



# Weighting climate models for hydrological projections: effects on contrasting hydroclimatic regions

Mariana Castaneda-Gonzalez<sup>1</sup> · Annie Poulin<sup>1</sup> · Rabindranarth Romero-Lopez<sup>2</sup> · Richard Turcotte<sup>3</sup>

Received: 9 June 2022 / Accepted: 21 November 2023 / Published online: 2 December 2023  
© The Author(s) 2023

## Abstract

Weighting climate models has recently become a more accepted approach. However, it remains a topic of ongoing discussion, especially for analyses needed at regional scales, such as hydrological assessments. Various studies have evaluated the weighting approaches for climate simulations. Yet, few case studies have assessed the impacts of weighting climate models on streamflow projections. Additionally, the methodological and location limitations of previous studies make it difficult to extrapolate their conclusions over regions with contrasting hydroclimatic regimes, highlighting the need for further studies. Thus, this study evaluates the effects of different climate model's weighting approaches on hydrological projections over hydrologically diverse basins. An ensemble of 24 global climate model (GCM) simulations coupled with a lumped hydrological model is used over 20 North American basins to generate 24 GCM-driven streamflow projections. Six unequal-weighting approaches, comprising temperature-, precipitation-, and streamflow-based criteria, were evaluated using an out-of-sample approach during the 1976–2005 reference period. Moreover, the unequal-weighting approaches were compared against the equal-weighting approach over the 1976–2005, 2041–2070, and 2070–2099 periods. The out-of-sample assessment showed that unequally weighted ensembles can improve the mean hydrograph representation under historical conditions compared to the common equal-weighting approach. In addition, results revealed that unequally weighting climate models not only impacted the magnitude and climate change signal, but also reduced the model response uncertainty spread of hydrological projections, particularly over rain-dominated basins. These results underline the need to further evaluate the adequacy of equally weighting climate models, especially for variables with generally larger uncertainty at regional scale.

**Keywords** Climate models weighting · Hydrological projections · Climate change impacts · Contrasting hydroclimatic regions · Hydrological ensemble weighting

---

✉ Mariana Castaneda-Gonzalez  
mariana.castaneda-gonzalez.1@ens.etsmtl.ca

<sup>1</sup> École de Technologie Supérieure, 1100 Notre-Dame Street West, Montreal H3C 1K3, Canada

<sup>2</sup> Universidad Veracruzana, Lomas del Estadio S/N, Zona Universitaria, 91000 Xalapa, Mexico

<sup>3</sup> Ministère du Développement Durable, Environnement Et Lutte Contre Les Changements Climatiques, 675 Boulevard René Lévesque Est, Quebec, QC G1R 5V7, Canada

## 1 Introduction

Hydrological projections issued from climate change impact assessments are often used to inform various applications, such as flood risk assessments (Salman and Li 2018). Such projections are commonly produced by using ensembles of post-processed (i.e., down-scaled and/or bias-adjusted) global or regional climate models' outputs to force one or multiple hydrological models. Large ensembles of global climate models (GCMs) and regional climate models (RCMs) are often recommended to produce ensembles of GCM- or RCM-driven streamflow projections that account for the uncertainty associated with climate modelling (Giuntoli et al. 2018; Kundzewicz et al. 2018). The management and assessment of these large ensembles of hydro-climatic projections is a topic of ongoing discussion, especially for decision-making purposes (Kiesel et al. 2020; Knutti et al. 2017; Pechlivanidis et al. 2018).

In hydrological impact studies, the dominant approach to manage and assess the large ensembles of hydro-climatic projections is the so-called climate models' democracy. This approach consists of giving identical weights to all the climate models of the ensemble, assuming that all members are equally plausible (Chen et al. 2017; Knutti et al. 2010). However, different studies have questioned the climate models' democracy approach and proposed weighting or sub-selecting ensemble members based on their adequacy to project a given variable for a specific purpose, especially for climate change impact studies at regional and local scales. The arguments supporting this approach include the following: (1) many climate models often share or duplicate processes representations (Abramowitz et al. 2019; Eyring et al. 2019; Knutti et al. 2017); (2) some climate models have shown more difficulties in representing mean regional climate than other climate models (Braconnot et al. 2012; Gleckler et al. 2008), especially over extremes (Do et al. 2020; Giuntoli et al. 2018). Therefore, weighting or sub-selecting climate models has gained acceptance in recent years, particularly for climate change impact studies (Eyring et al. 2019; Kiesel et al. 2020; Knutti et al. 2017).

Different approaches to weight or sub-select climate models have been proposed and assessed on ensembles of climate projections (e.g., Knutti et al. 2017; Räisänen et al. 2010; Sanderson et al. 2017; Xu et al. 2010). The main differences among them often include the criteria used to favor a given simulation, such as mean annual performance (e.g., Xu et al. 2010), independence (e.g., Knutti et al. 2017; Sanderson et al. 2017), and convergence (e.g., Giorgi and Mearns 2002) of one or various climate variables.

In hydrological impact studies, some approaches have assessed weighting or sub-selecting methods based on mean climate performance (e.g., Chen et al. 2017; Massoud et al. 2019, 2020; Padrón et al. 2019; Ruane and McDermid 2017), while few others have assessed weighting approaches based on mean streamflow (e.g., Wang et al. 2019; Yang et al. 2017). Among them, Padrón et al. (2019) evaluated the effects of weighting an ensemble of 36 GCM outputs on global precipitation projections. Using a modified Bayesian model averaging (BMA) method, higher weights were assigned to simulations with better performance against precipitation observations. Their results showed that projected precipitation extremes of the weighted ensemble were less pronounced in Europe, Southern Africa, and Western North America, but more pronounced in the Amazon compared to projections using equal-weighting. Thus, performance-based weighting was recommended, arguing that simulations that agree better with observations are likely to be more reliable. Focusing on GCM-driven streamflow projections at the basin scale, Chen et al. (2017) evaluated the effects of precipitation- and temperature-based weighting on

hydrological projections of a snow-dominated basin. Five weighting methods were tested over an ensemble of 28 GCM outputs with and without post-processing. Their results showed a limited impact of climate models weighting on the GCM-driven streamflow projections. Thus, the climate model's democracy approach was suggested.

More recently, Koluš et al. (2021) evaluated the sensitivity of water resources projections to climate models weighting over two basins in eastern Africa. Four weighting methods, based on climate model's performance, independence, and plausibility, were tested over an ensemble of 32 bias-adjusted GCM outputs and the 32 GCM-driven streamflow projections. Similar to Chen et al. (2017), small effects of climate model's weighting were observed on their risk assessments compared to the overall ensemble spread. However, the use of climate-based weighting for streamflow studies has been questioned due to the non-linear relationship between climate variables and streamflow (Knutti et al. 2017; Wang et al. 2019). Thus, the use of streamflow-based weighting methods has been suggested as an alternative (Kiesel et al. 2020; Wang et al. 2019; Yang et al. 2017). For instance, Yang et al. (2017) compared equal-weighting against the BMA method to weight an ensemble of global monthly runoff projections issued from 31 Earth System Models (ESMs) based on decadal mean runoff performance. Significant regional differences, including smaller runoff increases from the weighted projections compared to equal-weighting, were observed over northern latitudes, yet more pronounced runoff decreases in Amazonia and sub-Saharan Africa.

Using a different approach, Wang et al. (2019) evaluated the use of climate models weighting on GCM-driven hydrological projections simulated with a lumped hydrological model. Eight weighting methods, based on raw and bias-adjusted mean annual climate and GCM-driven streamflow performance, were tested on two different basins. The results showed that unequal weighting improved the simulation and reduced biases of the ensemble mean when using raw GCM outputs during the reference period. However, when the GCM-outputs were bias-adjusted, the impact of weighting was limited. Thus, equal weighting was still recommended. More recently, Kiesel et al. (2020) compared eight different methods to weight or sub-select 16 bias-adjusted climate model outputs from EURO-CORDEX over the Upper Danube basin. To evaluate the performance of weighting and sub-selection methods, the historical streamflow observations were divided into a reference and an evaluation period. The results revealed that the choice of method influenced the streamflow projections as much as the actual climate change signal. Moreover, their results suggested that methods maintaining more information showed better performance than methods sub-selecting a single best-performing model.

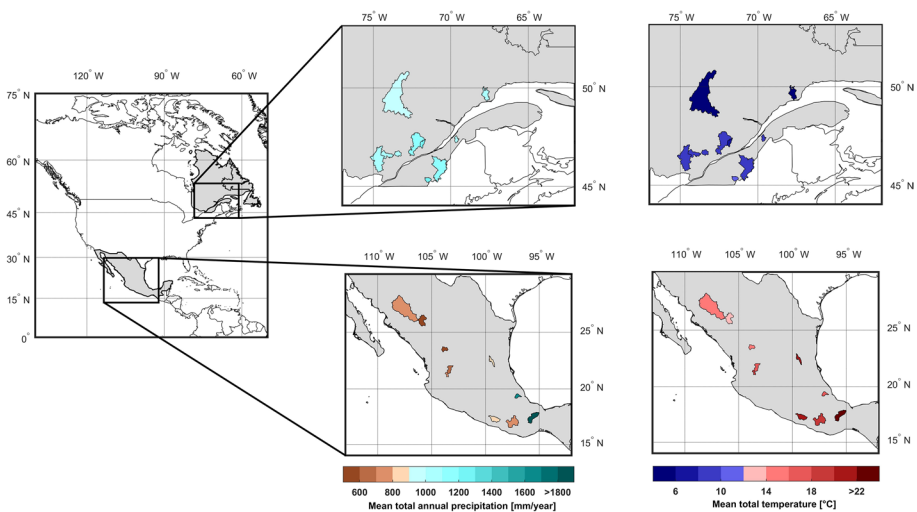
Contrasting conclusions are observed regarding the use of climate models weighting for hydrological projections. For instance, one of the main criticisms of weighting methods, similar to those of bias-adjustment approaches, is their strong reliance on climate simulations agreement with historical observations (Wootten et al. 2022). This can be problematic as good performance under historical conditions does not guarantee good performance in the future, especially if different climate conditions are expected in the future under climate change (Merz et al. 2011). At the same time, climate models that show strong difficulties in reproducing historical conditions at the regional scale can be hardly considered reliable for impact assessments. Currently, the literature shows that there is still no agreement on the method or approach to assign weights to climate models, especially for hydrological studies. This is particularly clear for studies at the regional or basin scale, where previous studies have shown different impacts depending on the region (Padrón et al. 2019; Yang et al. 2017). The complexity and variety of dominant processes at the basin scale make it hard to extrapolate previous analyses to regions with contrasting hydroclimatic regimes. Thus,

to further inform the ongoing discussion, the aim of this study is to evaluate the effects of weighting climate models on hydrological projections by analyzing the effects of different weighting criteria on future hydrological simulations over basins with contrasting hydro-meteorological regimes.

## 2 Study area and data

In this study, 20 North American basins with contrasting flood-generating processes were selected (see Fig. 1). The selected basins include 10 snow-dominated basins located in the Canadian province of Quebec and 10 rain-dominated basins located in the country of Mexico. The upper and lower panels of Fig. 1 show the mean total annual precipitation (mm) and mean annual temperature ( $^{\circ}\text{C}$ ) for the snow- and rain-dominated basins, respectively. As observed in Fig. 1, these two groups of basins represent different climatic regimes. Regarding precipitation, the 20 basins show mean total annual precipitation that varies from about 500 up to 2110 mm per year and mean annual temperatures ranging from about 5 up to 23.6  $^{\circ}\text{C}$ .

The basin data used in this study included daily historical records of minimum temperature, maximum temperature, precipitation, and streamflow. These datasets were obtained from the Hydrometeorological Sandbox – École de Technologie Supérieure (HYSETS) database (Arsenault et al. 2020). All meteorological and hydrometric data consisted of time series with a minimum length of 25 years between 1970 and 2013, depending on the availability of hydrometric data for each basin. The meteorological data used in this study consists of a grid with a resolution of  $1/16^{\circ}$  ( $\sim 6$  km), covering Mexico, the conterminous USA, and southern Canada (Livneh et al. 2015, 2013). For each basin, the data points within each catchment area were averaged to obtain a single time series per climate



**Fig. 1** Location of the 10 basins in Quebec (upper panels) and the 10 basins in Mexico (lower panels) used in the study. The upper and lower panels show the mean total annual precipitation (mm) and mean annual temperature ( $^{\circ}\text{C}$ ) for the snow-dominated and rainfall-dominated basins, respectively

variable. Further details about the characteristics of the basins are provided in Table S.1 in Online Resource 1.

### 3 Methodology

The methodology used in this study consisted of four main steps: (1) building an ensemble of GCM simulations, (2) calibrating and validating the hydrological model to couple it with the GCM-ensemble outputs, (3) applying six different climate model's weighting approaches, and (4) analyzing the resulting weighted streamflows for a reference period of 1976–2005 and two future periods of 2041–2070 and 2070–2099.

#### 3.1 Climate simulations

The first step involved selecting diverse raw GCM simulations. The choice to use raw GCM outputs was motivated by the intention to avoid the influences of bias-adjustment on the evaluation of weighting approaches. In the process of bias-adjustment, climate simulations are adjusted to align with observations. As a result, the weighting process is directly influenced, given that the same observations are subsequently used to determine the weights assigned to the climate simulations. This interaction might explain the limited impacts of weighting approaches observed in previous studies using bias-adjusted climate simulations. Additionally, it has been suggested that bias-adjustment performance over the reference period may not be preserved during future periods (Chen et al. 2020), making it problematic to interpret the results over different time frames. Thus, to avoid these issues, an ensemble of twenty-four raw GCM simulations issued from the Coupled Model Inter-comparison Project Phase 5 (CMIP5) database (Taylor et al. 2011) was selected. This climate ensemble includes outputs from fourteen different modeling centers and varied spatial resolutions, providing a diverse set of GCMs from the CMIP5 database. Please refer to Table S.2 in Online Resource 1 for further details on the GCM-ensemble members.

The outputs issued from each member of the GCM ensemble comprised three daily variables, (1) precipitation, (2) minimum temperature, and (3) maximum temperature. For each basin, the grid data points inside each catchment were spatially averaged using simple mean to obtain a single time series per climate variable. For basins with fewer than four grid points within their area, the closest surrounding points were included to ensure a minimal of four grid points per basin. The climate parameters issued from the selected points were averaged using the Thiessen's polygons method. Each dataset covered the reference period of 1976–2005 and two future horizons of 2041–2070 and 2070–2099 under the high-emission scenario, the Representative Concentration Pathway (RCP) 8.5.

#### 3.2 Hydrological modelling

The second methodological step comprised the calibration and validation of the selected hydrological model, which was then coupled with the GCM-outputs to produce the ensemble of daily GCM-driven streamflow simulations. In this study, the lumped empirical *GR4J* hydrological model (Perrin et al. 2003) combined with the snow module *CemaNeige* (Valéry et al. 2014) was used. *GR4J* is a simple rainfall-runoff model with four parameters. However, the snow accumulation, snowmelt, and evapotranspiration processes are not directly estimated by *GR4J*. Thus, the snow module *CemaNeige* and the Oudin

evapotranspiration formulation (Oudin et al. 2005) were added to allow the application of the hydrological model over the diverse study area. This addition introduced two parameters from the snow module, making a total of 6 parameters to calibrate (i.e., 4 from GR4J and 2 from CemaNeige). The inputs required by GR4J-CemaNeige consist of continuous series of daily precipitation, mean temperature, and the potential evapotranspiration calculated with the Oudin formulation. GR4J-CemaNeige combined with the Oudin formulation has demonstrated satisfactory performance over a diversity of basins and applications (e.g., Coron et al. 2012; Dallaire et al. 2021), supporting its use in this study.

The six parameters of the GR4J-CemaNige hydrological model were calibrated with the Shuffled Complex Evolution (SCE) algorithm (Duan et al. 1994) with the Kling-Gupta efficiency (KGE) criterion (Kling et al. 2012) as objective function. According to Knoben et al. (2019), KGE values larger than  $\approx -0.41$  indicate that the simulation has higher skill than the mean observations. For this study, KGE values above 0.5 were considered acceptable. The calibration and validation of the GR4J hydrological model consisted in splitting all available data into two parts. The first part was used for the calibration, while the second half was used for validation. In addition to the aforementioned validation, the GR4J model was evaluated under four contrasting climate conditions: (1) dry, (2) humid, (3) cold, and (4) warm climate conditions. These four 8-year-long periods were identified based on precipitation and temperature conditions. Mean total annual precipitation from each basin was ranked to identify dry and humid years, while mean annual temperature was used to identify cold and warm years. This calibration and validation approach is based on recent recommendations that highlight the importance of evaluating hydrological model parameterizations across contrasting conditions (Gelfan et al. 2020; Krysanova et al. 2018, 2020). The objective of this process is to ensure the robustness of the parameter sets for climate change impact assessment.

### 3.3 Weighting methods

The third step involved applying different climate- and streamflow-based weighting approaches. As previously stated, this study aims at evaluating the effects of different weighting criteria on hydrological projections. To address this aim, six unequal-weighting approaches were used, including three climate-based approaches comprising temperature and/or precipitation criteria, and three streamflow-based approaches. These weighting approaches were applied using the GCM outputs and GCM-driven streamflows simulated for reference period (i.e., 1976–2005). All unequal-weighting approaches used two main weighting methods, the reliability ensemble averaging (REA) and the upgraded REA (UREA).

The REA weighting method, developed by Giorgi and Mearns (2002), assigns weights to each member of a GCM simulations ensemble to minimize the contribution of members with poor performance. The performance criterion is based on reliability factors that consider (1) each simulation's fit to historical records and (2) a measure of the future projection convergence to the REA-weighted average. Both elements are often calculated with annual mean data. This method was initially applied over ensembles of GCM-temperature and -precipitation datasets yet recent applications have also used it over GCM-driven streamflow ensembles (e.g., Kiesel et al. 2020; Mani and Tsai 2017).

The UREA method, developed by Xu et al. (2010), proposed two major changes to the REA method. Instead of the future projection convergence criterion, the UREA method included multiple variables and statistics in the performance criteria. This allowed

including two or more variables, such as precipitation and temperature, as well as other statistics (e.g., interannual standard deviation and interannual coefficient of variation) to define the weights. The UREA method has been applied in various regions and studies for climate projections (e.g., Chen et al. 2017; Colorado-Ruiz et al. 2018; Singh and Achuta-Rao 2020) and GCM-driven streamflow projections (Wang et al. 2019).

Both the REA and UREA methods have been applied in different studies and have the advantage of including the calculation of uncertainty ranges, enabling the measurement and comparison of uncertainty spreads for all weighting approaches. More details on the weights and uncertainty ranges calculations with the REA and UREA methods can be found in Online Resource 2. It is important to underline that in this and following sections, uncertainty spread refers to the model response/output uncertainty spread.

The seven weighting approaches used in this study are the following:

**EW.** The equal weighting method assigns identical weights to all climate models.

**W1.** This approach uses the REA method to assign weights according to each climate model's historical mean annual temperature performance.

**W2.** This approach uses the REA method to assign weights according to each climate model's historical mean annual precipitation performance.

**W3.** This approach uses the UREA method to assign weights according to each climate model's historical mean annual precipitation and temperature performance.

**W4.** This approach uses the REA method to assign weights according to each climate model's historical mean annual streamflow performance.

**W5.** This approach uses the UREA method to assign weights according to each climate model's historical mean annual streamflow performance.

**W6.** This customized approach proposes using the UREA method to assign weights according to each climate model's historical mean seasonal streamflow performance. By using a seasonal-based criterion instead of the commonly used annual-based criterion, it is expected that the varying dominance of hydrological processes between seasons will be better considered and will be more adequate for hydrological studies.

The weights obtained for all methods and basins are presented in Figures S1 and S2 in Online Resource 4.

### 3.4 Data analysis

To investigate the effects of climate models' weighting on ensembles of hydrological projections, four metrics were selected and evaluated over the winter (December, January, February, DJF), spring (March, April, May, MAM), summer (June, July, August, JJA), and fall (September, October and November, SON) seasons for a reference period of 1976–2005 and two future periods of 2041–2070 and 2070–2099. Details on the calculation of each metric can be found in Online Resource 3.

The four metrics used in this study are the following:

*Mean annual and seasonal hydrograph representation.* This metric uses the normalized root mean squared error (NRMSE) criterion to compare each weighted-ensemble mean hydrograph against the observed mean hydrograph, calculated from historical records. This metric allows comparing the mean annual and seasonal streamflow representations between methods. To ensure an independent assessment of weighting methods, an



out-of-sample evaluation is conducted. Thus, the weights calculated for the 1976–1995 period are then evaluated for the period 1996–2005.

*Relative difference in mean seasonal peaks.* This metric compares the weighted-ensembles in terms of relative differences  $RD(\%)$  between the mean seasonal peaks of streamflow of a given unequal-weighting approach (i.e., W1–W6) and the equal-weighting approach (EW) used as reference. The aim is to evaluate the impact of unequal-weighting on mean seasonal peaks of streamflow compared to the equal-weighting approach over reference and future periods.

*Climate change signal.* This metric compares the climate change signals of all weighting approaches. The climate change signal of each approach is estimated by comparing the mean seasonal peak streamflow of a given approach during a future period against its mean seasonal peak streamflow during the reference period in terms of relative change (%).

*Seasonal streamflow spread.* This metric compares and measures the seasonal spread of the mean hydrographs estimated with the different weighting approaches. The standard deviation is used to quantify the seasonal spreads of the mean weighted hydrographs over the reference and future periods. The purpose is to compare the impacts of unequal- and equal-weighting approaches on the seasonal uncertainty spreads of the simulated streamflows.

## 4 Results

### 4.1 Hydrological model performance

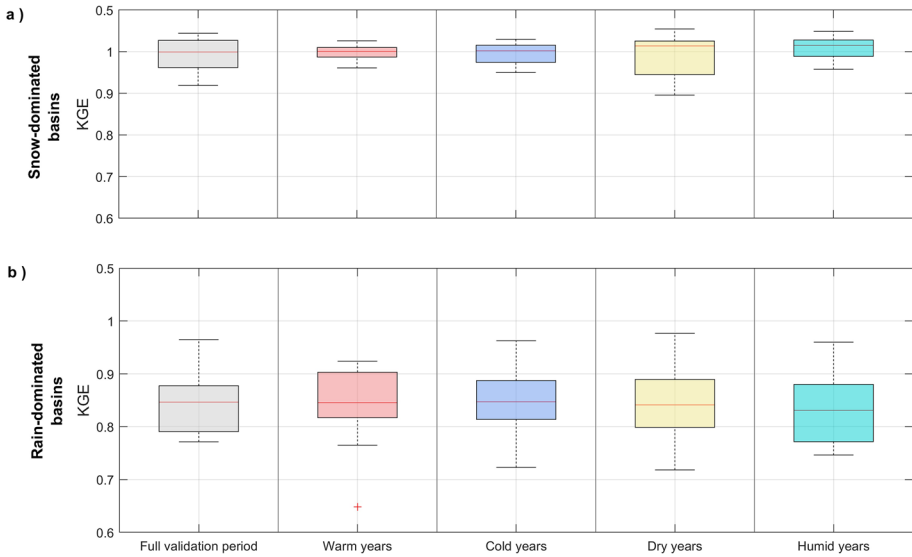
In this section, the validation results of the GR4J hydrological model are presented in Fig. 2. The figure shows the KGE values obtained from the hydrological model validation and evaluations over four climate contrasting periods for the snow- (panel a) and rain-dominated basins (panel b). The distributions of KGE-values for the evaluation under (1) the full validation period, (2) warm, (3) cold, (4) humid, and (5) dry 8-year periods, are presented from left to right (see Sect. 3.2 for details).

The distribution of KGE values obtained in the five model evaluation steps indicates a satisfactory performance overall, with median KGE values of over 0.7 for all cases. However, it is observed that the KGE values vary among the different contrasting climate conditions. For instance, GR4J tends to face more difficulties in simulating dryer conditions in the snow-dominated basins and humid conditions in the rain-dominated basins. The evaluation over the full validation period also demonstrates a satisfactory performance, with median KGE-values of 0.90 and 0.75 for the snow- and rain-dominated basins, respectively. These satisfactory performance metrics assure the use of the GR4J hydrological model for the subsequent methodological stages in this study.

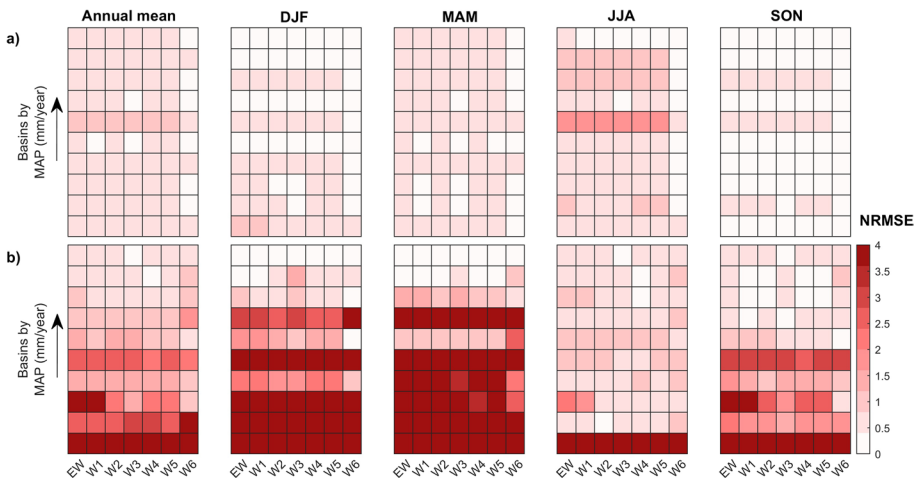
### 4.2 Mean annual and seasonal hydrograph representation

The comparison between mean annual and seasonal weighted and observed hydrographs was performed using the NRMSE over the period of 1995–2006. Figure 3 shows the NRMSE values obtained between the weighted and observed mean annual hydrographs for the snow- and rain-dominated basins on rows a and b, respectively. The NRMSE values are presented for the mean annual and seasonal hydrographs (from left to right). Each panel





**Fig. 2** Boxplots of the KGE values during the validation and evaluation over climate contrasting periods of the GR4J hydrological model are presented for the snow- and rain-dominated basins in **a** and **b**, respectively. From left to right, the results for the validation with the full-time series, and the evaluations under warm, cold, dry, and humid conditions are presented



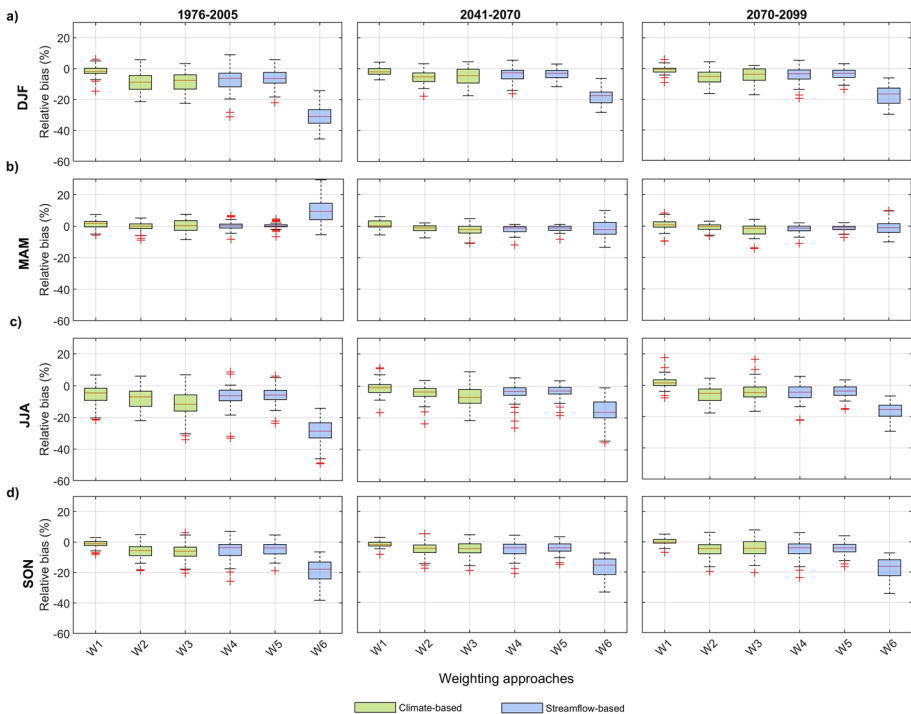
**Fig. 3** Normalized root mean squared error (NRMSE) values obtained from comparing weighted ensembles mean annual hydrograph against the observed mean annual hydrograph during the reference period of 1996–2005 for the snow-dominated basins (**a**) and the rainfall-dominated basins (**b**). Each panel shows climate weighting approaches on *x*-axis and basins on *y*-axis. Basins are ordered according to MAP, from lowest to highest

shows the weighting approaches on the *x*-axis and all basins are sorted by ascending mean annual precipitation (MAP, in mm/year) along the *y*-axis.

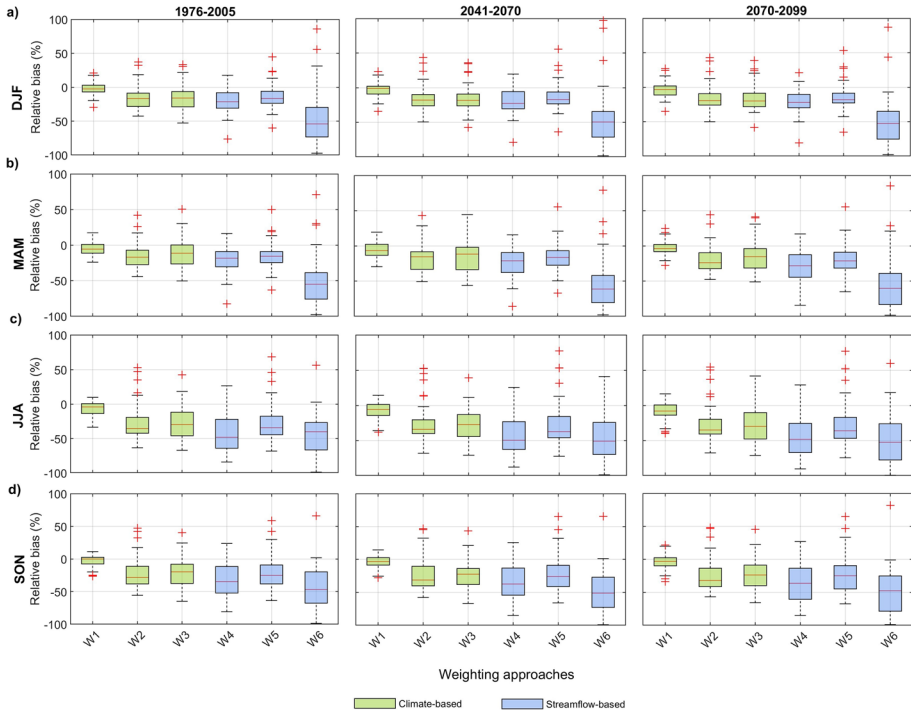
The results show that mean annual and seasonal hydrographs generally exhibit closer agreement with observations in snow-dominated basins across all weighting approaches. This is particularly evident in panel a, where almost all NRMSE values for the snow-dominated basins are lower compared to those of the rain dominated basins when comparing row by row. Among the weighting approaches, it is observed that W6 followed by W3, consistently display better annual and seasonal performance with lower NRMSE values than other approaches for most snow- and rain-dominated basins. Rain-dominated basins show more diverse results, with the driest basins showing generally larger NRMSE values, especially during dry months (DJF and MAM). Additionally, these drier basins show the overall worse mean annual hydrograph representation (All NRMSE values can be found in Tables S3 to S7 of Online Resource 4).

### 4.3 Impact on mean seasonal peak streamflow

Figures 4 and 5 show the relative bias between the mean seasonal peak streamflow of the unequally weighted ensembles (W1–W6) against the equally weighted (EW) ensemble



**Fig. 4** Relative bias (%) between the unequally weighted mean seasonal peak streamflows of the three climate-based (in green) and the three streamflow-based (in blue) ensembles and the equally weighted ensemble for the snow-dominated basins over the 1976–2005, 2041–2070, and 2070–2099 periods (from left to right panels). By row, results for the winter (a), spring (b), summer (c), and fall (d) months are presented



**Fig. 5** Relative bias (%) between the unequally-weighted mean seasonal peak streamflows of the three climate-based (in green) and the three streamflow-based (in blue) ensembles and the equally weighted ensemble for the rain-dominated basins over the 1976–2005, 2041–2070, and 2070–2099 periods (from left to right panels). By row, results for the winter (a), spring (b), summer (c), and fall (d) months are presented

over the snow- and rain-dominated basins, respectively. The results are organized by season (top to bottom) and time period (left to right).

Across most snow-dominated basins, the unequally weighted ensembles consistently show smaller mean seasonal peak streamflow values compared to the equally weighted ensemble during winter, summer, and fall seasons (panels a, c, and d) over both reference and future periods. The negative relative biases vary among weighting approaches, with the streamflow-based approach W6 showing the largest differences against EW. However, during the spring flood, a shift is observed, as more basins show a positive relative bias. This is especially observed during the reference period, where most basins show median relative biases of about +10% when using W6. It is also observed that relative differences decrease when moving into future periods across all seasons. Particularly with W6, several basins changed from positive to negative relative biases.

Over the rain-dominated basins (see Fig. 5), most basins show smaller mean seasonal peak streamflows with the six unequal-weighting approaches than with the equal-weighting approach. Similar to the snow-dominated basins, W6 shows the largest differences against EW. However, the rain-dominated basins show generally larger relative differences, with median values reaching up to −60%. During winter and spring months (panels a and b), W6 shows notably greater differences than the other

approaches. Conversely, during summer and fall months (panels c and d), both W4 and W6 show similar medians and distribution spreads, especially during the summer months.

Overall, the unequally weighted hydrological projections over snow-dominated regions showed, on average, median differences (i.e., relative biases) of about 3 to 5% with W1–W5, and 12 to 22% median differences with W6 across the different periods. Over the rain-dominated basins, the average of median differences showed values of about 16 to 18% with W1–W3, while approaches W4–W6 showed average differences of about 34 to 37%.

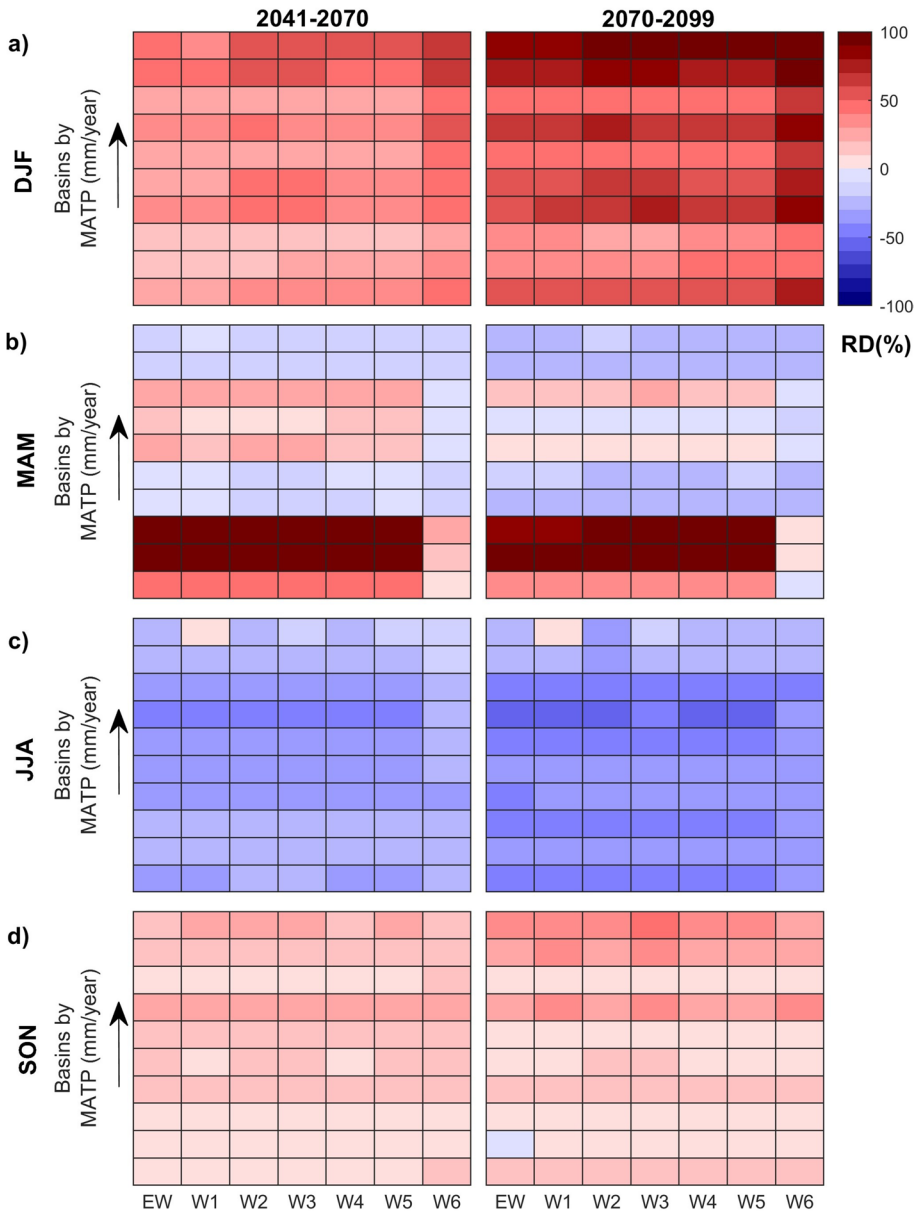
#### 4.4 Impacts on the climate change signal

Figures 6 and 7 show the climate change signals, measured as relative changes (%) between the reference and future mean seasonal peak streamflows, for all weighting approaches over the snow- and rain-dominated basins, respectively. The results are presented by season (top to bottom) and future horizon (left to right). Each panel displays the weighting approaches along the *x*-axis and the basins sorted by ascending mean annual total precipitation (MATP, in mm/year) along the *y*-axis. The climate change signals over the snow-dominated basins show relatively consistent outcomes across most weighting approaches. Nevertheless, W6 differs from the other approaches during specific seasons across both future horizons. This is especially observed during the spring flood (panel b), where W6 shows clear differences from the other approaches in certain basins. During winter and summer months (panels a and c, respectively), most approaches agree on the signal direction, showing overall peak flow increases in the winter and decreases in the summer. However, different signal magnitudes are observed, with certain approaches showing slightly larger peak flow increases in the winter (panel a) and smaller peak flow decreases in the summer (panel c), particularly W6. In the fall season (panel d), smaller differences between weighting approaches are observed between weighting approaches.

Over the rain-dominated basins (Fig. 7), a broader range of climate change signals is observed across basins, periods, and weighting approaches. No clear differences are observed between unequal-weighting approaches and equal-weighting, as climate change signals often disagree among various basins and future periods. This variability is particularly evident during summer and fall months (panels c and d). Nevertheless, it is observed that unequally weighted ensembles increased/decreased the magnitude of the climate change signal in some basins. In other words, peak flow increases or decreases projected with the equal-weighting approach are further accentuated with some unequal-weighting approaches (e.g., W6) in certain basins.

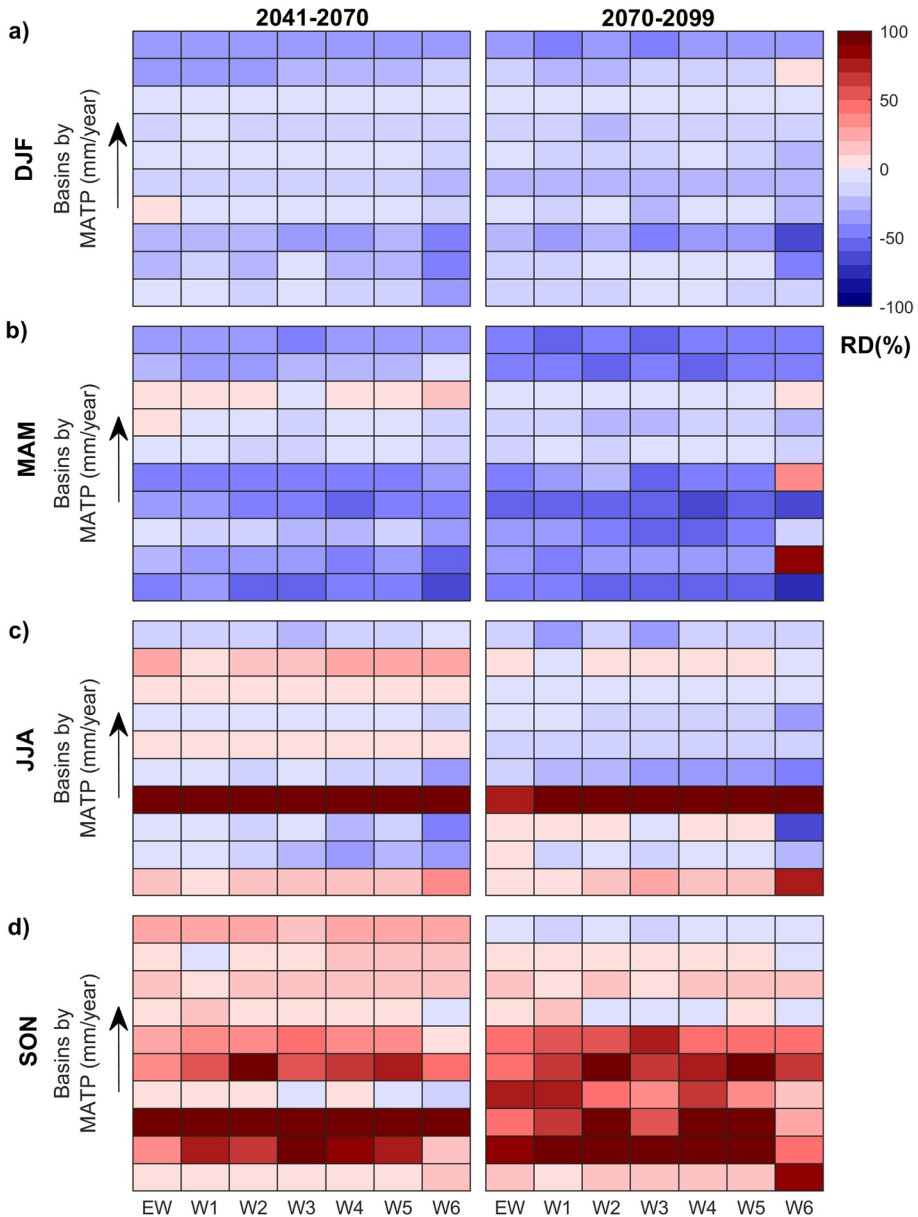
#### 4.5 Impacts on the streamflow-ensemble spread

Figures 8 and 9 show the boxplots of standard deviations (SD;  $\text{m}^3/\text{s}$ ) that quantify the ensemble spreads of the different weighted streamflow ensembles over the snow- and rain-dominated basins, respectively. The results are presented by season (top to bottom) and period (left to right). It is generally observed, over both figures, that all unequally weighted ensembles reduced the streamflow ensemble spread across all seasons. These effects are consistent over reference and future periods. Over the snow-dominated basins, the results show that unequal-weighting approaches can reduce the ensemble spread, with median standard deviations ranging from approximately 50 to 60% smaller compared to the EW

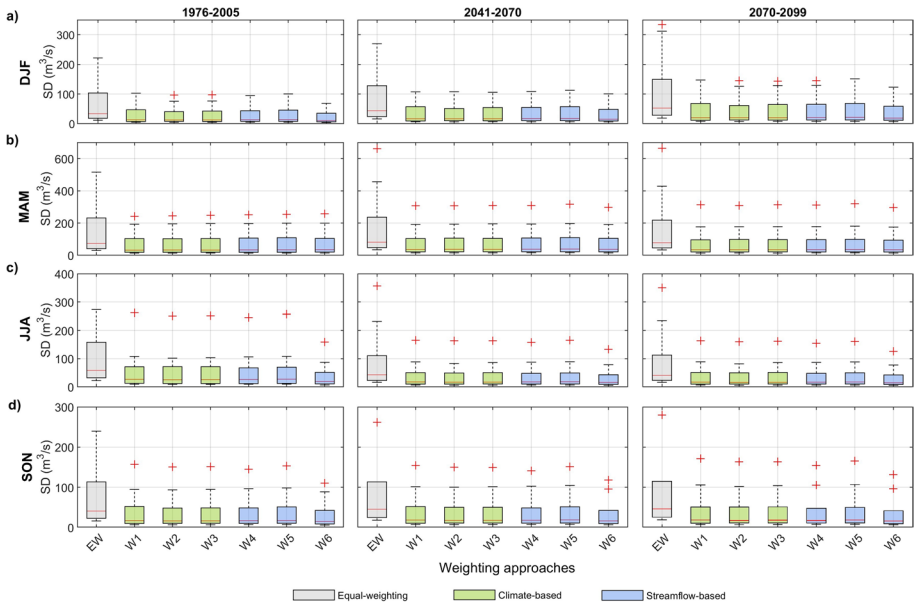


**Fig. 6** Climate change signal measured as relative changes (%) between the reference and future mean seasonal peak streamflows of all equally and unequally weighted ensembles for the snow-dominated basins over the 2041–2070 and 2070–2099 periods (from left to right panels). By row, results for the winter (a), spring (b), summer (c), and fall (d) months are presented. Shades of red indicate positive relative bias and shades of blue indicate negative relative biases

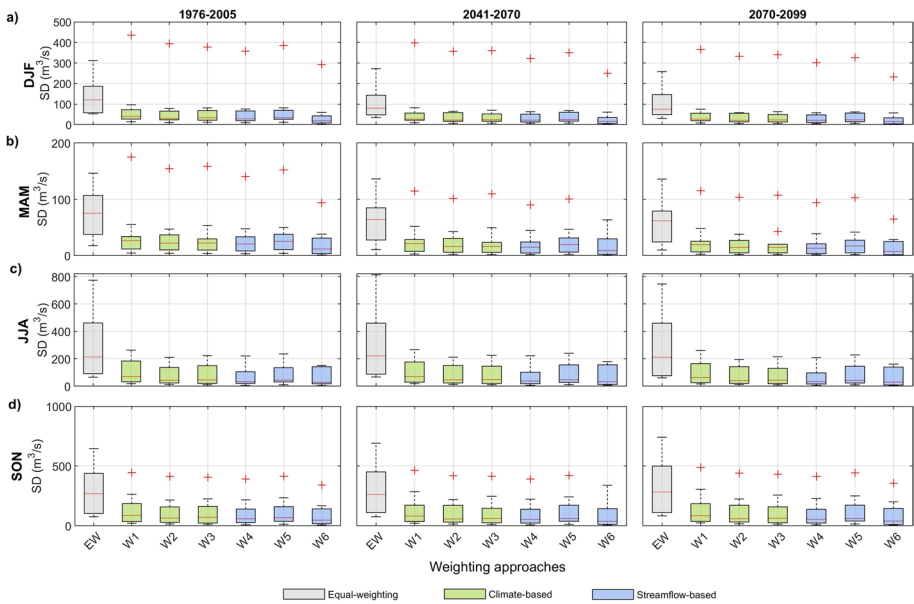
approach. These basins also show slightly smaller median standard deviations with the W6 approach, especially during summer and fall months (panels c and d). This behavior is also observed over the rain-dominated basins, with median standard deviations reaching values



**Fig. 7** Climate change signal measured as relative changes (%) between the reference and future mean seasonal peak streamflows of all equally and unequally weighted ensembles for the rain-dominated basins over the 2041–2070 and 2070–2099 periods (from left to right panels). By row, results for the winter (a), spring (b), summer (c), and fall (d) months are presented. Shades of red indicate positive relative bias and shades of blue indicate negative relative biases



**Fig. 8** Standard deviation ( $m^3/s$ ) of the equally weighted streamflow-ensemble (in grey), the three climate-based (in green), and the three streamflow-based (in blue) unequally weighted streamflow ensembles for the snow-dominated basins over the 2041–2070 and 2070–2099 periods (from left to right panels). By row, results for the winter (a), spring (b), summer (c), and fall (d) months are presented



**Fig. 9** Standard deviation ( $m^3/s$ ) of the equally weighted streamflow ensemble (in grey), the three climate-based (in green), and the three streamflow-based (in blue) unequally weighted streamflow ensembles for the rain-dominated basins over the 2041–2070 and 2070–2099 periods (from left to right panels). By row, results for the winter (a), spring (b), summer (c), and fall (d) months are presented



of up to 80% smaller than the ensemble spreads using the EW approach. These effects are consistently observed across all seasons and periods, especially with the W6 approach.

## 5 Discussion

### 5.1 Impacts of climate models weighting

Using large ensembles of climate simulations has become a standard approach for assessing climate change impacts on hydrology. To deal with these large ensembles, different studies have proposed weighting climate simulations based on their ability to reproduce the variable of interest. Thus, to further inform on the impacts of weighting strategies, different weighting approaches were tested on an ensemble of GCM-driven streamflow simulations over 20 basins with contrasting hydroclimatic regimes. The findings, in line with other studies (e.g., Kiesel et al. 2020; Wang et al. 2019), showed that weighting climate models can improve the representation of mean annual and seasonal hydrographs during the reference period. This improvement is evident in Fig. 3, where certain weighting approaches showed lower NRMSE values compared to the equal-weighting approach. Among the evaluated weighting approaches, W6 (the seasonal streamflow-based weighting approach) demonstrated a generally robust performance. This NRMSE reduction is not only observed over snow-dominated basins, where EW already showed a generally good hydrograph representation. It is also observed over warmer and dryer rain-dominated basins. It is important, however, to highlight that the selected methods and datasets add uncertainty to these findings. For instance, the calculation of weights over a limited period (1976–1995) and their subsequent evaluation over the 1996–2005 period add uncertainty to the present study, as natural variability is overlooked when using relatively short periods. Thus, future studies using longer evaluation periods are recommended.

The analyses also highlighted that weighting climate models can impact the magnitude of mean seasonal peak streamflows, climate change signal, and model response uncertainty spread of hydrological projections. In particular, unequally weighted hydrological projections showed peak streamflows of about 3 to 5% different than EW with W1–W5, and 12 to 22% different with W6, across snow-dominated basins. The rain-dominated basins showed larger peak streamflow differences of approximately 16 to 18% with W1–W3, and 34 to 37% with W4–W6. However, the direction and magnitude of these changes varied across seasons, periods, basins, and, notably, between weighting approaches. This was observed over Figs. 4 and 5, where W6 showed the largest mean peak differences in comparison to the equal-weighting approach.

The climate change signals showed diverse impacts across basins and weighting approaches. In certain snow-dominated basins, the climate change signal for the spring months changed from projected seasonal peak flow increases to decreases when using the W6 approach, while in others, the climate change signal was clearly attenuated (i.e., the three driest basins). These notable effects can be explained by the higher mean spring peak streamflows obtained during the reference period when using the W6 method (as observed in Fig. 4b), which consequently modified the difference against the projected peak streamflows. In contrast, larger peak flow increases and smaller peak flow decreases were observed during winter and summer months, respectively. Over the rain-dominated basins, no systematic effect of weighting was observed. However, climate change signals sometimes exhibited opposite projected changes between approaches, especially during the

flooding summer-fall months. Additionally, some basins showed peak flow decreases and increases compared to the ones projected with the EW approach. These results emphasize the fact that the choice of weighting approach can not only impact the magnitude of the signal, but also change the projected signal in the opposite direction.

The observed effects of weighting approaches on the projected signals can be attributed to the observed volatility of the hydrological change signal among different combinations of climate models (Melsen et al. 2018), as well as the various factors involved in the overall weighting process. For instance, one crucial aspect of weighting methods is the underlying assumption of stationary climate simulation performance. In comparison to bias-adjustment methods, climate simulation biases (in this case, weights) are assumed to be stationary. However, studies suggest that biases vary over time (Chen et al. 2020), and it can be expected that weights will also vary. Moreover, it has been suggested that climate change might impact streamflow seasonality, such as earlier snowmelt and/or shifted rain seasons (Eisner et al. 2017; Vormoor et al. 2015). This can pose a challenge when applying the seasonal weighting method W6, as future peak streamflows may occur in different months, each with different weights. This highlights the need to improve existing weighting methods by incorporating more flexible performance evaluations, such as distribution-based weighting.

Regarding the uncertainty-spread analysis, both basin groups showed clear reductions in seasonal spread of the streamflow ensembles when using unequal-weighting approaches compared to the seasonal spreads estimated with the EW approach across all seasons and periods. This outcome was expected, as some previous studies have shown reductions in both global and regional uncertainty spreads through the process of weighting climate model ensembles (Exbrayat et al. 2018; Multsch et al. 2015). These reductions were particularly pronounced over rain-dominated basins, where seasonal uncertainty spread reductions reached up to 80%. These generally larger reductions in seasonal spreads over rain-dominated basins, along with the larger peak streamflow differences described in Section 4.3, may be explained by the larger uncertainty associated with climate modelling in regions or seasons characterized by relatively heavy rainfall (Woldemeskel et al. 2016). Consequently, larger effects can be expected in areas or periods where streamflow is primarily driven by rainfall. However, it is important to highlight that this study was limited by the employed weighting methods (i.e., REA and UREA). Thus, if different methods are used to estimate uncertainty spreads, contrasting results can be expected.

## 5.2 Effects of climate model's weighting criteria

Precipitation-, temperature-, and streamflow-based criteria were used and compared across all basins, with some approaches relying on a single climate/streamflow variable while others incorporated multiple climate/streamflow variables. The results showed that the degree of impact from the weighting approaches varied depending on the selected criteria. This variability is clear when comparing W6 with the other approaches across the different analyses, as well as in Figures S1 and S2, this particular method exhibits more variability in weights compared to other approaches (refer to Online Resource 4). This weighting approach not only diminished the NRMSE against mean observations in comparison to all the other approaches during the reference period but also showed the largest differences against the equal-weighting approach. This behavior can be explained by the variable and multi-criteria used for evaluation. This approach is based on streamflow performance, which, in line with Wang et al. (2019), generally shows better annual streamflow

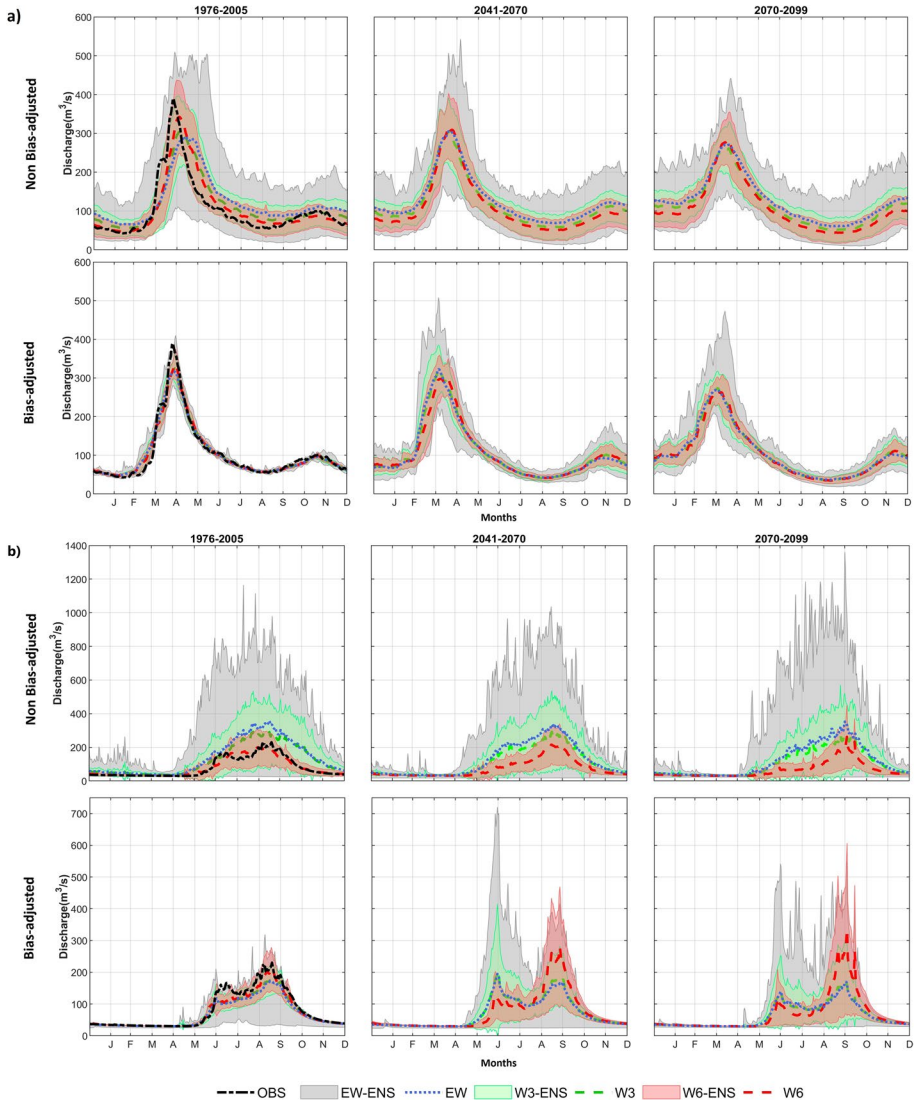
representation than climate-based weighting due to the non-linear relationship of climate variables and streamflow. Additionally, this approach does not solely rely on annual indicators but also incorporates multiple seasonal indicators. This means that climate models showing better seasonal agreement with observed streamflow seasonality will receive higher weights. While annual indicators have been previously used for weighting climate and streamflow projections, they can be misleading for streamflow due to the diverse processes that drive streamflow throughout the year. This effect is also observed in Fig. 3, where some basins show improved mean seasonal hydrograph representation with temperature- or precipitation-based weighting approaches. For instance, over snow-dominated basins, temperature-based approaches sometimes outperform others, as flood peaks are strongly influenced by temperature during spring melt. These results underscore the importance of selecting weighting criteria based on the needs of the studied variable.

However, it is important to note that while objective metrics are chosen to evaluate climate model performance, the selection of weighting criteria unavoidably involves subjectivity, as there is no universally accepted approach for such evaluation in the hydrological community. For instance, other criteria such as observed trend representation, spatial patterns, climate model independence, or signal-to-noise ratio have been suggested in the literature to evaluate the overall GCMs performance (Meher et al. 2017). However, these criteria were not incorporated into any of the studied approaches. Integrating these types of criteria with mean annual and seasonal representations can help identify more robust methods for ranking and weighting GCMs in climate change impact studies, relying not solely on performance metrics. For example, it is noteworthy that certain GCMs consistently received higher weights than others across the different weighting methods and basins in this study. These patterns are highlighted in Table S2, where the cumulative weights assigned to each GCM across all basins and weighting approaches indicated a general preference for GCMs #19 and #5 over the snow- and rain-dominated basins, respectively.

It is important to underscore that this study is limited by the criteria and methods chosen to evaluate and weight the GCMs. Therefore, further studies that assess the potential impacts of alternative criteria, as conducted in this study, are necessary to provide a clearer understanding of the adequacy of weighting criteria for specific purposes, as well as their implications for climate change impact studies.

### 5.3 The implications of bias-adjustment, hydrological modelling, and weighting

In this study, raw GCM outputs were combined with a lumped hydrological model to generate streamflow simulations that were used to weight the GCM-driven streamflow simulations. However, the absence of bias-adjustment can impact the results. To explore these potential issues, Fig. 10 presents a comparison of mean annual hydrographs for a snow- (panel a) and a rain-dominated basin (panel b), derived from raw and bias-adjusted GCM-outputs (upper and lower panels, respectively). Additional details regarding the two selected basins are provided in Table S.1 of Online Resource 1. For each basin, the uncertainty spreads and means of the mean annual hydrographs derived from EW, W3, and W6 are presented for both the reference and future periods. The W3 and W6 methods were chosen to compare one climate-based and one streamflow-based approach, respectively. In addition, these methods were identified as having some of the largest effects over both basin groups. Bias-adjustments were performed using the quantile mapping approach described in (Chen et al. 2020).



**Fig. 10** Mean annual hydrographs of one snow- (a) and one rain-dominated basins (b), estimated from raw and bias-adjusted GCM-outputs (upper and lower panels, respectively). For each basin, the EW, W3, and W6 mean annual hydrographs uncertainty spreads and means are presented for the reference and future periods (from left to right). Selected basins are highlighted in Table S.1 in Online Resource 1

Overall, Fig. 10 shows narrower spreads in the bias-corrected GCM-driven mean annual hydrographs. This behavior is clearly observed over both basins and all weighting approaches. However, it is notable that while the uncertainty ranges of W3 and W6 estimated from the raw GCM-driven mean annual hydrographs remain relatively constant over time, the bias-adjusted uncertainty ranges show clear spread changes over the future periods. Such differences were expected, as bias-adjustment methods may not preserve their performance over future periods (Chen et al. 2020).

Regarding the weighted means, the snow-dominated basin generally shows similar means among all methods, excluding the reference period where EW presents larger differences against W3 and W6. Nonetheless, similar means are observed when considering the raw and bias-adjusted GCM-driven hydrographs. Conversely, the rain-dominated basin shows more pronounced differences between weighting methods, particularly with W6. For instance, when examining the highest peak flow, both approaches (i.e., with and without bias-adjustment) show increases over future horizons. However, the bias-adjusted approach shows a clearly larger peak flow increase. This behavior can be linked to the bias-adjustment process, which adjusts the simulation to fit the highest observed peak within the reference period. These biases are assumed to remain constant throughout the future periods, potentially leading to these amplified peaks during the same months. Another contributing factor might be the coarse resolution of certain GCM/ESM simulations included in the study. Post-processing techniques serve not only to adjust inherent biases in climate simulations but also allow downscaling coarser GCM simulations, a process expected to be more adequate for basin-scale studies that unavoidably adds uncertainty to the results.

The larger effects and differences observed in the rain-dominated basin could also be linked to the greater uncertainty spreads in regions or seasons where streamflow production is influenced by convective precipitation events (Castaneda-Gonzalez et al. 2022). This larger spread becomes evident when comparing the snow- and rain-dominated basins.

The rationale behind using raw GCM outputs was to isolate the effects of weighting methods, solely focusing on quantifying their impacts on hydrological projections across various basins. Nonetheless, it is clear that this methodology produces interactions between raw GCM outputs and the calibrated hydrological model, which could impact the weighting process, particularly when strongly biased GCM outputs are fed to a calibrated hydrological model. Therefore, further studies that explore alternative methods to integrate these approaches could offer insights into more adequate applications for basin-scale assessments. For instance, using weighting methods to rank the most suitable climate simulations before and after bias-adjustments can be suggested, given that bias-adjustment methods can mask fundamental modelling deficiencies (Chen et al. 2021).

## 6 Conclusions

In this study, six approaches of unequal-weighting of climate models were tested and compared against the most common equal-weighting approach to analyze their impacts on hydrological projections. An ensemble of 24 raw GCM simulations was combined with a simple lumped hydrological model to produce 24 GCM-driven streamflows for 20 river basins located in Canada and Mexico for a reference and two future periods. Different weighting methods and criteria were first evaluated using the NRMSE metric in an out-of-sample approach during the reference period. Subsequently, all weighting methods were applied to explore their effects on hydrological projections across two distinct basin groups: (1) 10 rain-dominated and (2) 10 snow-dominated basins.

Overall, the results suggest that weighting climate models can impact the magnitude of projected mean seasonal peak streamflows, climate change signal, and uncertainty spread of the ensemble of river discharge projections. Although the primary objective of this study was not to evaluate the performance of different weighting methods, but rather to inform on the potential effects of their application on hydrological projections, various key findings emerged:

1. The unequal weighting of climate models can improve the representations of mean annual and seasonal hydrographs during the reference period. Among the different weighting approaches evaluated using an out-of-sample assessment during the reference period, the approach based on mean seasonal streamflows often outperformed the other approaches relying on mean annual streamflow, precipitation, or temperature data.
2. The choice of climate model's weighting approach can strongly impact the magnitude and direction of climate change signal in terms of mean seasonal peak streamflow. These effects vary across seasons and, more notably, between regions with different hydroclimatic regimes. Over both snow- and rain-dominated basins, the climate change signal often changed to an opposite direction during their main flood seasons (spring and summer-fall, respectively).
3. Rain-dominated basins generally exhibited larger impacts on mean seasonal peak streamflows and streamflow-ensemble spread when climate models were unequally weighted compared to snow-dominated basins.
4. Weighting climate simulations can lead to a reduction in the uncertainty spread of peak flow and streamflow projections.

While our study provides valuable insights, its limitations underscore the importance of further investigations to determine suitable approaches for weighting climate simulations in hydrological applications. For instance, future efforts might involve including more recent global/regional climate simulations at higher resolutions and/or physically based hydrological models to better gauge the adequacy of hydro-climatic simulations for hydrological applications. Additionally, it could be recommended that future studies focus on the development of weighting methods capable of estimating weighted uncertainty spreads. This would allow for a more thorough assessment of the potential reductions in uncertainty spreads.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10584-023-03643-9>.

**Author contribution** All the authors contributed in the initial design of the present study. Annie Poulin, Rabindranath Romero-Lopez, Richard Turcotte, and Mariana Castaneda-Gonzalez collected data and materials used in the study. All authors participated in designing the methodology and analyses. Mariana Castaneda-Gonzalez was in charge of writing the first draft of the manuscript. All authors reviewed and commented on the different versions to finally approve this final manuscript.

**Funding** This work was partially supported by the Consejo Nacional de Ciencia y Tecnología (CONACYT), the Fonds de recherche du Québec – Nature et technologies (FRQNT), and the Ministère de l'Économie, de la Science et de l'Innovation.

**Data availability** The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Declarations

**Ethics approval** Not applicable.

**Consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Competing interests** The authors declare no competing interests.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abramowitz G, Herger N, Gutmann E, Hammerling D, Knutti R, Leduc M, Lorenz R, Pincus R, Schmidt GA (2019) ESD Reviews: model dependence in multi-model climate ensembles: weighting, sub-selection and out-of-sample testing. *Earth Syst Dynam* 10:91–105
- Arsenault R, Brissette F, Martel J-L, Troin M, Lévesque G, Davidson-Chaput J, Gonzalez MC, Ameli A, Poulin A (2020) A comprehensive, multisource database for hydrometeorological modeling of 14,425 North American watersheds. *Sci Data* 7:243
- Braconnot P, Harrison SP, Kageyama M, Bartlein PJ, Masson-Delmotte V, Abe-Ouchi A, Otto-Bliesner B, Zhao Y (2012) Evaluation of climate models using palaeoclimatic data. *Nat Clim Chang* 2:417–424
- Castaneda-Gonzalez M, Poulin A, Romero-Lopez R, Turcotte R, Chaumont D (2022) Uncertainty sources in flood projections over contrasting hydrometeorological regimes. *Hydrol Sci J* 67:2232–2253
- Chen J, Brissette FP, Lucas-Picher P, Caya D (2017) Impacts of weighting climate models for hydro-meteorological climate change studies. *J Hydrol* 549:534–546
- Chen J, Brissette FP, Caya D (2020) Remaining error sources in bias-corrected climate model outputs. *Clim Change* 162:563–582
- Chen J, Arsenault R, Brissette FP, Zhang S (2021) Climate change impact studies: should we bias correct climate model outputs or post-process impact model outputs? *Water Resour Res* 57:e2020WR028638.
- Colorado-Ruiz G, Cavazos T, Salinas JA, De Grau P, Ayala R (2018) Climate change projections from Coupled Model Intercomparison Project phase 5 multi-model weighted ensembles for Mexico, the North American monsoon, and the mid-summer drought region. *Int J Climatol* 38:5699–5716
- Coron L, Andréassian V, Perrin C, Lerat J, Vaze J, Bourqui M, Hendrickx F (2012) Crash testing hydrological models in contrasted climate conditions: an experiment on 216 Australian catchments. *Water Resour Res* 48(5). <https://doi.org/10.1029/2011wr011721>
- Dallaire G, Poulin A, Arsenault R, Brissette F (2021) Uncertainty of potential evapotranspiration modelling in climate change impact studies on low flows in North America. *Hydrol Sci J* 66:689–702
- Do HX, Zhao F, Westra S, Leonard M, Gudmundsson L, Boulange JES, Chang J, Ciais P, Gerten D, Gosling SN, Müller Schmied H, Stacke T, Telteu CE, Wada Y (2020) Historical and future changes in global flood magnitude – evidence from a model–observation investigation. *Hydrol Earth Syst Sci* 24:1543–1564
- Duan Q, Sorooshian S, Gupta VK (1994) Optimal use of the SCE-UA global optimization method for calibrating watershed models. *J Hydrol* 158:265–284
- Eisner S, Flörke M, Chamorro A, Daggupati P, Donnelly C, Huang J, Hundecha Y, Koch H, Kalugin A, Krylenko I, Mishra V, Piniewski M, Samaniego L, Seidou O, Wallner M, Krysanova V (2017) An ensemble analysis of climate change impacts on streamflow seasonality across 11 large river basins. *Clim Change* 141:401–417
- Exbrayat JF, Bloom AA, Falloon P, Ito A, Smallman TL, Williams M (2018) Reliability ensemble averaging of 21st century projections of terrestrial net primary productivity reduces global and regional uncertainties. *Earth Syst Dynam* 9:153–165
- Eyring V, Cox PM, Flato GM, Gleckler PJ, Abramowitz G, Caldwell P, Collins WD, Gier BK, Hall AD, Hoffman FM, Hurtt GC, Jahn A, Jones CD, Klein SA, Krasting JP, Kwiatkowski L, Lorenz R, Maloney E, Meehl GA, Pendergrass AG, Pincus R, Ruane AC, Russell JL, Sanderson BM, Santer BD, Sherwood SC, Simpson IR, Stouffer RJ, Williamson MS (2019) Taking climate model evaluation to the next level. *Nat Clim Chang* 9:102–110
- Gelfan A, Kalugin A, Krylenko I, Nasonova O, Gusev Y, Kovalev E (2020) Does a successful comprehensive evaluation increase confidence in a hydrological model intended for climate impact assessment? *Clim Change* 163:1165–1185



- Giorgi F, Mearns LO (2002) Calculation of Average, Uncertainty Range, and Reliability of Regional Climate Changes from AOGCM Simulations via the “Reliability Ensemble Averaging” (REA) Method. *J Clim* 15:1141–1158
- Giuntoli I, Villarini G, Prudhomme C, Hannah DM (2018) Uncertainties in projected runoff over the conterminous United States. *Clim Change* 150:149–162
- Gleckler PJ, Taylor KE, Doutriaux C (2008) Performance metrics for climate models. *J Geophys Res: Atmos* 113(D6). <https://doi.org/10.1029/2007JD008972>
- Kiesel J, Stanzel P, Kling H, Fohrer N, Jähnig SC, Pechlivanidis I (2020) Streamflow-based evaluation of climate model sub-selection methods. *Clim Change* 163:1267–1285
- Kling H, Fuchs M, Paulin M (2012) Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J Hydrol* 424–425:264–277
- Knoben WJ, Freer JE, Woods RA (2019) Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. *Hydrol Earth Syst Sci* 23:4323–4331
- Knutti R, Furrer R, Tebaldi C, Cermak J, Meehl GA (2010) Challenges in Combining Projections from Multiple Climate Models. *J Clim* 23:2739–2758
- Knutti R, Sedláček J, Sanderson BM, Lorenz R, Fischer EM, Eyring V (2017) A climate model projection weighting scheme accounting for performance and interdependence. *Geophys Res Lett* 44:1909–1918
- Kolusu SR, Siderius C, Todd MC, Bhave A, Conway D, James R, Washington R, Geressu R, Harou JJ, Kashaigili JJ (2021) Sensitivity of projected climate impacts to climate model weighting: multi-sector analysis in eastern Africa. *Clim Change* 164:36
- Krysanova V, Donnelly C, Gelfan A, Gerten D, Arheimer B, Hattermann F, Kundzewicz ZW (2018) How the performance of hydrological models relates to credibility of projections under climate change. *Hydrol Sci J* 63:696–720
- Krysanova V, Hattermann FF, Kundzewicz ZW (2020) How evaluation of hydrological models influences results of climate impact assessment—an editorial. *Clim Change* 163:1121–1141
- Kundzewicz ZW, Krysanova V, Benestad RE, Hov Ø, Piniewski M, Otto IM (2018) Uncertainty in climate change impacts on water resources. *Environ Sci Policy* 79:1–8
- Livneh B, Rosenber EA, Lin C, Nijssen B, Mishra V, Andreadis KM, Maurer EP, Lettenmaier DP (2013) A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: update and extensions. *J Clim* 26:9384–9392
- Livneh B, Bohn TJ, Pierce DW, Munoz-Arriola F, Nijssen B, Vose R, Cayan DR, Brekke L (2015) A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950–2013. *Sci Data* 2:150042
- Mani A, Tsai FT-C (2017) Ensemble averaging methods for quantifying uncertainty sources in modeling climate change impact on runoff projection. *J Hydrol Eng* 22:04016067
- Massoud EC, Espinoza V, Guan B, Waliser DE (2019) Global climate model ensemble approaches for future projections of atmospheric rivers. *Earth’s Future* 7:1136–1151
- Massoud EC, Lee H, Gibson PB, Loikith P, Waliser DE (2020) Bayesian model averaging of climate model projections constrained by precipitation observations over the contiguous United States. *J Hydrometeorol* 21:2401–2418
- Meher JK, Das L, Akhter J, Benestad RE, Mezghani A (2017) Performance of CMIP3 and CMIP5 GCMs to simulate observed rainfall characteristics over the western Himalayan region. *J Clim* 30:7777–7799
- Melsen LA, Addor N, Mizukami N, Newman AJ, Torfs PJJF, Clark MP, Uijlenhoet R, Teuling AJ (2018) Mapping (dis)agreement in hydrologic projections. *Hydrol Earth Syst Sci* 22:1775–1791
- Merz R, Parajka J, Blöschl G (2011) Time stability of catchment model parameters: Implications for climate impact analyses. *Water Resour Res* 47
- Multsch S, Exbrayat JF, Kirby M, Viney NR, Frede HG, Breuer L (2015) Reduction of predictive uncertainty in estimating irrigation water requirement through multi-model ensembles and ensemble averaging. *Geosci Model Dev* 8:1233–1244
- Oudin L, Hervieu F, Michel C, Perrin C, Andréassian V, Anctil F, Loumagne C (2005) Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *J Hydrol* 303:290–306
- Padrón RS, Gudmundsson L, Seneviratne SI (2019) Observational constraints reduce likelihood of extreme changes in multidecadal land water availability. *Geophys Res Lett* 46:736–744
- Pechlivanidis IG, Gupta H, Bosshard T (2018) An information theory approach to identifying a representative subset of hydro-climatic simulations for impact modeling studies. *Water Resour Res* 54:5422–5435
- Perrin C, Michel C, Andréassian V (2003) Improvement of a parsimonious model for streamflow simulation. *J Hydrol* 279:275–289
- Räisänen J, Ruokolainen L, Ylhäisi J (2010) Weighting of model results for improving best estimates of climate change. *Clim Dyn* 35:407–422

- Ruane AC, McDermid SP (2017) Selection of a representative subset of global climate models that captures the profile of regional changes for integrated climate impacts assessment. *Earth Perspect* 4:1
- Salman AM, Li Y (2018) Flood Risk assessment, future trend modeling, and risk communication: a review of ongoing research. *Nat Hazard Rev* 19:04018011
- Sanderson BM, Wehner M, Knutti R (2017) Skill and independence weighting for multi-model assessments. *Geosci Model Dev* 10:2379–2395
- Singh R, AchutaRao K (2020) Sensitivity of future climate change and uncertainty over India to performance-based model weighting. *Clim Change* 160:385–406
- Taylor KE, Stouffer RJ, Meehl GA (2011) An overview of CMIP5 and the experiment design. *Bull Am Meteor Soc* 93:485–498
- Valéry A, Andréassian V, Perrin C (2014) ‘As simple as possible but not simpler’: what is useful in a temperature-based snow-accounting routine? Part 2–sensitivity analysis of the Cemaneige snow accounting routine on 380 catchments. *J Hydrol* 517:1176–1187
- Vormoor K, Lawrence D, Heistermann M, Bronstert A (2015) Climate change impacts on the seasonality and generation processes of floods – projections and uncertainties for catchments with mixed snow-melt/rainfall regimes. *Hydrol Earth Syst Sci* 19:913–931. <https://doi.org/10.5194/hess-19-913-2015>
- Wang HM, Chen J, Xu CY, Chen H, Guo S, Xie P, Li X (2019) Does the weighting of climate simulations result in a more reasonable quantification of hydrological impacts? *Hydrol Earth Syst Sci Discuss* 2019:1–29
- Woldemeskel FM, Sharma A, Sivakumar B, Mehrotra R (2016) Quantification of precipitation and temperature uncertainties simulated by CMIP3 and CMIP5 models. *J Geophys Res: Atmos* 121:3–17
- Wooten A, Massoud E, Waliser D, Lee H (2022) To weight or not to weight: assessing sensitivities of climate model weighting to multiple methods, variables, and domains. *Earth Syst Dynam Discuss* 2022:1–32
- Xu Y, Gao X, Giorgi F (2010) Upgrades to the reliability ensemble averaging method for producing probabilistic climate-change projections. *Climate Res* 41:61–81
- Yang H, Zhou F, Piao S, Huang M, Chen A, Ciais P, Li Y, Lian X, Peng S, Zeng Z (2017) Regional patterns of future runoff changes from Earth system models constrained by observation. *Geophys Res Lett* 44:5540–5549

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.