



# Understanding the influence of “hot” models in climate impact studies: a hydrological perspective

Mehrad Rahimpour Asenjan, Francois Brissette, Jean-Luc Martel, and Richard Arsenault

Hydrology, Climate and Climate Change Laboratory, École de technologie supérieure, Université du Québec, 1100 Notre-Dame Street West, Montréal, Quebec, H3C 1K3, Canada

**Correspondence:** Mehrad Rahimpour Asenjan (mehrad.rahimpour-asenjan.1@ens.etsmtl.ca)

Received: 9 February 2023 – Discussion started: 27 February 2023

Revised: 19 July 2023 – Accepted: 14 October 2023 – Published: 11 December 2023

**Abstract.** Efficient adaptation strategies to climate change require the estimation of future impacts and the uncertainty surrounding this estimation. Over- or underestimating future uncertainty may lead to maladaptation. Hydrological impact studies typically use a top-down approach in which multiple climate models are used to assess the uncertainty related to the climate model structure and climate sensitivity. Despite ongoing debate, impact modelers have typically embraced the concept of “model democracy”, in which each climate model is considered equally fit. The newer Coupled Model Intercomparison Project Phase 6 (CMIP6) simulations, with several models showing a climate sensitivity larger than that of Phase 5 (CMIP5) and larger than the likely range based on past climate information and understanding of planetary physics, have reignited the model democracy debate. Some have suggested that “hot” models be removed from impact studies to avoid skewing impact results toward unlikely futures. Indeed, the inclusion of these models in impact studies carries a significant risk of overestimating the impact of climate change.

This large-sample study looks at the impact of removing hot models on the projections of future streamflow over 3107 North American catchments. More precisely, the variability in future projections of mean, high, and low flows is evaluated using an ensemble of 19 CMIP6 general circulation models (GCMs), 5 of which are deemed hot based on their global equilibrium climate sensitivity (ECS). The results show that the reduced ensemble of 14 climate models provides streamflow projections with reduced future variability for Canada, Alaska, the Southeast US, and along the Pacific coast. Elsewhere, the reduced ensemble has either no impact or results in increased variability in future streamflow,

indicating that global outlier climate models do not necessarily provide regional outlier projections of future impacts. These results emphasize the delicate nature of climate model selection, especially based on global fitness metrics that may not be appropriate for local and regional assessments.

## 1 Introduction

Understanding the impact of climate change on water resources and hydrology is crucial for developing effective strategies for mitigation and adaptation (Eyring et al., 2019; Miara et al., 2017). The output of hydrological (e.g., Karlsson et al., 2016), water quality (Prajapati et al., 2023), and sediment transport (Sabokruhie et al., 2021) impact assessment studies is dependent on the choice of the future climate change projections. Hydrologists primarily use climate projection outputs from general circulation models (GCMs; e.g., Tabari, 2020) to study these impacts. The Coupled Model Intercomparison Project (CMIP) provides standardized metadata from coordinated simulations by different climate modeling groups (Meehl et al., 2007). The more recent Phase 6, CMIP6 (Eyring et al., 2016), is gradually replacing the widely used Phase 5, CMIP5, from the last decade (Hirabayashi et al., 2021; Martel et al., 2022; Zhang et al., 2023).

The concept of “model democracy” has been widely used in impact studies (e.g., Collins et al., 2013; IPCC, 2014), despite criticism (Knutti, 2010). This approach considers climate simulations to be independent and equally plausible, and it uses the ensemble mean and spread to define climate model uncertainty. Research has shown that the average of

equally weighted projections outperforms single models with respect to simulating mean climatic patterns (Chen et al., 2017; Reichler and Kim, 2008). However, this approach may be less effective for the CMIP6 ensemble, as the validity of some simulations is under question (Hausfather et al., 2022).

The CMIP6 ensemble includes a subset of “hot” models that predict greater warming than previous predictions made by CMIP5 (e.g., Kreienkamp et al., 2020). These hot models have a climate sensitivity that exceeds the expected plausible range, which is based on observations and our understanding of planetary physics. They also exhibit a higher equilibrium climate sensitivity (ECS), a measure of the steady-state temperature increase in the event of doubled carbon dioxide (CO<sub>2</sub>) concentrations in the atmosphere (Flynn and Mauritsen, 2020; Zelinka et al., 2020). The range of ECS values in CMIP6 models has increased to 1.8–5.6 °C compared with 2.1–4.7 °C in CMIP5, with an increase in the multi-model mean of 3.9 °C in CMIP6 from 3.3 °C in CMIP5 (Zelinka et al., 2020).

However, a plethora of evidence based on observations and our understanding of planetary physics indicates that we can confidently restrict the likely range of future warming trend and, more importantly, give less weight to extreme estimates (Liang et al., 2020; Tokarska et al., 2020). Recently, more research has been focused on constraining the ECS based on historical and paleoclimatic data (Knutti et al., 2017; Sherwood et al., 2020) or emergent constraints (Cox et al., 2018; Nijssen et al., 2020; Shiogama et al., 2022b). For example, Sherwood et al. (2020) used multiple lines of evidence and concluded that the likely (with a 66 % chance) ECS value is between 2.6 and 4.1 °C. Consequently, the most recent reports published by the Intergovernmental Panel on Climate Change (IPCC) have narrowed the likely ECS range to 2.5–4 °C (IPCC, 2023). It should be noted that the uncertainty surrounding the cooling impact (both direct and indirect) of aerosols on radiative forcing poses challenges in constraining future warming estimates (Bellouin et al., 2020; Forster et al., 2013; Smith et al., 2021). In essence, the current historical measurements do not provide a clear understanding of whether we are in a high-sensitivity, fast-warming scenario accompanied by strong contemporary aerosol cooling or if the situation is the opposite.

Climate change impact studies that include models with a high ECS may be biased and may overestimate the magnitude of impacts (Hausfather et al., 2022). Using the full ensemble of CMIP6 projections without restricting the hot models may no longer be the most appropriate option for impact studies (Ribes et al., 2021). Incorporating climate models with high sensitivity into impact studies may potentially lead to an overestimation of the overall economic consequences arising from future climate changes (Shiogama et al., 2022a). For instance, Shiogama et al. (2021) proposed a subset selection method that involved screening out hot models as the first step. On the other hand, Palmer et al. (2023) found that models with a higher sensitivity better represent

some key climatic processes over Europe. While they were unable to provide robust physical explanations for their findings, it is worth noting that hot models may provide valuable information at the regional scale and that this information may be more important than the global warming trend for impact modelers, adding to the complexity of selecting models for regional impact studies.

The decision to weight climate models for impact studies remains controversial, but it is difficult to ignore the potential pitfalls of using hot models in these studies (Hausfather et al., 2022). This study aims to evaluate how including or excluding hot models in a multi-model ensemble affects the results of a large-scale hydrological climate change impact study. This influence is measured in terms of the magnitude and uncertainty in various streamflow metrics for 3107 North American catchments.

## 2 Materials and methods

The data for this study were obtained from the HYSETS database, which contains hydrometeorological data from various sources for over 14 000 catchments in North America (Arsenault et al., 2020b). The database includes all necessary data for the reference period of this study, including catchment boundaries (in the form of shapefiles), streamflow observations, weather observations (from stations as well as multiple gridded and reanalysis datasets), and static catchment descriptors (such as area, slope, elevation, land-use fractions, and soil properties). This study used the ERA5 reanalysis dataset, which was found to be a reliable alternative to gauge observations in a previous large-scale comparison study over the same study area (Tarek et al., 2020), for meteorological data. To ensure representativeness, a subset of HYSETS catchments were selected using filters. First, catchments with drainage areas below 500 km<sup>2</sup> were excluded, as daily hydrological models would be inappropriate for modeling hydrological processes at smaller scales. Next, catchments required at least 10 years of data to ensure sufficient data to successfully calibrate hydrological models and bias-correct climate models. Overall, 3107 catchments were retained.

Table 1 presents the list of 19 CMIP6 GCMs selected for this study. This list includes five hot models, defined by an ECS greater than 4.1. These models are as follows: CanESM5 (ECS of 5.62), NESM3 (ECS of 4.68), IPSL-CM6A-LR (ECS of 4.52), EC-Earth3-Veg (ECS of 4.3), and EC-Earth3 (ECS of 4.2). This study will be able to compare the uncertainty generated by the entire ensemble (19 models) to that of a reduced ensemble (14 models) obtained by removing the 5 hot models.

The impact study in this paper uses a traditional top-down hydroclimatic modeling chain consisting of one shared socioeconomic pathway (SSP8.5), 19 CMIP6 GCMs, one bias correction method, and one hydrological model. The study

**Table 1.** The 19 GCMs selected in this study and their corresponding ECS values.

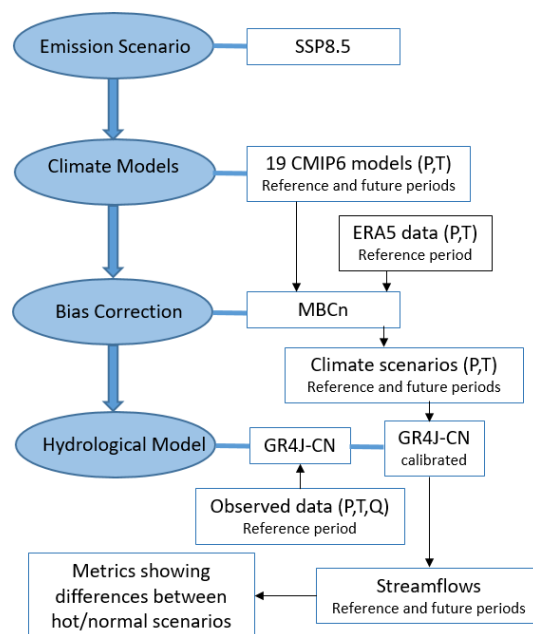
| GCM           | ECS               |
|---------------|-------------------|
| CANESM5       | 5.62 <sup>a</sup> |
| NESM3         | 4.68 <sup>a</sup> |
| IPSL-CM6A-LR  | 4.52 <sup>a</sup> |
| EC-Earth3-Veg | 4.3 <sup>a</sup>  |
| EC-Earth3     | 4.2 <sup>a</sup>  |
| ACCESS-ESM1-5 | 3.88 <sup>b</sup> |
| GFDL-CM4_gr1  | 3.89 <sup>b</sup> |
| GFDL-CM4_gr2  | 3.89 <sup>b</sup> |
| MRI-ESM2-0    | 3.14 <sup>a</sup> |
| MPI-ESM1-2-LR | 3.02 <sup>b</sup> |
| BCC-CSM2-MR   | 3.01 <sup>a</sup> |
| MPI-ESM1-2-HR | 2.98 <sup>b</sup> |
| FGOALS-g3     | 2.87 <sup>b</sup> |
| GFDL-ESM4     | 2.62 <sup>a</sup> |
| NorESM2-LM    | 2.60 <sup>a</sup> |
| MIROC6        | 2.57 <sup>a</sup> |
| NorESM2-MM    | 2.49 <sup>b</sup> |
| INM-CM5-0     | 1.92 <sup>a</sup> |
| INM-CM4-8     | 1.83 <sup>a</sup> |

<sup>a</sup> ECS values were taken from Tokarska et al. (2020). <sup>b</sup> ECS values were taken from Hausfather et al. (2022).

focuses solely on GCM uncertainty and does not consider other components, such as alternative SSPs, bias correction methods, or hydrological models, which would add uncertainty to future projections. These have been explored in previous studies (e.g., Wilby and Harris, 2006; Chen et al., 2011; Giuntoli et al., 2018; Troin et al., 2022) and are outside the scope of this work. The reference period is based on the 1971–2000 time frame, while the future climate is based on 2070–2099.

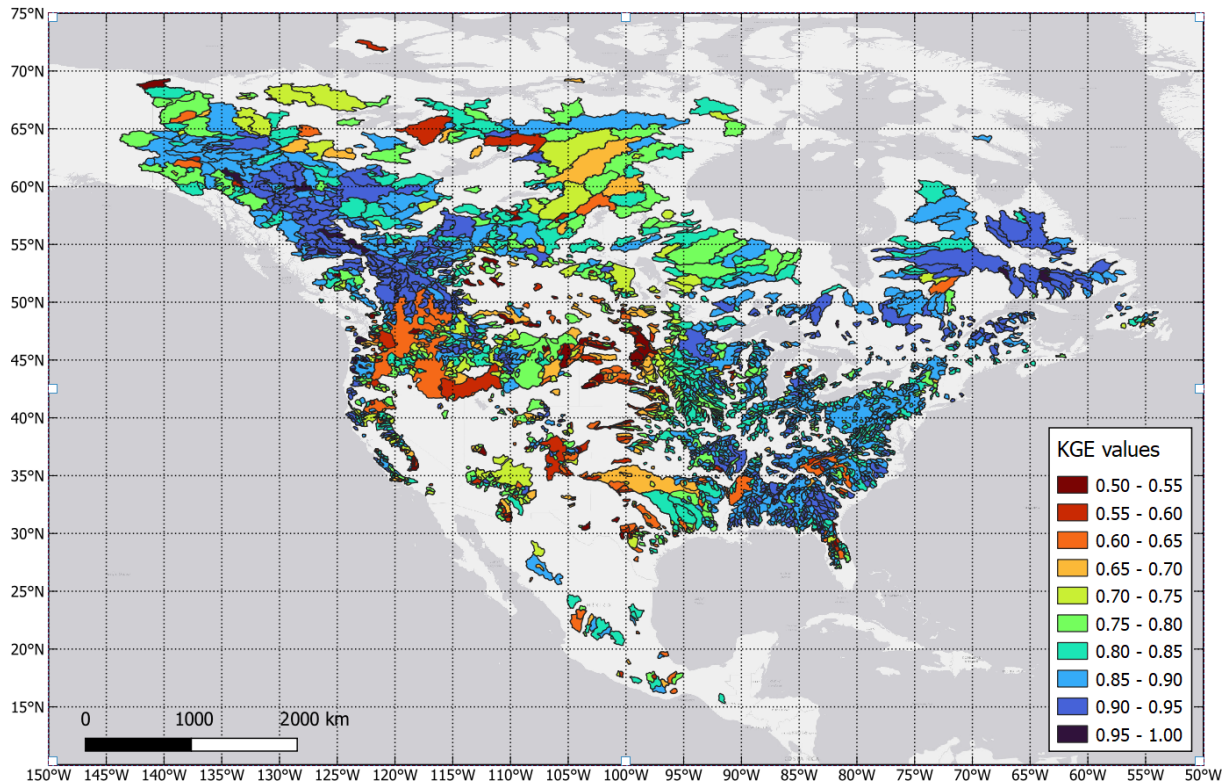
Figure 1 illustrates the methodological framework for each study catchment (Arsenault et al., 2020a). Precipitation and temperature data are first extracted from 19 CMIP6 climate models under the SSP8.5 scenario for both the reference and future periods. Using precipitation and temperature from the ERA5 reanalysis over the reference period, climate data are then bias-corrected using the multivariate bias correction (MBCn) method. These bias-corrected climate scenarios are subsequently employed as inputs for a calibrated hydrological model to compute streamflows. These computed streamflows are then employed to examine the impact of including (or not including) hot models in the impact study, using a set of defined metrics. Further details are provided below.

Climate models are mathematical representations of the Earth’s climate system, based on current understanding of its physics and chemistry. They are formulated using simplifying assumptions and parameterizations but may not fully capture the complexity of the real climate system due to limited

**Figure 1.** Methodological framework performed for each of the study catchments.

observations and understanding. As a result, climate models can be biased when compared with observations, due to factors such as model resolution, errors in reference datasets, and sensitivity to initial conditions. To ensure realistic impact simulations in impact studies, it is important to bias-correct climate model outputs. In this work, Cannon’s (2018) *N*-dimensional MBCn method was used to correct biases in daily precipitation and temperature. MBCn is considered the most advanced and efficient quantile-based multivariate bias correction method, as reported by studies such as Chen et al. (2018), Su et al. (2020), and Cannon et al. (2020). MBCn transfers the distribution of observational data to the corresponding distribution from the climate model while preserving the projection trends of the climate model simulation crucial for climate change impact studies (Maraun, 2016). No downscaling was performed because this study was conducted at the catchment scale.

In this study, the GR4J lumped rainfall–runoff model (Perin et al., 2003) was chosen to simulate streamflows. The model was selected due to the large number of catchments, which made it infeasible to use more complex, distributed models. Additionally, lumped models use averaged temperature and precipitation at the catchment scale, which is more consistent with the scale of GCMs, eliminating the need for downscaling. Lumped models have been shown to perform well with respect to simulating streamflows at catchment outlets (e.g., dos Santos et al., 2018; Reed et al., 2004). The GR4J model is simple, efficient, and shows high performance compared with other lumped conceptual models. It uses precipitation, potential evapotranspiration (PET), and catchment



**Figure 2.** Study catchment location. The color scale corresponds to the hydrological model Kling–Gupta efficiency (KGE) calibration score over the reference period. Only catchments with available data, a KGE values higher than 0.5, and an area larger than 500 km<sup>2</sup> were selected.

surface area as inputs. To account for snow accumulation in some catchments, the GR4J model is linked with the CemaNeige snow module (Valéry et al., 2014), resulting in a six-parameter model (GR4J-CN). The GR4J-CN model combination has been used in many studies, including climate change impact studies, and has been shown to perform well under a wide range of conditions (e.g., Riboust et al., 2019; Tarek et al., 2020; Wang et al., 2019). The calibration was performed using the Kling–Gupta efficiency (KGE) metric. The KGE metric (Gupta et al., 2009) directly combines the bias, ratio of variance, and correlation into a single metric. It provides a more robust and refined assessment of model performance when calibrating hydrological models, addressing the drawbacks of the Nash–Sutcliffe efficiency (NSE) metric (Nash and Sutcliffe, 1970; Knoben et al., 2019). Figure 2 presents the location of the 3107 retained catchments, which all have a KGE calibration value above 0.5.

The hydroclimatic modeling chain described above generated 19 different 30-year time series of daily streamflow for the 2070–2099 future period, each corresponding to one of the 19 GCMs listed in Table 1. Three streamflow metrics were extracted from each 30-year time series, representing mass balance ( $Q_{\text{mean}}$ ) and high ( $Q_{\text{max}}$ ) and low ( $Q_{\text{min}}$ ) flows:

- $Q_{\text{mean}}$  was obtained by averaging daily streamflow over the 30-year period;

- $Q_{\text{max}}$  was obtained by averaging the 30 annual maximum simulated streamflows;
- $Q_{\text{min}}$  was obtained by averaging the 30 annual minimum simulated streamflows.

These metrics will be used to assess the impact of removing hot climate models across a range of flow conditions.

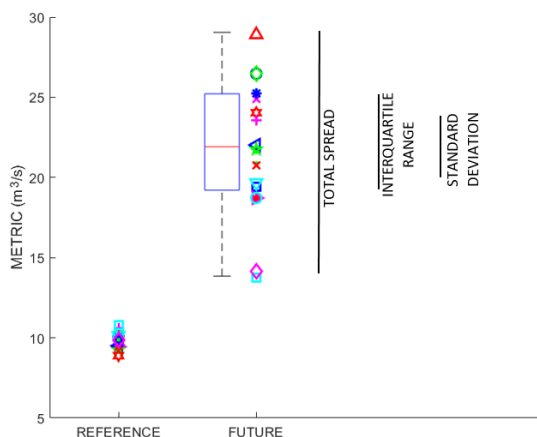
Figure 3 presents the three dispersion metrics used in this study to compare the spread (or uncertainty) of future projections of streamflow metrics. For the three streamflow metrics, 19 values from the original ensemble and 14 from the reduced ensemble for both the reference and future periods are extracted. The spread of the streamflow projections over the reference period is small, but it is not zero due to imperfect bias correction and the hydrology model’s strong nonlinear response to precipitation and temperature inputs. The spread is comparatively much larger in the future period, mainly due to differences in the sensitivity and structure of the climate models.

Total spread (TS) is defined as the full range of future streamflow responses:

$$TS = \text{metric}_{\text{max}} - \text{metric}_{\text{min}}. \quad (1)$$

The interquartile range (IQR) is defined as the distance between the 75th and 25th quantiles of the distribution, as shown by the blue rectangle in the box plot in Fig. 3.





**Figure 3.** Representation of the dispersion metrics used in this paper. Each marker represents one of the 19 climate models. METRIC refers to  $Q_{\text{mean}}$ ,  $Q_{\text{max}}$ , or  $Q_{\text{min}}$ , all in units of cubic meters per second ( $\text{m}^3 \text{s}^{-1}$ ).

$$\text{IQR} = Q_{75} - Q_{25} \quad (2)$$

Finally, the standard deviation ( $\sigma$ ) is the standard mathematical measure of dispersion. In the case of a normal distribution, the standard deviation and interquartile range are perfectly correlated, but this may not be the case for a skewed distribution.

All three metrics have units of cubic meters per second ( $\text{m}^3 \text{s}^{-1}$ ) and are, therefore, dependent on catchment size and, to a lesser extent, mean annual precipitation. To account for this, the metrics will be presented in a nondimensional form:

$$\text{TS}_{\text{nd}} = \frac{\text{TS}_{14}}{\text{TS}_{19}}, \quad (3)$$

where  $\text{TS}_{19}$  and  $\text{TS}_{14}$  represent the total spread for the full and reduced ensemble, respectively.  $\text{TS}_{\text{nd}}$  varies between 0 and 1:  $\text{TS}_{\text{nd}} = 1$  means that no reduction in total spread was obtained by removing the five warm models from the ensemble, whereas  $\text{TS}_{\text{nd}} = 0$  signifies that the total spread of the reduced ensemble has been totally eliminated.

Similarly, for the interquartile range ratio, we find the following:

$$\text{IQR}_{\text{nd}} = \frac{\text{IQR}_{14}}{\text{IQR}_{19}}. \quad (4)$$

However, in this case, the potential values vary in the  $0-\infty$  range. More practically, a value below 1 indicates that the IQR has been reduced by removing the five hot models from the ensemble, whereas a value larger than 1 shows the opposite. The latter is possible if the removed models are somewhat close to the median of the ensemble.

Finally, for the standard deviation the following ratio is used:

$$\sigma_{\text{nd}} = \frac{\sigma_{14}}{\sigma_{19}}, \quad (5)$$

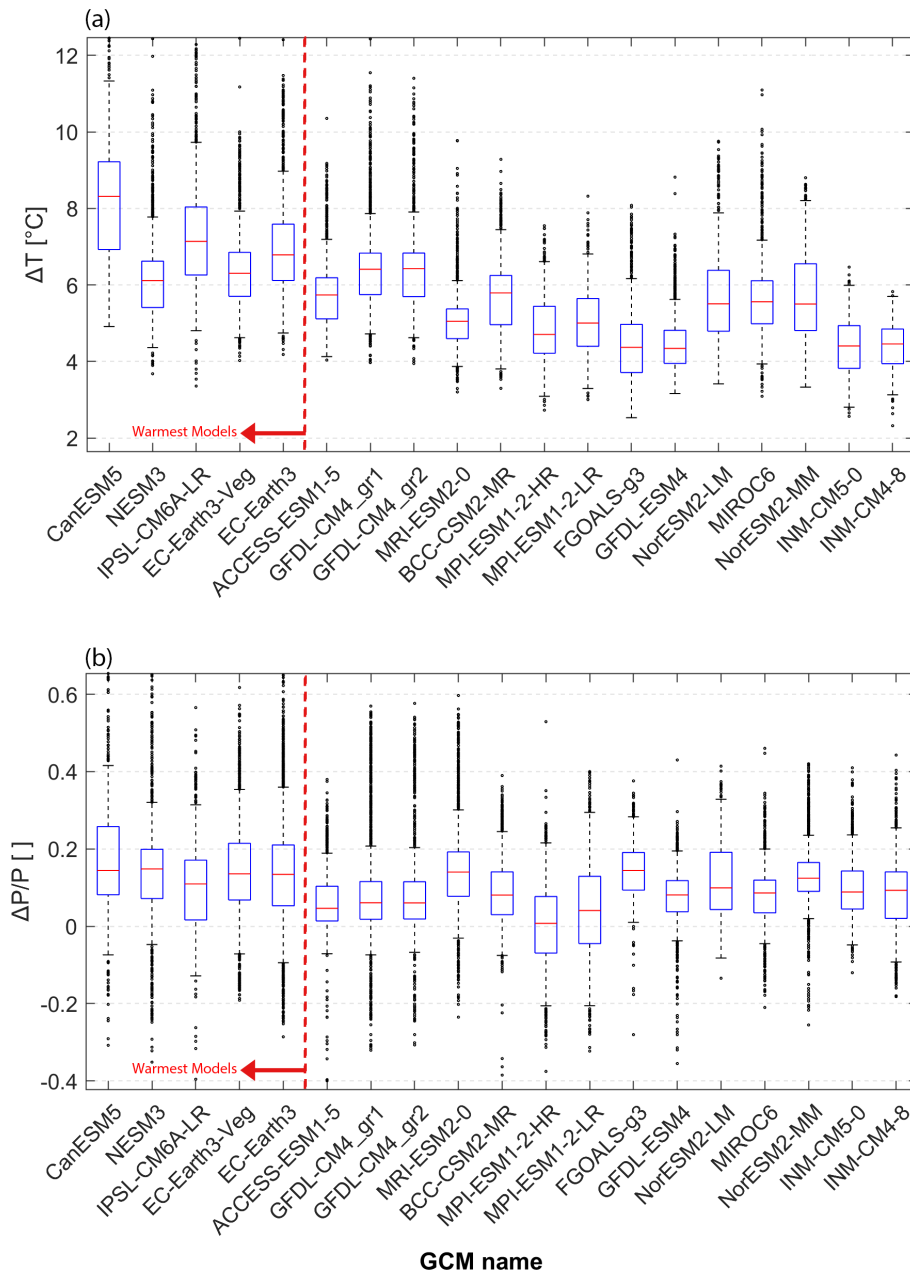
where a value below 1 indicates a smaller standard deviation for the reduced ensemble, and the opposite for a value above 1.  $\sigma_{\text{nd}}$  has the same possible range of values as  $\text{IQR}_{\text{nd}}$  ( $0-\infty$ ).

### 3 Results

Figure 4a presents the box plots of projected temperature increases for each of the 3107 catchments and for each climate model. The box plots provide a visual representation of key elements of the temperature increase distribution. The median of the distribution is shown as the red line near the center of the blue rectangle, which delimits the interquartile range ( $Q_{75}$  and  $Q_{25}$  for the upper and lower ends of the rectangle, respectively). The whiskers represent the 2.5th and 97.5th quantiles of the distribution, providing a 95 % coverage of the dataset. Quantiles below 2.5 and above 97.5 are shown as dots. Results indicate that the distribution of the projected temperature increases generally follows the same order as the ECS values presented in Table 1. However, there are some differences, which are not unexpected as global-scale ECS values are compared to regional-scale  $\Delta T$  values. The five hot models are ranked as the first, second, third, fifth, and sixth hottest regional models based on median values (considering that GFDL-CM4\_gr1 and GFDL-CM4\_gr2 – fourth and fifth, respectively – are actually the same model with different spatial resolutions).

Figure 4b presents the box plots of the projected changes in relative precipitation between the future and reference periods  $\left(\frac{P_{\text{fut}} - P_{\text{ref}}}{P_{\text{ref}}}\right)$ . The box plots depict the distribution of the projected precipitation changes for each of the 3107 catchments. Results indicate that the hot models, identified by their ECS values, are also among the models with the largest projected changes in relative precipitation. Specifically, the five hot models are all within the group of the eight wettest models. The models with more modest increases in precipitation (e.g., MPI-ESM and ACCESS) are also among the cooler models. This trend is expected, as a warmer atmosphere can hold more moisture (up to  $7 \% \text{ } ^\circ\text{C}^{-1}$ , according to the Clausius–Clapeyron relationship), leading to more precipitation. Increased precipitation may mitigate the anticipated impacts of warmer models, such as increased evapotranspiration.

In order to show regional patterns related to Fig. 4, Fig. 5 displays the mean  $\Delta T$  (Fig. 4a) and mean  $\Delta P/P$  (Fig. 4b) ratios between hot models and normal models. For temperature, a red color indicates that hot models are warmer than the other models on average. For precipitation, blue colors highlight increased precipitation in the hot models compared with the normal models. Overall, the hot global models exhibit a systematically larger temperature increase over the entire study domain. The hot models mostly exhibit increased precipitation compared with the normal models. However, the Pacific coast of the US as well as some catchments in

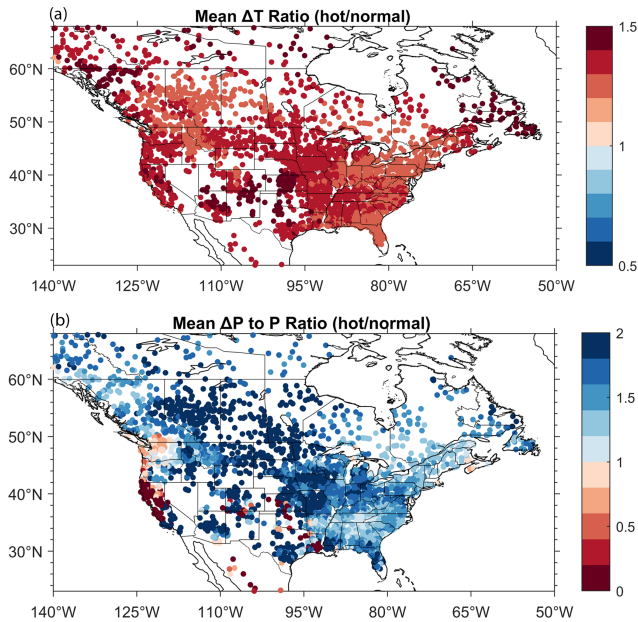


**Figure 4.** (a) Distribution of projected temperature increase ( $\Delta T$ ) and (b) projected relative annual precipitation increase ( $\Delta P/P$ ) for the 19 selected CMIP6 models for the 2070–2099 future period compared with the 1971–2000 reference period. Each box plot represents the distribution of the projected increases for the 3107 study catchments. The climate models are ordered in terms of their global-scale ECS values, starting with the largest on the left. The box plot whiskers correspond to the 2.5th and 97.5th quantiles, and a few catchment that were beyond the y axis limits are not shown.

the Southwestern US exhibit a decrease in precipitation according to the hot models. These observations underscore the regional variability in temperature and precipitation patterns when comparing hot and normal models.

Figure 6 presents the ratio of mean projected streamflow changes (hot models/normal models) for  $Q_{\text{mean}}$ ,  $Q_{\text{max}}$ , and  $Q_{\text{min}}$ . A blue color indicates larger projected streamflows by the hot models. Results show spatial patterns that differ

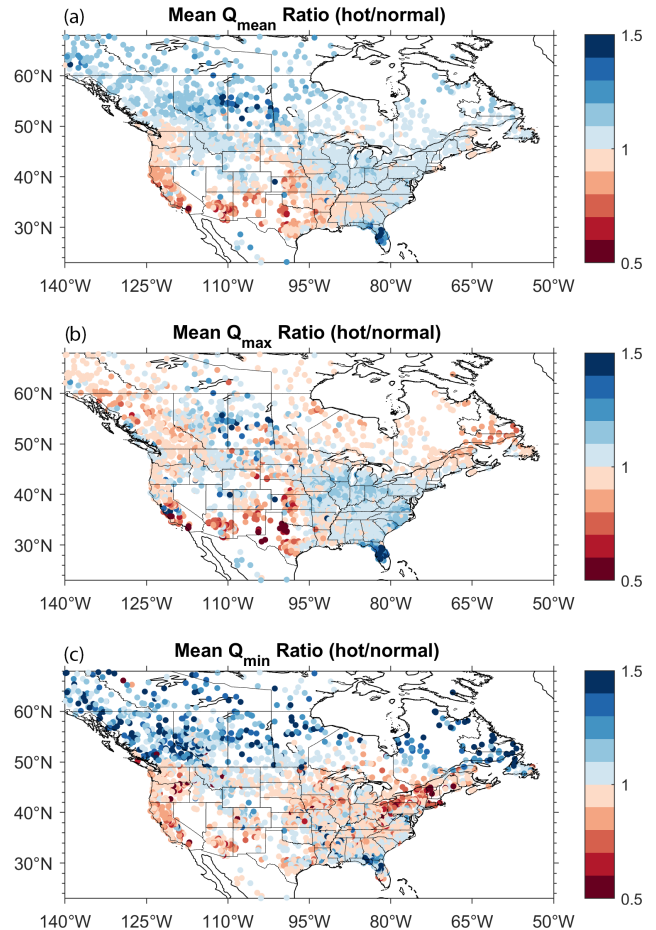
depending on the streamflow metrics. Hot models project higher mean flows over most of the study domain, except in the southwestern regions, where increased evapotranspiration nullifies potential increases in precipitation. For  $Q_{\text{max}}$ , increases are mostly localized in the Eastern US, whereas  $Q_{\text{min}}$  values are widely increasing in Canada and mostly decreasing in the US.



**Figure 5.** Mean  $\Delta T$  (a) and  $\Delta P/P$  (b) ratios (hot models to normal models). For  $\Delta T$ , a red color indicates that hot models are warmer, on average, than their normal (non-hot) counterparts. For  $\Delta P/P$ , a blue color shows that hot models are wetter than their normal (non-hot) counterparts. The graphs represent the differences computed between the future and reference periods.

Figure 7 presents the  $TS_{nd}$  for mean ( $Q_{mean}$ ), annual max ( $Q_{max}$ ), and min ( $Q_{min}$ ) streamflow obtained by removing the 5 hot models from the 19-member ensemble. A dark red color indicates no reduction in TS with the reduced ensemble, whereas lighter colors indicate a reduction. It can be seen that there is a clear spatial pattern that is relatively similar for all three streamflow metrics. The largest reductions in TS are seen in the northern regions as well as in the US Southeast, and along the US Pacific coast for  $Q_{mean}$  and  $Q_{min}$ . For all other regions of the US, no reduction in TS is observed. The reduced spread observed in the northern regions is smaller for  $Q_{max}$ . Despite these trends, a lot of variability remains present, with neighboring catchments sometimes showing contrasting behavior. More specifically, 57.0% of the catchments see a decrease in TS for  $Q_{mean}$ , 53.3% see a decrease for  $Q_{max}$ , and 61.7% see a decrease for  $Q_{min}$ .

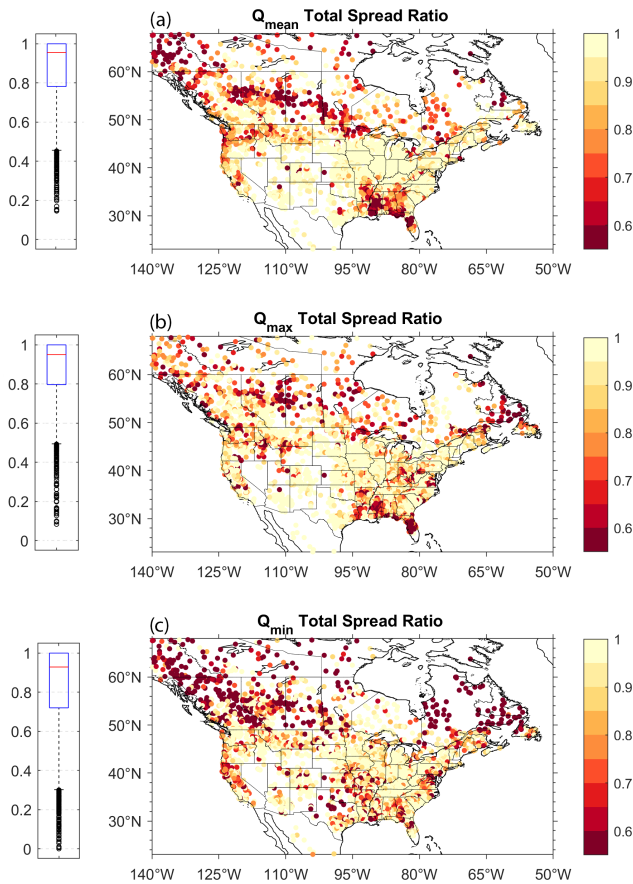
The data from Fig. 7 are shown in the form of box plots on the left side of each panel to better illustrate the range of TS reduction. This shows that the median  $TS_{nd}$  is relatively high for all three streamflow metrics: 0.96 for  $Q_{mean}$ , 0.95 for  $Q_{max}$ , and 0.93 for  $Q_{min}$ . This is primarily because a significant number of catchments see no reduction in TS (43%, 46.7%, and 38.3%, respectively). However, there is a significant reduction in TS observed in many catchments, and this decrease is strongly dependent on the geographical location of the catchments. Additionally, it can be seen that removing



**Figure 6.** Ratio of mean projected changes (hot models divided by normal models) for (a)  $Q_{mean}$ , (b)  $Q_{min}$ , and (c)  $Q_{max}$ . A blue color shows that hot models project larger streamflows than their normal (non-hot) counterparts.

the hot models has a greater impact on  $Q_{min}$  than on the other two metrics.

The  $TS_{nd}$  is heavily impacted by outliers and may not accurately represent the overall spread of models. Figure 8 presents the  $\sigma_{nd}$  for the three streamflow metrics. A red color ( $\sigma_{nd} > 1$ ) indicates that the model spread has increased following the removal of the hot models, whereas a blue color ( $\sigma_{nd} < 1$ ) corresponds to a decrease. Results indicate that removing the hot models consistently reduces  $\sigma_{nd}$  in Canada for  $Q_{mean}$  and  $Q_{min}$ , as well as to a lesser extent for  $Q_{max}$ . However, for the continental US, the results are more complex, with a lot of regional variability. Removing outlier models in the North Central, Northeastern, and Southwestern US results in an increase in  $\sigma_{nd}$  for both  $Q_{mean}$  and  $Q_{max}$ . Overall, as shown in the box plots in Fig. 8, removing the hot models likely reduces the spread in roughly two-thirds of catchments, while one-third see an increase. These values are larger than those obtained for TS. The trends seen in



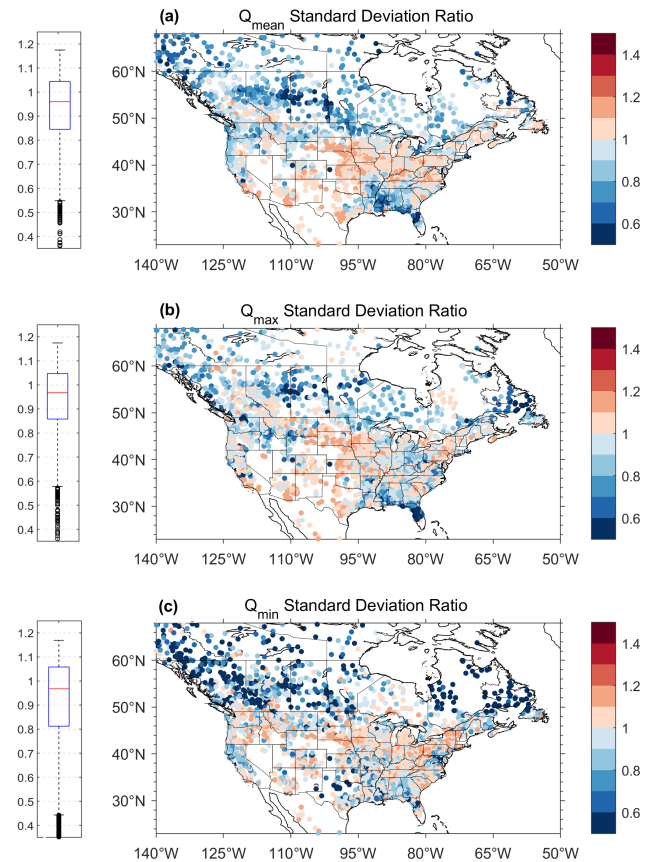
**Figure 7.** Total spread ratio ( $TS_{nd} = \frac{TS_{14}}{TS_{19}}$ ) for  $Q_{mean}$  (a),  $Q_{max}$  (b), and  $Q_{min}$  (c) resulting from the removal of the five hot models. Box plots are shown on the left side of each panel.

the  $IQR_{nd}$  are also very similar to those of  $\sigma_{nd}$  (see Figs. S1 and S2 in the Supplement).

#### 4 Discussion

Uncertainty is a key factor in assessing the impact of climate change. Different models and techniques, including various climate models, can lead to diverse climate projections and scenarios. Climate change interacts with other stressors, such as land-use change and population growth, in complex and unpredictable ways, making it important to accurately address uncertainty in climate impact studies to develop effective adaptation measures. Incorrectly representing uncertainty can lead to poor adaptation.

With the increased future temperatures, an intensification in the hydrological cycle is expected. However, this does not guarantee an automatic increase in water flow rates. This is because the rise in average temperature can also have a considerable impact on evapotranspiration. The outcome of these two factors working together is complex and varies based on the geographical location and primary cli-

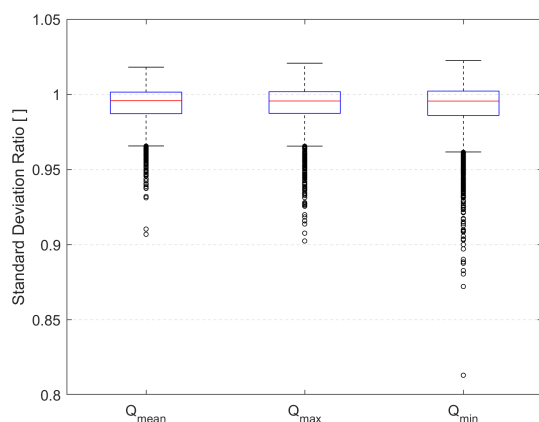


**Figure 8.** Standard deviation ratio ( $\sigma_{nd} = \frac{\sigma_{14}}{\sigma_{19}}$ ) for  $Q_{mean}$  (a),  $Q_{max}$  (b), and  $Q_{min}$  (c) resulting from the removal of the five hot models. Box plots are shown on the left side of each panel.

mate zones. The research paper indicates that regions characterized as hot tend to be associated with increased precipitation, further complicating the relationship between temperature and water flow.

Results show that removing the hot models is likely to reduce the spread of three streamflow metrics. Between 60% and 75% of catchments show a decrease in the spread of future streamflow projections, indicating that the hot models are outliers or are further from the mean than the average model. In such cases, keeping the hot models would result in an overestimation of future streamflow uncertainty. However, removing the hot models also led to an increase in the spread in certain regions, indicating overconfidence in the results. This means that, although they are outliers with respect to the ECS, the hot models may not be outliers with respect to impact studies. Generally, a reduction in spread was evident in northern regions, such as Canada and Alaska, as well as the coast of California and the Southeastern region of the US. Shiogama et al. (2022b) also concluded that the inclusion of hot models leads to an overestimation of annual mean precipitation increases in Alaska, Canada, and the





**Figure 9.** Box plots of the average standard deviation ratio for  $Q_{\text{mean}}$ ,  $Q_{\text{max}}$ , and  $Q_{\text{min}}$  resulting from the removal of five random models, after sampling 100 random combinations of five models.

western US, where there is a substantial decrease in the variability in streamflow metrics.

A reduction in the spread of future streamflow is expected when removing the hot models or reducing the number of climate models. A bootstrapping methodology was used to determine if the changes in spread were due to a reduction in the number of models. This was conducted by selecting a random sample of 14 (out of 19) models 100 times and computing the average standard deviation ratio. This was repeated for all catchments, and the aggregated results are shown in Fig. 9.

The results indicate that removing five random models results in a decrease in the standard deviation ratio almost 75 % of the time for all three streamflow metrics, but the median spread reduction ratio for this spread metric is extremely small (about 0.99 for all three streamflow metrics). This shows that removing the five hot models has a much larger impact than removing five random models. Therefore, the spread reduction observed in many catchments is not solely related to a reduction in the number of models.

At first glance, there is a strong physical reasoning for removing climate models with equilibrium climate sensitivity (ECS) values exceeding those expected from current data and the current understanding of planetary physics (Ribes et al., 2021; Shiogama et al., 2021). However, it should be noted that most impact studies are conducted at the regional or local scale, and these models may not be considered outliers at these scales. This study found that, although they may still be among the hottest in the study domain, globally hot models are not consistently the hottest, raising questions about whether their global behavior should automatically eliminate them from regional studies.

In this study, the climate performance of these models (such as their ability to represent climatic, hydroclimatic, or hydrological metrics) was not evaluated. The goal was to examine the impact of removing 5 hot models from a 19-

member ensemble. However, it is important to note that judging climate models based solely on their ECS values may result in the removal of models that have desirable characteristics at the regional scale (e.g., Palmer et al., 2023). Additionally, keeping hot models may also be useful from an impact perspective because they may provide a clearer picture of future changes, as internal variability is less likely to obscure changes. This is similar to the rationale behind using high-emission scenarios in impact studies, such as SSP8.5, even though they may not be considered realistic scenarios anymore (e.g., Hausfather and Peters, 2020). It is important to consider worst-case scenarios when analyzing potential outcomes, as high levels of greenhouse gas emissions or high model sensitivity, such as those projected in SSP8.5 or high-ECS models, are not unrealistic, even though they may be less likely. While it is valuable to consider these high-end scenarios, it should be made clear that they are indeed worst-case scenarios.

In this study, the question of whether to remove the hot models for impact studies is complex. Results showed that removing these models increased the future uncertainty in streamflow for about one-third of all catchments. This suggests that these hot outliers may not always be hydrological outliers when put through a hydrological modeling process. Hydrological models are well known for being highly nonlinear integrators of weather variables, such as temperature and precipitation, and these results align with findings from other studies that have demonstrated the complex relationship between climate model projections and hydrological projections (e.g., Chen et al., 2016; Ross and Najjar, 2019). The fact that the CMIP6 hot climate models tend to be wet models may also be a factor in these results, as increased evapotranspiration could be offset by increased precipitation, leading to somewhat average results for the wrong reasons.

The regional impact of model importance is also compared (see Figs. S3 and S4, which demonstrate the total spread ratio resulting from removing a single climate model and creating an 18-member ensemble). Removing CanESM5 leads to a clear reduction in the total spread in Alaska and the Yukon (for  $Q_{\text{mean}}$  and  $Q_{\text{min}}$ ) and in the Southeastern US for  $Q_{\text{max}}$ , indicating that CanESM5 is an outlier in these regions. Conversely, removing NESM3 does not result in significant decreases in spread over most of the study domain, as the high ECS value of NESM3 does not automatically translate into a correspondingly higher level of regional warming (see also Fig. 4), demonstrating that it is not an outlier in most regions. This underscores the strong regional differences among globally identified hot models.

The only uncertainty in this study is that originating from GCM/Earth system models. As stated earlier, in most impact studies, additional sources of uncertainty would also be incorporated. Additional greenhouse gas emission scenarios would be selected as well as other impact models (e.g., hydrology models). Downscaling and additional bias correction may be performed. These additional components

are likely to generate additional uncertainty which may, in some cases, dwarf that of climate models. As such, many of the differences observed between the original and reduced climate model ensembles in this paper may have little impact on the final uncertainty estimation. For example, for low flows, many studies have shown that most of the uncertainty lies within the hydrology models (e.g., Giuntoli et al., 2018; Krysanova et al., 2018; Trudel et al., 2017) and that removing climate models would have no impact on uncertainty.

The results show that there is no simple answer as to whether or not including hot models in climate change impacts studies. In the absence of any computational limitations, we would recommend using as many climate models as possible and then subsequently studying the impact of including/excluding hot models. If the selection of a subset of climate models is necessary (e.g., inability to use a large ensemble due to limited computational capability or the cost of running impact models) removing hot models may be a reasonable option. Evaluating climate model fitness for impact studies is a difficult endeavor, and, in addition to the ECS, additional performance metrics should also carefully be taken into account.

## 5 Conclusion

This study examines the impact of removing a subset of hot climate models on the spread of future projections of streamflow for 3107 North American catchments. Three streamflow metrics were considered: mean annual streamflow and the means of the respective annual maximum and minimum streamflow, over the reference period (1971–2000) and future period (2070–2099).

Hot climate models are determined based on their global equilibrium climate sensitivity (ECS), whereas impact studies typically focus on the local to regional scale. The hot climate models remain among the hottest in our regional evaluation, but they also tend to be among the wettest, potentially leading to a complex hydrological response.

Our research revealed mixed impacts of removing the hot climate models. A decrease in the variability in the projected streamflow metrics was generally observed in Canada and Alaska, the Southeast US, and the Pacific coast of the US. However, in other regions, removing the hot models resulted in no changes or, in some cases, even increases in the variability in the projected flows. This suggests that the hot models are not necessarily hydrological outliers, raising questions about using global performance metrics rather than regional metrics for model selection.

The findings of this study emphasize the importance of carefully selecting climate models and the potential risks of including inadequate models in impact studies. In the absence of constraints, it is recommended to use as many climate models as possible when determining impact uncertainty and to assess the impact of subsets of climate models

(based on high global equilibrium climate sensitivity or other performance metrics) a posteriori to evaluate the sensitivity of the impact model to climate model selection. These results highlight the need for further research on climate model fitness and the proper selection of model subsets for impact studies.

*Code and data availability.* The hydrometeorological data used in this study were obtained from the HYSETS database: <https://doi.org/10.17605/OSF.IO/RPC3W> (Arsenault et al., 2022). The CMIP6 GCM model outputs are available from the Earth System Grid Federation (ESGF) portal at Lawrence Livermore National Laboratory: <https://esgf-node.llnl.gov/search/cmip6/> (ESGF, 2022). The processed data and the codes used in this work are available from the authors upon reasonable request.

*Supplement.* The supplement related to this article is available online at: <https://doi.org/10.5194/hess-27-4355-2023-supplement>.

*Author contributions.* The experiments were designed by FB and carried out by MRA. The findings were analyzed and interpreted by MRA and FB. The paper was written by MRA and FB, with significant contributions from JLM and RA. JLM and RA also provided editorial feedback on the paper’s early drafts.

*Disclaimer.* Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

*Competing interests.* The contact author has declared that none of the authors has any competing interests.

*Financial support.* This research has been supported by the Natural Sciences and Engineering Research Council of Canada (grant no. RGPIN-2020-07242).

*Review statement.* This paper was edited by Efrat Morin and reviewed by Chris Huntingford and one anonymous referee.

## References

- Arsenault, R., Brissette, F., Chen, J., Guo, Q., and Dallaire, G.: NAC<sup>2</sup>H: The North American Climate Change and Hydroclimatology Data Set, *Water Resour. Res.*, 56, e2020WR027097, <https://doi.org/10.1029/2020WR027097>, 2020a.
- Arsenault, R., Brissette, F., Martel, J.-L., Troin, M., Lévesque, G., Davidson-Chaput, J., Gonzalez, M. C., Ameli, A., and Poulin, A.: A comprehensive, multisource database for hydrometeorological modeling of 14,425 North American watersheds, *Sci. Data*, 7, 243, <https://doi.org/10.1038/s41597-020-00583-2>, 2020b.
- Arsenault, R., Brissette, F., Martel, J.-L., Troin, M., Lévesque, G., Davidson-Chaput, J., Castañeda Gonzalez, M., Ameli, A., and Poulin, A.: HYSSETS – A 14425 watershed Hydrometeorological Sandbox over North America, OSF [data set], <https://doi.org/10.17605/OSF.IO/RPC3W>, 2022.
- Bellouin, N., Quaas, J., Gryspeerdt, E., Kinne, S., Stier, P., Watson-Parris, D., Boucher, O., Carslaw, K. S., Christensen, M., Daniau, A.-L., Dufresne, J.-L., Feingold, G., Fiedler, S., Forster, P., Gettelman, A., Haywood, J. M., Lohmann, U., Malavelle, F., Mauritsen, T., McCoy, D. T., Myhre, G., Mühlensstädt, J., Neubauer, D., Possner, A., Rugenstein, M., Sato, Y., Schulz, M., Schwartz, S. E., Sourdeval, O., Storelvmo, T., Toll, V., Winker, D., and Stevens, B.: Bounding Global Aerosol Radiative Forcing of Climate Change, *Re. Geophys.*, 58, e2019RG000660, <https://doi.org/10.1029/2019RG000660>, 2020.
- Cannon, A. J.: Multivariate quantile mapping bias correction: An  $N$ -dimensional probability density function transform for climate model simulations of multiple variables, *Clim. Dynam.*, 50, 31–49, <https://doi.org/10.1007/s00382-017-3580-6>, 2018.
- Cannon, A. J., Piani, C., and Sippel, S.: Bias correction of climate model output for impact models, in: *Climate Extremes and Their Implications for Impact and Risk Assessment*, Elsevier, 77–104, <https://doi.org/10.1016/B978-0-12-814895-2.00005-7>, 2020.
- Chen, J., Brissette, F. P., Poulin, A., and Leconte, R.: Overall uncertainty study of the hydrological impacts of climate change for a Canadian watershed, *Water Resour. Res.*, 47, W12509, <https://doi.org/10.1029/2011WR010602>, 2011.
- Chen, J., Brissette, F. P., and Lucas-Picher, P.: Transferability of optimally-selected climate models in the quantification of climate change impacts on hydrology, *Clim. Dynam.*, 47, 3359–3372, <https://doi.org/10.1007/s00382-016-3030-x>, 2016.
- Chen, J., Brissette, F. P., Lucas-Picher, P., and Caya, D.: Impacts of weighting climate models for hydrometeorological climate change studies, *J. Hydrol.*, 549, 534–546, <https://doi.org/10.1016/j.jhydrol.2017.04.025>, 2017.
- Chen, J., Li, C., Brissette, F. P., Chen, H., Wang, M., and Essou, G. R.: Impacts of correcting the inter-variable correlation of climate model outputs on hydrological modeling, *J. Hydrol.*, 560, 326–341, <https://doi.org/10.1016/j.jhydrol.2018.03.040>, 2018.
- Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichet, T., Gao, X., Gutowski Jr., W. J., Johns, T., Krinner, G., Shongwe, M., Weaver, A. J., Wehner, M., Allen, M. R., Andrews, T., Beyeler, U., Bitz, C. M., Bony, S., Booth, B. B. B., Brooks, H. E., Brovkin, V., Browne, O., Brutel-Vuilmet, C., Cane, M., Chadwick, R., Cook, E., Cook, K. H., Eby, M., Fasullo, J., Forest, C. E., Forster, P., Good, P., Goosse, H., Gregory, J. M., Hegerl, G. C., Hezel, P. J., Hodges, K. I., Holland, M. M., Huber, M., Joshi, M., Kharin, V., Kushnir, Y., Lawrence, D. M., Lee, R. W., Liddicoat, S., Lucas, C., Lucht, W., Marotzke, J., Masson-
- net, F., Matthews, H. D., Meinshausen, M., Morice, C., Otto, A., Patricola, C. M., Philippon, G., Rahmstorf, S., Riley, W. J., Saenko, O., Seager, R., Sedláček, J., Shaffrey, L. C., Shindell, D., Sillmann, J., Stevens, B., Stott, P. A., Webb, R., Zappa, G., Zickfeld, K., Joussaume, S., Mokssit, A., Taylor, K., and Tett, S.: Long-term Climate Change: Projections, Commitments and Irreversibility, in: *Climate Change 2013: The Physical Science Basis*, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, UK and New York, NY, USA, 1029–1136, <https://www.ipcc.ch/report/ar5/wg1> (last access: 12 November 2022), 2013.
- Cox, P. M., Huntingford, C., and Williamson, M. S.: Emergent constraint on equilibrium climate sensitivity from global temperature variability, *Nature*, 553, 319–322, <https://doi.org/10.1038/nature25450>, 2018.
- dos Santos, F. M., de Oliveira, R. P., and Mauad, F. F.: Lumped versus distributed hydrological modeling of the Jacaré-Guaçu Basin, Brazil, *J. Environ. Eng.*, 144, 04018056, [https://doi.org/10.1061/\(ASCE\)EE.1943-7870.0001397](https://doi.org/10.1061/(ASCE)EE.1943-7870.0001397), 2018.
- ESGF – Earth System Grid Federation: CMIP6 GCM data, <https://esgf-node.llnl.gov/search/cmip6/> (last access: 18 July 2022), 2022.
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geosci. Model Dev.*, 9, 1937–1958, <https://doi.org/10.5194/gmd-9-1937-2016>, 2016.
- Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., Collins, W. D., Gier, B. K., Hall, A. D., Hoffman, F. M., Hurtt, G. C., Jahn, A., Jones, C. D., Klein, S. A., Krasting, J. P., Kwiatkowski, L., Lorenz, R., Maloney, E., Meehl, G. A., Pendergrass, A. G., Pincus, R., Ruane, A. C., Russell, J. L., Sanderson, B. M., Santer, B. D., Sherwood, S. C., Simpson, I. R., Stouffer, R. J., and Williamson, M. S.: Taking climate model evaluation to the next level, *Nat. Clim. Change*, 9, 102–110, <https://doi.org/10.1038/s41558-018-0355-y>, 2019.
- Flynn, C. M. and Mauritsen, T.: On the climate sensitivity and historical warming evolution in recent coupled model ensembles, *Atmos. Chem. Phys.*, 20, 7829–7842, <https://doi.org/10.5194/acp-20-7829-2020>, 2020.
- Forster, P. M., Andrews, T., Good, P., Gregory, J. M., Jackson, L. S., and Zelinka, M.: Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models, *J. Geophys. Res.-Atmos.*, 118, 1139–1150, <https://doi.org/10.1002/jgrd.50174>, 2013.
- Giuntoli, I., Villarini, G., Prudhomme, C., and Hannah, D. M.: Uncertainties in projected runoff over the conterminous United States, *Climatic Change*, 150, 149–162, <https://doi.org/10.1007/s10584-018-2280-5>, 2018.
- Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, 377, 80–91, <https://doi.org/10.1016/j.jhydrol.2009.08.003>, 2009.
- Hausfather, Z. and Peters, G. P.: Emissions – the ‘business as usual’ story is misleading, *Nature*, 577, 618–620, <https://doi.org/10.1038/d41586-020-00177-3>, 2020.
- Hausfather, Z., Marvel, K., Schmidt, G. A., Nielsen-Gammon, J. W., and Zelinka, M.: Climate simulations:

- recognize the ‘hot model’ problem, *Nature*, 605, 26–29, <https://doi.org/10.1038/d41586-022-01192-2>, 2022.
- Hirabayashi, Y., Tanoue, M., Sasaki, O., Zhou, X., and Yamazaki, D.: Global exposure to flooding from the new CMIP6 climate model projections, *Sci. Rep.*, 11, 3740, <https://doi.org/10.1038/s41598-021-83279-w>, 2021.
- IPCC: Climate Change 2014: Synthesis Report, in: Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, IPCC, Geneva, Switzerland, <https://www.ipcc.ch/report/ar5/syr> (last access: 18 October 2022), 2014.
- IPCC: Intergovernmental Panel on Climate Change, Technical Summary, in: Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, 35–144, <https://doi.org/10.1017/9781009157896.002>, 2023.
- Karlsson, I. B., Sonnenborg, T. O., Refsgaard, J. C., Trolle, D., Børgesen, C. D., Olesen, J. E., Jeppesen, E., and Jensen, K. H.: Combined effects of climate models, hydrological model structures and land use scenarios on hydrological impacts of climate change, *J. Hydrol.*, 535, 301–317, <https://doi.org/10.1016/j.jhydrol.2016.01.069>, 2016.
- Knoben, W. J. M., Freer, J. E., and Woods, R. A.: Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores, *Hydrol. Earth Syst. Sci.*, 23, 4323–4331, <https://doi.org/10.5194/hess-23-4323-2019>, 2019.
- Knutti, R.: The end of model democracy?: An editorial comment, *Climatic Change*, 102, 395–404, <https://doi.org/10.1007/s10584-010-9800-2>, 2010.
- Knutti, R., Rugenstein, M. A. A., and Hegerl, G. C.: Beyond equilibrium climate sensitivity, *Nat. Geosci.*, 10, 727–736, <https://doi.org/10.1038/ngeo3017>, 2017.
- Kreienkamp, F., Lorenz, P., and Geiger, T.: Statistically Downscaled CMIP6 Projections Show Stronger Warming for Germany, *Atmosphere*, 11, 1245, <https://doi.org/10.3390/atmos11111245>, 2020.
- Krysanova, V., Donnelly, C., Gelfan, A., Gerten, D., Arheimer, B., Hattermann, F., and Kundzewicz, Z. W.: How the performance of hydrological models relates to credibility of projections under climate change, *Hydrolog. Sci. J.*, 63, 696–720, <https://doi.org/10.1080/02626667.2018.1446214>, 2018.
- Liang, Y., Gillett, N. P., and Monahan, A. H.: Climate Model Projections of 21st Century Global Warming Constrained Using the Observed Warming Trend, *Geophys. Res. Lett.*, 47, e2019GL086757, <https://doi.org/10.1029/2019GL086757>, 2020.
- Maraun, D.: Bias correcting climate change simulations – a critical review, *Curr. Clim. Change Rep.*, 2, 211–220, <https://doi.org/10.1007/s40641-016-0050-x>, 2016.
- Martel, J.-L., Brissette, F., Troin, M., Arsenaault, R., Chen, J., Su, T., and Lucas-Picher, P.: CMIP5 and CMIP6 Model Projection Comparison for Hydrological Impacts Over North America, *Geophys. Res. Lett.*, 49, e2022GL098364, <https://doi.org/10.1029/2022GL098364>, 2022.
- Meehl, G. A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J. F. B., Stouffer, R. J., and Taylor, K. E.: THE WCRP CMIP3 Multimodel Dataset: A New Era in Climate Change Research, *B. Am. Meteorol. Soc.*, 88, 1383–1394, <https://doi.org/10.1175/BAMS-88-9-1383>, 2007.
- Miara, A., Macknick, J. E., Vörösmarty, C. J., Tidwell, V. C., Newmark, R., and Fekete, B.: Climate and water resource change impacts and adaptation potential for US power supply, *Nat. Clim. Change*, 7, 793–798, <https://doi.org/10.1038/nclimate3417>, 2017.
- Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I – A discussion of principles, *J. Hydrol.*, 10, 282–290, [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6), 1970.
- Nijssen, F. J. M. M., Cox, P. M., and Williamson, M. S.: Emergent constraints on transient climate response (TCR) and equilibrium climate sensitivity (ECS) from historical warming in CMIP5 and CMIP6 models, *Earth Syst. Dynam.*, 11, 737–750, <https://doi.org/10.5194/esd-11-737-2020>, 2020.
- Palmer, T. E., McSweeney, C. F., Booth, B. B. B., Priestley, M. D. K., Davini, P., Brunner, L., Borchert, L., and Menary, M. B.: Performance-based sub-selection of CMIP6 models for impact assessments in Europe, *Earth Syst. Dynam.*, 14, 457–483, <https://doi.org/10.5194/esd-14-457-2023>, 2023.
- Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, *J. Hydrol.*, 279, 275–289, [https://doi.org/10.1016/S0022-1694\(03\)00225-7](https://doi.org/10.1016/S0022-1694(03)00225-7), 2003.
- Prajapati, S., Sabokruhie, P., Brinkmann, M., and Lindenschmidt, K.-E.: Modelling Transport and Fate of Copper and Nickel across the South Saskatchewan River Using WASP–TOXI, *Water*, 15, 2, <https://doi.org/10.3390/w15020265>, 2023.
- Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J., and Participants, D.: Overall distributed model intercomparison project results, *J. Hydrol.*, 298, 27–60, <https://doi.org/10.1016/j.jhydrol.2004.03.031>, 2004.
- Reichler, T. and Kim, J.: How Well Do Coupled Models Simulate Today’s Climate?, *B. Am. Meteorol. Soc.*, 89, 303–312, <https://doi.org/10.1175/BAMS-89-3-303>, 2008.
- Ribes, A., Qasmi, S., and Gillett, N. P.: Making climate projections conditional on historical observations, *Sci. Adv.*, 7, eabc0671, <https://doi.org/10.1126/sciadv.abc0671>, 2021.
- Riboust, P., Thirel, G., Le Moine, N., and Ribstein, P.: Revisiting a simple degree-day model for integrating satellite data: Implementation of SWE-SCA hysteresis, *J. Hydrol. Hydromech.*, 67, 70–81, <https://doi.org/10.2478/johh-2018-0004>, 2019.
- Ross, A. C. and Najjar, R. G.: Evaluation of methods for selecting climate models to simulate future hydrological change, *Climatic Change*, 157, 407–428, <https://doi.org/10.1007/s10584-019-02512-8>, 2019.
- Sabokruhie, P., Akomeah, E., Rosner, T., and Lindenschmidt, K.-E.: Proof-of-concept of a quasi-2d water-quality modelling approach to simulate transverse mixing in rivers, *Water*, 13, 307, <https://doi.org/10.3390/w13213071>, 2021.
- Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., Hegerl, G., Klein, S. A., Marvel, K. D., Rohling, E. J., Watanabe, M., Andrews, T., Braconnot, P., Bretherton, C. S., Foster, G. L., Hausfather, Z., von der Heydt, A. S., Knutti, R., Mauritsen, T., Norris, J. R., Proistosescu, C., Rugenstein, M., Schmidt, G. A., Tokarska, K. B., and Zelinka, M. D.: An Assessment of Earth’s Climate Sensitivity Using Multiple Lines of Evidence, *Rev. Geophys.*, 58, e2019RG000678, <https://doi.org/10.1029/2019RG000678>, 2020.
- Shiogama, H., Ishizaki, N. N., Hanasaki, N., Takahashi, K., Emori, S., Ito, R., Nakaegawa, T., Takayabu, I., Hijioka, Y., Takayabu, Y. N., and Shibuya, R.: Selecting CMIP6-Based Future Climate



- Scenarios for Impact and Adaptation Studies, SOLA, 17, 57–62, <https://doi.org/10.2151/sola.2021-009>, 2021.
- Shiogama, H., Takakura, J., and Takahashi, K.: Uncertainty constraints on economic impact assessments of climate change simulated by an impact emulator, *Environ. Res. Lett.*, 17, 124028, <https://doi.org/10.1088/1748-9326/aca68d>, 2022a.
- Shiogama, H., Watanabe, M., Kim, H., and Hirota, N.: Emergent constraints on future precipitation changes, *Nature*, 602, 612–616, <https://doi.org/10.1038/s41586-021-04310-8>, 2022b.
- Smith, C. J., Harris, G. R., Palmer, M. D., Bellouin, N., Collins, W., Myhre, G., Schulz, M., Golaz, J.-C., Ringer, M., Storelvmo, T., and Forster, P. M.: Energy Budget Constraints on the Time History of Aerosol Forcing and Climate Sensitivity, *J. Geophys. Res.-Atmos.*, 126, e2020JD033622, <https://doi.org/10.1029/2020JD033622>, 2021.
- Su, T., Chen, J., Cannon, A. J., Xie, P., and Guo, Q.: Multi-site bias correction of climate model outputs for hydro-meteorological impact studies: An application over a watershed in China, *Hydrol. Process.*, 34, 2575–2598, <https://doi.org/10.1002/hyp.13750>, 2020.
- Tabari, H.: Climate change impact on flood and extreme precipitation increases with water availability, *Sci. Rep.*, 10, 13768, <https://doi.org/10.1038/s41598-020-70816-2>, 2020.
- Tarek, M., Brissette, F. P., and Arsenault, R.: Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America, *Hydrol. Earth Syst. Sci.*, 24, 2527–2544, <https://doi.org/10.5194/hess-24-2527-2020>, 2020.
- Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F., and Knutti, R.: Past warming trend constrains future warming in CMIP6 models, *Sci. Adv.*, 6, eaaz9549, <https://doi.org/10.1126/sciadv.aaz9549>, 2020.
- Troin, M., Martel, J. L., Arsenault, R., and Brissette, F.: Large-sample study of uncertainty of hydrological model components over North America, *J. Hydrol.*, 609, 127766, <https://doi.org/10.1016/j.jhydrol.2022.127766>, 2022.
- Trudel, M., Doucet-Généreux, P. L., and Leconte, R.: Assessing river low-flow uncertainties related to hydrological model calibration and structure under climate change conditions, *Climate*, 5, 19, <https://doi.org/10.3390/cli5010019>, 2017.
- Valéry, A., Andréassian, V., and Perrin, C.: ‘As simple as possible but not simpler’: What is useful in a temperature-based snow-accounting routine? Part 2 – Sensitivity analysis of the Cemanige snow accounting routine on 380 catchments, *J. Hydrol.*, 517, 1176–1187, <https://doi.org/10.1016/j.jhydrol.2014.04.058>, 2014.
- Wang, H.-M., Chen, J., Xu, C.-Y., Chen, H., Guo, S., Xie, P., and Li, X.: Does the weighting of climate simulations result in a better quantification of hydrological impacts?, *Hydrol. Earth Syst. Sci.*, 23, 4033–4050, <https://doi.org/10.5194/hess-23-4033-2019>, 2019.
- Wilby, R. L. and Harris, I.: A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK, *Water Resour. Res.*, 42, W02419, <https://doi.org/10.1029/2005WR004065>, 2006.
- Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., Klein, S. A., and Taylor, K. E.: Causes of Higher Climate Sensitivity in CMIP6 Models, *Geophys. Res. Lett.*, 47, e2019GL085782, <https://doi.org/10.1029/2019GL085782>, 2020.
- Zhang, Y., Liu, H., Qi, J., Feng, P., Zhang, X., Liu, D. L., Marek, G. W., Srinivasan, R., and Chen, Y.: Assessing impacts of global climate change on water and food security in the black soil region of Northeast China using an improved SWAT-CO<sub>2</sub> model, *Sci. Total Environ.*, 857, 159482, <https://doi.org/10.1016/j.scitotenv.2022.159482>, 2023.