HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Helmholtz Centre for Environmental Research – UFZ | Permoserstrasse 15 | 04318 Leipzig/Germany

contact person:
Dr. Andreas Marx
Head of the Climate Office for
Central Germany
Department Computational
Hydrosystems
fon +49 341 235 1074
klima@ufz.de

Leipzig, 2017/12/18

Re-submission HESS-2017-485

Dear Louise Slater,

I would like to thank you and the reviewers again for the comments which helped to increase the overall quality of the manuscript.

Please find newly uploaded a revised manuscript, incorporating the proposed changes and additions (including the proposed changes to the figures, the discussion of model performance and limitations, methodological clarification, and terminology), as well as a a marked-up manuscript version (latexdiff file) showing the differences between the originally submitted and the revised manuscript below, as well as the detailed responses to the reviewers’ comments.

Kind regards,

Andreas Marx

Helmholtz Centre for
Environmental Research – UFZ

Company domicile: Leipzig

Permoserstrasse 15
04318 Leipzig | Germany
or
PF 500136
04301 Leipzig | Germany
Tel +49 341 235 0

info@ufz.de
www.ufz.de

Registration court: Leipzig district court
Commercial register No. B 4703

Chairman of the Supervisory Board:
MinDirig Wilfried Kraus

Management Board:
Prof. Dr. Georg Teutsch and
Prof. Dr. Heike Graßmann

Bank details:
HypoVereinsbank Leipzig
Sort code 860 200 86
Account No. 5080 186 136
Swift (BIC) code HYVEDEMM495
IBAN No. DE12860200865080186136
VAT No. DE 141 507 065
Tax No. 232/124/00416

Climate change alters low flows in Europe under a global warming of 1.5, 2, and 3 ~~degree global warming~~degrees

Marx Andreas[1], Kumar Rohini[1], Thober Stephan[1], Rakovec Oldrich[1,5], Wanders Niko[2,3], Zink Matthias[1], Wood Eric F.[3], Pan Ming[3], Sheffield Justin[4], and Samaniego Luis[1]

[1]Department Computational Hydrosystems, Helmholtz Centre for Environmental Research (UFZ), Leipzig, Germany
[2]Department of Physical Geography, Faculty of Geosciences, University Utrecht, The Netherlands
[3]Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, United States
[4]Geography and Environment, University of Southampton, Southampton, United Kingdom
[5]Faculty of Environmental Sciences, Czech University of Life Sciences, Czech Republic

Correspondence to: Andreas Marx (klima@ufz.de)

Abstract. There is growing evidence that climate change will alter water availability in Europe. Here, we investigate how hydrological low flows are affected under different levels of future global warming (i.e., 1.5, 2 and 3 K ~~)~~with respect to the pre-industrial period) in rivers with a contributing area of more than 1000 km^2. The analysis is based on a multi-model ensemble of 45 hydrological simulations based on three ~~RCPs (rcp2p6, rcp6p0, rcp8p5~~representative concentration pathways (RCP2.6, RCP6.0, RCP8.5), five Coupled Model Intercomparison Project Phase 5 (CMIP5~~GCMs (~~) general circulation models (GCMs: GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR, MIROC-ESM-CHEM, NorESM1-M) and three state-of-the-art hydrological models (HMs: mHM, Noah-MP, and PCR-GLOBWB). High resolution model results are available at ~~the unprecedented~~ a spatial resolution of 5 km across the ~~pan-European domain at~~ Pan-European domain at a daily temporal resolution. Low river flow is described as the percentile of daily streamflow that is exceeded 90% of the time. It is determined separately for each GCM/HM ~~combinations~~ combination and the warming scenarios. The results show that the low flow change signal amplifies with increasing warming levels. Low flows decrease in the Mediterranean, while they increase in the Alpine and Northern regions. In the Mediterranean, the level of warming amplifies the signal from -12% under 1.5 K~~to~~, compared to the baseline period 1971-2000, to -35% under a global warming of 3 K~~global warming~~, largely due to the projected decreases in annual precipitation. In contrast, the signal is amplified from +22% (1.5 K) to +45% (3 K) in the Alpine region ~~because of the reduced snow melt contribution~~due to changes in snow accumulation. The changes in low flows are significant for regions with relatively large change signals and under higher levels of warming. ~~Nevertheless~~However, it is not possible to distinguish climate induced differences in low flows between 1.5 and 2 K warming because of ~~the large variability inherent~~ (1) the large inter-annual variability which prevents distinguishing statistical estimates of period-averaged changes for a given GCM-HM combination, and (2) the uncertainty in the multi-model ensemble expressed by the signal-to-noise ratio. The contribution by the GCMs to the uncertainty in the model results is generally higher than the one by the HMs. However, the uncertainty due to HMs cannot be neglected. In the Alpine and Northern region as well as the Mediterranean, the uncertainty contribution by the HMs is partly higher than those by the GCMs due to different representations of processes such as snow, soil moisture and

evapotranspiration. Based on the analysis results, it is recommended (1) to use multiple HMs in climate impact studies and (2) to embrace uncertainty information on the multi-model ensemble as well as it's single members in the adaptation process.

1 Introduction

Hydrological drought is a slowly developing natural phenomenon than can occur anywhere, ~~independent~~ independently of the hydro-climatic regime (Van Loon, 2015). It is expressed as a deficiency in river discharge compared to the expected normal and is mainly caused by lower than average precipitation and soil moisture or strong increases in evapotranspiration. In addition to natural causes, human water use and reservoirs can significantly alter the drought signal in many places (Wanders and Wada, 2015). Droughts are rare events and can propagate from meteorological over soil moisture to hydrological droughts, finally resulting in socio-economic drought (Van Loon, 2015). Hydrological droughts affect the environment and cause damage to society and the economy. van Vliet et al. (2016) showed reduced potentials for thermoelectric power and hydropower generation under hydrological drought worldwide. In Europe, the 2003 drought and heatwave resulted in nearly -6.6% in hydropower and -4.7% in thermoelectric power generation. The total loss of the 2003 severe drought event was estimated to be ~~EUR~~ 8.7 billion Euro in Central and Southern Europe (EC, 2007). More recently, the 2015 drought event ~~(Van Lanen et al., 2016; Zink et al., 2016)~~ (Laaha et al., 2017; Van Lanen et al., 2016; Zink et al., 2016) in Central Europe also caused significant socio-economic and environmental problems. Economic losses due to droughts almost doubled between 1976-1990 and the 1991-2006 period to ~~about EUR~~ approximately 6.2 billion Euro per year. Social and environmental costs are often not considered (EC, 2007). A collection of hydrological drought impacts for Europe can be found in the European drought impact inventory (Stahl et al., 2016) sorted by impact categories, e.g., freshwater aquaculture and fisheries, energy and industry, waterborne transportation, public water supply or freshwater ecosystems. Furthermore, water quality is directly influenced by hydrological drought, e.g., in lowering the availability of the diluting medium water resulting in increasing pollutants concentrations.

Climate change is expected to alter the hydrological cycle throughout Europe. Temperature projections show significant warming for all emission scenarios over Europe. Southern Europe is the hotspot with strongest projected warming in summer, Northern Europe in winter time (Kovats et al., 2014). Jacob et al. (2014) projected mean annual precipitation to decrease under RCP4.5 mainly ~~on~~ in the Iberian Peninsula and Greece until the end of the century. It is expected that large ~~parts,~~ areas from the UK over France and Italy to the Balkan states will experience ~~no changes in the annual precipitation ,~~ almost no annual precipitation changes, whereas Central Europe and Northern Europe face precipitation increases. Under RCP8.5, the signal intensifies with an increase in large parts of Central Europe and Northern Europe of up to ~~about~~ approximately 25% and a decrease in Southern Europe. Meteorological droughts are projected to occur more frequently in the Mediterranean and to become less frequent in Scandinavia, with an intensification of the signal with increased warming levels (Stagge et al., 2015).

In the Paris Agreement of 2015, the Conference of the Parties of the United Nations Framework Convention on Climate Change emphasised ~~”~~"holding the increase in the global average temperature to well below 2~~-~~°C above pre-industrial levels and pursuing efforts to limit the temperature increase to 1.5~~-~~°C above pre-industrial levels~~” (UNFCCC, 2015, 2)~~ " (UNFCCC, 2015) and invited the Intergovernmental Panel on Climate Change (IPCC) to prepare a special report on the impacts of a global warming of 1.5~~-~~°C ~~global warming~~ in 2018. Notably, based on the estimated emissions over the past decades, it remains unclear if a limitation of global warming to two degrees or even three degrees can be achieved (Peters et al., 2012). Most climate impact studies in the past have focused on future time periods, e.g., changes until 2071-2100 under ~~differ- ent~~ different emission scenarios or representative concentration pathways (RCPs). Mitchell et al. (2016) argue that these studies are hardly usable for determining differences between warming levels, partly because of the large internal range of warming within the RCPs. Collins et al. (2013) reported the likely range of global warming for 2081-2100 relative to 1986-2005 ~~of~~ for the CMIP5 models with 0.3 K to 1.7 K under RCP2.6, 1.4 K to 3.1 K under RCP6.0, and 2.6 K to 4.8 K under RCP8.5.

In recent literature, several studies have investigated climate impacts on ~~the~~ low flows and hydrological droughts in Europe, focusing on differences ~~in drought characteristics~~ between historical and future time periods ~~(e.g. Forzieri et al., 2014)~~ (e.g., van Vliet et al., 2015; Wanders et al., 2015; Forzieri et al., 2014; Schneider et al., 2013). Recent assessment studies have changed their focus more towards analysing warming levels, covering the 2 degree goal (Roudier et al., 2016), comparing impacts between different levels of warming in selected river basins (Gosling et al., 2017), or focusing on runoff rather than streamflow (Donnelly et al., 2017). This study investigates projected changes in low streamflow, defined as Q90, representing daily streamflow exceeding 90% of the time, which has the potential to impact hydrological drought. Hydrological drought is associated with shortfalls on surface or subsurface water availability which can occur in low streamflow, groundwater, or reservoir levels. Changes in low flows analysed in this study can, but not always, result in drought. Exceptions are e.g., riverine based transportation, where streamflow values below a threshold level are defined as hydrological drought.

Whilst the climate and hydrological models in the available studies vary significantly as well as the formulation of low flow indices, similar patterns could be found. Decreasing river low flows are projected in ~~southern~~ Southern Europe and increasing low flows in ~~northern~~ Northern Europe. Nevertheless, there are limited studies reporting on changes in low flow conditions across Europe using an ensemble of GCM/HM simulations at high spatial resolution and for different warming levels. We fill this gap by ~~analyzing~~ analysing the changes in low flow conditions based on a large ensemble of hydrological simulations conducted at a high spatial resolution (5km) over Europe for different warming levels.

Specifically, we provide a comprehensive impact and uncertainty assessment for hydrological low flows across Europe under a global warming of 1.5, 2, and 3 K~~global warming~~. The study is based on a multi-member ensemble of high resolution simulations (5×5 km^2) from the EDgE project (http://edge.climate.copernicus.eu~~/~~, End-to-end Demonstrator for improved decision making in the water sector in Europe) which has been enlarged to 45 ensemble simulations consisting of three hydrological models (HMs) driven by five General Circulation Models (GCMs) under three RCPs. A consistent setup is achieved using identical meteorological input and land surface data to establish the three HMs. To investigate the usability of the simulation results, information on the robustness and uncertainty of projected changes as well as GCMs and HMs contributions to the overall uncertainty are discussed. The research questions aim to close a knowledge gap with respect to impacts of different levels of climate warming are as follows:

1. What is the magnitude and robustness of change in ~~hydrological droughts~~ low flows in Europe under a global warming of 1.5, 2, and 3~~K global warming~~ K?
2. Is there a significant difference in projected changes of ~~hydrological droughts~~ low flows between the three global warming levels?
3. How much do the GCMs and HMs contribute to the overall uncertainty for the particular warming levels?

2 Material and Methods

The study presented herein uses a consistent set of 45 high-resolution hydrological simulations based on five GCMs under three RCPs driving three HMs across Europe at a 5 km spatial resolution. ~~It aims~~ The aim is to provide a consistent framework using a compatible set of standardised forcings and initial conditions for the impact models to investigate ~~hydrological drought~~ low flow changes under different levels of warming. This multi-model ensemble has recently being used to analyse projected changes in river floods and high flows in Europe by Thober et al. (2017).

2.1 Climate and hydrologic models

Five CMIP5 General Circulation Models (GCMs: HadGEM2-ES, IPSL-CM5A-LR, MIROC-ESM-CHEM, GFDL-ESM2 and NorESM1-M) provided temperature and precipitation data to drive three hydrological models (HMs). Data for the time period 1950 to 2099 ~~with~~ at a daily time step ~~was~~ is available under three Representative Concentration Pathways (RCPs: 2.6, 6.0, and 8.5) from the ISI-MIP project (Warszawski et al., 2014, data available under doi:10.5880/PIK.2016.001). A trend-preserving bias-correction ~~was~~ is applied to GCM data by Hempel et al. (2013). GCM data at a horizontal resolution of 0.5° is hardly applicable to describe land surface processes on catchment scales in Europe. Therefore, this data has been ~~downscaled~~ disaggregated to 5×5 km^2 using External Drift Kriging (EDK) and the elevation as external drift ~~in~~ within the EDgE project. This interpolation technique accounts for altitude effects in temperature and precipitation and is widely applied in hydrological simulations (Zink et al., 2017). ~~Long-term trends are conserved~~ EDK adds sub-grid variability to the GCM fields, reflecting e.g., the altitude dependency of temperature. Methods such as EDK generally perform better in interpolating continuous meteorological variables compared to discontinuous variables such as precipitation. It is worth noting that the long-term trends are preserved using this interpolation technique. ~~The downscaled GCM datais used to drive~~ The variogram for EDK is estimated using the original E-OBS station data.

This meteorological data set at a spatial resolution of 5×5 km^2 is then used to force the three HMs: mHM, Noah-MP, and PCR-GLOBWB. Within the ~~EDgE-project, these~~ EDgE project, the HMs have been consistently set-up using the same land surface datasets (terrain, land cover, soil maps and geological information). Furthermore, a consistent external ~~routing~~ river flow routing scheme has been applied ~~for~~ to outputs of all HMs based on the multiscale Routing Model that has been developed originally for mHM (Samaniego et al., 2010). Ultimately, the differences in the hydrological simulations result from different process representations and parameterisations of the surface and subsurface in the HMs.

The HMs used in this study are grid-based distributed models grounded on numerical approximations of dominant hydrologic processes. The mesoscale Hydrological Model ~~(mHM)~~ (mHM, Samaniego et al., 2017b) has originally been developed in Central Europe and it uses the multiscale parameterisation technique, MPR (Samaniego et al., 2010; Kumar et al., 2013); that allows the model applicability at different spatial resolutions (1×1 km^2 to 50×50 km^2) and multiple locations without much of a calibration effort. The Noah-MP model was originally developed as land surface component of the 5th generation mesoscale model MM5 to enable climate predictions with physically based ensembles and represents both the terrestrial water and energy cycle (Niu et al., 2011). The ~~PC raster~~ PCRaster global water balance model (PCR-GLOBWB) was developed to represent the terrestrial water cycle with a special focus on groundwater and modelling water resources under water stress (Van Beek and Bierkens, 2008; Wanders and Wada, 2015). ~~All HMs have been calibrated in 9 diverse catchments located in Norway, Spainand UK to include~~

The three HMs used in this study are calibrated in nine near-natural European focus basins located in Spain, Norway and UK, which are selected based on the consultation with the user groups within the EDgE project. Besides these, we also include three more central EU catchments (located in France and Germany) to represent diversity in hydro-climatic regimes. All HMs parameters are calibrated such that the model simulations represent a range of ~~hydro-climatic regimes. The HMs were~~ hydrologic regimes, rather than tailored to any specific characteristics. This is done in a consistent manner so that the model simulations can be used for a range of indicators (including high, low, and average flows) within the EDgE project, resulting in slightly lower performances for low flows. We note that HMs could be calibrated to specific parts of the flow duration curve (FDC), however, this is not done in this study to avoid too specific tuning of the model simulations to those unique conditions and thereby losing valuable information on the entire FDC. In the current simulations human water management was not taken into account, since some models lack the ability to include these processes. Human water management can however have a significant impact on the low flow conditions, due to abstraction of additional water in drought condition or changes in reservoir management. As a result constraining the model to any specific low flow characteristic can result in a biased simulation. Also due to the similar reason we may expect a relatively lower model skill in matching observed low flow characteristics. The HMs are calibrated using observation-based E-OBS data (V12.0, Haylock et al., 2008) and automatic calibration ~~of~~ schemes are employed for mHM (Rakovec et al., 2016) and PCR-GLOBWB. Noah-MP has been calibrated manually adjusting the parameter for evaporation surface resistance based on the analysis by Cuntz et al. (2016).~~The assessment of climate change impacts is independent of whether impact models have been calibrated or not (Gosling et al., 2017). Nevertheless, temperature~~

Temperature and precipitation data from GCMs with coarse resolution have different statistical properties than interpolated observational datasets. To investigate if the observation-based calibration of the HMs is applicable to the ~~downscaled~~ disaggregated GCM data, model outputs are evaluated against ~~165~~ 357 gauging stations using the GCM forcing during the historic period ~~1971-2000~~ 1966-1995 (Fig. 1). The ~~analysis focused~~ stations and time period are selected to ensure the largest possible, complete dataset over 30 years. Their median basin area is 1680 km^2. The analysis focuses on matching the median of the 30 years annual percentile for low flows (Q90~~, see below)and average river discharge (Q50~~). The indicator for ~~hydrological drought (Q90)~~ low flows is used herein for the impact assessment studies as detailed ~~below~~ described in section 2.3.

The evaluation results show ~~a good agreement for both percentiles and all HMs driven by the 5 GCMs. The scatter around the 1:1 line (in Figure 1) is more related to the different HMs than to the different GCMs. For example, mHM shows the~~ overall an overestimation of observed Q90 by all HMs and GCMs. (Figure 1, lower left). This overestimation in the ensemble average is mainly the result of the overestimation by the HMs PCR-GLOBWB and Noah-MP simulations, while the mHM runs show only a slight overestimation and result in closest correspondence to the observed values~~for both low flow (Q90) and median flow(Q50). Noah-MP has an overall larger scatter than mHM. For Q90, this scatter is distributed around the 1:1 line. For Q50, a slight overestimation is observed (~~. Nevertheless, it cannot be concluded that mHM performs best due to the neglection of human activities in many basins (abstraction as well as e.g., ensuring minimum ecological flow). Well-calibrated HMs do not necessarily mean that future simulated discharge under a changed climate can be reproduced satisfactorily (Vaze et al., 2010). Furthermore, the selection of HMs may have a larger effect than the calibration of parameters in hydrological climate impact studies (Mendoza et al., 2015). The spatial pattern of the relative bias for the multi-model ensemble average is shown in Fig. 1 ~~d). PCR-GLOBWB slightly overestimates both indices - the scatter is , however, comparable to the one observed for mHM.~~ (lower right). It is important to assert that this spatial pattern differs significantly between the HMs while the climate change signal for low flow projections in this study (see section 3) is remarkably similar across all three HMs.

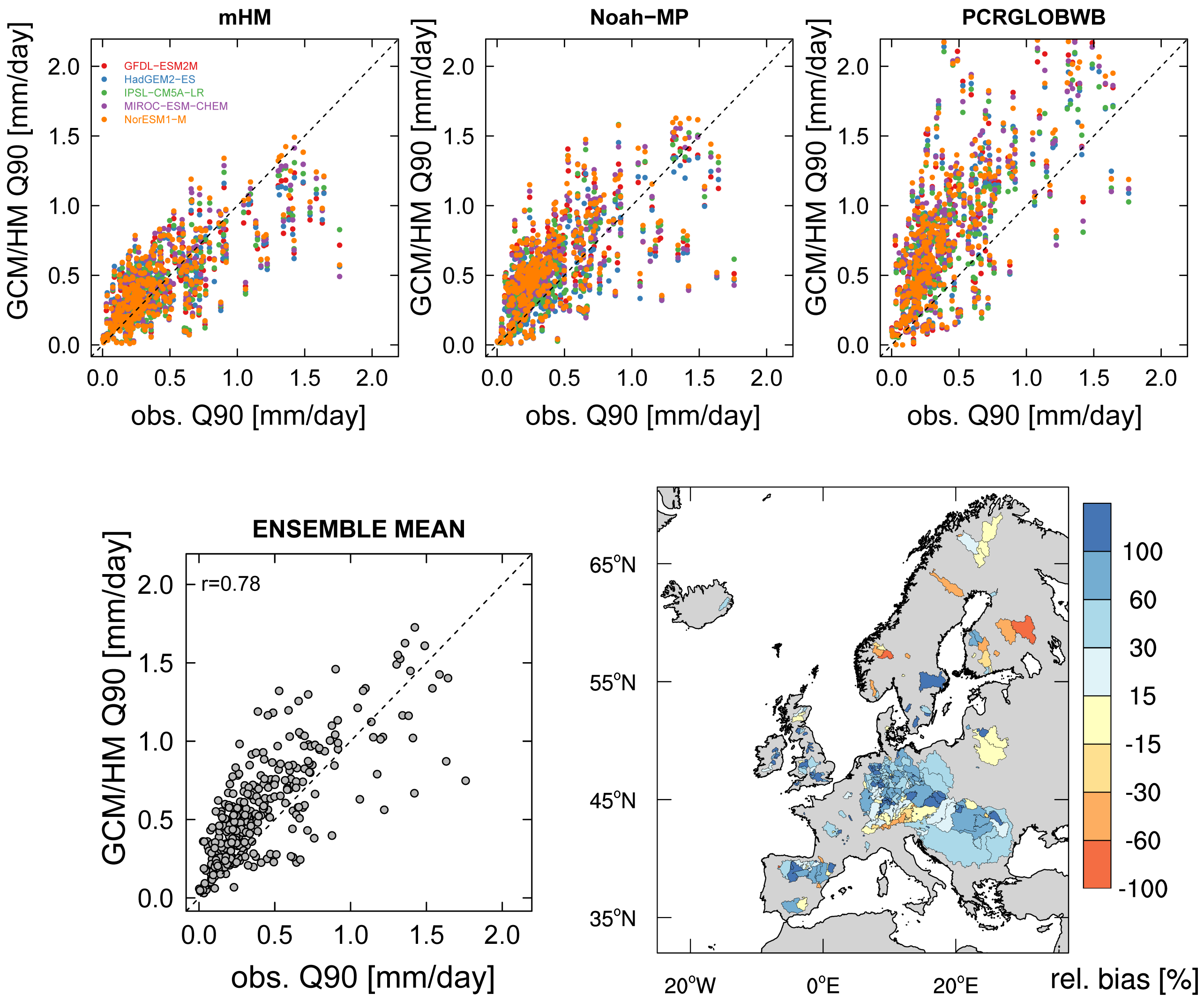

Figure 1. Scatter plot between observed low flow and ~~(~~GCM-HM ~~)~~ simulated low flow (Q90) ~~and median flow (Q50)~~ over ~~165~~ 357 gauges across Europe. Simulated values correspond to the median of the annual estimates calculated for the historical time-period ~~1971-2000~~1966-1995. The colours of the dots denote the five GCMs used to drive the hydrologic models mHM (left column), Noah-MP (middle column) and PCR-GLOBWB (right column). The location of the ~~gauges~~ basins and the spatial pattern of the relative bias is ~~displayed in Fig. 2 (~~shown on the lower right~~side)~~.

2.2 Determination of 1.5, 2, and ~~3 K~~ 3-K time periods

The five CMIP5 GCMs used in this study have different sensitivities to climate forcing. The development of annual global temperature varies significantly over time between the models and RCPs. Therefore, the time period with a mean global warm-

Table 1. Determination of 1.5, 2, and 3 K time periods for different GCM/RCP combinations. A time sampling approach was used comparing 30-year running means to the period 1971-2000 with an assumed warming of 0.46 K to pre-industrial conditions.

Warming level	RCP	GFDL-ESM2M	HadGEM2-ES	IPSL-CM5A-LR	MIROC-ESM-CHEM	NorESM1-M
1.5 K	2.6	-	2007-2036	2008-2037	2006-2035	2047-2076
	6.0	2040-2069	2011-2040	2009-2038	2012-2041	2031-2060
	8.5	2021-2050	2004-2033	2006-2035	2006-2035	2016-2045
2 K	2.6	-	2029-2058	2060-2089	2023-2052	-
	6.0	2060-2089	2026-2055	2028-2057	2028-2057	2054-2083
	8.5	2038-2067	2016-2045	2018-2047	2017-2046	2031-2060
3 K	2.6	-	-	-	-	-
	6.0	-	2056-2085	2066-2095	2055-2084	-
	8.5	2067-2096	2035-2064	2038-2067	2037-2066	2057-2086

ing of 1.5, 2, and 3 K with respect to pre-industrial condition ~~,~~ also varies between the GCM simulations. ~~In this study~~Here, a time sampling method is used to determine the time-period for different levels of global warming (James et al., 2017). This approach has been used to investigate climate impacts over Europe for a global warming of 2 ~~K global warming~~ K (Giannakopoulos et al., 2009; Vautard et al., 2014) and for global differential impacts between a warming of 1.5 K and 2~~K warming (Schleussner et al., 2016). In this study,~~ K (Schleussner et al., 2016). 30-year running mean global temperatures are compared to those of the 1971-2000 period in the GCM simulations. The latter period corresponds to a global warming of 0.46 K ~~with~~ (average value from three estimations with a spread between 0.437 K and 0.477 K) with respect to pre-industrial condition (Vautard et al., 2014). The first 30-year period with a global warming crossing one of the three warming levels (1.5, 2, 3 K) is then determined for each of the 15 GCM/RCP combinations. The identified 30-year time-period for the corresponding GCM/RCP combination is shown in Table 1. It is ~~worth-noting~~ worth noting that for some of the combinations, we could not identify any 30-year period for ~~the selected warming levels~~a selected warming level. For example, none of the GCM simulations crossed the ~~3.0~~ 3 K warming level under the RCP2.6 over the entire simulation period up to 2099.

Available methods for identifying regional climate responses to global warming targets face advantages and disadvantages (James et al., 2017). Limitations in the time sampling method occur in the direct comparison between different warming levels because the number of ensemble members varies. Available simulations reduce from 14 under 1.5 K warming over 13 under 2 K to 8 simulations under a global warming of 3 K~~global warming~~. Furthermore, the annual temperature within future 30-year periods may be pathway dependent, e.g., a rapid or slower warming. This may influence the results in climate impact simulations. Nevertheless, the time sampling method ~~poses the advantage of~~ is advantageous creating a large ensemble of simulations, which is essential to determine differences between warming levels (Mitchell et al., 2016).

2.3 Low flow indicator used, uncertainty metrics, and spatial aggregation of results

The impact of climate change is quantified for low flows. Commonly, the ~~70th-90th~~ 70th to 90th percentile of exceedance is used to define hydrological droughts for rivers with perennial type streamflow (Fleig et al., 2006). Within the framework of the EDgE project, the co-production with stakeholders from the water sector in Norway, Spain and the UK resulted in Q90 (daily flows exceeded 90% of the time) as ~~hydrological drought~~ low flow index. The Q90 is estimated for each calendar year over a given 30-year period, and the median of Q90 is subsequently calculated from the respective 30 samples as a final indicator. We ~~use the period~~ recognise that the use of a calendar year may influence our results in snow-influenced catchments where the low flow period may span over two consecutive years. To assess possible consequences, we compared the annual results against simulations for the winter half year and found only minor changes in overall results, especially in snow dominated regions. Further seasonal assessment is not performed in this study. We use the period 1971-2000 as a reference for the estimation of climate impacts and the relative changes in Q90 ~~was~~ is estimated with respect to this reference period for different warming levels. ~~Notably the 1971-2000 is the last 30-year period considered in the historical simulations of GCMs in the CMIP5 project.~~

~~European macro-regions (left side) used in the IPCC AR5 (Kovats et al., 2014) based on an environmental stratification after Metzger et al. (2005) (Source: own graphics based on GIS data provided by Marc J. Metzger, University of Edinburgh. The data is remapped to the 5 km grid used in this study). The location of the 162 gauges used for the validation of the GCM/HM simulations is shown on the right side.~~ ~~To~~ The non-parametric Wilcoxon rank-sum test is applied to account for the robustness of the results~~, the non-parametric Wilcoxon rank-sum test is applied~~. The null hypothesis of equal means between the climate periods per GCM-HM simulation is tested at 5% significance, which has been applied in Gosling et al. (2017) among others. Based on the ensemble of Wilcoxon rank-sum tests, the robustness is estimated following the IPCC AR4 procedure presented in Solomon et al. (2007). Robustness is computed as the percentage of projections showing a significant change. Important thresholds are ~~<~~less than 33% for unlikely and ~~>~~greater than 66% for likely changes, representing the percentage of ~~simulations in the ensemble~~ ensemble simulations showing a significant change. Significance here does not account for the sign or magnitude of change.

The signal to noise ratio (SNR) is commonly used to quantify the uncertainty in hydrological extremes studies (Prudhomme et al., 2014; Hall et al., 2014; Giuntoli et al., 2015)~~, here~~. Here, the SNR is computed as the median divided by the inter-quantile range (i.e., the difference between the 25th and 75th percentile). It has been acknowledged in recent literature that both GCMs and HMs contribute to the uncertainty in projected changes (Gosling et al., 2017; Donnelly et al., 2017; Hattermann et al., 2017). In this study, the sequential sampling approach of Samaniego et al. (2017a), following Schewe et al. (2014), is applied. In this approach, the uncertainty due to GCM is estimated by first fixing a HM and then calculate the range ~~of Q90~~ (max-min) of Q90 changes corresponding to five GCMs outputs. Repeat the previous step for all other remaining HMs. Finally, estimate the average of ensemble ranges that would then represent the uncertainty due to GCMs. Likewise, the same steps could be repeated by fixing the GCM and calculating the range statistics over the HMs to represent the uncertainty component due to HMs. We use the bootstrap technique to account for different sample size of GCM and HMs; and perform the sequential uncertainty assessment with three GCMs and HMs outputs over the 1000 realizations.

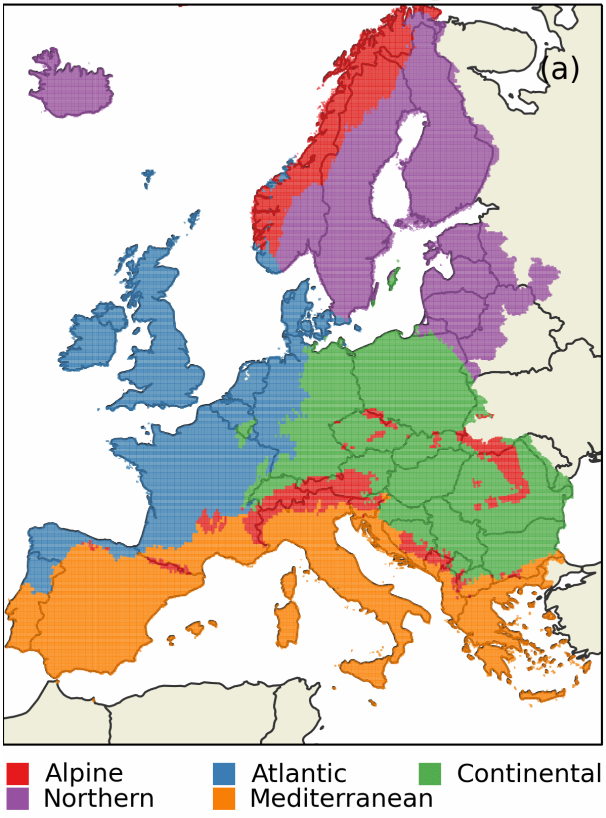

Figure 2. European macro-regions used in the IPCC AR5 (Kovats et al., 2014) based on an environmental stratification after Metzger et al. (2005) (Source: own graphics based on GIS data provided by Marc J. Metzger, University of Edinburgh. The data is remapped to the 5 km grid used in this study).

To account for regional differences in climate impacts, the results ~~in this study~~ of our analyses are displayed over Europe and additionally aggregated for five different regions (Fig. 2~~, left side~~). These macro-scale regions have been used in the latest IPCC WGII report ~~in the Europe chapter~~ for Europe (Kovats et al., 2014) and were originally identified based on the environmental stratification presented in Metzger et al. (2005), using a principal component analysis accounting for 20 different environmental variables. Furthermore, the low flow impact assessment ~~studies~~ carried out here is limited to river basins with upstream ~~area greater than 10000~~areas greater than 1000 km^2. Smaller (and headwater) basins are not considered here as to limit the delineation errors of river network in the runoff routing scheme ~~.~~ (see e.g., Fig. 3 for the resulting river network).

Table 2. Relative changes [%] in streamflow Q90 between the ~~1980s~~ past (1971-2000) and different warming levels averaged over IPCC AR4 Europe regions shown in Fig. 2.

Warming level	Absolute warming	Alpine	Atlantic	Continental	Northern	Mediterranean
1.5 K	1.04 K	22.2	-7.3	-4.1	8.4	-12.0
2 K	1.54 K	29.6	-10.0	-4.5	15.9	-16.3
3 K	2.54 K	44.8	-21.6	-19.1	24.1	-35.1

3 Results and Discussion

3.1 ~~Hydrological drought~~ Changes in low flows under different levels of warming compared to 1971-2000

The change signal in ~~hydrological droughts~~ low flows gets stronger with increased levels of warming in most parts of Europe (Fig. 3, left row). An amplification in decreasing low flows can be identified in the Iberian Peninsula, the south-western part of France, and southeast Europe including Greece and the Balkan states. On the contrary, large parts of the Alps and Scandinavia face an intensification of increasing low flow signal with higher levels of warming. The region from Germany over Poland to the Baltic countries shows generally very small changes, and the sign of change in low flows alters with increased warming. Under a global warming of 1.5 K~~global warming~~, the mean change in streamflow Q90 over Europe is ~~about~~ approximately zero (Fig. 3, upper left), but with large spatial differences between the IPCC AR5 Europe regions and with different directions of change. The regional low flow statistics are based on the average of all the grid cells per region.

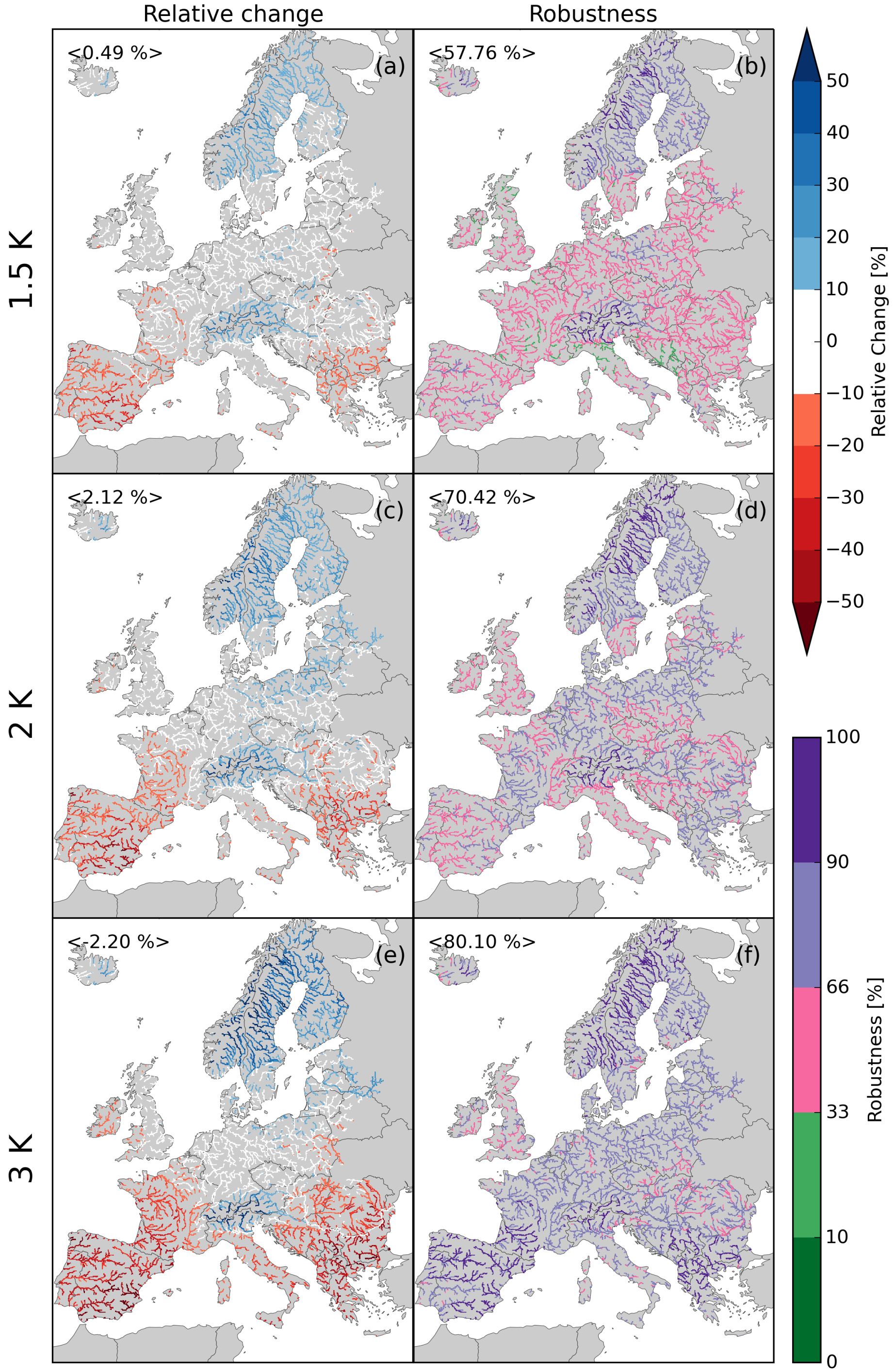

Figure 3. Change in multi-model ensemble mean ~~hydrological drought~~ low flow [%] under different warming levels compared to the 1971-2000 baseline (left) and robustness (right). The latter is expressed by the percentage of simulations based on a Wilcoxon rank sum test with 5% significance level. An agreement of ~~>~~more than 66% in the ensemble is classified as ”likely” change. The values given in the upper left of the subplots is the continental average along the river network for all grid cells with a contributing area greater than 1000 km^2.

~~About~~ Approximately half of the rivers in Europe show decreases in low flows under 1.5 K warming, with an hotspot in the Iberian Peninsula region and the strongest decrease in the Mediterranean [-12% over the whole area] and the Atlantic region [-7%] (Tab. 2). ~~Contrarily~~On the contrary, increases in low ~~flows~~ flow are expected in the Alpine [+22%] and Northern ~~area~~ areas [+8%]. This occurs mainly due to ~~change~~ changes in snow accumulation and melt, and ~~, consequently~~ , consequently results in higher winter low flows. The Continental area shows overall the smallest changes with both positive and negative values, but less than 10% even under a global warming of 2 K~~global warming~~.

More regions in Europe show significant changes in low flow with an increased level of warming (Fig. 3, left row). Robustness is expressed as the percentage of ~~models~~ simulations passing the Wilcoxon rank sum test at 5%. Under a warming level of 1.5 K~~warming, overall around~~ , approximately 57% of the ensemble simulations show significant changes. Highest values are found in snow-dominated regions (e.g., Alpine and Northern region). Under a warming level of 2 K~~warming~~, the percentage of ensemble simulations with significant changes increases to ~~about~~ approximately 70%, being distributed equally over Europe~~.~~ ~~Under~~ , and this number increases to 80% for a warming level of 3 K~~global warming~~. Under a global warming of 3 K, the agreement among the ensemble simulations increases to overall 80%. The strongest regional change is found in the Mediterranean, with likely changes across 31% of the river basins under a warming level of 1.5 K~~warming~~, 64% under 2 K and 90% under 3 K~~warming~~, respectively. ~~It can generally be stated that the~~ The significance is highest in regions with strong (positive and negative) change signals. Nevertheless, there are exceptions, e.g., under 2 K warming the signal for the Mediterranean might be stronger, but it is less robust than that for the Atlantic.

The results presented here confirm those found in earlier studies for low flow and hydrological drought projections across Europe. Forzieri et al. (2014), for example, gave an overview on projected changes in average 7-day minimum flows until the end of the century under the SRES A1B scenario. A single HM was selected for the analysis in that study which was then driven by 12 regional climate model (RCM) precipitation and temperature dataset. The analysis showed that streamflow droughts become more severe and persistent in ~~southern~~ Southern Europe, while droughts decrease in northern and northeastern parts of Europe. Wanders et al. (2015) found similar patterns over Europe using 5 GCMs and a single HM, with a clear influence of ~~decrease~~ decreasing snow accumulation in ~~northern~~ Northern Europe and an increase ~~drought impact~~ in drought impacts in the Mediterranean. Recently, Gosling et al. (2017) investigated changes in hydrologic droughts under a global warming of 1, 2 and 3 K ~~global warming~~ over large river catchments (~~>~~greater than 50000 km^2) including ~~the~~ two European basins - the Central European Rhine and Mediterranean ~~Targus~~ Tagus River. They used Q95 as a low flow indicator, based on the same 5 GCMs applied in our study with an ensemble of global as well as catchment hydrological models. Nevertheless, the results from both studies are comparable under a global warming of 2 K and 3~~K global warming, with decreasing~~ K, with projected decrease in low flows in the Rhine and ~~Targus~~ Tagus River. Low flow (Q90) in this study under a warming level of 2 K ~~warming~~ is almost unchanged in the Rhine, ~~decreasing~~ and up to -11% under 3 K~~warming~~. The more pronounced low flow decrease is found in the ~~Targus~~ Tagus River showing -16% under 2 K and -33% under a global warming of 3 K~~global warming~~. The GCMs used in van Vliet et al. (2015) are also identical to those used in this study. However, the HMs E-HYPE (Donnelly et al., 2016) and VIC (Cherkauer et al., 2003) were used to simulate the changes in Q90 for RCP2.6 and RCP8.5 for the 2050s and 2080s. Overall, the spatial pattern of changes in flow indicator fits to our results quite well, likewise the amplification of the signal over time until the end of the century was ~~found~~also found in both studies. The strongest reductions in ~~hydrological droughts were~~ low flows are exhibited in Southern Europe and related to decreasing annual precipitation. The spatial pattern under a global warming of 2 ~~K global warming compare~~ K compares well with those reported by Roudier et al. (2016) for low flows with a 10-year return ~~periods~~period. Notably, the underlying model ensemble consists of 11 bias-corrected RCMs and two hydrologic models, which are different from those used in this study. They found a 15% reduction in low flows for the Mediterranean, which is very similar to the 16% reduction found in this study. Although the results on the climate induced change in low flows presented herein are generally comparable to other studies, we provide new spatially explicit information on low flows under different levels of warming over Europe.

~~Notably, our~~ Our study shows contrasting results for the Mediterranean region compared to Donnelly et al. (2017) under different levels of warming. At a global warming of 3 K~~global warming~~, large decreases up to -35% and high robustness (very likely) are observed here, whereas no projected changes in absolute grid-specific runoff values with little robustness was reported by Donnelly et al. (2017). These differences can be explained through methodological choices on low flow indices used between the two studies. The relative changes in the routed river low flow quantified here is more informative for water resources assessments compared to the absolute changes of grid-specific runoff. This holds especially true in ~~these dry~~ drier regions, which are characterized by very small ~~Q10~~ Q90 runoff values. From a practitioner point of view, our study highlights the need for adaptation to climate induced ~~hydrological droughts~~ low flows in these regions, which would not be concluded based on the metrics reported in Donnelly et al. (2017).

It is observed that the changes in river low flows can be explained to a large ~~extend~~ extent by the median change in annual precipitation over all levels of global warming (Fig. 4). To investigate the influence of precipitation on ~~hydrological droughts~~low flows, we compare the relative change of Q90 discharge to the changes in the annual total precipitation over the 30-years for different levels of warming. The Mediterranean region shows the strongest decrease in precipitation and low flows among all warming levels. The correlation coefficient between changes in annual precipitation and Q90 increases from 0.45 under 1 K to 0.62 under 3 K level of warming. Notably, the increased spread ~~around the 1:1 line under the~~ in the median changes of annual total precipitation and simulated river low flows under the warming level of 3~~K warming level also~~ K contribute to higher ~~r^2 values~~ correlation compared to other warming levels. Furthermore, we observe a relatively stronger correspondence between changes in annual total precipitation and low flow indicator in ~~the~~ river basins characterized by projected decrease in low flows. The r^2 value rises from 0.61 to 0.77 with an increase in a global warming level from 1.5 to 3 K; compared to an increase of 0.45 to 0.65 for the same warming levels in river basins showing projected increase in low flows. Overall, the Continental and Atlantic ~~region~~ regions show the smallest changes in precipitation and low flows. In the Northern region, the projected increases in changes of both variables are highest. In this region, the relationship between precipitation and low flows is the weakest ~~, exemplified in~~ as exemplified by the low r^2 values for the positive precipitation changes. This can be explained due to the increasing influence of snow processes, accumulation as well as snow melt. This holds also true for catchments ~~>~~greater than 1000 km^2 in Alpine regions (not displayed here).

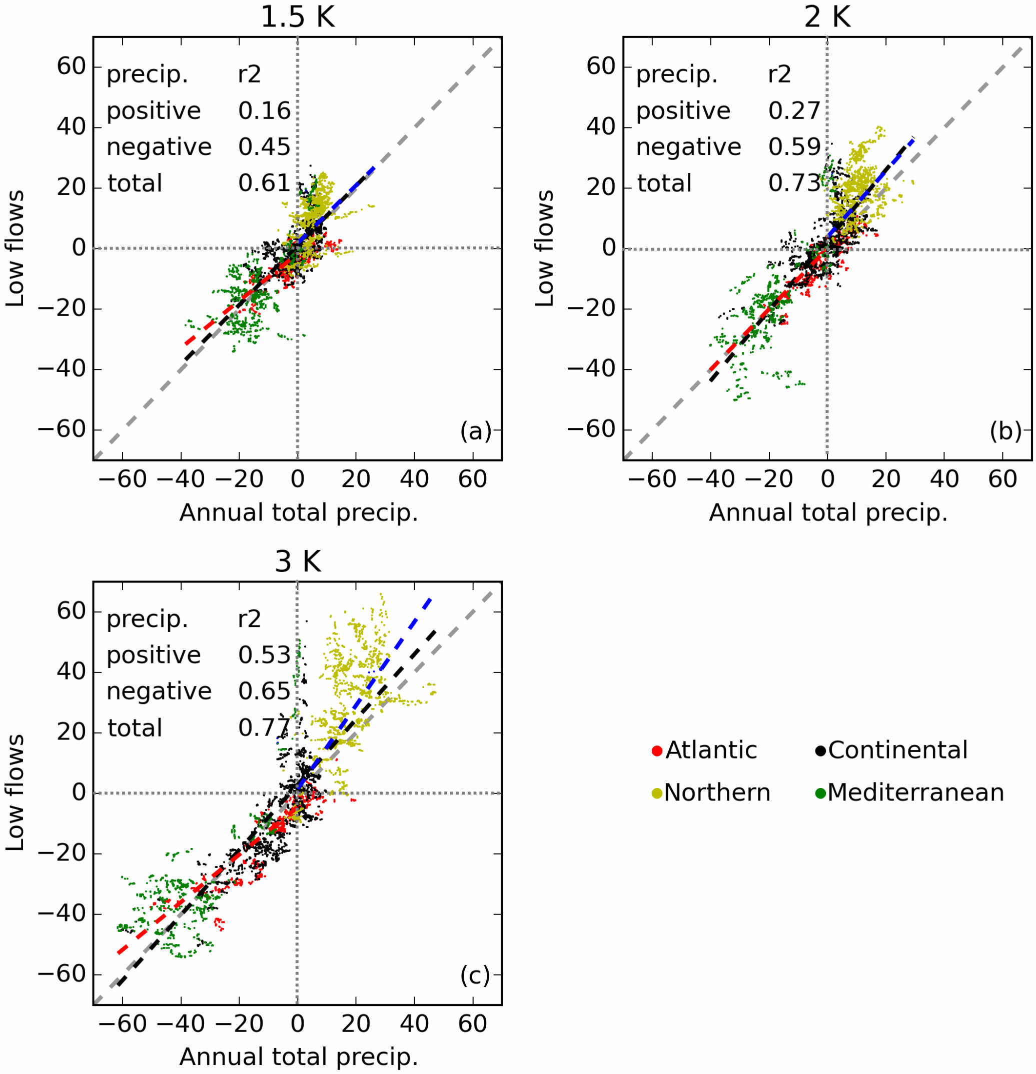

Figure 4. Relationship between the median changes in the annual total precipitation and simulated river low flows (Q90) under a global warming of 1.5 K (a), 2 K (b) and 3 K (c)~~global warming~~. ~~Only~~ Anomalous to other results shown in this study, only river grid cells from basins ~~>~~ greater than 10000 km^2 are shown for clarity in the figure. Results are similar to those including river grid cells with contributing areas greater than 1000 km^2. Linear regression lines are shown for positive values (blue dashed), negative values (red dashed) and all data points (black dashed). The Alpine with overall smaller catchment sizes is not included, but shows a similar ~~behavior~~ behaviour to the basins in the Northern region. All changes are expressed as multi-model ensemble mean changes (GCM/HM combinations for low flows and GCMs for annual precipitation).

Table 3. Relative changes averaged over regions [%] in multi-model ensemble mean low flow indicator (Q90) between different levels of global change.

Warming level	Absolute warming	Alpine	Atlantic	Continental	Northern	Mediterranean
1.5 K → 2 K	0.5 K	8.6	-1.1	-0.3	10.7	-6.6
2 K → 3 K	1.0 K	17.0	-9.0	-12.3	13.5	-16.0
1.5 K → 3 K	1.5 K	23.9	-12.9	-12.2	22.6	-24.0

Under a warming level of 3 K~~of warming~~, we identified a larger spread between total annual precipitation and low flows. In the Northern area, this can be explained due to higher temperatures which could then lead to less snow accumulation and increased winter low flows. In contrast, higher temperatures combined with lower than average annual precipitation in the Mediterranean ~~tend to~~ result in higher evapotranspiration and decreased low flows. ~~Other studies have also found~~ Our results agree with other studies reporting about the general relationship between precipitation and low flow changes ~~(e.g. Forzieri et al., 2014; van Vliet et al., 2015; Gosling et al., 2017)~~(e.g., Forzieri et al., 2014; van Vliet et al., 2015; Gosling et al., 2017) , even though relating precipitation and low flows in different ways. ~~Although the results on the climate induced change in low flows presented herein are generally comparable to other studies, we provide new spatially explicit information on hydrological droughts under different levels of warming across Europe.~~ In the following section, the differences between ~~political~~ policy relevant levels of warming are examined.

3.2 Differences in low flows between ~~hydrological droughts under~~ different future levels of warming

One of the objectives of this study is to analyse differences in the change signal and the sensitivity of the low flow changes to different levels of global warming. This provides additional information compared to the results presented above. Both of these results, in combination, are important for the discussion on mitigation targets and for adaptation planning in accordance with the Paris agreement (UNFCCC, 2015). With increased levels of global warming from 1.5 to 2 K, 2 to 3 K and 1.5 to 3 K, an amplification of the change signal in low flow is expected over a large part of Europe (Fig. 5, panels a, c, and e). This holds especially true in regions with relatively big positive and negative changes in low flows. The overall robustness of the low flow changes in Europe increases with increasing temperature differences between the global warming levels (Fig. 5, panels b, d, and f).

The changes in streamflow Q90 between 1.5 K and 2 K warming are generally small with few rivers exhibiting changes larger than 10% in magnitude. The pattern is similar to the one shown in Fig. 3, ~~highlighting~~ which highlights that the sign of change is conserved in areas with relatively large changes (more than ±10%), even under the ~~relatively~~ small warming of only 0.5 K. ~~Nevertheless, these results~~These results, however, are not robust. None of the rivers show likely changes, meaning that less than 66% of the ensemble simulations are significant at the river grid cell level. Moreover, most parts of Europe show changes marked as unlikely with a total agreement of only 15% over Europe and all simulations. The regional changes in low flows between the two warming levels are also small (see Tab. 3). The Atlantic and Continental area show ~~an almost~~ almost an unchanged situation. The Northern region exhibits the largest increase with 11%, and the Mediterranean faces -7% decrease in low flows averaged over the considered stratified region.

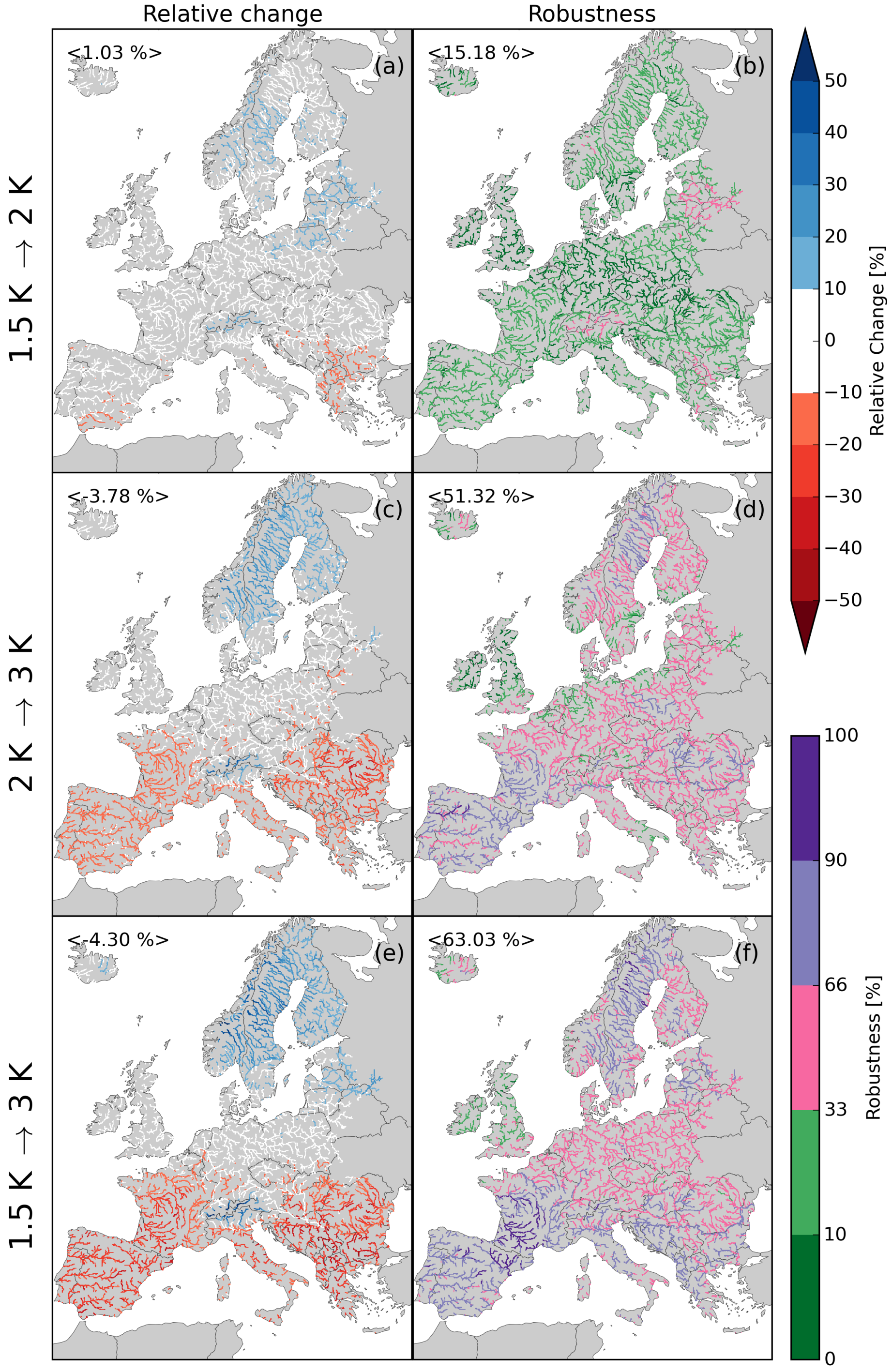

Figure 5. Relative change [%] in multi-model ensemble median Q90 between different levels of warming (left) and robustness of the signal between those (right). The latter is expressed by the percentage of simulations based on a Wilcoxon rank sum test with 5% significance level. The values given in the upper left of the subplots is the continental average along the river network for all grid cells with a contributing area greater than 1000 km^2.

The robustness ~~given in (~~results presented in Fig. 3 ~~,~~ (panels b, d, f) alone ~~does~~ do not allow for determining warming level thresholds of change in low flow indicator. Therefore, we included the robustness of the change between the warming levels in this section. Combining the information in Fig. 3 (b,d) with Fig. 5 (b), we see robust changes between the past time period and a 2 K warmer world. The information of non-significant differences between 1.5 K and 2 K warming allows for the conclusion that the majority of change already happens before reaching a warming level of 1.5 K~~warming level~~. Limiting climate change to a warming level of 1.5 K ~~warming level~~ in comparison to 2 K has only a limited effect on low flows. ~~Nevertheless,~~ These results point out that an even lower mitigation goal would be needed for regions where substantial negative impacts occur. Low flow changes between 2 K and 3 K warming (Fig. 5, panels c and d) are more pronounced with large parts of the Central Alps and Scandinavia showing an increase of more than 10% in low flows. On the contrary, most regions on the Iberian Peninsula, France, Italy, the Balkan states and Greece face a decrease of more than 10% low flow. The strongest increase is projected for the Alpine region (+17%), ~~while~~ and the strongest decrease for the Mediterranean (~~-24~~-16%). Overall, half of the simulations show robust changes over Europe with large regional differences. Likely changes are found in the southwest of Europe, northern Norway and the Balkan states. It is worth ~~emphasizing~~ emphasising that the differences between a global warming of 2 K and 3 K ~~global warming~~ in low flows are substantial. These changes are on top of those projected between 1971-2000 and a 2 K warming, where already 70% of the simulations show significant changes (Fig. 5 d). As a result, the increase in low flows in the Alpine and Northern regions could, e.g., in combination with increased future annual precipitation in the GCMs (see Fig. 4), lead to a higher hydropower potential~~, while~~. On the contrary, a further decrease of available water ~~in~~ (in low flows as well as annual precipitation) in the Mediterranean may pose additional water stress in that area. ~~This highlights that~~ Although human influences such as reservoir management or human water demand are not considered in this study, different regional adaptation options should be considered depending on whether the world warms 2 K or 3 K. This holds also true for the more pronounced warming between 1.5 and 3 K, (Fig. 5, panels ~~e-f~~e and f) where the regional changes in low flows as well as the robustness amplify compared to 2 and 3 K warming. These results also highlight the non-linear sensitivity of changes in low flows to different levels of global warming. For example with long-lasting infrastructure or long planning horizons, adaptation strategies should be put in place now, without waiting for the 3 K level to be reached or not.

Overall, the robustness in the change signal rises with increased temperature differences between the warming levels. Based on the results of the multi-model assessment conducted here, ~~distinguished climate change impacts on~~ significant differences in low flows between the ~~political~~ policy relevant 1.5 K and 2 K warming could not be identified. Little differences between these two warming levels have been observed because of the high variability among the ~~ensemble members~~GCM/HM simulations. The multi-model variability is further analysed in detail in the following section.

3.3 ~~Methodological uncertainties and~~ Uncertainty contributions from GCMs and HMs

To provide a comprehensive picture over uncertainties, the signal to noise ratio (SNR) is investigated additionally to the robustness of the change signal based on the Wilcoxon rank sum test presented in sections 3.1 and 3.2. Furthermore, the uncertainty contribution of the GCMs and HMs for different levels of warming is also investigated.

Table 4. Dimensionless uncertainty contribution of GCMs and HMs averaged over the stratified European regions described in section 2.3.

Warming level	European regions				
	Alpine	Atlantic	Continental	Northern	Mediterranean
GCM uncertainty					
1.5 K	27.3	25.8	35.2	31.2	31.4
2 K	32.1	31.6	44.7	43.7	38.5
3 K	52.1	32.9	48.3	63.4	31.3
HM uncertainty					
1.5 K	26.7	19.8	21.1	31.9	25.0
2 K	33.4	24.0	25.3	39.6	30.2
3 K	55.6	31.4	31.1	55.1	34.7

Under a global warming level of 1.5 K and 2~~K warming~~ K, large parts of Europe exhibit substantial uncertainty, expressed as the SNR (Fig. 6, panels a and b). It is ~~expressed~~ estimated as the ensemble median divided by the ensemble inter-quartile range (Giuntoli et al., 2015). Using the inter-quartile range partly accounts for outliers in the ensemble simulations. The SNR is small for changes in low flows under a warming level 1.5 K ~~warming~~ and increases with further warming. These results are similar to the increasing changes and robustness of the simulations with the increased warming level ~~shown previously~~ as also previously discussed in Figure 3. Under a global warming level of 1.5 K~~warming~~, the spatial patterns of SNR and robustness coincide between the different methods (Fig. 3 b compared to Fig. 6 a). Nevertheless, a direct comparison of the uncertainty patterns under higher levels of warming between SNR and robustness ~~lead~~ leads to different conclusions in some regions. As an example, large parts of Germany show a robust change under 2 K and 3 K warming (Fig. 3, panels d and f) whereas the SNR is smaller than 0.8 over the same regions indicating a high uncertainty. This occurs because the Wilcoxon rank sum test is performed for each ensemble member separately, and the result is independent of the sign of change and absolute value. ~~Contrarily~~On the contrary, the SNR shows the uncertainty among the ensemble members and depends on the variability between those ensemble simulations. Additionally, thresholds selected for rejecting results or marking them as uncertain have greater influence on the presented results in both methods. This highlights that the uncertainty information conveyed strongly depends on the metrics selected to represent them. In other words, the robustness indicates that most ensemble members project significant changes in Germany, but there is disagreement among them indicated by a low signal to noise ratio. ~~Uncertainties should be considered in adaptation planning, e.g. in deciding to use climate impact simulations to determine regional vulnerability quantitatively or qualitatively.~~

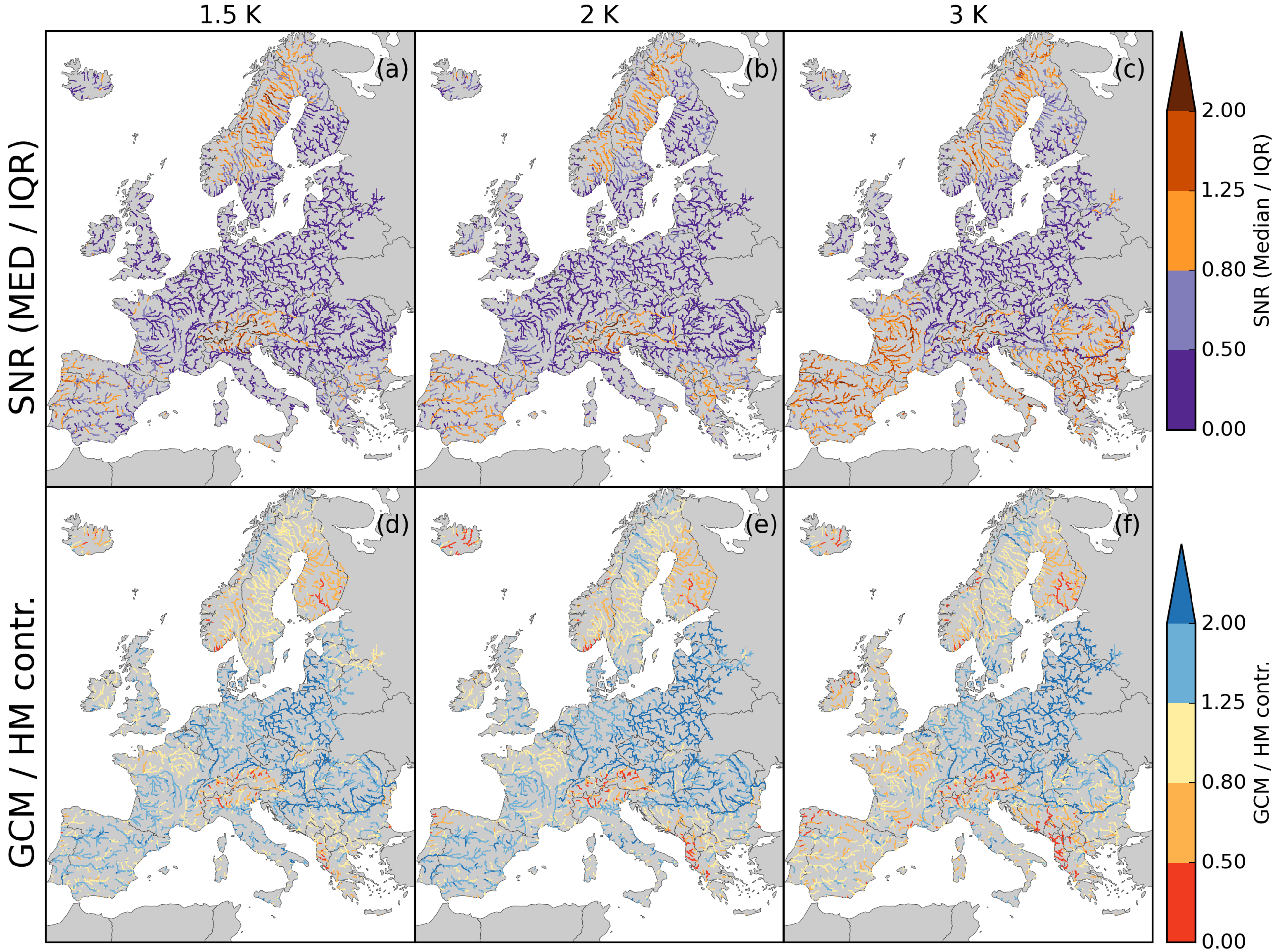

Figure 6. The upper row (a-c) shows the signal-to-noise ratio (ensemble median divided by the inter-quartile range) for the change in low flows (Q90) between the 1980s and 1.5 K (a), 2 K (b) and 3 K (c) warming. The relative uncertainty contribution of GCMs and HMs is shown in the lower row (d-f) for the three warming levels. Low values of GCM/HM indicate large HM uncertainty, values larger than one indicate a ~~domination~~ higher contribution of the ~~GCM contribution~~ GCMs to the total uncertainty.

The SNR results presented here are in line with the findings for the Rhine and ~~Targus~~ Tagus River in Gosling et al. (2017). A comparison to other studies like Forzieri et al. (2014) or Roudier et al. (2016) is in this case difficult because those studies used different metrics to describe uncertainty and, consequently, the patterns in those studies vary significantly from the patterns shown here.

~~Total uncertainties in low flow projections is separated into~~ GCM and HM contributions ~~using~~ to total uncertainty separated with the sequential sampling method ~~proposed in Schewe et al. (2014). The results~~ (Samaniego et al., 2017a) are shown in Fig. 6 (d-f)~~and spatially aggregated~~ , and the spatially aggregated results over the IPCC Europe regions in Tab. 4. The ~~overall~~ uncertainty rises with higher levels of warming for both sources of uncertainty because of two reasons. The GCM uncertainty increases because a 30 year period reaching a 3 K warming often has a strong temperature period (with higher than average annual temperature) within this 30 year period. On the contrary, GCM runs under the RCP2.6 often stabilise around a global warming level of 1.5 K. This pathway dependency of GCM runs influences the variability of the results with expectedly higher variability in the former case (James et al., 2017). The HM uncertainty increases with global warming because certain regions might cross thresholds. For example, parts of France might move from a energy-limited to a water-limited regime. The contribution of the GCMs to the ~~uncertainty over Europe is about~~ overall uncertainty across Europe is approximately 21% higher under a global warming level of 1.5 K, 25% higher under 2 K, and only 10% higher under a global warming of 3 ~~K global warming~~ K in comparison to the HM contribution. ~~The uncertainty rises with higher levels of warming for both sources of uncertainty. This may be related to "how" a level of warming is reached - i.e., the pathway dependency. Within a 30-year period, the temperature equilibrium could already have been reached with almost constant annual temperatures. This would result in different climate impacts compared to a constant rise of annual temperatures within a 30-year period with the same overall level of warming (James et al., 2017). HMs are the major source of uncertainty~~ This decrease of GCM/HM contribution can be mostly attributed to the Mediterranean and Atlantic regions (in particular France). In these dry regions, the different representations of evaporation using temperature-based potential evapotranspiration used in mHM and PCR-GLOBWB will lead to a different evaporative response compared to explicitly solving the full energy-balance of the land surface as in Noah-MP. Furthermore, HMs contribution to the total uncertainty is regionally higher than average in the Alpine ~~region~~and Northern regions, where snow accumulation and melt play an important role (Fig. 6 panels d-f). Snow processes are treated differently between the HMs, which explains the relatively high uncertainties in the Northern and Alpine area. Both mHM and PCR-GLOBWB use a temperature based conceptual degree-day method for snow processes, whereas the NOAH-MP model employs an energy balance scheme to resolve the snow accumulation and melt processes. In the Atlantic and Continental ~~region, uncertainty is dominated by the GCMs~~ regions, GCM uncertainty is higher under all levels of warming. One reason is that the lower quantiles of summer precipitation in CMIP5 simulations are generally underestimated and have a large spread in Central Europe (Liu et al., 2014). In the Rhine River basin, the spread in summer precipitation ~~in~~ across the five GCMs used in this study ~~was~~ is highest compared to other seasons (Krysanova and Hattermann, 2017). Remarkably, within the summer season the spread was higher under RCP8.5 compared to RCP2.6. Furthermore, HMs ~~show equal performance~~ generally show nearly similar skill in humid areas where most of the models have been developed and calibrated (Huang et al., 2017). The Northern area shows a nearly similar contribution in GCMs and HMs. In the Mediterranean, the uncertainty due to the HMs rises with increased warming. Reasons ~~are~~ for such behavior could be the increased importance of the soil moisture and resulting actual evapotranspiration as well as infiltration treatment, which differs substantially between the HMs. For example, mHM uses separate storages for actual evapotranspiration and different runoff components (fast and slow interflow and baseflow components), whereas actual evapotranspiration and runoff depend on the same storages in Noah-MP leading to a higher inter-variable dependency. This suggests that differences in soil ~~representation~~ and runoff representations within a model can have a significant effect on the simulation of future ~~hydrological droughts~~ low flows, and can have a significant impact on the trend signal, as ~~confirmed~~ also had been previously noted by Wanders and Van Lanen (2015).

The procedure to differentiate between GCM and HM uncertainty has previously been presented in Samaniego et al. (2017a).

They used six HMs forced with bias-corrected outputs from five GCMs under two RCPs set up in seven large river basins worldwide for the period 1971-2099. ~~Similarly to~~ Similar to the findings of this study, they ~~found~~ also reported that uncertainty for a runoff index increases with time which corresponds to increased warming. Furthermore, the GCMs generally dominate the HMs uncertainty in ~~droughts~~low flows. Nevertheless, they also agree on the fact that the uncertainty contribution of the HMs depends on the hydro-climatic regime. Similarly, Vetter et al. (2015) used the ANOVA method to distinguish between different sources of uncertainty, including RCP uncertainty, which is not separately investigated here. For low flows, they came up with a 70% contribution of RCPs on the drought impacts, with RCP uncertainty rising until the end of the 21st century. This may be explained due to the widening temperature range in the RCPs over time, which is not comparable to our approach of using a time sampling approach to identify different warming levels (Collins et al., 2013).

Overall, the regions showing higher uncertainty contribution from GCMs exhibited comparably lower SNR, indicating a significant variability in the GCM projections that are propagated through the HMs to the low flow signal. Furthermore, the contribution of the GCMs to the total uncertainty ~~dominates~~ is higher than the contribution of HMs over Europe. Nevertheless, the influence of HMs cannot be neglected and outperforms the uncertainties in GCMs in some regions and depending on the warming level. Our results therefore strongly suggest the use of multiple hydrologic models for climate change impact assessment studies for future low flow projections, and that the use of single hydrologic models may provide misleading results.

4 Summary and Conclusions

Climate change is projected to alter low flows expressed as the Q90 indicator in Europe under a global warming of 1.5, 2 and 3 K~~global warming~~. The magnitude of changes as well as the robustness in 45 member multi-model ensemble is amplified with increased levels of warming. Higher levels of warming therefore demand more distinctive adaptation actions. The mountainous regions in Europe show the strongest low flow increase from 22% under 1.5 K to 45% under a warming of 3 K~~warming~~. Continental Europe faces slight decreases in low flows. Higher decreases are expected in the Mediterranean (up to -35% under 3 K warming) and the Atlantic. We conclude that a warming level of 3~~K warming level~~ K will impose higher water stress over a large part of the Mediterranean, an area which already suffers from limited water resources that makes adaptation necessary. Further limitations in water availability may result in new managing challenges for water resource managers and policy makers, including the management of competition for water resources between sectors.

The projected changes in Q90 across Europe between the reference period (1971-2000) and a warming level of 1.5 K~~warming level~~, as well as between a global warming of 1.5 K and 2 K ~~global warming~~ are generally small with a low robustness and a small signal to noise ratio. It is not possible to distinguish climate impacts between a global warming of 1.5 K and 2 K~~global warming~~. Nevertheless, some hotspot regions show changes greater than ± 10% between all warming levels investigated in this study. It would be misleading to conclude that mitigation of greenhouse gases is not needed. It is revealed here that large parts of the change in the climate induced low flow signal between the reference period and a ~~2 K~~ global warming level of 2 K already happens before reaching the warming level of 1.5 K~~warming level~~, specifically in the Alpine regions, Northern Europe and the Mediterranean. Therefore, mitigating climate change even below the 1.5 degree goal (UNFCCC, 2015) would be necessary to reduce negative drought impacts in hotspot regions like the Mediterranean.

The results shown here are independent of the uncertainty in emission scenarios. On the other hand, the uncertainty of the determination in the time periods for different warming levels is introduced. Generally, the robustness in the simulations and signal to noise ratio in the ensemble rise with increased warming and with the magnitude of change. As a result, regions with relatively large changes in ~~hydrological droughts~~ low flows show a relatively low uncertainty in the results and have therefore the highest need to adapt to changing conditions. It is observed here that the selection of metrics to define uncertainty strongly influences the result. Here, we use the combination of robustness covering the significance in the change for every single ensemble member together with SNR pointing at the variability and strength of the signal for the overall ensemble. Uncertainties should be considered in adaptation planning, e.g., in deciding to use climate impact simulations to determine regional vulnerability quantitatively or qualitatively. We conclude that the combination of different kinds of information, namely the change signal, the robustness and SNR, ~~support~~ should be used in the adaptation process. These can be used to decide e.g., on the adaptation need or if a quantitative or qualitative approach should be chosen for the estimation of regional vulnerability to climate change. It is observed that the GCMs ~~generally dominate over the HMs~~ contribution to the overall uncertainty is higher than the HM contribution across Europe and that the HM contribution to total uncertainty rises with increased warming. This is related to the exhibited strong correspondence between the changes in ~~the~~ mean annual total precipitation and streamflow Q90, which is ~~most pronounced for~~ strongest in lower warming levels and in the Atlantic and Continental Europe. Nevertheless, the HM contribution cannot be neglected and in some regions, it is higher than the GCM contribution especially in the Alpine, Northern and Mediterranean, with rising global temperatures. The main reasons are the rising importance of hydrologic process description of snow, soil moisture and evapotranspitation, and infiltration. We conclude that climate change studies focusing on river low flows should employ large multi-model ensembles including multiple driving climate models as well as multiple impact models to provide a comprehensive analysis of model uncertainty.

Competing interests. The authors declare that no competing interests are present.

Acknowledgements. This study has been funded within the scope project HOKLIM (www.ufz.de/hoklim) by the German ministry for education and research (BMBF, grant number 01LS1611A). This study has been partially funded by the Copernicus Climate Change Service. The European Centre for Medium Range Weather Forecasts implements this service and the Copernicus Atmosphere Monitoring Service on behalf of the European Commission. We would like to thank all the colleagues contributing to the EDgE (http://edge.climate.copernicus.eu/) project. Furthermore, we acknowledge the funding from NWO Rubicon 825.15.003. The ENSEMBLES data used in this work was funded by the EU FP6 integrated project ensembles (contract number 505539) whose support is gratefully acknowledged. We acknowledge the E-OBS dataset from the EU FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (http://www.ecad.eu). We would like to thank people from various organizations and projects for kindly providing us the data which ~~were~~ are used in this study, which includes ISI-MIP, JRC, ESA, NASA, USGS, GRDC, BGR, UNESCO, ISRIC, EEA, EWA & CEDEX.

References

Cherkauer, K. A., Bowling, L. C., and Lettenmaier, D. P.: Variable infiltration capacity cold land process model updates, Global and Planetary Change, 38, 151 – 159, doi:http://dx.doi.org/10.1016/S0921-8181(03)00025-0, http://www.sciencedirect.com/science/article/pii/S0921818103000250, project for Intercomparison of Land-surface Parameterization Schemes, Phase 2(e), 2003.

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A., and Wehner, M.: Long-term Climate Change: Projections, Commitments and Irreversibility, book section 12, p. 1029–1136, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, doi:10.1017/CBO9781107415324.024, www.climatechange2013.org, 2013.

Cuntz, M., Mai, J., Samaniego, L., Clark, M., Wulfmeyer, V., Branch, O., Attinger, S., and Thober, S.: The impact of standard and hard-coded parameters on the hydrologic fluxes in the Noah-MP land surface model, Journal of Geophysical Research: Atmospheres, 121, 10,676–10,700, 2016.

Donnelly, C., Andersson, J. C., and Arheimer, B.: Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe, Hydrological Sciences Journal, 61, 255–273, doi:10.1080/02626667.2015.1027710, http://dx.doi.org/10.1080/02626667.2015.1027710, 2016.

Donnelly, C., Greuell, W., Andersson, J., Gerten, D., Pisacane, G., Roudier, P., and Ludwig, F.: Impacts of climate change on European hydrology at 1.5, 2 and 3 degrees mean global warming above preindustrial level, Climatic Change, 19, 1–14, 2017.

EC: Water Scarcity and Droughts. Second Interim report, Tech. rep., DG Environment, European Commission, http://ec.europa.eu/environment/water/quantity/pdf/comm_droughts/2nd_int_report.pdf, 2007.

Fleig, A. K., Tallaksen, L. M., Hisdal, H., and Demuth, S.: A global evaluation of streamflow drought characteristics, Hydrology and Earth System Sciences, 10, 535–552, doi:10.5194/hess-10-535-2006, http://www.hydrol-earth-syst-sci.net/10/535/2006/, 2006.

Forzieri, G., Feyen, L., Rojas, R., Flörke, M., Wimmer, F., and Bianchi, A.: Ensemble projections of future streamflow droughts in Europe, Hydrology and Earth System Sciences, 18, 85–108, doi:10.5194/hess-18-85-2014, http://www.hydrol-earth-syst-sci.net/18/85/2014/, 2014.

Giannakopoulos, C., Sager, P. L., Bindi, M., Moriondo, M., Kostopoulou, E., and Goodess, C.: Climatic changes and associated impacts in the Mediterranean resulting from a 2 °C global warming, Global and Planetary Change, 68, 209 – 224, doi:http://dx.doi.org/10.1016/j.gloplacha.2009.06.001, http://www.sciencedirect.com/science/article/pii/S0921818109001131, 2009.

Giuntoli, I., Vidal, J. P., Prudhomme, C., and Hannah, D. M.: Future hydrological extremes: the uncertainty from multiple global climate and global hydrological models, Earth System Dynamics, 6, 267–285, 2015.

Gosling, S. N., Zaherpour, J., Mount, N. J., Hattermann, F. F., Dankers, R., Arheimer, B., Breuer, L., Ding, J., Haddeland, I., Kumar, R., Kundu, D., Liu, J., van Griensven, A., Veldkamp, T. I. E., Vetter, T., Wang, X., and Zhang, X.: 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, Climatic Change, 141, 577–595, doi:10.1007/s10584-016-1773-3, http://dx.doi.org/10.1007/s10584-016-1773-3, 2017.

Hall, J., Arheimer, B., Borga, M., Brazdil, R., Claps, P., Kiss, A., Kjeldsen, T. R., Kriauciuniene, J., Kundzewicz, Z. W., Lang, M., Llasat, M. C., Macdonald, N., McIntyre, N., Mediero, L., Merz, B., Merz, R., Molnar, P., Montanari, A., Neuhold, C., Parajka, J., Perdigao, R. A. P., Plavcova, L., Rogger, M., Salinas, J. L., Sauquet, E., Schaer, C., Szolgay, J., Viglione, A., and Bloeschl, G.: Understanding flood regime changes in Europe: a state-of-the-art assessment, Hydrology and Earth System Sciences, 18, 2735–2772, 2014.

Hattermann, F. F., Krysanova, V., Gosling, S. N., Dankers, R., Daggupati, P., Donnelly, C., Flörke, M., Huang, S., Motovilov, Y., Buda, S., Yang, T., Müller, C., Leng, G., Tang, Q., Portmann, F. T., Hagemann, S., Gerten, D., Wada, Y., Masaki, Y., Alemayehu, T., Satoh, Y., and Samaniego, L.: Cross-scale intercomparison of climate change impacts simulated by regional and global hydrological models in eleven large river basins, Climatic Change, 141, 561–576, 2017.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, Journal of Geophysical Research: Atmospheres (1984–2012), 113, 2008.

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

Huang, S., Kumar, R., Flörke, M., Yang, T., Hundecha, Y., Kraft, P., Gao, C., Gelfan, A., Liersch, S., Lobanova, A., Strauch, M., van Ogtrop, F., Reinhardt, J., Haberlandt, U., and Krysanova, V.: Evaluation of an ensemble of regional hydrological models in 12 large-scale river basins worldwide, Climatic Change, 141, 381–397, doi:10.1007/s10584-016-1841-8, http://dx.doi.org/10.1007/s10584-016-1841-8, 2017.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.-F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., and Yiou, P.: EURO-CORDEX: new high-resolution climate change projections for European impact research, Regional Environmental Change, 14, 563–578, doi:10.1007/s10113-013-0499-2, http://dx.doi.org/10.1007/s10113-013-0499-2, 2014.

James, R., Washington, R., Schleussner, C.-F., Rogelj, J., and Conway, D.: Characterizing half-a-degree difference: a review of methods for identifying regional climate responses to global warming targets, Wiley Interdisciplinary Reviews - Climate Change, 8, 2017.

Kovats, R., Valentini, R., Bouwer, L., Georgopoulou, E., Jacob, D., Martin, E., Rounsevell, M., and Soussana, J.-F.: Europe, In: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change[Barros, V.R., C.B. Field, D.J. Dokken, M.D. Mastrandrea, K.J. Mach, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.)], pp. 1267–1326, https://ipcc.ch/pdf/assessment-report/ar5/wg2/WGIIAR5-Chap23_FINAL.pdf, 2014.

Krysanova, V. and Hattermann, F. F.: Intercomparison of climate change impacts in 12 large river basins: overview of methods and summary of results, Climatic Change, 141, 363–379, doi:10.1007/s10584-017-1919-y, http://dx.doi.org/10.1007/s10584-017-1919-y, 2017.

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

Laaha, G., Gauster, T., Tallaksen, L. M., Vidal, J.-P., Stahl, K., Prudhomme, C., Heudorfer, B., Vlnas, R., Ionita, M., Van Lanen, H. A. J., Adler, M.-J., Caillouet, L., Delus, C., Fendekova, M., Gailliez, S., Hannaford, J., Kingston, D., Van Loon, A. F., Mediero, L., Osuch, M., Romanowicz, R., Sauquet, E., Stagge, J. H., and Wong, W. K.: The European 2015 drought from a hydrological perspective, Hydrology and Earth System Sciences, 21, 3001–3024, doi:10.5194/hess-21-3001-2017, https://www.hydrol-earth-syst-sci.net/21/3001/2017/, 2017.

Liu, Z., Mehran, A., Phillips, T., and AghaKouchak, A.: Seasonal and regional biases in CMIP5 precipitation simulations, Climate Research, 60, 35–50, doi:10.3354/cr01221, http://www.int-res.com/abstracts/cr/v60/n1/p35-50/, 2014.

Mendoza, P. A., Clark, M. P., Mizukami, N., Newman, A. J., Barlage, M., Gutmann, E. D., Rasmussen, R. M., Rajagopalan, B., Brekke, L. D., and Arnold, J. R.: Effects of Hydrologic Model Choice and Calibration on the Portrayal of Climate Change Impacts, Journal of Hydrometeorology, 16, 762–780, doi:10.1175/JHM-D-14-0104.1, https://doi.org/10.1175/JHM-D-14-0104.1, 2015.

Metzger, M. J., Bunce, R. G. H., Jongman, R. H. G., Mücher, C. A., and Watkins, J. W.: A climatic stratification of the environment of Europe, Global Ecology and Biogeography, 14, 549–563, doi:10.1111/j.1466-822X.2005.00190.x, http://dx.doi.org/10.1111/j.1466-822X.2005.00190.x, 2005.

Mitchell, D., James, R., Forster, P. M., Betts, R. A., Shiogama, H., and Allen, M.: Realizing the impacts of a 1.5 °C warmer world, Nature Clim. Change, 6, 735–737, http://dx.doi.org/10.1038/nclimate3055, commentary, 2016.

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, Journal of Geophysical Research, 116, 2011.

Peters, G. P., Andrew, R. M., Boden, T., Canadell, J. G., Ciais, P., Le Quéré, C., Marland, G., Raupach, M. R., and Wilson, C.: The challenge to keep global warming below 2 °C, Nature Climate Change, 3, 4 EP –, http://dx.doi.org/10.1038/nclimate1783, 2012.

Prudhomme, C., Giuntoli, I., Robinson, E. L., Clark, D. B., Arnell, N. W., Dankers, R., Fekete, B. M., Franssen, W., Gerten, D., Gosling, S. N., Hagemann, S., Hannah, D. M., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y., and Wisser, D.: Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment, Proceedings of the National Academy of Sciences, 111, 3262–3267, doi:10.1073/pnas.1222473110, http://www.pnas.org/content/111/9/3262.abstract, 2014.

Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., Attinger, S., Schäfer, D., Schrön, M., and Samaniego, L.: Multiscale and multivariate evaluation of water fluxes and states over european river Basins, Journal of Hydrometeorology, 17, 287–307, 2016.

Roudier, P., Andersson, J. C. M., Donnelly, C., Feyen, L., Greuell, W., and Ludwig, F.: Projections of future floods and hydrological droughts in Europe under a +2°C global warming, Climatic Change, 135, 341–355, doi:10.1007/s10584-015-1570-4, http://dx.doi.org/10.1007/s10584-015-1570-4, 2016.

Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resources Research, 46, 2010.

Samaniego, L., Kumar, R., Breuer, L., Chamorro, A., Flörke, M., Pechlivanidis, I. G., Schäfer, D., Shah, H., Vetter, T., Wortmann, M., and Zeng, X.: Propagation of forcing and model uncertainties on to hydrological drought characteristics in a multi-model century-long experiment in large river basins, Climatic Change, 141, 435–449, doi:10.1007/s10584-016-1778-y, http://dx.doi.org/10.1007/s10584-016-1778-y, 2017a.

Samaniego, L., Kumar, R., Mai, J., Zink, M., Thober, S., Cuntz, M., Rakovec, O., Schäfer, D., Schrön, M., Brenner, J., Demirel, C. M., Kaluza, M., Langenberg, B., Stisen, S., and Attinger, S.: mesoscale Hydrologic Model, doi:10.5281/zenodo.1069203, https://doi.org/10.5281/zenodo.1069203, 2017b.

Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N. W., Clark, D. B., Dankers, R., Eisner, S., Fekete, B. M., Colón-González, F. J., Gosling, S. N., Kim, H., Liu, X., Masaki, Y., Portmann, F. T., Satoh, Y., Stacke, T., Tang, Q., Wada, Y., Wisser, D., Albrecht, T., Frieler, K., Piontek, F., Warszawski, L., and Kabat, P.: Multimodel assessment of water scarcity under climate change, Proceedings of the National Academy of Sciences, 111, 3245–3250, doi:10.1073/pnas.1222460110, http://www.pnas.org/content/111/9/3245.abstract, 2014.

Schleussner, C.-F., Lissner, T. K., Fischer, E. M., Wohland, J., Perrette, M., Golly, A., Rogelj, J., Childers, K., Schewe, J., Frieler, K., Mengel, M., Hare, W., and Schaeffer, M.: Differential climate impacts for policy-relevant limits to global warming: the case of 1.5 °C and 2 °C, Earth System Dynamics, 7, 327–351, doi:10.5194/esd-7-327-2016, http://www.earth-syst-dynam.net/7/327/2016/, 2016.

Schneider, C., Laizé, C. L. R., Acreman, M. C., and Flörke, M.: How will climate change modify river flow regimes in Europe?, Hydrology and Earth System Sciences, 17, 325–339, doi:10.5194/hess-17-325-2013, https://www.hydrol-earth-syst-sci.net/17/325/2013/, 2013.

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L.: Climate Change 2007 - The Physical Science Basis , Contribution of Working Group I to the Fourth Assessment Report of the IPCC, Cambridge University Press, Cambridge, United Kingdom and New York, USA, 2007.

Stagge, J., Rizzi, J., Tallaksen, L. M., and Stahl, K.: Future Meteorological Drought: Projections of Regional Climate Models for Europe. DROUGHT-R&SPI deliverable D3.3, http://www.eu-drought.org/media/default.aspx/emma/org/10859960/DROUGHT+RSPI+Technical+Report+No.+25+-+Future+Meteorological+Drought+Projections+of+Regional+Climate.pdf, 2015.

Stahl, K., Kohn, I., Blauhut, V., Urquijo, J., De Stefano, L., Acácio, V., Dias, S., Stagge, J. H., Tallaksen, L. M., Kampragou, E., Van Loon, A. F., Barker, L. J., Melsen, L. A., Bifulco, C., Musolino, D., de Carli, A., Massarutto, A., Assimacopoulos, D., and Van Lanen, H. A. J.: Impacts of European drought events: insights from an international database of text-based reports, Natural Hazards and Earth System Sciences, 16, 801–819, doi:10.5194/nhess-16-801-2016, http://www.nat-hazards-earth-syst-sci.net/16/801/2016/, 2016.

Thober, S., Kumar, R., Wanders, N., Marx, A., Pan, M., Rakovec, O., Samaniego, L., Sheffield, J., Wood, E. F., and Zink, M.: Multi-model ensemble projections of European river floods and high flows at 1.5, 2, and 3 degree global warming, Environmental Research Letters, pp. 1–22, doi:https://doi.org/10.1088/1748-9326/aa9e35, 2017.

UNFCCC: Adoption of the Paris agreement, Proposal by the President, Tech. rep., United Nations, Geneva, Switzerland, 2015.

Van Beek, L. and Bierkens, M.: The Global Hydrological Model PCR-GLOBWB: Conceptualization, Parameterization and Verification, Tech. rep., Department of Physical Geography, Utrecht University, Utrecht, The Netherlands, http://vanbeek.geo.uu.nl/suppinfo/vanbeekbierkens2009.pdf, 2008.

Van Lanen, H. A., Laaha, G., Kingston, D. G., Gauster, T., Ionita, M., Vidal, J.-P., Vlnas, R., Tallaksen, L. M., Stahl, K., Hannaford, J., Delus, C., Fendekova, M., Mediero, L., Prudhomme, C., Rets, E., Romanowicz, R. J., Gailliez, S., Wong, W. K., Adler, M.-J., Blauhut, V., Caillouet, L., Chelcea, S., Frolova, N., Gudmundsson, L., Hanel, M., Haslinger, K., Kireeva, M., Osuch, M., Sauquet, E., Stagge, J. H., and Van Loon, A. F.: Hydrology needed to manage droughts: the 2015 European case, Hydrological Processes, 30, 3097–3104, doi:10.1002/hyp.10838, http://dx.doi.org/10.1002/hyp.10838, hYP-16-0162, 2016.

Van Loon, A.: Hydrological drought explained, Wiley Interdisciplinary Reviews: Water, 2, 359–392, doi:10.1002/wat2.1085, http://dx.doi.org/10.1002/wat2.1085, 2015.

van Vliet, M. T., Donnelly, C., Strömbäck, L., Capell, R., and Ludwig, F.: European scale climate information services for water use sectors, Journal of Hydrology, 528, 503 – 513, doi:http://dx.doi.org/10.1016/j.jhydrol.2015.06.060, http://www.sciencedirect.com/science/article/pii/S0022169415004734, 2015.

van Vliet, M. T. H., Sheffield, J., Wiberg, D., and Wood, E. F.: Impacts of recent drought and warm years on water resources and electricity supply worldwide, Environmental Research Letters, 11, 124 021, http://stacks.iop.org/1748-9326/11/i=12/a=124021, 2016.

Vautard, R., Gobiet, A., Sobolowski, S., Kjellstrom, E., Stegehuis, A., Watkiss, P., Mendlik, T., Landgren, O., Nikulin, G., Teichmann, C., and Jacob, D.: The European climate under a 2 degrees C global warming, Environmental Research Letters, 9, 2014.

Vaze, J., Post, D., Chiew, F., Perraud, J.-M., Viney, N., and Teng, J.: Climate non-stationarity – Validity of calibrated rainfall–runoff models for use in climate change studies, Journal of Hydrology, 394, 447 – 457, doi:https://doi.org/10.1016/j.jhydrol.2010.09.018, http://www.sciencedirect.com/science/article/pii/S0022169410005986, 2010.

Vetter, T., Huang, S., Aich, V., Yang, T., Wang, X., Krysanova, V., and Hattermann, F.: Multi-model climate impact assessment and intercomparison for three large-scale river basins on three continents, Earth System Dynamics, 6, 17–43, doi:10.5194/esd-6-17-2015, https://www.earth-syst-dynam.net/6/17/2015/, 2015.

Wanders, N. and Van Lanen, H. A. J.: Future discharge drought across climate regions around the world modelled with a synthetic hydrological modelling approach forced by three general circulation models, Natural Hazards and Earth System Sciences, 15, 487–504, doi:10.5194/nhess-15-487-2015, https://www.nat-hazards-earth-syst-sci.net/15/487/2015/, 2015.

Wanders, N. and Wada, Y.: Human and climate impacts on the 21st century hydrological drought, Journal of Hydrology, 526, 208 – 220, doi:http://dx.doi.org/10.1016/j.jhydrol.2014.10.047, http://www.sciencedirect.com/science/article/pii/S0022169414008427, 2015.

Wanders, N., Wada, Y., and Van Lanen, H. A. J.: Global hydrological droughts in the 21st century under a changing hydrological regime, Earth System Dynamics, 6, 1–15, doi:10.5194/esd-6-1-2015, https://www.earth-syst-dynam.net/6/1/2015/, 2015.

Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., and Schewe, J.: The Inter-Sectoral Impact Model Intercomparison Project (ISI–MIP): Project framework, Proceedings of the National Academy of Sciences, 111, 3228–3232, doi:10.1073/pnas.1312330110, http://www.pnas.org/content/111/9/3228.abstract, 2014.

Zink, M., Samaniego, L., Kumar, R., Thober, S., Mai, J., Schäfer, D., and Marx, A.: The German drought monitor, Environmental Research Letters, 11, 074 002, http://stacks.iop.org/1748-9326/11/i=7/a=074002, 2016.

Zink, M., Kumar, R., Cuntz, M., and Samaniego, L.: A high-resolution dataset of water fluxes and states for Germany accounting for parametric uncertainty, Hydrology and Earth System Sciences, 21, 1769–1790, doi:10.5194/hess-21-1769-2017, https://www.hydrol-earth-syst-sci.net/21/1769/2017/, 2017.

Author comment: Interactive comment on "Climate change alters low flows in Europe under a 1.5, 2, and 3 degree global warming" by Andreas Marx et al.; Anonymous Referee #1

Marx Andreas[1]

[1]Department Computational Hydrosystems, Helmholtz Centre for Environmental Research (UFZ), Leipzig, Germany

Correspondence to: Andreas Marx (klima@ufz.de)

We thank the reviewer for the time and effort in commenting on our manuscript. We provide responses to each individual point below. For clarity, comments are given in normal font, and our responses are given as blue text.

The authors present a comprehensive study of change in low flows for Europe using downscaled GCM output fed into three different hydrologic models. I am happy to recommend publishing of the manuscript subject to maybe some clarifications.

This paper is looking at changes in the percentile (as the abstract says) ? but the introduction is focuses on droughts. As it is currently phrased I am not sure I feel comfortable with research Question 1. I think this should be changed to say it is looking at changes in low flows. The introduction needs some text to relate drought to low flows. I understand that at the bottom of page 4 it is stated that Q90 is the drought metric but this comes too late in the piece.

Thank you for this helpful comment. We agree to include the clear differentiation between the terms "hydrological drought" and "low flow" and we will adapt research question 1, accordingly. We suggest to include the paragraph:

"This study investigates low streamflow, defined as Q90, representing daily streamflow exceeding 90% of the time, which has the potential to impact hydrological drought. Hydrological drought is associated with shortfalls on surface or subsurface water availability which can occur in low streamflow, groundwater, or reservoir levels. Changes in low flows shown in this study can, but will not in every case, result in drought. Exceptions are e.g. riverine based transportation, where streamflow values below a threshold level are defined as hydrological drought."

I think there are a few papers that could be cited in the introduction, for example, Hall et al. (2014); 10.5194/hess-17-325-2013; and a recent article that looks at the sensitivity of flows to temperature 10.1038/s41598-017-81-1084.

We agree to include 10.5194/hess-17-325-2013 in the introduction. Two other suggested paper are deemed beyond the scope of this manuscript.

I did find it odd that a lot of material was introduced in the discussion on Page 10 and Page 17/18. Given it is relevant I think the introduction needs to (at least briefly) incorporate these references to put this works novelty in context.

Thank you for your suggestion. We will extend introduction with the studies mentioned in the discussion.

Could the bias correction be elaborated in a sentence or two because the choice of bias correction can make a huge difference to the results? Especially if the fo- cus is drought, authors need to correct for low-frequency variability biases - see 10.1016/j.jhydrol.2016.04.018.

There is a huge number of bias correction methods available, all facing advantages and disadvantages (e.g. https://doi.org/10.1016/j.jhydrol.2012.05.052) for hydrological impact studies. The advantage of the method applied in our manuscript (Hempel et al.) is that it is trend preserving, which is of major importance for climate impact studies. We would have a different opinion, if this is meant by the comment that "authors need to correct for low-frequency variability biases" in daily precipitation and temperature, because they are hardly directly linkable to low flow events.

We refer to the statement in in Donnelly et al (2017): "Cannon et al. (2015) and Maurer and Pierce (2014) showed that approaches like the quantile mapping used here can change the climate signal in the raw CM output significantly. Nevertheless, it is still unclear which methods give the most realistic climate change projections.".

Worth noting we are tracking for higher increases than 3 degrees probably: 10.1038/nclimate1783

Agreed.

Can the results in Table 1 be verbally contrasted with land predictions for Europe (i.e. will Europe heat up more or less than the global average). The IPCC reports will have this.

This is a good suggestion, but out of scope of this manuscript. Europe warms faster than the global mean, which has been visualised (for the underlying 5 GCM simulations) in http://edge.climate.copernicus.eu/Apps/#climate-change

I am pretty sure that the low flow statistics in Table 2 are based on average of all the grid cells in a region but I am not sure. This could be mentioned in the text.

We we will reformulate accordingly.

Figure 4 ? not really clear to me what the blue dashed line indicates. I think the lines need to be described in the legend.

Agreed. The dashed lines show two regressions (for positive and for negative deviations).

It is a bit hard to assess Table 3 because the step changes aren't linear. You could compare the following: Table1 Row 1 (0-1.5K) increase equivalent to 22, -7, -4, 8, -12 changes and comparing to Row 3 in Table 2 (again a 1.5 K increase but now from 1.5 to 3K) of 24, -13, -12, 23, -23.

For clarity, we suggest to include a row "absolute warming" (tab. 2: 1.04 K, 1.54 K, 2.54 K; tab. 3: 0.5 K, 1 K, 1.5 K). Comparison of results in tab. 2 and 3 rarely gives added value to the manuscript as changes are nonlinear with warming and regionally different - this is already reflected.

It was not clear to me how the GCM and HM signal-to-noise ratio was split.

The SNR was calculated for the combined GCM/HM runs (no splitting). If you refer to the GCM/HM uncertainty, the approach is described on page 8, line 2-8 in detail.

Abstract Line 5: Unprecedented is a strong word and I would remove it.

For Europe, there is no study available using a multi-model ensemble with 45 members including three impact models for low flows and at a high spatial resolution of $5\times\ 5\ km^2$. We think this justifies the usage of the term "unprecedented".

Page 8 Line 4: Typo. "...by first fixing a HM and then calculating the range of Q90 (max-min) corresponding to give GCM outputs and repeating the previous step ..."

Agreed.

Author comment: Interactive comment on "Climate change alters low flows in Europe under a 1.5, 2, and 3 degree global warming" by Andreas Marx et al.; Anonymous Referee #2

Marx Andreas[1]

[1]Department Computational Hydrosystems, Helmholtz Centre for Environmental Research (UFZ), Leipzig, Germany

Correspondence to: Andreas Marx (klima@ufz.de)

We thank the reviewer for the time and effort in commenting on our manuscript. We provide responses to each individual point below. For clarity, comments are given in normal font, and our responses are given as blue text.

General comments

This manuscript explores the impact of climate change on low river flows in Europe using a multi-model GCM and hydrological model (HM) ensemble under three global warming scenarios. The use of this ensemble allows the authors to assess the range of uncertainty in projections and the relative contributions of GCMs and HMs. Overall, it is an interesting and informative study, well-written and clear, supported by appropriate figures and references. There are some questions surrounding catchment selection for model validation and the general omission of smaller catchments, as well as the extent to which conclusions can be drawn on drought when analysing only flow percentiles. However, once these and some other interpretational aspects are addressed, I would recommend this study for publication.

Thank you for the overall positive feedback. From the altogether three reviews, we realised that the information given on calibration and validation needs to be extended in the manuscript. Smaller catchments have not generally been omitted. The catchment size at a horizontal resolution of 5×5 km^2 is limited by the DEM in determining the catchment boundaries. Therefore, the results and conclusions in this study are based on catchments (or better river grid cells) with a contributing area >1000 km^2 have been used for the study, and these are shown in figure 3 and figure 5.

The selection of catchments >10000 km^2 in figure 1 and 4 in the first version of the manuscript has been done for clarity reasons. This will be changed for validation figure 1. Notably, for the validation (see attached figure) we selected 357 basins based on daily streamflow data availability (selection criteria; complete dataset of 30 yrs, 1966-1995, this time period would change (old: 1971-2000) because it resulted into largest sample size). Their median basin area is 1680 km^2.

Furthermore, the terms "hydrological drought" and "low flow" will be better specified and used in a coherent way.

Specific comments

Evaluating model performance (Page 4, line 30 to Page 5, line 4): I think more interpretation is required of Fig 1. There are only nine lines devoted to this, and I am not sure that I entirely agree with the assessment that "results show a good agreement" without some caveats. Low flows for PCR-GLOBWB and median flows for Noah-MP are systematically over-estimated across almost all catchment sizes, and there is a systematic under-estimation of low flows for Noah-MP. Whilst no-one is expecting perfect model results, there should be more attention given to the validation, as well as additional text in the discussion on the potential influence of model performance on the conclusions drawn.

Thank you for pointing this out. We suggest to include the following paragraph on the calibration of the hydrological models using observed meteorological forcing data (which focussed on headwater catchments). We suggest to include the paragraph: "The three HMs used in this study were calibrated in nine European focus basins located in Spain, Norway and UK, which were selected based on the consultation with the user groups within the EDgE project. Besides these, we also included three more central EU catchments (located in France and Germany) to represent diversity in hydro-climatic regimes. All HMs parameters were calibrated such that the model simulations represent a range of hydrologic regimes, rather than tailored to any specific characteristics. This was done in a consistent manner so that the model simulations can be used for a range of indicators (including high, low, and average flows) within the EDgE project. We recognize that HMs could be calibrated to a specific streamflow characteristic (in this case to low flows), but this was not considered within this study. We also note that the HMs do not consider human management effects in this study which could have substantial effect during the low flow times - as a result constraining the model to any specific low flow characteristic can result in a biased simulations. Also due to the similar reason we may expect a relatively lower model skill in matching the observed low flow characteristic."

The text "results show a good agreement" was written behind the background that GCM data for the time period 1971-2000, which differs from the observed weather in that period, was used to drive the HMs for the validation against simulated Q90 values. We agree that the discussion should be extended. Furthermore, based on Reviewer 3, Figure 1 will be re-drawn using specific discharge to remove the basin-scale dependency of the data (see above and attached figure). We acknowledge the adressed systematic biases and see only limited influence on the future results and conclusions drawn. It is not possible to determine if a model that fits perfectly in the past is also able to produce perfect results under changed climate conditions. Consequently, it cannot be concluded that imperfect models are not useable for estimating future (relative) changes.

Catchment selection for validation (Figure 2): There is no information on how or why these catchments were selected for validation. It would appear that a number of nested sub-catchments of relatively few large rivers have been selected (i.e. multiple downstream stations on the Rhone, Loire, Ebro, etc.) There is also no information on from where the river flow data were sourced. Data are freely available for some regions where the models are not evaluated but for which results are presented.

Agreed. The selection is partly explained in the general comments answer. We selected 357 basins based on data availability (selection criteria; complete dataset of 30 yrs, 1966-1995). Their median area is 1680 km^2. Fig. 2 will be adapted (see attached figure).

Omission of catchments <10,000km2 (Page 8, lines 13-15): Perhaps this argument explains the selection of catchments in Fig 2? I am not convinced that modelled data at 5km spatial resolution cannot resolve the river flow network of catchments <10,000km2. The authors highlight the "unprecedented" (Abstract) 5km spatial resolution and on a number of occasions high-

light the "spatially explicit information" in this study, but removing smaller catchments seems not to capitalise on this. This section also says that such catchments will be removed, but the maps displayed in Fig 3 onwards all feature a river flow network which contains routed flows for catchments less than 10,000km2, in which the network appears to be relatively well defined. All of this is relevant also in relation to the comment above on model performance at the lower end of the flow regime across all HMs (Fig 1). Catchments <10,000km2 also omitted from Fig 4; are the results similar?
This is not the case, explanations have been given in the author replies above.

Drought or low flows (throughout manuscript): There is some inconsistency between the use of 'drought' and 'low flows'. This paper analyses changes in median annual Q90 flows, which allows conclusions to be drawn on climate change impacts on low flows but not necessarily drought. The authors use low flows and drought at times interchangeably, including in the research questions and conclusions.
Thank you for this helpful comment. We agree do include the clear differentiation between the terms "hydrological drought" and "low flow" and we will adapt research question 1, respectively. We suggest to include the paragraph:
"This study investigates low streamflow, defined as Q90, representing daily streamflow exceeding 90% of the time, which has the potential to impact hydrological drought. Hydrological drought is associated with shortfalls on surface or subsurface water availability. These can occur i.e. in low streamflow, groundwater, or reservoir levels. Changes in low flow shown in this study can, but will not in any case, result in drought. Exceptions are e.g. riverine based transportation, where streamflows below a threshold level are defined as hydrological drought."

Robustness (Page 10, line 6): There is detail on the hotspots of changes in low flows, but in the end the low robustness means that for the Mediterranean / Atlantic, changes are not 'likely' (as defined by the authors) for most of these areas for either 1.5K or 2K. In fact, the signal for the Mediterranean might be stronger than that for the Atlantic, but it is less robust than the Atlantic. Statements like "Nevertheless, these results are not robust" (Page 13, lines 17-18) could be useful here.
We will modify the text as suggested.

Uncertainty from GCMs or HMs: There are a number of statements on Page 18 that need to be clarified in relation to Table 4. "HMs are the major source of uncertainty in the Alpine region" – GCMs and HMs are closer together in Alpine compared with other regions, but the numbers in Table 4 are similar for GCMs and HMs across all warming levels, and GCMs are higher for 1.5K. "The Northern area shows a nearly similar contribution in GCMs and HMs" – so does Alpine (see above), and GCMs and HMs are even more comparable for 2K and 3K in Alpine than in Northern. "In the Mediterranean, the uncertainty due to the HMs rises with increased warming" – this is true for all regions. It is also strong to say that GCMs "dominates" total uncertainty for Europe (Page 18, line 33), especially given the negligible differences between GCMs and HMs for two of the five regions.
Agreed. We will modify the text as suggested.

Technical corrections

Agreed. We will implement them in the revised manuscript.

Page 2, line 32: "differ- ent" to "different"

Page 4, line 11 - Page 5, line 4: Very lengthy paragraph could be better structured and split into multiple shorter paragraphs.

Figure 1: Useful to have a legend for colour based on GCM as there are some systematic patterns.

Figure 1 will be re-drawn (attached) to remove the basin-scale dependency of the data (based on comments of Reviewer 3, see above).

Page 11, line 9: "Q10" should be "Q90"?

Page 11, line 11: "to a large extent"

Page 15, line 11: Mediterranean should be "(-16%)" not "(-24%)", reading off Table 3 for 2K to 3K?

Page 17, line 1 (and throughout): "Targus" should be "Tagus"?

Table 4: It's more editorial, but Fig 6 discussed before Table 4 despite being featured afterwards.

Author comment: Interactive comment on "Climate change alters low flows in Europe under a 1.5, 2, and 3 degree global warming" by Andreas Marx et al.; Anonymous Referee #3

Marx Andreas[1]

[1]Department Computational Hydrosystems, Helmholtz Centre for Environmental Research (UFZ), Leipzig, Germany

Correspondence to: Andreas Marx (klima@ufz.de)

We thank the reviewer for the time and effort in commenting on our manuscript. We provide responses to each individual point below. For clarity, comments are given in normal font, and our responses are given as blue text.

This manuscript deals with a multi-GCM and multi-hydrological models assessment of changes in low flows across Europe between a present-day period (1971-200) and 3 different global warming levels: 1.5K, 2K, and 3K (and between them as well). It therefore contributes to document the effects of climate change on low-flow hydrology in Europe in the context of the Paris Agreement. This manuscript thus deals with a topical and important topic, and fits well into the scope of HESS. It is generally well structured and written, and conclusions are generally well supported by results shown. I have however two main comments (as well as specific comments) detailed below that should be addressed before the manuscript is published in HESS.

Thank you very much for the extensive commenting and feedback which will help to increase the quality of the manuscript.

1 Main comments

1.1 Hydrological calibration and simulation over influenced catchments The calibration details (specific comment #9 and #10) as well as the validation results (specific comments #12, #13, #14) do not give enough confidence on the quality of hydrological modeling, and highlights the issue of calibrating and/or validating – seemingly natural-catchment-only – models against highly influenced catchments like the Ebro or the Rhône, especially for low flows. First there is not enough information on the calibration process, and even the catchments used for that are not identified. Second, validation is done for a large part over influenced catchments, and also over ensembles of highly nested catchments. Both points should be reconsidered in a future revision of the manuscript.

We acknowledge these facts. The information given on calibration and validation will be modified and extended in the manuscript. More information is given in the specific comments.

1.2 Scale of catchments selected for presenting and averaging results There are numerous inconsistencies throughout the manuscript in terms of the minimum catchment size used for presenting results (and giving averaged figures), see specific comments #20, #21, #22, #24, #27. Addressing this comment may imply reformatting all results, but this is also intrinsically linked to main comment #1. Indeed, the manuscript state that the runoff routing scheme prevent using results for catch- ments smaller than 10000 km2, and near-natural catchments are usually only smaller than that in Europe. In parallel, maps of results are given over a river network en- compassing drained areas much small than the indicated threshold. This thus shades doubts (maybe unjustified, but is has to be demonstrated) on the validity of models and results, together with issues highlighted in main comment #1.

The information on catchment sizes will be extended in the manuscript. Smaller catchments <10000 km^2 have not been omitted. The catchment size at a spatial resolution of 5×5 km ^{2} is limited e.g. by the DEM in determining the catchment boundaries. Therefore, results from catchments (or better river grid cells) with a contributing area >1000 km^2 have been used in the study, and these are shown in figure 3 and figure 5 and have been used for drawing conclusions. The selection of catchments >10000 km^2 in figure 1 and 4 has been done for clarity reasons (clearness of the figures) only. This will be changed for validation figure 1. More information is given in the specific comments.

2 Specific comments

1. P1L2, “1.5, 2 and 3 K”: please specify that this is with respect to the preindustrial period

Agreed.

2. P1L10, “-12%”: What is the baseline period here? This is all the more important that there could easily be confusion with the baseline used for the global warming level (see above).

Agreed. Baseline period for determining relative changes is 1971-2000. This information will be included.

3. L11-12: this sentence is ambiguous. Less snowmelt may imply less streamflow in some conditions (e.g. constant liquid precipitation or declining total precipitation). Please rephrase and make it clearer.

Agreed. We will modify the text as suggested.

4. P1L13-14: This sentence is also quite ambiguous. What is exactly preventing dis- tinguishing between 1.5 and 2K warming effects? Is it the interannual variability which prevents distinguishing statistically estimates of period-averaged changes- for a given GCM-HM combination? Or is it the uncertainty due to the multimodel ensemble that prevents distinguishing ensembles of multimodel period-average estimates between present and future? Or both? Please make it clear here.

Agreed. It is both and we will include the information on uncertainty.

5. P2L11: The low-flow component of the 2015 drought event has been specifically studied by Laaha et al. (2017). I believe this reference is worth adding to the manuscript.

Agreed.

6. P2L21-27: What is the time slice that corresponds to the quantitative and quali- tative results recalled here? Please make it clearer.

1971-200. We sugegst to phrase "until the end of the century."

7. P4L7-10: First, this interpolation step should not be called downscaling as the latter refers to methods that actually add information for each day (either through regional climate models or empirical-statistical downscaling models) to the larger scale GCM fields. I would therefore strongly recommend using “disaggregating” or “disaggregated” instead of “downscaling” or “downscaled” in the manuscript.

We will modify the text as suggested.

8. P4L7-10: Second, this interpolation step should be better documented here, in order for the reader to understand the advantages and shortcomings of this approach, which are essential for assessing the quality of subsequent hydrological simulations. This interpolation step should ideally be assessed using a global reanalysis against high-resolution gridded datasets, like RCMs are actually as- sessed (see e.g. Kotlarski et al., 2017, among many others, for a recent exam- ple). This would critically allow distinguishing errors coming from (1) the spatial interpolation technique (and their large-scale forcings), and (2) the hydrological models. Please at least add some comments on that in the manuscript. Plus, the reference used for this interpolation technique is incomplete in the list of ref- erences.

We will extend the text as suggested. Assessing the meteorological input fields against other sources is out of the scope of this study. The missing reference will be added.

9. P4L24: What are these 9 catchments? Please provide some more information (location, surface, etc.). Are they near-natural or influenced catchments?

#9, #10 and #11 are commented together under #11

10. P4L24-26: What is the period used for calibration? And what are the calibration criteria (for both automatic and manual calibration)? Are they specific for low flows? Please carefully specify all this in the manuscript.

#9, #10 and #11 are commented together under #11

11. P4L27-29, “The assessment ... (Gosling et al., 2017)”: This is a very strong statement, which I tend to disagree with at least as a general conclusion. This is moreover hardly supported by the reference given in the manuscript, which compares global hydrological models and catchment hydrological models for the Rhine and Tagus (and other catchments, but not located in Europe). Results for a low flow in- dicator (Q95) show a large divergence of the two types of models with increasing global warming level (Gosling et al., 2017, their Fig. 2). As a conclusion, I would therefore strongly recommend removing this statement from the manuscript.

Comments #9, #10 and #11 are interrelated and are answered connectedly.

It is important to recognise that all HMs applied are well-established, widely applied and have been used (and calibrated) for Europe in former studies referred to in the manuscript. Furthermore, additional calibration for the three HMs was done in focus basins. Nevertheless, the validity of calibrated parameters may be limited in CC studies (Vaze et al., 2010) and the results in multi-HM climate impact studies may be less influenced by the calibration than by the model-structure of the HMs. This could be shown e.g. in Mendoza et al. (2015).

We agree to #16 that the statement "The assessment ... is independent" is too strong, we would remove this statement with the citation Gosling 2017 and replace it with "Well-calibrated HMs do not necessarily mean that future discharge under a changed climate can be reproduced satisfactorily (Vaze et al., 2010). Furthermore, the selection of HMs may have a larger effects than calibration in hydrological climate impact studies (Mendoza et al., 2015)".

Considering the HMs calibration (#9 and #10) we suggest to extend the manuscript and include the paragraph: "Furthermore, the three HMs used in this study were calibrated in nine near-natural European focus basins located in Spain, Norway and UK, which were selected based on the consultation with the user groups within the EDgE project. Besides these, we also included three more central EU catchments (located in France and Germany) to represent diversity in hydro-climatic regimes. All HMs parameters were calibrated such that the model simulations represent a range of hydrologic regimes, rather than tailored to any specific characteristics. This was done in a consistent manner so that the model simulations can be used for a range of indicators (including high, low, and average flows) within the EDgE project. HMs could be calibrated to specific parts of the flow duration curve (FDC), however, this was not done in this study to avoid too specific tuning of the model simulations to those unique conditions and thereby losing valuable information on the entire FDC.

In the current simulations human water management was not taken into account, since some models lack the ability to include these processes and one focus on this work is on determining the HM uncertainty in low flow conditions. Human water management can however have a significance impact on the low flow conditions, due to abstraction of additional water in drought conditions or changes in reservoir management - as a result constraining the model to any specific low flow characteristic can result in a biased simulations. Also due to the similar reason we may expect a relatively lower model skill in matching the observed low flow characteristic."

12. P4L33-P5L4 and Figure 1: The assessment of HMs is very light and not strongly supported by Fig. 1. Indeed, this figure is potentially misleading, as it basically only checks that catchments have equally small/large indicators (Q90 or Q50) for both observations and simulations, which is mainly driven by the size of the catchment. I would therefore recommend using a different and more informative representation of differences, preferably in terms of relative errors (in percents), and also preferably as maps in order to show the potential spatial pattern in errors. This representation would also greatly help in comparing present-day errors to relative changes presented later in the manuscript. I personally would not give too much credit for a model showing for a given location present-day errors as large as 3K future changes...

We agree to change the metric to remove the catchment size effect. Therefore, we would use specific discharge [mm/d], include the information on the HM ensemble mean, and show the relative bias spatially distributed (see attached figure). We tend to disagree with the statement " I personally would not give too much credit for a model showing for a given location present-day errors as large as 3K future changes...". The comparison shown here is an "honest" one because the HMs are driven with GCM input for a time period in the past. This is usually not shown in climate impact studies. Furthermore, assuming a constant error or bias over time in the GCM-HM simulations would result in perfect study results. Therefore, we trust the uncertainty measures presented in this study (SNR combined with robustness) more. Considering the relative biases shown in the attached figure it is important to notice that the spatial pattern is very different from the climate change signal (Fig. 3 and 5). It would be critical if these patterns would match.

13. Figure2, right: This figure shows the location of validation gauges used in Fig. 1. First, it shows that many points in Fig. 1 comes from the same rivers and are necessarily highly correlated, which inherently bring some bias to the results that should be representative of the whole Europe. I would strongly recommend removing redundant points scatter plots like presented in Fig. 1. This would not be a problem however with suggested spatial representations (cf. above).

The validation gauges have been changed and more gauges are included in the revised manuscript. This will be changed (attached figure), and additionally, spatial representations are shown.

14. Figure 2, right: The second point is that several validation gauges are located on highly influenced rivers. For example, the Ebro river (Spain) is heavily influenced by water abstractions for irrigation, and the seasonal regime of the Rhône river (France) is heavily influenced by all the hydropower reservoirs located in the Alps (and other surrounding mountain ranges). There are many other cases that can be spotted on the map. As a consequence, observed streamflow indicators for low flows simply cannot be compared to natural (i.e. without human influence) hydrological simulations for these catchments. A good fit to observations may indeed reveal that physical parameters in HMs are tweaked to compensate for no representation of human influence. This may not be a problem in itself (at least for practical modeling purposes if not scientifically satisfactory) if human influences would not have changed and would not change in the future. Which has happened and definitely will. As a conclusion, I would strongly recommend using only near-natural catchments as validation (and also calibration) gauges for natural hydrological modeling (as I suppose it is the case in the manuscript, even if some HMs considered may represent human influences). A number of reference hydrometric networks have recently been developed at the country scale (Hannaford and Marsh, 2008; Giuntoli et al., 2013; Murphy et al., 2013), and one should take advantage of these. Note that these networks overlap for some countries (but not for some other) with stations tagged “climate sensitive” in the Global Runoff data Centre.

Calibration in HMs has been performed using headwater catchments and no heavily human influenced basin was included. It would generally be a good idea to use “climate sensitive” stations only. Unluckily, these are not uniformly distributed all over Europe, but only available in selected countries. Esp. in the Mediterranean area there is no such station available. We would consider this comment in future studies in case an area-wide coverage of climate sensitive stations is available.

15. P6L3: The 0.46K figure has uncertainties attached to it, according to the reference cited (Vautard et al., 2014). Please do mention these uncertainties in the manuscript, with possibly additional references that provides 1971-2000 estimates of global warming level.

Agreed. "The warming of 0.46 K in an average value from three estimations with a spread between 0.437 K and 0.477 K."

16. P6L20: The use of calendar year is not entirely satisfactory for computing Q90 in snow-influenced catchments where the low-flow period (or one of the low-flow periods, which is a more difficult situation) may span two calendar years. Please consider changing the calculation procedure or at least justify this approximation.

We will extend the text including the limitation mentioned.

17. P7L8-9: Please mention here (rather than in the results section) that the robustness is compute as the percentage of projections showing a significant change.

Agreed.

18. Table 2: Please make clear that “1980s” refers to the 1971-2000 period.

Agreed.

19. P8L3-9: I don’t really understand this peculiar choice of method for computing the relative contributions of uncertainty from GCMs and HMs. Many studies demonstrated that simple Analysis of Variance (ANOVA) approaches are perfectly suited to this case, and it has been recently widely applied to compute contribution from GCMs and HMs (see e.g. Giuntoli et al., 2015; Vetter et al., 2017, among many others), even by some of the authors of the present manuscript (Mishra et al., 2015).

ANOVA approaches can critically take account of GCM/HMs interactions, which is presumably not the case of the method used, and of the different sizes of fixed effects. The set-up is here rather simple compared to more complex ones that consider unbalanced number of runs from each GCMs and/or multiple sources of uncertainty (see e.g. Addor et al., 2014; Vidal et al., 2016). I therefore strongly recommend using a simple two-way ANOVA approaches for the present study, or at least check current results against a simple two-way ANOVA approach. Indeed, I am unsure of how this sequential sampling approach relates to the more traditional ANOVA approach, and what their respective underlying hypotheses are. I would welcome some online discussion on this.

The rationale of not using ANOVA and the description of the sequential sampling procedure similar to that proposed by (Schewe et al., 2014) was explained in (Samaniego et al., 2017). In short, standard parametric procedures, such as the Analysis of Variance (ANOVA), require assumptions of normality to estimate significance levels (the F and the t-student test require that the underlying variable is normally distributed). The low-flow statistics estimated in this study are non-normal and hence standard methods are not appropriate. For the estimation of the relative contributions of uncertainty from GCMs and HMs we use the range of the ensemble instead of the variance as suggested by Schewe et al. The confidence interval and the significance level of variability is estimated, in this case, with a non-parametric method (basically it is bootstrapping). This method is called sequential sampling in Samaniego et al. 2016. because it includes a "sampling with replacement" procedure to generate confidence intervals for the range statistic. Moreover, a non- parametric (bootstrapping) procedure is preferred here to reduce the effects of the biased variance estimation due to the small sample size.

20. P8L13-15: First, this should come much earlier in the manuscript. Second, this is not consistent with maps of streamflow changes that seemingly include results for catchments with a surface lower than 10000 km2. This should be clarified. This is closely linked to specific comment #14.

First: Agreed. Second: explained in 1.2. The information that river grid cells with a contributing area >1000 km^2 are used in the results section will be included.

21. Figure 3 (and Fig. 5 and Fig. 6). See comment above. Plus, the figure indicated above each map is seemingly a continental average of the plotted value along the river network. First, this should be clarified. Second, this value is closely related to the choic3-15, e of the minimal catchment surface area considered. Values would be very different if, as stated P8L1only catchments with an area larger than 10000 km2 would be considered. Please make all these statement and results consistent across the manuscript.

Agreed.

22. P10L3-4: This statement is somewhat inconsistent with the choice of the calendar year use for the calculation of Q90. Please clarify this in the manuscript.

Agreed. See #16

23. P10L7, “models”: I presume this should be “simulations”.

Yes, right.

24. P11L1, “new spatially explicit information”. This is again contradictory with the 10000 km2 statement. Cf. comments above.

See general comment 1.2

25. P11L16-17.This sentence is ambiguous. The increased spread along the 1:1 line (i.e. when smaller and larger values are considered) does indeed contribute to a higher coefficient of determination, which is not the case for the spread across (i.e. with higher residuals from) the 1:1 line. Please rephrase.

Agreed.

26. Figure 4: Several presumably regression lines are given on the graph. Please either define and comment them, or remove them. Also, please add lines delimiting the quadrants.

The regression lines are shown for positive and negative values. Modifications will be done as proposed.

27. Figure 4: The legend states that only catchments with a surface area higher than 10000 km2 are considered. This is again not consistent with values provided by other figures.

This is true and was done with respect to the clearness of the figure. A comparison to surface areas higher than 1000 km^2 showed similar results.

28. P13L3-5: This is already written P10L35-P11L2. And this is commented in specific comment #24.

See #24

29. Title of Section 3.2: The difference between section 3.1 and section 3.2 are not understandable based on this title, and the reader may be unsettled at this point as I was. There should be something of a “between the levels of warming” somewhere. Please rephrase.

Agreed.

30. Figure 5. Cf. comment #21.

Agreed.

31. P15L15-16: The increase in winter low flows would not necessarily lead to a higher hydropower potential. It actually depends on the evolution of total precipitation. And the possible evolution of hydropower production would depend on the type of reservoir management, as well as management rules constrained by possible other water usages (sustaining summer low flows downstream, irrigation, recreation, etc.). Moreover, a decrease in low flows does not necessarily imply a decrease in overall water availability average over the year, and the water stress is conditional on the respective weight of water availability and water demand for a given time. So I would recommend adapting the statements according to the above comments.

Agreed.

32. P15L17-18: I however completely agree with the need of regional adaptation options. Except that adaptation strategies should be put in place now, without waiting for the 3K level to be reached or not.

Agreed. We will include a sentence on this.

33. P16L6, “the result is independent of the sign of change”: Well, this is a potentially serious issue. Indeed, how to interpret a situation where e.g. out of15 projections, 5 give a significant upward change, 5 other no significant change, and the last 5 a significant downward change? I would recommend interpreting this situation with particularly no robust signal! So please make clearer in the manuscript all the different possible cases and the way to interpret them. An alternative for presenting robustness would be the one used in the IPCC AR5 WGI report, i.e. the percentage of projections agreeing on the sign of the change.

Fully agree with first statement. This is the reason why we suggest to use a combination of SNR and robustness. E.g. using the IPCC AR5 WGI report would give an information similar to SNR, but without a significance information.

34. P16L11-13: I totally agree with this sentence, but it comes here out of the blue. Please consider moving it to the introduction, discussion, or conclusion.

The sentence will be added in the conclusions.

35. Figure 6. The choice of colour breaks is here particularly unfortunate here. For the SNR, I would appreciate having a break in value 1, in order to see where the median change is higher than the uncertainty in projections. For the ratio of GCM to HM uncertainty contribution, this is all the more important to see where this crosses the 1 value. An alternative would be to use bivariate colour scales (Teuling et al., 2011) to jointly plot the evolution of both sources of uncertainty.

We understand that it is naturally to expect color breaks at 1 and we also used these at a previous version of this Figure. The purpose of this section is to highlight the substantial uncertainties associated with the results. For this reason, we decided to use a range of plus/minus 20% around 1. to mark regions where the contribution by GCMs and HMs (GCM/HM contr.) are of the same order of magnitude (please note that 0.8 and 1.25 are the inverse of each other for multiplication). It is measleading to distinguish a value slightly higher than 1. from one slightly lower (e.g., 1.04 from 0.98). Given the uncertainty in the analysed dataset, we only consider a higher contribution by either GCMs or HMs if it is at least 20% higher than the other. Similar arguments hold for the signal to noise ratio. We added a sentence to clarify these points (see p. 15 line 10). Using a bivariate color scheme following Teuling et al. 2011 is a possible alternative for the presentation of Figure 6. This color scheme would show the values in absolute terms rather then relative to each other. It would be possible to distinguish high and low values, but it would be harder to see which of the two sources of uncertainty is higher. We also think that showing the signal to noise ratio already allows to identify regions with high and low uncertainty and, additionally, providing absolute values is not required. Proposed paragraph to include: "It is worth noting that we have chosen the color scheme in Figure 6 in a way that regions where the SNR is within 20% around 1. have the same colors. These regions have a signal, which is of similar magnitude as the uncertainty. Different colors are used to mark regions where the the signal is more than 20% higher or lower than the uncertainty."

36. P17L4-5: This exact sentence has already been written P8L3-4, and commented above (comment #19)

The sentence will be rephrased.

37. P17L8-P18L3: I am more or less OK with what is written here, but I do not understand why this would imply that the ratio of HM contribution to GCM contribution is higher at the 3K level. Please provide some explanations in the manuscript. Couldn't this be related to timing of threshold crossing in HM behavior that would differ from one HM to another, e.g. going from energy-limited to water-limited evaporation process?

Answered under #38

38. P18L4-20: This whole paragraph tends to support the above hypothesis. This should be related in the manuscript to recent uncertainty decomposition results obtained for a catchment located in the Southern Alps. It showed that the increasing spread of changes in future low flows by different HMs is linked to increasing spread in simulated evaporation and snow water equivalent (Vidal et al., 2016).

We agree with the reviewer that the explanation that we provided is only valid for the increase of the uncertainty of GCMs and that other factors such as the one mentioned by the reviewer influence the increase of HM uncertainty. We rephrased this paragraph to be more explicit about the different sources influencing the uncertainty contribution paragraph starting at p. 17, l. 7:

Total uncertainties in low flow projections is separated into GCM and HM contributions using the sequential sampling method proposed in Schewe et al. (2014). The results are shown in Fig. 6 (d-f) and spatially aggregated over the IPCC Europe regions in Tab. 4. The uncertainty rises with higher levels of warming for both sources of uncertainty because of two reasons. The GCM uncertainty increases because a 30 year period reaching a 3 K warming often has a strong temperature period within this 30 year period. Contrarily, GCM runs under RCP 2.6 often stabilise around 1.5 K global warming. This pathway dependency of warming influences the variability of the results with expectedly higher variability in the former case (James et al., 2017). The HM uncertainty increase with global warming because certain regions might crosse thresholds. For example, parts of France might move from a energy-limited to a water-limited regime. Overall, the contribution of the GCMs to the uncertainty over Europe is about 21% higher under 1.5 K, 25% higher 10 under 2 K and only 10% higher under 3 K global warming in comparison to the HM contribution. This decrease of GCM/HM contribution can be mostly attributed to the Mediterranean and Atlantic region (in particular France). In these dry regions, the different representations of evaporation using temperature-based potential evapotranspiration used in mHM and PCR-GLOBWB lead to different responses than explicitly solving the full energy-balance of the land surface as in Noah-MP.

39. P19L28-30, “We conclude. . . support the adaptation process.” Well, this is actually only a wish. Nothing in the paper allows asserting that, even I personally hope this is the case. So please rephrase.

Agreed.

3 Technical corrections

Technical corrections hereafter will be adressed according to the reviewers suggestions.

1. P1L5, “unprecedented”: it is a bit far-fetched, given that (1) GCM forcings are only disaggregated to this resolution without adding any downscaling information, and (2) results are seemingly partly given only for catchments >10000 km2 (P8L13- 15).
2. P1L6: “combination”
3. P2L2: “independently”?
4. P2L22-24: I believe that the sentence is not grammatically correct.
5. P2L30: “2”? in reference (UNFCC, 2015)
6. P2L34: “because of”
7. P3L3: please check missing or incorrect “the”
8. P3L8: “southern Europe”
9. P11L11: “extent”

10. P13L5: “political” -> “policy”. Also P15L23.
11. P15L22, “distinguished”: please rephrase.
12. P15L24, “ensemble members”: Please clarify what they are.
13. P20L1, “pronounced”: What is? Please rephrase.
14. P21L5-8: Wrong formatting, cf. IPCC report citation rules.
15. P23L34: line feed
16. P24L25-26: extra information to be removed

References

Mendoza, P. A., Clark, M. P., Mizukami, N., Newman, A. J., Barlage, M., Gutmann, E. D., Rasmussen, R. M., Rajagopalan, B., Brekke, L. D., and Arnold, J. R.: Effects of Hydrologic Model Choice and Calibration on the Portrayal of Climate Change Impacts, Journal of Hydrometeorology, 16, 762–780, doi:10.1175/JHM-D-14-0104.1, https://doi.org/10.1175/JHM-D-14-0104.1, 2015.

Samaniego, L., Kumar, R., Breuer, L., Chamorro, A., Flörke, M., Pechlivanidis, I. G., Schäfer, D., Shah, H., Vetter, T., Wortmann, M., and Zeng, X.: Propagation of forcing and model uncertainties on to hydrological drought characteristics in a multi-model century-long experiment in large river basins, Climatic Change, 141, 435–449, doi:10.1007/s10584-016-1778-y, http://dx.doi.org/10.1007/s10584-016-1778-y, 2017.

Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N. W., Clark, D. B., Dankers, R., Eisner, S., Fekete, B. M., Colón-González, F. J., Gosling, S. N., Kim, H., Liu, X., Masaki, Y., Portmann, F. T., Satoh, Y., Stacke, T., Tang, Q., Wada, Y., Wisser, D., Albrecht, T., Frieler, K., Piontek, F., Warszawski, L., and Kabat, P.: Multimodel assessment of water scarcity under climate change, Proceedings of the National Academy of Sciences, 111, 3245–3250, doi:10.1073/pnas.1222460110, http://www.pnas.org/content/111/9/3245.abstract, 2014.

Vaze, J., Post, D., Chiew, F., Perraud, J.-M., Viney, N., and Teng, J.: Climate non-stationarity – Validity of calibrated rainfall–runoff models for use in climate change studies, Journal of Hydrology, 394, 447 – 457, doi:https://doi.org/10.1016/j.jhydrol.2010.09.018, http://www.sciencedirect.com/science/article/pii/S0022169410005986, 2010.