CrossMark

Uncertainties in projected runoff over the conterminous United States

Ignazio Giuntoli [1,2,3] • Gabriele Villarini [3] • Christel Prudhomme [2] • David M. Hannah [1]

Abstract

Projections of runoff from global multi-model ensembles provide a valuable basis for the estimation of future hydrological extremes. However, projections suffer from uncertainty that originates from different error sources along the modeling chain. Hydrological impact studies have generally partitioned these error sources into global impact and global climate model (GIM and GCM, respectively) uncertainties, neglecting other sources, including scenarios and internal variability. Using a set of GIMs driven by GCMs under different representative concentration pathways (RCPs), this study aims to partition the uncertainty of future flows coming from GIMs, GCMs, RCPs, and internal variability over the CONterminous United States (CONUS). We focus on annual maximum, median, and minimum runoff, analyzed decadally over the twenty-first century. Results indicate that GCMs and GIMs are responsible for the largest fraction of uncertainty over most of the study area, followed by internal variability and to a smaller extent RCPs. To investigate the influence of the ensemble setup on uncertainty, in addition to the full ensemble, three ensemble configurations are studied using fewer GIMs (excluding least credible GIMs in runoff representation and GIMs accounting for vegetation and CO_2 dynamics), and excluding intermediate RCPs. Overall, the use of fewer GIMs has a minor impact on uncertainty for low and medium flows, but a substantial impact for high flows. Regardless of the number of pathways considered, RCPs always play a very small role, suggesting that improvement of GCMs and GIMs and more informed ensemble selections can yield a reduction of projected uncertainties.

✉ Ignazio Giuntoli
   i.giuntoli@bham.ac.uk

[1]   School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham B15 2TT, UK

[2]   Centre for Ecology and Hydrology, Wallingford, Oxfordshire OX10 8BB, UK

[3]   IIHR-Hydroscience & Engineering, University of Iowa, Iowa City, IA, USA

1 Introduction

A changing climate and an increasing world population make the knowledge of future hydrology ever more valuable to inform adaptation strategies that will have to deal with unprecedented pressures on food production and exposure to water-related hazards (e.g., Lavell et al. 2012). In this direction, runoff projections from multi-model ensemble experiments like WaterMIP (Haddeland et al. 2011) and ISI-MIP (Warszawski et al. 2014) are increasingly used in climate impact studies, but their utility is undermined by the large uncertainties that originate in the different components of the modeling chain. There is indeed a consensus on the growing need to well characterize uncertainty both to inform the selection/ design of multi-model ensembles and to improve the components of the modeling chain (e.g., Northrop 2013). Uncertainty in climate projections comes from three main sources: the models, the scenarios, and the internal variability (Hawkins and Sutton 2009; Tebaldi and Knutti 2007). Model uncertainty, or response uncertainty, results from different models yielding different responses to the same external forcing owing to differences in the physical and numerical formulations employed. Scenario uncertainty originates from the limited knowledge of the external factors that influence the climate system, for instance trajectories of greenhouse gases, land use change, and ozone concentrations in the stratosphere. Internal variability is the natural variability of the climate system without external forcing and is caused by non-linear dynamical processes intrinsic to the atmosphere, the ocean, and the coupled ocean-atmosphere system (e.g., Deser et al. 2012a).

Dominant sources of uncertainty in climate projections depend on the variable of interest. Precipitation projections are generally dominated by global climate model (GCM) uncertainty and internal variability rather than scenarios (e.g., Hawkins and Sutton 2011; Deser et al. 2012a; Pendergrass et al. 2015). For runoff projections, while GCMs play a large role, the global impact models (GIMs—which simulate the water cycle at the land surface) can outweigh the GCMs in the contribution to uncertainty, especially in those areas where storage-release processes (e.g., snow-ice) present a challenge (e.g., Hagemann et al. 2013; Giuntoli et al. 2015a). In particular, the choice of representative concentration pathways (RCPs) has a more systematic impact on temperature than on precipitation change. With the exception of a few studies at regional scales (Camici et al. 2017; Peleg et al. 2017; Vidal et al. 2016), little is known about the contribution to the uncertainty in runoff projections coming from internal variability and scenarios as well as the interplay with the other two sources (i.e., GCMs and GIMs).

If the uncertainty from model response (GCMs and GIMs) can potentially be reduced through improvement of the models, and if emission scenarios can be better constrained, internal variability is still unlikely to be reduced because of the inherently unpredictable nature of unforced climate fluctuations beyond a decade (Deser et al. 2012b). Because of internal variability, climate projections of different variables can be inherently uncertain in many parts of the world, with locations that are subject to large internal variability (e.g., the city of Seattle, USA, in Deser et al. (2012b)) that are bound to be affected by an irreducible share of uncertainty in climate projections. An important role in this is played by the GCMs, whose biases in representing the climate (e.g., annual cycles of temperature and precipitation, storm tracks, climate variability patterns) largely influence uncertainty (McSweeney et al. 2015) that is then cascaded to the GIMs.

Uncertainty is generally characterized by partitioning the variance of the ensemble spread into different components using statistical frameworks like analysis of variance (ANOVA; e.g., Yip et al. 2011; Sansom et al. 2013). When runs with different initial conditions are

unavailable, internal variability can be sampled as a measure of the noise in the projections throughout the runs as in the framework proposed by Hawkins and Sutton (2009) and used in Orlowsky and Seneviratne (2013), among others.

To better constrain future projections of runoff, it is important to quantify major sources of uncertainty not only at one extreme (e.g., flood or drought), but across the runoff spectrum. To this aim, this study examines the partitioning of uncertainty in annual maximum, medium, and minimum flows from the ISI-MIP runoff projections. This multi-model ensemble has been studied globally for low (Prudhomme et al. 2014; Schewe et al. 2014; Giuntoli et al. 2015a), medium (Davie et al. 2013), and high (Dankers et al. 2013; Giuntoli et al. 2015a; Dottori et al. 2018) flows, comparing runoff metrics from future and past periods. We focus on the CONterminous United States (CONUS) using a fractional change approach to assess how the uncertainty evolves transiently throughout the twenty-first century (e.g., Hawkins and Sutton 2011; Hingray and Saïd 2014). In particular, we analyze high, medium, and low flows jointly, showing how uncertainties differ across indices; moreover, in addition to GCMs and GIMs, we include the contribution of RCPs and internal variability.

We quantify uncertainty using both absolute and relative changes from the baseline period to understand if considerable differences in using either of the two exist. The use of relative changes is generally preferred for reducing the influence of historical biases in the GCMs and GIMs (e.g., meteorological forcing, simplifications in the model structure, and scale issues) that can propagate and amplify into the future (Sperna Weiland et al. 2012). However, when model runs are affected by zero-rich time series, absolute changes represent a more tractable approach.

We use the "one model one vote" approach without implementing any model weighting scheme (not suitable for the ISI-MIP Fast Track runs, which were not forced with observed data in the control period and would thus hinder robust estimates). With this approach, we also investigate whether the ensemble configuration influences the partition of uncertainty by excluding models—or "culling" as discussed in, e.g., Overland et al. (2011) and Thibeault and Seth (2014)—on the basis of credibility in medium/low flow representation. Furthermore, we analyze uncertainty excluding the so-called biome models (GIMs that include CO_2 and vegetation dynamics) and then excluding intermediate RCPs (4.5 and 6.0). This is done to assess to what extent the uncertainty share changes using all available or fewer GIMs and RCPs with the aim to suggest a better use of resources making model runs and an improved choice of ensembles in hydrological impact studies.

2 Data and methods

2.1 Simulated runoff

We use simulations of daily unrouted runoff from the ISI-MIP Fast Track comprising nine GIMs driven by five bias-corrected CMIP5 (fifth Coupled Model Inter-comparison Project; Taylor et al. 2012) GCMs in their control (1971–2005) and future (2006–2099) period, under four RCP scenarios (RCPs 2.6, 4.5, 6.0, and 8.5).

The GCM outputs (i.e., the climate forcing used as input for the GIMs) have a spatial resolution ranging from $1.875° \times 1.25°$ for HadGEM2-ES to $2.8° \times 2.8°$ grid for MIROC-ESM-CHEM (Maloney et al. 2014) and have been bi-linearly interpolated to a common $0.5° \times 0.5°$ grid and bias-corrected towards the observation-based WATCH Forcing Data (WFD) as detailed in Hempel et al. (2013). These GCMs have been evaluated by McSweeney and Jones

(2016) and belong to different clusters in the model family tree of Knutti et al. (2013), indicating good independence among models.

The daily runoff output comes from nine GIMs with a spatial resolution of $0.5° \times 0.5°$ (except JULES, whose runs were regridded to $0.5°$ from $1.25° \times 1.875°$). GIMs can be broadly grouped as hydrological (H08, MPI-HM, PCRGlobWB, WBM, MacPDM, VIC, MATSIRO) and biome (LPJmL and JULES; they are also hydrological models, but include effects of vegetation and CO_2 dynamics on runoff) models. Please refer to Table S.1 and Section S.1 in the SI for further details on the GIMs, and to Section S.2 for details on their evaluation.

Every GIM was run with every GCM under four RCPs. We use all the simulations available to maximize the sample for a total of 156 runs (MacPDM, VIC, and MATSIRO lack RCPs 4.5 and 6.0 runs except when forced by the HadGEM2-ES GCM).

2.2 Hydrological indices

We consider three hydrological indices describing high-, median-, and low-flow regime for the period 2006–2099 over the CONUS (approximately 5700 land cells)—the area bounded in longitude by $-135°$ and $-60°$ E, and in latitude by $54°$ and $24°$ N. To this aim, over the period 1971–2099, we extracted from the daily runoff runs: (i) the Annual Maximum (AMax), (ii) the Annual Median (AMed), and (iii) the Annual Mininum (AMin) daily runoff values. We use these indices decadally, as for temperature and precipitation in Hawkins and Sutton (2009) and Pendergrass et al. (2015), thereby reducing the noise in the signal of the ensemble spread and the contribution of internal variability to uncertainty. Internal variability would otherwise be a dominant component at the annual scale, with its contribution decreasing for increasing spatial and temporal aggregation scales. Hence, we run a 10-year moving average on the annual indices obtaining decadal AMax, AMed, and AMin series for the period 1975–2095.

2.3 Partition of uncertainty

We quantify four sources of uncertainty: GCMs, GIMs, RCPs, and internal variability (henceforth called IVar), following the statistical framework proposed by Hawkins and Sutton (2009). The uncertainty is expressed by the variance of multi-model anomalies averaged for each source (except for IVar). We used the same partition approach in two variants: (i) using raw anomalies—henceforth referred to as absolute change or AC; (ii) using relative anomalies—henceforth referred to as percentage change or PC. The two variants are detailed in Section S.3 of the SI.

In the application of the partition via PC, a limitation exists when zero or very small values in the reference period mean are encountered. These small values can be common for AMin and AMed, and are therefore screened out of the analysis to prevent them from being biased towards large uncertainty improperly (as in, e.g., Baker and Huang 2012). Masks vary with the index considered (please refer to Section S.4 Table S2 in the SI for details on the masking procedure).

We express transient uncertainty over the entire domain on a gridcell basis, as well as over regions characterized by aggregating gridcells into nine climate homogeneous areas (http://www.ncdc.noaa.gov/monitoring-references/maps/us-climate-regions.php).

For a selected gridcell in the northeast, Fig. 1 shows the overall spread of the transient runs from 2006 to 2095 and the corresponding fractional variance (uncertainty) for both AC and PC partitions. For consistency, we show the fractional uncertainty excluding the two GIMs with
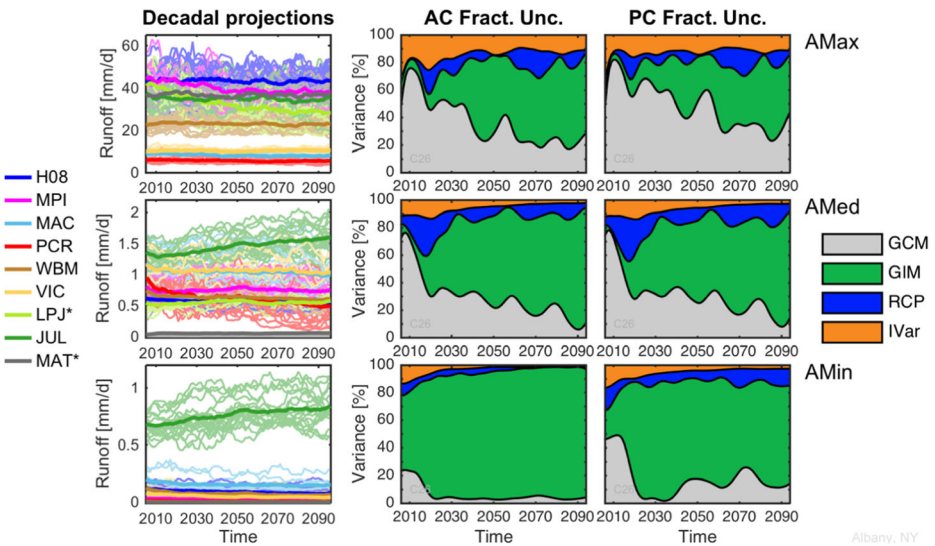
Fig. 1 Decadally averaged AMax (top panels), AMed (mid panels), and AMin (lower panels) projections (left) colored according to GIM (thick lines are averages per GIM) and corresponding fractional uncertainty using absolute change (second column) and percentage change (third column), for a selected gridcell (42.7° N–73.9° E; Albany, NY). The two low credibility GIMs (*) are displayed in the first column but are not considered in the fractional uncertainty

very small values (LPJmL and MATSIRO) as it will be discussed in the next section. AMax shows a large share of uncertainty from GCMs at the beginning of the run that decreases rapidly in the first decade and continues to decrease until the end of the run in favor of the GIMs; on the other hand for AMin, the GIM share rises quickly and explains (around 2020) virtually all of the variance in the projections. Similar to AMax, AMed depicts a declining GCM contribution, although the RCP uncertainty share is generally larger. For these three indices all throughout the run, RCP and the IVar keep a constant and marginal contribution to total variance relative to the GIM and GCM contributions. The differences in the magnitude of the projections across GIMs can be considerable and this explains how they dominate future uncertainty. AC and PC shares are very similar in all three indices.

2.4 Ensemble configurations

The uncertainty partition was carried out over the entire ensemble and over selected subsets obtained by excluding specific GIMs and intermediate RCPs (Table S3) for a total of four configurations, each including five GCMs:

1. *oE*, ensemble of opportunity (all runs available)—156 runs (9 GIMs, 4 RCPs)
2. *cE*, credibility ensemble (without low credibility GIMs)—124 runs (7 GIMs, 4 RCPs)
3. *cE-noB*, *cE* without biome GIMs—104 runs (6 GIMs, 4 RCPs)
4. *cE-noIR*, *cE* without the intermediate RCP 4.5 and 6.0–70 runs (7 GIMs, 2 RCPs)

First, we examine the uncertainty in the general case of using all of the runs available, *oE* (Table S3—first column), i.e., the general case that allows for sampling the largest spread provided by the ISI-MIP dataset. Secondly, we analyze the credibility ensemble, *cE*, having excluded GIMs

that show limitations simulating realistic low and medium runoff. Indeed, two GIMs have shown a disproportionate number of days with null runoff (Giuntoli et al. 2015a, b), notably in the western USA (Fig. S3; in Fig. S4 results are shown for Europe for comparison). We define low credibility GIMs as those with more than 40% of land gridcells in the domain of study with more than 40% of the days in the year with null runoff on average, calculated across the five GCMs in the control period 1971–2005. Consequently, MATSIRO and LPJmL are identified as less credible, with 50 and 53% of land gridcells with more than 40% with null runoff days, respectively. JULES is retained with 27% of land with more than 40% of null runoff days on average, with gridcells located mostly in the arid US southwest. Thirdly, we analyze the effect of excluding the biome GIMs, *cE-noB*; and lastly, we examine the effect of excluding intermediate RCP runs, *cE-noIR*.

3 Results

We present the partition of uncertainty into four sources for decadally averaged transient runs of AMax, AMed, and AMin for the different ensemble setups. We map the relative contributions to the total uncertainty over the entire domain at five time slices from 2010 to 2090 showing results for AC and PC (in SI).

3.1 *oE*

Throughout the twenty-first century and using the full ensemble, the uncertainty in AMax (Fig. 2a) is explained mostly by the GCMs and GIMs, with the latter increasing their contribution to the total uncertainty from the west of the domain spreading to the northwest (around 2030) and to the upper Midwest and northeastern USA; this leads to a reduced importance of the GCMs over the twenty-first century, especially for the northern half of the domain. The share of IVar is large at the beginning of the century over the domain but declines steadily with the exception of the south and southeast, with a contribution to the total variance
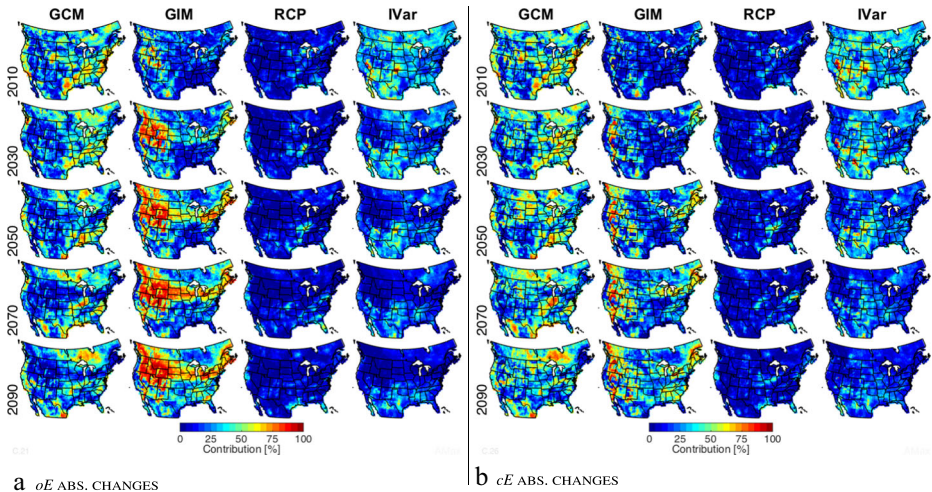

Fig. 2 Relative contribution (on 100%) to uncertainty in Amax: on the left **a** the full ensemble (*oE*); on the right **b** the credibilty ensemble (*cE*). The layout: in column, from left to right the uncertainty sources: GCMs, GIMs, RCPs, and internal variability; in row, from top to bottom the time slices: 2010, 2030, 2050, and 2070. These results refer to the absolute changes

(15–30%) comparable to that of the GIMs. In contrast and as seen in Fig. 1, the contribution to the total uncertainty from RCPs is very limited compared to GIMs and GCMs, with shares ranging from 5 to 20% of the total variance. In particular, both RCPs and IVar show the smallest contributions in the northern part of the domain (see mean relative contributions per region in Fig. S5a).

AMed (Fig. 3a) shows results similar to AMax: GCMs and GIMs are by far the major contributors, with GIMs increasing their contribution to uncertainty over the domain during the twenty-first century. This is particularly true in the northern and eastern parts of the CONUS, even though shares are more balanced between the two sources, without the clear predominance of GIM contribution as seen, for instance, in the western USA for AMax. Also, GCMs show larger contributions to the total variance both at the beginning of the run and, in general, in the western and northwestern USA compared to AMax (Fig. S6). Interestingly, in the US west and southwest, IVar has high shares from as much as 40% in 2006 to 15–20% in 2095. This is due to the large fluctuations in the projections that make up large deviations from the smooth fit, thereby making the contribution of IVar to uncertainty substantial (as seen in Las Vegas in Fig. S2).

For AMin (Fig. 4a), the similar contribution to the uncertainty by GCMs and GIMs seen for AMax and AMed only holds at the beginning of the twenty-first century, in which IVar, especially through the Rocky Mountains, has a substantial contribution to uncertainty. GIMs increase steadily their share throughout the period. In particular, the GCMs lose their share to the advantage of GIMs in the east, while in the west, IVar shows considerable percentages of about 40%, which slowly reduce to 15–20% at the end of the run. The strong role that IVar plays in the southwest (in the area spanning from the lower Rocky Mountains to the arid areas of the south) can partly be attributed to the difficulty of both GCMs and GIMs in simulating runoff in mountaineous and arid areas, where indeed AMin time series generally suffer from poor simulations with anomalous erratic departures from zero or very low values. With the exception of the US southeast around 2030–2040 (~20%), RCPs contribute only marginally to the total variance (Fig. S7).
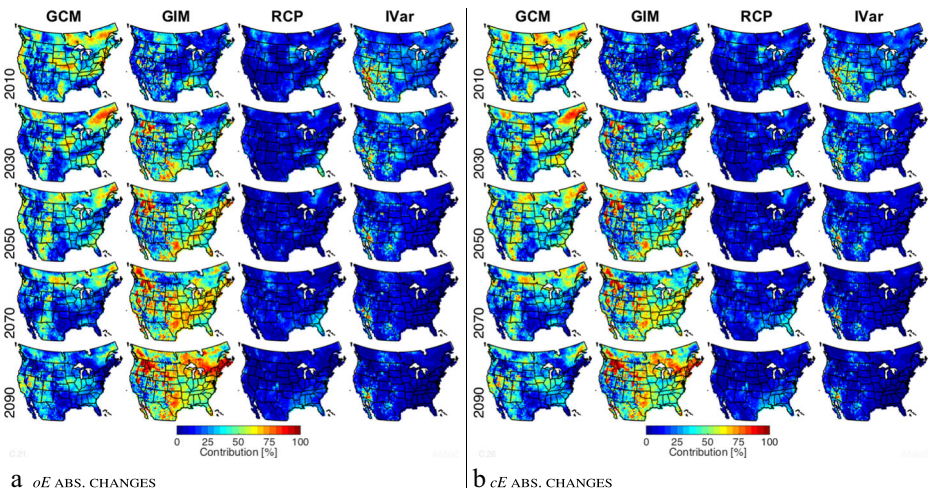
Fig. 3 Same as Fig. 2 but for AMed. **a** *oE*. **b** *cE*. These results refer to the absolute changes

a *oE* ABS. CHANGES   b *cE* ABS. CHANGES
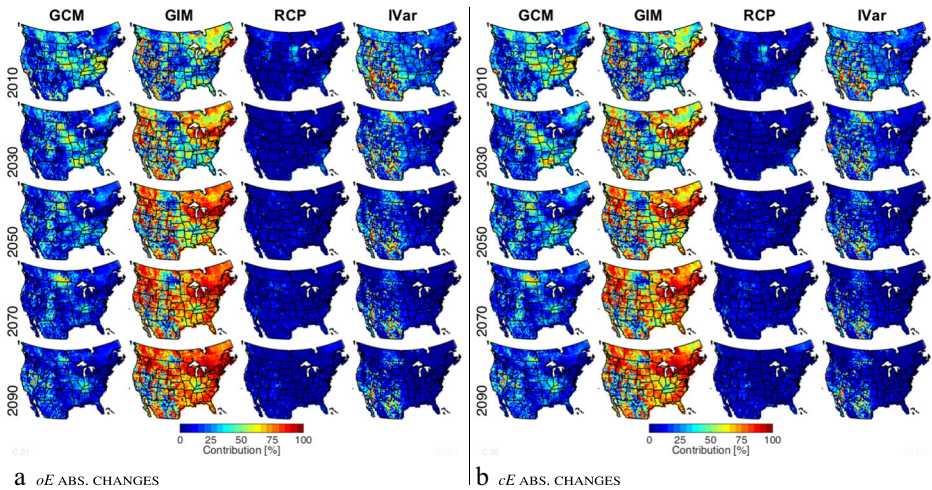
Fig. 4 Same as Fig. 2 but for AMin. **a** *oE*. **b** *cE*. These results refer to the absolute changes

Results with PC in AMax (Fig. S9a) are very similar to those obtained with AC, although the western part of the domain is partly masked. There are slight increases (5 to 15%) in GCM and IVar uncertainty shares that are taken from the GIM, except for the eastern part of the domain (Fig. S5b). The heavy masking occurring with the use of the *oE* makes AMed and AMin not comparable.

3.2 *cE*

The exclusion of two of the GIMs (MATSIRO and LPJmL) from the full ensemble has an effect on both relative contributions and total variance. For AMax in particular, patterns are markedly different: the GIM source ceases to dominate over most of the domain (Fig. 2b), and the strong contribution in the northern half is limited to the west and northeast of the domain in favor of the uncertainties in the GCMs. Compared to the *oE*, the lower variance of the *cE* is reflected in lower total variance in all regions, particularly the northern ones (Fig. S8a). When we focus on AMed (Fig. 3b), GCM and RCP contributions grow from 5 to 10% over the entire domain, with a reduction in GIM contribution. There are large fractions of GCM uncertainty at the beginning of the run over most of the domain that decline in favor of the GIM uncertainty, which explains most of the uncertainty at the end of the run. Similar to AMed, AMin (Fig. 4b) has close shares of fractional variance to the *oE* (Fig. 4a) all throughout the run.

The total variance for *cE* is close to and only slightly higher than that of *oE* (Fig. S8a), suggesting that MATSIRO and LPJmL projections lie within the range of the other GIMs and that the exclusion of less credible GIMs does not yield lower total variances.

Results for PC in all three indices (Fig. S9b-S14b) depict the same patterns seen for AC. For AMed, the increase in RCP contribution over time in the southeast is even clearer with PC.

3.3 *cE-noB*

If, in addition to MATSIRO and LPJmL, we exclude JULES, the resulting ensemble (*cE-noB*) depicts marked differences especially for AMax. Indeed, LPJmL and JULES account for

vegetation dynamics and varying CO_2 (absent in the other GIMs) and tend to project higher runoff depending on the region.

Consistent with what discussed for *cE*, when we focus on AMax, the *cE-noB* (Fig. S15a) depicts even lower fractions of GIM uncertainty, in favor of the GCM and IVar. This is especially true in the southern parts of the study region. Both *cE* and *cE-noB* have in common the exclusion of LPJmL, which is responsible for contributing the high variance in the northern half of the domain. Substantial changes in the proportion of the total variance result from the exclusion of the three GIMs: the high GIM uncertainty in the northwest and north seen with the *oE* is no longer present, and GCMs become the dominant source over the entire domain during the whole period. Except for the south and southeast where the GIM contribution was already small (~ 15–30%), GCM and IVar increase their shares. This shows that biome GIMs bring about greater variance to the ensemble for the peak flows, with the exception of the southern/southeastern CONUS (Fig. S8a). It is worth noting that, over the regions covering the southern half of the USA, the RCP contribution to uncertainty is similar to that of GIMs (~ 10–20%) and sometimes greater, reaching 30% in the south and southeast. This suggests that low credibility and biome GIMs can be considered as outliers in the GIMs' spread; therefore, their exclusion facilitates an emerging signal of the RCP in these regions.

Changes in proportions from *oE* to *cE* and *cE-noB* for AMax are not as pronounced as those obtained for AMed (Fig. S16) and AMin (Fig. S17) for which we see a clear shift from the GIMs' to the GCMs' contribution. Namely, for AMed, there is a marked gain in GCM contribution, especially in the northern regions.

The influence of biome GIMs on AMin's total variance (Fig. S8) is fairly weak from the west to the south of the study region, where these models tend to simulate zero runoff over extended periods of time, leading to lower variance when included in the *oE*. Elsewhere (Fig. S17), GCMs increase their share by about 10%. As seen for AMed, RCP's contribution increases (especially in the eastern part of the domain) to the same level of the GCMs; this holds even though GIMs still remain the dominant source of uncertainty. The inclusion of biome models in the ensemble brings about larger shares of GIM uncertainty. In particular, when JULES is present, we see a larger and increasing GIM share in the southern half of the domain. Results for PC in AMax (Fig. S15b), AMed (Fig. S16b), and AMin (Fig. S17b) are in line with those seen for AC.

3.4 *cE-noIR*

We also tested the impacts of excluding the intermediate RCPs (4.5 and 6.0) on the model spread. For AMax (Fig. S18a—versus all RCPs in Fig. 2b), differences are barely noticeable, while for AMed (Fig. S19a—versus all RCPs in Fig. 3b), the RCP fraction of total variance has higher values in the eastern USA when compared to the *cE*, especially in the south and southeast USA after 2050. Similarly, for AMin (Fig. S20a—versus all RCPs in Fig. 4b), the RCP proportions increase at the end of the run, especially in the east. In essence, all indices show proportions similar to the ensemble with four RCPs, although these proportion tend to systematically fluctuate throughout the period, perhaps owing to the use of fewer projections (i.e., decreased information). Overall, the use of fewer RCPs leads to the same results for PC and AC for all indices (Fig. S18b-S20b).

4 Discussion

GCMs and GIMs are responsible for most of the uncertainty in runoff projections throughout the twenty-first century over the CONUS. This is consistent with Wada et al. (2013), who analyzed projections of irrigated water demand.

If we consider the physical processes pertaining the study region, Villarini (2016) gives an overview of the meteorological patterns that influence flood seasonality in the USA: the North American monsoon and the North Pacific tropical cyclones in Arizona and New Mexico; snowmelt in the north-central USA; thunderstorms and mesoscale convective systems towards the Gulf of Mexico; North Atlantic tropical cyclones in Florida; extratropical cyclones and atmospheric rivers in the central and western USA; snowmelt, rain on snow/ice in the northeastern USA; tropical and extratropical systems along the US East Coast. Berghuijs et al. (2016) report that the variability of annual flows is much larger for the more arid catchments of the central USA, and that snow controls the flood response in the Rockies and in the northern states. Moreover, evaporation-controlled soil moisture plays a dominant role over most of the USA.

We can summarize the main contributing factors affecting the uncertainty in snow-melt, storage-release processes, and soil moisture accounting for the GIMs, potentially affecting the AMin and AMed; precipitation variability, intensity, and spatial/temporal distribution for the GCMs, potentially affecting AMax (for its sudden onset and short duration); uncertainty in CO_2 emissions and in the way a warmer world changes the climate for RCPs; and the natural variability in the climate and the hydrology for IVar.

Large differences in parameterization and model structure exist across GIMs. Notably, GIMs accounting for vegetation and CO_2 dynamics tend to produce increased (reduced) runoff in the east (southwest) of the USA, while two GIMs have shown unrealistic runoff representation over the west-southwest USA. Given the variety of GCMs and GIMs forming the ensemble of projections, we have limited scope to reconcile how the main physical processes play out in the uncertainty contributions and their evolution in time. Clearly, orography represents a challenge for both GCMs and GIMs, and this may explain the higher fractional contribution of internal variability over the Rocky Mountains. Another difficult setting for both GIMs and GCMs is represented by arid regions (e.g., US southwest) as reported by Vetter et al. (2015), Wanders et al. (2015), Ward et al. (2013), and Huang et al. (2017), among others. GIMs have also difficulties in the cold regions (e.g., US Midwest or the northern Great Plains), where snow and ice dynamics play a role in the lag at which runoff is generated from melting, and generally in those areas where storage-release processes have a strong influence on runoff (Trigg et al. 2016; Giuntoli et al. 2015b; Hagemann et al. 2013).

In addition to the structural model differences of GIMs and GCMs, more uncertainty lies in the grid transformation from the coarser resolution of the GCMs to the finer resolution of the GIMs, and in the bias correction of the climate variables. Even though the bias correction may affect our results by reducing inter-GCM and RCP variability (Hempel et al. 2013), we are unable to quantify these impacts because of the lack of non-bias-corrected runs in our dataset. Hempel et al. (2013) also note that, while trends and long-term means are well represented, the ISI-MIP bias correction approach can present limitations with regard to the adjustment of variability; this is because it corrects daily data with respect to the monthly mean, neglecting the variability at other time scales (e.g., weekly and monthly variability of precipitation), which can affect drought and flood simulation.

We have tested to what extent the uncertainty contributions change using different ensemble setups to understand the effect of including all runs or excluding those deemed less credible. Appraisal of model credibility is arbitrary and much depends on the variable of interest and the metric used. Among the models we considered, McSweeney and Jones (2016) found that the five GCMs available represent CMIP5 GCMs reasonably well, although they argue that twice as many CMIP5 GCMs would be necessary to adequately represent the range

of temperature and precipitation changes globally. This suggests that the GCM uncertainty could be larger if more GCMs were available.

While there has been a consensus on using all GCMs, the other studies using this dataset have occasionally excluded some of the nine GIMs from the ISI-MIP ensemble. This choice was generally dictated by tractability of the data with respect to the metric used rather than by model selection based on evaluation or credibility criteria. For instance, Dankers et al. (2013) employed all nine GIMs when working on high flows; on the other hand, Prudhomme et al. (2014) excluded LPJmL and MATSIRO whose runs would poorly adapt to the threshold level method used to quantify drought changes. Giuntoli et al. (2015a) excluded LPJmL, MATSIRO, and JULES for the joint assessment of high and low flows, with the purpose of keeping a consistent GIM set for both flow quantiles. One merit of our study was to reveal the difference it makes to use both reduced ensemble setups and the whole ensemble.

This study clearly shows that the contribution of RCPs to uncertainty is small compared to the other sources, consistent with previous studies (Arnell and Gosling 2013; Orlowsky and Seneviratne 2013; Tang and Lettenmaier 2012; Wada et al. 2013). However, we have shown how this weak contribution holds for runoff indices going from minimum to maximum runoff, regardless of time and of whether less credible and biome GIMs are excluded from the analysis.

Finally, lacking run replicates, we assess uncertainty from internal variability on the basis of the distance of the departures from to the smooth fit and we assume it to be constant throughout the period of analysis. However, were run replicates available, and therefore, e.g., a four-way ANOVA feasable, we could expect results to change quantitatively, but we have no reason to expect them to change qualitatively. It should be noted that efforts in overcoming the assumption of independence among uncertainty sources can be found in the literature (e.g., Fatichi et al. 2016; Hingray and Saïd 2014).

We quantified uncertainty using both absolute and percentage changes finding similar results (similar to Good et al. 2016 and Hawkins and Sutton 2011 for precipitation), although for PC, the presence of zero/near-zero values that produce unduly very large values of variance have caused the filtering out of parts of the domain. Large variance can also arise from the presence of isolated peaks in the time series, a common caveat for AC and PC, which leads to the improper attribution of the uncertainty to the internal variability, as it may happen in snow-ice-dominated areas of the central USA.

Moreover, in the interpretation of the results, one should bear in mind that fewer models from the reduced ensembles (*cE*s) do not necessarily lead to a smaller spread and to a lower uncertainty (i.e., uncertainty can increase as a result of computing the variance on the same range of values but with fewer projections). For this reason the *oE*, which includes GIMs with lower credibility in medium and low flows, has a lower degree of uncertainty than the *cE* as it includes projections that have low variance (for being near zero most of the time), producing the apparent beneficial effect of reducing the uncertainty, while it is just an artifact.

5 Conclusions

The contributions of GCMs, GIMs, RCPs, and internal variability to the uncertainties in projected runoff vary regionally and through time depending on the runoff index over the CONUS throughout the twenty-first century.

Regardless of the runoff index, GCMs and GIMs account for the largest uncertainty in decadal runoff projections, followed by IVar and ultimately by RCPs, whose contribution is the smallest, except in the southeast USA for AMed and AMin. In particular, uncertainty in AMax is increasingly

dominated by GIMs in the western and northeastern CONUS, while in the east, the GCMs' contributions decrease their share over the twenty-first century. Although regional patterns differ, similar to AMax, the uncertainty in AMed sees an increase in GIM uncertainty with a concurrent decrease of GCMs' contribution over the twenty-first century. For AMin, uncertainty is mainly governed by GIMs, the predominant source over the CONUS, in particular towards the end of the twenty-first century.

Analysis of subsets of the ensemble (considering fewer GIMs) has delivered essentially the same results in the case of AMin and AMed in comparison to the full ensemble. Conversely, GIMs' uncertainty is markedly lower when fewer projections are used for AMax, in particular when biome models are excluded. This is due to an attenuation of the higher peaks introduced by the biome GIMs (LPJmL and JULES), which simulate increased runoff (thus variance), especially in the northeastern USA.

As hinted by previous studies at the regional (Gudmundsson et al. 2012; Giuntoli et al. 2015b) and global (Prudhomme et al. 2014; Giuntoli et al. 2015a; Zaherpour et al. 2018) scales, some GIMs have difficulties in simulating runoff, especially low flows. Over our study area, this difficulty is most evident in the US south-southwest, and it explains the higher fractions of variance explained by IVar.

Finally, we have tested the influence of excluding two intermediate RCPs, as it can occur when modeling centers contribute to fewer scenarios to reduce the computational burden. We found that the use of the higher (RCP 8.5) and lower (RCP 2.6) scenarios returns essentially the same results of using all RCPs. Overall, the variety of GIMs used makes it difficult to identify key drivers responsible for how the uncertainty is partitioned and evolves over time across these models.

To conclude, the quest for reducing uncertainty by choosing impact models that are deemed best or more credible is of little relevance for medium and low flows over the CONUS, for which uncertainty remains largely the same over time regardless of the ensemble selected. Overall, promising margins for reducing uncertainties will come from improved constraints on the GCMs and GIMs, rather than scenarios. Moreover, the limited impact of internal variability indicates that the majority of the uncertainties are indeed predictable, and not chaotic and irreducible. Our results underline the importance of improving global models (GCMs and GIMs) via continual efforts in model evaluations at the regional or continental scales and adopting common frameworks and metrics for facilitating comparisons between ensemble products and uncertainty quantifications.

References

Arnell NW, Gosling SN (2013) The impacts of climate change on river flow regimes at the global scale. J Hydrol 486:351–364. https://doi.org/10.1016/j.jhydrol.2013.02.010
Baker NC, Huang H (2012) A comparison of absolute and relative changes in precipitation in multimodel climate projection. Atmos Sci Lett 13:174–179. https://doi.org/10.1002/asl.388

Berghuijs WR, Woods RA, Hutton CJ, Sivapalan M (2016) Dominant flood generating mechanisms across the United States. Geophys Res Lett:1–9. https://doi.org/10.1002/2016GL068070.

Camici S, Brocca L, Moramarco T (2017) Accuracy versus variability of climate projections for flood assessment in central Italy. Clim Chang 141:273–286. https://doi.org/10.1007/s10584-016-1876-x

Dankers R et al (2014) First look at changes in flood hazard in the Inter-Sectoral Impact Model Intercomparison Project ensemble. Proc Natl Acad Sci USA 111:3257–3261. https://doi.org/10.1073/pnas.1302078110

Davie JCS et al (2013) Comparing projections of future changes in runoff from hydrological and biome models in ISI-MIP. Earth Syst Dyn 4:359–374. https://doi.org/10.5194/esd-4-359-2013

Deser C, Phillips A, Bourdette V, Teng H (2012a) Uncertainty in climate change projections: the role of internal variability. Clim Dyn 38:527–546. https://doi.org/10.1007/s00382-010-0977-x

Deser C, Knutti R, Solomon S, Phillips AS (2012b) Communication of the role of natural variability in future North American climate. Nat Clim Chang 2:775–779. https://doi.org/10.1038/nclimate1562

Dottori F et al (2018) Increased human and economic losses from river flooding with anthropogenic warming. Nat Clim Chang 20:9039. https://doi.org/10.1038/s41558-018-0257-z

Fatichi S et al (2016) Uncertainty partition challenges the predictability of vital details of climate change. Earth Future 4:240–251. https://doi.org/10.1002/2015EF000336

Giuntoli I, Vidal J-P, Prudhomme C, Hannah DM (2015a) Future hydrological extremes: the uncertainty from multiple global climate and global hydrological models. Earth Syst Dyn 6:267–285. https://doi.org/10.5194/esd-6-267-2015

Giuntoli I, Villarini G, Prudhomme C, Mallakpour I, Hannah DM (2015b) Evaluation of global impact models' ability to reproduce runoff characteristics over the Central United States. J Geophys Res Atmos 120:9138–9159. https://doi.org/10.1002/2015JD023401

Good P, Booth BBB, Chadwick R, Hawkins E, Jonko A, Lowe JA (2016) Between a first and second 2 K of global warming. Nat Commun 7:1–8. https://doi.org/10.1038/ncomms13667

Gudmundsson L et al (2012) Comparing large-scale hydrological model simulations to observed runoff percentiles in Europe. J Hydrometeorol 13:604–620. https://doi.org/10.1175/JHM-D-11-083.1

Haddeland I et al (2011) Multimodel estimate of the global terrestrial water balance: setup and first results. J Hydrometeorol 12:869–884. https://doi.org/10.1175/2011JHM1324.1

Hagemann S et al (2013) Climate change impact on available water resources obtained using multiple global climate and hydrology models. Earth Syst Dyn 4:129–144. https://doi.org/10.5194/esd-4-129-2013

Hawkins E, Sutton R (2009) The potential to narrow uncertainty in regional climate predictions. Bull Am Meteorol Soc 90:1095–1107. https://doi.org/10.1175/2009BAMS2607.1

Hawkins E, Sutton R (2011) The potential to narrow uncertainty in projections of regional precipitation change. Clim Dyn 37:407–418. https://doi.org/10.1007/s00382-010-0810-6

Hempel S, Frieler K, Warszawski L, Schewe J, Piontek F (2013) A trend-preserving bias correction—the ISI-MIP approach. Earth Syst Dyn 4:219–236. https://doi.org/10.5194/esd-4-219-2013

Hingray B, Saïd M (2014) Partitioning internal variability and model uncertainty components in a multimember multimodel ensemble of climate projections. J Clim 27:6779–6798. https://doi.org/10.1175/JCLI-D-13-00629.1

Huang S et al (2017) Evaluation of an ensemble of regional hydrological models in 12 large-scale river basins worldwide. Clim Chang 141:399–400. https://doi.org/10.1007/s10584-016-1895-7

Knutti R, Masson D, Gettelman A (2013) Climate model genealogy: generation CMIP5 and how we got there. Geophys Res Lett 40:1194–1199. https://doi.org/10.1002/grl.50256

Lavell A, Oppenheimer M, Diop C, Hess J, Lempert R, Li J, Muir-Wood R, Myeong S (2012) Climate change: new dimensions in disaster risk, exposure, vulnerability, and resilience. In: Field CB, Barros V, Stocker TF, Qin D, Dokken DJ, Ebi KL, Mastrandrea MD, Mach KJ, Plattner G-K, Allen SK, Tignor M, Midgley PM (eds) Managing the risks of extreme events and disasters to advance climate change adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC). Cambridge University Press, Cambridge and New York, pp 25–64

Maloney ED et al (2014) North American climate in CMIP5 experiments: part III: assessment of twenty-first-century projections*. J Clim 27:2230–2270. https://doi.org/10.1175/JCLI-D-13-00273.1

McSweeney CF, Jones RG (2016) How representative is the spread of climate projections from the 5 CMIP5 GCMs used in ISI-MIP? Clim Serv 1:24–29. https://doi.org/10.1016/j.cliser.2016.02.001

McSweeney CF, Jones RG, Lee RW, Rowell DP (2015) Selecting CMIP5 GCMs for downscaling over multiple regions. Clim Dyn 44:3237–3260. https://doi.org/10.1007/s00382-014-2418-8

Northrop PJ (2013) Comments on "A simple, coherent framework for partitioning uncertainty in climate predictions". J Clim 26:4375–4376. https://doi.org/10.1175/JCLI-D-12-00527.1

Orlowsky B, Seneviratne SI (2013) Elusive drought: uncertainty in observed trends and short- and long-term CMIP5 projections. Hydrol Earth Syst Sci 17:1765–1781. https://doi.org/10.5194/hess-17-1765-2013

Overland JE, Wang M, Bond NA, Walsh JE, Kattsov VM, Chapman WL (2011) Considerations in the selection of global climate models for regional climate projections: the Arctic as a case study*. J Clim 24:1583–1597. https://doi.org/10.1175/2010JCLI3462.1

Peleg N, Blumensaat F, Molnar P, Fatichi S, Burlando P (2017) Partitioning the impacts of spatial and climatological rainfall variability in urban drainage modeling. Hydrol Earth Syst Sci 21:1559–1572. https://doi.org/10.5194/hess-21-1559-2017.

Pendergrass AG, Lehner F, Sanderson BM, Xu Y (2015) Does extreme precipitation intensity depend on the emissions scenario? Geophys Res Lett 42:8767–8774. https://doi.org/10.1002/2015GL065854

Prudhomme C et al (2014) Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment. Proc Natl Acad Sci U S A 111:3262–3267. https://doi.org/10.1073/pnas.1222473110

Sansom PG, Stephenson DB, Ferro CAT, Zappa G, Shaffrey L (2013) Simple uncertainty frameworks for selecting weighting schemes and interpreting multimodel ensemble climate change experiments. J Clim 26: 4017–4037. https://doi.org/10.1175/JCLI-D-12-00462.1.

Schewe J et al (2014) Multimodel assessment of water scarcity under climate change. Proc Natl Acad Sci 111: 3245–3250. https://doi.org/10.1073/pnas.1222460110

Sperna Weiland FC, van Beek LPH, Kwadijk JCJ, Bierkens MFP (2012) Global patterns of change in discharge regimes for 2100. Hydrol Earth Syst Sci 16:1047–1062. https://doi.org/10.5194/hess-16-1047-2012

Tang Q, Lettenmaier DP (2012) 21st century runoff sensitivities of major global river basins. Geophys Res Lett 39. https://doi.org/10.1029/2011GL050834

Taylor IH, Burke E, McColl L, Falloon P, Harris GR, McNeall D (2012) Contributions to uncertainty in projections of future drought under climate change scenarios. Hydrol Earth Syst Sci Discuss 9:12613–12653. https://doi.org/10.5194/hessd-9-12613-2012

Tebaldi C, Knutti R (2007) The use of the multi-model ensemble in probabilistic climate projections. Philos Trans A Math Phys Eng Sci 365:2053–2075. https://doi.org/10.1098/rsta.2007.2076

Thibeault JM, Seth A (2014) A framework for evaluating model credibility for warm-season precipitation in northeastern North America: a case study of CMIP5 simulations and projections. J Clim 27:493–510. https://doi.org/10.1175/JCLI-D-12-00846.1

Trigg MA et al (2016) The credibility challenge for global fluvial flood risk analysis. Environ Res Lett 11: 094014. https://doi.org/10.1088/1748-9326/11/9/094014

Vetter T, Huang S, Aich V, Yang T, Wang X, Krysanova V, Hattermann F (2015) Multi-model climate impact assessment and intercomparison for three large-scale river basins on three continents. Earth Syst Dyn 6:17–43. https://doi.org/10.5194/esd-6-17-2015

Vidal J-P, Hingray B, Magand C, Sauquet E, Ducharne A (2016) Hierarchy of climate and hydrological uncertainties in transient low-flow projections. Hydrol Earth Syst Sci 20(9):3651–3672. https://doi.org/10.5194/hess-20-3651-2016

Villarini G (2016) On the seasonality of flooding across the continental United States. Adv Water Resour 87:80–91. https://doi.org/10.1016/j.advwatres.2015.11.009

Wada Y et al (2013) Multimodel projections and uncertainties of irrigation water demand under climate change. Geophys Res Lett 40:4626–4632. https://doi.org/10.1002/grl.50686

Wanders N, Wada Y, Van Lanen HAJ (2015) Global hydrological droughts in the 21st century under a changing hydrological regime. Earth Syst Dyn 6:1–15. https://doi.org/10.5194/esd-6-1-2015

Ward PJ, Jongman B, Weiland FS, Bouwman A, van Beek R, Bierkens MFP, Ligtvoet W, Winsemius HC (2013) Assessing flood risk at the global scale: model setup, results, and sensitivity. Environ Res Lett 8:044019. https://doi.org/10.1088/1748-9326/8/4/044019

Warszawski L, Frieler K, Huber V, Piontek F, Serdeczny O, Schewe J (2014) The Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP): project framework. Proc Natl Acad Sci U S A 111:3228–3232. https://doi.org/10.1073/pnas.1312330110

Yip S, Ferro CAT, Stephenson DB, Hawkins E (2011) A simple, coherent framework for partitioning uncertainty in climate predictions. J Clim 24:4634–4643. https://doi.org/10.1175/2011JCLI4085.1.

Zaherpour J et al (2018) Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts. Environ Res Lett 13:065015. https://doi.org/10.1088/1748-9326/aac547