LETTER • OPEN ACCESS

Intercomparison of regional-scale hydrological models and climate change impacts projected for 12 large river basins worldwide—a synthesis

To cite this article: Valentina Krysanova *et al* 2017 *Environ. Res. Lett.* **12** 105002

View the article online for updates and enhancements.

Related content

Environmental Research Letters

LETTER

CrossMark

Intercomparison of regional-scale hydrological models and climate change impacts projected for 12 large river basins worldwide—a synthesis

Valentina Krysanova[1,19], Tobias Vetter[1], Stephanie Eisner[2,3], Shaochun Huang[1,4], Ilias Pechlivanidis[5], Michael Strauch[6], Alexander Gelfan[7], Rohini Kumar[6], Valentin Aich[1,8], Berit Arheimer[5], Alejandro Chamorro[9], Ann van Griensven[10,11], Dipangkar Kundu[12], Anastasia Lobanova[1], Vimal Mishra[13], Stefan Plötner[14], Julia Reinhardt[1], Ousmane Seidou[15], Xiaoyan Wang[16], Michel Wortmann[1,17], Xiaofan Zeng[18] and Fred F Hattermann[1]

1   Potsdam Institute for Climate Impact Research (PIK), Telegraphenberg, A31, 14473, Potsdam, Germany
2   Norwegian Institute of Bioeconomy Research, PO Box 115, 1431 Ås, Norway
3   Center for Environmental Systems Research, University of Kassel, Wilhelmshöher Allee 47, 34109 Kassel, Germany
4   The Norwegian Water Resources and Energy Directorate (NVE), PO Box 5091, Majorstua, 0301 Oslo, Norway
5   Swedish Meteorological and Hydrological Institute, Norrköping, SE-601 76, Sweden
6   UFZ—Helmholtz Center for Environmental Research, Permoserstr. 15, 04318 Leipzig, Germany
7   Water Problems Institute of RAS, 3 Gubkina Str., 119333 Moscow, Russia
8   World Meteorological Organization, 7 bis Avenue de la Paix, 1211 Geneva, Switzerland
9   Justus-Liebig-University Giessen, Heinrich-Buff-Ring 26, 35392, Giessen, Germany
10 Vrije Universiteit Brussel (VUB), Brussel, Belgium
11 UNESCO-IHE Institute for Water Education, Delft, Netherlands
12 Centre for Carbon, Water and Food, Sydney Institute of Agriculture, The University of Sydney, 1 Central Avenue, Eveleigh 2205, NSW, Australia
13 Department of Civil Engineering, Indian Institute of Technology (IIT) Gandhinagar, Palaj, Gandhinagar, Gujarat, 382355, India
14 Institute of Water Resources Management, Leibniz University of Hannover, Appelstr. 9A, 30167 Hannover, Germany
15 Department of Civil Engineering, University of Ottawa, 161 Louis Pasteur Office A113, Canada
16 State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Center for Global Change and Water Cycle, Hohai University, Nanjing 210098, People's Republic of China
17 Department of Geography, University College London, Gower Street, London WC1E 6BT, United Kingdom
18 School of Hydropower & Information Engineering, Huazhong University of Science and Technology, 1037 Road Luoyu, Hongshan District, Wuhan 430074, People's Republic of China
19 Author to whom any correspondence should be addressed.

E-mail: krysanova@pik-potsdam.de

Supplementary material for this article is available online

Abstract

An intercomparison of climate change impacts projected by nine regional-scale hydrological models for 12 large river basins on all continents was performed, and sources of uncertainty were quantified in the framework of the ISIMIP project. The models ECOMAG, HBV, HYMOD, HYPE, mHM, SWAT, SWIM, VIC and WaterGAP3 were applied in the following basins: Rhine and Tagus in Europe, Niger and Blue Nile in Africa, Ganges, Lena, Upper Yellow and Upper Yangtze in Asia, Upper Mississippi, MacKenzie and Upper Amazon in America, and Darling in Australia. The model calibration and validation was done using WATCH climate data for the period 1971–2000. The results, evaluated with 14 criteria, are mostly satisfactory, except for the low flow. Climate change impacts were analyzed using projections from five global climate models under four representative concentration pathways. Trends in the period 2070–2099 in relation to the reference period 1975–2004 were evaluated for three variables: the long-term mean annual flow and high and low flow percentiles Q_{10} and Q_{90}, as well as for flows in three months high- and low-flow periods denoted as HF and LF. For three river basins: the Lena, MacKenzie and Tagus strong trends in all five variables were found (except for Q_{10} in the MacKenzie); trends with moderate certainty for three to five variables were confirmed for the Rhine, Ganges and Upper Mississippi; and increases in HF and LF were found for the Upper Amazon, Upper Yangtze and Upper Yellow. The analysis of projected streamflow seasonality demonstrated

increasing streamflow volumes during the high-flow period in four basins influenced by monsoonal precipitation (Ganges, Upper Amazon, Upper Yangtze and Upper Yellow), an amplification of the snowmelt flood peaks in the Lena and MacKenzie, and a substantial decrease of discharge in the Tagus (all months). The overall average fractions of uncertainty for the annual mean flow projections in the multi-model ensemble applied for all basins were 57% for GCMs, 27% for RCPs, and 16% for hydrological models.

1. Introduction

A rigorous quantification of climate change impacts in the water sector under different radiative forcing scenarios and levels of global warming is necessary for creating appropriate adaptation policies and strategies. It is usually done by driving global or regional climate models (GCMs or RCMs) with scenarios of future radiative forcing (representative concentration pathways, RCPs). Climate model outputs are usually bias-corrected to match observed or reanalysis climate data in the historical period. The resulting climate datasets are then used to drive hydrological models (HMs) to provide an assessment of expected changes (see methodology description in Krysanova *et al* 2016, Olsson *et al* 2016).

In the last decade numerous impact studies used ensembles of climate scenarios but only one impact model, and recently also sets of impact models started to be applied. Previously, intercomparison of impacts using multiple HMs has been done for the water sector applying mainly global hydrological models (e.g. Haddeland *et al* 2014, Dankers *et al* 2014, Schewe *et al* 2014), and studies using regional-scale models have also appeared (e.g. Vetter *et al* 2015). The study focuses were different: Haddeland *et al* (2014) analysed and compared climate change and direct human impacts on the water cycle; Dankers *et al* (2014) studied potential climate impacts on flood hazards, and indicated large uncertainties and disagreements even on the sign of change for some individual river basin; Schewe *et al* (2014) analysed water resources and water scarcity in a warmer world; and Vetter *et al* (2015) studied impacts on mean discharge and extremes and evaluated related uncertainties for three large river basins on three continents.

The fundamental differences between the global and regional (or basin-scale) HMs are their low and fine spatial resolutions, respectively, including the resolution of input data, and their approaches to calibration/validation: the global HMs are usually not calibrated, whereas for the regional HMs calibration is a must. The global-scale modelling results are often considered as not credible at the river basin scale (Dankers *et al* 2014, Kundzewicz *et al* 2017, Hattermann *et al* 2017), where the impacts actually happen, and where adaptation strategies should be designed and applied. Their low credibility at the basin scale is mainly due to poor performance in the historical

period, often contradicting change signals and large uncertainties of projections.

Therefore, our study aims to narrow the gap by providing more robust and credible climate impact results for the regional scale using *calibrated and validated basin-scale models*. Namely, the purpose is to provide a comprehensive intercomparison of impacts simulated by nine state-of-the-art regional-scale hydrological models driven by an ensemble of up-to-date climate scenarios from five GCMs for 12 large river basins located on all continents. The multi-model framework is then used to quantify sources of uncertainties in the ensemble. The obtained results could be used for developing adaptive management strategies.

The following specific objectives were pursued: (a) the evaluation of performance of HMs in the historical period, (b) the quantitative assessment of climate change impacts on mean river discharge and extremes looking for robust trends, and (c) the evaluation of uncertainties from three major sources: RCPs, GCMs and HMs. The study also allowed detecting weaknesses of climate and hydrological models in specific regions or for some variables, substantially contributing to uncertainties in the projections.

The analysis was performed in the framework of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) (www.isi-mip.org; Warszawski *et al* 2014) by an international team of regional-scale hydrological modellers. The detailed results on intercomparison of climate change impacts on river discharge, evapotranspiration and extremes for all 12 or a subset of river basins are presented in papers of a special issue (SI) recently published in *Climatic Change* (see editorial paper, Krysanova and Hattermann 2017). The intention of this synthesis paper is to provide a summary of major results obtained based on the SI papers and analysis performed beyond.

2. Data and methods

2.1. Hydrological models

In this study eight basin-scale hydrological models: ECOMAG (Motovilov *et al* 1999, 2013), HBV (Bergström and Forsman 1973), HYMOD (Boyle 2001), HYPE (Lindström *et al* 2010), mHM (Samaniego *et al* 2010, Kumar *et al* 2013), SWAT (Arnold *et al* 1998), SWIM (Krysanova *et al* 1998) and VIC (Liang *et al* 1994), and WaterGAP3

Table 1. Overview of main characteristics of 12 river basins and modelling case studies (X) performed with hydrological models. Sign (X + X) means that the same model was applied twice, by two modelling groups. The average temperature and precipitation are estimated from the WATCH data (Weedon *et al* 2011) in the period 1971–2000.

Basin	Rhine	Tagus	Niger		Blue Nile		Ganges	Yellow	Yangtze	Lena	Darling	MacKenzie	Mississippi	Amazon
Continent	Europe	Europe	Africa		Africa		Asia	Asia	Asia	Asia	Australia	N. America	N. America	S. America
Gauge	Lobith	Almourol	Lokoja	Koulikoro	Khartoum	El Deim	Farakka	Tangnaiha	Cuntan	Stolb	Louth	Ar. Red River	Alton	SP Olivenca
Drainage area, th. km2	161	67.5	2074	120	325	176	835	121	867	2460	489	1660	444	991
Av. Altitude, m a.s.l.	497	471	415	462	1295	1857	1060	4125	2903	620	314	1437	305	1131
Av. Temperature, deg C	8.7	14	27.7	26.5	23.3	21	21.1	-2	6.8	-10.2	19.2	-4.3	7.3	21.7
Av. Precipitation, mm/yr	1038	671	625	1495	1067	1300	1173	506	768	384	590	435	967	2122
Runoff coefficient	0.44	0.23	0.12	0.18	0.11	0.21	0.40	0.33	0.51	0.52	0.01	0.39	0.27	0.69
VIC	X	X	X	X	X	X	X	X	X	X	X		X	X
SWIM	X	X	X	X	X	X	X	X	X	X			X	X
WaterGAP3	X	X	X	X	X	X	X	X		X			X	X
mHM	X			X	X	X	X	X			X		X	X
HYMOD	X + X			X	X		X + X	X			X		X + X	X
HBV	X + X	X		X		X	X	X	X				X	X
SWAT				X		X			X		X		X	X
HYPE	X	X	X				X			X		X		
ECOMAG										X		X		
Case studies	9	5	4	7	5	6	8	6	4	5	4	2	8	7

(Verzano 2009) suitable for both global and regional scales were applied. Three models (VIC, mHM and WaterGAP3) were applied to grid cells with sub-grid heterogeneity accounting methods, five models (ECOMAG, HBV, HYPE, SWAT and SWIM) disaggregated basins into subbasins and hydrological response units based on topography, land use and soil classes, and one model (HYMOD) was lumped. The models used two to six climate parameters as input (see table 2 in Krysanova and Hattermann (2017) for more details).

Information on the modelling protocol including model descriptions can be found in Krysanova and Hattermann (2017).

2.2. River basins

Twelve large river basins located on six continents (table 1) were selected for intercomparisons in this study. The Tagus basin is the smallest, while the Niger and Lena are the largest. Due to complex geomorphological structures and numerous anthropogenic alterations in the Amazon, Mississippi, Yangtze and Yellow, only the less human-influenced upper parts of the basins were considered in this study. For simplicity, we will omit 'Upper' in the names of these basins later in the text. For the Niger and Blue Nile, two gauges were considered.

The study basins cover a range of geographical zones considering climate, topography and continental distribution. Five basins (Amazon, Lena, MacKenzie, Yellow and Yangtze) are characterized by prevailing natural land cover: forest and/or grassland (≥ 66%). A substantial share of cropland (38–65%) can be found in other five basins (Ganges, Blue Nile, Tagus, Mississippi and Rhine), crops and grassland occupy a half of the Niger drainage area, and 44% of the Darling basin is covered by pastures and rangeland.

The basins are located in different climate zones: from tropical wet (Amazon) and humid subtropical (Ganges) to Mediterranean (Tagus) and semiarid (Darling), and from temperate (Rhine and Mississippi) to highlands (Yellow) and subarctic (MacKenzie and Lena). The average annual temperature exceeds 20 °C in four basins, and it is below zero in three basins. The average annual precipitation is the highest in Amazon (> 2000 mm), and it is below 500 mm in two Arctic basins.

Three of the basins (Darling, Blue Nile and Niger) are characterized by a relatively low annual average runoff with runoff coefficients ≤ 0.12, while the Amazon, Ganges and Rhine have the highest runoff coefficients (table 1). The diversity of meteorological and runoff characteristics in the selected basins confirms that they represent a variety of climatic and runoff generation conditions of the globe.

In total, 80 modelling case studies (table 1) were used in both model evaluation and comparison of impacts and uncertainties. Due to restricted resources, it was not possible to apply every model to every basin. The impact assessment was driven by outputs from five GCMs available for four RCPs, i.e. 20 hydrological model runs were performed for every case study in table 1, leading in total to 1600 time series for the analysis.

2.3. Data

Mostly common sources of geospatial data across river basins and models were used, with some variation between models and regions. The global digital elevation model (DEM) constructed from the Shuttle Radar Topography Mission (http://srtm.csi.cgiar.org/) at 3 arc seconds resolution (∼90 m) was used for ten basins, except the Lena and MacKenzie. For the latter two a hydrologically adjusted DEM from USGS (Hydro 1K, https://lta.cr.usgs.gov/HYDRO1K) was applied. The Global Land Cover 2000 map (GLC, http://bioval.jrc.ec.europa.eu/products/glc2000/products.php) produced by the EC Joint Research Centre with 22 land cover types was used. Soil parametrization was done using data from the Harmonized World Soil Database (www.cnrm.meteo.fr/gmme/PROJETS/ECOCLIMAP) at 1 km resolution

and the Digital Soil Map of the World (www.fao.org/waicent/FaoInfo/Agricult/AGL/AGLL/dsmw.htm) based on the FAO/UNESCO Soil Map of the World.

The HMs were driven by the daily WATCH forcing data (Weedon *et al* 2011) with 0.5° × 0.5° resolution for their evaluation in the historical period 1971–2000. For the Amazon, where a systematic underestimation of precipitation was found, a correction method accounting for high resolution climatologies and cloud water interception was developed and suggested for future studies (Strauch *et al* 2017).

The observed daily (for the Lena, Amazon, Darling, Mississippi, Rhine, Niger, MacKenzie and Tagus) or monthly (Ganges, Blue Nile) discharge data from the Global Runoff Data Centre (GRDC), and daily data from national sources (for the Yellow and Yangtze: from China Hydrological Yearbooks), where GRDC data were not available, were used for comparison with the simulated discharge, mostly in the period 1971–2000. For the Blue Nile and Mississippi, shorter time series were available and used for model evaluation, and for the Ganges (monthly discharge data available for 1949–1973 only) the evaluation period was shifted to 1961–1973. In most cases, human influences were not considered. WaterGAP3 was applied with and without consideration of human water management, but for the intercomparison only model runs without management were used, for consistency.

2.4. Climate projections

The climate model data originate from the Coupled Model Intercomparison project (CMIP5, Taylor *et al* 2012, http://cmip-pcmdi.llnl.gov/cmip5). The outputs of five GCMs: GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR, MIROC-ESM-CHEM, and NorESM1-M (see references in the supplementary material available at stacks.iop.org/ERL/12/105002/mmedia) simulated under four RCPs: RCP2.6, RCP4.5, RCP6.0 and RCP8.5 were selected in the ISIMIP project to represent the full range of changes in global mean temperature and precipitation in CMIP5 (see Warszawski *et al* 2014 for more details). The climate projections were bias-corrected to match the WATCH forcing data using a trend-preserving method (Hempel *et al* 2013). Climate statistics and (later) simulation outputs for the reference period (1981–2010) were compared with those for three scenario periods: near-future (2006–2035), mid-century (2036–2065) and end-century (2070–2099). For analysis of trends the reference period was 1975–2004.

The evaluation of the climate model projections for our basins is briefly described in the supplementary material.

2.5. Analysis methods

2.5.1. Evaluation of hydrological model performance

Fourteen numerical criteria (see table A1 in supplementary) were selected to assess HM performance depending of the simulated variable under consideration:

- monthly hydrograph: Nash–Sutcliffe efficiency (NSE: Nash and Sutcliffe, 1970), the modified Kling–Gupta efficiency (KGE: Kling *et al* 2012), volumetric efficiency (VE: Criss and Winston, 2008) and percent bias in discharge (PBIAS);
- long-term mean seasonal dynamics (or the annual cycle of discharge): the Pearson's correlation coefficient (r) and relative difference in standard deviation;
- flow duration curves (FDC): percent biases in FDC mid-segment slope, high-segment volume (corresponding to the highest 2, 5 and 10% of flow) and low-segment volume (corresponding to the lowest 30% of flow, related to baseflow);
- extreme flows: percent biases of 10- and 30 year flood and low flow return intervals (ΔFlood and ΔLF) obtained by fitting the generalized Pareto distribution (Coles 2001) to the peaks over threshold (high flow) or by fitting the generalized extreme value distribution (Coles 2001; Huang *et al* 2013) to the annual minimum 7 day mean flows; and the NSE criterion on inverse flows (NSEIQ) for the low flow evaluation.

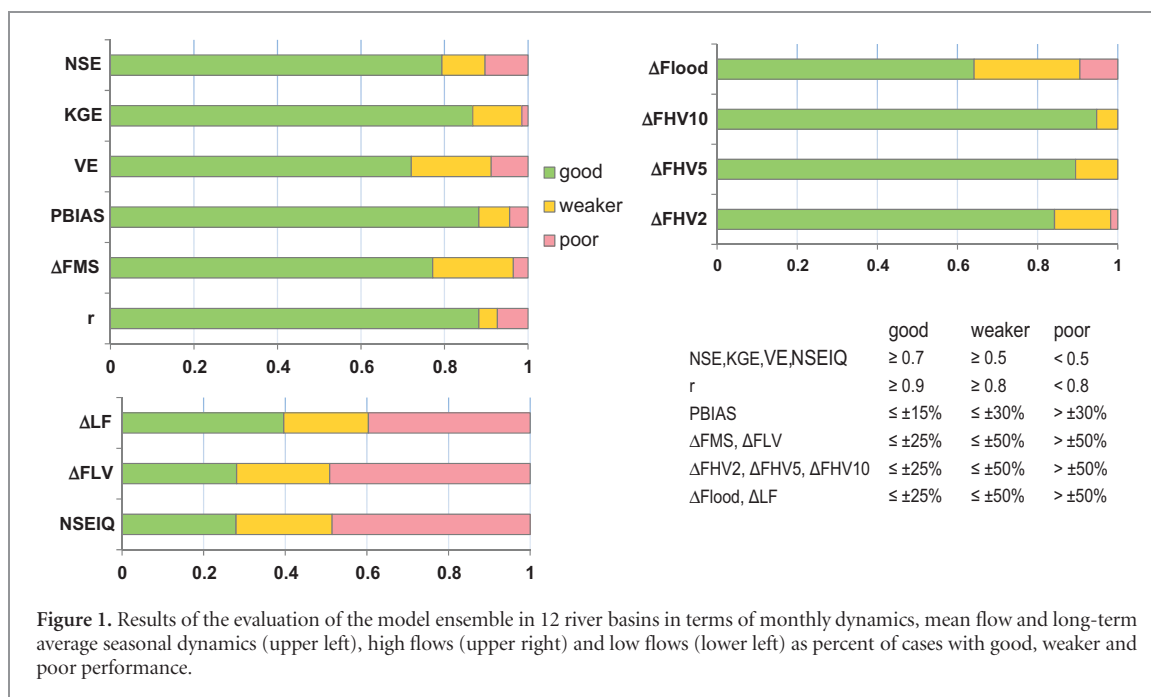### 2.5.2. Evaluation of seasonal dynamics

The simulated long-term average daily river discharge was analysed for three periods: reference, mid-century and end-century. The relative changes between the reference and future periods were calculated using simulations driven by the same GCM.

The seasonal dynamics were analysed qualitatively for changes between periods and RCPs, as well as quantitatively for spreads (or variability) and seasonal shifts. The mean relative spreads (in %) were calculated as $(Q_{25} - Q_{75})/Q_{50}*100$, where Q_{50}, Q_{25} and Q_{75} are runoff quantiles, averaged over all days of the year and compared between basins and time periods. Temporal shifts of the high-flow season were determined by finding the 14 day period with the highest discharge volume based on the ensemble mean, and by comparing the mid-day of the 14 day period for the reference and end-century periods under RCPs 2.6 and 8.5.

2.5.3. Analysis of projected changes in mean, high and low flows

Changes in the projected runoff were analysed for the annual mean flow (MF) and two annual runoff quantiles representing high flow (Q_{10}) and low flow (Q_{90}), as well as for three months high- and low-flow periods (denoted as HF and LF).

The trends in MF, Q_{10} *and* Q_{90} were evaluated statistically between the median of 30 annual values of each variable in the reference period (1975–2004) and medians of the future 30 years periods starting in 2008, considering 63 future periods in total, the first being 2008–2037 and the last 2070–2099. The statistical

Figure 1. Results of the evaluation of the model ensemble in 12 river basins in terms of monthly dynamics, mean flow and long-term average seasonal dynamics (upper left), high flows (upper right) and low flows (lower left) as percent of cases with good, weaker and poor performance.

significance of trend was estimated at the 0.05 significance level with the Wilcoxon signed-rank test using the *R* statistical software. The analysis was performed separately for each basin, three variables, RCP scenario, driving GCM and HM.

To analyse changes in HF and LF, the 30 day moving averages of runoff (MAR) of the long-term average dynamics in the reference and end-century periods were calculated for every model run. Maximum and minimum of MAR (MAR_{max} and MAR_{min}) were found, and 90 d high/low flow periods centred around MAR_{max} and MAR_{min} were extracted. The average monthly MAR values in these high/low flow periods, denoted as HF and LF values, were compared between the end-century and reference periods, and percent changes were calculated. For every basin and RCP, $N \times 5 \times 3$ values of percentage changes were obtained for HF and LF, where N is the number of HMs applied for this basin, and 5 and 3 correspond to 5 GCMs and 3 months. This allowed estimating (a) shares of positive and negative changes, and (b) shares of cases exceeding ± 5% change. In case the shares were higher than thresholds of 0.65 and 0.75, correspondingly, we could state that increase or decrease in HF or LF is projected (in the latter case certainty is higher). This analysis was performed for each basin, two variables and four RCP scenarios.

2.5.4. Uncertainty analysis
Three sources of uncertainty (from RCPs, GCMs and HMs) in the projected annual mean flow and two annual runoff quantiles Q_{10} and Q_{90} were evaluated using the ANOVA method (Bosshard *et al* 2013) by splitting variances into the contributing sources and interaction terms. As these three factors have different sample sizes (e.g. for the Amazon: five GCMs, four

RCPs and seven HMs), a subsampling was used to avoid biases. More details can be found in Vetter *et al* (2017).

The uncertainty related to input data (topography, land use, soil, etc.) was not accounted for in this study, but could be recommended for future studies.

3. Results and discussion

3.1. Evaluation of models' performance
The evaluation of nine HMs was done using the performance criteria described in section 2.5.1 for all basins. Note that only two of these criteria (NSE, PBIAS) were used for calibration of the models. The aggregated results for criteria targeted on monthly dynamics (NSE, KGE, VE, PBIAS), mean flow (ΔFMS), long-term average seasonal dynamics (r), high flows (ΔFlood, ΔFHV10, ΔFHV5, ΔFHV2) and low flows (ΔLF, ΔFLV, NSEIQ) are shown as percent of all simulated cases with a good, moderate and poor model performance in figure 1, separately for every criterion. More detailed results for single models and criteria, on which these aggregated results are based on, can be found in Huang *et al* (2017). The line charts of the long-term average seasonal dynamics simulated by our models in 12 basins can be seen in Huang *et al* (2017), and, in comparison with the global model outputs, in Hattermann *et al* (2017).

The model performance for monthly dynamics is quite good: according to three of four criteria, more than 80% of simulations are above the 'good' threshold. The same can be stated for seasonal dynamics: coefficient of correlation is above 0.9 for 88% of all model runs, and bias in standard deviation (not shown in figure 1) is below 20% in 72% of all simulations. The results for high flows are also satisfactory, especially for the high-segment volume of FDC
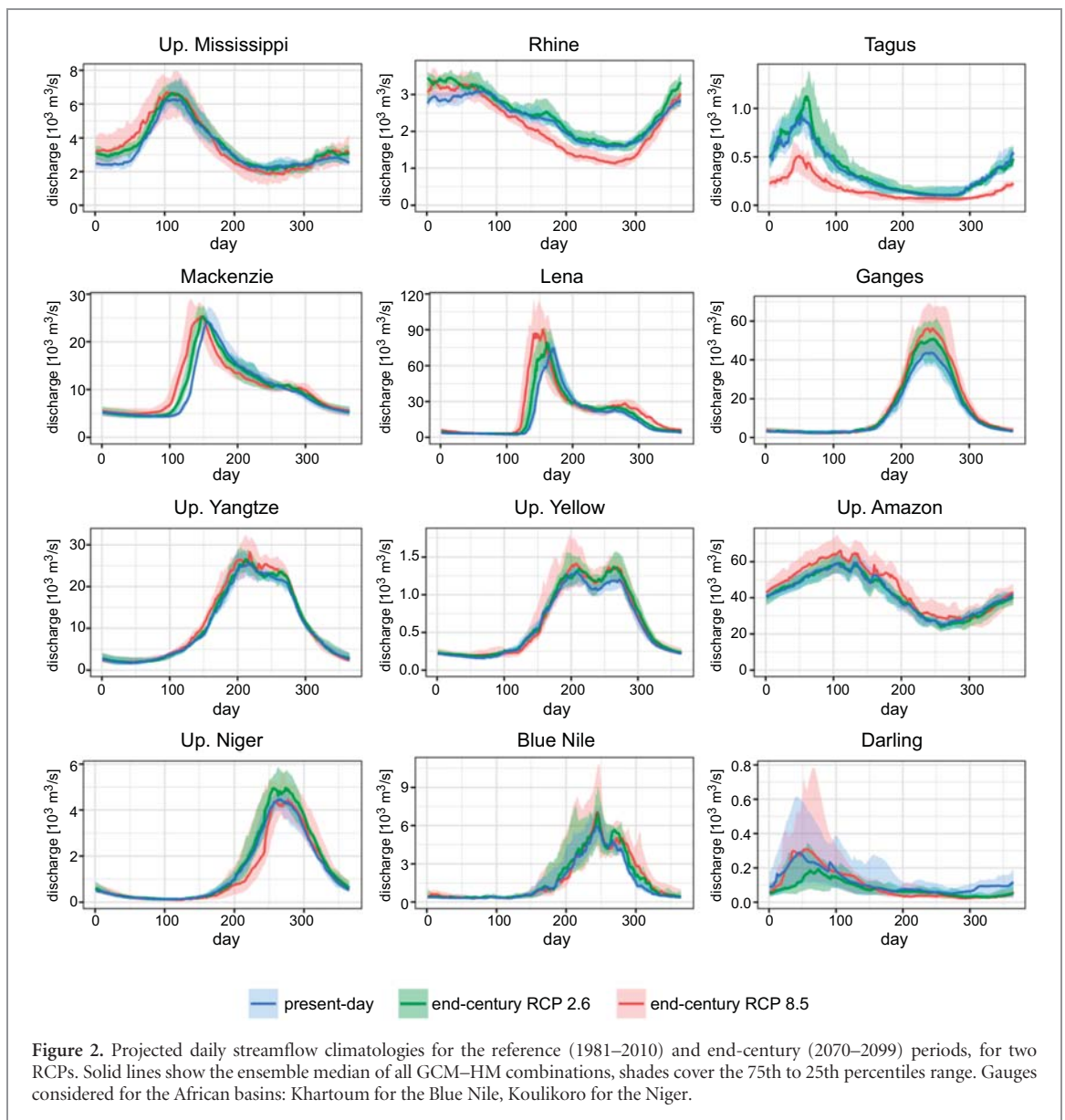
Figure 2. Projected daily streamflow climatologies for the reference (1981–2010) and end-century (2070–2099) periods, for two RCPs. Solid lines show the ensemble median of all GCM–HM combinations, shades cover the 75th to 25th percentiles range. Gauges considered for the African basins: Khartoum for the Blue Nile, Koulikoro for the Niger.

corresponding to the highest 2, 5 and 10% of flow, and a slightly weaker performance for extreme floods. However, the simulated low flows show higher biases, and 40%–50% of all simulations are in the 'poor' range, indicating the need for improving model structure and parameterization in this respect.

3.2. Impacts on seasonal dynamics
The projected streamflow seasonality was analysed qualitatively and quantitatively. Figure 2 shows the annual cycle of streamflow with the *daily* time step for the reference period and projections for two RCPs in the end-century, and analysis based on it follows.

When comparing ensemble median streamflow under RCP8.5 at end-century with that in the reference period, the following patterns emerge:

- snowmelt flood peaks are amplified and shifted to earlier dates in the Lena and MacKenzie, accompanied by lower runoff levels in summer;

- streamflow volumes in high-flow period increase in the Amazon, Ganges, Yangtze and Yellow;

- streamflow volumes decrease in the Tagus (all months), and during the high flow onset in the Niger;

- partial sub-seasonal increases and decreases are observed for the Mississippi and Rhine;

- only minor changes occur in the Darling and Blue Nile.

There are substantial differences in median streamflow climatology between the two RCPs in eight of 12 cases: Rhine, Tagus, Niger, Lena, MacKenzie, Ganges, Amazon and Darling (figure 2).

The mean relative spreads (or variabilities, see definition in section 2.5.2) are lower than 30% for the reference period and under RCP2.6 in the end-century for the Rhine (the lowest), Amazon, MacKenzie, Yellow and Mississippi; they range from 40% to 58% for the Yangtze, Lena and Ganges; range from 55% to 90% for the Tagus, Niger and Blue Nile; and exceed 100%

Table 2. Evaluation of trends in Q_{10}, mean flow (MF), Q_{90} (upper panel) and changes in three-months high and low flows (HF, LF, lower panel) by the end of the century for 12 river basins under RCPs 4.5 and 8.5. Upper panel: trends evaluated on significance for all HMs. Lower panel: if the share of outputs exceeding ±5% change is higher than 75%, this is indicated by dark blue or orange and thick arrow (interpreted as increase/decrease in HF/LF with a high certainty), and if the share of positive/negative outputs is higher than 65%, this is indicated by light blue or orange and thin arrow (interpreted as increase/decrease in HF/LF with a moderate certainty). The results in the upper panel are based on a part of table 1 in Vetter *et al* (2017), format is slightly modified. © Springer Science+Business Media, Dordrecht 2016. With permission of Springer.

Legend (symbols):
- ‡ All stat. sign. positive trends
- + At least 1 sign. positive trend and at least 1 insign. trend
- O No one trend ist stat. sign.
- ┃ At least 1 sign. negative trend and at least 1 insign. trend
- ┃┃ All stat. sign. negative trends
- × At least 1 sign. positive trend and 1 sign. negative trend

Legend (arrows):
- ← >.75 exceed +5%
- ← >.65 positive
- → >.75 exceed -5%
- → >.65 negative

for the Darling. The spreads are negatively correlated with the runoff coefficients of the basins (exponential regression, $R^2 = 0.84$), which is most probably related to the larger uncertainty of the multi-model ensemble in dry areas, which is often the case in hydrological modelling (e.g. Nicolle *et al* 2014, Donnelly *et al* 2016).

Substantial temporal shifts of the high-flow season (see section 2.5.2 for definition) caused by earlier snow melt were found in the two Arctic basins. In the Lena, the ensemble median of projected snow-melt peak advances from June 18th (reference) to June 8th under RCP2.6 and May 30th under RCP8.5 (which corresponds to a shift by nearly three weeks). For the MacKenzie a shift of about two weeks is projected, from June 6th in the reference period to May 24th under RCP8.5 (May 31st under RCP2.6). Similar results were found in other snow-dominated regions (e.g. Bergström *et al* 2001, Andréasson *et al* 2004).

According to simulation results, the spring peak will occur six days later under RCP2.6, and four days earlier under RCP8.5 in the Tagus (from February 17th to February 23rd and February 12th, respectively) attributable to changes in the temporal precipitation pattern.

The largest shift was estimated for the Rhine: from present-day March 17th to February 3rd and January 12th under RCPs 2.6 and 8.5, respectively, at the end-century. This finding is very likely related to the combined effect of increasing winter precipitation and rising winter temperatures, i.e. a large share of the surplus precipitation is not stored in the snowpack but discharged immediately. However, the estimated shift depends on the length of the chosen 'window', and the pattern of high-flow period in the Rhine (almost steady level during about 80 days in winter) indicates that the estimated large shift for the Rhine involves uncertainty.

In the Niger, the onset of the high-flow season is shifted under RCP8.5 due to a delayed onset of the rainy season. We quantified the effect by extracting the day when discharge exceeds $0.33*Q_{max}$ for the first time. This indicator is unusual, however we could not apply classical FDC indices since the Niger has a very strong seasonality with a prolonged low-flow season. The following estimates were obtained from the ensemble median: the onset of the high-flow season shifts from July 31st in present-day to August 15th under RCP8.5, indicating a two-week delay at the end-century. For the other basins no significant shifts were found.

The analyses of seasonality with the *monthly* resolution can be found in Eisner *et al* 2017.

3.3. Projected changes in mean, high and low flows

The long-term annual mean flow and quantiles Q_{10} and Q_{90} were analysed for statistically significant trends as described in section 2.5.3, and results for the end of the century 2070–2099 in relation to the reference period 1975–2004 for two RCPs are presented

in table 2 (upper panel). More detailed results on trends for the same three variables and all four RCPs can be found in Vetter *et al* (2017). Other results on trends in hydrological extremes are described in Pechlivanidis *et al* (2017).

The analysis of trends under RCP8.5 shows the following:

- robust positive trends for three variables in the Lena and for MF in the MacKenzie (Gelfan *et al* 2017); and robust negative trends for three variables in the Tagus, all - with a high certainty.

- positive trends for Q_{90} in the MacKenzie, Q_{10} and MF in the Ganges, Q_{10} in the Rhine and Mississippi; and negative trends for MF and Q_{90} in the Rhine, all—with a moderate certainty.

The same tendencies are visible under RCP4.5, but in some cases they are weaker (Tagus) or stronger (Ganges).
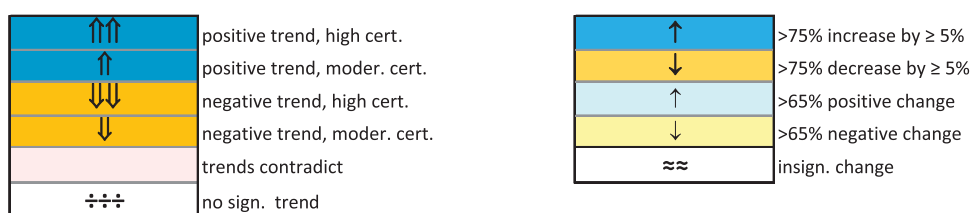
In addition, analysis of changes in discharge during the high-flow and low-flow periods of three months was performed for all basins as described in section 2.5.3, and shares of positive/negative changes and changes exceeding ±5% were calculated. The tails with larger shares are shown in table 2 (lower panel). As we see from table 2, tendencies in HF and LF mostly follow robust trends in Q_{10} and Q_{90} shown in the upper panel, but they are not identical. So, under RCP8.5 in the Amazon, Yangtze and Yellow both HF and LF increase with a moderate to high certainty (compare with results for the Yangtze in Su *et al* 2017), LF decreases in the Mississippi, and HF increases in the MacKenzie (compare with results in 3.2). In two basins, Rhine and Mississippi, runoff is projected to increase in the high-flow period, and decrease in the low-flow period under RCP8.5. The differences can be explained by the fact that statistical significance of a trend is a 'stronger' criterion, and therefore for some basins it was not found, despite of distinct changes in high/low flow periods confirmed by most of the simulation runs.

The results on trends in three variables and changes in HF and LF by the end of the century for RCP8.5 scenario are summarized in table 3, where positive and negative trends confirmed by most of models are presented. For three river basins, the Lena, MacKenzie and Tagus, strong trends in all five variables were found, except for Q_{10} in the MacKenzie; trends with moderate certainty for three to five variables were confirmed for the Rhine, Ganges and Mississippi; and positive trends in HF and LF were found for the Amazon, Yangtze and Yellow. For the Blue Nile no clear trends were identified (Teklesadik *et al* 2017), and for the Niger and Darling only potential changes in LF with a moderate certainty could be stated.

In addition, a summary on main findings in associated papers focusing on one to seven river basins published in *Climatic Change* can be found in the supplementary material. For example, they include

Table 3. Summary of results on trends* in mean flow (Q_{50}), high and low percentiles (Q_{10} and Q_{90}) and changes in three-month high and low flows by the end of this century for 12 river basins based on simulations of nine hydrological models driven by five GCMs under RCP8.5.

River basin	Statistically significant trends			Changes in high and low flows	
	Trend in Q10	Trend in Q50	Trend in Q90	3 mon. HF period	3 mon. LF period
Lena	⇑⇑⇑	⇑⇑⇑	⇑⇑⇑	↑	↑
MacKenzie		⇑⇑⇑	⇑⇑⇑	↑	↑
Ganges	⇑	⇑		↑	↓
Amazon				↑	↑
Yangtze				↑	↑
Yellow				↑	↑
Mississippi	⇑	÷÷÷	÷÷÷	↑	↓
Rhine	⇑	⇓	⇓	↑	↓
Tagus	⇓⇓⇓	⇓⇓⇓	⇓⇓⇓	↓	↓
Darling				≈≈	↓
Niger				≈≈	↑
Blue Nile				≈≈	≈≈

⇑⇑⇑	positive trend, high cert.		↑	>75% increase by ≥ 5%	
⇑	positive trend, moder. cert.		↓	>75% decrease by ≥ 5%	
⇓⇓⇓	negative trend, high cert.		↑	>65% positive change	
⇓	negative trend, moder. cert.		↓	>65% negative change	
	trends contradict		≈≈	insign. change	
÷÷÷	no sign. trend				

*Trends with the high certainty: ≥ 80% of all results are statistically significant; Trends with the moderate certainty: ≥ 80% show trends in this direction, corresponding to either ++/+ or —/– in Table 2, upper part.

analysis of flow regimes under a warmer climate using indices of hydrological alteration (Wang *et al* 2017), and a multi-model assessment of sensitivity of evapotranspiration and a proxy for available water to climate change (Mishra *et al* 2017).

3.4. Sources of uncertainty

There are three major sources of uncertainty in the projected annual mean flow and two runoff quantiles Q_{10} and Q_{90}: from RCPs, GCMs and HMs, and they were evaluated in our study as described in section 2.5.4. The obtained results for three variables ordered by fractions of uncertainty related to GCMs are presented in figure 3. It is evident that the largest fraction of uncertainty is related to driving GCMs, followed by RCPs, and the smallest fraction is related to HMs. However, contribution of hydrological models in the overall uncertainty is higher for the low flow quantile, Q_{90}, compared to other two variables (compare with similar results in Samaniego *et al* 2017 and Pechlivanidis *et al* 2017), which could be connected to rather poor model performance for low flow. The HM-related uncertainty is also quite high in the snow-dominated Upper Yellow basin (see more details in Vetter *et al* 2017 and Giuntoli *et al* 2015).

Table 4 summarizes results on the fractions of uncertainty presented in figure 3 in a qualitative form. It shows the prevailing contribution of GCMs to uncertainty in most cases (except Tagus and Lena), and explains cases of absent robust trends (Blue Nile, Niger, Darling) by a very high uncertainty due to

GCMs. Pechlivanidis *et al* (2017) further show that the uncertainty (both related to climate and hydrological models) is generally higher in the dry than in wet basins, and according to Samaniego *et al* (2017), the HM-related uncertainties cannot be neglected for hydrological drought projections.

As estimated in Vetter *et al* (2017), the overall fractions of uncertainty for the annual mean flow projections in the multi-model ensemble runs averaged over 12 basins were 57% for GCMs, 27% for RCPs, and 16% for HMs. More details on uncertainty evaluation for these 12 basins can be found in Vetter *et al* (2017).

The uncertainty due to internal climate variability was not considered here for the following reason: the analysis by Hawkins and Sutton (2009) has shown that its importance increases at shorter time scales, but for the decadal time scale and regional scale from about 2000 km the climate model uncertainty prevails over the internal climate variability.

Comparing the basin characteristics (mean precipitation, runoff coefficient, table 1) and qualitative results on trend analysis, high- and low-flow periods (table 3) and fractions of uncertainty (table 4) we can see distinct similarities between the Yangtze and Yellow (all indicators), Lena and MacKenzie (except trends in Q_{10} and dominant uncertainty sources), Rhine and Mississippi (except trends in MF and Q_{90}), as well as Niger, Blue Nile and Darling (except LF and precipitation). The remaining three basins, Tagus, Ganges and Amazon, represent single specific cases.
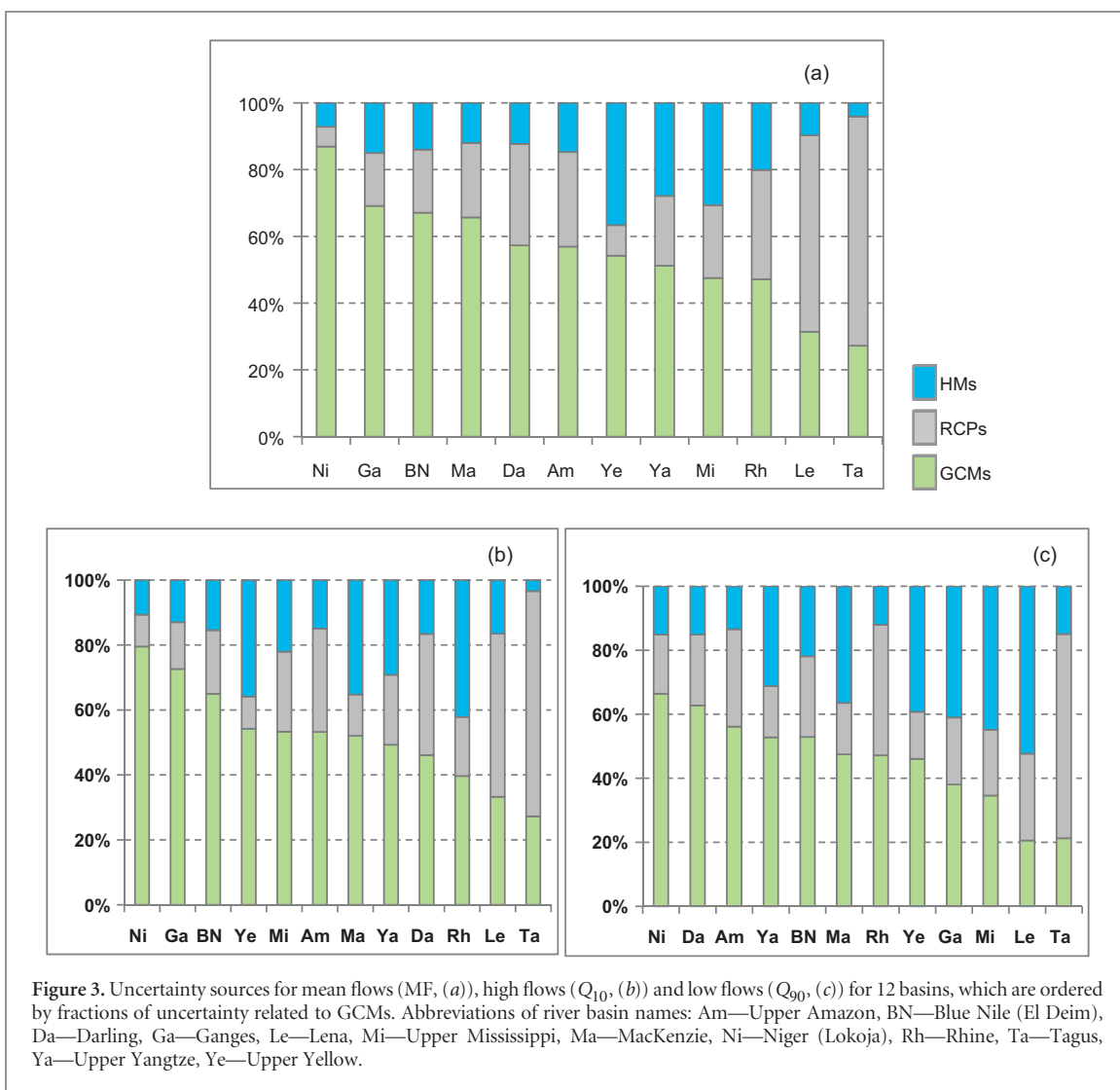
Figure 3. Uncertainty sources for mean flows (MF, (*a*)), high flows (Q_{10}, (*b*)) and low flows (Q_{90}, (*c*)) for 12 basins, which are ordered by fractions of uncertainty related to GCMs. Abbreviations of river basin names: Am—Upper Amazon, BN—Blue Nile (El Deim), Da—Darling, Ga—Ganges, Le—Lena, Mi—Upper Mississippi, Ma—MacKenzie, Ni—Niger (Lokoja), Rh—Rhine, Ta—Tagus, Ya—Upper Yangtze, Ye—Upper Yellow.

Table 4. Summary of results on evaluation of sources of uncertainty related to trends in annual mean flow (Q_{50}) and two annual runoff quantiles representing high flow and low flow (Q_{10} and Q_{90}).

River basin	Uncertainty sources								
	Q10			Q50			Q90		
	GCM	RCP	HM	GCM	RCP	HM	GCM	RCP	HM
Lena	X	XXX		X	XXXX			X	XXX
MacKenzie	XXX		XX	XXXXX			XXX		XX
Ganges	XXXXX			XXXXX			XX		XX
Amazon	XXX	X		XXXX	X		XXXX	X	
Yangtze	XXX		X	XXX		X	XXX		X
Yellow	XXX		XX	XXX		XX	XXX		XX
Mississippi	XXX	X		XXX		X	X		XXX
Rhine	XX		XX	XXX	X		XXX	XX	
Tagus	X	XXXXX		X	XXXXX			XXXX	
Darling	XXX	XX		XXXX	X		XXXX		
Niger	XXXXX			XXXXX			XXXXX		
Blue Nile	XXXXX			XXXXX			XXX	X	

XXXXX	>65%
XXXX	>55%
XXX	>45%
XX	>35%
X	>25%
	<25%

4. Summary and conclusions

This synthesis paper describes *one of the first comprehensive studies providing a multi-model inter-comparison of climate change impacts in the water sector using regionally calibrated/validated hydrological models* driven by an ensemble of climate projections for 12 large river basins worldwide. The multi-model design enabled to provide robust results for some of the basins, and helped to identify sources of uncertainty and needs for model improvement. The cases of missing robust trends (or agreement on no trend) can be explained by a very high uncertainty due to GCMs. Overall, the distribution of changes varies with the basin's hydro-climatic characteristics and climate projections. This study narrows the existing gap in knowledge by applying the multi-impact-model approach based on regional-scale calibrated and validated models for the impact assessment in large river basins on all continents.

Most applied models could adequately reproduce monthly discharge, average seasonal dynamics, moderate and high flows in the basins, but simulation of low flows appeared to be more problematic. We think that *using models after checking their performance in the historical period and selecting only models with good performance allows us to provide more robust and credible results* (Krysanova *et al* 2017), compared with an opposite case when impact models are applied without any evaluation of their performance. Though a good or satisfactory model performance does not guarantee its reliability for future projections, especially for the far future with high climate change signals, it definitely increases acceptance of results by decision makers and water managers.

Our study showed that *there is a room for improvement of hydrological model performance,* in particular for the low flow simulation. A more comprehensive, spatially-distributed calibration and rigorous evaluation of model performance for a proxy climate could further improve the credibility of hydrological impact simulations. A proxy of the future climate can be constructed by either considering historical periods which bear similarity to the projected future climate or other locations in the same geographical zone with a climate similar to the expected one. *A large uncertainty due to driving climate scenarios*, especially in some regions (African basins, Darling in Australia) hampered the identification of robust trends, and further efforts of climate modellers are needed to solve this problem.

Though the fractional uncertainty from the regional-scale HMs is the smallest in the overall results, it is still notable for some variables and basins. Therefore, finally, we advocate greater usages of multi-model ensembles of the calibrated regional hydrological models for impact assessment to provide more robust and credible results.

References

Andréasson J, Bergström S, Carlsson B, Graham L P and Lindsröm G 2004 Hydrological change—climate change impact simulations for Sweden *AMBIO* **33** 228–34

Arnold J G, Srinivasan R, Muttiah R S and Williams J R 1998 Large area hydrologic modeling and assessment part I: model development *J. Am. Water Resour. Assoc.* **34** 73–89

Bergström S and Forsman A 1973 Development of a conceptual deterministic rainfall-runoff model *Nord. Hydrol.* **4** 147–70

Bergström S, Carlsson B, Gardelin M, Lindström G, Pettersson A and Rummukainen M 2001 Climate change impacts on runoff in Sweden—assessments by global climate models, dynamical downscaling and hydrological modelling *Clim. Res.* **16** 101–12

Bosshard T, Carambia M, Goergen K, Kotlarski S, Krahe P, Zappa M and Schär C 2013 Quantifying uncertainty sources in an ensemble of hydrological climate impact projections *Water Resour. Res.* **49** 1523–36

Boyle D P 2001 Multicriteria calibration of hydrological models *PhD Dissertation* University of Arizona, Tucson

Coles S 2001 *An Introduction to Statistical Modeling of Extreme Values* (London: Springer)

Criss R E and Winston W E 2008 Do Nash values have value? Discussion and alternate proposals *Hydrol. Process* **22** 2723–5

Dankers R *et al* 2014 First look at changes in flood hazard in the inter-sectoral impact model intercomparison project ensemble *Proc. Natl Acad. Sci.* **111** 3257–61

Donnelly C, Andersson J C M and Arheimer B 2016 Using flow signatures and catchment similarities to evaluate a multi-basin model (E-HYPE) across Europe *Hydrol. Sci. J.* **61** 255–73

Eisner S *et al* 2017 An ensemble analysis of climate change impacts on streamflow seasonality across 11 large river basins *Clim. Change* **141** 401–17

Gelfan A *et al* 2017 Climate change impact on the water regime of two great Arctic rivers: modeling and uncertainty issues *Clim. Change* **141** 499–515

Giuntoli I *et al* 2015 Future hydrological extremes: the uncertainty from multiple global climate and global hydrological models *Earth Syst. Dyn.* **6** 267–85

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

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

Hawkins E and Sutton R 2009 The potential to narrow uncertainty in regional climate predictions *Bull. Am. Meteorol. Soc.* **90** 1095–1107

Hempel S, Frieler K, Warszawski L, Schewe J and Piontek F 2013 A trend-preserving bias correction—the ISIMIP approach *Earth Syst. Dyn. Discuss.* **4** 49–92

Huang S, Krysanova V and Hattermann F F 2013 Projection of low flow conditions in Germany under climate change by combining three RCMs and a regional hydrological model *Acta Geophys.* **61** 151–93

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

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

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. Model.* **106** 261–89

Krysanova V, Kundzewicz Z W and Piniewski M 2016 Assessment of climate change impacts on water resources *Handbook of Applied Hydrology* 2nd edn, V P Singh (ed) (New York: McGraw-Hill) ch 148

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 *et al* 2017 How the performance of hydrological models relates to credibility of projections under climate change *Hydrol. Sci. J.* submitted

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* 2017 Differences in flood hazard projections in Europe—their causes and consequences for decision making *Hydrol. Sci. J.* **62** 1–14

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

Lindström G, Pers C P, Rosberg R, Strömqvist J and Arheimer B 2010 Development and test of the HYPE (hydrological predictions for the environment) model—a water quality model for different spatial scales *Hydrol. Res.* **41** 295–319

Mishra V *et al* 2017 Multimodel assessment of sensitivity and uncertainty of evapotranspiration and a proxy for available water resources under climate change *Clim. Change* **141** 451–65

Motovilov Yu G, Gottschalk L, Engeland K and Rodhe A 1999 Validation of a distributed hydrological model against spatial observation *Agric. Forest Meteorol.* **98–99** 257–77

Motovilov Yu G 2013 ECOMAG: a distributed model of runoff formation and pollution transformation in river basins solution *Understanding Freshwater Quality Problems in a Changing World. Proc. of H04, IAHS-IAPSO-IASPEI Assembly* B Arheimer, A Collins, V Krysanova, E Lakshmanan, M Meybeck and M Stone (eds) vol **361** (Gothenburg: IAHS Publications) pp 227–34

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

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

Olsson J *et al* 2016 Hydrological climate change impact assessment at small and large scales: key messages from recent progress in Sweden *Climate* **4** 39

Pechlivanidis I G *et al* 2017 Analysis of hydrological extremes at different hydro-climatic regimes under present and future conditions *Clim. Change* **141** 467–81

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

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

Strauch M *et al* 2017 Adjustment of global precipitation data for enhanced hydrologic modeling of tropical Andean watersheds *Clim. Change* **141** 547–60

Su B *et al* 2017 Impacts of climate change on streamflow in the upper Yangtze river basin *Clim. Change* **141** 533–46

Teklesadik A D *et al* 2017 Inter-model comparison of hydrological impacts of climate change on the upper Blue Nile basin using ensemble of hydrological models and global climate models *Clim. Change* **141** 517–32

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

Verzano K 2009 Climate change impacts on flood related hydrological processes: further development and application of a global scale hydrological model *Reports on Earth System Science, 71-2009* (Hamburg: Max Planck Institute for Meteorology)

Vetter T, Huang S h, Aich V, Yang T, Wang X, Krysanova V and Hattermann F 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

Wang X *et al* 2017 Analysis of multi-dimensional hydrological alterations under climate change for four major river basins in different climate zones *Clim. Change* **141** 483–98

Warszawski L, Frieler K, Huber V, Piontek F, Serdeczny O and Schewe J 2014 The inter-sectoral impact model intercomparison project (ISI-MIP): project framework *Proc. Natl Acad. Sci.* **111** 3228–32

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** 823–48