.

The differences in flood characteristics between the pre-industrial and historical periods found in this study clearly show that the flood conditions in the historical period cannot fully represent the conditions of the pre-industrial level. Thus, the pre-industrial level should be specifically investigated to fulfill the requests of the Paris agreement.

Climate change effects at the 1.5, 2.0 and 3.0 K warming levels

Figure 3 also shows the flood characteristics under future scenarios, which can be compared to the pre-industrial level. In total, there are 21, 15 and 6 future simulations from three HMs driven by 7, 5 and 2 climate projections corresponding to 1.5, 2.0 and 3.0 K warming, respectively. Due to the small sample sizes, we plotted the colored points corresponding to all future scenarios, and superimposed the mean values as black points. Here, we mainly compare the ensemble mean values in the scenarios to the mean values of the grey boxes in the piControl period and focus on the results that are statistically significant (the black points with stars). We notice that the statistically significant assessments are not robust for the 3.0 K scenarios due to the small sample size.

Regarding the flood timing, floods tend to occur 11 d earlier in the Rhine under the 1.5 K scenario, and 13 d earlier in the Mississippi under the 3.0 K scenario. This may be due to earlier snowmelt under a warmer climate. In the Niger with summer flooding, the flood timing is likely to shift by 4 d later under the 2.0 and 3.0 K scenarios, but the differences are not statistically significant.

There is a pronounced increase (by 13%–16%) in the 100 year flood magnitude under the 3.0 K scenario in the Rhine and Mississippi. In the Yellow, increases by 11% and 31% are projected under the 1.5 and 3.0 K scenarios, respectively. Flood magnitude shows a small decrease (−5%) in the Niger under the 1.5 K scenario, and a more substantial decrease (−20%) under the 2.0 and 3.0 K scenarios. However, most of the changes are not statistically significant. In addition, there are a few floods in the Rhine and Yellow in future, which overshot the highest floods in the pre-industrial period. This indicates that very extreme floods that never happened under piControl could occur in future.

Based on simulations conducted here, floods are projected to occur more frequently (7%–11%) in the Rhine under the 1.5 and 2.0 K scenarios, and less frequently (10%–22%) in the Mississippi under all scenarios. Besides, projections under the 3.0 K scenario show lower frequency (by about 7%) for the Rhine. We observe only minor changes in the Yellow river under the 1.5 K scenario. Some numbers of flood occurrence in three basins (except Mississippi) overshoot the highest records in the pre-industrial period.

Agreement/disagreement of general circulation models and HMs on trends

The trends in the AMD, AMD dates and the number of independent peaks per year were tested for the period 1911–2099 using the Mann–Kendall test. The Sen's slopes were calculated to indicate the positive or negative trends. The outputs of trend analysis were analyzed regarding agreement of the forcing GCMs and HMs on the direction and significance of trend. Table 3(a) presents results showing five distinct cases: strong, moderate and weak trends, no trend, and disagreement on trends, and table 3(b) presents the same results but grouped by HMs. Namely, five distinct cases are distinguished:

1) strong agreement on positive/negative trend, if outputs of all three HMs for a certain GCM (a) or outputs driven by all four GCMs for a certain HM (b) show statistically significant positive/negative trend (cases +3/+4, −3/−4);
2) moderate agreement on positive/negative trend, if outputs of two HMs/driven by two GCMs show significant trends and others: insignificant trend (+2, −2),
3) weak trend, if only one model shows significant trend, and others: insignificant (+1, −1),
4) agreement on no trend, if all models show insignificant trends (0), and
5) disagreement, if at least one model shows a significant positive trend and at least one model shows a significant negative trend.

The results reveal that most of GCMs and HMs agree on increasing level and frequency of floods for the Rhine under RCP2.6 (except for frequency under HadGEM2). There is also certain agreement on increasing level of floods in the Yellow under RCP2.6, but it is weaker than for the Rhine. Climate models disagree on flood trends in the Mississippi: both level and frequency of floods increase under GFDL, and decrease under three other GCMs, whereas timing is shifted to earlier dates according to most of model outputs. In other cases: for the Niger under both RCPs and for the Rhine and Yellow under RCP6.0 there is disagreement of GCMs and partly also HMs on trends.

Table 3(b) presents the same results but grouped by HMs (with cases +4, −4 added, as four runs driven by four GCMs are presented here for every HM), where the same tendencies are visible. However, a comparison of (a) and (b) parts of table 3 reveals a much higher agreement of the HMs on trends (a: only 13% of cases (grey boxes) show disagreement) than the GCMs (b: 69% of cases show disagreement). The higher agreement implies lower uncertainty contributed by the HMs to the overall results, which is in line with previous assessment of uncertainty sources in ISIMIP for 12 large river basins using more GCMs, HMs and four RCPs (Vetter *et al* 2017).

Table 3. Agreement of hydrological models and forcing general circulation models (GCMs) on trends in flood timing (fl. timing: annual maximum discharge date), flood level (fl. level: annual maximum discharge) and flood frequency (fl. freq.: number of independent peaks over threshold per year) in the period 1911–2099: (a) results are grouped by GCMs, (b) results are grouped by hydrological models (HM). Five cases with different levels of agreement/disagreement are explained in the text.

GCMs	RCPs	Rhine			Yellow			Mississippi			Niger		
		fl. timing	fl. level	fl. freq.	fl. timing	fl. level	fl. freq.	fl. timing	fl. level	fl. freq.	fl. timing	fl. level	fl. freq.
(a)													
GFDL	2.6	−2[d]	+3[f]	+3	+3	+1[g]	−1[h]	−3[a]	+3	+3	−1	+3	X[b]
Had	2.6	+3	+3	X	X	+2[c]	0[i]	−3	−3	−2[d]	X	+3	−3
IPSL	2.6	−3	+2	+2	+3	+2	+2	−2	−2	X	+3	−3	−3
MIROC	2.6	−2	+3	+3	X	+2	+3	−1	−2	−2	−1	−1	+1
GFDL	6.0	−3	X	X	+1	+1	+2	+3	+3	+3	−2	+3	+2
Had	6.0	+3	X	−2	−2	−3	−3	−2	−3	X	X	X	−2
IPSL	6.0	+2	−2	X	−2	+3	+3	−3	−3	−2	+3	−3	−3
MIROC	6.0	−3	+3	+3	−3	−3	+3	−3	−2	−3	+1	+3	+2
(b)													
HMs	RCPs	Rhine			Yellow			Mississippi			Niger		
		fl. timing	fl. level	fl. freq.	fl. timing	fl. level	fl. freq.	fl. timing	fl. level	fl. freq.	fl. timing	fl. level	fl. freq.
mHM	2.6	X	+4[e]	+4	X	+3	+3	−3	X	X	X	X	X
SWIM	2.6	X	+4	X	X	+3	+3	−3	X	X	+3	X	−3
VIC	2.6	X	+3	+4	X	+1	X	−3	X	+2	X	X	X
mHM	6.0	X	+3	+3	−3	X	X	X	X	X	X	X	X
SWIM	6.0	X	X	X	X	X	X	X	X	X	+3	X	−1
VIC	6.0	X	X	X	−2	X	X	X	X	X	X	X	X

[a] 3 sig neg.
[b] 2 sig neg, 1 insign.
[c] 1 sig neg, 2 insign.
[d] 3 insign.
[e] 1 sig pos, 2 insign.
[f] 2 sig pos, 1 insign.
[g] 3 sig pos.
[h] 4 sig pos.
[i] Disagreement.

Discussion

One uniqueness of this study is inclusion of multiple subperiods to assess the flood characteristics in the pre-industrial and historical periods. The ANOVA analysis confirms the importance of selection of subperiods and indicates that using a single short reference period is not sufficient for the flood frequency analysis. Willner *et al* (2018) also considered the natural variability of floods by deriving the median and uncertainty from 12 independent 34-year pre-industrial periods. In our study, if we would also have selected the independent 31-year subperiods instead of using the moving window method, we could obtain two sets of periods: (1) 1675–1705, 1706–1736, and so on; and (2) 1662–1692, 1693–1723, and so on. Figure S9–S11 in supplementary show the comparisons of results for the flood characteristics derived based on the ensembles of all multiple 31-year periods and the ensembles of two sets of periods. In general, there are only minor differences between the median results and some notable differences between the 25th percentile and 75th percentiles of the three ensemble results. However, the highest values are often different between the two subsets and the full ensemble always shows larger ranges than the two subsets. Hence, the ensemble using the moving window method provides the reliable median values, even if the selected periods are correlated. It also shows the complete results with the most extreme conditions, which is important for worse case analyzes, compared to the subset periods. In addition, selection of subsets is still arbitrary and may lead to different changes in some specific cases.

This study also highlights that HMs are one of the major sources of variance for flood simulations in the pre-industrial period, and it could be partly due to their bias in the validation period. Specifically in this study, better validation results could be achieved by using observational climate data rather than reanalysis data, improving description of flood generation processes in HMs and calibrating the models for flood peaks specifically. In addition, it is worth mentioning that land use change and human influences were not considered in our modeling, which would not only affect the validation results but also future projections. A more comprehensive and robust assessment on floods could be achieved by considering these drivers and by using more HMs and GCM simulations.

The time-sampling approach to determine the 1.5, 2.0 and 3.0 K climate scenarios has been recently applied to investigate the climate impact on water resources and hydrological extremes at different global warming levels (Donnelly *et al* 2017, Gosling *et al* 2017, Marx *et al* 2018, Samaniego *et al* 2018, Thober *et al* 2018). Despite its advantages of being simple and easily derived from available (CMIP5 simulations) datasets, the approach has also certain limitations (see James *et al* (2017) for an in-depth discussion on this topic). The first limitation is that it generates different number of ensemble members for different warming scenarios. For example, we have only two simulations for the 3.0 K scenario, and therefore we could hardly perform statistically significant assessments and produce robust recommendations for this warming level. The second disadvantage is that this approach is sensitive to the multi-decadal natural variability. As the natural variability is strong in both the piControl and historical periods, it should not be ignored in climate impact studies, especially for flood frequency assessments. We assume that the range of natural variability is mostly covered for the 1.5 and 2.0 K scenarios, because the total scenario periods contain 217 years (7 times 31) and 155 years (5 times 31), respectively. The mean flood characteristics may be biased by the limited sample under the 3.0 K scenario. Hence, the results for the 1.5 and 2.0 K scenarios are more robust than for the 3.0 K scenario in this study. Nevertheless, we could find some floods that overshoot the historical records in the Rhine, and all floods in the Yellow were higher than the mean pre-industrial level (figure 3) under 3.0 K despite the small sample size. This may imply a higher risk of extreme floods under the higher-end warming scenarios. The third disadvantage of the time-sampling approach lies in an assumption that the implications of the GMT increments will be the same regardless of the emissions pathway. James and Washington (2013) showed the time-sampling approach is robust to the rate of anthropogenic forcing while greenhouse gases are still rising. Furthermore, Donnelly *et al* (2017) compared the impact of the high and low RCP ensembles for changes in annual mean and maximum runoff across Europe under the 2 K warming level and found that the differences due to RCPs are generally smaller than the difference between warming levels. However, it is worth-noting that the time-sampling approach cannot provide reliable scenarios if there is a lag in the response to anthropogenic forcing or changes due to emission reductions. In addition, this method would not be appropriate for the long-term impacts such as glacier storage.

It is not straightforward to compare the current results with others in existing literature analyzing the impacts at the 1.5, 2.0 and 3.0 K warming levels due to differences in model setting, data and methods. The major difference is that other studies (Hirabayashi *et al* 2013, Thober *et al* 2018, Willner *et al* 2018) did comparisons only for the annual high flows (Q_{10}, Q_{max}) to recent historical decades (one 30-year reference period only), and not to the long pre-industrial period. Donnelly *et al* (2017) compared the future high flow projections to the pre-industrial level from 1881 to 1910 and (Willner *et al* 2018) used one historical period but considered the natural variability derived from the pre-industrial period. However, if we compare the historical level with different warming levels (mostly 2.0 or 3.0 K) in our modelling, our results broadly corroborate with findings of their results, where the

projected flood magnitude in the 21st century is expected to increase in the Yellow and Rhine River basins.

Conclusions

This study within the framework of ISIMIP provides a first consistent view of flood characteristics in the long pre-industrial and historical periods for four large river basins and projects the climate change impacts on floods at the warming levels of 1.5, 2.0 and 3.0 K above the pre-industrial level. Our study differs from previous climate impact studies on floods by considering the long pre-industrial period as a reference and evaluating the effects of selection of 31-year subperiods (i.e. natural variability). The contribution of individual GCMs and HMs to overall variance is substantial for the simulation of flood timing, magnitude and frequency in the pre-industrial period, indicating the importance of using ensembles of GCMs and HMs for impact assessment. In addition, the selection of 31-year subperiods is an important contributor to variance for flood frequency, confirming that the flood regimes can vary over several decades. Thus, it is necessary to account for the natural variability in the impact assessments, especially for flood frequency.

Despite the strong variance of projections, there are some pronounced changes indicated by the mean values in the historical period accounting for all GCMs, HMs and selection of subperiods. There is a clear shift towards later flood timing in the Yellow and Mississippi and lower frequency of floods in the Rhine, Yellow and Mississippi. Thus, the historical period cannot fully represent the pre-industrial level and it should be used cautiously as reference to fulfill the request by the Paris agreement. In the future, floods are projected to occur earlier in the Rhine (6–11 d) and less frequently in the Mississippi (10%–22%). The strongest changes in flood magnitude are projected under the 3.0 K scenario: an increase by 13%–16% in the Rhine and Mississippi and by 31% in the Yellow, and a decrease by 20% in the Niger. Under the 1.5 and 2.0 K scenarios, changes in flood magnitude are within 11%. However, the changes are statistically significant only for the Rhine and Yellow under the 3.0 K scenario. There are few very extreme events under future warming which overshoot the highest floods in the pre-industrial period, and some numbers of flood occurrence overshoot the highest records in the pre-industrial period (except for Mississippi). This indicates that more extreme and frequent floods are possible to occur under a warmer climate.

The results on trend analysis in general agree with the previous assessment of impacts corresponding to 1.5, 2.0 and 3.0 K scenarios. The evaluation of trend analysis results regarding agreement of forcing GCMs and HMs revealed a much higher agreement of HMs than GCMs on the simulated trends. The comparison of results on trend analysis grouped by GCMs and by HMs confirms that uncertainty related to GCMs is much higher than the HMs uncertainty.

The ensemble sizes of the 1.5, 2.0 and 3.0 K scenarios are different in this study. The results may be biased, especially for the 3.0 K scenario, due to the small number of samples. More robust results could be expected with larger ensembles from the high-end RCP scenarios and consideration of other impacts, e.g. land use change and human influences.

Acknowledgements

The authors would like to thank the ISIMIP phase 2b to provide the GCM simulations. This work is partially funded by the Helmholtz Alliance Climate Initiative REKLIM (www.reklim.de).

ORCID iDs

Shaochun Huang https://orcid.org/0000-0001-7426-5181
Rohini Kumar https://orcid.org/0000-0002-4396-2037
Oldrich Rakovec https://orcid.org/0000-0003-2451-3305
Valentin Aich https://orcid.org/0000-0002-3699-3775
Luis Samaniego https://orcid.org/0000-0002-8449-4428
Stefan Liersch https://orcid.org/0000-0003-2778-3861

References

Aboubacar A 2007 *Niger River Basin Atlas* (Niamey, Niger: Niger Basin Authority)
Aich V, Kone B, Hattermann F F and Paton E N 2016 Time series analysis of floods across the niger river basin *Water-Sui* **8** 165
Aich V *et al* 2014 Comparing impacts of climate change on streamflow in four large African river basins *Hydrol. Earth Syst. Sci.* **18** 1305–21
Alfieri L *et al* 2017 Global projections of river flood risk in a warmer world *Earths Future* **5** 171–82
Alfieri L, Dottori F, Betts R, Salamon P and Feyen L 2018 Multi-model projections of river flood risk in Europe under global warming *Climate* **6** 16
Arnell N W and Gosling S N 2016 The impacts of climate change on river flood risk at the global scale *Clim. Change* **134** 387–401
Bartholome E and Belward A S 2005 GLC2000: a new approach to global land cover mapping from Earth observation data *Int. J. Remote Sens.* **26** 1959–77
Bloschl G *et al* 2017 Changing climate shifts timing of European floods *Science* **357** 588–90
Bosshard T *et al* 2013 Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections *Water Resour. Res.* **49** 1523–36
Bronstert A *et al* 2007 Multi-scale modelling of land-use change and river training effects on floods in the Rhine basin *River Res. Appl.* **23** 1102–25
Calton B, Schellekens J and Martinez-de la Torre A 2016 Water Resource Reanalysis v1: Data Access and Model Verification Results (Version v1.02). Zenodo (https://doi.org/10.5281/zenodo.57760)

Coles S 2001 *An Introduction to Statistical Modeling of Extreme Values* (London: Springer)
Collins W J *et al* 2011 Development and evaluation of an Earth-System model-HadGEM2 *Geosci. Model. Dev.* **4** 1051–75
Dee D P *et al* 2011 The ERA-Interim reanalysis: configuration and performance of the data assimilation system *Q. J. R. Meteor. Soc.* **137** 553–97
Donnelly C *et al* 2017 Impacts of climate change on European hydrology at 1.5, 2 and 3 degrees mean global warming above preindustrial level *Clim. Change* **143** 13–26
Dottori F *et al* 2018 Increased human and economic losses from river flooding with anthropogenic warming *Nat. Clim. Change* **8** 781
Dufresne J L *et al* 2013 Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5 *Clim. Dyn.* **40** 2123–65
Dunne J P *et al* 2012 GFDL's ESM2 global coupled climate-carbon earth system models: I. Physical formulation and baseline simulation characteristics *J. Clim.* **25** 6646–65
FAO/IIASA/ISRIC/ISS-CAS/JRC 2009 Harmonized World Soil Database (version 1.1). (Rome: FAO)(Laxenburg, Austria: IIASA)
Frieler K *et al* 2017 Assessing the impacts of 1.5 degrees C global warming - simulation protocol of the inter-sectoral impact model intercomparison project (ISIMIP2b) *Geosci. Model. Dev.* **10** 4321–45
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 degrees C, 2 degrees C and 3 degrees C *Clim. Change* **141** 577–95
GRDC 2018 The World-Wide Repository of river Discharge Data and Associated Metadata 56068 (Koblenz, Germany: Federal Institute of Hydrology (BfG))
Guha-Sapir D, Hoyois P H, Wallemacq P and Below R 2017 *Annual Disaster Statistical Review 2016 The Numbers and Trends.* (Brussels: CRED)
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
Hall J *et al* 2014 Understanding flood regime changes in Europe: a state-of-the-art assessment *Hydrol. Earth Syst. Sci.* **18** 2735–72
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. Change* **141** 561–76
Hempel S, Frieler K, Warszawski L, Schewe J and Piontek F 2013 A trend-preserving bias correction - the ISI-MIP approach *Earth Syst. Dyn.* **4** 219–36
Hipel K W and McLeod A I 1994 *Time Series Modelling of Water Resources and Environmental Systems* (Amsterdam: Elsevier) p 1012
Hirabayashi Y *et al* 2013 Global flood risk under climate change *Nat. Clim. Change* **3** 816–21
Huang S C *et al* 2017 Evaluation of an ensemble of regional hydrological models in 12 large-scale river basins worldwide *Clim. Change* **141** 381–97
Hurkmans R T W L, De Moel H, Aerts J C J H and Troch P A 2008 Water balance versus land surface model in the simulation of Rhine river discharges *Water Resour. Res.* **44** W01418
Intergovernmental Panel on Climate Change (IPCC) 2012 *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change* (New York: Cambridge University Press)
James R and Washington R 2013 Changes in African temperature and precipitation associated with degrees of global warming *Clim. Change* **117** 859–72
James R, Washington R, Schleussner C F, Rogelj J and Conway D 2017 Characterizing half-a-degree difference: a review of methods for identifying regional climate responses to global warming targets *Wires Clim. Change* **8** e457
Jones P W 1999 First- and second-order conservative remapping schemes for grids in spherical coordinates *Mon. Weather Rev.* **127** 2204–10
Krysanova V *et al* 2018 How the performance of hydrological models relates to credibility of projections under climate change *Hydrol. Sci. J.* **63** 696–720
Krysanova V and Hattermann F F 2017 Intercomparison of climate change impacts in 12 large river basins: overview of methods and summary of results *Clim. Change* **141** 363–79
Krysanova V, Mueller-Wohlfeil D I and Becker A 1998 Development and test of a spatially distributed hydrological/water quality model for mesoscale watersheds *Ecol. Modell.* **106** 261–89
Kumar R, Samaniego L and Attinger S 2013 Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations *Water Resour. Res.* **49** 360–79
Kundzewicz Z W *et al* 2014 Flood risk and climate change: global and regional perspectives *Hydrol. Sci. J.* **59** 1–28
Kundzewicz Z W *et al* 2017 Differences in flood hazard projections in Europe - their causes and consequences for decision making *Hydrol. Sci. J.* **62** 1–14
Lange S 2016 EartH2Observe, WFDEI and ERA-Interim data Merged and Bias-corrected for ISIMIP (EWEMBI), GFZ Data Services (https://doi.org/10.5880/pik.2016.004)
Liang X, Lettenmaier D P, Wood E F and Burges S J 1994 A simple hydrologically based model of land surface water and energy fluxes for general circulation models *J. Geophys. Res.* **99** 14415–28
Madsen H, Lawrence D, Lang M, Martinkova M and Kjeldsen T R 2014 Review of trend analysis and climate change projections of extreme precipitation and floods in Europe *J. Hydrol.* **519** 3634–50
Mallakpour I and Villarini G 2015 The changing nature of flooding across the central United States *Nat. Clim. Change* **5** 250–4
Marx A *et al* 2018 Climate change alters low flows in Europe under global warming of 1.5, 2, and 3 °C *Hydrol. Earth Syst. Sci.* **22** 1017–32
Mitchell D *et al* 2016 Realizing the impacts of a 1.5 degrees C warmer world *Nat. Clim. Change* **6** 735–7
Nash I E and Sutcliffe I V 1970 River flow forecasting through conceptual models part I—A discussion of principles *J. Hydrol.* **10** 282–90
Niger Basin Authority 2008 NIGER-HYCOS http://nigerhycos.abn.ne
Peel M C, Finlayson B L and McMahon T A 2007 Updated world map of the Koppen–Geiger climate classification *Hydrol. Earth Syst. Sci.* **11** 1633–44
Petrow T and Merz B 2009 Trends in flood magnitude, frequency and seasonality in Germany in the period 1951–2002 *J. Hydrol.* **371** 129–41
Samaniego L, Kumar R and Attinger S 2010 Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale *Water Resour. Res.* **46** W05523
Samaniego L *et al* 2017 Propagation of forcing and model uncertainties on to hydrological drought characteristics in a multi-model century-long experiment in large river basins *Clim. Change* **141** 435–49
Samaniego L *et al* 2018 Anthropogenic warming exacerbates European soil moisture droughts *Nat. Clim. Change* **5** 1117–21
Snedecor G W and Cochran G W 1989 *Statistical Methods* 8th ed. (Ames, IA: Iowa State University Press)
Stackhouse P W *et al* 2011 24.5 year SRB data set released *GEWEX News* (http://www.gewex.org/gewex-content/files_mf/1432209318Feb2011.pdf) **21** 10–2
Taylor K E, Stouffer R J and Meehl G A 2012 An overview of CMIP5 and the experiment design *Bull. Am. Meteorol. Soc.* **93** 485–98
Thober S *et al* 2018 Multi-model ensemble projections of European river floods and high flows at 1.5, 2, and 3 degrees global warming *Environ. Res. Lett.* **13** 014003
US Geological Survey 2016 Explanations for the National Water Conditions https://water.usgs.gov/nwc/explain_data.html

United Nations Framework Convention on Climate Change (UNFCCC) 2015 Paris Climate Change Conference – November 2015, COP 21 Adopt. Paris Agreement (https://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf)

Vetter T *et al* 2015 Multi-model climate impact assessment and intercomparison for three large-scale river basins on three continents *Earth Syst. Dyn.* **6** 17–43

Vetter T *et al* 2017 Evaluation of sources of uncertainty in projected hydrological changes under climate change in 12 large-scale river basins *Clim. Change* **141** 419–33

Ward P J *et al* 2015 Usefulness and limitations of global flood risk models *Nat. Clim. Change* **5** 712–5

Warszawski L *et al* 2014 The Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP): Project framework *Proc. Natl Acad. Sci. USA* **111** 3228–32

Watanabe M *et al* 2010 Improved Climate Simulation by MIROC5. Mean States, Variability, and Climate Sensitivity *J. Clim.* **23** 6312–35

Weedon G P *et al* 2014 The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data *Water Resour. Res.* **50** 7505–14

Willner S N, Levermann A, Zhao F and Frieler K 2018 Adaptation required to preserve future high-end river flood risk at present levels *Sci. Adv.* **4** eaao1914

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

Yang T *et al* 2014 Climate change and probabilistic scenario of streamflow extremes in an alpine region *J. Geophys. Res. Atmos.* **119** 8535–51