Climatic Change
DOI 10.1007/s10584-016-1829-4

Cross-scale intercomparison of climate change impacts simulated by regional and global hydrological models in eleven large river basins

F. F. Hattermann[1] · **V. Krysanova**[1] · **S. N. Gosling**[2] · **R. Dankers**[3] · **P. Daggupati**[4] · **C. Donnelly**[5] · **M. Flörke**[6] · **S. Huang**[1] · **Y. Motovilov**[7] · **S. Buda**[8] · **T. Yang**[9,10] · **C. Müller**[1] · **G. Leng**[11] · **Q. Tang**[12] · **F. T. Portmann**[13] · **S. Hagemann**[14] · **D. Gerten**[1,15] · **Y. Wada**[16,17,18,21] · **Y. Masaki**[19] · **T. Alemayehu**[20] · **Y. Satoh**[21] · **L. Samaniego**[22]

Received: 8 January 2016 / Accepted: 6 October 2016

Abstract Ideally, the results from models operating at different scales should agree in trend direction and magnitude of impacts under climate change. However, this implies that the sensitivity to climate variability and climate change is comparable for impact models designed for either scale. In this study, we compare hydrological changes simulated by 9 global and 9

This article is part of a Special Issue on "Hydrological Model Intercomparison for Climate Impact Assessment" edited by Valentina Krysanova and Fred Hattermann.

Electronic supplementary material The online version of this article (doi:10.1007/s10584-016-1829-4) contains supplementary material, which is available to authorized users.

✉ F. F. Hattermann
hattermann@pik-potsdam.de

1 Potsdam Institute for Climate Impact Research, Potsdam, Germany

2 School of Geography, University of Nottingham, Nottingham, UK

3 Met Office, Exeter, UK

4 University of Guelph, Guelph, Canada

5 Swedish Meteorological and Hydrological Institute, Norrköping, Sweden

6 Center for Environmental Systems Research, University of Kassel, Kassel, Germany

7 Water Problems Institute of Russian Academy of Science, Moscow, Russia

8 National Climate Center, China Meteorological Administration, Beijing, China

9 State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing, China

Published online: 04 January 2017

regional hydrological models (HM) for 11 large river basins in all continents under reference and scenario conditions. The foci are on model validation runs, sensitivity of annual discharge to climate variability in the reference period, and sensitivity of the long-term average monthly seasonal dynamics to climate change. One major result is that the global models, mostly not calibrated against observations, often show a considerable bias in mean monthly discharge, whereas regional models show a better reproduction of reference conditions. However, the sensitivity of the two HM ensembles to climate variability is in general similar. The simulated climate change impacts in terms of long-term average monthly dynamics evaluated for HM ensemble medians and spreads show that the medians are to a certain extent comparable in some cases, but have distinct differences in other cases, and the spreads related to global models are mostly notably larger. Summarizing, this implies that global HMs are useful tools when looking at large-scale impacts of climate change and variability. Whenever impacts for a specific river basin or region are of interest, e.g. for complex water management applications, the regional-scale models calibrated and validated against observed discharge should be used.

1 Introduction

Climate change is a global phenomenon, but its impacts are manifested at the regional scale (IPCC 2013). A global view on climate change impacts is important to quantify the aggregated effects, and developments at the global scale can influence driving forces in the region under study. The regional scale, on the other hand, is where most adaptation measures are planned and implemented and where interaction with affected stakeholders is most intense (Krysanova et al. 2005; Hattermann et al. 2011). As a result, both global and regional studies provide valuable information for decision-making and scientific understanding. The cross-scale interaction makes it important to bridge the scales in impact assessment and to compare the sensitivity of impact

[10] Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Beijing, China

[11] Joint Global Change Research Institute, Pacific Northwest National Laboratory, Richland, WA, USA

[12] Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China

[13] Institute of Physical Geography, Johann Wolfgang Goethe-University Frankfurt am Main, Frankfurt, Germany

[14] Max Planck Institute for Meteorology, Hamburg, Germany

[15] Geography Department, Humboldt-Universität zu Berlin, Berlin, Germany

[16] NASA Goddard Institute for Space Studies, New York, NY, USA

[17] Center for Climate Systems Research, Columbia University, New York, NY, USA

[18] Faculty of Geosciences, Utrecht University, Utrecht, The Netherlands

[19] Center for Global Environmental Research, National Institute for Environmental Studies, Tsukuba, Japan

[20] Vrije Universiteit Brussel, Brussel, Belgium

[21] International Institute for Applied Systems Analysis, Laxenburg, Austria

[22] UFZ-Helmholtz Centre for Environmental Research, Leipzig, Germany

models of both scales to climate variability and change. Further, a comparison of regional and global hydrological models across a wide range of river basins provides a framework to test the consistency between the different scales of analysis and identifying a need for improvement.

Global hydrological models (Glob-HMs) are usually designed to supply consistent impact assessment for the continental and global scales. These models often compromise the model performance at the scale of individual catchments for the sake of overall model performance (Gosling and Arnell 2011; Müller-Schmied et al. 2014). Regional or catchment-scale hydrological models (Cat-HMs) are typically more streamlined to the specific characteristics of the catchment under investigation, e.g. through local input data that better describe local conditions, calibration to observations and implementation of regionally important hydrological features (Koch et al. 2013; Hattermann et al. 2006). There is no strict border between "purely" global and "purely" regional models. More and more hydrological features are implemented in global models, and model advancement and increase in computational power have led to the development that some global models are applied at the regional scale with higher resolution (e.g. WaterGAP3, Eisner (2016); DBH, Tang et al. 2007), while some regional models are applied at the continental scale (e.g. HYPE, Donnelly et al. 2015; Racovec et al. 2016a, b).

The way we distinguish the global and regional models in our study is that the former were applied for all continents with a spatial resolution of 0.5° without calibration (with the exception of WaterGAP2), while the regional models were applied for 11 large-scale river basins with a finer spatial resolution and were calibrated to observed discharge (see more details in Krysanova and Hattermann, this special issue SI). In this study, we make use of global and regional HM output data (Table A1 in the Annex) uploaded in the framework of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP, Schellnhuber et al. 2014; Warszawski et al. 2014). ISIMIP is a community-driven modelling effort bringing together impact modelers across sectors and scales to create consistent and comprehensive projections of impacts at different levels of global warming, based on the Representative Concentration Pathways (RCPs, van Vuuren et al. 2011) and Shared Socio-Economic Pathways (SSPs) scenarios (IPCC 2013).

In our study, we investigate the consistency of climate change impacts on the long-term average seasonal dynamics of discharge in 11 large-scale river basins (see Table A2 in Annex), covering the main climatic zones and hydrological regimes on all continents. It was not possible to apply all regional models to all basins, because implementation of new model set-ups is work intensive and exceeded the capacity of the regional team.

To our knowledge, this is one of the first comprehensive cross-scale inter-comparisons of multiple hydrological models considering river basins on *all* continents, although there have been cross-scale model intercomparisons involving fewer models and basins (Gosling et al. 2011; Piniewski et al. 2014). Responses to climate change in hydrological extremes and mean annual runoff of the same HMs are reported in another cross-scale paper by Gosling et al. (2016). The most recent comparison of Glob-HMs was conducted within the framework of ISIMIP and described by Schewe et al. 2014; Dankers et al. 2014; Prudhomme et al. 2014; Haddeland et al. 2014; Davie et al. 2013; Wada et al. 2013 and Portmann et al. 2013. Model intercomparisons for the regional scale are described in Breuer et al. 2009; Bosshard et al. 2013; Chen et al. 2013 and Vetter et al. 2014.

2 Methods, models, river basins and climate data

2.1 Models

In total, outputs from 9 Glob-HMs (CLM, DBH, H08, LPJmL, Mac-PDM.09, MATSIRO, MPI-HM, PCR-GLOBWB, WaterGAP2) and 9 Cat-HMs (ECOMAG, HBV, HYMOD, HYPE, mHM, SWAT, SWIM, VIC, WaterGAP3) are considered in this study. Annex Table A1 lists the models and references where more information on them can be found. While the global models consistently simulate hydrological processes and river routing with a spatial resolution of 0.5°, different approaches are used by the regional models: regular grids (e.g. VIC and WaterGAP3) and disaggregation schemes with subbasins and hydrological response units (SWIM, HYPE and SWAT). More information on basic processes represented in the models is given in Annex Table A3. All models simulate the full water cycle, with daily precipitation and temperature as main inputs, calculation of evapotranspiration, infiltration, generation of runoff, and application of a routing scheme to transfer the locally generated runoff along the river network to the outlet. Some of the models include more processes such as lake dampening of flow, regulation of flow, wetlands and more.

Table A2 illustrates which hydrological models were applied in which of the eleven river basins. While the Glob-HMs provided outputs for each river basin, only a subset of Cat-HMs was applied in most cases, due to the workload associated with model set-ups and calibration in catchments. More information about the regional models, the calibration process and the validation results can be found in Krysanova and Hattermann and in Huang et al. 2016.

The Glob-HMs are operated at the same spatial resolution as the provided climate data (0.5°), whereas further model-specific interpolation of climate data to the subbasin scale was necessary to run the regional models. In addition, some of the regional models corrected precipitation and temperature during interpolation taking into account elevation. Following the approach accepted in ISIMIP, all models were applied and their results analysed in this study disregarding their specific model set-up and performance in the reference period.

2.2 River basins

Eleven river basins were selected for this cross-scale comparison to cover the most important climate zones and hydrological regimes worldwide. The maps in Figure A1 show their location, and Table A2 summarizes some of their characteristics (Annex). More information about these river basins is given in Krysanova and Hattermann (this SI). The upper parts of several basins (Mississippi, Amazon, Yangtze, Yellow, Niger and Blue Nile) were chosen because they have no or minor influence of human management, thus making it possible to compare close-to-natural discharge.

2.3 Data

To obtain a coherent impact model intercomparison, the models are driven by climate forcing data from the same source and for the same periods. For the analysis of model performance under current conditions, all models were forced by global WATCH Forcing Data (WFD, Weedon et al. 2011), daily 0.5 by 0.5° gridded meteorological data covering the period 1971–2000. The CMIP5 climate scenario data (Taylor et al. 2012) used in this study were provided by ISIMIP. Five Earth System Models (HadGEM2-ES, IPSL-CM5A-LR, MIROC-ESM-

CHEM, GFDL-ESM2M, NorESM1-M) which have been bias-corrected using a trend-preserving method (Hempel et al. 2013), were applied. In this study, only the high-end scenario RCP8.5 was used. For more information about the climate scenario simulations for the individual river basins see Krysanova and Hattermann (this SI).

Other important input data for hydrological models are soil, land cover, elevation and hydrological information such as the river network. In most cases, they were taken from globally available data sources (see Table 2 in Krysanova and Hattermann, this SI), and in some cases regional-scale models used different spatial data they were originally implemented with. Observed discharge time series were provided by the Global Runoff Data Centre (GRDC 2013) or country specific agencies.

3 Results

3.1 Model performance during the reference period

3.1.1 Comparison of simulated and observed discharges

The so-called "validation" runs for comparison of the hydrological models were done for the gauging stations listed in Table A2 for the whole reference period 1971–2000. In advance, all Cat-HMs were calibrated against observed discharge and afterwards validated in a split sample mode, i.e. validating the model using discharge data of a time period different from calibration, normally with an 8–10 year period for calibration, depending on data availability. The Glob-HMs were not calibrated, except WaterGAP2 (which was calibrated against long-term average monthly discharge for a number of gauges worldwide).

Figure 1 visualizes the long-term average monthly dynamics of discharge for 1971–2001 simulated by Cat-HMs and Glob-HMs at the downstream gauges of the eleven basins, and Table 1 provides quantitative assessment. In general, Cat-HMs reproduce the observed long-term average seasonal dynamics of discharge well, with narrow ranges of uncertainty. This is partly so because the minimizing volume error is generally a calibration target of regional models. Results of the Glob-HMs in most cases show much higher uncertainty ranges in terms of deviation from the mean, and often a considerable bias towards observed data, mostly too high discharge, e.g. for the rivers Rhine, Tagus, Upper Mississippi, Upper Niger, Blue Nile, Ganges and Darling. The best performance of the mean of the nine Glob-HM results is for the Upper Yellow, followed by the Upper Yangtze (Fig. 1).

The Darling is an extreme case, with a strong overestimation of the long-term average seasonal dynamics by Glob-HMs, while the Cat-HMs perform better but not as well as for the other basins (see also Table 1). A possible reason for the poor results especially in the Darling and in other arid and semi-arid climates may be the low runoff coefficient (i.e. the fraction of precipitation that reaches the basin outlet) because even a small underestimation of evapotranspiration (or overestimation of precipitation in the forcing) may lead to large overestimation of river discharge. Also, lots of unregulated and regulated water abstractions are reported for the Darling, including water harvesting (Kingsford 2000, Thoms and Sheldon 2000), which were not considered in the modelling for this study.

In the Upper Amazon, all models underestimate discharge in May and June, due to underestimation of precipitation in the rainy season in the driving WATCH ERA-40 data (Strauch et al. 2016). In the Lena, the inclusion of soil freezing and thawing is very likely to

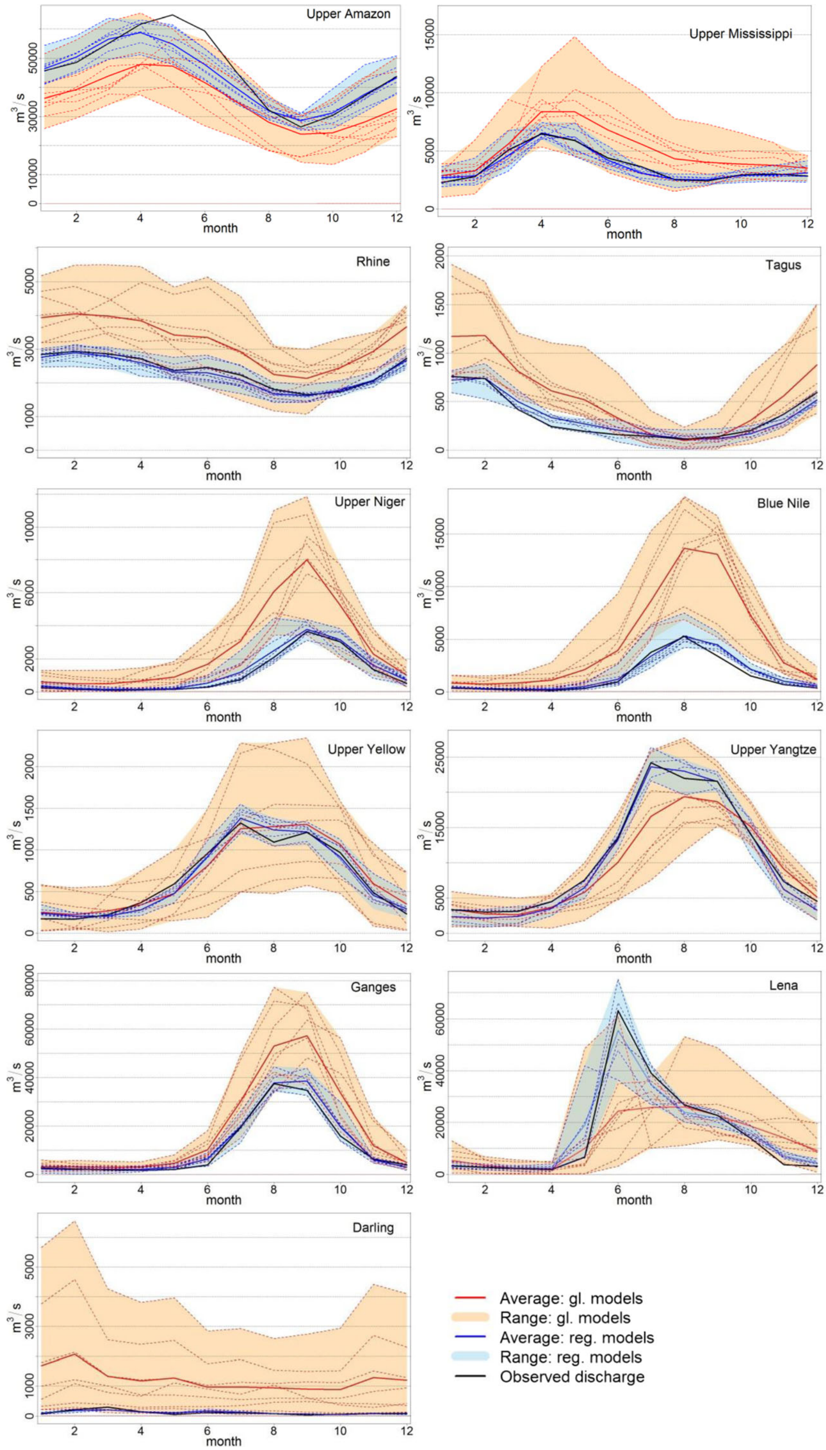

Fig. 1 Comparison of the observed and simulated long-term average monthly seasonal dynamics of river discharge for 1971–2000 as modelled by the Cat-HMs and Glob-HMs in the selected 11 river basins

influence river discharge as it results in river runoff generation during spring-summer period and a high runoff peak, a process considered only in the ECOMAG model and, by a static permafrost mask, in MPI-HM.

Table 1 provides quantitative results of the model comparison shown in Fig. 1: the correlation coefficient (r) between the simulated and observed mean annual cycles of the years 1971–2000, the bias in standard deviation ($\Delta\sigma$) in percent and the d-factor as a measure of uncertainty.

According to the thresholds, high correlation (≥0.9) was found for 10 basins (all except the Darling) for means of Cat-HMs, but for only 4 out of 11 basins for means of Glob-HMs; and low bias in standard deviation (< ±15 %) was found in 9 cases for means of Cat-HMs, but only in one case for means of Glob-HMs. The values of d-factor below 1 denoting a low uncertainty related to observations (see Abbaspour et al. 2007) were found for 9 basins with Cat-HMs, but only for one case with Glob-HMs.

There are several reasons which could be responsible for large biases in HMs: a) model structure, b) its parametrization, c) spatial resolution, d) biased input data, and e) calibration. In our case, biased input data (precipitation) caused underestimation of discharge in the Upper Amazon in May and June (see Strauch et al. 2016), and missing calibration of the eight Glob-HMs used in Table 1 is possibly the main reason of the often high bias of their discharge results.

3.1.2 Sensitivity of modelled river discharge to climate variability

We investigated the sensitivity of discharge simulated by the Glob-HMs and Cat-HMs to climate variability by calculating the anomalies of annual precipitation and annual discharge for the reference period 1971–2000 and fitting outputs from the two model sets to a nonlinear

Table 1 Performance of the global and regional model sets considering reproduction of the long-term average seasonal dynamics (monthly values) in the period 1971–2000 using WATCH data as climate input. Indicators are the correlation coefficient r between simulated and observed monthly values (r), percent bias in standard deviation ($\Delta\sigma$) (see Equation 1 in Annex) in columns 2, 3, 7 and 8, and d-factor as a measure of uncertainty (Abbaspour et al. 2007) in columns 4 and 9. Usually the thresholds $r \geq 0.9$ and $\Delta\sigma < \pm 15$ % denote a good performance (Huang et al., this SI). High average fit ($r \geq 0.9$, $\Delta\sigma < \pm 15$ %, d-factor < 1) is indicated by shading. The percentage share of models with a moderate fit of $r > 0.8$ and $\Delta\sigma < \pm 30$ % is shown in columns 5, 6, 10 and 11 (single model evaluation)

	Cat-HMs					*Glob-HMs*				
Basin	*Average dynamics: corr. coef. r*	*Average dynamics: bias in STD Δσ*	*d-factor*	*Share of models with r>0.8, in %*	*Share of models with Δσ:< 30, in %*	*Average dynamics: corr. coef. r*	*Average dynamics: bias in STD Δσ*	*d-factor*	*Share of models with r>0.8, in %*	*Share of models with Δσ:< 30, in %*
Column #	2	3	4	5	6	7	8	9	10	11
Rhine	0.95	1.9	1.08	100	78	0.87	68	4.60	88	50
Tagus	0.96	-5.4	0.75	100	60	0.91	67	2.86	100	50
U. Niger	0.96	7.3	0.72	100	100	0.89	116	2.59	75	13
Blue Nile	0.97	4.5	0.65	100	83	0.93	187	3.23	100	13
Lena	0.92	-10.6	0.51	80	100	0.61	66	1.23	38	0
U. Yellow	0.97	4.5	0.55	100	100	0.89	7.2	2.34	75	25
U. Yangtze	0.99	7.2	0.37	100	100	0.90	-16	0.99	88	50
Ganges	0.98	7.4	0.45	100	100	0.95	60	1.32	100	38
Darling	0.83	-29.5	0.68	50	50	0.34	431	47.2	0	38
U. Mississippi	0.92	2.0	1.13	88	88	0.80	59	3.61	50	25
U. Amazon	0.90	-16.5	0.89	83	100	0.87	-25	1.99	100	50

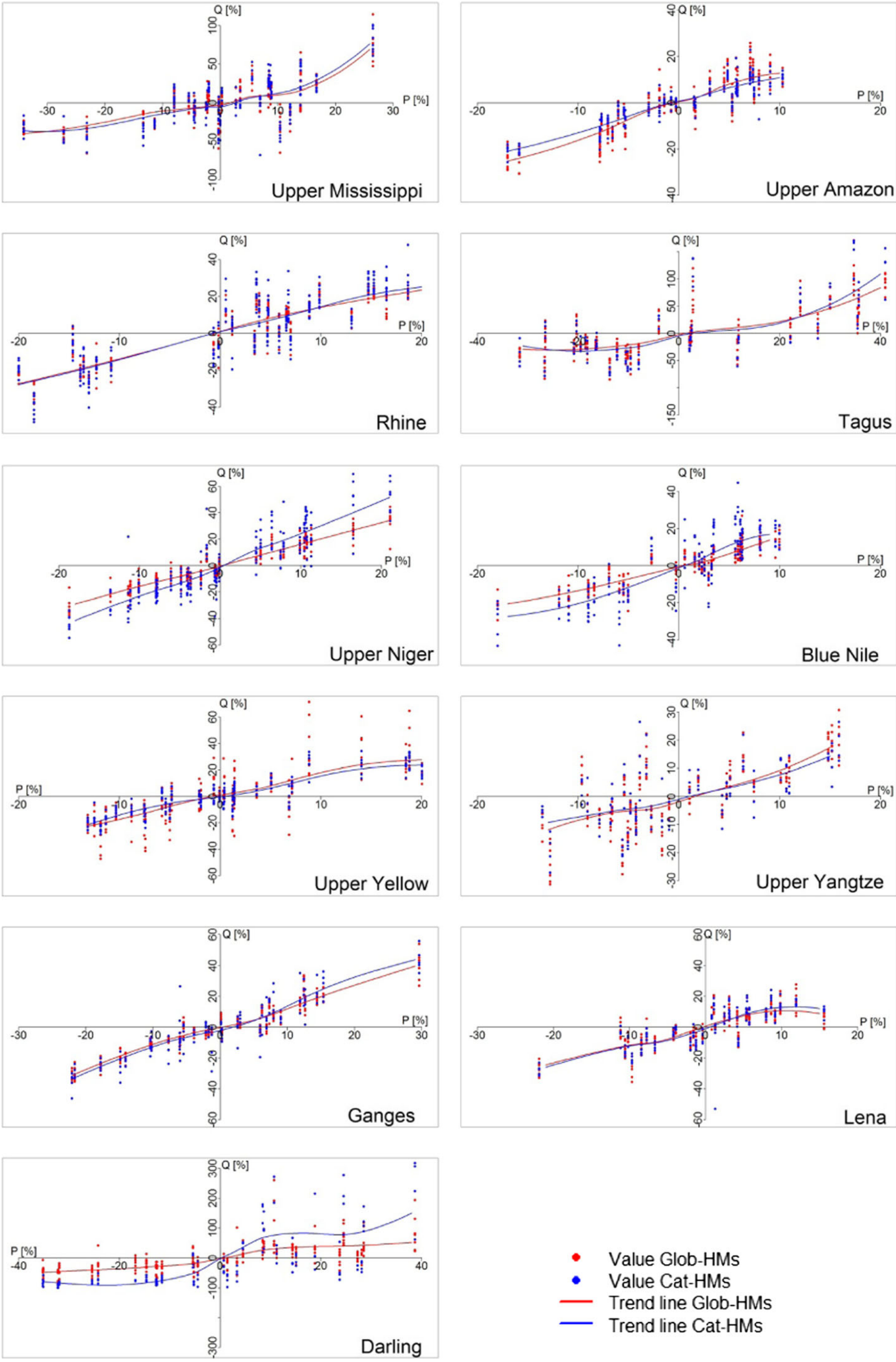

Fig. 2 Sensitivity of annual discharge simulated by Glob-HMs and Cat-HMs to annual variability in precipitation for the 11 basins: anomalies in discharge (y-axis) versus anomalies in precipitation (x-axis) in the period 1971–2000 in percent. The lines were calculated using the LOESS technique

regression (Fig. 2). The anomalies are defined as the differences between the annual values for each year and the long-term average annual values over the period 1971–2000.

The lowest variability in precipitation was found for the Upper Amazon and Blue Nile basins with anomalies ranging from −20 to +10 %. The largest variability in precipitation was found for the Darling and Tagus basins with annual precipitation anomalies ranging from −40 to +40 %. The latter two are the driest regions considered, and they are consequently also the basins that show the highest variability in discharge (from less than −80 % to more than 150 % in the case of Tagus and from −100 % to more than 300 % in the case of Darling). The lowest variability in discharge was found for the Upper Amazon and the Upper Yangtze (between −30 and +30 %).

The correlation of changes in precipitation to changes in discharge has mostly close-to-linear character, only the Lena, Upper Mississippi, Tagus and Darling show more nonlinear responses (Fig. 2). A positive anomaly in precipitation greater than 10 % usually produces a positive anomaly in discharge, but a smaller increase in rainfall may be associated with a decrease in discharge in single model runs (e.g. when in the specific application evapotranspiration increases more than precipitation).

The coefficient of determination R^2 of the fitted curves (see Table A4) is high for the Ganges, Upper Niger, Upper Amazon, Rhine and Blue Nile (both model types) in connection with their mostly high precipitation and runoff coefficients, and much lower for the Tagus, Upper Mississippi, Upper Yangtze and Darling. In general, there is no clear and distinct relation to the runoff coefficient, but interesting is that the single R^2 values of the two model sets are comparable.

A robust conclusion that can be drawn from Fig. 2 is that no systematic differences in Glob-HM and Cat-HM sensitivities to climate variability can be observed, only the Darling River (where also the bias in discharge is highest for both model ensembles), as well as the Upper Niger and Blue Nile rivers show larger deviations.

3.2 Climate change impacts on seasonal flows

Comparison of the climate change impacts simulated by Glob-HMs and Cat-HMs was done for the high-end scenario RCP8.5 by comparing the *differences in long-term average monthly discharges* between the periods 2071–2099 and 1971–2000 in *terms of medians and uncertainty ranges from the two HM sets* (Figs. 3, A3 and Table 2).

While temperature increases in all basins under scenario conditions, trends in precipitation are diverse (Krysanova and Hattermann, this SI). In general, the rivers showing the strongest overall decrease in mean seasonal discharge are the Tagus, Rhine and Darling, whereas increases are most pronounced for the Ganges, Lena and Upper Amazon. The changes in medians without uncertainty ranges are shown additionally for 6 basins in Figure A3 in the Annex.

As one can see in Fig. 3, similar to the model validation against observed discharge (Fig. 1), the Glob-HMs mostly span much wider ranges, especially for the Tagus, Upper Niger and Darling. The medians of the simulated changes of the two model ensembles are visually comparable in Fig. 3 for some basins (e.g. for the Rhine and Ganges), but differ in other cases (see Figure A3). In some cases, for example the Upper Mississippi, Upper Niger and Upper Yangtze Rivers, the uncertainty in changes from the Glob-HMs is very large compared to the average changes. Therefore, a more formal analysis of similarity of the long-term average discharges from the two ensemble results (Fig. 3) was done using the non-parametric Wilcoxon signed-rank test, with a confidence level of 95 % and in the two-sided mode (see Annex). The hypothesis of similarity of the population mean ranks (i.e. of the signals of change) was confirmed in five cases of eleven (Rhine, Upper Niger, Ganges, Upper Mississippi and Lena) by this test.

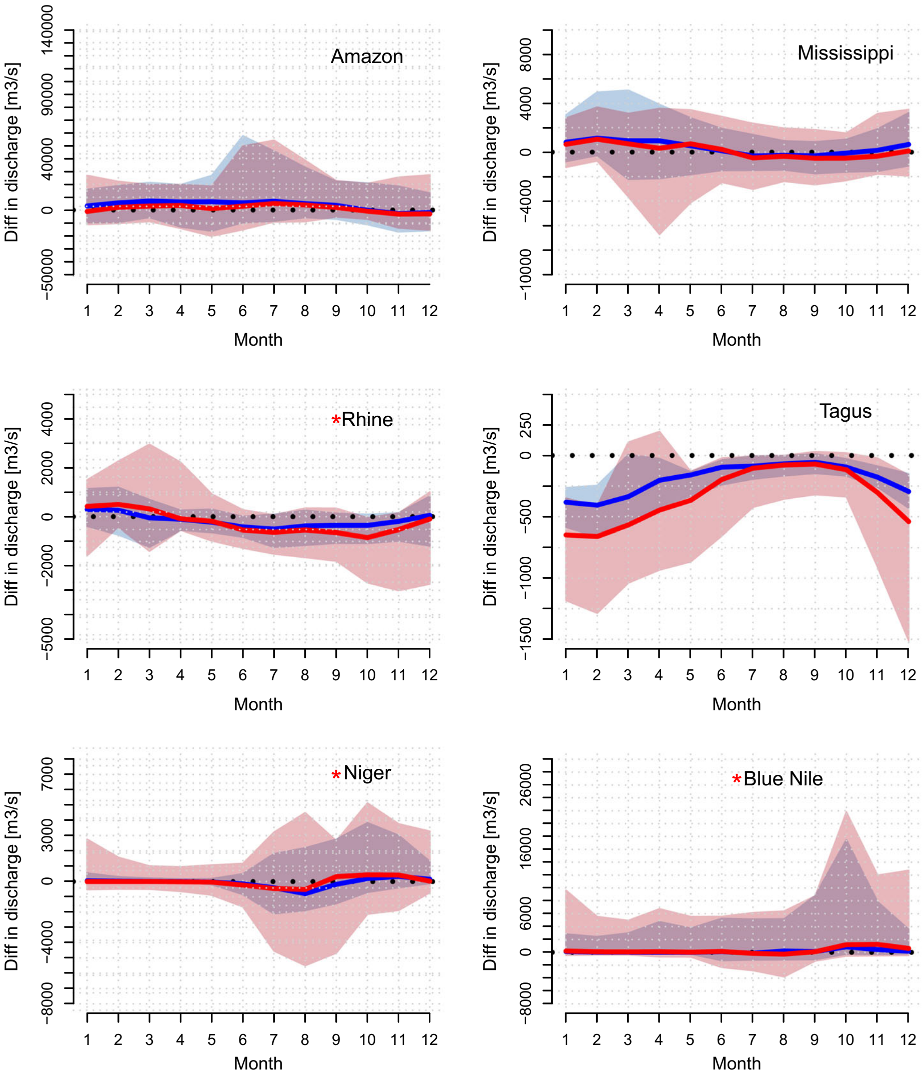

Fig. 3 Comparison of climate change impacts on the long-term average monthly discharge modelled by the Glob-HMs and by Cat-HMs driven by 5 GCMs (scenario RCP8.5) for the period 2071–2099 compared to the reference period 1971–2000. *Red stars* indicate that the medians are not distinguishable with the confidence level of 95 % following the two-sided Wilcoxon signed-rank test

In addition, the change signals in terms of means and medians (presented in Fig. 3) as well as spreads and the ratio of spreads to the mean were estimated (Table 2, columns 2–9) and analyzed. The last two columns provide a qualitative estimation of similarity. As we see from this table, the means and medians are well comparable for the Ganges and Lena (though the shapes of seasonal dynamics for the Lena are different, see Fig. 3), and the differences are not large for the Rhine and Blue Nile. For the remaining seven basins differences are higher than 70 %, and in three cases they are very high (Upper Niger, Upper Yangtze and Darling). The spreads from Glob-HM simulations are higher than those from Cat-HMs in 10 cases out of 11. The spreads are well comparable in four cases: for the Lena, Upper Amazon and the two Chinese basins.

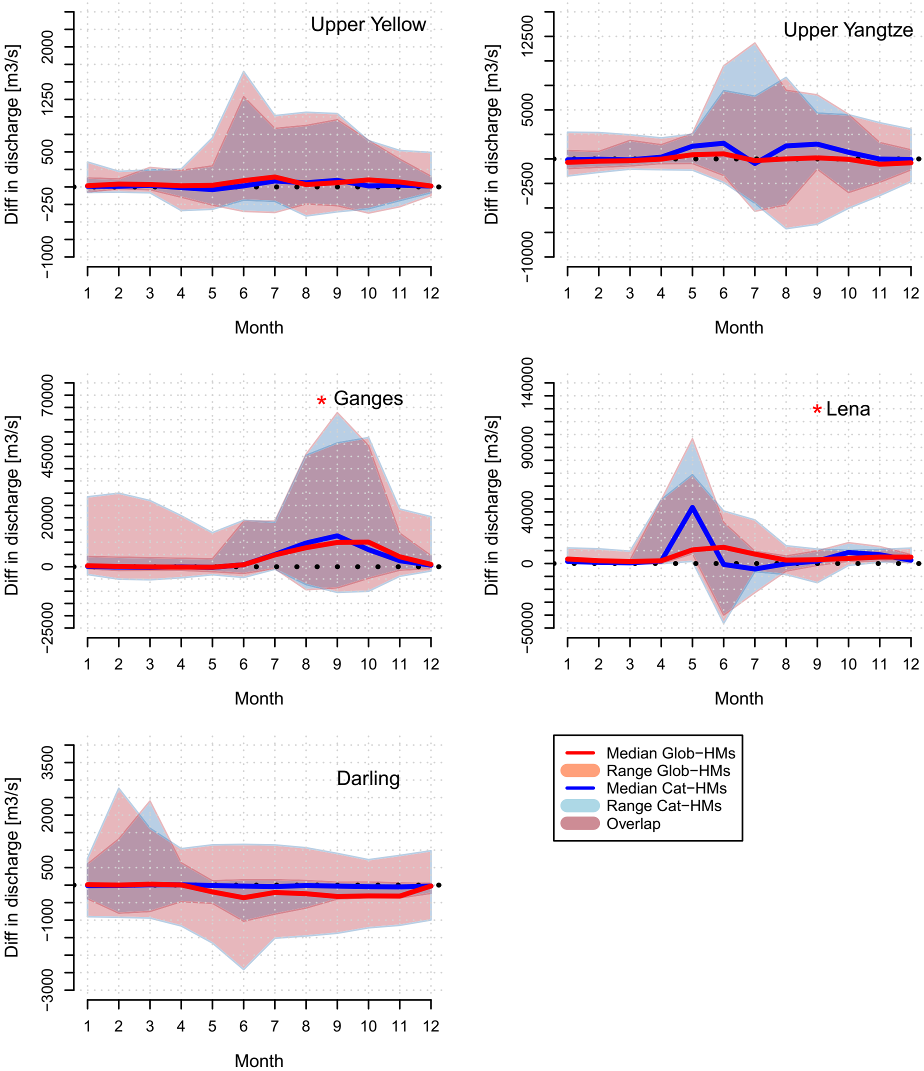

Fig. 3 (continued)

It is important to mention that the large uncertainty ranges in Fig. 3 are the combined effects of global climate model and hydrological model uncertainty. The uncertainty related only to HMs can be seen in Figure A4 (evaluation in Table A5), where results driven only by one climate model, GFDL-ESM2M, are presented as partial results from Fig. 3, confirming higher uncertainty related to Glob-HMs compared to Cat-HMs.

4 Discussion

The results suggest that the model sensitivity of Cat-HMs and Glob-HMs to climate variability is comparable in most cases (Fig. 2). The spreads of the annual discharges for each model set

Table 2 Comparison of differences in seasonal dynamics of discharge (in m^3/s) between end of the century and reference period (RCP 8.5) simulated by Glob-HMs and Cat-HMs (as in Fig. 3) in terms of annual means, medians and spreads

	Global models				Regional models				Comparison of means, medians and spreads				
	Change signal (seas. mean)	Change signal (seas. median)	av. spread	av. spread/ mean	Change signal (seas. mean)	Change signal (seas. median)	av. spread	av. spread/ mean	Difference in means (abs. values, without sign)	Difference in medians (abs. values, without sign)	Spread (Glob HM)/Spread (Cat-HM)	Similarity of means & medians	Similarity of spreads
Data unit	m^3/s	m^3/s	m^3/s	–	m^3/s	m^3/s	m^3/s	–	–	–	%	–	–
Rhine	−278	−237	2067	7.4	−169	−164	1154	6.8	Glob-HM: 64 % higher (neg)	Glob-HM: 45 % higher (neg)	179	+/−	---
Tagus	−366	−341	652	1.8	−205	−196	244	1.2	Glob-HM: 79 % higher (neg)	Glob-HM: 74 % lower (neg)	267	---	---
Upper Niger	25	−5	3841	153	137	−66	1746	12.7	Cat-HM: 5.5 times larger (pos)	Cat-HM: 13 times larger (neg)	220	---	---
Blue Nile	1131	274	7763	6.9	838	159	5413	6.4	Glob-HM: 35 % higher (pos)	Glob-HM: 72 % higher (pos)	143	+/−	+/−
Ganges	5101	3206	21789	4.3	4161	3132	16373	3.9	Glob-HM: 23 % higher (pos)	Glob-HM: 2 % higher (pos)	133	++	+
Lena	5923	4934	24365	4.1	6211	5239	21492	3.4	Cat-HM: 5 % higher (pos)	Cat-HM: 6 % higher (pos)	113	++	++
Upper Yangtze	9	−73	5861	651	813	487	5253	6.4	Cat-HM: 116 times higher	Cat-HM and Glob-HM: diff. signs	112	---	++
Upper Yellow	88	53	648	7.4	61	19	681	11.1	Glob-HM: 44 % higher (pos)	Glob-HM: 2.8 times larger (pos)	95	---	++
Darling	−196	−161	2084	10.7	−45	−21	854	18.9	Glob-HM: 4.4 times larger (neg)	Glob-HM: 7.7 times larger (neg)	244	---	---
Upper Mississippi	25	126	4859	194	405	353	3324	8.2	Cat-HM: 16 times larger (pos)	Cat-HM: 180 % higher (pos)	146	---	+/−
Upper Amazon	2928	1311	33640	11.5	5271	3849	32051	6.1	Cat-HM: 80 % higher (pos)	Cat-HM: 2.9 times larger	105	---	++

The change signals (columns 2, 3, 6, 7) are calculated by averaging 12 values of the long-term mean or median values. The signs in the last two columns: ++ similar (differ < 25 %); + quite similar (differ < 40 %); +/− moderately different (difference 40–70 %); — strongly different (difference > 70 %)

in Fig. 2 have similar ranges, because models from both scales use the same climate input and comparable equations to calculate potential evapotranspiration (see Table A3), also indicating that the model sensitivity to climate variability is comparable and not altered by calibration of the Cat-HMs.

The comparable model sensitivity is in contrast to the fact that the long-term average seasonal dynamics of discharge simulated by the Glob-HMs often show large biases compared to the observed values when driven with WATCH data for the reference period 1971–2000 (Fig. 1, Table 1), whereas Cat-HMs show a better reproduction of the observed seasonal discharge.

When looking at climate change impacts, the highly aggregated outputs such as the long-term monthly averages of the two model sets show visually comparable shapes for many of the 11 catchments (Fig. 3) with large uncertainty bounds stemming from GCMs and HMs. However, a more formal statistical analysis of similarity of the mean ranks (i.e. signals of change in discharge) from the two ensembles confirms the hypothesis of similarity by the Wilcoxon test only in five cases of eleven (Fig. 3). Also the analysis of differences in means, medians and spreads (Table 2) reveals many differences between the two HM ensembles and only in two cases (Ganges and Lena) results agree in all three criteria.

While the focus here was on absolute changes in discharge, for many applications it might be sufficient to evaluate relative changes only (Schewe et al. 2014), or in the case of floods or droughts to use extreme value statistics (Feyen et al. 2012; Hattermann et al. 2014; Gosling et al. 2016). Figure A2 (Annex) shows the relative changes in discharge under climate change for three basins where the absolute results of the two model sets showed stronger differences, the Mississippi, Yangtze and Darling. Especially for the Darling, the similarity of results from the two model ensembles increases.

The results presented here generally support those ones from an earlier multi-scale hydrological model intercomparison (Gosling et al. 2011), which showed that Glob-HMs can be useful tools for understanding catchment-scale hydrological responses to climate change, if *ensemble mean impacts* on average annual flows, sign of change, or the seasonal cycle are of interest. However, the fact that ensemble medians of both model sets are comparable in some of the basins while single models (especially the global ones) often generate high uncertainty ranges proves that there is a real benefit in using a multi-model ensemble, as also reported in previous studies (e.g. Hagemann et al. 2013). In addition, this allows the model-related uncertainty to be quantified.

Most practitioners would certainly prefer a lower uncertainty in scenario results, while it might be of interest in some cases to screen a larger range of uncertainty, for example when planning sensitive infrastructure in riverine areas. Generally, models overestimating runoff by far during the reference period might also do so under climate change conditions.

In most cases, when simulated water components are used in subsequent management applications, accuracy of the data is important, for example in the case of water availability per capita, hydropower production, flood protection and crop production. In these cases, outputs of uncalibrated models or, more generally, models with poor performance should be used with care. For water resources applications, changes in many components of the water cycle also within the catchment may be equally important, and in this case a multi-site and multi-criteria validation is necessary (Hattermann et al. 2005). However, in some cases, the good model performance we observed for the Cat-HMs could be a sign of over-calibration, e.g. where hydrological processes are influenced by management which was not included.

Calibration of hydrological models is complex and the stability of calibrated parameters over time (and into the future) may be questionable (Merz et al. 2011). However, the fact that a model can adequately respond to the climatic variability within the historical period lends more trust to the projections using this specific model.

In general, the multi-model approach in climate impact assessment using several GCMs and HMs allows a) to find more robust results confirmed by most of the models, b) to identify uncertainty sources for every basin/region, c) to find out in which regions improvement of which tools is necessary, and d) to compare approaches disregarding and considering model performance.

5 Conclusions

Our study is, to our knowledge, one of the first comprehensive cross-scale hydrological model intercomparisons, based on application of 9 global and 9 regional hydrological models in 11 large scale river basins. Some of the results were as to be expected: Glob-HMs, mostly uncalibrated, often show a large bias in the long-term average seasonal discharge when results are compared against observations, although they do in many cases reproduce the intra-annual variability well. More surprising is the fact that the sensitivity of models of both scales to climate variability (evaluated for model ensembles) is quite similar in most basins. The simulated impacts in terms of seasonal dynamics show that medians are to a certain extent comparable in some basins but show differences in others, and spreads related to global models are mostly larger. The hypothesis of similarity of signals of change from the two ensembles is confirmed statistically only in five cases of eleven .

This study was limited to analysis of river discharge at the outlet of large scale river basins, an indicator for changes in the water balance of large regions. In follow-up investigations, more attention should be given to improving performance of global models, spatially-distributed calibration of models, analysis of other components of the water cycle, and also to other sources of uncertainty in scenario analysis, such as the emission scenarios and the driving global climate models.

Acknowledgments This work has been conducted under the framework of ISIMIP, funded by the German Federal Ministry of Education and Research (BMBF) with project funding reference number 01LS1201A. Responsibility for the content of this publication lies with the authors. We acknowledge the World Climate Research Programme's Working Group on Coupled Modeling, which is responsible for CMIP, and we thank the respective climate modelling groups for producing and making available their model output. We also acknowledge the support of the Global Runoff Data Centre and of the Environment Research and Technology Development Fund (S-10) of the Ministry of the Environment, Japan. This work was supported by Japan Society for the Promotion of Science KAKENHI (16H06291), and the Program for Risk Information on Climate Change supported by the Ministry of Education, Culture,Sports, Science, and Technology-Japan (MEXT).

References

Abbaspour KC, Yang J, Maximov I, Siber R, Bogner K, Mieleitner J, Zobrist J, Srinivasan R (2007) Modelling of hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. J Hydrol 333:413–430

Bosshard T, Carambia M, Goergen K, Kotlarski S, Krahe P, Zappa M, Schaer C (2013) Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections. Water Resour Res 49:1523–1536

Breuer L, Huisman JA, Willems P, Bormann H, Bronstert A, Croke BFW, Frede HG, Graeff T, Hubrechts L, Jakeman AJ, Kite G, Lanini J, Leavesley G, Lettenmaier DP, Lindstroem G, Seibert J, Sivapalan M, Viney NR (2009) Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM). I: Model intercomparison with current land use. Adv Water Resour 32:129–146

Chen X, Yang T, Wang X, Xu C-Y, Yu Z (2013) Uncertainty intercomparison of different hydrological models in simulating extreme flows. Water Resour Manag 27(5):1393–1409

Dankers R, Arneli NW, Clark DB, Falloon PD, Fekete BM, Gosling SN, Heinke J, Kim H, Masaki Y, Satoh Y, Stacke T, Wada Y, Wisser D (2014) First look at changes in flood hazard in the inter-sectoral impact model intercomparison project ensemble. PNAS 111(9):3257–3261

Davie JCS et al (2013) Comparing projections of future changes in runoff from hydrological and biome models in ISI-MIP. Earth Syst Dynam. doi:10.5194/esdd-4-279-2013

Donnelly C, Andersson J, Arheimer B (2015) Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe. Hydrol Sci J. doi:10.1080/02626667.2015.1027710

Eisner S (2016) Comprehensive evaluation of the WaterGAP3 model across climatic, physiographic, and anthropogenic gradients. PhD thesis. University of Kassel, Kassel, Germany. Available at http://nbn-resolving.de/urn:nbn:de:hebis:34-2016031450014

Feyen L, Dankers R, Bódis K, Salamon P, Barredo JI (2012) Fluvial flood risk in Europe in present and future climates. Clim Chang 112(1):47–62. doi:10.1007/s10584-011-0339-7

Gosling SN, Arnell NW (2011) Simulating current global river runoff with a global hydrological model: model revisions, validation, and sensitivity analysis. Hydrol Process 25(7):1099–1085. doi:10.1002/hyp.7727

Gosling SN, Taylor RG, Arnell NW, Todd MC (2011) A comparative analysis of projected impacts of climate change on river runoff from global and catchment-scale hydrological models. Hydrol Earth Syst Sci 15:279–294. doi:10.5194/hess-15-279-2011

Gosling SN, Zaherpour J, Mount NJ, Hattermann F, Dankers R, Arheimer B, Breuer L, Ding Haddeland, I, Kumar R, Kundu D, Liu J, van Griensven A, Veldkamp T, Vetter T, Wang X, Zhang X (2016) A comparison of changes in river runoff from multiple global and catchment-scale hydrological models under global warming scenarios of 1°C, 2°C and 3°C. Clim Chang. doi:10.1007/s10584-016-1773-3

GRDC, BfG The GRDC - Global Runoff Database. Available at: http://www.bafg.de/nn_266934/GRDC/EN/01__GRDC/03__Database/database__node.html?__nnn=true. Accessed 7 Feb 2013

Haddeland I, Heinke J, Biemans H, Eisner S, Flörke M, Hanasaki N, Konzmann M, Ludwig F, Masaki Y, Schewe J, Stacke T, Tessler ZD, Wada Y, Wisser D (2014) Global water resources affected by human interventions and climate change. PNAS 111(9):3251–3256. doi:10.1073/pnas.1222475110

Hagemann S, Chen C, Clark DB, Folwell S, Gosling SN, Haddeland I, Hanasaki N, Heinke J, Ludwig F, Voss F, Wiltshire AJ (2013) Climate change impact on available water resources obtained using multiple global climate and hydrology models. Earth Syst Dynam 4:129–144. doi:10.5194/esd-4-129-2013

Hattermann FF, Wattenbach M, Krysanova V, Wechsung F (2005) Runoff simulations on the macroscale with the ecohydrological model SWIM in the Elbe catchment—validation and uncertainty analysis. Hydrol Process 19(3):693–714

Hattermann FF, Krysanova V, Habeck A, Bronstert A (2006) Integrating wetlands and riparian zones in river basin modelling. Ecol Model 199:379–392

Hattermann FF, Weiland M, Huang S, Krysanova V, Kundzewicz ZW (2011) Model-supported impact assessment for the water sector in central Germany under climate change—a case study. Water Resour Manag 25(13):3113–3134

Hattermann FF, Huang S, Burghoff O, Willems W, Österle H, Büchner M, Kundzewicz ZW (2014) Modelling flood damages under climate change conditions—a case study for Germany. Nat Hazards Earth Syst Sci 14(12):3151–3168, http://www.nat-hazards-earth-syst-sci.net/14/3151/2014/nhess-14-3151-2014.html

Hempel S, Frieler K, Warszawski L, Schewe J, Piontek F (2013) A trend-preserving bias correction—the ISI-MIP approach. Earth Syst Dyn Discuss 4:49–92. doi:10.5194/esdd-4-49

Huang S, et al (2016) Evaluation of an ensemble of regional hydrological models with respect to simulation of seasonal dynamics, floods and droughts in 12 large river basins. Clim Chang. doi:10.1007/s10584-016-1841-8

IPCC (2013) Climate change 2013: the physical science basis—summary for policymakers. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. – Intergovernmental Panel on Climate Change. IPCC Secretariat

Kingsford RT (2000) Ecological impacts of dams, water diversions and river management on floodplain wetlands in Australia. Austral Ecology 25:109–127. doi:10.1046/j.1442-9993.2000.01036.x

Koch H, Liersch S, Hattermann FF (2013) Integrating water resources management in eco-hydrological modelling. Water Sci Technol 67(7):1525–1533

Krysanova V, Hattermann FF, Wechsung F (2005) Development of the ecohydrological model SWIM for regional impact studies and vulnerability assessment. Hydrol Process 19:763–783

Merz R, Parajka J, Blöschl G (2011) Time stability of catchment model parameters – implications for climate impact analyses. Water Resour Res 47(2). doi:10.1029/2010WR009505

Müller-Schmied H, Eisner S, Franz D, Wattenbach M, Portmann FT, Flörke M, Döll P (2014) Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. Hydrol Earth Syst Sci 18:3511–3538

Piniewski M, Laizé CLR, Acreman MC, Okruszko T, Schneider C (2014) Effect of climate change on environmental flow indicators in the Narew Basin, Poland. J Environ Qual 43(1):155–167. doi:10.2134/jeq2011.0386

Portmann FT, Döll P, Eisner S, Flörke M (2013) Impact of climate change on renewable groundwater resources: assessing the benefits of avoided greenhouse gas emissions using selected CMIP5 climate projections. Environ Res Lett 8(2):024023

Prudhomme C, Giuntoli I, Robinson EL, Clark DB, Arneli NW, Dankers R, Fekete BM, Franssen W, Gerten D, Gosling SN, Hagemann S, Hannah DM, Kim H, Masaki Y, Satoh Y, Stacke T (2014) Hydrological droughts in the 21st century: hotspots and uncertainties from a global multi-model ensemble experiment. PNAS 111(9):3262–3267

Rakovec O, Kumar R, Mai J, Cuntz M, Thober S, Zink M, Attinger S, Schäfer D, Schrön M, Samaniego L (2016a) Multiscale and multivariate evaluation of water fluxes and states over European river basins J. Hydrometeorol. 17(1):287–307

Rakovec O, Kumar R, Attinger S, Samaniego L, (2016b) Improving the realism of hydrologic model functioning through multivariate parameter estimation Water Resour. Res. 52 (10):7779–7792

Schellnhuber HJ, Frieler K, Kabat P (2014) The elephant, the blind, and the ISI-MIP. PNAS 111(9):3225–3227

Schewe J, et al (2014) Multi-model assessment of water scarcity under climate change. PNAS 111(2014). doi:10.1073/pnas.1222460110

Strauch M, Kumar R, Eisner S, Mulligan M, Reinhardt J, Santini W, Vetter T, Friesen J (2016) Adjustment of global precipitation data for enhanced hydrologic modeling of tropical Andean watersheds. Clim Chang (Online first): doi:10.1007/s10584-016-1706-1

Tang Q, Oki T, Kanae S, Hu H (2007) The influence of precipitation variability and partial irrigation within grid cells on a hydrological simulation. J Hydrometeorol 8:499–512. doi:10.1175/JHM589.1

Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. Bull Am Meteorol Soc 93:485–498

Thoms MC, Sheldon F (2000) Water resource development and hydrological change in a large dryland river: the Barwon–Darling River, Australia. J Hydrol 228(1–2):10–21. doi:10.1016/S0022-1694(99)00191-2

van Vuuren D, Edmonds J, Kainuma M, Riahi K, Thomson A, Hibbard K, Hurtt G, Kram T, Krey V, Lamarque J-F, Masui T, Meinshausen M, Nakicenovic N, Smith S, Rose S (2011) The representative concentration pathways: an overview. Clim Chang 109:5–31. doi:10.1007/s10584-011-0148-z

Verzano K (2009) Climate change impacts on flood related hydrological processes: Further development and application of a global scale hydrological model, no. 71–2009 in Reports on Earth System Science, 71. Max Planck Institute for Taylor K E, Stouffer R J and Meehl G A (2012) An overview of CMIP5 and the experiment design. Bull Am Meteorol Soc 93:485–498

Vetter T, Huang S, Aich V, Yang T, Wang X, Krysanova V, Hattermann FF (2014) Multi-model climate impact assessment and intercomparison for three largescale river basins on three continents. Earth Syst Dyn Discuss 5(2):849–900

Wada Y, Wisser D, Eisner S, Flörke M, Gerten D, Haddeland I, Hanasaki N, Masaki Y, Portmann FT, Stacke T, Tessler Z, Schewe J (2013) Multi-model projections and uncertainties of irrigation water demand under climate change. Geophys Res Lett 40(17):4626–4632

Warszawski L, Frieler K, Huber V, Piontek F, Serdeczny O, Schewe J (2014) The inter-sectoral impact model intercomparison project (ISI-MIP): project framework. PNAS 111(9):3228–3232

Weedon GP, Gomes S, Viterbo P, Shuttleworth WJ, Blyth E, Osterle H, Adam JC, Bellouin N, Boucher O, Best M (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 12:823–848. doi:10.1175/2011 JHM1369.1