Peak Runoff Timing is Linked to Global Warming Trajectories

Donghui Xu[1*], Valeriy Y. Ivanov[1*], Xiuyuan Li[2], Tara J. Troy[3]

[1]Univeristy of Michigan-Ann Arbor

[2]Lehigh University

[3]Department of Civil Engineering, University of Victoria, Victoria, BC Canada

Corresponding authors: Donghui Xu (donghui.xu@pnnl.gov),

Valeriy Y. Ivanov (ivanov@umich.edu)

† Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI 48109, tel.: 734-730-8326.

Key Points:

- Climate multi-model ensemble projects change of peak annual runoff timing over the continental U.S. during the 21st century
- Spatial patterns of peak runoff timing earlier onset as well as delay are more pronounced for higher future greenhouse concentrations
- Springtime shifts in the dates of maximum snow accumulation and soil moisture wetness are associated with changes in peak annual runoff timing

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2021EF002083.

Abstract

The earth's hydroclimate is continuing to change, and the corresponding impacts on water resource space-time distribution need to be understood to mitigate their socioeconomic consequences. A variety of ecosystem services, transport processes, and human activities are synced with the *timing* of peak annual runoff. To understand the influence of changing hydroclimate on peak runoff dates across the continental U.S., we downscaled outputs of ten Global Circulation Models for different future scenarios. Our results quantify robust spatial patterns of both negative (up to 3-5 weeks) and positive (up to 2-4 weeks) shifts in the dates of peak annual runoff occurrence by the end of this century. In snowmelt-dominated areas, annual maxima are projected to shift to earlier dates due to the corresponding changes in snow accumulation timing. For regions in which the occurrence of springtime extreme soil wetness shifts to later time, we find that peak annual runoff is also projected to be delayed. These patterns of runoff timing change tend to be more pronounced for projections of higher greenhouse concentration in the future.

Plain Language Summary

The occurrence of peak annual runoff characterizes the major phase of watershed surface hydrology. Many natural dynamics and human activities are synced with the *timing* of its occurrence, ranging from ecosystem services and channel transport of sediments and contaminants to reservoir refilling and management. The sensitivity of peak annual runoff *timing* to changing hydroclimate remains unknown. In this work, we identify how peak annual runoff occurrence will change in the future over the continental U.S. using outputs of several climate models. Spatial patterns of the change show both earlier (by up to 3-5 weeks) and delayed (up to 2-4 weeks) occurrence of peak runoff. We attribute these timing changes to the shifts in snowmelt and springtime soil moisture processes. Specifically, areas in which snowmelt drives watershed hydrology exhibit earlier dates of maximum snow accumulation and peak runoff. In regions where peak runoff is projected to occur later, we find a tendency for later occurrence of full saturation conditions. Earlier and later peak runoff occurrence can potentially lead to competing water use interests and aggravating concerns for aquatic environments and their ecosystem services.

Keywords: Climate change (1807); peak runoff (1817); surface hydrology; climate model projections (1847); uncertainty (1873); human activities

1.0 Introduction

Surface water is an essential source of freshwater, whose variability has profound impacts on the life of humanity (*Hall et al.*, 2014). Surface water peak flows can result in flooding – the most impactful natural hazard of all weather-related events in terms of fatalities and material costs (*Doocy et al.*, 2013). But high streamflow also replenishes reservoirs, carries and deposit nutrients in floodplains, can be the source of tremendous useable energy, and is an important source of irrigation for agriculture in arid areas. Additionally, the diversity of fish communities is closely related to the streamflow seasonality (*Knight et al.*, 2014). Understanding patterns of surface flows in space and time is therefore crucial for flood control, water supply, crop yield, ecosystem services, water quality control, and hydropower generation (*Kemter et al.*, 2020). Streamflow characteristics, such as the magnitude, frequency, and seasonality, can be affected by human-induced land use and climate change that both intensify the global hydrologic cycle (*Bosmans et al.*, 2017; *Winsemius et al.*, 2016). Stemming from observation-based studies and climate model projections, analyses of the sign and magnitude of peak annual streamflow changes in the historical period and the future remain controversial (*Greve et al.*, 2018; *Gudmundsson et al.*, 2019; *Hirsch and Ryberg*, 2012; *Lins and Slack*, 2005; *Mallakpour and Villarini*, 2015; *Milly et al.*, 2005; *Yang et al.*, 2017; *Zhai et al.*, 2020). Nonetheless, there is high confidence that the frequency of extreme floods associated with annual streamflow maxima has increased over most regions, and this trend is likely to continue in the future (*Arnell and Gosling*, 2016; *Hirabayashi et al.*, 2013; *Hirsch and Archfield*, 2015; *Milly et al.*, 2002; *Slater and Villarini*, 2016; *Swain et al.*, 2020). A number of studies have also addressed the question of streamflow seasonality shifts due to impact of non-stationary climate on maximum annual streamflow occurrence (*Bloschl et al.*, 2017; *Clow*, 2010; *Cunderlik and Ouarda*, 2009; *Dudley et al.*, 2017; *Villarini*, 2016). Focusing on historical trends using gagelevel data, their principal conclusions are that many watersheds have already experienced a significant shift in annual maximum streamflow timing. However, an open question is whether streamflow seasonality will change in the coming decades, and if so, which factors would be the main drivers.

It is vital to understand the key governing processes that determine the major phase of watershed streamflow in order to understand its future shifts. Several studies have reported substantial variability in the seasonality of maximum annual flows over the continental U.S. and attributed it to distinct differences in flood-generating mechanisms (*Berghuijs et al.*, 2016; *Villarini*, 2016). Specifically, precipitation and antecedent soil water conditions were identified as key factors explaining the occurrence of highest flows over the central U.S. (*Slater and Villarini*, 2017) and western coastal areas (*Berghuijs et al.*, 2016; *Ye et al.*, 2017). In the western mountainous areas (*Li et al.*, 2017; *Yan et al.*, 2019) and the northeastern U.S. (*Hodgkins et al.*, 2003), snowmelt was determined to be the dominant driver of runoff. Climate change can directly or indirectly affect precipitation, soil moisture, and snowmelt processes, with consequences to flood seasonality across regions with distinct dominant runoff generating mechanisms, triggering implications for hydropower, agriculture, and aquatic ecosystem services. For example, numerous studies reported that trends of increasing temperature in regions with snowmelt-driven hydrology have already resulted in earlier annual peak streamflow (*Barnett et al.*, 2005; *Clow*, 2010; *Hodgkins et al.*, 2003; *Kam et al.*, 2018; *Regonda et al.*, 2005; *Stewart et al.*, 2005). Trends and interpretations in regions with other processes of dominant hydrological influence are cumbersome to disentangle and projections into the future are also subject to this large attribution uncertainty. In this study, we address knowledge gaps related to the understanding of future changes in peak runoff seasonality at the U.S. continental (CONUS) scale. Specifically, we assess the likelihood of changes in peak runoff timing during the 21st century based on daily runoff projections that are outputs of ten General Circulation Models (GCMs) from the fifth phase of the Coupled Model Intercomparison Project (CMIP5). The sensitivity of GCM-modeled runoff to temperature is not well constrained, which can result in significant uncertainty for future projections (*Lehner et al.*, 2019). To enhance confidence of the projection and in order to reduce GCM biases, we apply the Bayesian weighting averaging (BWA) method of *Smith et al.* (2009) to produce multi-model ensemble estimates that rely on model performance over the control period and model projection convergence in the future to assign model weights. The product of *Livneh et al.* (2013) is used in this Bayesian framework to reduce biases of GCM runoff estimates. Using the downscaled estimates of future runoff, we aim to identify patterns of peak runoff timing change under the different CO_2 emission scenarios and carry out analysis that identifies main drivers of the projected changes.

2.0 Methods

2.1 Runoff historical data and projections

Long-term estimates of daily runoff (surface water yield per unit area) provided by *Livneh et al.* (2013) are used in this study as true "observations" within the Bayesian framework of multi-model downscaling to reduce projection biases. Daily runoff is obtained as output of the Variable Infiltration Capacity (VIC) model (*Liang et al.*, 1994) forced with precipitation and temperature, at the spatial resolution of $1/16^{\circ} \times 1/16^{\circ}$.

Realizations from ten General Circulation Models developed in different institutions were downloaded from the CMIP5 database (http://pcmdi9.llnl.gov/). Only one GCM version is chosen for each institution (see Table 1) to reduce the dependence within the multi-model ensemble. GCMs selected in this study satisfy the criteria of availability of daily runoff outputs and completeness of spatial coverage over the contiguous U.S. Emission scenarios corresponding to the Representative Concentration Pathway (RCP; *van Vuuren et al.*, 2011) 4.5 and 8.5 are used to represent medium and most pessimistic predictions of greenhouse gas concentration in the future.

Because GCM outputs and the runoff dataset of *Livneh et al.* (2013) have different meshes, they were converted to the same $1^{\circ} \times 1^{\circ}$ resolution for analysis convenience. We first re-mapped all GCM outputs to $1/16^{\circ} \times 1/16^{\circ}$ resolution with the nearest neighbor method. Then, both GCM and the runoff data layers were aggregated by averaging over grid cells falling inside each $1^{\circ} \times 1^{\circ}$ cell of the analyzed product set.

Table 1. List of CMIP5 models used in this study

No.	Institution	Model Name	Resolution (lon x lat)
1	Beijing Climate Center	bcc-csm1-1	128×64
2	Euro-Mediterranean Centre on Climate Change	CMCC-CM	480×240
3	National Center for Meteorological Research, Météo-Franch and CNRS laboratory	CNRM-CM5	256×128
4	Commonwealth Scientific and Industrial Research Organization – Queensland Climate Change Centre of Excellence	CSIRO-Mk3-6-0	192×96

5	Institute of Numerical Mathematics of the Russian Academy of Sciences	Inmcm4	180×120
6	Institute of Atmospheric Physical and Centre for Earth System Science	FGOALS-g2	128×60
7	Model for Interdisciplinary Research on Climate	MIROC5	256×128
8	Max Planck Institute for Meteorology	MPI-ESM-MR	192×96
9	Meteorological Research Institute	MRI-CGCM3	320×160
10	Norwegian Climate Center	NorESM1-M	144×96

2.2 Multi-variate Bayesian Weighting Averaging (BWA)

It has been established in the literature that making future projections based on a multi-model ensemble is preferred over inferences based on single-model outputs (*Knutti et al.*, 2010; *Tebaldi and Knutti*, 2007) due to potentially high biases of any given model. Biases of GCM projections in climate variables (e.g., temperature and precipitation) can be significant (*Knutti et al.*, 2010; *Xu et al.*, 2018), and therefore they must be addressed before any robust conclusion on climate change can be drawn. The Bayesian weighted averaging (BWA) approach of *Smith et al.* (2009); *Tebaldi et al.* (2004); *Tebaldi et al.* (2005) has grown in popularity as a sufficiently general tool to assess climate change uncertainties from multiple model projections with minimum subjective assumptions. This approach is derived from the Reliability Ensemble Average method introduced by *Giorgi and Mearns* (2002) to integrate model outputs, such that the model weights are based on model performance in the past period with historical observations and model output convergence in the future period. The first version of BWA was univariate such that each location was considered separately, creating solutions informed by the local model performance (*Tebaldi et al.*, 2005). In cases of large model-observation differences, this version could produce problematic posterior distributions (*Smith et al.*, 2009; *Xu et al.*, 2018). To extend the approach utility, *Smith et al. (2009)* proposed a multivariate version of BWA that simultaneously considers a set of model outputs in multiple regions. Model weights therefore rely on its performance in all regions and locations considered, which ensures a more robust model skill evaluation given site-to-site variation of uncertainties. Additionally, this method requires fewer parameters in calculating the posterior distributions than the univariate version and is thus more computationally efficient. Readers are referred to *Smith et al.* (2009) for a detailed derivation, and only a brief description of the formulation is introduced here.

Smith et al. (2009) postulated that the *j*th climate model projections in the past and future in the *i*th region are denoted as X_{ij} and Y_{ij}, with $i = 1,...,R$, $j = 1,\ldots,M$, where R is the total number of regions considered and M is the total number of models in an ensemble. X_{i0} is the associated historical observation for the same past period. It is assumed that observations and projections are random Gaussian variables that are distributed as:

$$X_{i0} \sim N[\mu_0 + \zeta_i, \lambda_{0i}{}^{-1}], \tag{1}$$

$$X_{ij} \sim N[\mu_0 + \zeta_i + \alpha_j, (\eta_{ij}\phi_i\lambda_j)^{-1}], \tag{2}$$

$$Y_{ij} \mid X_{ij} \sim N[\nu_0 + \zeta_i' + \alpha_j' + \beta_i(X_{ij} - \mu_0 - \zeta_i - \alpha_j), (\eta_{ij}\theta_i\lambda_j)^{-1}], \tag{3}$$

where λ_{0i} is the inverse of variance of X_{i0} based on observational data. The other parameters are assumed to have the following prior distributions, all are mutually independent:

$$\mu_0, \nu_0, \zeta_i, \zeta_i', \beta_0, \beta_i \sim U(-\infty, \infty), \tag{4}$$

$$\theta_i, \phi_i, \psi_0, \theta_0, c, a_\lambda, b_\lambda \sim G[a,b], \tag{5}$$

$$\lambda_j \mid a_\lambda, b_\lambda \sim G[a_\lambda, b_\lambda], \tag{6}$$

$$\eta_{ij} \mid c \sim G[c,c], \tag{7}$$

$$\alpha_j \mid \psi_0 \sim N[0, \psi_0^{-1}], \tag{8}$$

$$\alpha_j^{'} \mid \alpha_j, \beta_0, \theta_0, \psi_0 \sim N[\beta_0 \alpha_j, (\theta_0 \psi_0)^{-1}]. \tag{9}$$

Conventionally, $G[a,b]$ denotes the gamma distribution with the shape parameter a and the rate parameter b. The parameters μ_0 and ν_0 are interpreted as the global means, ζ_i and $\zeta_i^{'}$ are the differences from the global mean defined for a specific region '*i*', and α_j and $\alpha_j^{'}$ represent the global biases for a specific model '*j*' for the past and future periods, respectively. In terms of the variance assumption in the above equations, λ_j represents the inverse of the variance of the *j*th model, ϕ_i represents the inverse of the variance for the *i*th region in the past, and θ_i represents the inverse of the variance at *i*th region in the future. The introduction of η_{ij} here is to guarantee that climate models have different patterns of output variance in different regions. The uniform distribution is selected over $(-\infty, \infty)$, and a, b and c are set to 0.01 to ensure that all of the priors are uninformative. The other three hyperparameters β_0, θ_0, ψ_0 are used to define the common distribution of climate models. The analytical forms of the joint posterior distributions are unknown, but closed-forms of each marginal posterior distribution are derived in the appendix of *Smith et al.* (2009). In practice, the Markov Chain Monte Carlo (MCMC) process is used to estimate the posterior distributions (*Smith et al.*, 2009). Note that the parameter λ_{0i} capturing historical variability of peak runoff timing is accounted for in this methodology to represent 'noise' in the peak time occurrence: larger 'noise' implies less confidence in the distributions of model-observation biases and thus this will cause the posterior distribution of peak runoff timing change to have larger variance. We further note that the random variables α_j and $\alpha_j^{'}$ representing model biases additionally account for the uncertainty of biases in GCM model outputs and their larger variances (assessed via the MCMC process) will yield higher 'noise' in the projections of timing of peak runoff (see SM. 2 robustness metric).

2.3 Adaption of BWA to peak runoff timing

GCMs estimate runoff (i.e., water excess in a model grid cell), not streamflow (i.e., the flow rate at a given point in a channel network). Consequently, in this study we use annual peak runoff as an indicator of the occurrence of major hydrological phase, rather than annual peak streamflow used in previous observation-driven studies. Runoff routing to channel network and in-channel wave transformation can introduce additional uncertainty since the coarse spatial resolution of GCM computational mesh cannot represent these processes and the resultant runoff-streamflow basin lag. However, a comparison between the high-resolution *Livneh et al.* (2013) runoff dataset and streamflow measured at USGS gauges across CONUS illustrates that the correlation between the average annual runoff and streamflow is high both in terms of magnitude and timing (Figure. S1). This suggests that shifts in the timing of both variables in the future period should be also correlated (although this is apparently impossible to verify). We further emphasize that peak runoff timing is not equivalent to peak streamflow timing. An apparent advantage is that runoff projections from grid-based model outputs allow us to study runoff spatial variability over the entire U.S. continent, without the need to explicitly include the effects of water management, which is necessary for point-scale streamflow analysis. The quality of daily runoff product of *Livneh et al.* (2013) used in the Bayesian framework to reduce biases has been verified (SM. 1).

The occurrence date of annual peak runoff is the variable of interest inferred from GCM outputs. Daily GCM runoff outputs are used to derive the annual peak runoff timing, and Day of Year (DOY) is used to represent its occurrence date, where January 1st corresponds to 1 and December 31st to 365 (or 366 during a leap year). The original BWA cannot be applied directly to DOY due to its circular nature. To resolve this issue, we use the differences between the modeled and observed dates as the variable of interest in BWA to convert the circular variable to a linear variable:

$$\tilde{X}_{ij} = X_{ij} - X_{i0}, \tag{10}$$

$$\tilde{Y}_{ij} = Y_{ij} - X_{i0}. \tag{11}$$

where $\tilde{X}_{ij}$ and $\tilde{Y}_{ij}$ respresent the deviations from the observed peak runoff timing (X_{i0}) for the jth model at ith location for the control period and future period, respectively. An example of the conversion is given in SM. 3.

The Bayesian posteriors of multi-model ensemble mean of runoff peak timing are constructed using outputs of selected GCMs (see Table 1) for the control and future periods. The control period is defined as 1961-1990, and two future periods selected in this study are 2041-2070 (mid-century) and 2071-2100 (end-century). Two CO_2 emission scenarios, RCP 4.5 and RCP 8.5 (*Rogelj et al.*, 2012), are used here to represent the different possible trajectories of the global climate evolution. We use the differences of the mean peak annual runoff timing estimated from the Bayesian posteriors for future and control periods to make inferences on the change of runoff seasonality timing caused by the global change. The robustness metric of *Knutti and Sedláček (2012)* accounting for the uncertainty of GCM projections is used to calculate the strength of the change signal of the multi-model mean (see SM. 2). Higher robustness of the inferred change will depend on the peak timing variability over historical period, both in terms of observations and model simulations, model vs. observation differences (i.e., model biases), and the degree of convergence of modeled outputs for both historical and future periods.

2.4 Dates of maximum precipitation and snowpack, and the distribution of soil moisture saturation

Precipitation is an obvious driver of many hydrologic dynamics. We compute shifts in the dates of maximum 1, 3, 5, and 7-day accumulated precipitation by taking the difference of the multi-model date averages (equal GCM weights) for the future and control periods.

We use the occurrence time of maximum annual snow water equivalent (maxSWE) from the selected set of GCMs to identify the onset of snow melting phase. Only cells with maxSWE higher than 15 $[kg/m^2]$ (i.e.,15 $[mm]$ liquid water depth) are analyzed to ensure sufficient snow accumulation prior to snowmelt. We compute the change of the maxSWE mean date by taking the difference of the multi-model date averages (equal GCM weights) for the future and control periods.

Daily soil moisture over the top 10 cm depth from the selected set of GCMs is used to develop a distribution of springtime dates of extreme wetness. We first use the maximum soil moisture over the selected 30-year periods (control or future) to identify the soil saturation limit θ_{sat}. We then construct empirical cumulative density function (CDF) of the dates between February 1st and May 31st when soil moisture is higher than $0.95 * \theta_{sat}$, using both control and future periods based on the outputs of all GCMs (see Table 1). Only late winter - spring period is considered since the robustness metric of *Knutti and Sedláček (2012)* for changes in peak runoff timing exhibits high values (> 0.6) during this interval only. The difference of days between the two CDFs corresponding to the median values (i.e., CDF at 0.5) is used to represent the shift of the distribution centroid of extreme springtime soil wetness in the future.

3.0 Results

3.1 Changes of peak annual runoff timing

The peak annual runoff over the continental U.S. exhibits clear regional patterns (Figure S2). Figure 1 illustrates the change of the mean timing of annual peak runoff between the future and the control periods inferred from the multi-model BWA posterior distributions. We present four cases corresponding to two future periods and two emission scenarios. The grid cells with high confidence of the change inference based on the robustness metric of *Knutti and Sedláček (2012)* are highlighted. Higher robustness means that the project runoff changes are more significant than the model noise and historical variability (Figure S3), i.e., the associated projection uncertainty is smaller. The fractions of the CONUS area in Figure 1 showing grid cells with high robustness changes for these four time periods are (a). 9.3%, (b). 10.2%, (c). 10.7%, and (d). 17.2%, implying that the higher the greenhouse gas concentrations changes (and, correspondingly, the higher the projected temperature increases), the more consistent and significant runoff peak timing changes projected by GCMs. The spatial patterns of robust changes are similar across all four scenarios. Specifically, the regions with winter snowpack, such as the Rocky Mountains and New England, are projected to have annual peak runoff shift to earlier dates, by up to 3-5 weeks. Peak runoff is likely to be delayed by up to 2-4 weeks in the Midwest region, southern Florida, and parts of the west coast, where soil moisture has been argued to be the key factor in peak runoff formation (*Ivancic and Shaw*, 2015). The change in the west of Gulf Coast region has a high uncertainty due to the poorly pronounced period of peak runoff, since highest runoff can occur at any time of a year. The changes have different signs for the upper Missouri basin region, when comparing the results for the end of century RCP 8.5 scenario with the other three cases, but the spread of model projections likely cause this since the inference robustness is not high.

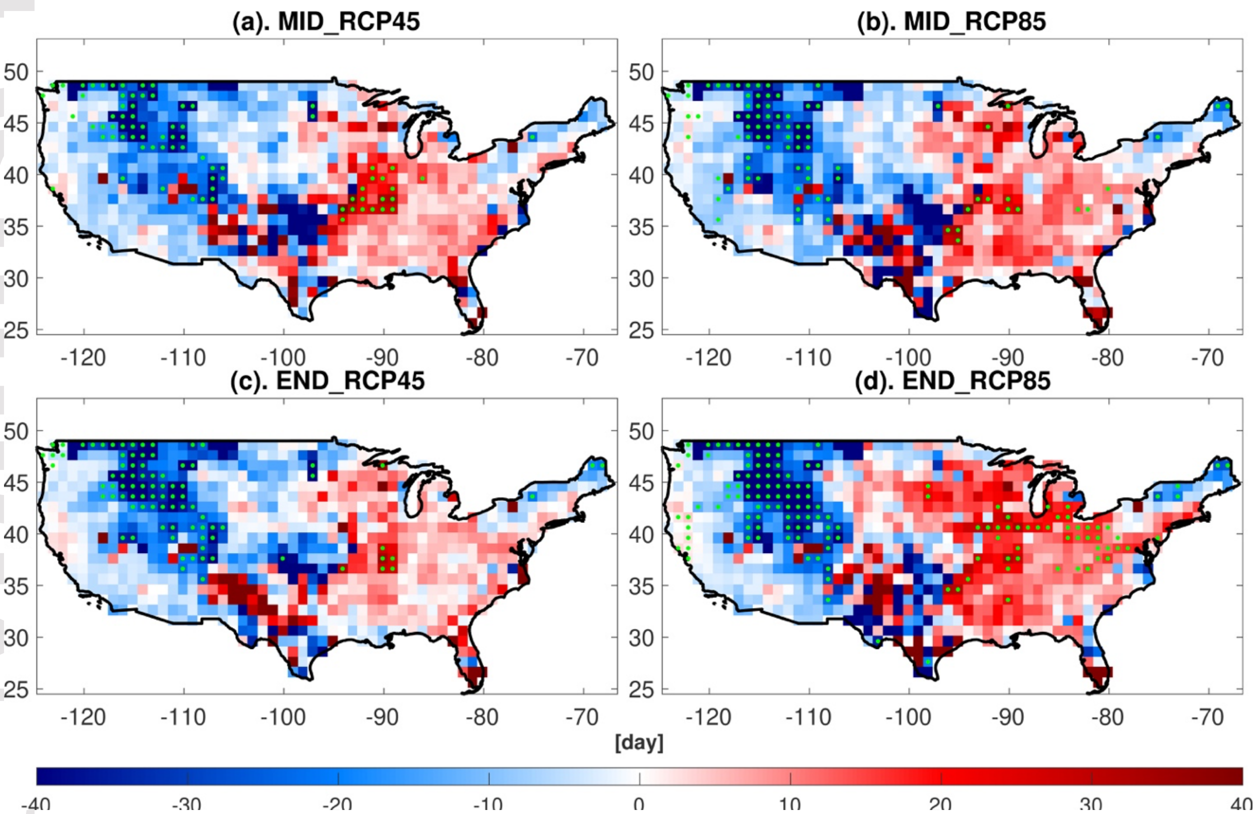

Figure 1. Change of the mean date of annual peak runoff occurrence between the control (CTL) and future periods (FUT). The difference (FUT – CTL) is estimated using the dates of maximum likelihood from BWA posterior distributions for the two periods. The grid cells with inference of high robustness (SM.2, metric of *Knutti and Sedláček* (2012) higher than 0.6) are stippled with green points. "MID" (subplots (a) and (b)) represents the date difference with respect to 2041-2070 and 1961-1990 periods, and "END" (subplots (c) and (d)) represents the difference with respect to 2071-2100 and 1961-1990 periods. Daily runoff product (SM.1) of *Livneh et al.* (2013) and outputs from ten GCMs are used to construct the BWA posterior. All of the results are shown at $1^{\circ}\times 1^{\circ}$ resolution.

3.2 Attribution of the change in peak annual runoff timing

To develop an attribution of the patterns of peak runoff timing change in Figure 1, we investigate outputs of daily precipitation, surface snow accumulation, and top layer (0-10 cm) soil moisture from the same CMIP5 multi-model ensemble. Figure 2a shows the changes of annual peak daily precipitation timing for the end of the century RCP 8.5 scenario (the other cases can be found in Figure S4). While extreme heavy precipitation (e.g., corresponding to return periods larger than 100 years) is generally associated with long-term maximum annual runoff (*Smith et al.*, 2013), changes of the mean timing of peak daily annual precipitation cannot explain the change in the mean timing of peak annual runoff (Figure S5a). Likewise, shifts in maximum 3-day, 5-day, and 7-day accumulated precipitation also were not found to be related to the inferred changes in the peak runoff seasonality (not shown). This is consistent with previous studies that relied on stream gauge data to demonstrate that snowpack dynamics and antecedent soil wetness can play more critical roles in generating peak annual streamflow (*Ivancic and Shaw*, 2015), with the exception for urban areas where heavy rainfall was identified to be the primary factor (*Sharma et al.*, 2018).

The change of maxSWE mean date illustrates the predominantly earlier dates of maximum snow accumulation in the future (Figure 2b for RCP8.5 end-of-century; Figure S6 for all of the future cases). As the *delayed* peak runoff cannot be attributed to the changes of maxSWE timing (Figure S5b), we explore the possibility of impact of maxSWE date change on *earlier* timing of peak runoff only (i.e., blue cells with green circles in Figure 1d). For all the four future scenarios, a positive relationship between the peak runoff and the peak maxSWE timing change indicates a coherent shift of both to earlier dates (Figure 3, blue squares). The high correlation also implies causation as the shifts are projected to occur in regions dominated by snowpack (Figure 2b) and snowmelt process is the dominant runoff generation mechanism, i.e., the earlier start of snowmelt is related to the earlier phase of runoff production via well-understood, physically meaningful processes.

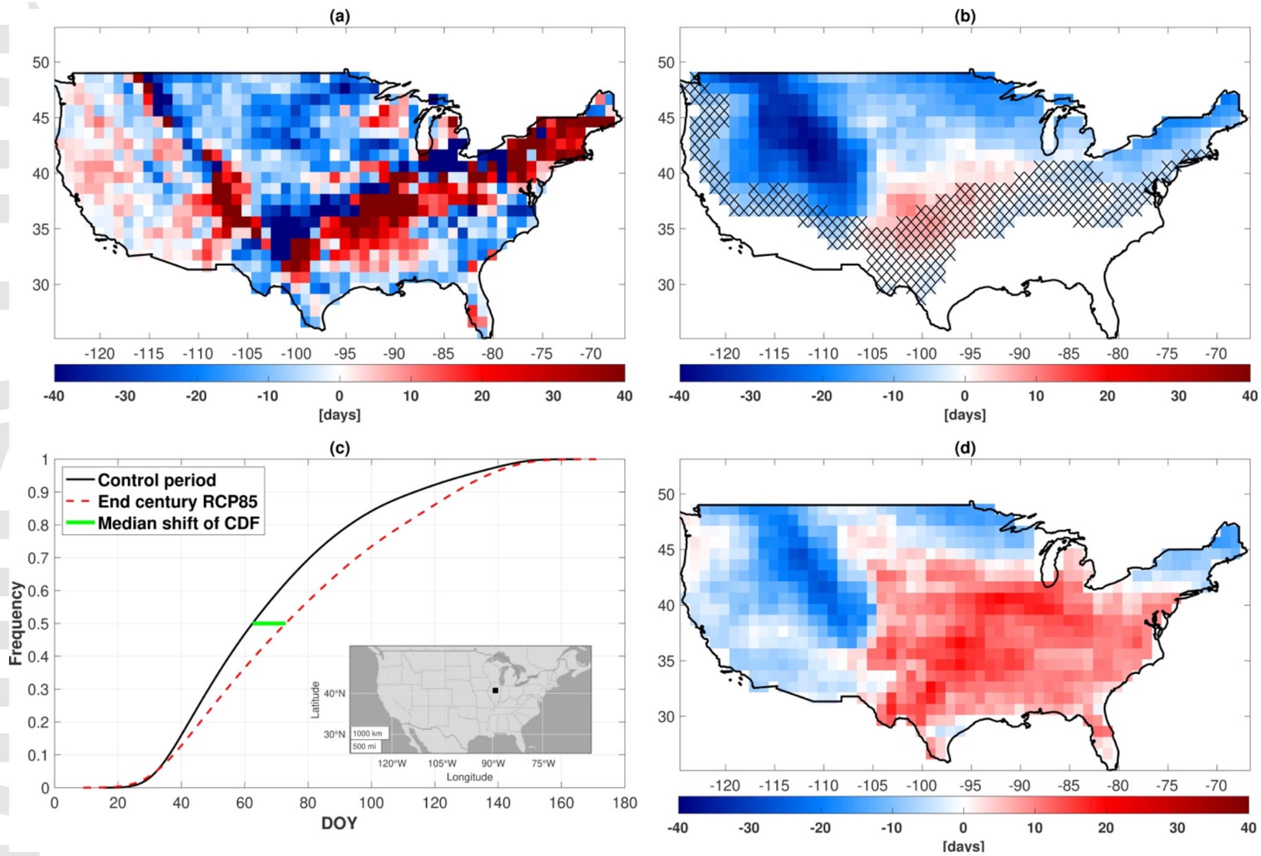

Figure 2. Change of precipitation, snowpack, and spring soil moisture seasonality (RCP8.5 scenario). (a) The difference of annual timing of peak precipitation between the end-of-century and the control period. (b) The difference of annual timing of maxSWE between the end-of-century and the control period. The white areas along the southern and western coasts represent negligible snow accumulation in the control period (i.e., maxSWE < 15 mm). Hatching marks areas in which snow accumulation becomes negligible in the future. (c) Empirical cumulative density functions (CDFs) of the dates between February 1 (DOY=32, 'DOY' – day of year) and May 31 (DOY=151) on which soil moisture is 95%-100% of its saturation limit. GCM outputs during the control period (solid black line) and the end of century period (red dashed line) are used. The CDFs are constructed for an exemplary grid cell (with the robustness metric of the peak timing change > 0.6) indicated with the black square in the inset. The solid green line represents the shift between the two CDFs at their median values, i.e., the difference represents the date change of the distribution centroid of springtime extreme soil wetness. Subplot (d) illustrates the shift of the centroid of springtime wetness illustrated in (c) between the end-of-century and the control period over the CONUS area.

The projections of daily water content in the top 10 cm of soil are used to investigate the impact of soil wetness on later peak runoff occurrence (i.e., red cells with green circles in Figure 1d). Unlike precipitation and snow, soil moisture is bounded by the saturation limit θ_{sat}, reaching this limit many times in a given year. Consequently, we identified all dates when soil moisture exceeded 95% of θ_{sat} in GCM outputs for both the control and future periods to construct their empirical cumulative density function (Sec. 2.4). As an example, Figure 2c illustrates CDFs inferred from multi-model projections for control and future periods for a grid cell with the delayed peak runoff in the end-century RCP 8.5 scenario. What is apparent in this illustration is that nearly the entire CDF of the days of extreme spring wetness in future shifts to a later time of the year, as compared to the control period. This delay reflects the combined control of precipitation, evapotranspiration, and snowmelt on soil wetness due to the persistence property of soil moisture (*Ghannam et al.*, 2016).

While peak annual runoff may correspond to any day on the CDF of dates of extreme springtime wetness, we calculate the difference between the median CDF values to assess the interval between the two distribution centroids. Figure 2d illustrates these differences over the CONUS area for the end-of-century RCP8.5, which yields a positive relationship with the shift of annual peak runoff timing to *later* dates only (Figure 3d, red squares). The relationship is relatively insensitive to the choice of the CDF quantile (e.g., using 25% and 75% in Figure S.7 leads to similar inferences). By taking the difference of the dates at 50% of CDF, we infer the shift of springtime soil wetness centroid. However, the occurrence of peak runoff cannot be related to the occurrence of extreme wetness dates in any straightforward fashion, i.e., peak runoff can theoretically occur on any date of springtime soil saturation conditions. Specifically, Figure 4 shows that during the control period peak annual runoff occurred on average around DOY 50 (i.e., 32% of the CDF); it shifts to DOY 65 (43% of the CDF) for future conditions. Furthermore, the results for the other future projection scenario (i.e., RCP4.5) and period (i.e., mid-century) show similar patterns of the change (Figure 3a-c and Figure S.8).

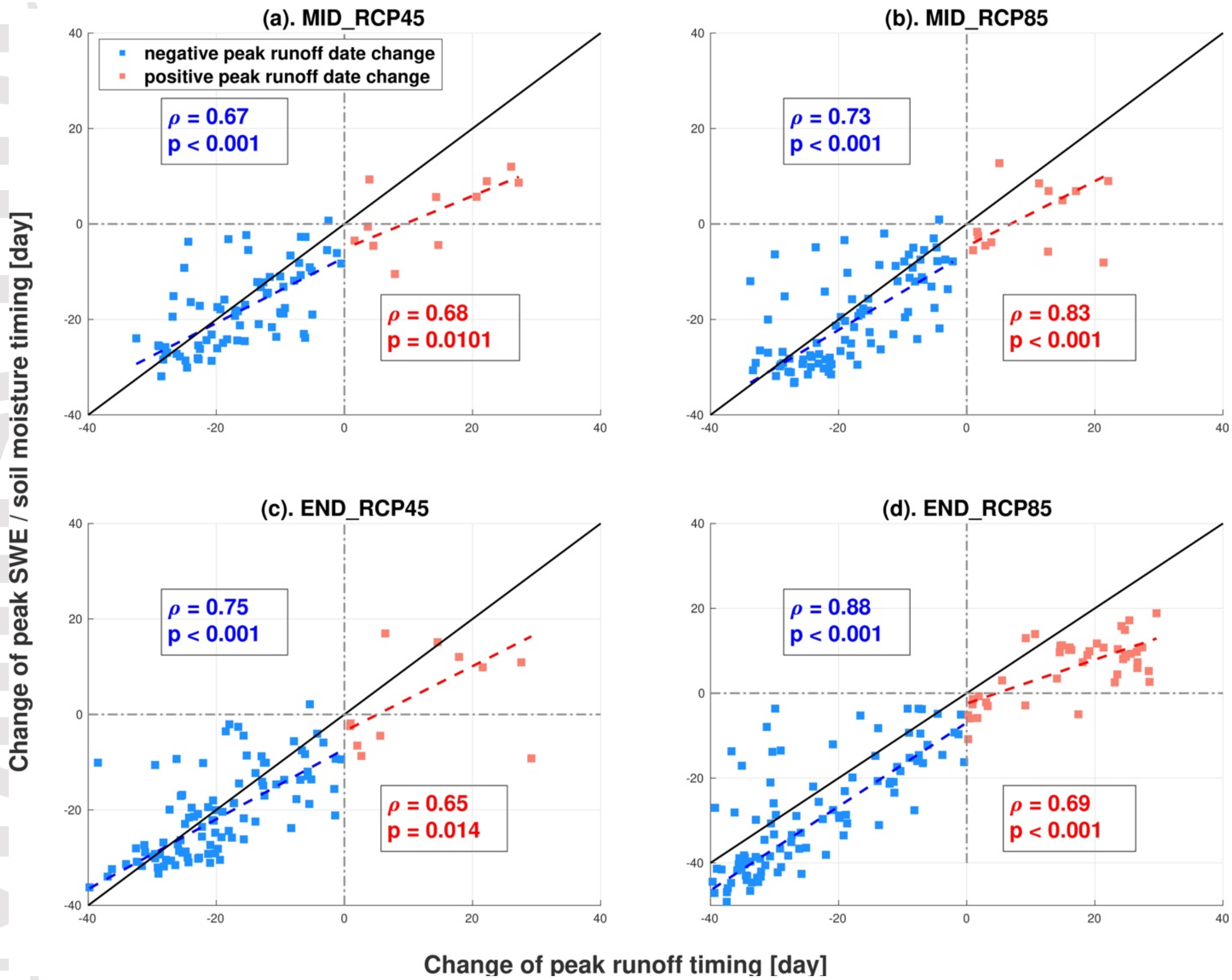

Figure 3. Attribution of the change in mean timing of annual peak runoff. Regressions between the peak annual runoff timing change and the change of the date of maximum snow water equivalent (blue squares), and the shift of centroid date of extreme spring soil wetness (red squares) for (a). the mid-of-century, RCP4.5 scenario, (b) the mid-of-century, RCP8.5 scenario, (c) the end-of-century, RCP4.5 scenario, and (d) the end-of-century, RCP8.5 scenario. Only the results for locations with the change robustness metric larger than 0.6 for peak annual runoff timing are presented. The peak runoff timing changes are calculated using the multi-model ensemble mean with equal weights assigned to each GCM to ensure a consistent comparison with the changes in the peak SWE and soil moisture timing. The grey line represents the 1:1 reference line, and the blue and red dashed lines are the linear least-squares regression lines. ρ is the correlation coefficient and p is the corresponding p-value.

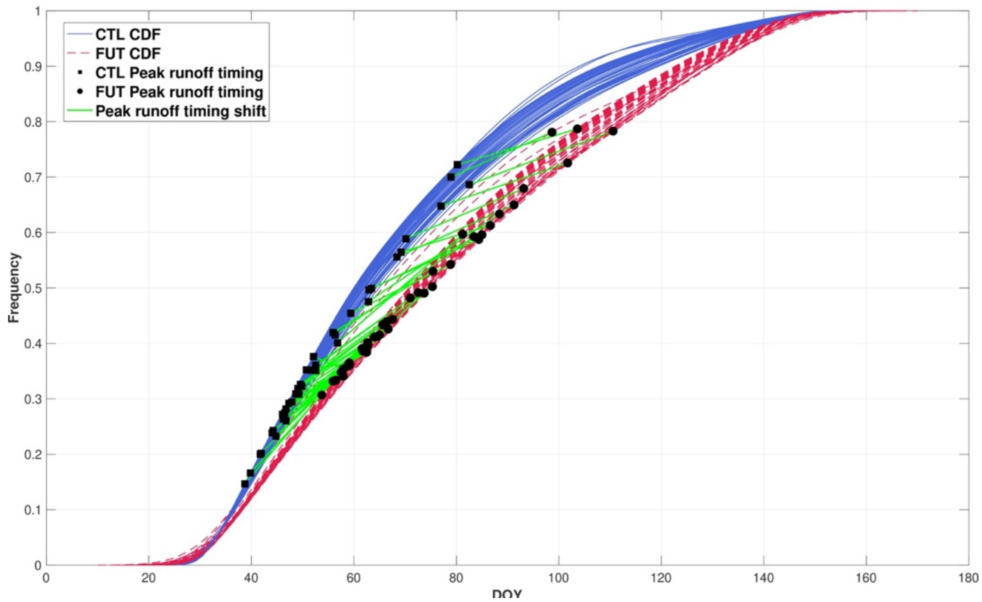

Figure 4. CDFs of dates of extreme springtime soil wetness and the shifts of peak annual runoff timing (RCP8.5, end-of-century scenario). The blue solid lines (red dashed lines) represent CDFs of dates of soil saturation for each cell in Figure 2d with delayed peak annual runoff (red cells with green circles - high robustness) for control period (future period). Black squares (black) circles are the corresponding peak annual runoff occurrence dates from multi-model mean for the control period (future period). The green solid lines illustrate the shift of peak annual runoff timing for all examined cells.

We additionally note that the *negative* changes of soil wetness timing also exhibit (weaker) correlation with the *negative* changes of peak runoff timing (Figure S5c). However, these projected shifts of runoff timing to earlier dates are located in regions dominated by snowpack runoff generation (Figure 1). Therefore, changes in snowmelt timing are expected to contribute to changes in spring soil moisture dynamics, triggering collinear effects between the two predictors: the timing of maxSWE and the centroid date of soil saturation during spring period.

4.0 Discussions and conclusions

In this study, we focus on linking peak runoff seasonality with changes in the climate system. In summary, our results show clear spatial patterns of peak annual runoff timing change over the continental U.S. caused by the projected global climate change that drives changes in the physical processes of land-surface hydrology. We find that snowmelt will occur earlier in the future and this will cause a shift of peak annual runoff to earlier dates, with the median of 2.7 (RCP4.5) to 3.9 (RCP8.5) weeks by the end of the century in regions where snowmelt is the dominant runoff generating mechanism. In other regions, where climate projections yield a robust signal of delay in peak annual runoff timing with the median of 1.6 (RCP4.5) to 2.6 (RCP8.5) weeks by the end of the century, we uncover the importance of soil wetness during spring period; we find that there is an overall shift of extreme soil wetness conditions to later dates. Such shifts in the *timing* of extreme soil moisture conditions may correspond to various expressions of the soil moisture process (e.g., conceptual illustrations in Figure S9), e.g., they *may* correspond to specific changes in its first and higher-order moments. However, while we note that springtime moisture conditions are projected to be drier (e.g., by ~3%, end-of-century, RCP8.5) and exhibit higher variance (~7%), we do not find a strong relationship between changes in these two moments and changes in peak runoff timing. Since the distribution of soil moisture is always positively skewed, the change in these moments may be insufficient to represent the change in peak runoff timing, which is likely to be affected by extremes of soil moisture process. Further attribution analysis is warranted.

We find that all the changes are projected to be more pronounced and more robust by the end of the 21st century if the current greenhouse gas emission levels are maintained, since RCP8.5 represents the "business as usual" scenario (*van Vuuren et al.*, 2011). Such changes can pose serious challenges to the human activities and natural environment, since they are adapted to the historical runoff seasonality (*Bloschl et al.*, 2017). For example, nearly three quarters of water supply in the western United States are driven by snowmelt (*Dettinger*, 2005) and the 3-5 week earlier peak runoff can result in competing water use interests: prioritizing reservoir storage can conflict with ensuring sufficient flows for salmon migration (*Dudley et al.*, 2017). Likewise, a 2-

4 week delay in springtime extreme wetness conditions in the U.S. Midwest may imply late crop planting and a delay in springtime fertilizer applications; when combined with high flows and warmer summer conditions, this can pose threat to aquatic environments and their ecosystem services (*Michalak et al.*, 2013).

This study analyses runoff rather than streamflow because streamflow is not available in GCMs' outputs. Despite the correlation of the two for the historical period (Figure S1b), caution must be exercised in interpretation of the study results. Specifically, while robust changes of the former in the future are detected, this study does not present objective evidence that the timing of peak streamflow will be impacted in the same fashion. To investigate the change of peak streamflow timing, a hydrodynamic model is needed to route runoff. However, modeling this process will introduce additional uncertainties from unavoidable errors in representation of drainage network and channel geometry, and specification of "effective" friction properties of the land-surface at the scale of GCM grid cell of several hundred square kilometers, etc. There is currently no objective way of accounting for these additional uncertainties and thus projections of streamflow metrics into future will likely remain elusive.

Furthermore, the timing of peak *daily average* runoff can be different from the timing of peak *instantaneous* runoff. Conceptually, the difference between the two would be characteristic of systems in which peak runoff is controlled by extreme rainfall. The latter is not well captured by GCMs (*Dai*, 2006; *Stephens et al.*, 2010) and thus in these systems one expects low convergence of GCM outputs. We do however identify regions with high robustness of the change (Figure 1), implying that runoff dynamics bear a signature of day-to-day persistence reflecting their driving processes (*Berghuijs et al.*, 2016; *Ye et al.*, 2017). Arguably, this suggests that in these regions the timing of annual peak of *instantaneous* runoff coincides with that of annual peak of *daily average* runoff.

While the Bayesian method was applied to reduce the multi-model ensemble uncertainty, the approach does not automatically guarantee the uncertainty of future runoff projection to be well constrained, which represents a limitation of this study. In fact, the uncertainty of BWA projections can be large for many grid cells, as the Bayesian posterior of the peak runoff timing change can span a wide range: from negative to positive values (Figure 5). We acknowledge that it is not reliable to draw any conclusion for locations with such a high uncertainty. However, grid cells with high values of the robustness metric exhibit consistent bounds (i.e., either positive or negative) informed by a narrower model spread, indicating superior agreement among the models. This supports the high confidence placed by the analysis on cells with high robustness in the multi-model ensemble. While not entirely impossible, the signal of high robustness is unlikely to be merely fortuitous as the presented results on change of peak runoff timing make perfect physical sense.

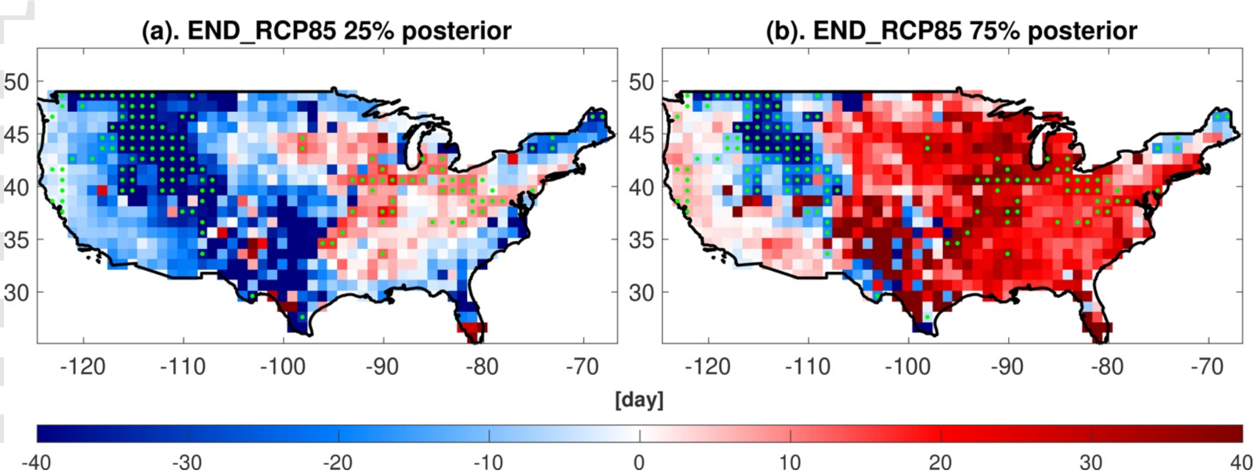

Figure 5. Uncertainty bounds for the change of the mean date of annual peak runoff occurrence. (a). 25%, and (b) 75% of the BWA posterior distribution of the change of mean date

of annual peak runoff occurrence between the end-of-century and the control period, the RCP 8.5 projection. The green dots denote cells with high robustness metric (as identified in Figure 1d).

We acknowledge that the real-world impacts of climate change on runoff generation are complicated and controlled by many factors at the scales of their governing physical processes. Specifically, with their simplified runoff generation mechanisms, current GCM versions can realistically mimic only major phases of runoff due to the input of rain or meltwater in excess of soil saturation. GCM land-surface modules are one-dimensional representations of hydrology over large areas of a grid cell that grossly simplifies spatial variations of land-surface conditions. They cannot capture vital details of the other types of runoff generation such as those controlled by hillslope hydrology and surface-groundwater interactions (*Beven*, 2012; *Bisht et al.*, 2018), soil structure (*Or*, 2020), snow redistribution across landscape in areas of complex topography (*Chegwidden et al.*, 2020), or mosaic of landuse variations such as those due to urbanization (*McGrane*, 2016). While relevant processes and their controlling factors can be captured by detailed models of watershed hydrology stemming from the first principles (*Fatichi et al.*, 2012; *Ivanov et al.*, 2008; *Kim et al.*, 2012; *Maxwell et al.*, 2014), these models cannot be operated at global scales. This is because of the infeasibly enormous computational demand implied by the high spatial resolution and time stepping required for appropriate solution of the governing partial differential equations (*Fatichi et al.*, 2016). Therefore, suitable simplifications (known as "parameterizations") of processes (e.g., surface and groundwater flow, snow) and/or controlling factors (e.g., topography, soil structure, landuse) continue to be necessary for GCMs. Correspondingly, recent developments targeting to improve the representation and realism of hydrological physical processes in land-surface models have included surface water dynamics (*Ekici et al.*, 2019), land-river interactions (*Chaney et al.*, 2020; *Decharme et al.*, 2019), parameterizations of sub-grid topography (*Tesfa et al.*, 2020), variable soil thickness (*Brunke et al.*, 2016), and variably saturated flow dynamics with groundwater (*Bisht et al.*, 2018). While comprehensive offline assessments have been carried out, these developments have not yet been directly implemented in GCMs; further studies are necessary to better understand the sensitivity (*Dwelle et al.*, 2019) of the modeled runoff dynamics to the inclusion of new parameterizations and their parameters. On a related note, confirmation (*Oreskes et al.*, 1994) of model parameters is another vital step to improve the skill of runoff generation simulations (*Huang et al.*, 2013; *Troy et al.*, 2008) that has been long overlooked. In summary, many efforts have been dedicated to improving the realism of large-scale hydrological process and robustness of runoff projections. Continued efforts will need focus on sensitivity of GCM runoff generation to the inclusion of new processes and key controlling factors. It will be necessary to understand whether they lead to the improved space-time representation of runoff process and GCM agreement with large-scale hydrological models that have more sophisticated physical representation of the governing processes.

Acknowledgements

We acknowledge the modeling groups listed in Table 1, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modelling (WGCM) for making the CMIP5 multi-model dataset available. We also thank the Office of Support, U.S. Department of Energy for providing the support for this dataset. We are grateful to Dr. Ben Livneh for making the daily runoff data accessible and downloaded from ftp://livnehpublicstorage.colorado.edu/public/Livneh.2013.CONUS.Dataset/. The scripts used in this study to process the data can be found at https://zenodo.org/record/5042381#.YNtR5xNKjxU. D. Xu was supported by the Dow Sustainability Fellowship at the University of Michigan (https://sustainability.umich.edu/dow). V.Y. Ivanov was partially supported by the NSF grants 1725654 and 1754163. We acknowledge constructive criticism of two anonymous reviewers that helped improve this manuscript.

Competing interests:
The authors declare no competing interests.

References

Abramowitz, M. (1974), *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables*, Dover Publications, Inc.

Arnell, N. W., and S. N. Gosling (2016), The impacts of climate change on river flood risk at the global scale, *Climatic Change*, *134*(3), 387-401.

Barnett, T. P., J. C. Adam, and D. P. Lettenmaier (2005), Potential impacts of a warming climate on water availability in snow-dominated regions, *Nature*, *438*(7066), 303-309.

Berens, P. (2009), CircStat: A MATLAB Toolbox for Circular Statistics, *2009*, *31*(10), 21.

Berghuijs, W. R., R. A. Woods, C. J. Hutton, and M. Sivapalan (2016), Dominant flood generating mechanisms across the United States, *Geophysical Research Letters*, *43*(9), 4382-4390.

Beven, K. (2012), *Rainfall-Runoff Modeling: The Primer 2nd Edition*.

Bisht, G., W. J. Riley, G. E. Hammond, and D. M. Lorenzetti (2018), Development and evaluation of a variably saturated flow model in the global E3SM Land Model (ELM) version 1.0, *Geosci. Model Dev.*, *11*(10), 4085-4102.

Bloschl, G., et al. (2017), Changing climate shifts timing of European floods, *Science*, *357*(6351), 588-590.

Bosmans, J. H. C., L. P. H. van Beek, E. H. Sutanudjaja, and M. F. P. Bierkens (2017), Hydrological impacts of global land cover change and human water use, *Hydrol. Earth Syst. Sci.*, *21*(11), 5603-5626.

Brunke, M. A., P. Broxton, J. Pelletier, D. Gochis, P. Hazenberg, D. M. Lawrence, L. R. Leung, G.-Y. Niu, P. A. Troch, and X. Zeng (2016), Implementing and evaluating variable soil thickness in the Community Land Model, version 4.5 (CLM4. 5), *Journal of Climate*, *29*(9), 3441-3461.

Chaney, N. W., L. Torres-Rojas, N. Vergopolan, and C. K. Fisher (2020), Two-way coupling between the sub-grid land surface and river networks in Earth system models, *Geosci. Model Dev. Discuss.*, *2020*, 1-31.

Chegwidden, O. S., D. E. Rupp, and B. Nijssen (2020), Climate change alters flood magnitudes and mechanisms in climatically-diverse headwaters across the northwestern United States, *Environmental Research Letters*, *15*(9), 094048.

Clow, D. W. (2010), Changes in the Timing of Snowmelt and Streamflow in Colorado: A Response to Recent Warming, *Journal of Climate*, *23*(9), 2293-2306.

Cunderlik, J. M., and T. B. M. J. Ouarda (2009), Trends in the timing and magnitude of floods in Canada, *J Hydrol*, *375*(3), 471-480.

Dai, A. (2006), Precipitation Characteristics in Eighteen Coupled Climate Models, *Journal of Climate*, *19*(18), 4605-4630.

Decharme, B., C. Delire, M. Minvielle, J. Colin, J.-P. Vergnes, A. Alias, D. Saint-Martin, R. Séférian, S. Sénési, and A. Voldoire (2019), Recent Changes in the ISBA-CTRIP Land Surface System for Use in the CNRM-CM6 Climate Model and in Global Off-Line Hydrological Applications, *J Adv Model Earth Sy*, *11*(5), 1207-1252.

Dettinger, M. D. (2005), Changes in Streamflow Timing in the Western United States in Recent Decades—from the National Streamflow Information Program: U.S.*Rep.*, Geological Survey Fact Sheet 2005–3018.

Doocy, S., A. Daniels, S. Murray, and T. D. Kirsch (2013), The human impact of floods: a historical review of events 1980-2009 and systematic literature review, *PLoS Curr*, *5*.

Dudley, R. W., G. A. Hodgkins, M. R. McHale, M. J. Kolian, and B. Renard (2017), Trends in snowmelt-related streamflow timing in the conterminous United States, *J Hydrol*, *547*, 208-221.

Dwelle, M. C., J. Kim, K. Sargsyan, and V. Y. Ivanov (2019), Streamflow, stomata, and soil pits: Sources of inference for complex models with fast, robust uncertainty quantification, *Adv Water Resour*, *125*, 13-31.

Ekici, A., H. Lee, D. M. Lawrence, S. C. Swenson, and C. Prigent (2019), Ground subsidence effects on simulating dynamic high-latitude surface inundation under permafrost thaw using CLM5, *Geosci. Model Dev.*, *12*(12), 5291-5300.

Fatichi, S., V. Y. Ivanov, and E. Caporali (2012), A mechanistic ecohydrological model to investigate complex interactions in cold and warm water-controlled environments: 1. Theoretical framework and plot-scale analysis, *J Adv Model Earth Sy*, *4*(2).

Fatichi, S., et al. (2016), An overview of current applications, challenges, and future trends in distributed process-based models in hydrology, *J Hydrol*, *537*, 45-60.

Fekete, B. M., C. J. Vörösmarty, and W. Grabs (2002), High-resolution fields of global runoff combining observed river discharge and simulated water balances, *Global Biogeochemical Cycles*, *16*(3), 15-11-15-10.

Ghannam, K., T. Nakai, A. Paschalis, C. A. Oishi, A. Kotani, Y. Igarashi, T. o. Kumagai, and G. G. Katul (2016), Persistence and memory timescales in root-zone soil moisture dynamics, *Water Resources Research*, *52*(2), 1427-1445.

Giorgi, F., and L. O. Mearns (2002), Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the "reliability ensemble averaging'' (REA) method, *Journal of Climate*, *15*(10), 1141-1158.

Greve, P., L. Gudmundsson, and S. I. Seneviratne (2018), Regional scaling of annual mean precipitation and water availability with global temperature change, *Earth Syst. Dynam.*, *9*(1), 227-240.

Gudmundsson, L., M. Leonard, H. X. Do, S. Westra, and S. I. Seneviratne (2019), Observed Trends in Global Indicators of Mean and Extreme Streamflow, *Geophysical Research Letters*, *46*(2), 756-766.

Hall, J. W., D. Grey, D. Garrick, F. Fung, C. Brown, S. J. Dadson, and C. W. Sadoff (2014), Coping with the curse of freshwater variability, *Science*, *346*(6208), 429.

Hirabayashi, Y., R. Mahendran, S. Koirala, L. Konoshima, D. Yamazaki, S. Watanabe, H. Kim, and S. Kanae (2013), Global flood risk under climate change, *Nature Climate Change*, *3*(9), 816-821.

Hirsch, R. M., and K. R. Ryberg (2012), Has the magnitude of floods across the USA changed with global CO2 levels?, *Hydrolog Sci J*, *57*(1), 1-9.

Hirsch, R. M., and S. A. Archfield (2015), Not higher but more often, *Nature Climate Change*, *5*, 198.

Hodgkins, G. A., R. W. Dudley, and T. G. Huntington (2003), Changes in the timing of high river flows in New England over the 20th Century, *J Hydrol*, *278*(1-4), 244-252.

Huang, M., Z. Hou, L. R. Leung, Y. Ke, Y. Liu, Z. Fang, and Y. Sun (2013), Uncertainty Analysis of Runoff Simulations and Parameter Identifiability in the Community Land Model: Evidence from MOPEX Basins, *Journal of Hydrometeorology*, *14*(6), 1754-1772.

Ivancic, T. J., and S. B. Shaw (2015), Examining why trends in very heavy precipitation should not be mistaken for trends in very high river discharge, *Climatic Change*, *133*(4), 681-693.

Ivanov, V. Y., R. L. Bras, and E. R. Vivoni (2008), Vegetation-hydrology dynamics in complex terrain of semiarid areas: 1. A mechanistic approach to modeling dynamic feedbacks, *Water Resources Research*, *44*(3).

Kam, J., T. R. Knutson, and P. C. D. Milly (2018), Climate Model Assessment of Changes in Winter–Spring Streamflow Timing over North America, *Journal of Climate*, *31*(14), 5581-5593.

Kemter, M., B. Merz, N. Marwan, S. Vorogushyn, and G. Blöschl (2020), Joint Trends in Flood Magnitudes and Spatial Extents Across Europe, *Geophysical Research Letters*, *47*(7), e2020GL087464.

Kim, J., A. Warnock, V. Y. Ivanov, and N. D. Katopodes (2012), Coupled modeling of hydrologic and hydrodynamic processes including overland and channel flow, *Adv Water Resour*, *37*, 104-126.

Knight, R. R., J. C. Murphy, W. J. Wolfe, C. F. Saylor, and A. K. Wales (2014), Ecological limit functions relating fish community response to hydrologic departures of the ecological flow regime in the Tennessee River basin, United States, *Ecohydrology*, *7*(5), 1262-1280.

Knutti, R., and J. Sedláček (2012), Robustness and uncertainties in the new CMIP5 climate model projections, *Nature Climate Change*, *3*(4), 369-373.

Knutti, R., R. Furrer, C. Tebaldi, J. Cermak, and G. A. Meehl (2010), Challenges in Combining Projections from Multiple Climate Models, *Journal of Climate*, *23*(10), 2739-2758.

Lehner, F., A. W. Wood, J. A. Vano, D. M. Lawrence, M. P. Clark, and J. S. Mankin (2019), The potential to reduce uncertainty in regional runoff projections from climate models, *Nature Climate Change*, *9*(12), 926-933.

Li, D., M. L. Wrzesien, M. Durand, J. Adam, and D. P. Lettenmaier (2017), How much runoff originates as snow in the western United States, and how will that change in the future?, *Geophysical Research Letters*, *44*(12), 6163-6172.

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges (1994), A simple hydrologically based model of land surface water and energy fluxes for general circulation models, *Journal of Geophysical Research: Atmospheres*, *99*(D7), 14415-14428.

Lins, H. F., and J. R. Slack (2005), Seasonal and regional characteristics of US streamflow trends in the United States from 1940 to 1999, *Phys Geogr*, *26*(6), 489-501.

Livneh, B., E. A. Rosenberg, C. Lin, B. Nijssen, V. Mishra, K. M. Andreadis, E. P. Maurer, and D. P. Lettenmaier (2013), A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States: Update and Extensions, *Journal of Climate*, *26*(23), 9384-9392.

Mallakpour, I., and G. Villarini (2015), The changing nature of flooding across the central United States, *Nature Climate Change*, *5*, 250.

Maxwell, R. M., et al. (2014), Surface-subsurface model intercomparison: A first set of benchmark results to diagnose integrated hydrology and feedbacks, *Water Resources Research*, *50*(2), 1531-1549.

McGrane, S. J. (2016), Impacts of urbanisation on hydrological and water quality dynamics, and urban water management: a review, *Hydrological Sciences Journal*, *61*(13), 2295-2311.

Michalak, A. M., et al. (2013), Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions, *Proceedings of the National Academy of Sciences*, *110*(16), 6448.

Milly, P. C. D., K. A. Dunne, and A. V. Vecchia (2005), Global pattern of trends in streamflow and water availability in a changing climate, *Nature*, *438*(7066), 347-350.

Milly, P. C. D., R. T. Wetherald, K. A. Dunne, and T. L. Delworth (2002), Increasing risk of great floods in a changing climate, *Nature*, *415*(6871), 514-517.

Or, D. (2020), The Tyranny of Small Scales—On Representing Soil Processes in Global Land Surface Models, *Water Resources Research*, *56*(6).

Oreskes, N., K. Shrader-Frechette, and K. Belitz (1994), Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences, *Science*, *263*(5147), 641.

Regonda, S. K., B. Rajagopalan, M. Clark, and J. Pitlick (2005), Seasonal Cycle Shifts in Hydroclimatology over the Western United States, *Journal of Climate*, *18*(2), 372-384.

Rogelj, J., M. Meinshausen, and R. Knutti (2012), Global warming under old an new scenarios using IPCC climate sensitivity range estimates, *Nature Climate Change*, *2*(4), 248-253.

Sharma, A., C. Wasko, and D. P. Lettenmaier (2018), If Precipitation Extremes Are Increasing, Why Aren't Floods?, *Water Resources Research*, *54*(11), 8545-8551.

Slater, J. L., and G. Villarini (2017), Evaluating the Drivers of Seasonal Streamflow in the U.S. Midwest, *Water*, *9*(9).

Slater, L. J., and G. Villarini (2016), Recent trends in U.S. flood risk, *Geophysical Research Letters*, *43*(24), 12,428-412,436.

Smith, J. A., M. L. Baeck, G. Villarini, D. B. Wright, and W. Krajewski (2013), Extreme Flood Response: The June 2008 Flooding in Iowa, *J Hydrometeorol*, *14*(6), 1810-1825.

Smith, R. L., C. Tebaldi, D. Nychka, and L. O. Mearns (2009), Bayesian Modeling of Uncertainty in Ensembles of Climate Models, *Journal of the American Statistical Association*, *104*(485), 97-116.

Stephens, G. L., T. L'Ecuyer, R. Forbes, A. Gettelmen, J.-C. Golaz, A. Bodas-Salcedo, K. Suzuki, P. Gabriel, and J. Haynes (2010), Dreary state of precipitation in global models, *Journal of Geophysical Research: Atmospheres*, *115*(D24).

Stewart, I. T., D. R. Cayan, and M. D. Dettinger (2005), Changes toward Earlier Streamflow Timing across Western North America, *Journal of Climate*, *18*(8), 1136-1155.

Swain, D. L., O. E. J. Wing, P. D. Bates, J. M. Done, K. A. Johnson, and D. R. Cameron (2020), Increased Flood Exposure Due to Climate Change and Population Growth in the United States, *Earth's Future*, *8*(11), e2020EF001778.

Tebaldi, C., and R. Knutti (2007), The use of the multi-model ensemble in probabilistic climate projections, *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences*, *365*(1857), 2053-2075.

Tebaldi, C., L. O. Mearns, D. Nychka, and R. L. Smith (2004), Regional probabilities of precipitation change: A Bayesian analysis of multimodel simulations, *Geophysical Research Letters*, *31*(24).

Tebaldi, C., R. L. Smith, D. Nychka, and L. O. Mearns (2005), Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multimodel ensembles, *J Climate*, *18*(10), 1524-1540.

Tesfa, T. K., L. R. Leung, and S. J. Ghan (2020), Exploring Topography-Based Methods for Downscaling Subgrid Precipitation for Use in Earth System Models, *Journal of Geophysical Research: Atmospheres*, *125*(5), e2019JD031456.

Troy, T. J., E. F. Wood, and J. Sheffield (2008), An efficient calibration method for continental-scale land surface modeling, *Water Resources Research*, *44*(9).

van Vuuren, D. P., et al. (2011), The representative concentration pathways: an overview, *Climatic Change*, *109*(1-2), 5-31.

Villarini, G. (2016), On the seasonality of flooding across the continental United States, *Adv Water Resour*, *87*, 80-91.

Winsemius, H. C., et al. (2016), Global drivers of future river flood risk, *Nature Climate Change*, *6*(4), 381-385.

Xu, D., V. Y. Ivanov, J. Kim, and S. Fatichi (2018), On the use of observations in assessment of multi-model climate ensemble, *Stochastic Environmental Research and Risk Assessment*.

Yan, H., N. Sun, M. Wigmosta, R. Skaggs, L. R. Leung, A. Coleman, and Z. Hou (2019), Observed Spatiotemporal Changes in the Mechanisms of Extreme Water Available for Runoff in the Western United States, *Geophysical Research Letters*, *0*(0).

Yang, H., F. Zhou, S. L. Piao, M. T. Huang, A. P. Chen, P. Ciais, Y. Li, X. Lian, S. S. Peng, and Z. Z. Zeng (2017), Regional patterns of future runoff changes from Earth system models constrained by observation, *Geophysical Research Letters*, *44*(11), 5540-5549.

Ye, S., H.-Y. Li, L. R. Leung, J. Guo, Q. Ran, Y. Demissie, and M. Sivapalan (2017), Understanding Flood Seasonality and Its Temporal Shifts within the Contiguous United States, *Journal of Hydrometeorology*, *18*(7), 1997-2009.

Zhai, R., F. Tao, U. Lall, B. Fu, J. Elliott, and J. Jägermeyr (2020), Larger Drought and Flood Hazards and Adverse Impacts on Population and Economic Productivity Under 2.0 than 1.5°C Warming, *Earth's Future*, *8*(7), e2019EF001398.