CrossMark

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

Simon N. Gosling[1] · Jamal Zaherpour[1] ·
Nick J. Mount[1] · Fred F. Hattermann[2] ·
Rutger Dankers[3] · Berit Arheimer[4] · Lutz Breuer[5,6] ·
Jie Ding[7,8] · Ingjerd Haddeland[9] · Rohini Kumar[10] ·
Dipangkar Kundu[11] · Junguo Liu[7,12] ·
Ann van Griensven[13,14] · Ted I. E. Veldkamp[15] ·
Tobias Vetter[2] · Xiaoyan Wang[16] · Xinxin Zhang[12]

Abstract We present one of the first climate change impact assessments on river runoff that utilises an ensemble of global hydrological models (Glob-HMs) and an ensemble of catchment-scale hydrological models (Cat-HMs), across multiple catchments: the upper Amazon, Darling, Ganges, Lena, upper Mississippi, upper Niger, Rhine and Tagus. Relative changes in simulated mean annual runoff (MAR) and four indicators of high and low extreme

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

✉ Simon N. Gosling
simon.gosling@nottingham.ac.uk

[1] School of Geography, University of Nottingham, Nottingham NG7 2RD, UK

[2] Potsdam Institute for Climate Impact Research, Telegrafenberg, Potsdam A62, D-14473, Germany

[3] Met Office, Exeter, UK

[4] Swedish Meteorological and Hydrological Institute (SMHI), Norrköping 60176, Sweden

[5] Institute for Landscape Ecology and Resources Management (ILR), Research Centre for BioSystems, Land Use and Nutrition (iFZ), Justus Liebig University, Giessen, Germany

[6] Centre for International Development and Environmental Research (ZEU), Justus Liebig University, Giessen, Germany

[7] School of Environmental Science and Engineering, South University of Science and Technology of China, Shenzhen 518055, China

🙋 Springer

flows are compared between the two ensembles. The ensemble median values of changes in runoff with three different scenarios of global-mean warming (1, 2 and 3 °C above pre-industrial levels) are generally similar between the two ensembles, although the ensemble spread is often larger for the Glob-HM ensemble. In addition the ensemble spread is normally larger than the difference between the two ensemble medians. Whilst we find compelling evidence for projected runoff changes for the Rhine (decrease), Tagus (decrease) and Lena (increase) with global warming, the sign and magnitude of change for the other catchments is unclear. Our model results highlight that for these three catchments in particular, global climate change mitigation, which limits global-mean temperature rise to below 2 °C above preindustrial levels, could avoid some of the hydrological hazards that could be seen with higher magnitudes of global warming.

1 Introduction

Article 2 of the 2015 United Nations Framework Convention on Climate Change (UNFCCC) Paris Agreement includes an action to limit any future increase in global-mean temperature to well below 2 °C above pre-industrial levels and to pursue efforts to limit to 1.5 °C, recognising that this would significantly reduce risks and impacts of climate change (UNFCCC 2015). However, if the latest Government climate action pledges from 185 countries were implemented then global-mean warming would still reach 2.7 °C (CAT 2015). Whilst this is almost 1 °C lower than an alternative future in which only existing policies remain enacted (CAT 2015), it misses the UNFCCC targets. There is therefore significant interest in understanding the potential impacts of different amounts of global-mean warming.

Whilst the UNFCCC target is framed in terms of global-mean temperature, the impacts will be felt heterogeneously across the world (Arnell et al. 2016) and one of the key impacts of global warming will be on water resources (Arnell and Gosling 2013). Within this context, we present an assessment of the impact of different levels of global warming on river runoff, focusing on eight major river catchments across the world. The assessment has two main aims.

The first aim is to understand if there are systematic differences between projections of runoff change from two ensembles that are comprised of different types of hydrological model. Here, 'type' of model refers to whether a model is designed to operate at the global-scale

[8]   Institute of Water Resources Management, Hydrology and Agricultural Hydraulic Engineering, Leibniz Hannover University, Hannover 30167, Germany

[9]   Norwegian Water Resources and Energy Directorate, Oslo, Norway

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

[11]   Department of Environmental Sciences, Faculty of Agriculture and Environment, The University of Sydney, Sydney, Australia

[12]   School of Nature Conservation, Beijing Forestry University, Beijing 10083, China

[13]   Department of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, Brussels, Belgium

[14]   UNESCO-IHE Institute for Water Education, Delft, Netherlands

[15]   Vrije Universiteit, Institute for Environmental Studies (IVM), Amsterdam, Netherlands

[16]   State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Center for Global Change and Water Cycle, Hohai University, Nanjing 210098, China

(global-scale hydrological model, Glob-HM) or at the catchment-scale (catchment-scale hydrological model, Cat-HM). Previous assessments have employed *either* ensembles comprised exclusively of Cat-HMs (e.g. Dams et al. 2015) *or* Glob-HMs (including global land surface models; e.g. Haddeland et al. 2011; Prudhomme et al. 2014) to assess the impact of climate change on runoff. Whilst some studies have employed one Glob-HM and one Cat-HM (Gosling et al. 2011), or one Glob-HM with two Cat-HMs (Thompson et al. 2013), none have compared impacts across two large multi-model ensembles of Glob-HMs and Cat-HMs. This is in part due to the significant resources that are required to setup and calibrate Cat-HMs. Knowledge of the relative spreads in projections from the two types of ensemble can help users decide which types of model to use in their assessments and can also provide decision-makers with an appreciation of the possible range in impacts that can arise from using different tools available to the hydrological modelling scientific community.

The second aim is to investigate the effect of different amounts of global-mean warming on runoff for each catchment. We investigate the impacts of three global warming scenarios from a consistent set of global climate model (GCM) projections: 1, 2 and 3 °C warming above pre-industrial levels. The eight catchments are of significance to human settlements and major ecosystems: the Upper Amazon, Darling, Ganges, Lena, Upper Mississippi, Upper Niger, Rhine and Tagus. The approach allows us to explore the extent to which current runoff will be affected if the UNFCCC 2 °C target is met, and what the impacts might be if it is missed and global-mean warming rises to 3 °C.

In this context the application of a Glob-HM ensemble and Cat-HM ensemble to assess global warming impacts across several catchments is a significant advancement because with a companion study that we conducted in parallel (Hattermann et al. 2016) it is the first time that impacts across two large multi-model ensembles of Glob-HMs and Cat-HMs have been compared. Up until now the two types of ensemble have been considered only independently. This advancement has been afforded by a large international collaboration of multiple Glob-HM and Cat-HM modelling groups brought together under the framework of the Inter-Sectoral Impact Model Inter-comparison Project (ISIMIP) over several years (Warszawski et al. 2014).

2 Methods

2.1 Study catchments

The eight large (>50,000 km^2) river catchments (Table 1; see Online Resource 1 for a map) cover all continents and several global hydro-regions defined by Meybeck et al. (2013), including Equatorial (Upper Amazon), Southern Mid-Latitudinal (Darling), Northern Mid-Latitudinal (Ganges, Upper Mississippi, Rhine, Tagus), Boreal (Lena) and Northern Sub-tropical (Upper Niger).

Gauging stations were identified for each catchment for which the Global Runoff Data Centre (GRDC) held observed discharge data, which were then later used to calibrate the Cat-HMs used in this analysis (Krysanova and Hattermann, this special issue). Availability of this data meant that sub-catchments had to be modelled instead of entire systems (Online Resource 1) and for three catchments (Upper Amazon, Upper Mississippi and Upper Niger) only the upper parts were modelled because of complex geomorphological structures and numerous human alterations further downstream, which would require extended input datasets and

Table 1 Catchments, gauges, upstream area of gauge, and the Glob-HMs and Cat-HMs that comprised each ensemble (shaded)

	Catchment	Upper Amazon	Darling	Ganges	Lena	Upper Mississippi	Upper Niger	Rhine	Tagus
	Gauge	Sao Paulo De Olivenca	Louth	Farakka	Stolb	Alton	Koulikoro	Lobith	Almourol
	Upstream drainage area [km²]	990,781	489,300	835,000	2,460,000	444,185	120,000	160,800	67,490
Cat-HM ensemble	HBV (Bergstrom and Forsman 1973)								
	HYMOD (Boyle 2001)								
	mHM (Kumar et al. 2013; Samaniego et al. 2010)								
	SWIM (Krysanova et al. 1998)								
	VIC-Cat-HM [*] (Liang et al. 1994)								
	WaterGAP3 [+] (Verzano 2009)								
	HYPE (Lindstrom et al. 2010)								
	SWAT (Arnold et al. 1993)								
	Number of Cat-HM simulations	7	3	6	3	7	6	7	3
Glob-HM ensemble	LPJmL (Bondeau et al. 2007)								
	Mac-PDM.09 (Gosling and Arnell 2011)								
	MPI-HM (Hagemann and Dümenil 1997)								
	PCR-GLOBWB (Wada et al. 2014)								
	WBMplus (Wisser et al. 2010)								
	H08 (Hanasaki et al. 2008)								
	MATSIRO (Pokhrel et al. 2011)								
	VIC-Glob-HM [*] (Liang et al. 1994)								
	DBH (Tang et al. 2007)								
	Number of Glob-HM simulations	9	9	9	9	9	9	9	9
	Total number of simulations	16	12	15	12	16	15	16	12

[*] VIC-Cat-HM was developed from a Glob-HM counter-part (VIC–Glob-HM). VIC-Cat-HM was calibrated for each catchment and run in "catchment mode" similar to previous applications (Abdulla et al. 1996), whereas VIC-Glob-HM was not calibrated to the study catchments and it was run in "global mode". Thus they can be considered as different types of models. However, the Cat-HM and Glob-HM versions of this model are not completely independent but neither are any of the hydrological models because they all share some similar model parameterisations (e.g. method for calculating potential evapotranspiration)

[+] The WaterGAP model can operate at the global-scale but the version employed here (WaterGAP3) operates on a 5 arcminute global grid, specifically so that it can be run at the catchment-scale. This is a much finer resolution than what the global-scale models used in the analysis operate at (0.5°x0.5°), so it is classed as a catchment-scale model in this analysis

significantly more resource intensive modelling approaches (Krysanova and Hattermann, this special issue).

2.2 Hydrological models

A total of nine Glob-HMs and eight Cat-HMs participated in the assessment (Table 1). The Cat-HMs comprised one ensemble and the Glob-HMs another. Not all Cat-HMs were run for every catchment, however, because unlike Glob-HMs they need to be calibrated on a catchment-by-catchment basis and run for each catchment individually, which is a resource- and time-intensive activity. This means that the Glob-HM ensembles contain nine simulations per catchment/scenario/GCM combination while the Cat-HM ensembles contain from 3 to 7 (all eight Cat-HMs were not run for any of the catchments - the number of Cat-HM simulations for each catchment is listed in Table 1). The Cat-HMs were calibrated and the performance of the calibration evaluated in a separate validation period (Huang et al. 2016) while the Glob-HMs were generally not calibrated to catchment-specific conditions. Water management and other human alterations on the fluvial system were not modelled by all Glob-HMs and Cat-HMs. Flows for every catchment were extracted and processed from the model simulated fields of daily discharge (m^3/s) time series and then converted to daily runoff (mm/day) using the upstream drainage area for each simulation.

We did not weight or exclude individual hydrological models based upon their performance at simulating present-day runoff (Hattermann et al. 2016) and we treated all hydrological models as being independent even though many share similar model parameterisations (e.g. method for calculating potential evapotranspiration; Online Resource 2).

All the models share similarities in their fundamental approach to modeling land-surface hydrological processes, such as simulating the land surface water balance, representing soil and vegetation across the catchment, calculating evapotranspiration, and applying routing schemes to transfer locally generated runoff over the catchment to the catchment outlet. However, the Glob-HMs and Cat-HMs differ in a fundamental way, which is the spatial scale at which the models represent hydrological processes and in turn the water balance resolved. All the Glob-HMs applied here operate with a 0.5°x0.5° spatial resolution grid across the global land-surface. The Cat-HMs employ various approaches. Three Cat-HMs run on a grid (mHM, VIC-Cat-HM, WaterGAP3), while four operate by splitting the catchment into sub-catchments and hydrological response units (HBV, HYPE, SWAT and SWIM), and one treats the entire catchment as a single entity (HYMOD).

An evaluation of the performance of the Glob-HMs and Cat-HMs across 11 catchments (Hattermann et al. 2016) shows that the correlation coefficients between simulated and observed long-term average seasonal dynamics, averaged over all models, is greater than 0.9 in 10 catchments for the Cat-HMs and in 4 catchments for the Glob-HMs (the coefficients are greater than 0.8 in 11 (Cat-HMs) and 8 (Glob-HMs) catchments). While the sensitivity of the Glob-HMs and Cat-HMs to observed climate variability is in general similar, the Glob-HM ensemble mean shows a large positive bias to observed data in annual flows for almost all 11 catchments and the spread across Glob-HMs is wider than for the Cat-HMs in the historical period (Hattermann et al. 2016). Detailed descriptions of the Glob-HMs and Cat-HMs employed in this study are provided by the respective references to each model cited in Table 1 and in Online Resource 2, whilst brief summaries are provided by Krysanova and Hattermann (this special issue).

2.3 Global warming scenarios

Daily input data on the climate variables required by each hydrological model (Online Resource 2) were extracted from five GCM simulations: HadGEM2-ES, IPSL-CM5A-LR, MIROC-ESM-CHEM, GFDL-ESM2 and NorESM1-M; all run under the RCP8.5 emissions pathway (Riahi et al. 2011) for the period 1971-2099 because this is the only pathway for which all five GCMs reach 3 °C by the end of their simulation period (year 2100). All variables were bias-corrected towards the WATCH observation-based dataset (Weedon et al. 2011), using an established method (Hempel et al. 2013), specifically designed to preserve long-term trends in temperature and precipitation projections to facilitate climate change impact assessments.

Daily hydrological simulations were performed with each Glob-HM and Cat-HM, for each catchment, using the daily climate data from each GCM as input (five simulations per Glob-HM/Cat-HM). Summaries of temperature and precipitation change are reported by Krysanova and Hattermann (this special issue). Daily discharge for the 31-year periods centred on the year whose global-mean temperature corresponds to different levels of global-mean warming relative to pre-industrial (1 °C, 2 °C, 3 °C) were extracted from each Glob-HM/Cat-HM simulation (Online Resource 3). At 3 °C the periods are centred around 2050 for three GCMs and around 2075 for two GCMs. Also extracted from each Glob-HM/Cat-HM simulation was the period 1980–2010 to represent the "present-day" (which corresponds to 0.6 °C above pre-industrial). This approach allowed us to estimate the effect of different amounts of global-mean warming (relative to pre-industrial) on catchment runoff, with the runoff changes characterised as changes relative to present-day (note that global-mean warming translates into different levels of warming across the different catchments).

2.4 Hydrological indicators

Five hydrological indicators were calculated from the daily timeseries of runoff:

1. Mean annual runoff (MAR): the mean over 31-years of the total daily runoff for each year in the timeseries.
2. Q5: the magnitude of daily runoff that is exceeded 5 % of the time in the timeseries, and thus an indicator of high flow.
3. Q95: the magnitude of daily runoff that is exceeded 95 % of the time in the timeseries, and thus an indicator of low flow.
4. The magnitude of maximum daily runoff associated with the 2, 5, 10, 20, 25, and 50-year return periods (see Online Resource 4 for specific details).
5. The magnitude of the minimum 7-day moving average of runoff associated with the 2, 5, 10, 20, 25, and 50-year return periods (see Online Resource 4 for specific details).

3 Results

3.1 Statistical distributions of changes in MAR, Q5 and Q95 with global warming

The median values of the two ensembles, corresponding to Glob-HMs and Cat-HMs, mostly respond consistently with each other to global warming (Fig. 1). We applied a Wilcoxon rank sum test to the 31-year median values of change in each hydrological indicator relative to
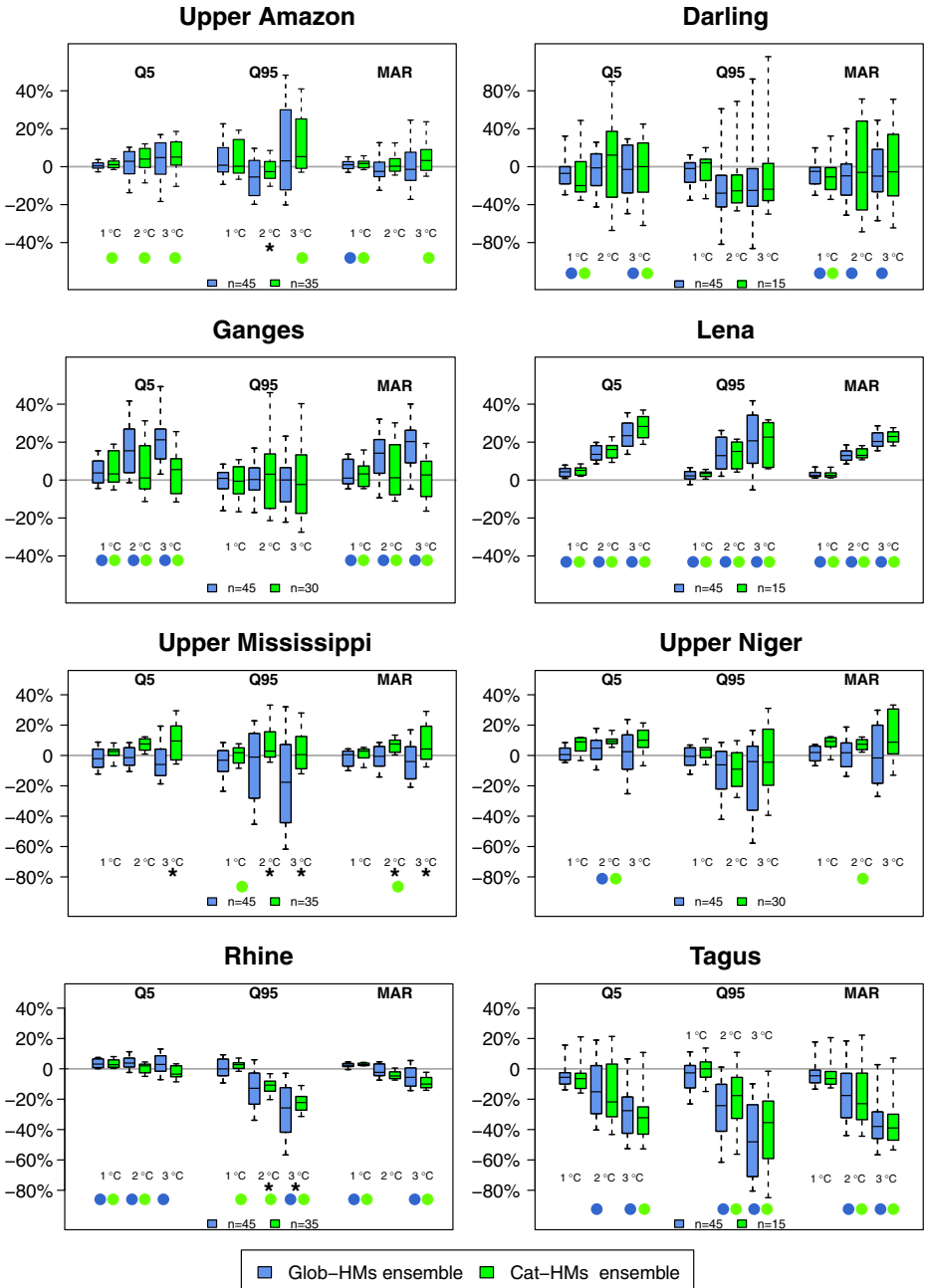
Fig. 1 Change (%; vertical axis) from present-day in three hydrological indicators (Q5, Q95, MAR) with three global warming scenarios (horizontal axis) for eight catchments. The box-whiskers show the 10th, 25th, 50th, 75th and 90th percentiles of the distribution of changes in each hydrological indicator, for the Glob-HM and Cat-HM ensembles respectively (n denotes size of the ensemble, i.e. the number of GCM-hydrological model combinations). Asterisks denote where the Wilcoxon rank sum test rejects the null hypothesis of equal medians between the two ensembles at the 0.05 significance level. Filled circles denote where the projections under climate change for the models that comprise each ensemble represent a significant change from present-day according to a paired-samples t-test (0.05 significance level)

present (MAR, Q5 and Q95, respectively) as estimated by each ensemble. This was to test whether the median values for each ensemble were statistically different from each other. Each ensemble was treated independently (Muerth et al. 2013) because although the driving GCMs were the same, the models that yield the runoff simulations are independent. In the majority of catchments (5/8) there is no statistically significant ($p < 0.05$) difference between the medians of the two ensembles, for all three hydrological indicators and all global warming scenarios. Only for one catchment (Upper Mississippi) is there a significant difference for all three hydrological indicators.

Whilst there is some consistency in the median values between the two ensembles, the spreads are generally wider for the Glob-HM ensemble than the Cat-HM ensemble for the majority of catchments (Upper Amazon, Lena, Upper Mississippi, Upper Niger, Rhine, Tagus). It should also be noted that in some cases the medians of the two model ensembles differ in sign, so that one ensemble projects an increase in runoff and the other a decrease (for instance for MAR in Upper Amazon, Upper Mississippi, and Upper Niger).

Some catchments show a clear trend in their response to increases in global warming, with large differences between 2 and 3 °C that are consistent across both ensembles, as well as significant changes in hydrological indicators from present-day (assessed by a paired-samples t-test between the ensemble projections for each magnitude of global warming and the present-day simulations; Fig. 1). Examples include the Rhine, where there is an increased risk of decreases in low flows, with the median Q95 change escalating from around −11 % at 2 °C to −23 % at 3 °C (the change relative to present-day at 3 °C is statistically significant for both ensembles for Q95, $p < 0.05$). The risk of increases in high flows increases for the Lena, where the change in median Q5 intensifies from around +17 % (2 °C) to +26 % (3 °C), with a similar magnitude increase in MAR also (the changes are statistically significant for both ensembles at all magnitudes of global warming for all hydrological indicators, $p < 0.05$). The Tagus experiences declines in MAR and an enhanced risk of decreases in low flows, with changes in both MAR and Q95 of around −20 % at 2 °C, and then declining even further to around −40 % at 3 °C (the changes relative to present-day at 2 and 3 °C are statistically significant for both ensembles for MAR and Q95, $p < 0.05$). On the other hand, the Upper Mississippi and Upper Niger experience relatively small changes in the medians of hydrological indicators between 2 and 3 °C and the magnitude of changes relative to present are infrequently statistically significant.

## 3.2 Probability of changes in MAR, Q5 and Q95 with global warming

As an alternative approach to investigate the distribution of impacts from the Glob-HM and Cat-HM ensembles (as in Fig. 1), we calculated the probability of impacts of different magnitudes occurring from the impact distributions of the two ensembles (Fig. 2). For example, if 18 out of 45 simulations in an ensemble show that MAR increases by up to 30 % with global warming, then the probability of this magnitude of change is 40 %. Strictly, the probabilities should be interpreted as an indicator of the level of model agreement within an ensemble, not the level of certainty of a particular outcome occurring.

The two ensembles show similar responses to global warming for all three hydrological indicators for the Ganges, Lena and Tagus. However, for some catchments there is strong agreement between the ensembles for some indicators (such as increases in MAR and Q5 for the Upper Amazon; and decreases in MAR for the Rhine) but not for other indicators in the same catchment.
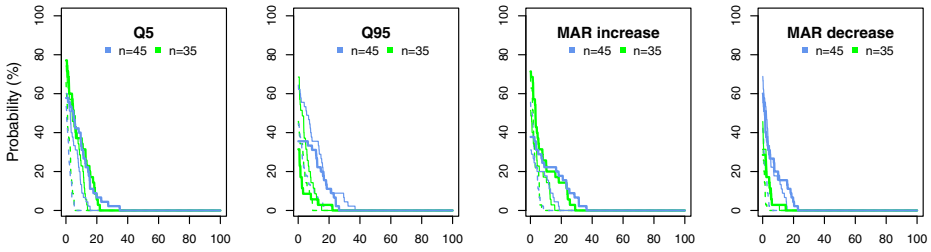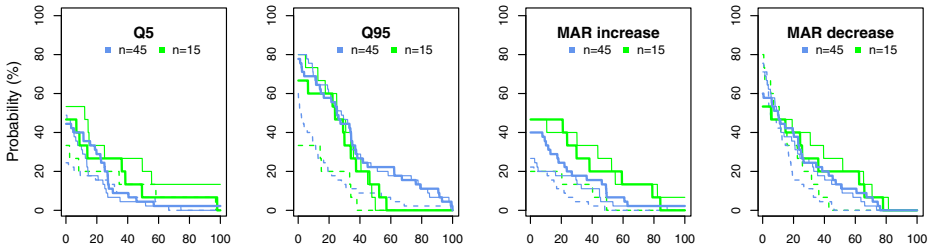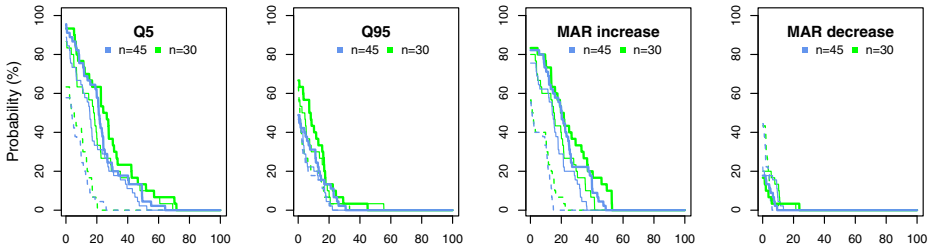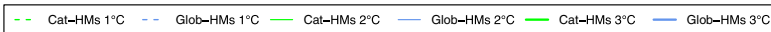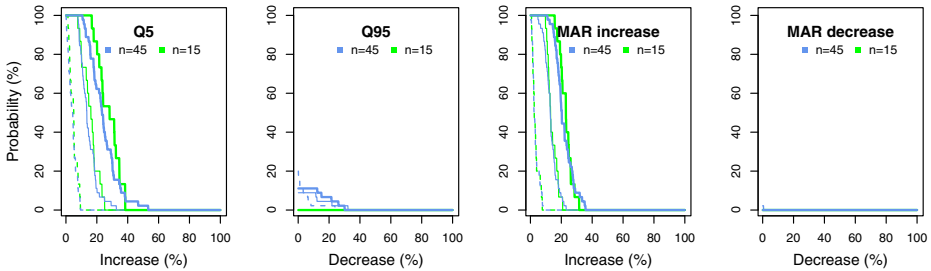
Fig. 2 The probability (%, vertical axes) under three global warming scenarios, of present-day Q5 increasing, Q95 decreasing, MAR increasing, and MAR decreasing, by different magnitudes (1–100 % of present day values, in 0.1 % increments; horizontal axes). Probability is calculated from the distribution of changes in each hydrological indicator as simulated by the Glob-HM and Cat-HM ensembles respectively (n denotes size of the ensemble)
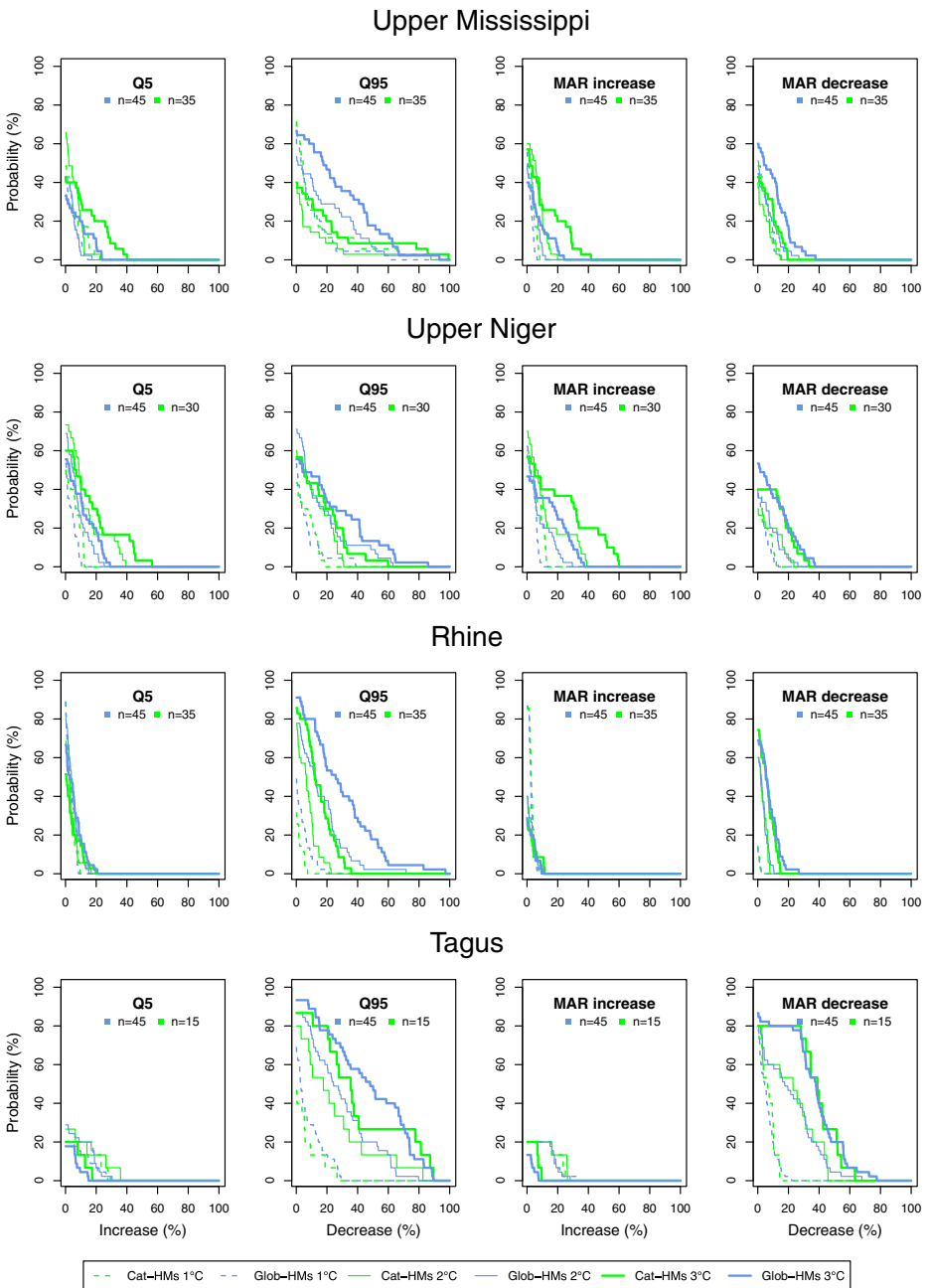
Fig. 2 (continued)

The probabilities of large decreases (>20 %) in Q95 are greater with the Glob-HM ensemble than with the Cat-HM ensemble for the majority (5) of catchments. For example at 3 °C, for the Darling, Upper Mississippi, Upper Niger, and Rhine, the probabilities of Q95
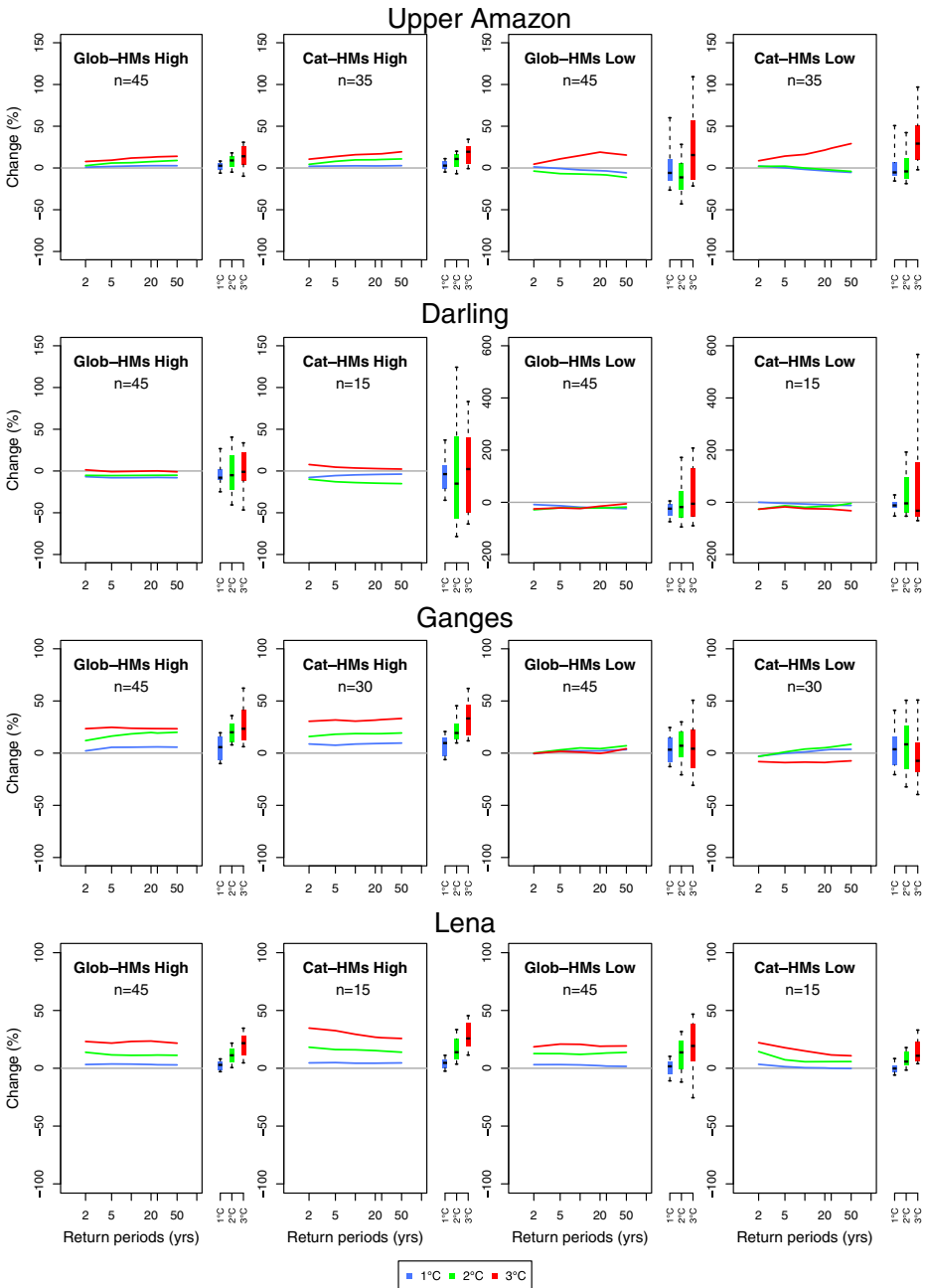
Fig. 3 The change relative to present-day (vertical axes) of maximum daily runoff ("High") and minimum 7-day moving average of runoff ("Low") associated with 2, 5, 10, 20, 25, and 50-year return periods (horizontal axes; log-scale) for the Glob-HM ensemble and Cat-HM ensemble. The lines show the median for each ensemble and the box-whiskers show the 10th, 25th, 50th, 75th and 90th percentiles of the distribution of changes for the 50-year return period. The absolute values for each scenario and present-day (i.e. not change) are shown in Online Resource 5
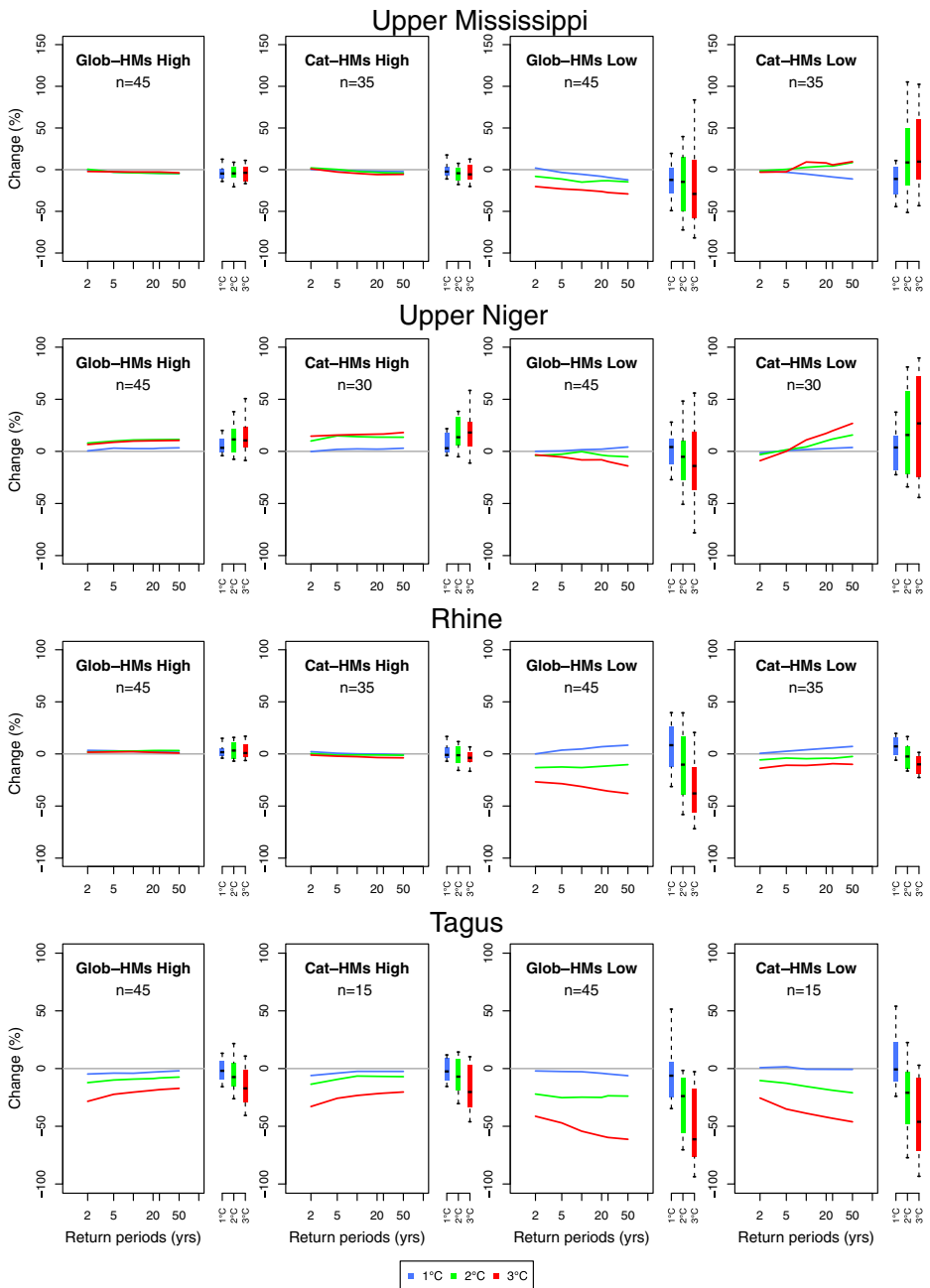
Fig. 3 (continued)

decreasing by up to 40 % according to the Glob-HM ensemble (Cat-HM ensemble) are: 26 % (20 %), 31 % (9 %), 24 % (7 %), and 29 % (0 %), respectively. This reverses in some cases for

large (>20 %) increases in Q5 though, where the probabilities are higher for the Cat-HM ensemble (Darling, Upper Mississippi, Upper Niger).

The probability of large decreases in MAR (>20 %) with 3 °C warming simulated by the Glob-HMs (Cat-HMs) are substantial for the Darling, Upper Niger and Tagus: 40 % (40 %), 18 % (20 %), and 80 % (80 %), respectively. These probabilities decline considerably when global warming is limited to 2 °C for the Upper Niger (from around 18 % at 3 °C to 4 % at 2 °C) and Tagus (from around 80 % at 3 °C to 47 % at 2 °C) (but not for the Darling where at 2 °C the probability is similar to at 3 °C). Likewise, substantial probabilities of large increases in MAR (>20 %) for the Ganges, Lena and Upper Amazon are considerably reduced at lower levels of global warming.

3.3 Changes in magnitude of return period runoff

The box-whiskers in Fig. 3 show that the Glob-HM and Cat-HM ensembles in most cases project changes that are similar in magnitude to each other. Two exceptions are for the Upper Mississippi and Upper Niger, where the Glob-HM ensemble projects declines in the magnitude of minimum flows associated with return periods greater than 5-years, while the Cat-HM ensemble projects increases. Also, for some catchments the spread of the Glob-HM ensemble is wider than for the Cat-HM ensemble (Upper Amazon, Rhine, Tagus). Whilst it appears for some catchments as though global warming has little effect on minimum (Darling, Ganges) and maximum flows (Upper Mississippi, Rhine), the medians disguise the large spreads of the ensembles.

Both ensembles project clear increases in the magnitude of maximum flows with global warming for the Upper Amazon, Ganges, and Lena, and decreases in the magnitude of minimum flows under 2 and 3 °C for the Rhine, and Tagus (Fig. 3), which are consistent with projected changes in Q5 and Q95 (Fig. 1). Some of the impacts intensify substantially with each degree of global warming. For example, the median changes from the Glob-HM ensemble in the 50-year return period minimum flow for 2 °C (and 3 °C) warming is −30 % (−60 %) for the Tagus and −10 % (−35 %) for the Rhine. Notably, the Upper Amazon shows a small decrease in the magnitude of minimum flows at the lower levels of global warming but a strong increase at 3 °C warming, due to a significant increase in precipitation between 2 and 3 °C warming (Krysanova and Hattermann, this special issue).

4 Discussion

4.1 Similarities and differences between the impacts simulated by the two ensembles

Our primary aim was to compare projections of runoff change from two ensembles of different types of hydrological models. We tested this by comparing impacts from an ensemble of Glob-HMs with an ensemble of Cat-HMs, both run with consistent input data from five GCMs.

The median estimates from both ensembles generally show that there is similarity in the effect of global warming on the hydrological indicators considered. This supports earlier work that showed climate change projections from a single Glob-HM were generally similar to those from a single Cat-HM (Gosling et al. 2011) but here the result is based upon a significantly more robust approach that employed ensembles of multiple Glob-HMs and Cat-HMs. There are a number of subtleties that

underlay the general conclusion that the two ensembles produce similar projections, however.

Firstly, the spread in projections of any one hydrological indicator for each ensemble is usually larger than the difference between the two ensemble medians. This means that the choice of hydrological models that comprise an ensemble, and the GCMs that provide the input data, are an important factor in determining the spread of the projections from the ensemble. Other work has quantified the relative contributions of hydrological models and GCMs to ensemble spread (Hagemann et al. 2013) with evidence suggesting that their relative contributions vary spatially across the globe (Dankers et al. 2014). Only three Cat-HMs comprised the Cat-HM ensemble for three catchments (Darling, Lena, Tagus). Whilst the inter-ensemble comparison would have been more consistent had we omitted these three catchments so that a similar number of models were included in each ensemble, our preference was to include as many catchments as possible to cover as many hydro-regions as possible. Clearly there is a loss of potential information where only three Cat-HMs are used compared to where six or seven are used. Thus we acknowledge that for these three catchments the Cat-HM ensemble spread could have been different if other Cat-HMs had been included in the ensemble.

Secondly, the difference in medians between the two ensembles is comparable to the climate change impact (difference between 1 and 3 °C) for the Ganges, Upper Mississippi and Upper Niger and there is sometimes a clear difference in sign of change between the ensembles at 3 °C warming (Upper Amazon MAR, Upper Mississippi MAR and Q5, Upper Niger MAR and Rhine Q5).

Thirdly, there are a small number of cases where the difference between the two ensembles could be large in absolute terms (mm), despite the Wilcoxon rank sum test indicating that the differences are not significant ($p < 0.05$), namely for the projections of Q5 and MAR for the Ganges.

Whilst Hattermann et al. (2016) investigated the relative abilities of the Glob-HMs and Cat-HMs to simulate present-day runoff in terms of long-term average seasonal dynamics, we did not weight and/or include/exclude individual models accordingly from each ensemble. This is because we wanted to assess impacts irrespective of individual model performance. A separate study is currently weighting models according to their performance. All hydrological models were treated as being independent even though many share similar model parameterisations (e.g. method for calculating potential evapotranspiration; Online Resource 2).

Despite these caveats the results highlight that the two modelling communities (Glob-HM and Cat-HM) seem to be approaching each other, both in terms of simulation results and also in their conceptual approach to model development. This is because some Cat-HMs can now be run at the continental scale with high resolution, such as a version of HYPE (E-HYPE, different from what was applied here; Donnelly et al. 2016; Hundecha et al. 2016) that can be run across around 35,000 sub-catchments in Europe, whilst some Glob-HMs are now starting to develop rigorous calibration procedures (Müller Schmied et al. 2014).

4.2 Comparing ensemble spread

Whilst the medians of the two ensembles generally display similar responses to global warming, the spread in projections from the Glob-HM ensemble is wider than from the Cat-HM ensemble in the majority of cases. A possible reason is that the Cat-HMs are constrained (calibrated) to specific catchment conditions (e.g., to reproduce observed discharge

characteristics) (Krysanova and Hattermann, this special issue), which means that they perform better than the Glob-HMs under present-day climate conditions (Hattermann et al. 2016).

It is also possibly the result of sampling bias introduced by the different sizes of the two ensembles, where the relatively smaller number of Cat-HMs represents only a limited section of the true hydrological model spread that would arise from using all the Cat-HMs that are in existence (Butts et al. 2004). Indeed, our study is not exhaustive in its inclusion of all available Glob-HMs and Cat-HMs, so the spread of the ensembles are likely underestimated.

However, even if we had included every Glob-HM and Cat-HM in existence, it remains possible that all models may miss (or not represent well enough) certain key processes, for instance the response of glaciers to global warming and in turn on runoff – thus even a "complete" ensemble may not sample the true spread (this also applies to the ensemble of GCMs that provide input to the hydrological models). Nevertheless, our study and a companion study that we conducted in parallel (Hattermann et al. 2016) are the first multi-catchment assessments to date of the impacts of climate change derived from multiple catchment-scale and global-scale hydrological models.

4.3 The impacts of global warming

We sought to demonstrate the effects of different amounts of global-mean warming on runoff for each catchment. Both ensembles clearly show that the Tagus and Rhine could see large increases in the magnitude of impacts between 1, 2, and 3 °C warming for all the considered hydrological indicators. Both catchments may experience declines in the magnitude of low and high flows and MAR with increasing warming (only at 3 °C with the Cat-HMs for high flows for the Rhine) due to large declines in precipitation (Tagus) or little change (Rhine) combined with increases in evapotranspiration due to warmer temperatures that are all simulated consistently by the five GCMs (Krysanova and Hattermann, this special issue). This supports earlier impact assessment studies conducted in these catchments (Kilsby et al. 2007; Kwadijk and Rotmans 1995; Middelkoop et al. 2001).

We also found clear evidence for increases in MAR and high flows for the Lena. The application of different snowmelt schemes by the hydrological models (see Online Resource 2) has little bearing on the projections. Temperature for the Lena rises most out of the eight study catchments under each 1 °C of global-mean warming (Krysanova and Hattermann, this special issue) – this in turn increases snowmelt runoff, confirming previous climate change assessments (Arora and Boer 2001) and observed changes in response to recent warming (Yang et al. 2002).

Our analysis shows no conclusive effects of global warming on runoff for the other catchments because the median changes are often close to zero with the 25th and 75th percentiles of the impacts distributions on either side.

The high spread seen for the Darling arises from it being a challenging catchment to model as it has a very low runoff coefficient and transmission losses are not well represented by many hydrological models (Gosling and Arnell 2011). Moreover, the performance of the Glob-HM ensemble has not been evaluated for its low flow simulation performance in the study catchments. Elsewhere it has been shown that Glob-HMs tend to struggle to simulate low flows (Giuntoli et al. 2015; Gudmundsson et al. 2012).

The large spread for the Ganges may be explained by the models' varying abilities to represent and account for (or not) dynamical changes in glacier mass balance with global

warming, for instance earlier work showed that Glob-HMs differ significantly in their simulations of snow accumulation in the Himalayan region (Haddeland et al. 2011).

It is well documented that the method for calculating potential evapotranspiration plays a significant role in determining simulated surface runoff changes (Haddeland et al. 2011; Thompson et al. 2013) and this can partly explain the large spread in simulated runoff for the Upper Amazon, Ganges, Upper Niger and Upper Mississippi. The hydrological models employ different approaches to simulate potential evapotranspiration (see Online Resource 2) which means that whilst the GCMs consistently simulate relative increases in both precipitation and temperature with global warming for these catchments (Krysanova and Hattermann, this special issue), there is diversity in the simulated surface runoff response.

The overall approach tied the analysis to prescribed global temperature changes so that the time periods for each magnitude of warming vary across GCMs instead of being fixed (Online Resource 3). This yields more robust regional patterns of change than if fixed periods are used because some of the uncertainty related to the GCMs' climate sensitivity is removed (Vautard et al. 2014). Although the approach is established (Arnell et al. 2014; Gosling et al. 2010; Roudier et al. 2015; Schewe et al. 2014) it does come with the caveat that it largely overlooks the rate at which, and when, each global-warming level is reached. Timing, as well as the magnitude of warming, is important (Lenton 2011), because adaptation is generally more straightforward and less costly with more time and natural systems may respond differently to slower rates of climate change. Some studies (Scholze et al. 2006) have overcome this by grouping simulations from multiple GCMs into categories of warming (e.g. <2 °C, 2–3 °C and >3 °C) for a fixed time period (e.g. 2071–2100) but this can reduce the precision and range for which global warming impacts can be assessed, especially where simulations from only a limited number of GCMs are available.

5 Conclusion

Along with a companion study that we conducted in parallel (Hattermann et al., this special issue) we have presented the first multi-catchment assessment of the impacts of climate change derived from multiple catchment- and global-scale hydrological models. Our results shed new light on the utility of global-scale hydrological models to understand changes in runoff at the catchment-scale because the median estimates of change from ensembles of Glob-HMs and Cat-HMs generally showed similar effects of global warming on the considered hydrological indicators. We did find some distinct differences though, such as a wider spread in impacts from the Glob-HMs for several catchments.

Although the sign and magnitude of change with global warming for some catchments is unclear (Upper Amazon, Darling, Ganges, Upper Niger and Upper Mississippi) we found more compelling evidence for the Rhine, Tagus and Lena. For these latter catchments in particular, it is clear that stringent global climate change mitigation action in line with Article 2 of the Paris Agreement (UNFCCC 2015) could have significant benefits in terms of avoiding some of the hydrological hazards that could be seen with higher magnitudes of global warming such as 3 °C.

A rationale for our study was that knowledge of the relative spreads in projections from the two types of ensemble could provide decision-makers with an appreciation of the possible range in impacts that can arise from using different tools available to the hydrological

modelling scientific community. In this context a decision-maker might ask: "If I am to conduct a climate impact study in e.g. the Mississippi catchment, which model type would I favor for assessing future runoff?" In light of our results, we would advise against favoring any one type of ensemble. Whilst the Cat-HM ensemble has been shown to perform better than the Glob-HM ensemble when simulating present-day discharge (Hattermann et al., this special issue), this does not guarantee higher credibility of projections under climate change. To this end we would encourage decision-makers to consider the full range of spreads from both the Glob-HM and Cat-HM ensembles, since these represent the spread of possible impacts according to several different types of models used by the hydrological modelling scientific community.

References

Abdulla FA, Lettenmaier DP, Wood EF, Smith JA (1996) Application of a macroscale hydrologic model to estimate the water balance of the Arkansas-Red River Basin. J Geophys Res: Atmos 101:7449–7459

Arnell NW, Gosling SN (2013) The impacts of climate change on river flow regimes at the global scale. J Hydrol 486:351–364

Arnell NW, Brown S, Gosling SN, Hinkel J, Huntingford C, Lloyd-Hughes B, Lowe JA, Osborn T, Nicholls RJ, Zelazowski P (2014) Global-scale climate impact functions: the relationship between climate forcing and impact. Clim Chang 134:475–487

Arnell NW, Brown S, Gosling SN, Gottschalk P, Hinkel J, Huntingford C, Lloyd-Hughes B, Lowe JA, Nicholls RJ, Osborn TJ, Osborne TM, Rose GA, Smith P, Wheeler TR, Zelazowski P (2016) The impacts of climate change across the globe: a multi-sectoral assessment. Clim Chang 134:457–474

Arnold JG, Allen PM, Bernhardt G (1993) A comprehensive surface-groundwater flow model. J Hydrol 142:47–69

Arora VK, Boer GJ (2001) Effects of simulated climate change on the hydrology of major river basins. J Geophys Res: Atmos 106:3335–3348

Bergstrom S, Forsman A (1973) Development of a conceptual deterministic rainfall-runoff model. Hydrol Res 4:147–170

Bondeau A, Smith PC, Zaehle S, Schaphoff S, Lucht W, Cramer W, Gerten D, Lotze-Campen H, MÜLler C, Reichstein M, Smith B (2007) Modelling the role of agriculture for the 20th century global terrestrial carbon balance. Glob Chang Biol 13:679–706

Boyle DP (2001) Multicriteria calibration of hydrological models. PhD Dissertation, University of Arizona, Tucson, Arizona, USA

Butts MB, Payne JT, Kristensen M, Madsen H (2004) An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. J Hydrol 298:242–266

CAT (2015) Climate Action Tracker

Dams J, Nossent J, Senbeta TB, Willems P, Batelaan O (2015) Multi-model approach to assess the impact of climate change on runoff. J Hydrol 529(Part 3):1601–1616

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. Hydr. Sciences Journal 6:255-273

Giuntoli I, Villarini G, Prudhomme C, Mallakpour I, Hannah DM (2015) Evaluation of global impact models' ability to reproduce runoff characteristics over the central United States. J Geophys Res: Atmos 120:9138–9159

Gosling SN, Arnell NW (2011) Simulating current global river runoff with a global hydrological model: model revisions, validation, and sensitivity analysis. Hydrol Process 25:1129–1145

Gosling SN, Bretherton D, Haines K, Arnell NW (2010) Global hydrology modelling and uncertainty: running multiple ensembles with a campus grid. Philos Trans Royal Soc Lond A: Math Phys Eng Sci 368:4005–4021

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

Gudmundsson L, Tallaksen LM, Stahl K, Clark DB, Dumont E, Hagemann S, Bertrand N, Gerten D, Heinke J, Hanasaki N, Voss F, Koirala S (2012) Comparing large-scale hydrological model simulations to observed runoff percentiles in Europe. J Hydrometeorol 13:604–620

Haddeland I, Clark DB, Franssen W, Ludwig F, Voß F, Arnell NW, Bertrand N, Best M, Folwell S, Gerten D, Gomes S, Gosling SN, Hagemann S, Hanasaki N, Harding R, Heinke J, Kabat P, Koirala S, Oki T, Polcher J, Stacke T, Viterbo P, Weedon GP, Yeh P (2011) Multimodel estimate of the global terrestrial water balance: setup and first results. J Hydrometeorol 12:869–884

Hagemann S, Dümenil L (1997) A parametrization of the lateral waterflow for the global scale. Clim Dyn 14:17–31

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 Dyn 4:129–144

Hanasaki N, Kanae S, Oki T, Masuda K, Motoya K, Shirakawa N, Shen Y, Tanaka K (2008) An integrated model for the assessment of global water resources – Part 1: model description and input meteorological forcing. Hydrol Earth Syst Sci 12:1007–1025

Hattermann FF, Krysanova V, Gosling SN, Dankers R, Daggupati P, Donnelly C, Flörke M, Huang Sh, Motovilov Yu, Buda S, Yang T, Müller C, Leng G, Tang Q, Portmann FT, Hagemann S, Gerten D, Wada Y, Masaki Y, Alemayehu T, Satoh Y, Samaniego L (2016) Cross-scale intercomparison of climate change impacts simulated by regional and global hydrological models in eleven large river basins. Clim Chang. doi:10.1007/s10584-016-1829-4

Hempel S, Frieler K, Warszawski L, Schewe J, Piontek F (2013) A trend-preserving bias correction-the ISI-MIP approach. Earth Syst Dyn 4:219–236

Huang S, Kumar R, Flörke M, Yang T, Hundecha Y, Kraft P, Gao C, Gelfan A, Liersch S, Lobanova A, Strauch M, Ogtrop FV, Reinhardt J, Haberlandt U, Krysanova V (2016) Evaluation of an ensemble of regional hydrological models in 12 large-scale river basins worldwide. Clim Chang. doi:10.1007/s10584-016-1841-8

Hundecha Y, Arheimer B, Donnelly C, Pechlivanidis I (2016) A regional parameter estimation scheme for a pan-European multi-basin model. Journal of Hydrology: Regional Studies, 6: 90-111 doi:10.1016/j.ejrh.2016.04.002

Kilsby CG, Tellier SS, Fowler HJ, Howels TR (2007) Hydrological impacts of climate change on the Tejo and Guadiana Rivers. Hydrol Earth Syst Sci 11:1175–1189

Krysanova V, Müller-Wohlfeil D-I, Becker A (1998) Development and test of a spatially distributed hydrological/water quality model for mesoscale watersheds. Ecol Model 106:261–289

Kumar R, Samaniego L, Attinger S (2013) Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations. Water Resour Res 49:360–379

Kwadijk J, Rotmans J (1995) The impact of climate change on the river rhine: a scenario study. Clim Chang 30:397–425

Lenton TM (2011) Beyond 2°C: redefining dangerous climate change for physical systems. Wiley Interdiscip Rev Clim Chang 2:451–461

Liang X, Lettenmaier DP, Wood EF, Burges SJ (1994) A simple hydrologically based model of land surface water and energy fluxes for general circulation models. J Geophys Res: Atmos 99:14415–14428

Lindstrom G, Pers C, Rosberg J, Stromqvist J, Arheimer B (2010) Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales. Hydrol Res 41:295–319

Meybeck M, Kummu M, Dürr HH (2013) Global hydrobelts and hydroregions: improved reporting scale for water-related issues? Hydrol Earth Syst Sci 17:1093–1111

Middelkoop H, Daamen K, Gellens D, Grabs W, Kwadijk JCJ, Lang H, Parmet BWAH, Schädler B, Schulla J, Wilke K (2001) Impact of climate change on hydrological regimes and water resources management in the Rhine Basin. Clim Chang 49:105–128

Muerth MJ, Gauvin St-Denis B, Ricard S, Velázquez JA, Schmid J, Minville M, Caya D, Chaumont D, Ludwig R, Turcotte R (2013) On the need for bias correction in regional climate scenarios to assess climate change impacts on river runoff. Hydrol Earth Syst Sci 17:1189–1204

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

Pokhrel Y, Hanasaki N, Koirala S, Cho J, Yeh PJF, Kim H, Kanae S, Oki T (2011) Incorporating anthropogenic water regulation modules into a land surface model. J Hydrometeorol 13:255–269

Prudhomme C, Giuntoli I, Robinson EL, Clark DB, Arnell NW, Dankers R, Fekete BM, Franssen W, Gerten D, Gosling SN, Hagemann S, Hannah DM, Kim H, Masaki Y, Satoh Y, Stacke T, Wada Y, Wisser D (2014) Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment. Proc Natl Acad Sci 111:3262–3267

Riahi K, Rao S, Krey V, Cho C, Chirkov V, Fischer G, Kindermann G, Nakicenovic N, Rafaj P (2011) RCP 8.5—A scenario of comparatively high greenhouse gas emissions. Clim Chang 109:33–57

Roudier P, Andersson JM, Donnelly C, Feyen L, Greuell W, Ludwig F (2015) Projections of future floods and hydrological droughts in Europe under a +2°C global warming. Climatic Change:1-15

Samaniego L, Kumar R, Attinger S (2010), Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resour Res 46: W05523

Schewe J, Heinke J, Gerten D, Haddeland I, Arnell NW, Clark DB, Dankers R, Eisner S, Fekete BM, Colón-González FJ, Gosling SN, Kim H, Liu X, Masaki Y, Portmann FT, Satoh Y, Stacke T, Tang Q, Wada Y, Wisser D, Albrecht T, Frieler K, Piontek F, Warszawski L, Kabat P (2014) Multimodel assessment of water scarcity under climate change. Proc Natl Acad Sci 111:3245–3250

Scholze M, Knorr W, Arnell NW, Prentice IC (2006) A climate-change risk analysis for world ecosystems. Proc Natl Acad Sci 103:13116–13120

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

Thompson JR, Green AJ, Kingston DG, Gosling SN (2013) Assessment of uncertainty in river flow projections for the Mekong River using multiple GCMs and hydrological models. J Hydrol 486:1–30

UNFCCC (2015) Adoption of the Paris Agreement. Proposal by the President. Proposal by the President. Available from: http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf. United Nations, Geneva, Switzerland

Vautard R, Gobiet A, Sobolowski S, Kjellström E, Stegehuis A, Watkiss P, Mendlik T, Landgren O, Nikulin G, Teichmann C, Jacob D (2014) The European climate under a 2 °C global warming. Environ Res Lett 9:034006

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. Max Planck Institute for Meteorology, Hamburg, Germany

Wada Y, Wisser D, Bierkens MFP (2014) Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources. Earth Syst Dyn 5:15–40

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

Weedon GP, Gomes S, Viterbo P, Shuttleworth WJ, Blyth E, Österle 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

Wisser D, Fekete BM, Vörösmarty CJ, Schumann AH (2010) Reconstructing 20th century global hydrography: a contribution to the Global Terrestrial Network- Hydrology (GTN-H). Hydrol Earth Syst Sci 14:1–24

Yang D, Kane DL, Hinzman LD, Zhang X, Zhang T, Ye H (2002) Siberian Lena River hydrologic regime and recent change. J Geophys Res: Atmos 107:4694