

Manuscript No: hess-2018-513

Title: Process-Based Flood Frequency Analysis in an Agricultural Watershed Exhibiting Nonstationary Flood Seasonality

Authors: Guo Yu et al.

---

Dear Editor and Reviewers,

We would like to thank the Editor and three anonymous reviewers for providing critical and helpful comments. The central goal has been to create a well-organized paper that highlights the opportunity and challenge with process-based flood frequency analysis (FFA) approaches. The reviewers have contributed greatly to that goal and we have carefully considered each of their criticisms.

The changes that we have made on the basis of these criticisms constitute a major overhaul of the original manuscript, including:

- 1) revisit our model calibration procedure and include the snowpack routine to obtain acceptable performance;
- 2) add a section on model validation to the revised manuscript;
- 3) highlight the impacts of changing flood seasonality on FFA by adding Fig.8 to the revised manuscript;
- 4) add the analysis of pros and cons on process-based FFA approaches to the conclusion section;
- 5) revised/restructured the focus of the introduction and conclusion to clarify the objectives of this study.

Therefore, we would appreciate that the reviewers grant us another through reading.

Sincerely,

Guo Yu

### *Common Responses to three reviewers*

Responses are provided in blue and proposed revision are in Red. Original reviewer comments are in black. Line and page numbers refer to the original manuscript.

Based upon comments from all three reviewers, we have revisited our model calibration procedure and have been able to obtain acceptable performance from the snowpack routine. This involved a “2-step” calibration process in which warm season processes are calibrated first, and then “warm season parameters “ are held constant during subsequent calibration of snowpack-related parameters. This recalibration of HBV is done using both CPC and Stage IV rainfall. We have also added a section on model validation to the revised manuscript, again based on comments from all three reviewers requesting additional validation results. Since all three reviewers provided critiques on these topics, we discuss these two changes before addressing specific comments from individual reviewers.

We have revised model calibration part in the original manuscript, P9, line 15-24, to:

We calibrated the HBV models using both CPC and Stage IV rainfall, and most parameters are the same for CPC- and Stage IV-based models, except for three snow routine parameters (TT, CFMAX, SFCE) and three recession coefficients (K0, K1, K2), allowing for the variability of model parameters for different climate conditions. For each model setup, we first calibrated the model with snowpack routine “turned off” (by setting TT parameter to a very low value) to obtain parameters that can simulate summer floods adequately. Then, keeping these optimized non-snow routine parameters unchanged, we calibrated the snow routine parameters.

To determine the optimized model parameter sets in each procedures, we followed the Genetic Algorithm and Powell (GAP) optimization method as presented by Seibert (2000), which is briefly summarized here. First, 5000 parameter sets are randomly generated from a uniform distribution of the values of each parameter (Table 1), which were then applied to the HBV model in order to maximize Kling Gupta Efficiency (Gupta et al., 2009) of simulated daily discharge. After the GAP has finished, the optimized parameter set were fine-tuned using Powell’s quadratic convergent method (Press, 1996) with 1000 additional runs. Lastly, the optimized parameter set was manually adjusted to improve the fits between observed and simulated annual peak flow (see Lamb, 1999). More elaborate calibration and uncertainty estimation procedures such as Generalized Likelihood Uncertainty Estimation (GLUE; Beven and Binley, 1992; Beven, 1993; Beven and Binley, 2014) could be used, but are outside the scope of our study.

After calibration, HBV (two different parameter sets) was used to perform CS with historical CPC and Stage IV rainfall and temperature data to derive long-term simulated soil moisture and snowpack values, which are usually difficult to obtain via measurement. We “pair” samples of these initial conditions with synthetic rainfall events, as described in Sect. 4.2 and Sect. 4.3.

**Table 1.** Overview of HBV model parameters and prior parameter boundaries.

Parameter	Description	Units	Min value	Max value
<b>Snow Routine</b>				
TT	Threshold temperature for liquid and solid precipitation	°C	-3	3
CFMAX	Degree-day factor	mm d <sup>-1</sup> °C <sup>-1</sup>	0.5	4
SFCF	Snowfall correction factor	-	0.5	1.2
CFR	Refreezing coefficient	-	0.01	0.1
CWH	Water holding capacity of the snow storage	-	0.1	0.3
<b>Soil Moisture Routine</b>				
FC	Maximum soil moisture storage (field capacity)	mm	100	550
LP	Relative soil water storage below which AET is reduced linearly	-	0.3	1
BETA	Exponential factor for runoff generation	-	1	5
<b>Response Routine</b>				
PERC	Maximum percolation from upper to lower groundwater box	mm d <sup>-1</sup>	0	10
UZL	Threshold of upper groundwater box	mm	0	50
K0	Recession coefficient 0	d <sup>-1</sup>	0.5	0.9
K1	Recession coefficient 1	d <sup>-1</sup>	0.15	0.5
K2	Recession coefficient 2	d <sup>-1</sup>	0.01	0.15
<b>Routing Routine</b>				
MAXBAS	Length of triangular weighting function	d	1	2.5

We have also added “Section 5.2 Model Validation” by modifying the original paper, P13-14, to:

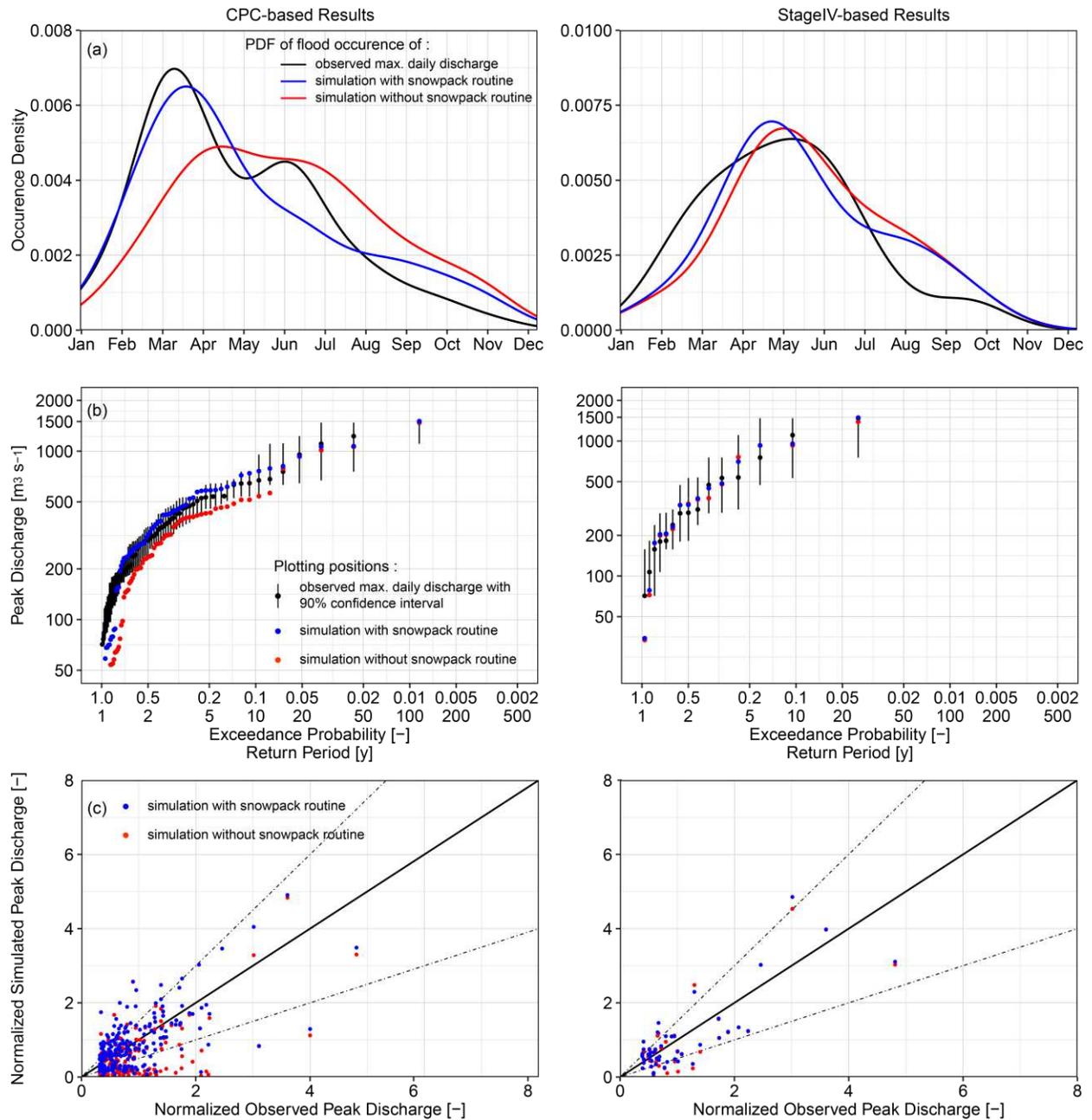
## 5.2 Model Validation

We validated the performance of HBV continuous simulation with respect to flood seasonality, frequency of annual daily discharge maxima, and normalized peak flow (i.e. the simulated or observed daily discharge divided by the 2-year flood), using both Stage IV and CPC as precipitation inputs (Fig. 4). We also validated two structures: one with and the other without the HBV snowpack module. The purpose for this latter validation effort is to highlight the importance of proper process representation (and subsequent validation) in process-based FFA.

Simulated flood seasonality varies substantially during the CPC period of record (1948-2016) depending on the inclusion of the snowpack routine. Differences are less for the Stage IV period of record (2002-2016), due to the decreasing role of snowpack in deriving the floods in recent years (Fig. 4a). In both cases, the seasonality of flooding simulated using HBV is improved with the inclusion of the snowpack module, with a higher (lower) frequency of springtime (summertime) floods which more closely resembles observations. Empirical (i.e. plotting position-based) distributions for the simulated annual daily discharge maxima are mostly within the 90% confidence interval (obtained by nonparametric bootstrap) of the observations (Fig.

4b). The CPC-based simulations differ considerably depending on the inclusion of the snowpack module for more common events, but differences in simulated maxima vanish as flood magnitude increases (e.g. AEP<0.1). This is because the most extreme flood events occur later in the season and are thus independent of snowpack or snowmelt processes. Differences are generally negligible between Stage IV-based simulations with and without snowpack, since floods in this shorter, more recent period are generally driven by summertime thunderstorms. These findings are consistent with the general understanding of the regional seasonality of flooding in the region, as discussed in Sect. 5.1.

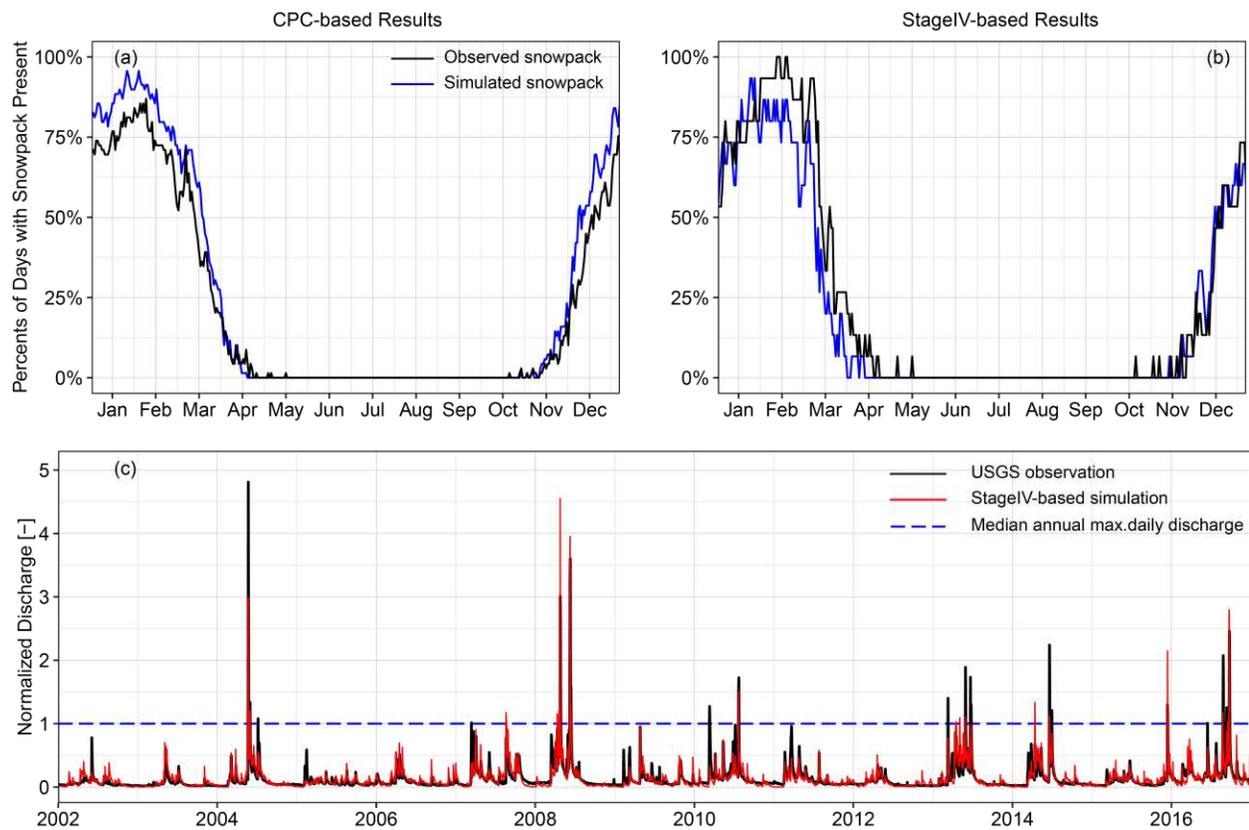
We compared all simulated and observed flood peaks that can be associated with a USGS observed daily streamflow value that is at least three times the mean annual daily discharge (Fig. 4c). When associating simulated and observed flood peaks, we look within a 2-day window to allow for modest errors in simulated flood peak timing. All peaks in Fig. 4c are normalized by the median annual (i.e. 2-year) flood, which, as a rule of thumb, can be considered as the “within bank” threshold. Again, HBV with the snowpack routine outperforms the model without it, especially for the small to modest flood events in CPC-based simulations. The model without snowpack routine underestimate the small to modest flood events in two cases due to the neglect of water flux from potential snowmelt. While modest scatter exists in the Stage IV-based simulated peaks, there is no obvious systematic bias with event magnitude when the snowmelt routine is included.



**Figure 4.** HBV model validation for flood seasonality (a), frequency of annual max. daily discharge (b) and normalized peak flow (c). For each panel, the corresponding model validation is performed against CPC- (1948-2016) and StageIV-based (2002-2016) simulation and the results derived from HBV model with (without) snowpack routine are shown in blue (red). The 90% confidence interval for observed max. daily discharge (empirical distribution) is derived using the bootstrapping approach. Peak discharge is defined as a data point with USGS observed value that is at least three times the average observations, and peak discharge are normalized by the median of annual daily discharge maxima (i.e. the 2-year flood). Straight black lines indicate 1:1 correspondence, while dashed lines indicate the envelope within which the modeled values are within 50% of observed.

We also validate HBV's snowpack routine using observed GHCN daily snow depth for two simulation periods (Fig. 5a, 5b) and using USGS daily streamflow observations for Stage IV-based period (Fig. 5c). Because of their differing spatial resolutions and physical representations, point-scale GHCN daily snow

depths cannot be directly or quantitatively compared to the watershed-scale snow water equivalent simulated by HBV. Therefore, we validate the snowpack simulation in terms of the snowpack occurrence, defined as the number of occurrences where snow is present on a particular date divided by the total number of years in the historical record. For example, there are 50 days where snowpack is present on January 1st in the 69-year period from 1948-2016, based on GHCN observations and thus the corresponding occurrence rate is 0.72 (50 divided by 69). The HBV model with the snowpack routine captures the central tendency of observed snowpack dynamics, showing that snowpack frequently exists from early November to mid-February, with frequency of snow decreasing from late February until disappearing in early April.



**Figure 5.** The comparison of percent of days with snowpack present between observations and simulations (a, b) and hydrograph validation for StageIV-based simulation (c). For each day within a year, the percent of snowpack existing days is calculated as the ratio of the number of years when snowpack is present to the total years (69 years for CPC and 15 years for StageIV). Observed and simulated hydrograph are normalized by the median annual flood, which is indicated by the dashed blue line.

Model hydrograph validation is provided in Fig. 5c for the Stage IV period (2002-2016), when major flooding occurred throughout Iowa. Model performance shows no obvious evidence of systematic bias in the streamflow simulations. Although flood seasonality derived by Stage IV-based simulation differs slightly from observations (Fig. 4b), these mismatches are associated with flood events smaller than the median annual flood (blue dash line in Fig. 5c). Stage IV-based simulations do not show bias flood magnitude in late summer. In other words, remaining biases in terms of flood seasonality generally

correspond with frequent, small-magnitude events that are typically of less interest in FFA. We therefore conclude that the HBV model with snowpack is generally suitable for subsequent process-based FFA.

## *Replies to the comments of Anonymous Referee #1*

Responses are provided in blue and proposed revision are in Red. Original reviewer comments are in black. Line and page numbers refer to the original manuscript.

The authors explore the utility of hydrological simulations driven by stochastically transposed rainfall fields in deriving flood frequency over a watershed that experiences nonstationarities. Their results highlight the importance of considering changing flood seasonality in flood frequency analysis. While process-based approaches have a fair amount of advantages, their shortcomings are also quite obvious, for instance, mode uncertainty in both parameters and model structure, representation of synthetic rainfall scenarios, etc. As a hydrologist, I would still favor statistical approaches if the gauging record is good (as is the case in this paper). This being said, I would suggest the authors focus on explaining the importance of changing flood seasonality in flood frequency, but rather demonstrating the superiority of process-based approaches to other FFA methods (which is not, as far as I can see).

We thank the reviewer for these useful critiques, that have been very helpful in improving the paper. We fully agree that, particularly in situations of plentiful stream gage observations, statistical approaches are generally preferable. It was never our intention to suggest that our approach is superior to such methods. We note in the original manuscript (P4, line 23-26), however, that there have been prior studies that have demonstrated situations in which rainfall-runoff modeling approaches of various kinds can outperform statistical methods. This, combined with the relative immaturity of rainfall-runoff model-based FFA approaches compared with statistical methods, suggests that additional research, of the kind we present here, can and should be done. As the reviewer stresses, one of the things that such research can point to is the importance of processes and their changes (e.g. seasonal to interannual). In our revised manuscript, we have attempted to emphasize our viewpoint on these issues more clearly. Example revisions to this effect include:

- Include the snowpack routine in the HBV model for both CPC- and StageIV-based simulations.
- Modify the model calibration part (Chapter 4.1) in the original manuscript.
- Add a new section for model validation.
- Address the importance of changing flood seasonality in flood frequency.

We have analyzed two sets of CPC-based results, one for 1948-2016 and the other for 2002-2016 to demonstrate how the changes in flood agents affect the FFA results. We have added the following part to Sect.5.3, P17, line 21 of the original manuscript.

To demonstrate that the discrepancies between the process-based FFA results generated using CPC and using StageIV are driven by changes in flood agents, rather than by differences in model structure (i.e. parameter values), we compared FFA results generated using CPC-based for 1948-2016 and 2002-2016, in terms of event rainfall, initial soil moisture, flood type and peak magnitude (Fig. 8). From 2002-2016 (Fig. 8b), there are fewer flood events driven by snowmelt or rain-on-snow but more driven by rainfall, particularly large magnitude flood events (over 1000 m<sup>3</sup>/s). In addition, some of the rainfall driven floods

(upper left of Fig. 8b) from 2002-2016 indicates high initial soil moisture, which are in accordance with the significant increasing trend of annual precipitation (Table 2). In general, changes in individual flood agents and their interactions can affect flood frequency. Process-based approaches can help illuminate these changes.

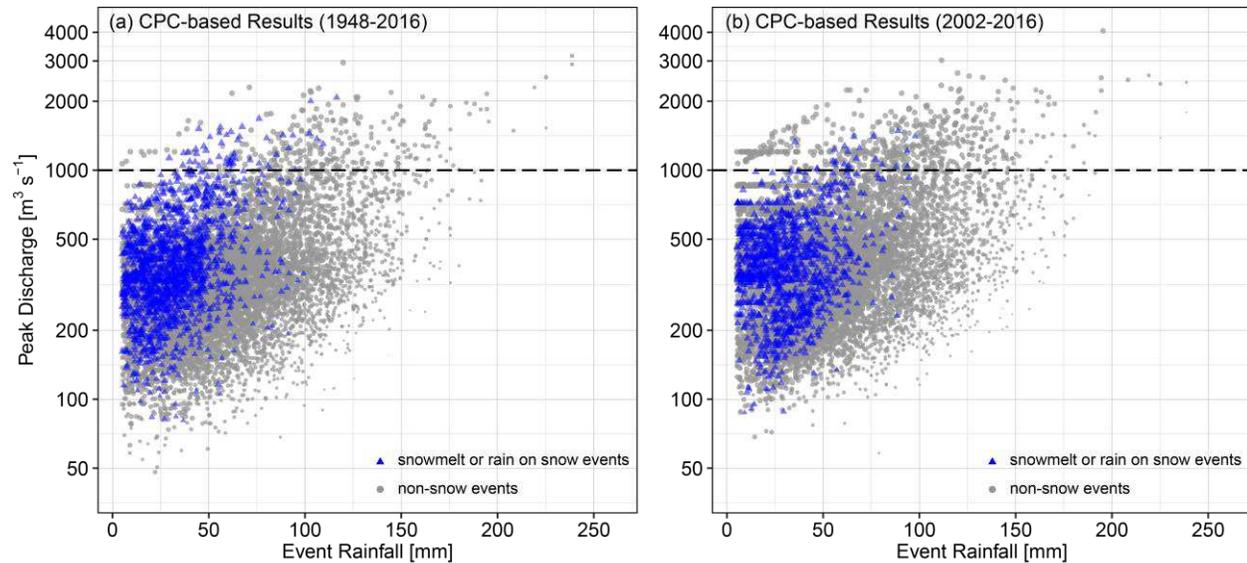


Figure 8. The simulated flood magnitude using CPC rainfall during 1948-2016 (a) and 2002-2016 (b) period, and corresponding antecedent conditions sampled from the continuous simulation. The blue triangles represent the snow related flood events (e.g. snowmelt or rain on snow) and grey dots represents the non-snow related flood events (e.g. rainfall driven). The size of the triangles or dots indicate the antecedent soil moisture with higher value in larger shape. The black dash line indicates the 1000m<sup>3</sup>/s flood magnitudes.

Specific comments 1-1: An important part is missing from the present paper is model validation. Evidence needs to be explicitly presented to show the capability of long-term model simulations in capturing, for instance, flood seasonality, as well as other features (distribution of annual maximum discharge). This can be done by adding simulation results into Figure 3b and Figure 5a.

We thank the reviewer for this suggestion, which was also voiced by the other reviewers. We have include the model validation, as shown at the beginning of this response, to further demonstrate the capability of long-term simulation in capturing the flood seasonality, high flow magnitude and distribution of annual maximum discharge. We hope the reviewers find it to be more convincing that the limited validation that we included in the original manuscript.

Specific comments 1-2: The authors show a larger frequency of floods during post-summer season in their simulations, could this be possibly related to the positive model biases in representing rainfall-runoff processes during this season? The reliability of process-based approaches in FFA builds on decent model simulations. The authors should spend additional efforts in demonstrating this in the paper. This can be done by providing a quantitative assessment of the model performance.

The hydrograph validation plot (Figure. 5b), along with the flood seasonality validation plot (Figure. 4a) shows that the HBV model with snowpack routine can capture the observed flood seasonality and daily streamflow in the long-term simulation. Although model simulates more flood events in late summer (August-September), it is not biased in terms of late summer flood magnitude. Therefore, we believe these simulated extreme late summer flood events (over 1500 m<sup>3</sup>/s) are associated with the regional late-summer storm events in Iowa, rather than model bias.

Specific comments 1-3: Another question about the simulation, how is channel flow represented/considered in the analyses. Antecedent streamflow in the channels can be an important element in representing antecedent watershed wetness, in addition to soil moisture, that plays a role in streamflow simulation.

The reviewer is correct in general that this should be considered within our framework. The HBV model, however, does not need to sample channel flow (streamflow) for the antecedent conditions. The following equations show how the HBV model calculates the streamflow.

$$Q[t] = Q_0[t] + Q_1[t] + Q_2[t]$$

$$Q_0[t] = K_0 * MAX(SUZ[t - 1] + recharge[t] + excess[t] - UZL, 0)$$

$$Q_1[t] = K_1 * (SUZ[t - 1] + recharge[t] + excess[t] - UZL - PERC)$$

$$Q_2[t] = K_2 * (SLZ[t - 1] + PERC)$$

Where conceptually,

$Q[t]$  is the current time streamflow

$Q_0[t], Q_1[t], Q_2[t]$  are the current time overland flow, intermediate flow and baseflow

$recharge[t], excess[t]$  are the current time flux to groundwater and excess runoff, all of which depend on the soil moisture at previous time step

$SUZ[t - 1], SLZ[t - 1]$  are the water level in upper and lower groundwater box at the previous time step

$K_0, K_1, K_2, UZL, PERC$  are model parameters

In general, the current time overland flow ( $Q_0[t]$ ) and intermediate flow ( $Q_1[t]$ ) only depend on soil moisture and water level in the upper groundwater box at the previous time step while the current time baseflow  $Q_2[t]$  depends on the water level in the lower groundwater box at the previous time step. The more details on HBV model structure can be found in the HBV references in original manuscript, P3, line 4-7.

Specific comments 2-1: The representation of synthetic rainfall fields is another key in process-based FFA approaches. The authors mentioned that they chose 'most intense rainfall events' within a prescribed domain. How exactly do they define "most intense rainfall events"? Please explain.

The RainyDay software selects the most intense rainfall events within the transposition domain, in terms of rainfall accumulation of duration  $t$  and with the same size, shape, and orientation of

the watershed. For example, the principal axis of the Turkey River watershed in this study is oriented roughly northwest-southeast and has an area of 4002 km<sup>2</sup>. In this case, the 450 selected storms from the historical rainfall data are those associated with the 450 highest 96-hour rainfall accumulations over an area of 4002 km<sup>2</sup> with the same shape and orientation as the Turkey River watershed.

We have modified P10, line 13-14 to:

These intense storms are in terms of 96-hour rainfall accumulation and have the same size, shape, and orientation of the Turkey River watershed, which is oriented roughly northwest-southeast and with an area of 4002 km<sup>2</sup>. In order to avoid overlapping storms, these selected events must be separated by at least 24 hours.

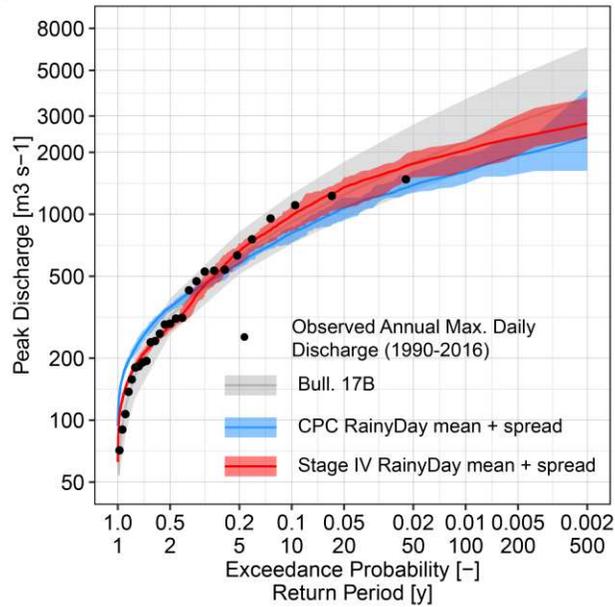
Specific comments 2-2: The authors use the word “realistic” throughout the paper which is inappropriate or miss-leading. They are using synthetic rainfall fields, even though based on real storm events. Please modify.

We believe that our word choice is reasonable when referring the SST-based rainfall fields. They require no parameterization or assumption regarding their spatial or temporal structure (only their starting location is changed), and thus are objectively more realistic than more conventional stochastic rainfall generators. The “realistic” claim would be admittedly more suspect in an environment with complex terrain features (e.g. mountains, coastlines) where both radar estimates and transposition of rainfall fields would be more suspect. Most references on SST in original manuscript, P9, line 30-31, also used word “realistic rainfall”.

Specific comments 3: The authors show flood frequency estimates in modern times using Stage IV rainfall fields, and the results match well with gauging records. How about the performance of CPC rainfall in estimating flood frequency?

The RainyDay based FFA using CPC-Unified rainfall data from 2002 to 2016 closely resembles the Stage IV-based FFA, as we mentioned in original manuscript, P16, line 12-15. Regardless, we have added a supplementary plot showing the CPC, Stage IV and Bull.17B based FFA for the modern time (2002-2016).

Supplementary Fig. 1 shows two features that result using CPC data. First, the extreme tail is underestimated, relative to the Stage IV-based simulations and the statistical approach. CPC is known to contain errors in the extreme tail, due to gage undercatch, insufficient gage density to properly sample convective rain cells, and spatial averaging of such cells over large areas, which effectively reduces peak rainfall depths. Second, CPC overestimates the magnitude of more frequent events. This is likely the result of its coarse spatial resolution, which will “smear” rainfall over larger areas (i.e. entire ~600 km<sup>2</sup>) grid cells when it should be more localized. This would serve to increase the likelihood of rainfall over the watershed, albeit at relatively lower depths/intensities. Thus, if one is to restrict the time period of the rainfall data to recent years (for example, the 2002-2016 time period for which Stage IV is available), then Stage IV would likely be better. As an aside, this belief that Stage IV is preferable to other precipitation datasets in the United States is widely shared in the satellite precipitation community, where Stage IV is often used as a validation dataset.



**Supplementary Figure 1.** Three peak discharge analyses for Turkey River at Garber, IA: RainyDay with Stage IV (2002-2016) and CPC-(2002-2016) rainfall and USGS frequency analyses (1990-2016) using Bulletin 17B methods. Shaded areas denote the ensemble spread (RainyDay-based results) and the 90% confidence intervals (Bulletin 17B-based analysis), respectively. All observed annual daily streamflow maxima from 1990 to 2016 are shown in black dots.

Specific comments 4: An interesting finding in the paper is described in P17 Line 15-20, but needs to be rephrased. We can see summer floods dominate the upper tail of flood frequency in this region, even though they do not occur as frequent as spring floods. The distribution derived from gauging records is still the ‘truth’ anyway. Under-representation of summer floods is a pretty common feature of flood peak distributions in the US. I would suggest the authors to provide a brief diagnostic summary of the most extreme flood events in this region.

This is a good suggestion and the newly-added model validation section includes seasonal validations (5.2 Model Validation), as shown at the beginning of this response. Model validations, with respect to flood seasonality, normalized peak flow and hydrograph, show that HBV does not show bias flood magnitude in late summer.

A summary of the most extreme flood events in Iowa are provided in the section 5.1 of the original manuscript, and is provided here: “Flood peak distributions in Iowa “mixtures” of two basic types. Spring floods are associated with springtime rains, high soil moisture, and potentially snowmelt. Summer floods are associated with convective systems. The latter have been shown to significantly affect the upper tail of the flood peak distribution (Villarini et al, 2011) who showed that about 40% of the largest flood peaks are during the May-July period in Iowa. It is important that any process-based FFA approach capture the influence of this mixture on the flood frequency curve.” This does not imply that individual gage records are “the truth”, only the best representative of it that we have. Thus, discrepancies between model-based approaches and such as ours and observational records warrant further attention.

Specific comments 5: The authors compared simulation results using model with and without snow module, and suggest in the paper that “the modeler must either have sufficient data to diagnose such issues or have sufficient prior knowledge.” (P18 Line 14). I would believe a snow module should be needed in simulation hydrological regimes in this region (dominant spring floods in flood frequency). We cannot simply opt out the snow module by simply checking the simulation. What prior knowledge do the authors have? I would suggest the authors to examine the observed snow climatology over this region, and more ideally, carry out detailed diagnostic analyses of flood agents in this region.

This is a very useful critique. We took this advice into account and developed a new calibration approach that avoids some of the pitfalls that we encountered using more standard calibration techniques. As shown above, we validate this new calibration with respect to flood seasonality, hydrograph, normalized peak flow and snowpack. We finally conclude that the snowpack routine of HBV is indeed important in this study region for this application. We appreciate the insistence of all reviewers in this regard, since it has led to a stronger and more defensible methodology.

Specific comments 6-1: P22 Line5-7, it is not true that conventional statistical FFA methods underestimate flood frequency. At this stage, I would still believe statistical estimates are the ground truth, which enables the evaluation of the process-based approach. The authors do not show updated Bulletin 17B curves using the 1990-2016 flood records in Figure 5, which I would suggest to update.

Figure 5 in the original manuscript shows that conventional FFA methods (defined here as usage of stationary statistical distributions fitted to the period of record using a standard fitting software) underestimate flood frequency beyond the 2-5 year recurrence interval. The statistical fits shown in Figure 5 are included to emphasize that we neglect nonstationarity (as is typically done in FFA practice) at our peril, and usage of “old” data in the face of pronounced hydrologic change can produce incorrect results. We therefore must contend that statistical estimates in such situations should not be considered “ground truth.” Bulletin 17B-based results using 1990-2016 flood peaks are shown in the figure above (see responses to specific comment 3). This fits the observed flood peaks well, as one would expect, though obviously subject to substantial uncertainty for low AEP events due to the short fitting period. Other methods, such as nonstationary FFA, could be used, but our goal is not to prove the superiority of one method or another, but rather to highlight some important issues regarding flood physical processes, their changes, and the resulting implications for flood frequency, issues which are generally ignored in conventional analysis.

As I have mentioned earlier in general comments, it is not wise for the authors to demonstrate the dominating superiority of process-based FFA approaches in this paper, at least for this region. Process-based approach, as presented in this paper (hydrological model + SST), can be highly recommended in poorly gauged watersheds. For poorly-gauged watersheds, however, another issue arises as how to obtain a large ensemble of antecedent watershed wetness conditions used in event-based model simulations. The authors need to provide a discussion about both pros and cons of the proposed approach.

Again, our intention was not to argue for the superiority of process-based methods, and we regret that we gave the reviewer that impression. We have modified the manuscript to make more clear

the point that we are attempting to highlight the importance of flood processes and their changes in “shaping” flood frequency, and show an approach that can begin to account for such processes and their changes—though more work is needed, and is ongoing within our research group and elsewhere. Additionally, we agree with the reviewer that a brief discussion about both pros and cons of our framework is necessary.

We have revised the last paragraph of the conclusion to:

A number of issues remain that make broader usage of our process-based framework challenging. Perhaps the biggest limitation of process-based approaches is the necessity of discharge observations, which are central to both identifying hydrologic changes and to calibrate and validate the hydrologic model. Thus, usage of the approach in ungauged basins may not produce satisfactory results. This issue is fundamental to other FFA techniques as well. Statistically-based discharge analyses, for example, similarly rely on streamflow observations, while design storm approaches also require hydrologic model calibration.

Our framework highlights the opportunity and challenge with process-based FFA approaches; namely, that progress on understanding and estimating flood frequency and how it is evolving in an era of unprecedented changes in land use and climate requires better understanding of how the underlying physical processes, and the interactions between them, are changing. Poor model representation of key hydrological processes, however, can lead to incorrect conclusions about present or future flood frequency. Despite the challenge, we share the view of Sivapalan and Samuel (2009) that process-based approaches hold great potential for advances in FFA research and practice, particularly in projecting the future FFA when coupled with high resolution climate model. We do not propose that process-based approaches should necessarily supplant more conventional discharge-based analyses, and discharge observations were central to our present study. Rather, we anticipate a gradual “merging” of statistical and process-based stochastic simulation techniques as well as of the associated observations and synthetic data.

I have a couple of additional comments on word expressions, paragraph organizations, etc., but they can wait till the second round of review. The paper can be a worthwhile contribution to the literature subject to major revisions.

We look forward to further feedback from the reviewer. We have also made minor modification to the structure and word choice in the revised version.

## *Replies to the comments of Anonymous Referee #2*

Responses are provided in blue and proposed revision are in Red. Original reviewer comments are in black. Line and page numbers refer to the original manuscript.

The work presents an investigation of flood frequency in the Turkey River basin in the Midwestern United States. The proposed framework, referred to as “process-based” FFA, uses stochastic storm transposition to generate synthetic storms and a lumped hydrologic model to simulate discharge at the outlet of the basin. The authors carry out a series of simulations and corresponding analyses of flood frequency to investigate the impact of seasonality in FFA and potential changes between past and present conditions. Overall, the work has several nice features and the questions posed by the authors are interesting. However, I have some major concerns about certain elements of the proposed framework that need to be addressed before the work can be considered for publication. I provide below major and minor comments that will hopefully help.

We thank the reviewer for these useful critiques, which have been very helpful in improving the manuscript.

Major comments 1: My first and most important concern about the proposed work is related to the choice of the hydrologic model used. The authors mention in different sections themselves that using a lumped model has several limitations. It is good that they acknowledge this limitation themselves but this does not solve the problem. In fact, based on statements as in Line 13, Page 15 “We did not use the snowpack routine...it was shown to produce unrealistic streamflow results” and given that snow processes are important in the selected basins, one immediately recognizes that the choice of the model is not appropriate. If we combine this with the author’s statement in conclusions “L22-23, page 22: Poor model representation of key hydrological processes, however, can lead to incorrect conclusions about present and future flood frequency”...I am very skeptical about the conclusions derived based on this model’s results. If the model cannot represent well snow processes (particularly flooding due to rain on snow, which should be important in the area) then I fear that the “process-based” FFA is flawed. In this case, the work should be presented at most as a sensitivity analysis and statements such as L1, P22 “helps shed light on the physical processes that shape flood frequency” should be rephrased accordingly.

This is a valid criticism and we thank the reviewer. We hope that the added model calibration and validation, as shown in the beginning of this response, addresses most of the reviewer’s present concern. As shown, we have devised a new calibration approach that provided acceptable performance while included the snowpack routine in the HBV model, since we agree with the reviewer that snow processes are potentially important elements of flooding in the region and should not be omitted.

Major comments 2: The calibration and validation of the model lacks clarity. Which forcing was used to calibrate the model? And how the model was validated? These points are not clear in section 4.1. Then in section 5.2 L13,P15 “Different HBV parameters are used...” suggests that separate parameterization was used for the different precipitation forcing but no evidence is

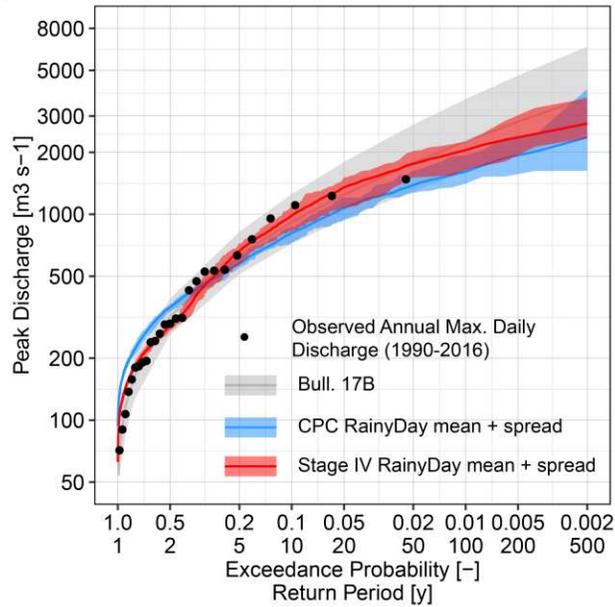
provided on a) the validation of the model for the two dataset and b) the variability in model parameters. For the later, if the parameters are significantly different, it will highlight further problems with the approach since this will mean that CPC HBV and CPC-Stage IV simulations treat hydrological processes differently (i.e. may give more weight to different processes in each case). This needs to be investigated and clearly explained in order to understand whether the results can be considered “realistic” or are results of a numerical exercise that mixes two different things.

We hope the updated model calibration can help reviewers find our process-based FFA to be less speculative and more convincing. While ideally model parameters could remain constant regardless of the rainfall dataset used, this is generally not good modeling practice, since rainfall error structures can differ substantially between datasets. For example, due to its much coarser spatial resolution, CPC, even when used in a lumped model, will produce more frequent light rain and lower extremes than Stage IV. Therefore, we believe that calibration for individual input datasets is a necessary evil. Our future research will use distributed physics-based models in place of HBV, and hopefully this is less of an issue in such models.

Major comments 3: For the results in Fig. 5 right panel: Do you use soil moisture years prior to 1990 for the StageIV process-based approach? Also, you should apply the Bull. 17B for the two periods (1933-1989 and 1990-2016) and add them on the graph for comparison.

We did not use the soil moisture prior to 1990 for the Stage IV-based simulation. The antecedent conditions for Stage IV-based simulation are only sampled from continuous simulation of Stage IV period, which is 2002-2016. We have not applied the Bull.17B method to annual daily streamflow maxima for 1933-1989 period because we have not investigated any RainyDay-based simulation for the corresponding time. However, we have added a supplementary plot showing the CPC, Stage IV and Bull.17B based FFA for the modern time (2002-2016), similar to what this reviewer and reviewer 1 suggest.

Supplementary Fig. 1 shows that process-based FFA using CPC precipitation from 2002-2016 closely resembles the Stage IV-based FFA, suggesting that rainfall differences, rather than model structures, are the primary drivers of the differences in this figure. It also shows two features that result using CPC data. First, the extreme tail is underestimated, relative to the Stage IV-based simulations and the statistical approach. CPC is known to contain errors in the extreme tail, due to gage undercatch, insufficient gage density to properly sample convective rain cells, and spatial averaging of such cells over large areas, which effectively reduces peak rainfall depths. Second, CPC overestimates the magnitude of more frequent events. This is likely the result of its coarse spatial resolution, which will “smear” rainfall over larger areas (i.e. entire ~600 km<sup>2</sup> grid cells) when it should in reality be more localized. This would serve to increase the likelihood of rainfall over the watershed, albeit at relatively lower depths/intensities. Thus, if one is to restrict the time period of the rainfall data to recent years (for example, the 2002-2016 time period for which Stage IV is available), then Stage IV would likely be a better choice.



**Supplementary Figure 1.** Three peak discharge analyses for Turkey River at Garber, IA: RainyDay with Stage IV (2002-2016) and CPC-(2002-2016) rainfall and USGS frequency analyses (1990-2016) using Bulletin 17B methods. Shaded areas denote the ensemble spread (RainyDay-based results) and the 90% confidence intervals (Bulletin 17B-based analysis), respectively. All observed annual daily streamflow maxima from 1990 to 2016 are shown in black dots.

Minor comments 1: P1, L18: “a watershed that is undergoing significant climatic... change”. Is the climatic change at the scale of the watershed only? Consider revising.

We have revised this sentence to:

The methodology is applied to the Turkey River watershed in the Midwestern United States, which is undergoing significant climatic and hydrologic change.

Minor comments 2: P16, L2: “but higher estimates” should be “but gives higher estimates”?

Correct. We have modified that sentence to “but yields higher estimates for rarer events”.

Minor comments 3: Fig.6: Improve caption. What is the upper and what the lower panel?

This figure has been updated.

Minor comments 4: P18L13: “processes in her” should be “processes in his/her”

We have updated the text.

### *Replies to the comments of Anonymous Referee #3*

Responses are provided in blue and proposed revision are in Red. Original reviewer comments are in black. Line and page numbers refer to the original manuscript.

This combination of continuous and event based modelling is a quite novel idea and provides a flexible framework for DFFA. The application of the methods seems sound, the research is done systematically and the paper reads quite well. However, I do have some concerns regarding the selection of the hydrological model, the selection of two precipitation data sets and some of the conclusions. I will detail these below in the major comments, followed by some minor comments. The paper is worth to be published after major revision.

We thank the reviewer for these useful critiques, which have been very helpful in improving the paper. We address these issues more deeply in specific responses, but generally speaking: 1.) in the revised manuscript, we have reintroduced the snowpack routine in the HBV and calibrate and validate the model more carefully. We discussed the model validation with respect to the flood seasonality, peak flow, snowpack, and hydrographs. 2.) we discuss the limitations of CPC precipitation data and the reason why we include the Stage IV precipitation data in this process-based FFA framework. 3.) we provide a short summary of the pros and cons of the proposed FFA framework.

Major comments 1: The selection of the lumped HBV model is not plausible to me, especially given that a) the snow routine is not working and b) the high resolution StageIV rainfall data cannot be utilized by this lumped model.

Since we have updated the HBV model by including the snowpack routine and validated the model as shown in the beginning of this response, we hope the reviewer finds the selection of the lumped HBV model to be more convincing. It also should be noted that, the process-based FFA methodology employed in this study could be coupled with other (sophisticated) hydrologic models, as we mentioned in the original manuscript, P9, line 10, and, in fact, that is our next research direction. Nonetheless, after decades of research, lumped models have still proven to be very useful in a variety of hydrologic fields including flood applications and research. One challenge that we faced in this study was how to quickly implement and evaluate modifications and additions to the methodology, which can be much slower and more challenging using a more sophisticated distributed model.

We respectfully disagree that the Stage IV rainfall data cannot be utilized by a lumped model. Regardless of model choice, Stage IV precipitation data is generally better than CPC data in the study region, in terms of accuracy-this is evident, for example, in the fact that the satellite precipitation community routinely uses Stage IV and related gage-corrected radar products, rather than CPC, to validate satellite rainfall estimates. CPC is known to contain errors in the extreme tail, due to gage undercatch, insufficient gage density to properly sample convective rain cells, and spatial averaging of such cells over large areas, which effectively reduces peak rainfall depths. Second, CPC overestimates the magnitude of more frequent events. This is likely the result of its coarse spatial resolution, which will "smear" rainfall over larger areas (i.e. entire ~600 km<sup>2</sup>) grid cells when it should be more localized. This would serve to increase the likelihood of rainfall over

the watershed, albeit at relatively lower depths/intensities. Thus, if one is to restrict the time period of the rainfall data to recent years (for example, the 2002-2016 time period for which Stage IV is available), then Stage IV would likely be better. It is true that the lumped model cannot “leverage” the rainfall spatial structure embedded in Stage IV, but it still benefit from its improved accuracy.

Major comments 2: The application of two rainfall data sets is not plausible and also quite confusing for the reader since a) the Stage IV rainfall data observation period (2002-2016) is covered also by the CPC rainfall data observation period (1948-2016), b) a lumped hydrological model cannot really benefit from high resolution rainfall data (see 1) and c) the hydrological simulation results for both rainfall data sets seem to be very similar (as the authors state on page 16, lines 12-13). I would recommend to do all the simulations with the CPC rainfall if the hydrological model is not changed. If a more suitable hydrological model is selected the two data sets might be kept in the study but the differences in hydrological response using the two data sets for the same time period (2002-2016) need also to be demonstrated and discussed.

We feel that including the Stage IV-based simulation in this case study is important in two respects: 1.) As mentioned in the response to comment 1, we believe the Stage IV precipitation data has high accuracy than CPC. As an aside, this belief that Stage IV is preferable to other datasets when long records are not required is widely shared in the satellite precipitation validation community, where Stage IV is often used as a validation dataset. 2.) We also want to highlight that using only 15 years of rainfall records, our process-based approach can produce accurate estimates of “present-day” flood frequency.

In addition, we have analyzed two CPC-based results from 1948-2016 and 2002-2016 to demonstrate how the changes in flood agents affect the FFAs. We have added the following part to Sect.5.3, P17, line 21 of the original manuscript.

To demonstrate that the discrepancies between the process-based FFA results generated using CPC and using StageIV are driven by changes in flood agents, rather than by differences in model structure (i.e. parameter values), we compared FFA results generated using CPC-based for 1948-2016 and 2002-2016, in terms of event rainfall, initial soil moisture, flood type and peak magnitude (Fig. 8). From 2002-2016 (Fig. 8b), there are fewer flood events driven by snowmelt or rain-on-snow but more driven by rainfall, particularly large magnitude flood events (over 1000 m<sup>3</sup>/s). In addition, some of the rainfall driven floods (upper left of Fig. 8b) from 2002-2016 indicates high initial soil moisture, which are in accordance with the significant increasing trend of annual precipitation (Table 2). In general, changes in individual flood agents and their interactions can affect flood frequency. Process-based approaches can help illuminate these changes.

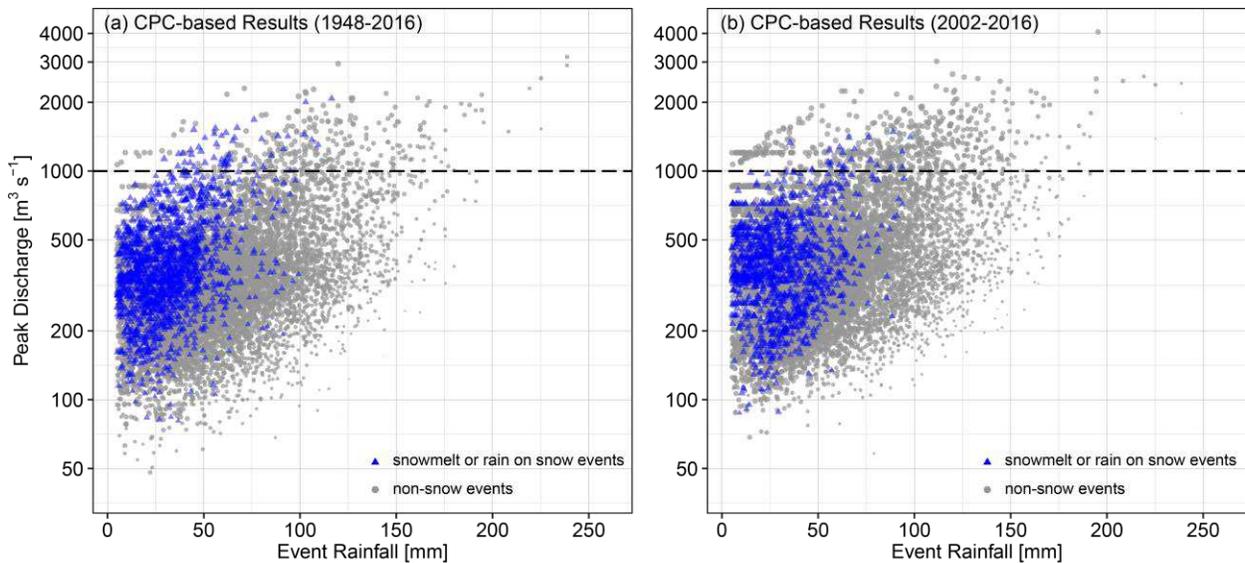


Figure 8. The simulated flood magnitude using CPC rainfall during 1948-2016 (a) and 2002-2016 (b) period, and corresponding antecedent conditions sampled from the continuous simulation. The blue triangles represent the snow related flood events (e.g. snowmelt or rain on snow) and grey dots represents the non-snow related flood events (e.g. rainfall driven). The size of the triangles or dots indicate the antecedent soil moisture with higher value in larger shape. The black dash line indicates the 1000m<sup>3</sup>/s flood magnitudes.

Major comments 3: The application of a model without snow routine for a catchment with significant snow processes doesn't make sense to me. This way the advantage of process based flood frequency analysis (FFA) is partly lost; obtaining the correct hydrological response for the wrong reason is not satisfying. I am not convinced that the non-stationarity in seasonality is only due to changed soil moisture conditions from rainfall. Temporarily shifted snow dynamics might play a role as well

After taking the reviewers' comments into account very seriously, we recalibrate our model with snowpack routine "turned on" and validate it with respect to flood seasonality, hydrograph, normalized peak flow and snowpack. We finally conclude that the snowpack routine of HBV is indeed important in this study region.

Major comments 4: I would be careful with the conclusion, that only with this DFFA method nonstationarity in seasonality can be handled well. Also, non-stationary seasonal FFA approaches are available employing mixed distributions for getting final design values. This needs to be briefly discussed.

We appreciate the comment. Certainly seasonality could be considered using other approaches, though mixture distribution approaches may still not elucidate the fundamental drivers that "shape" flood frequency, even if they can provide good end results. We are not aware of such approaches being used in widespread practice, at least in the United States. Nonetheless, we had added a brief comment in this regard to the conclusions in acknowledgement of this criticism.

We have revised the first paragraph of Section 6 on P21, line 13-15, to :

It must be noticed that the statistical approaches coupling with flood seasonality indices can also investigate the impacts of seasonality on FFA and improve the flood frequency estimation in a regional scale (Ouarda et al., 2006). Our aim is to estimate flood quantiles by reconstructing meteorological and hydrological processes and their interactions, providing an alternative approach which is also well-suited to nonstationary environments (see also Sivapalan and Samuel, 2009).

Major comments 5: This combination of continuous and event based modelling is a good idea. However, there is an important limitation which should at least be mentioned. The framework provides only one possible realization of initial conditions. Nature is more variable. Stochastic rainfall models producing continuous rainfall don't pose this limitation on hydrology.

Each event-based simulation is randomly paired with initial conditions drawn from a continuous simulation (15 years in the case of Stage IV, 69 years for CPC). Thus, we would argue that a large number of possible realizations of initial conditions are used. We would direct the reviewer to Section 4.3. If the reviewer finds this description incomplete, we would appreciate suggestions for how we can make this point more clear. Though we have not tested rigorously, we would guess that relatively short records (say, 15 years) of continuous simulations are sufficient to obtain enough variability in initial conditions. Compared with rainfall, soil moisture (which is bounded between 0 and saturation) and springtime snowpack have thinner tails and thus easier to represent in our framework by sampling from relatively short continuous simulation.

We agree that continuous stochastic rainfall models also have the ability to produce a wide range of pre-event conditions, though it is likely nontrivial to properly calibrate their seasonality with respect to the extreme tail of precipitation-demanding long training datasets.

Minor comments 1: Page 2, line 4: This sentence is confusing. I am assuming you mean '... statistical analysis of observed streamflow, design storms !and! continuous simulation !or! other so called "derived" or "process based" methods'.

Correct. We have modified this sentence to:

Most existing FFA methods belong to one of three approaches: statistical analysis of streamflow observations, design storms, and continuous simulation or other so-called "derived" or "process-based" methods.

Minor comments 2: Page 4, lines 15-17: This sentence seems not to be complete.

We apologize for this. We have revised this sentence to:

Wright et al. (2014a) discusses additional design storm shortcomings including time of concentration concepts, in greater detail, while also pointing out that design storm approaches (like other hydrological model-based FFA) can incorporate future projections in land use and rainfall more explicitly than can statistical discharge-based methods.

Minor comments 3: Page 10, steps 3 and 4: I would stress that the 30 storms per year are randomly transposed over the domain, only sometimes hitting the catchment and sometimes not. They are not all transposed on the catchment, which would lead to an overestimation of the flood frequency. The reader not familiar with your method might misunderstand that.

The reviewer is correct. We have added this sentence to P10, line 22.

It must be noted that some of the  $k$  transposed storms may not “hit” Turkey River watershed, and thus their calculated watershed rainfall are zero.

Minor comments 4: Page 11, lines 8-9: The selection of the largest event per year for FFA might also be misunderstood. Here, it also needs to be considered that many of the 30 events do not produce any flood if they do not hit the catchment (see comment 3).

We hope the response to previous comments also addresses this one.

Minor comments 5: Page 14: line 2: Should it not be “... but overestimates for  $p_e < 0.3$  ...”

We assume the reviewer mean Page 16, line 2. We have revised this sentence to:

The Stage IV-based flood frequency curve agrees reasonably well with the discharge-based FFA for  $p_e > 0.3$  (left panel of Fig. 6), but yields higher estimates for rarer events.

Minor comments 6: Fig. 5: Why did you select the period 1990 – 2016 and not 1980 or 1970 as starting year? This needs to be justified.

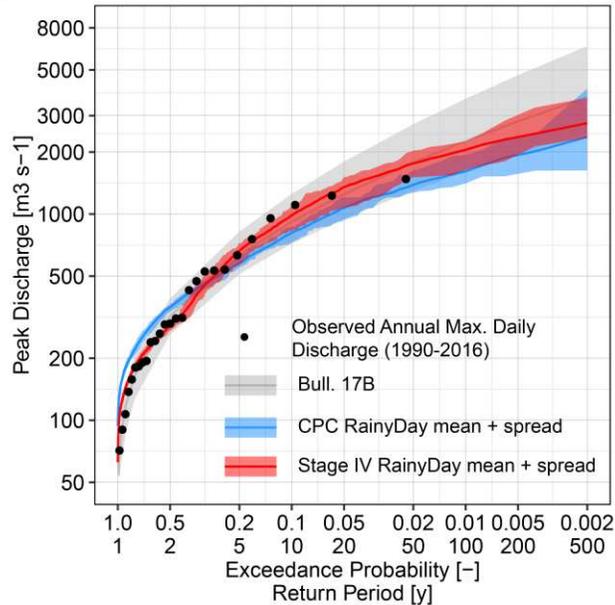
We have not performed any statistical test (e.g. Pettitt test) to determine this change point. However, an “eyeball test” of annual daily discharge maxima (Fig. 1a) from the original manuscript indicates the apparent elevated flood activity during 1990-2016 period. Our arguments do not hinge on a precise determination of when floods in Turkey River began to change, which in any event has likely been a gradual change.

Minor comments 7: Fig. 5: I would also add a statistical analysis (Bull 17.b) for the contemporary period (1990-2016) for comparison.

We have added a supplementary plot showing the CPC, Stage IV and Bull.17B based FFA for the modern time (2002-2016), as other reviewers have suggested.

Supplementary Fig. 1 shows that process-based FFA using CPC precipitation from 2002-2016 closely resembles the Stage IV-based FFA, suggesting that rainfall differences, rather than model structures, are the primary drivers of the differences in this figure. It also shows two features that result using CPC data. First, the extreme tail is underestimated, relative to the Stage IV-based simulations and the statistical approach. CPC is known to contain errors in the extreme tail, due to gage undercatch, insufficient gage density to properly sample convective rain cells, and spatial

averaging of such cells over large areas, which effectively reduces peak rainfall depths. Second, CPC overestimates the magnitude of more frequent events. This is likely the result of its coarse spatial resolution, which will “smear” rainfall over larger areas (i.e. entire ~600 km<sup>2</sup> grid cells) when it should be more localized. This would serve to increase the likelihood of rainfall over the watershed, albeit at relatively lower depths/intensities. Thus, if one is to restrict the time period of the rainfall data to recent years (for example, the 2002-2016 time period for which Stage IV is available), then Stage IV would likely be better.



**Supplementary Figure 1.** Three peak discharge analyses for Turkey River at Garber, IA: RainyDay with Stage IV (2002-2016) and CPC-(2002-2016) rainfall and USGS frequency analyses (1990-2016) using Bulletin 17B methods. Shaded areas denote the ensemble spread (RainyDay-based results) and the 90% confidence intervals (Bulletin 17B-based analysis), respectively. All observed annual daily streamflow maxima from 1990 to 2016 are shown in black dots.

Minor comments 8: Fig. 6: There is no description neither in legend nor in figure caption about the source of the two figures. I assume they stem from different precipitation data sets.

We have updated this figure.

# Process-Based Flood Frequency Analysis in an Agricultural Watershed Exhibiting Nonstationary Flood Seasonality

Guo Yu<sup>1</sup>, Daniel B. Wright<sup>1</sup>, Zhihua Zhu<sup>2</sup>, Cassia Smith<sup>3</sup>, Kathleen D. Holman<sup>4</sup>

<sup>1</sup>Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, 53706, USA

5 <sup>2</sup>Department of Water Resources and Environment, Sun Yat-sen University, Guangzhou, 510275, China

<sup>3</sup>Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, 15213, USA

<sup>4</sup>[US Department of the Interior, Bureau of Reclamation, Denver, 80225, USA](#)

~~<sup>4</sup>[Research and Development Office, Bureau of Reclamation, Denver, 80225, USA](#)~~

Correspondence to: Zhihua Zhu (zzhu264@wisc.edu)

10 **Abstract.** Floods are the product of complex interactions ~~of~~among processes including ~~rainfall~~precipitation, soil moisture, and watershed morphology. Conventional flood frequency analysis (FFA) methods such as design storms and discharge-based statistical methods offer few insights into these process interactions and how they “shape” the probability distributions of floods. Understanding and projecting flood frequency in conditions of nonstationary hydroclimate and land use requires deeper understanding of these processes, some or all of which may be changing in ways that will be undersampled in observational records. This study presents an alternative “process-based” FFA approach that uses stochastic storm transposition to generate large numbers of realistic rainstorm “scenarios” based on relatively short rainfall remote sensing records. Long-term continuous hydrologic model simulations are used to derive seasonally varying distributions of watershed antecedent conditions. We couple rainstorm scenarios with seasonally appropriate antecedent conditions to simulate flood frequency. The methodology is applied ~~in~~to the 4002 km<sup>2</sup> Turkey River watershed in the Midwestern United States, ~~a watershed that~~which is undergoing significant climatic and hydrologic change. We show that using only 15 years of rainfall records, our methodology can produce ~~more~~more-accurate estimates of “present-day” flood frequency ~~than is possible using longer discharge or rainfall records~~. We found that shifts in the seasonality of soil moisture ~~conditions~~, snow, and extreme rainfall in Turkey River exert important controls on flood frequency. We also demonstrate that process-based techniques may be prone to errors due to inadequate representation of specific seasonal processes within hydrologic models. ~~If s~~Such mistakes are ~~avoidable~~avoided, however, ~~and our process-based approaches can~~ may provide a ~~clearer~~useful pathway toward understanding current and future flood frequency in nonstationary conditions ~~compared with more conventional methods~~ and thus be valuable for supplementing existing FFA practices.

## 1 Introduction

30 Riverine floods, among the most common natural disasters worldwide, are the product of complex interactions between heavy rainfall, watershed and river channel morphology, and antecedent (i.e. initial) conditions including soil moisture and snowpack.

Their impacts are projected to increase in the future due to hydrometeorological factors (e.g. Hyndman, 2014) and increased human development in flood prone areas (e.g. Ntelekos et al., 2010; Ceola et al., 2014; Prosdocimi et al., 2015). Estimating the relationships between flood likelihood and severity is central to flood risk management and infrastructure design; these relationships are typically represented by flood frequency distributions (or curves), while the broad family of procedures used to derive them is termed flood frequency analysis (FFA). Most existing FFA methods belong to one of three approaches: statistical analysis of ~~observed~~-streamflow observations, design storms, ~~or-and~~ continuous simulation ~~and-or~~ other so-called “derived” or “process-based” methods. Each has strengths and shortcomings, which are briefly summarized in Sect. 2 (see Wright et al., 2014a for a distinct summary).

FFA is challenging even in stationary (i.e. unchanging) watershed and hydroclimatic conditions due to the scarcity of observations of large floods and the associated factors that generate them (Stedinger and Griffis, 2011). The role of soil moisture in flood frequency, for example, is very important (Berghuijs et al., 2016), but poorly understood due to a lack of long-term observations. Furthermore, the individual and joint flood causative factors will evolve as a watershed undergoes changes in land use or hydroclimate (Machado et al., 2015). Leading causes of change (i.e. nonstationarity) include human intervention through land use change or reservoir construction (Konrad and Booth, 2002; Schilling and Libra, 2003; Villarini et al., 2009), natural climate variability (Enfield et al., 2001; Jain and Lall, 2000) and anthropogenic climate change driven by increased-increasing greenhouse gases-gas concentrations (Milly et al., 2008; Hirsch and Ryberg, 2012). Combinations of these will lead to nonstationary flood frequency, a challenge for which the bulk of existing FFA methods are ill-suited (El Adlouni et al., 2007; Gilroy and McCuen, 2012).

In this study, we present an alternative FFA methodology that aims to “construct” the flood frequency curve through a combination of observations, stochastic methods, and hydrological modeling that generates and combines the causative factors (i.e. processes) such as rainfall and soil moisture that produce floods. This concept is not new, and has traditionally be called “derived FFA” (e.g. Eagleson, 1972; Franchini et al., 2005; Haberlandt, 2008), though we prefer the more descriptive term “process-based FFA” (after Sivapalan and Samuel, 2009; see Clark et al., 2015a, 2015b and Lamb et al., 2016; who discuss somewhat similar techniques). Sivapalan and Samuel (2009) argue in favor of process-based approaches in the face of nonstationary conditions, though they do not actually lay out a specific FFA procedure.

We ~~apply our present such a~~ process-based procedure, and apply it methodology to an agricultural watershed in the Midwestern United States that is undergoing substantial seasonal hydroclimatic and hydrologic changes that have led to nonstationary flood frequency. We ~~will show that process-based FFA this procedure may hold better prospects than other methods in this watershed and more broadly, and~~ is useful for deciphering the underlying physical processes that drive flooding, as well as drivers of flood frequency their changes in this watershed. (The reader is directed to Sivapalan and Samuel (2009) for a strong argument in favor of process based approaches in the face of nonstationary conditions, though they do not actually lay out a specific FFA procedure.) Our methodology underscores the importance of seasonality in the joint contributions of rainfall ~~and~~ soil moisture, and snow to flood frequency. To our knowledge, this study is the first to explore the role that seasonal changes in hydroclimatic and hydrologic processes play in nonstationary flood frequency, though other studies have explored the

~~importance of such processes in flood occurrence more generally (e.g. Berghuijs et al., 2016). We also argue that any process-based FFA approach will require careful consideration of seasonality.~~

The structure of the paper is as follows: Section 2 briefly reviews ~~the~~ three ~~broad types of~~ forementioned FFA approaches. Section 3 introduces the study region, watershed, and hydrometeorological data. Section 4 outlines the process-based FFA methodology used in this study, including the hydrological model, the stochastic storm transposition (SST) procedure used to derive the synthetic rainfall scenarios, and elements of both continuous and event-based rainfall-runoff simulation. The nonstationary hydroclimate of the study watershed and trends in relevant hydrometeorological variables are analyzed in Sect. 5.1. Model validation is presented in Sect. 5.1. Process-based FFA results are presented and compared with “conventional” statistical estimates in Sect. 5.23. Simulated flood seasonality is explored in Sect. 5.34. The relationships between rainfall and simulated peak discharge quantiles ~~are~~ examined in Sect. 5.45. Section 6 includes a summary and concluding remarks.

## 2 Review of FFA Approaches

### 2.1 Discharge-based Statistical Approaches

Statistical FFA approaches involve fitting a statistical distribution to extreme discharge observations and extrapolating this distribution to estimate quantiles such as the 100-year or 500-year discharge. While these approaches utilize direct observations of flooding (e.g. peak discharge or volume), long streamflow records at or near the given river cross section are needed for reliable quantile estimates. Such records are lacking in many locations, even in developed countries. Statistical approaches are limited by the available observations; thus, the estimation distribution may not represent the “true” (unknown) distribution of possible outcomes (Linsley, 1986; Klemeš, 1986, 2000a, 2000b). In principle, regionalized Regional FFA methods are able to improve quantile estimates both at gaged and ungaged locations (Dawdy et al., 2012); ~~however,~~ they make assumptions, however, regarding the transferability of regional information to specific locations and ~~can,~~ in doing so, may neglect key geophysical processes that dominate the spatiotemporal variability of floods (Ayalew and Krajewski, 2017).

Though streamflow observations are the result of a range of complex factors including rainfall, soil moisture, and channel routing, without concurrent observations of these “upstream” variables, neither streamflow observations nor distributions fitted to them provide much insight into flood causes. Long-term records of such variables, particularly soil moisture, are virtually nonexistent. There have been numerous examples within the FFA literature pointing to situations in which discharge-based analyses are-can be inferior to ~~others-those~~ based on hydrologic modeling, including cases of basin storage “discontinuities” (Rogger et al., 2012), reservoirs (Ayalew et al., 2013), and land use change (Cunha et al., 2011).

Finally, most statistical FFA methods assume that the magnitude of extreme flood events and quantiles are stationary. This assumption conflicts with numerous examples in which hydrological records exhibit various types of nonstationarity (e.g. Potter, 1976; Villarini et al., 2009; Douglas et al., 2000; Franks and Kuczera, 2002)(~~e.g. Salas and Obeysekera, 2014; Potter, 1976; Villarini et al., 2009; Douglas et al., 2000; Franks and Kuczera, 2002~~). Though nonstationary statistical FFA techniques do exist (e.g. Cheng et al., 2014; Gilleland and Katz, 2016; Serago and Vogel, 2018), they face severe limitations extrapolating

to future conditions (Luke et al., 2017; Sivapalan and Samuel, 2009; Stedinger and Griffis, 2011) since they rarely consider the fundamental physical causes of change.

## 2.2 Design Storm Approaches

5 Design storm (DS) approaches use idealized rainfall scenarios of a given return period as inputs to a ~~calibrated~~-hydrological model to simulate flood peaks. DS is widely used in practice due to its simplicity (Cudworth, 1989; Kjeldsen, 2007; Ball et al., 2016). To some extent, the flood-producing physical processes are captured via the hydrological model, which also provides a complete simulated flood hydrograph, as opposed to only the peak discharge or volume provided by statistical approaches. However, DS approaches rely on at least three major assumptions: (1) point-based rainfall intensity-duration-frequency (IDF) estimates (which are subject to some of the same aforementioned statistical and data availability issues as  
10 flood discharges) can be converted into hyetographs using dimensionless temporal rainfall distributions and into basin-averaged estimates using area reduction factors (e.g. Svensson and Jones, 2010); (2) IDF estimates, based on annual rainfall maxima, produce flood peaks which are quantiles of the distributions of flood annual maxima-~~distribution~~; and (3) there is a 1:1 equivalence between rainfall and simulated discharge quantiles (i.e. return periods or recurrence intervals), for example, a 100-year idealized rainfall event will produce a reasonable estimate of the 100-year peak discharge. The last of these  
15 assumptions discounts the possibility that watershed initial conditions such as soil moisture and snowpack can modulate the transformation of rainfall quantiles into discharge quantiles.

These assumptions are not without their shortcomings. Wright et al. (2014b), for example, showed significant disparities between observed point and basin-averaged rainfall extremes that cannot be captured using conventional ARF concepts. Using design storm in conjunction with a derived distribution approach, Viglione and Blöschl (2009) and Viglione et al. (2009),  
20 ~~meanwhile~~, demonstrated that the ratio of rainfall return period to flood peak return period is controlled by storm duration, a runoff coefficient (which is related to antecedent conditions), and a runoff threshold effect-~~using design storm approach in conjunction with derived distribution approach~~. ~~These initial~~Antecedent conditions can vary substantially by season, meaning that high soil moisture may only ~~very~~-infrequently coincide with extreme rainfall. Wright et al. (2014a) discusses additional design storm shortcomings in greater detail, including time of concentration concepts, while also pointing out that design storm  
25 approaches (like other hydrologic model-based FFA) can incorporate future projections in land use and rainfall more explicitly than can statistical discharge-based methods.

## 2.3 Continuous Simulation and Process-Based FFA Approaches

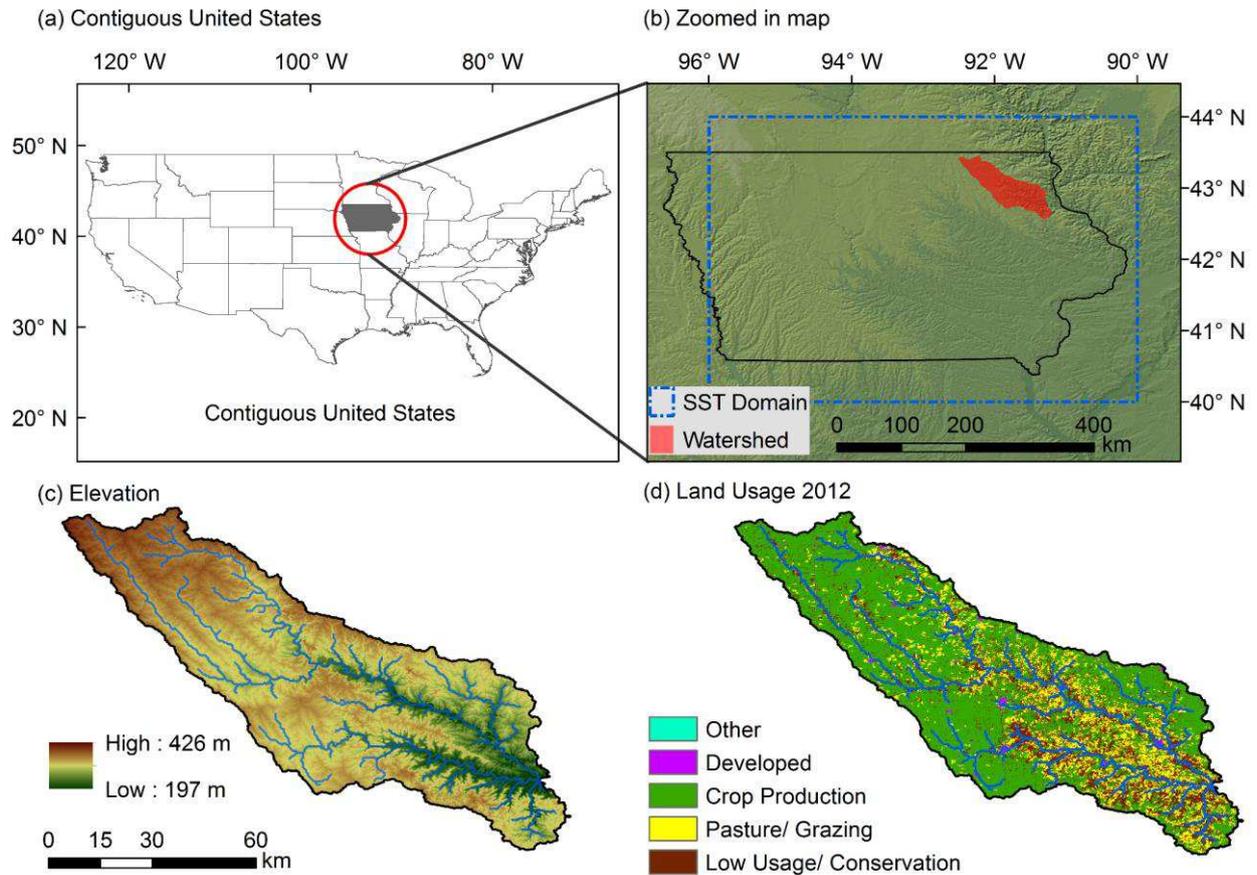
Continuous simulation (CS) and process-based approaches to FFA leverage the potential benefits of hydrological models while minimizing the simplifying assumptions of DS methods. ~~These~~-CS approaches typically use long series of historical or  
30 stochastically generated rainfall, temperature, and occasionally other meteorological variables as ~~hydrological~~-model inputs, to simulate long discharge time series. Peak flows can be extracted from these series and the flood frequency distribution can be obtained. Thus, event rainfall return period and duration and antecedent conditions do not need to be specified and the

equality between ~~the~~ rainfall and discharge return period is not assumed (Calver et al., 1999, 2009). In addition, projections of future flood frequency can be developed by incorporating general circulation model (GCM) rainfall and temperature projections into the input meteorological series (Gilroy and McCuen, 2012; Rashid et al., 2017). On the other hand, CS approaches are limited by the general lack of reliable long-term time series of extreme rainfall and other meteorological data (Blazkova and Beven, 1997, 2002, 2009) and, in the case of sophisticated distributed approaches, by potentially high computational demands (Li et al., 2014; Peleg et al., 2017). Stochastic rainfall generation techniques typically struggle to produce with producing the extremes that are critical for flooding (e.g. Cameron et al., 2000; Furrer and Katz, 2008), and training such models for locations with rainfall nonstationarities and strong seasonal variations is nontrivial. Camici et al., (2011) and Li et al. (2014) present ~~FFA~~-process-based FFA approaches that couple long CS simulation results with event-based simulations.

One argument in favor of CS and process-based approaches is that the complex joint relationships between flood drivers such as rainfall and soil moisture are resolved within the modeling framework and thus do not rely on users' assumptions. We demonstrate that caution is needed in the representation of seasonality; to briefly summarize, it is critical that both seasonality in input variables as well as seasonally varying processes within the model be "correct." Without verifying this, process-based approaches may produce ~~incorrect results, or, as shown in Sect. 5,~~ seemingly correct results as a result of incorrect methods.

### 3 Study Region and Data

The study watershed of Turkey River (~~Fig. 1~~) is situated in northeastern Iowa (~~Fig. 1a, 1b~~), ~~and the~~ The portion upstream of the US Geological Survey (USGS) stream gage at Garber (~~ID code~~ gage number 05412500) has a drainage area of 4002 km<sup>2</sup>, ~~with E~~ elevations ranging from approximately 426 m above sea level (masl) in the west to 197 masl at the stream gage (~~Fig. 1c~~) site. ~~The s~~ Streams at in the upper part of the catchment have relatively ~~low~~ mild slopes, while the channels and hillslopes in the lower part are steeper. Soils ~~in Turkey River~~ are mainly loams and silts (IFC, 2014). According to USGS 2012 National Land Cover Dataset (NLCD), ~~the~~ the Turkey River watershed is predominantly agricultural, with less than 2% urban land cover (~~Fig. 1d~~). Comparisons of NLCD from 1992, 2001, 2006, and 2012 indicates that land uses have not evolved significantly over time (results not shown), though the hydrologic impacts of subsurface tile drainage, which has become ubiquitous throughout the region, are poorly understood and could exert meaningful influence on flooding (see, e.g. Schilling et al., 2014).

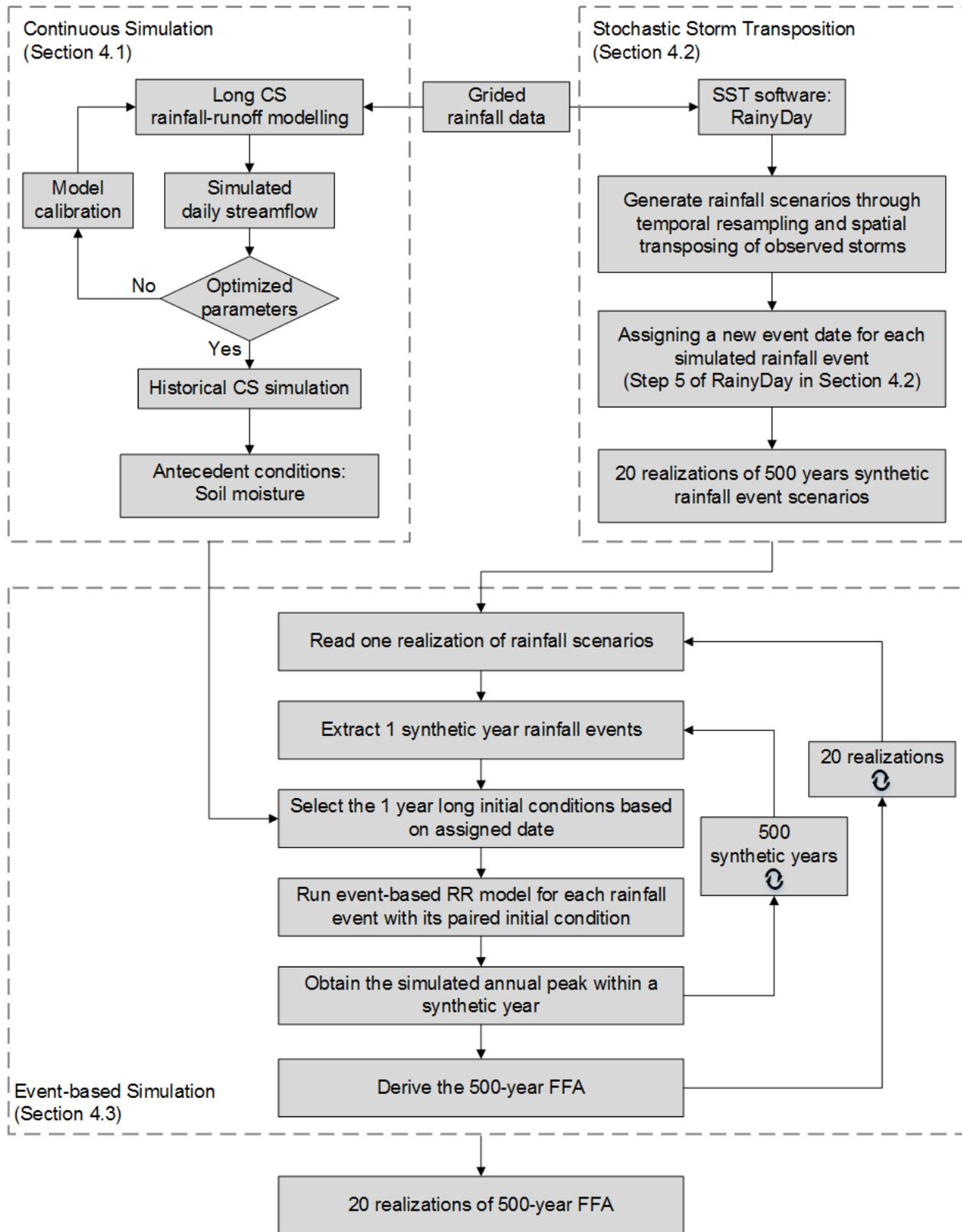


**Figure 1.** Study region. (a) Contiguous-Conterminous United States; with the state of Iowa is highlighted in grey. (b) The-Z-zoomed-in map shows showing Iowa (black outline) and the Turkey River watershed (red) and the extent of the stochastic storm transposition region (blue dash line). (c, d) Turkey River watershed showing land surface elevation (based on the USGS National Elevation Dataset) and land use (based on the USGS 2012 NLCD), respectively.

We use daily discharge observations for 84 years (1933-2016) from the USGS streamgauge at Garber, Iowa (USGS gage identifier 05412500) to understand the hydroclimatology of flooding and to validate our FFA results. Daily discharge observations for 69 years (1948-2016), in conjunction with Global Historical Climate Network (GHCN) daily temperature and snow data are used to configure, calibrate, and validate the hydrological model, as described in Sect. 4.1. The CPC US Unified (CPC-Unified; Chen et al., 2008) and Stage IV (Lin and Mitchell, 2005) precipitation data, available through the National Oceanic and Atmospheric Administration, are used for rainfall analyses. CPC-Unified provides daily, 0.25° rainfall estimates interpolated from rain gage observations, while Stage IV provides hourly, approximately 4 km estimates by merging data from rain gages and the National Weather Service Next-Generation Radar network (NEXRAD; Crum and Alberty, 1993). In this study, analyses based on Stage IV use data from 2002-2016, while long-term analyses based on CPC-Unified use data from 1948-2016.

#### 4 Methodology

The FFA approach presented in this study combines continuous simulation (CS), stochastic storm transposition (SST) using the RainyDay software, and event-based simulation. CS provides large samples of seasonally varying antecedent conditions, ~~namely including~~ soil moisture and snowpack. SST produces large numbers of synthetic rainfall scenarios, ~~including realistic estimates of rainfall space-time structure.~~ Together, these drive event-based simulations to generate the synthetic flood peaks that are used to derive flood frequency distributions. The approach is illustrated schematically in Fig. 2 and summarized in the following subsections.



**Figure 2.** Flow chart showing the process-based FFA approach. Dotted outlines delineate components associated with subsections 4.1, 4.2 and 4.3.

#### 4.1 Hydrological Model, Calibration, and Continuous Simulation

We used the lumped Hydrologiska Byråns Vattenavdelning (HBV) model (Bergström, 1992, 1995; Lindström et al., 1997).

5 HBV has been widely used to study hydrological response in United States (Vis et al., 2015; Niemeyer et al., 2017) and other regions of the world (Harlin and Kung, 1992; Osuch et al., 2015; Seibert, 2003; Chen et al., 2012). The “HBV-Light” version (–henceforth referred to as HBV; Seibert and Vis, 2012) version is used in this study, and consists of four main routines: snowpack, soil moisture, catchment response, and runoff routing routine. The model HBV simulates daily discharges based on time series of precipitation and air temperature, as well as estimates of long-term daily potential evapotranspiration. A list of model parameters is shown in Table 1.

10 The process-based FFA methodology employed in this study could be used coupled with any other hydrological models. Utilizing a distributed hydrological model would allow for more realistic representation of important characteristics like changing land use, rainfall spatiotemporal structure, and flood wave attenuation in river channels. Other models and could operate at higher (i.e. subdaily) temporal resolution in terms of inputs, model time steps, and outputs. We selected HBV at the daily time step due to its simplicity, computational speed, and its ability to represent conceptually multiple watershed hydrological processes.

15 We calibrated separate HBV models using both CPC and Stage IV rainfall. Most parameter values were the same for CPC- and Stage IV-based models except for three snow routine parameters (TT, CFMAX, SFCF) and three recession coefficients (K0, K1, K2), allowing for the variability of model parameters for different climate conditions. For each model setup, we first calibrated the model with snowpack routine “turned off” (by setting TT parameter to a very low value) to obtain parameters that can simulate summer floods adequately. Then, keeping these optimized non-snow routine parameters unchanged, we calibrated the snow routine parameters.

20 To determine the optimized model parameter sets in each procedures, we followed the Genetic Algorithm and Powell (GAP) optimization method as presented by Seibert (2000), which is briefly summarized here. First, 5000 parameter sets are randomly generated from a uniform distribution of the values of each parameter (Table 1), which were then applied to the HBV model in order to maximize Kling Gupta Efficiency (Gupta et al., 2009) of simulated daily discharge. After the GAP has finished, the optimized parameter set were fine-tuned using Powell’s quadratic convergent method (Press, 1996) with 1000 additional runs. Lastly, the optimized parameter set was manually adjusted to improve the fits between observed and simulated annual peak flow (see Lamb, 1999). More elaborate calibration and uncertainty estimation procedures such as Generalized Likelihood Uncertainty Estimation (GLUE; Beven and Binley, 1992; Beven, 1993; Beven and Binley, 2014) could be used, but are outside the scope of our study.

30 The two different HBV models were then used to perform CS with historical CPC and Stage IV rainfall and temperature data to derive long-term simulated soil moisture and snowpack values, which are usually difficult to obtain via measurement. We

“pair” samples of these initial conditions with synthetic rainfall events to simulate hypothetical floods, as described in Sect. 4.2 and Sect. 4.3.

We calibrated the model automatically via 5000 continuous simulations from 1948–2016 by maximizing Kling and Gupta Efficiency (Gupta et al., 2009) using the Genetic Algorithm and Powell optimization method (Seibert, 2000). After the genetic algorithm has finished, 1000 additional runs are performed for fine tuning using Powell’s quadratic convergent method (Press, 1996). Lastly, the optimized parameter set is manually adjusted to improve the fits between observed and simulated annual peak flow (see Lamb, 1999). More elaborate calibration and uncertainty estimation procedures such as Generalized Likelihood Uncertainty Estimation (GLUE; Beven and Binley, 1992; Beven, 1993; Beven and Binley, 2014) could be used, but are outside the scope of our study. After calibration, HBV was used to perform CS with historical CPC and Stage IV rainfall and temperature data to derive long term simulated soil moisture and snowpack values, which are usually difficult to obtain via measurement. We “pair” samples of these initial conditions with synthetic rainfall events, as described in Sect. 4.2 and Sect. 4.3.

**Table 1.** Overview of HBV model parameters and upper and lower parameter limits used for calibration.

<u>Parameter</u>	<u>Description</u>	<u>Units</u>	<u>Min value</u>	<u>Max value</u>
<u>Snow Routine</u>				
<u>TT</u>	<u>Threshold temperature for liquid and solid precipitation</u>	°C	<u>-3</u>	<u>3</u>
<u>CFMAX</u>	<u>Degree-day factor</u>	mm d <sup>-1</sup> °C <sup>-1</sup>	<u>0.5</u>	<u>4</u>
<u>SFCF</u>	<u>Snowfall correction factor</u>	=	<u>0.5</u>	<u>1.2</u>
<u>CFR</u>	<u>Refreezing coefficient</u>	=	<u>0.01</u>	<u>0.1</u>
<u>CWH</u>	<u>Water holding capacity of the snow storage</u>	=	<u>0.1</u>	<u>0.3</u>
<u>Soil Moisture Routine</u>				
<u>FC</u>	<u>Maximum soil moisture storage (field capacity)</u>	mm	<u>100</u>	<u>550</u>
<u>LP</u>	<u>Relative soil water storage below which AET is reduced linearly</u>	=	<u>0.3</u>	<u>1</u>
<u>BETA</u>	<u>Exponential factor for runoff generation</u>	=	<u>1</u>	<u>5</u>
<u>Response Routine</u>				
<u>PERC</u>	<u>Maximum percolation from upper to lower groundwater box</u>	mm d <sup>-1</sup>	<u>0</u>	<u>10</u>
<u>UZL</u>	<u>Threshold of upper groundwater box</u>	mm	<u>0</u>	<u>50</u>
<u>K0</u>	<u>Recession coefficient 0</u>	d <sup>-1</sup>	<u>0.5</u>	<u>0.9</u>
<u>K1</u>	<u>Recession coefficient 1</u>	d <sup>-1</sup>	<u>0.15</u>	<u>0.5</u>
<u>K2</u>	<u>Recession coefficient 2</u>	d <sup>-1</sup>	<u>0.01</u>	<u>0.15</u>
<u>Routing Routine</u>				
<u>MAXBAS</u>	<u>Length of triangular weighting function</u>	d	<u>1</u>	<u>2.5</u>

## 4.2 Stochastic Storm Transposition

Stochastic storm transposition (SST) is a bootstrap method to generate realistic probabilistic rainfall scenarios through temporal resampling and spatial transposing of observed storms from the surrounding region. SST ~~is a bootstrap method that aims to~~ effectively “lengthens” the rainfall record ~~by performing via~~ “space-for-time substitution,” ~~within a rigorous probabilistic framework.~~ Unlike rainfall IDF curves, SST can preserve observed rainfall space-time structure, and, unlike design storm methods, obviates the need to equate rainfall duration to catchment response time (Wright et al., 2013, 2014a, 2014b). Alexander (1963), Foufoula-Georgiou (1989), and Fontaine and Potter (1989) provide general descriptions of SST. Wilson and Foufoula-Georgiou (1990) apply the method for regional rainfall frequency analysis while Gupta (1972), Franchini et al. (1996), England et al. (2014) and Nathan et al. (2016) use it for FFA.

5 Wright et al. (2013) used SST with a 10-year high-resolution radar rainfall dataset to estimate spatial IDF relationships. Wright et al. (2014a) used this approach with a physics-based distributed hydrologic model for FFA in a heavily urbanized watershed, demonstrating its usefulness in evaluating multi-scale flood response.

RainyDay is an open-source, Python-based SST software that couples SST methods with rainfall remote sensing data. A more detailed description can be found in Wright et al. (2017); not all of its features are used in this study. The following steps describe how RainyDay is ~~modified and~~ used ~~in this study here~~:

1. We define a 6-degree (longitude) by 4-degree (latitude) geographic transposition domain (40° to 44° N, 90° to 96° W; blue dash line of Fig. 1 inset) which encompasses the Turkey River watershed. This same domain was used in Wright et al (2017) and, importantly for the SST approach, extreme rainfall properties are roughly homogeneous within it.
2. The RainyDay software creates a “storm catalog” from 15 years of Stage IV (69 years of CPC) rainfall-precipitation data that consists of the 450 (2070) most intense rainfall-precipitation event within the transposition domain. These intense storms are in terms of 96-hour rainfall accumulation and have the same size, shape, and orientation of the Turkey River watershed, which is oriented roughly northwest-southeast and with an area of 4002 km<sup>2</sup>. In order to avoid overlapping storms, these selected events must be separated by at least 24 hours. These storms have a maximum duration of 96 hours and must be separated by at least 24 hours. Storms that exhibit “radar artifacts” such as major bright band contamination or beam blockage are excluded from subsequent steps.
3. The RainyDay software generates a Poisson-distributed integer  $k$  that represents a “number of storms per year.” The rate parameter  $\lambda$  of this Poisson distribution is calculated by dividing the total number of rainfall events in the storm catalog by the number of years in the historical rainfall record (*e. g. i. e.*  $\lambda = 450/15 = 30.0$  storms per year).
4. RainyDay randomly selects  $k$  storms from the storm catalog and transposes the associated rainfall fields within the transposition domain by an east-west distance  $\Delta x$  and a north-south distance  $\Delta y$ , where  $\Delta x$  and  $\Delta y$  are ~~drawn~~ drawn from a ~~two-dimensional~~ dimensional Gaussian kernel density estimate based on the locations of the original storms in the storm catalog. For each of the  $k$  transposed storms, the time series of rainfall over the Turkey River watershed is computed. It must be noted that some of the  $k$  transposed storms may not “hit” Turkey River watershed, and thus their calculated

watershed rainfall are zero. Steps 3 and 4 can be understood as temporal resampling of observed rainfall events to “synthesize” a hypothetical year of rainfall events over the transposition domain and, by extension, over the watershed. Although the rainfall events for the “synthetic” year do not form a continuous series, the dates associated with each observed storm event are recorded, thus facilitating seasonally-consistent flood simulations.

- 5 5. All  $k$  events within a synthetic year are assigned a new, randomly selected year from 1948-2016 (2002-2016) for CPC (Stage IV) rainfall data, which used to select antecedent conditions. This ensures that the  $k$  rainfall events are all “embedded” within a single realistic annual representation of watershed conditions. This ensures that “wet” and “dry” years in terms of snowpack and soil moisture can potentially produce wet or dry years of flood response. Antecedent conditions are randomly selected from  $\pm 7$  within seven days of the updated storm date to ensure realistic seasonality of storms and watershed conditions. A storm that occurred on 15 July 15, 2016, for example, could be paired with initial conditions selected from a day date ranging between 8-22 July 8-22 from a randomly selected year, while the remaining  $k-1$  events would be paired with seasonally appropriate initial conditions from that the same selected year.
- 10
6. RainyDay repeats Steps 3-5 500 times to create one realization of 500 synthetic years of rainfall events for Turkey River. Twenty such realizations of 500 synthetic years each are generated. Unlike in the existing version of RainyDay, all rainfall events within a synthetic year are retained for subsequent event-based flood simulations, since the modulating effects of antecedent conditions mean that the largest rainfall event in a given year does not necessarily produce that year’s largest flood peak (this possibility is explored in Sect. 5.4).
- 15

### 4.3 Event-Based Flood Simulation

Using the seasonally-consistent “paired” SST-based rainfall events and watershed initial conditions derived from CS (Sect. 20 4.12) and SST-based rainfall events (Sect. 4.2), HBV simulates the “event peak” (the maxima daily discharge). The largest “event peak” peak among the  $k$  events within that comprise a synthetic year represents the simulated annual maximum daily streamflow. This process is repeated for all 500 synthetic years within each realization, resulting in 500 annual maximum streamflow values, which are then ranked in descending magnitude. The annual exceedance probability  $p_e$  (i.e. the probability in a given year that an event of equal or greater intensity-magnitude will occur) of each maximum streamflow are calculated 25 by dividing its rank by 500 (the total number of simulated annual maximum daily streamflow). The 20-twenty realizations provide estimates of variability for each flood quantile.

## 5. Results

### 5.1 Hydroclimatology and Nonstationarity

Four distinct time periods (Fig. 3a) are used-considered for analyzing the changing hydroclimatology in Turkey River: the 30 USGS daily mean streamflow period of record (1933-2016), a more recent period of apparent elevated flood activity (1990-

2016), the period of the Stage IV rainfall record (2002-2016), and the period of the CPC rainfall record (1948-2016). Results here and in subsequent subsections “align” with one or more of these time periods.

The hydroclimate of Turkey River is changing, as shown ~~by using~~ the Mann-Kendall (MK) test for monotonic trends (Mann, 1945) ~~and the Thiel-Sen estimator (Sen, 1968), a nonparametric methods used~~ to determine trend direction and significance magnitude (i.e. slope), respectively (Table 2). Since 1948, annual precipitation and discharge show significant increases ( $p < 0.05$ ) and their variability has also increased ~~(Table 1)~~, while annual maximum daily discharge has decreased, though not significantly. It is important to note, however, that there are two counteracting seasonal trends (see also Fig. 3a): annual daily discharge maxima have decreased significantly in March-April, ~~but has have decreased-increased somewhat in significantly while~~ May-September ~~has increased somewhat~~. Thus, the lack of statistically significant change in flood magnitude annual maximum daily discharge in Turkey River ~~at the annual scale~~ masks changes in the seasonality of flooding.

**Table 2.** Mann-Kendall trend ~~(two-sided)~~ test (two sided) for hydrological variables.  $p$ -values are given in parentheses; bolded values are significant at the 5% level. The Analyses of trends in variances refers to examine changes in the absolute values of residuals associated obtained from a linear regression using with the Thiel-Sen estimator (Sen, 1968).

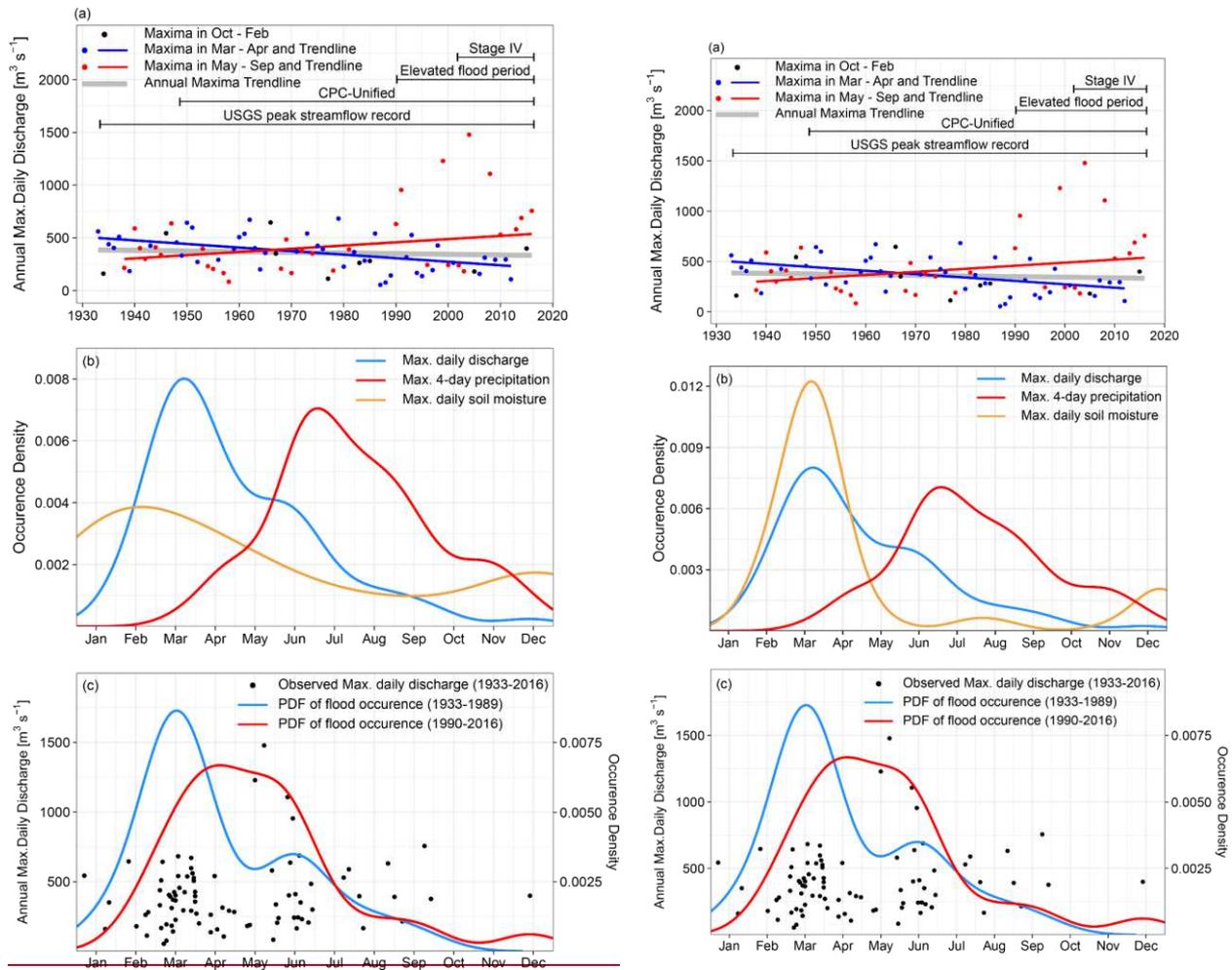
Data	Time Range	Trend
<b>Annual Discharge</b>	<b>1933-2016</b>	<b>↑ (0.001)</b>
Annual Max. Daily Discharge	1933-2016	↓ (0.447)
Variance of Annual Max. Daily Discharge	1933-2016	↑ (0.056)
<b>Annual Max. Daily Discharge in March-April</b>	<b>1933-2016</b>	<b>↓ (0.002)</b>
Annual Max. Daily Discharge in May-September	1933-2016	↑ (0.089)
<b>Annual Precipitation</b>	<b>1948-2016</b>	<b>↑ (0.003)</b>
Annual Max. Daily Precipitation	1948-2016	↑ (0.362)
Annual Max. 4-day Precipitation	1948-2016	↑ (0.419)
Annual Mean Temperature	1948-2016	↓ (0.462)
March-May Mean Temperature	1948-2016	↑ (0.443)

We examine this flood seasonality, both in observations and in our continuous HBV simulations (Fig. 3b). The seasonal distribution of flood occurrence for 1948-2016 shows a March-April maximum, with elevated flood activity continuing through May and June. This is distinct from, though overlaps somewhat with the seasonality of both the ~~4~~four-day annual maxima of rainfall, which occur most frequently in the June-September period, and simulated daily annual maxima soil moisture, which only tend to occur in ~~January~~March-April. These results highlight that flood activity is the product of seasonal variations in

both soil moisture and rainfall. ~~4(Four-day rainfall shown in Fig. 3b since it~~ is used in SST; seasonality in ~~1one-day rainfall is~~ ~~very-similar;~~ (results not shown).

5 The March-April peak of flood occurrence corresponds with relatively high soil moisture associated with snowmelt, rain on or frozen soil, ~~or-and~~ frequent spring rains. The secondary peak of flood occurrence in May-June is associated with larger flood magnitudes (including the ~~largest-flood event of record, in 2004~~) due to ~~recent-severe-organized~~ thunderstorms ~~systems~~. ~~For instance, w~~Widespread flooding in Iowa in June 2008 showed that ~~such~~ thunderstorm systems make critical contributions to the upper tail of flood peak distributions ~~in the region~~ (Smith et al., 2013). Although the frequent ~~heavy-rainfall events in~~ August-~~or~~September ~~heavy rainfall events evident in Fig. 3b~~ have not triggered any ~~of the~~ recorded annual flood peaks in Turkey River, our process-based FFA demonstrates that they may still relevant to current and future flood frequency, as shown in Sect. 5.~~43~~.

10 The largest annual maxima (over  $800 \text{ m}^3/\text{s}^{-1}$ ) occur in May-July (Fig. 3c), consistent with the broader climatology of flooding in Iowa (Smith et al., 2013; Villarini et al., 2011). Furthermore, both the seasonality and magnitude of flood peaks have shifted since approximately 1990 (Fig. 3a, 3c), with March-April (May-September) floods decreasing (increasing) in magnitude, leading to a shift in the seasonality of the overall distribution of annual maxima daily streamflow from a high in March prior to 1990 to a prolonged high from April-~~to~~ June post-1990. Although the small sample size of the annual maxima daily discharge during this elevated ~~1990-2016~~ late-spring-~~and~~ summertime flood period ~~(1990-2016)~~ may affect the reliability of the derived ~~PDF-distribution~~ of flood occurrence, Park and Markus (2014) ~~also~~ reported a significant shift toward summertime flooding in the nearby Pecatonica River. Statistically based FFA (including nonstationary methods) based on annual maxima discharges may fail to capture the impact of this shifting seasonality on flood frequency.



**Figure 3.** (a) Linear trends for two groups of annual maxima daily discharge: March-April floods (blue) and May-September floods (red) using the nonparametric Thiel-Sen estimator (Sen, 1968). The October-February maxima daily discharge are in black dots and its trend line is not calculated because only nine annual maxima are occur during this period. The trend line for the “overall” annual maxima time series (i.e. disregarding seasonality) is in wide grey line. The four critical time ranges are shown in black lines. (b) Occurrence densities of the date during the year for the observed annual daily maxima discharge, observed annual 4-day maxima precipitation, and simulated annual daily maxima soil moisture in Turkey River watershed from 1948 to 2016. (c) The magnitude and the date during the year for annual flood peaks in Turkey River at Garber is in (black dots) and sample-sample probability density functions (PDFs) for flood events in different periods (1933-1989, blue; 1990-2016, red) are shown. In this study, all probability densities for occurrence date are estimated using a Gaussian kernel smoother/smoothing.

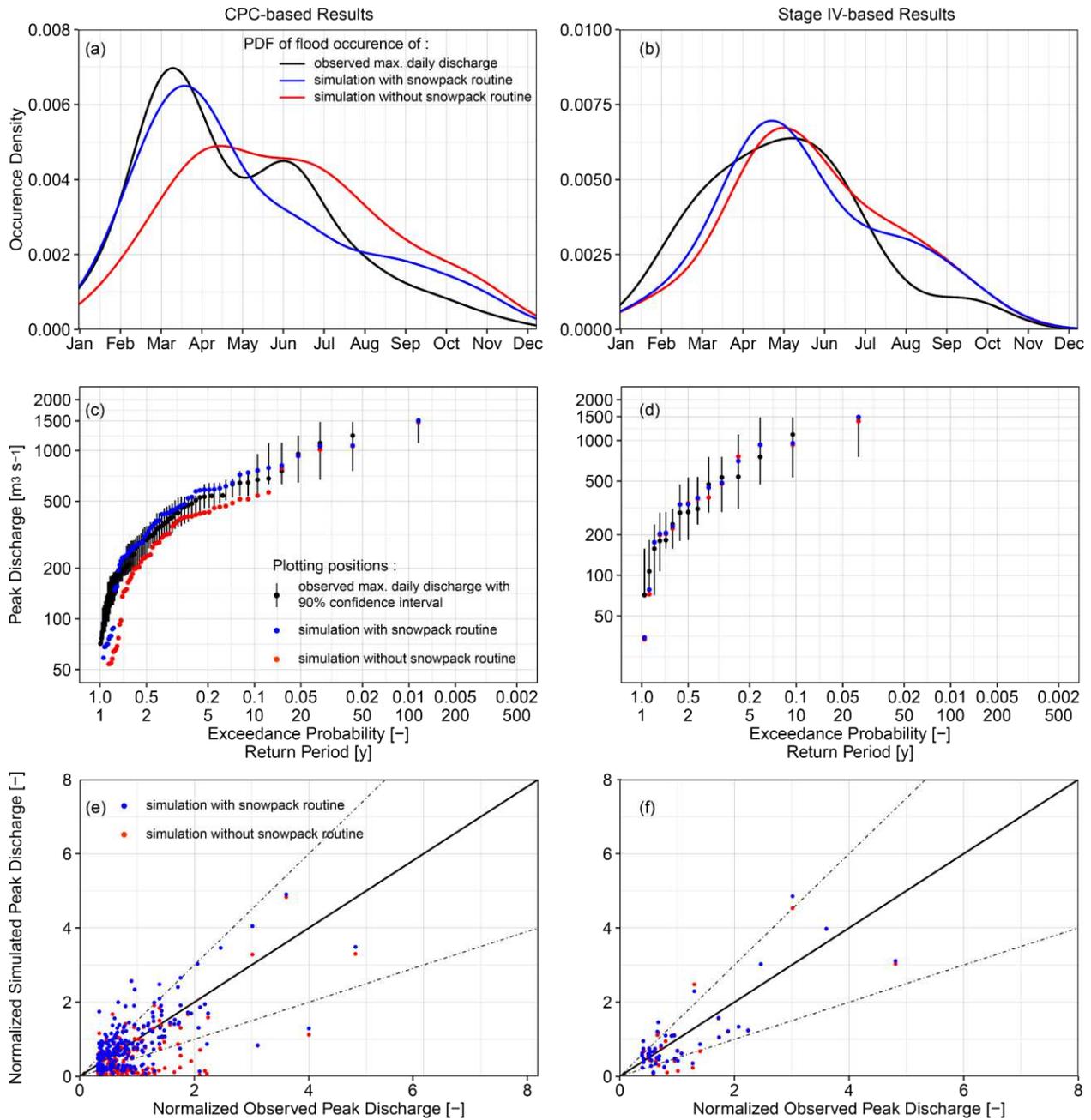
## 5.2 Model Validation

We validated the performance of continuous HBV simulations with respect to flood seasonality, frequency of annual daily discharge maxima, and normalized peak flow (i.e. the simulated or observed daily discharge divided by the 2-year flood), using both Stage IV and CPC as precipitation inputs (Fig. 4). We also validated two model structures: one with and the other

without the HBV snowpack module. The purpose for this latter validation effort is to highlight the importance of proper process representation (and subsequent validation) in process-based FFA.

5 Simulated flood seasonality varies substantially during the CPC period of record (1948-2016) depending on the inclusion of the snowpack routine (Fig. 4a). Differences are less for the Stage IV period of record (2002-2016), due to the decreasing role of snowpack in deriving the floods in recent years (Fig. 4b). In both cases, the seasonality of flooding simulated using HBV is improved with the inclusion of the snowpack module, with a higher (lower) frequency of springtime (summertime) floods which more closely resembles observations. Empirical (i.e. plotting position-based) distributions for the simulated annual daily discharge maxima are mostly within the 90% confidence interval (obtained by nonparametric bootstrap) of the observations (Fig. 4c, 4d). CPC-based simulation results differ considerably depending on the inclusion of the snowpack module for more common events, but differences in simulated maxima vanish as flood magnitude increases (e.g. AEP<0.1). This is because the most extreme flood events occur later in the season and are thus independent of snowpack or snowmelt processes. Differences are generally negligible between Stage IV-based simulations with and without snowpack, since floods in this more recent period are generally driven by summertime thunderstorms. These findings are consistent with the general understanding of the regional seasonality of flooding in the region, as discussed in Sect. 5.1.

15 We compared all simulated and observed flood peaks that can be associated with a USGS observed daily streamflow value that is at least three times the mean annual daily discharge (Fig. 4e, 4f). When associating simulated and observed flood peaks, we look within a 2-day window to allow for modest errors in simulated flood peak timing. All peaks in Figs. 4e and 4f are normalized by the median annual (i.e. 2-year) flood, which, as a rule of thumb, can be considered as the “within bank” threshold. Again, HBV with the snowpack routine outperforms the model without it, especially for the small to modest flood events in CPC-based simulations. The model without snowpack underestimates small to modest flood events in two cases due to the neglect of potential snowmelt contributions. While modest scatter exists in the Stage IV-based simulated peaks, there is no obvious systematic bias with event magnitude when the snowmelt routine is included. The good performance of the Stage IV simulations suggests that, when focusing on the recent period of elevated flood activity, Stage IV may be a more suitable rainfall input than CPC-Unified. In addition, CPC rainfall is known to contain errors in the extreme tail, due to gage  
20 “undercatch”, insufficient gage density to properly sample convective rain cells, and spatial averaging of such cells over large areas, which effectively reduces peak rainfall depths.  
25

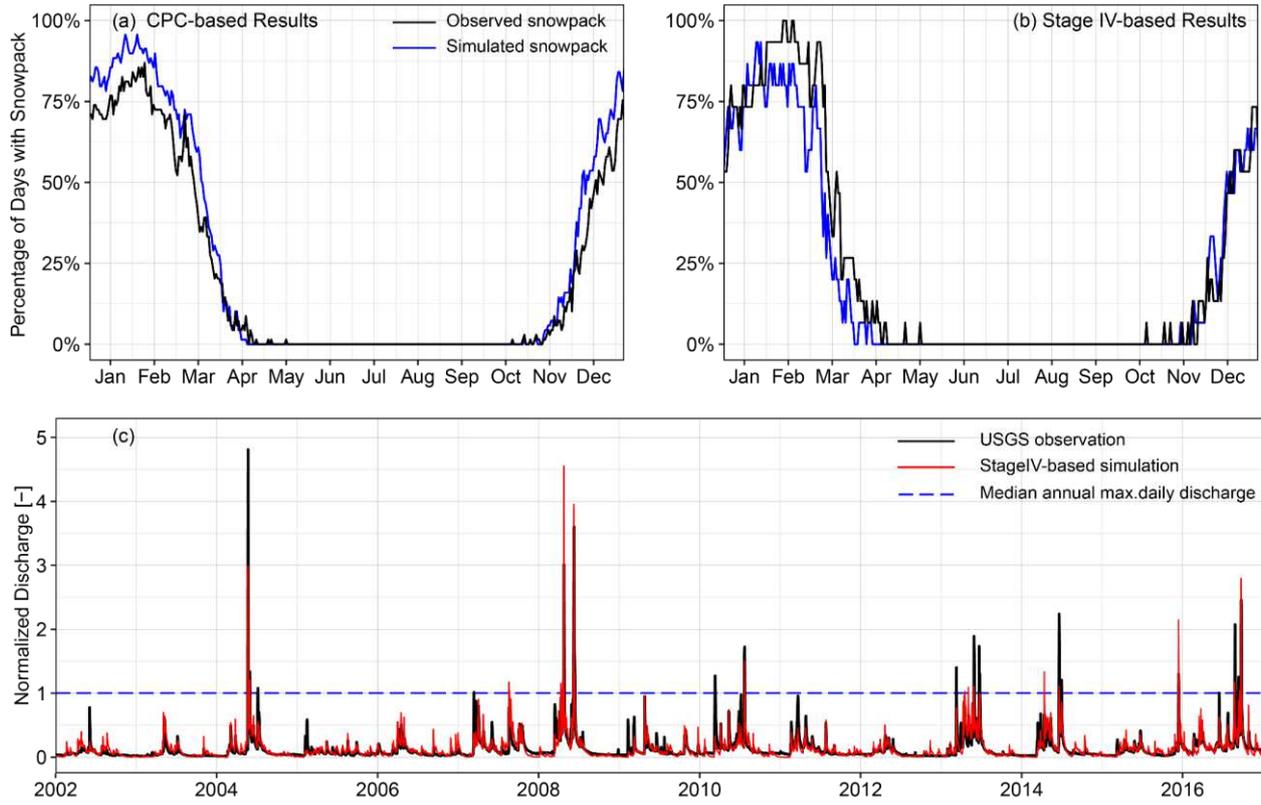


**Figure 4.** HBV model validation for flood seasonality (a, b), frequency of annual max. daily discharge (c, d) and normalized peak flow (e, f) for CPC and Stage IV-based continuous simulations. Model validation is performed for HBV simulations with and without using CPC for 1948–2016 (panels a, c, e) and Stage IV for 2002–2016 (panels b, d, f). The 90% confidence intervals for the empirical distributions of observed maximum daily discharges (c, d) are derived using nonparametric bootstrapping. Flood peak discharge in (e) and (f) is defined as a data point with USGS observed value that is at least three times the average observations. Peak discharges are normalized by the median of annual daily discharge maxima (i.e. the 2-year flood). Straight solid black lines indicate 1:1 correspondence, while dashed lines denote an envelope within which the modeled values are within 50% of observed.

5

We also validate HBV's snowpack routine using observed GHCN daily snow depth for two simulation periods (Fig. 5a, 5b) and using USGS daily streamflow observations for Stage IV-based period (Fig. 5c). Because of their differing spatial resolutions and physical representations, point-scale GHCN daily snow depths cannot be directly compared to the watershed-scale snow water equivalent simulated by HBV. Instead, we validate snowpack simulations in terms of the snowpack occurrence, defined as the number of nonzero snowpack on a particular date divided by the total number of years in the historical or simulated record. For example, there are 50 days in the GHCN observations when snowpack is present on January 1st in the 69-year period from 1948-2016, thus the occurrence rate is 0.72 (50 divided by 69). The HBV model with the snowpack routine captures the central tendency of observed snowpack dynamics, showing that snowpack frequently exists from early November to mid-February, with frequency of snow decreasing from late February until disappearing in early April.

5



10

**Figure 5.** Percentage of days with nonzero snowpack present in observations and simulations (a, b) and hydrograph validation for Stage IV-based simulation (c). For each day within a year, the percent with nonzero snowpack is calculated as the ratio of the number of years in which snowpack is present on that day to the total years (69 years for CPC and 15 years for Stage IV). Observed and simulated hydrographs are normalized by the median annual flood, which is indicated by the dashed blue line.

15

Model hydrograph validation is provided in Fig. 5c for the Stage IV period (2002-2016), when major flooding occurred throughout Iowa. Model performance shows no obvious evidence of systematic bias in the streamflow simulations (see also Fig. 4f). Although flood seasonality derived from Stage IV-based simulation differs slightly from observations (see also Fig. 4a), these mismatches are associated with flood events smaller than the median annual flood (blue dash line in Fig. 5c). Stage

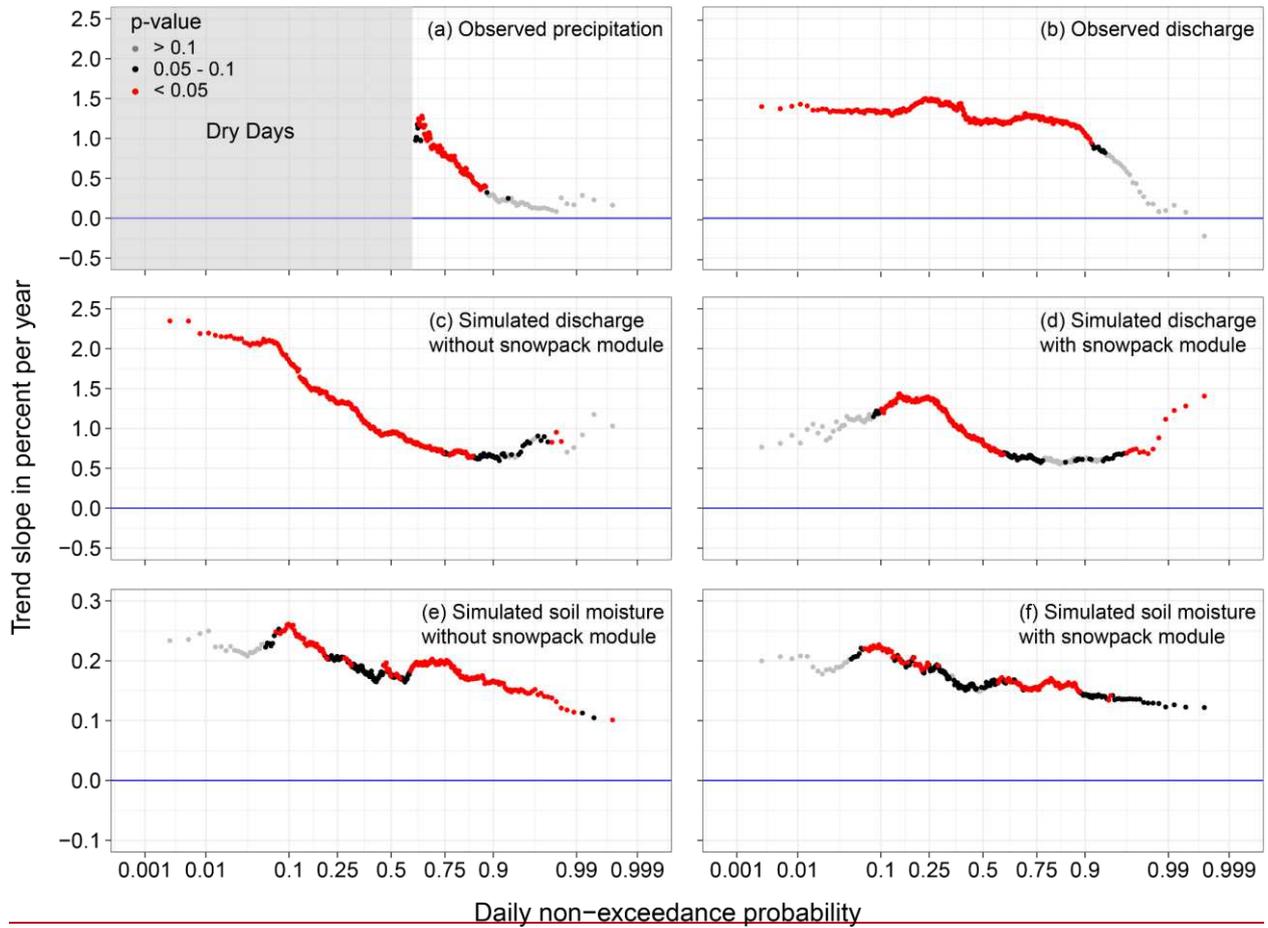
IV-based simulations do not show bias flood magnitude in late summer. In other words, remaining biases in terms of flood seasonality generally correspond with frequent, small-magnitude events that are typically of less interest in FFA. We therefore conclude that the HBV model with snowpack is generally suitable for subsequent process-based FFA.

As mentioned in Sect. 2.3, caution is needed when using hydrological models for process-based FFA. The hydrological model must be faithful, to a reasonable degree, to observed nonstationarities, while the importance of soil moisture in flooding implies that processes such as subsurface flow and storage beyond the event scale must also be adequately represented. “Quantile-Kendall plots” (Hirsch and De Cicco, 2015) for observed and simulated daily streamflows from 1948 to 2016 highlight this (Fig. 4). Each point on the plot is a trend slope computed for the 1948–2016 period using the Thiel-Sen estimator for a given quantile of the variable in question, while the color of the point indicates the significance of the trend computed using the MK test. For instance, the point at the far left (right) is the first (365<sup>th</sup>) order statistic, which is the annual minimum (maximum). The plots are useful for displaying long-term trends across the entire distribution.

Trends in observed streamflow (Fig. 4b) below the 90<sup>th</sup> percentile are largely positive (around 1.5% per year) and significant at the 5% level. Beyond the 90<sup>th</sup> percentile, the trend is less significant. Quantile-Kendall plots for simulated daily streamflow derived and without the HBV snowpack routine (Fig. 4c, 4d) reveal that simulated streamflow trends without the snowpack routine more closely resemble the observed trends, since the trend slope generally decreases with increasing quantile and the significance at high quantiles is generally low. (Note that model calibration is performed separately for both simulations, meaning the model parameters differ between them.) The plot of simulated streamflow with the snowpack routine differs substantially from observations, including a significant 1.5% per year increase in simulated annual maxima which contrasts with an insignificant 0.2% observed decrease. While a different hydrological model structure would produce different outcomes, these results highlight that certain process representations in models may produce undesirable results that could propagate through to FFA.

We also show Quantile-Kendall plots for observed daily precipitation (Fig. 4a) and simulated daily soil moisture derived without (with) the snowpack routine (Fig. 4e, 4f). Both Quantile-Kendall plots for observed precipitation and simulated soil moisture exhibit positive trend slopes that decrease moderately with increasing quantile. It can be inferred from these plots that the increases in precipitation and soil moisture appear to result in an increase in low and moderate flows across the Turkey River watershed, though their implications for flood seasonality are less clear.

Previous studies also have shown increases in annual and seasonal precipitation and streamflow totals as well as changes in the frequency of intense rain events and the seasonality of timing of precipitation in the Midwestern United States and have suggested potential causes including large-scale climate variability and climate warming (e.g. Gupta et al., 2015; Mallakpour and Villarini, 2016, 2015; Park and Markus, 2014; Yang et al., 2013). Specific attribution of the changes in Turkey River is beyond the scope of this study, but these trends nonetheless highlight the potential challenge and important considerations for FFA in a changing hydroclimate.



**Figure 4.** Quantile-Kendall plots for observed precipitation (a), observed daily streamflow (b), simulated daily streamflow with (without) snowpack routine (c, d), and simulated daily soil moisture with (without) snowpack routine (e, f) for 1948-2016. The color represents the p-value for the Mann-Kendall test. Red indicates a trend that is significant at 0.05 level. Black indicates an attained significance between 0.05 and 0.1. Grey dots indicate trends that are not significant at the 0.1 level.

5

### 5.3.2 Flood Frequency Analyses

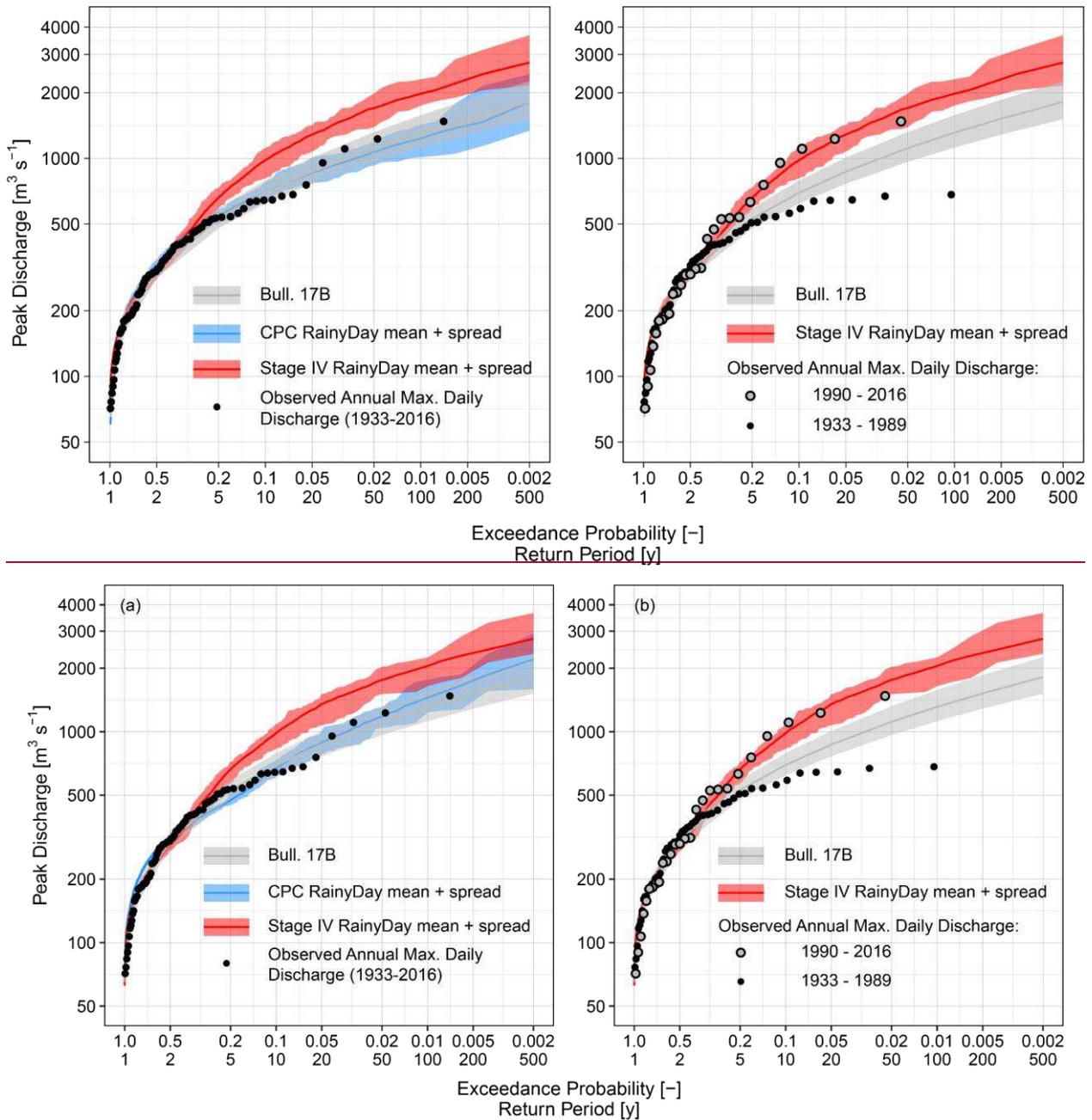
RainyDay-based FFA—flood frequency distributions for Turkey River at Garber using both Stage IV and CPC rainfall datasets/precipitation are compared with the distribution based on statistical analyses of discharge observations-based FFA using 1933-2016 USGS annual maxima daily streamflows (Fig. 56). The latter is derived-estimated using the HEC-SSP software (Bartles et al., 2016), which implements methods from USGS-Bulletin 17B (Interagency Advisory Committee on Water Data, 1982) using “station skew” to fit the log-Pearson Type III distribution. Observed annual maxima daily streamflow maxima from 1933 to 2016 are also shown, where plotting position ( $p_e$ ) is estimated using the Cunnane plotting position (Cunnane, 1978). As mentioned above, Ddifferent HBV parameters are used for the Stage IV and CPC-based simulations;

10

~~this is necessary due to different-the differing time range-periods and error properties in-of these two precipitation datasets. We did not use the snowpack routine in HBV to generate the results in Fig. 5; this routine was shown to produce unrealistic streamflow results (Fig. 4c, 4d).~~

5 The Stage IV-based flood frequency curve agrees reasonably well with the ~~Bulletin 17B results discharge-based FFA~~ for  $p_e > 0.3$  (left panel of Fig. 56), but ~~yields~~ higher estimates for ~~rarer events  $p_e < 0.3$~~ . The CPC-based curve, on the other hand, matches closely with ~~the discharge-based curve~~ Bulletin 17B. The Stage IV analyses use shorter but more recent (2002-2016) meteorological and hydrological records than the other frequency curves. When ~~the~~ streamflow observations are divided into two groups (1933-1989 and 1990-2016), it becomes clear that the recent peak flood observations align well with the Stage IV-based SST results (right panel of Fig. 56). This, along with the increasing trend of annual mean precipitation and discharge shown in the previous subsection, suggests that, despite the relatively short (15-year) rainfall record used, the Stage IV-based driven process-based FFA adequately reflects flood frequency in the wetter recent climate (a similar result is shown in Wright et al., 2017), while the ~~longer CPC-based-based~~ and Bulletin 17B methods, both based on much longer data records discharge-based methods fail to do so.

15 ~~Considering Fig. 3 and Fig. 4, the-The~~ results shown in Fig. 5-6 suggest that the recent shift from spring to summer flood activity is accompanied by a substantial shift in the flood frequency distribution. The close agreement between ~~CPC-based and dischargeprocess-based~~ results using CPC and the statistically-based analysis using Bulletin 17B FFA suggests that even in stationary situations with long records, ~~the~~ statistical methods do not necessarily produce superior results to process-based approaches. ~~We also derived the RainyDay-basedProcess-based~~ FFA using CPC Unified rainfall data precipitation from 2002 ~~to\_ 2016 and it~~ closely resembles the Stage IV-based FFA (results not shown), ~~pointing to~~ suggesting that rainfall ~~temporal process~~ nonstationarity, rather than differences between different rainfall input datasets, ~~as-are~~ the primary drivers of the differences in the CPC-based and Stage IV-based results in the left panel of Fig. 56. ~~The following subsections explore the hydrologic processes that are embedded within these process-based flood frequency curves.~~



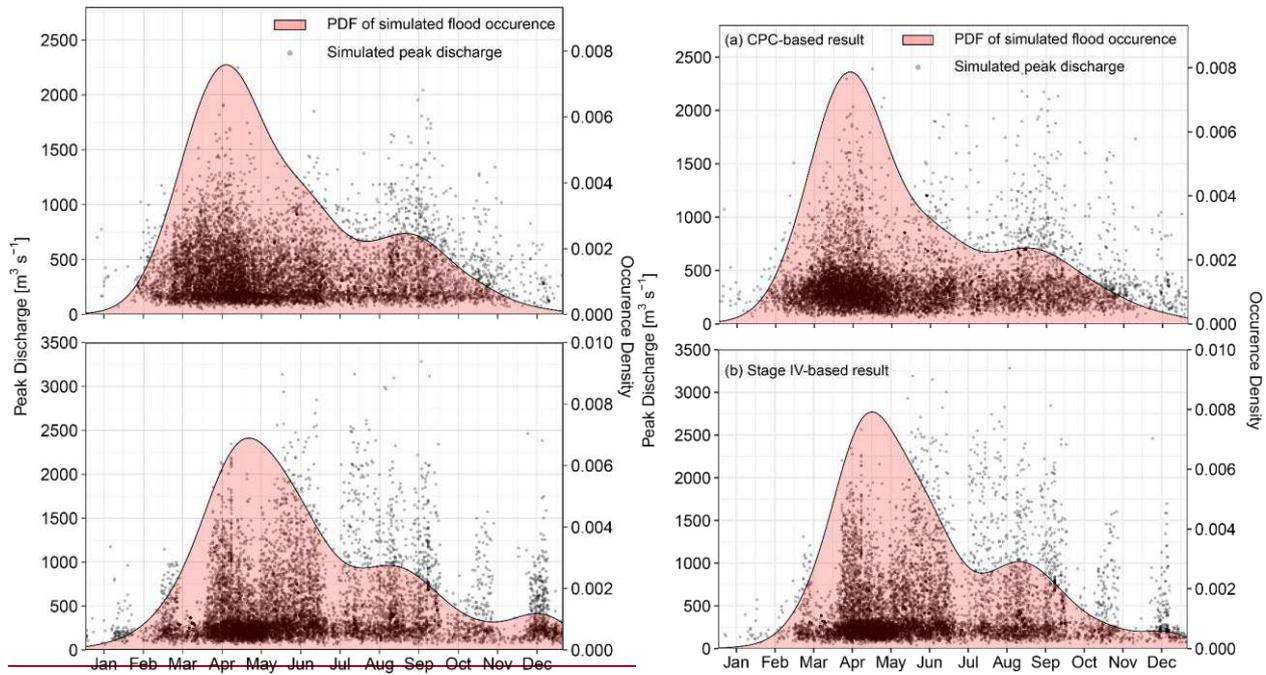
5 **Figure 56. Three peak** Peak discharge analyses for Turkey River at Garber, IA. **(a)** RainyDay with Stage IV (2002-2016) and CPC-(1948-2016) rainfall and USGS frequency analyses using Bulletin 17B methods. **All observed USGS annual maxima daily streamflow from 1933 to 2016 are also shown.** Shaded areas denote the ensemble spread (RainyDay-based results) and the 90% confidence intervals (Bulletin 17B-based analysis), respectively. **All observed annual maxima daily streamflow from 1933 to 2016 are shown in one group in the left panel, (b) Same as (a), but are with the USGS observations divided separated into two pre-1990 and post-1990 groups, in the right panel. Stage IV and Bulletin 17B curves are identical in the two panels, replotted to highlight recent changes in flood frequency.**

### 5.4.3 Simulated Flood Seasonality

As shown in Sect. 5.1, the recent climatology of flooding in Turkey River watershed shows a peak in flood activity-occurrence during March-April, with elevated activity (including high-magnitude events) continuing through July, reflective of the regional flood “mixture distribution” (e.g. Smith et al., 2011). March-April flooding is associated with springtime rains, high soil moisture, and potentially snowmelt processes, while May-July flooding results from warm-season organized thunderstorm systems. It is important that any process-based FFA approach capture the influence of this mixture on the flood frequency curve.

The seasonal distribution of simulated flood occurrence and magnitude using Stage IV- and CPC-based results show that most simulated floods in our process-based approach occur between March and June (Fig. 67), in accordance with observed annual maxima daily discharge (Fig. 3c; see also Fig. 7b). The peak of occurrence using Stage IV is shifted several weeks later than the CPC-based results, which agrees with the recent shift in seasonality of flood observations shown in the Fig. 3c. Although most-many simulated events still occur around-in April, our results show that the largest peaks occur later, in May-September. This is consistent with Villarini et al. (2011), who showed that summertime-warm season organized convective systems are responsible for some of the largest peaks in Iowa.

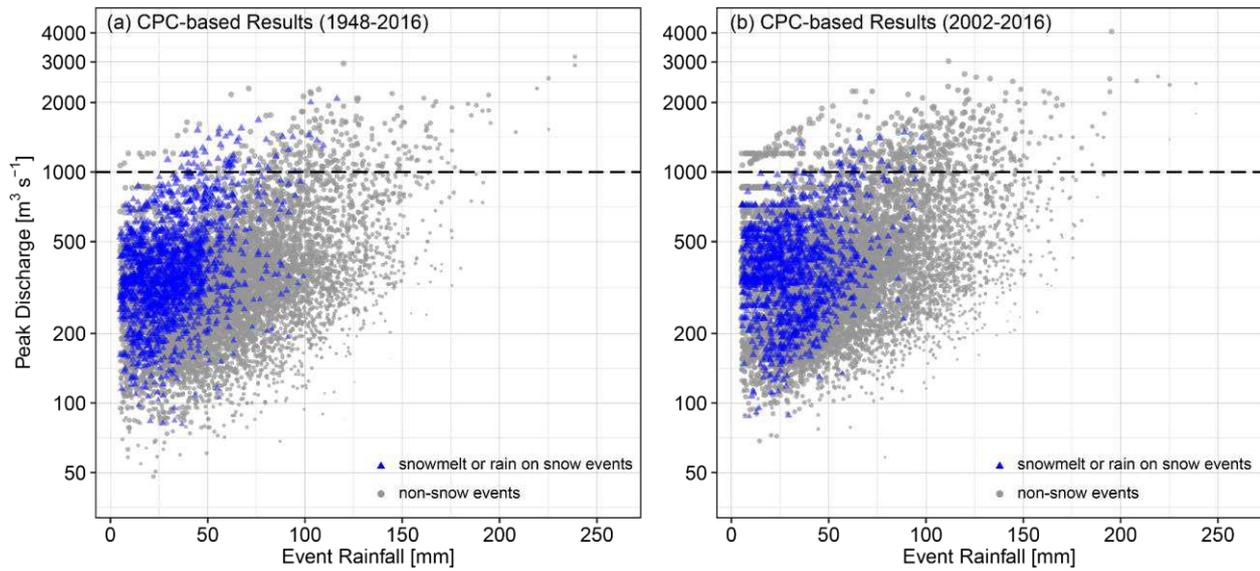
Figure 6 Our process-based results shows that rainfall-events-around-August-September storms have the potential to cause severe flooding (Fig. 7), despite the lack of observed-August-large floods during this time of year peaks in Turkey River the stream gage record. The Stage IV- and CPC-based storm catalogs generated by RainyDay, which includes major storms from the surrounding region, includes-including several large late-summer storm-events; capable of producing substantial flood response, and which produce-indeed to induce large floods within the-our process-based analysis. This suggests that the general lack of major late-summer floods in the watershed’s observational records for Turkey River may not be a feature of the “true” (unknown) distribution of flooding in the watershed, but rather due to limited size of the observational record of undersampling of this distribution in the observed flood record. This result is supported by regional analysis or-of regional floods observations (Villarini et al. 2011), and points to the potential for SST to improve representations of seasonal variations in extreme rainfall relative to local observations understanding of flood frequency seasonality relative to discharge-based approaches alone.



**Figure 67.** Time of occurrence during the year for simulated peak discharge in Turkey River at Garber using (a) CPC and (b) Stage IV.

To demonstrate that the discrepancies between the process-based FFA results generated using CPC and using Stage IV- are driven by changes in physical processes, rather than by differences in model structure (i.e. parameter values), we compared FFA results generated using CPC-based for 1948-2016 and 2002-2016, in terms of event rainfall, initial soil moisture, flood type and peak magnitude (Fig. 8). Compared with the 1948-2016 period (Figure. 8a), there are fewer flood events driven by snowmelt or rain-on-snow during 2002-2016 (Fig. 8b) but more driven by rainfall. This is particular true for flood events (larger than  $1000 \text{ m}^3 \text{ s}^{-1}$ ). In addition, some of the rainfall-driven floods from 2002-2016 were caused by relatively low rainfall but high initial soil moisture, in accordance with the significant increasing trend of annual precipitation and discharge (Table

10 2).



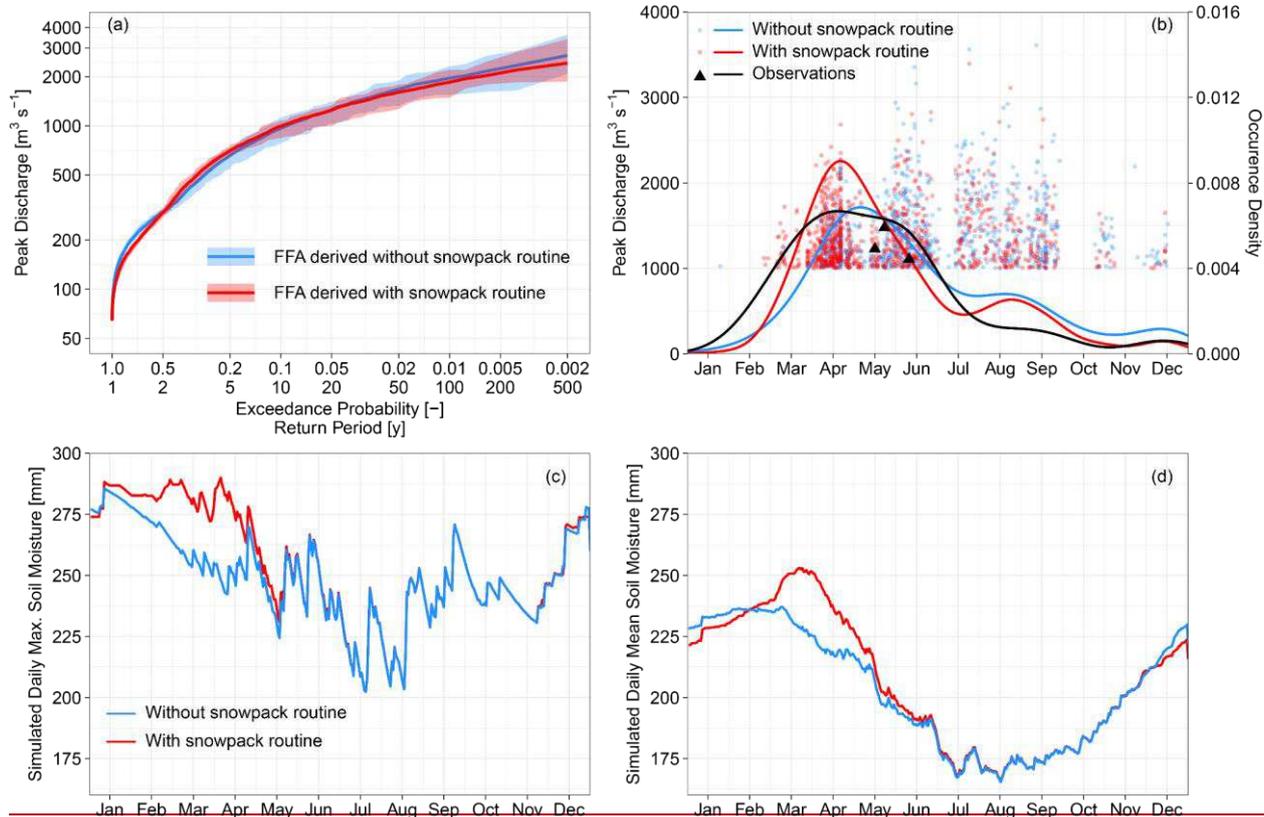
**Figure 8.** The simulated flood magnitude using CPC rainfall during 1948-2016 (a) and 2002-2016 (b) periods, and corresponding antecedent conditions. The blue triangles denote the snow related flood events (e.g. snowmelt was nonzero in the simulation) and grey dots represent the non-snow related flood events (e.g. rainfall driven). The size of the triangles or dots indicate the antecedent soil moisture with higher value in larger shape. The black dash line indicates the  $1000 \text{ m}^3 \text{ s}^{-1}$  flood magnitudes.

5

We also examined how hydrological model process representation influence FFA results. We showed previously that the HBV snowpack routine produces trends in simulated daily streamflow that are less realistic than simulated trends without the routine (Fig. 4c, 4d). Interestingly, however, we found very similar flood frequency curves regardless of whether the snowpack routine is used (Fig. 7a) despite very different simulated seasonality (Fig. 7b). With snowpack routine, our approach simulates many large floods (over  $1000 \text{ m}^3/\text{s}$ ) between February and April (Fig. 7b). This is due to high March-April soil moisture value (Fig. 7e, 7d) associated with snowmelt, which increases the probability of flood occurrence during this period. The flood seasonalities derived from historical observations and from the simulated results without the snowpack module, meanwhile, do not exhibit the very frequent April floods that are present in the simulations with the snowpack routine. This example shows that process-based frequency analyses can be subject to issues related poor hydrological model process representation which can produce “correct” results for the wrong reasons. This implies that the modeler must either have sufficient data to diagnose such issues (as we have done here) or have sufficient prior knowledge of the seasonally varying flood processes in her study area to recognize such pitfalls.

10

15

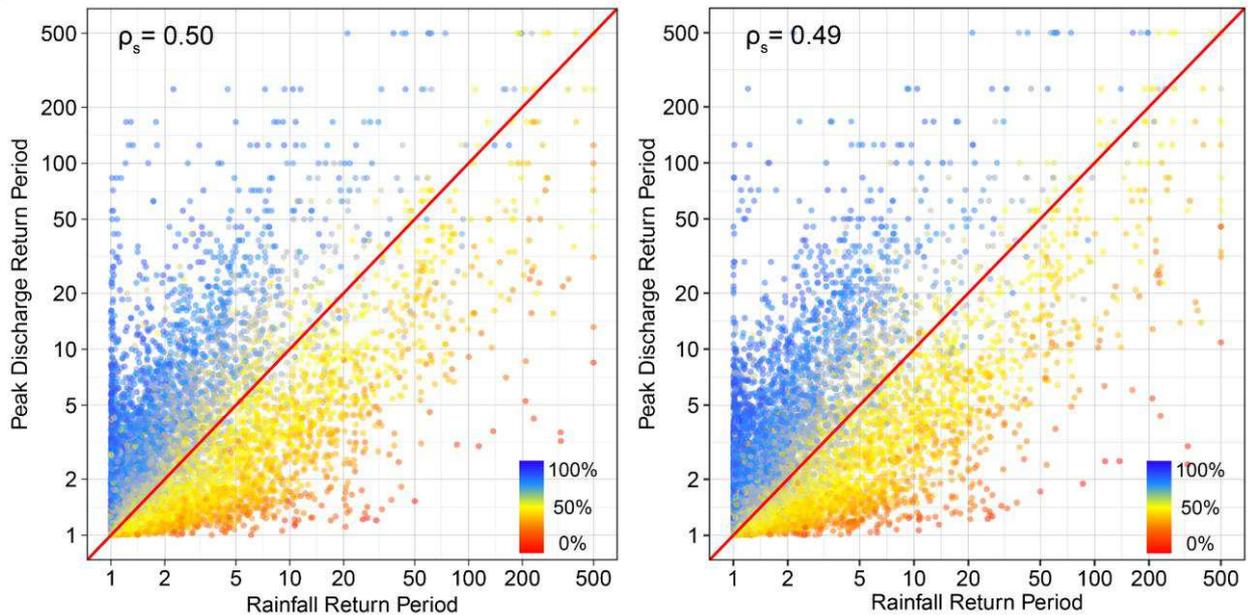


**Figure 7.** Flood frequency curves (a) and simulated floods seasonality (b) derived by RainyDay-based approach with Stage IV precipitation with (red) or without (blue) the HBV snowpack routine. Simulated discharge larger than 1000 m<sup>3</sup>/s (dots) and observed peak streamflow (1990-2016) larger than 1000 m<sup>3</sup>/s (triangles) are shown; discharges below 1000 m<sup>3</sup>/s are omitted for clarity. The maximum (c) and mean (d) of simulated soil moisture for each day of the year are shown.

The results shown in Fig. 7 also illustrate a key issue in FFA using both statistical approaches and process-based methods: flood quantiles, though the product of physical processes, reveal little about the underlying processes. This is particularly problematic in changing hydroclimatic or watershed conditions, because nonstationary behavior is likely the result of seasonal shifts in one or more processes. Failure to recognize shifts could lead to incorrect predictions of future conditions. For example, our findings using the HBV snowpack routine predict that most floods are due to high springtime soil moisture due to snowmelt (Fig. 7c, 7d). If we were to project future flood frequency in a warming climate, we might conclude that these spring floods will diminish in importance and thus the tail of the flood distribution will decrease in magnitude. Observations, in contrast, show that an important shift toward summertime flooding has occurred, which may imply the opposite behavior in the tail of the flood distribution in Turkey River since warm season convective rainfall extremes are predicted to increase (e.g. Prein et al., 2016).

### 5.54 Comparison ~~between of~~ rainfall and peak discharge quantiles

We examined the relationships between the return periods of 96-hour basin-averaged rainfall accumulations and simulated peak discharge for Turkey River at Garber using Stage IV-based results (Fig. 89; CPC-based results show similar patterns and thus are not shown here). Antecedent soil moisture for each simulated event is also shown. Similar to Wright et al. (2014a), Fig. 8-9 shows that simulated peak discharge quantiles can differ substantially from the rainfall quantiles of the rainfall that produce them. For instance, 500-year ( $p_e = 0.002$ ) rainfall events can cause simulated peak discharges ranging from ~~211-~~ year ( $p_e = 0.04891$ ) to 500-year ( $p_e = 0.002$ ), corresponding to a range in peak discharge of ~~890-1072~~ to ~~2990-2743~~  $\text{m}^3 \text{s}^{-1}$ . ~~Results indicate that the P~~ peak discharge quantiles are always larger (in terms of return period) than the quantiles of rainfall quantiles that produced them in wet antecedent soil moisture conditions, while the reverse is true in for dry conditions. These results also demonstrate that the DS assumption of 1:1 equivalency between rainfall and peak discharge quantiles does not hold in Turkey River. Rainfall spatial variability and drainage network structure, which are ignored in this study due to the lumped (i.e. non-distributed) nature of HBV, further complicate the relationship between rainfall and discharge quantiles.



**Figure 89.** Relationships between rainfall ~~return periods~~ and simulated peak discharge return periods estimated via our RainyDay (process)-based method using Stage IV rainfall data. Spearman rank correlation  $\rho_s$  is given. ~~Shading e~~Color indicates the normalized modeled antecedent soil moisture value, ~~which is~~ calculated as  $Normalized\ soil\ moisture = \frac{soil\ moisutre - min.soil\ moisutire}{field\ capacity - min.soil\ moisutire} * 100\%$ .

We further examine the relationship between annual rainfall and annual flood peak maxima. In Sect. 2.2, we pointed out that DS methods utilize IDF curves, which are usually estimated using annual maxima from rain gage records and which depict, to estimate quantiles from the distribution of annual ~~discharge peaks~~rainfall maxima. DS methods use quantiles from this distribution to generate flood estimates, implicitly assuming that annual rainfall maxima produce annual discharge maxima.

In our process-based FFA approach, ~~however,~~ we do not assume that annual discharge maxima are the result of the largest

rainfall event of the year. Rather, lower-magnitude rainfall events, combined with high soil moisture, could produce the highest discharge. ~~Table 3~~~~Table 3~~~~Table 3~~ shows the percentage of annual peak flow driven by annual maximum gains with increasing return period for both CPC-based and Stage IV-based results. For simulated peak flow with  $p_e > 0.01$ , a large portion of simulated annual peak flow is not caused by the annual maximum rainfall. For rarer peak flows ( $p_e \leq 0.01$ ), over 90% of these flood events are driven by the annual maximum rainfall, pointing to the fact that the tail of flood peaks is driven by extreme rainfall, with antecedent conditions playing a modulating role.

**Table 3.** Percentages of simulated annual ~~peak-maxima daily flows~~ driven by ~~annual maximum~~ 96-hour rainfall ~~annual maximum~~.

Return Period	Driven by Annual Maximum Rainfall	
	CPC-based results	Stage IV-based results
1-2	<del>29</del> <u>24</u> %	<del>36</del> <u>37</u> %
2-5	<del>46</del> <u>32</u> %	<del>48</del> <u>45</u> %
5-10	<del>59</del> <u>39</u> %	<del>66</del> <u>67</u> %
10-20	<del>70</del> <u>48</u> %	<del>75</del> <u>77</u> %
20-50	<del>76</del> <u>60</u> %	80%
50-100	<del>81</del> <u>72</u> %	<del>88</del> <u>84</u> %
100-200	<del>90</del> <u>77</u> %	<del>97</del> <u>85</u> %
200-500	<del>98</del> <u>93</u> %	<del>93</del> <u>95</u> %

## 6 Summary and Conclusions

Interactions between rainfall, land cover, river channel morphology, and watershed antecedent conditions are important drivers of flood response. Standard approaches to estimate extreme flood quantiles (termed flood frequency analysis; FFA), however, ~~often~~ take a superficial view of these interactions, as argued in Sect. 2. This study presents an alternative FFA framework that combines elements of observational analysis, stochastic rainfall generation, and continuous and event-scale ~~rainfall-runoff~~~~hydrologic~~ simulation. We apply the framework to Turkey River, an agricultural watershed in the Midwestern United States that is undergoing significant hydroclimatologic and hydrologic changes which is increasing the magnitude of the largest flood events and shifting their occurrence from the spring to summer.

We use Stochastic Storm Transposition (SST) to create and resample from “storm catalogs” developed from both 15 years of high-resolution bias-corrected radar rainfall ~~dataset~~ and ~~from~~ 69 years of gridded rain gage observations to produce large numbers of rainfall scenarios for Turkey River. ~~These scenarios, when coupled with seasonally realistic watershed conditions,~~

can help to reconstruct the seasonal and secular variations in meteorological and hydrological processes and their interactions, providing an alternative FFA approach which is well-suited to nonstationary environments (see also Sivapalan and Samuel, 2009). While statistical approaches can in principle be applied to investigate the impacts of seasonality on FFA (e.g. Ouarda et al., 2006), such methods still do not directly provide process-level understanding of the factors that “shape” flood frequency.

5 Unlike design storm approaches to FFA, the synthetic rainfall scenarios derived by the SST-based procedure do not require any assumptions regarding the spatial and temporal structure of rainfall, since they are driven by the structure and variability of historical observed storms. ~~Unlike design storm approaches to FFA, the synthetic rainfall scenarios derived by the SST-based procedure do not require any assumptions regarding the spatial and temporal structure of rainfall, since they are driven by the structure and variability of historical observed storms. Unlike discharge-based statistical analyses, our approach helps~~  
10 ~~shed light on the physical processes that shape flood frequency. Resampling and spatial transposing of observed rainstorms from the surrounding region makes it feasible to generate extreme precipitation scenarios using relatively short rainfall records. In nonstationary rainfall conditions, recent rainfall data can produce more realistic rainfall scenarios and flood quantile estimates than methods that rely on longer records.~~

Our analyses show that using the most recent 15 years of rainfall can produce realistic “present-day” flood quantile estimates  
15 that reflect the nonstationarities in rainfall and watershed conditions. ~~Use~~ The use of longer records, both within our procedure and conventional statistical FFA methods, leads to underestimates of current flood frequency due to their inability to represent recent shifts in flood activity in Turkey River. Our results challenge some common FFA assumptions, including the design storm presumption that rainfall annual maxima produce discharge annual maxima and the assumption of 1:1 equivalence in rainfall and flood quantiles. We paint a more complex picture in Turkey River, in which the shifting seasonality in rainfall and  
20 watershed conditions combine to shape the flood frequency.

Spatial variability in rainfall structure, soil moisture, land use and watershed morphology, which are ignored in this study due to the use of a lumped hydrological model, ~~as~~ add further complexity to the flood-generating processes. ~~However,~~ †The proposed framework can be employed with more sophisticated distributed hydrological models, thus facilitating the examination of rainfall spatial variability and its interactions with other factors (e.g. heterogeneous watershed characteristics and river network processes; Zhu et al., 2018; Viglione et al., 2010b, 2010a). This coupling may prove particularly useful for  
25 FFA in large watersheds in which there is a practically infinite number of different combinations of such spatially and temporally varying processes ~~factors~~ that could produce floods—a population that is almost certain to be undersampled in stream gage records and poorly served by design storm assumptions.

A number of issues remain that make broader usage of our process-based framework challenging. Perhaps the biggest  
30 limitation of process-based approaches is the necessity of discharge observations, which are central to both identifying hydrologic changes and to calibrate and validate the hydrologic model. Thus, usage of the approach in ungaged basins may not produce satisfactory results. This issue is fundamental to other FFA techniques as well. Statistically-based discharge analyses, for example, similarly rely on streamflow observations, while design storm approaches also require hydrologic model calibration.

We also note that caution is needed when attempting to employ process-based FFA. We were able to produce very similar flood frequency distributions using our approach, regardless of whether or not the HBV hydrologic model's snowpack routine was "turned on" or off (results omitted for brevity), despite very different simulated seasonality of flooding. This highlights that process-based frequency analyses can be influenced by poor model process representation that can lead to seemingly "correct" results for the wrong reasons. This implies that the modeler must have sufficient data and experience to recognize such issues. It also illustrates a key issue in FFA using both statistical approaches and process-based methods: flood quantiles, though the product of interactions between physical processes, reveal relatively little about those underlying processes that produce them. This is particularly problematic in changing hydroclimatic or watershed conditions, because nonstationary behavior is likely the result of seasonal shifts in one or more processes that may affect flooding in ways that are not well-reflected in observational records. Our results showing that major floods could occur in Turkey River in the late summer under current hydroclimatic conditions, despite their absence in the instrumental record, is one example of this. Failure to recognize and model such shifts could lead to results for past or present flood conditions that appear to be correct, but that may lead to incorrect inferences about future conditions.

In summary, our framework and results highlights the opportunity and challenge with process-based FFA approaches; namely, that progress on understanding and estimating flood frequency and how it is evolving in an era of unprecedented changes in land use and climate requires better understanding of how the underlying physical processes, and the interactions between them, are changing. Poor model representation of key hydrological processes, however, can lead to incorrect conclusions about present or future flood frequency. Despite the challenge, we share the view of Sivapalan and Samuel (2009), however, that process-based approaches hold great potential for advances in FFA research and practice, particularly in projecting future flood hazards in conjunction with data and modeling advances in the climate science community. We do not propose that process-based approaches should necessarily supplant more conventional discharge-based analyses, and acknowledge that discharge observations are essential in such studies. Rather, we anticipate a gradual "merging" of statistical and process-based stochastic simulation techniques as well as of the associated observations and synthetic data.

### Software and model code

The RainyDay software is available at Github ~~container~~ (<https://github.com/danielbwright/RainyDay2>), and a web-based version of RainyDay is available at ~~Daniel. W.D. Wright~~'s research group website (<http://her.cce.wisc.edu/projects/rainyday>).

### Competing interests

The authors declare that they have no conflict of interest.

## Acknowledgements

~~Gue. Yu~~'s and ~~Daniel. W.D. Wright~~'s contributions were supported by the U.S. National Science Foundation (NSF) Hydrologic Sciences Program (award number 1749638) and by the Bureau of Reclamation Research and Development Office Project ID 1735, which also supported ~~Kathleen. H.K. Holman~~'s contributions. ~~Cassia. S.C. Smith~~'s contributions were supported by the  
5 U.S. National Aeronautics and Space Administration's MUREP Institutional Research Opportunity. ~~Zhibua. Z.Z. Zhu~~'s contributions were supported by Sun Yat-sen University. This study used computing resources and assistance from the UW-Madison Center For High Throughput Computing (CHTC), which is supported by UW-Madison, the Advanced Computing Initiative, the Wisconsin Alumni Research Foundation, the Wisconsin Institutes for Discovery, and NSF, and is an active member of the Open Science Grid, which is supported by NSF and the U.S. Department of Energy. We thank M. Booij and  
10 M. Vis for their guidance on the HBV model.

## References

- Alexander, G. N.: Using the probability of storm transposition for estimating the frequency of rare floods, *J. Hydrol.*, 1, 46–57, 1963.
- Ayalew, T., Krajewski, W. and Mantilla, R.: Exploring the Effect of Reservoir Storage on Peak Discharge Frequency, *J. Hydrol. Eng.*, 18(12), 1697–1708, doi:10.1061/(ASCE)HE.1943-5584.0000721, 2013.
- 15 Ayalew, T. B. and Krajewski, W. F.: Effect of River Network Geometry on Flood Frequency: A Tale of Two Watersheds in Iowa, *J. Hydrol. Eng.*, 22(8), 06017004, doi:10.1061/(ASCE)HE.1943-5584.0001544, 2017.
- Ball, J., Babister, M., Nathan, R., Weeks, W., Weinmann, E. and Retallick, M.: Australian Rainfall and Runoff: A Guide to Flood Estimation, edited by I. Testoni, Commonwealth of Australia (Geoscience Australia), 2016.
- 20 Bartles, M., Brunner, G., Fleming, M., Faber, B. and Slaughter, J.: HEC-SSP Statistical Software Package Version 2.1, Computer Program Documentation, US Army Corps of Engineers, Institute for Water Resources Hydrologic Engineering Center (HEC), 609 Second Street Davis, CA 95616-4687., 2016.
- Berghuijs, W. R., Woods, R. A., Hutton, C. J. and Sivapalan, M.: Dominant flood generating mechanisms across the United States: Flood Mechanisms Across the U.S., *Geophys. Res. Lett.*, 43(9), 4382–4390, doi:10.1002/2016GL068070, 2016.
- 25 Bergström, S.: The HBV Model: Its Structure and Applications, Swedish Meteorological and Hydrological Institute (SMHI), Hydrology, Norrköping., 1992.
- Bergström, S.: The HBV model (Chapter 13), in *Computer Models of Watershed Hydrology*, edited by V. P. Singh, pp. 443–476, Water Resources Publications, Highlands Ranch, Colorado, USA., 1995.
- Beven, K.: Prophecy, reality and uncertainty in distributed hydrological modelling, *Adv. Water Resour.*, 16(1), 41–51, doi:10.1016/0309-1708(93)90028-E, 1993.
- 30 Beven, K. and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Process.*, 6(3), 279–298, doi:10.1002/hyp.3360060305, 1992.

- Beven, K. and Binley, A.: GLUE: 20 years on: GLUE: 20 YEARS ON, *Hydrol. Process.*, 28(24), 5897–5918, doi:10.1002/hyp.10082, 2014.
- Blazkova, S. and Beven, K.: Flood frequency prediction for data limited catchments in the Czech Republic using a stochastic rainfall model and TOPMODEL, *J. Hydrol.*, 195(1–4), 256–278, 1997.
- 5 Blazkova, S. and Beven, K.: Flood frequency estimation by continuous simulation for a catchment treated as ungauged (with uncertainty), *Water Resour. Res.*, 38(8), 14-1-14–14, doi:10.1029/2001WR000500, 2002.
- Blazkova, S. and Beven, K.: A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic, *Water Resour. Res.*, 45(12), doi:10.1029/2007WR006726, 2009.
- 10 Calver, A., Lamb, R. and Morris, S. E.: River flood frequency estimation using continuous runoff modelling, *Proc. Inst. Civ. Eng. - Water Marit. Energy*, 136(4), 225–234, 1999.
- Calver, A., Stewart, E. and Goodsell, G.: Comparative analysis of statistical and catchment modelling approaches to river flood frequency estimation: River flood frequency estimation, *J. Flood Risk Manag.*, 2(1), 24–31, doi:10.1111/j.1753-318X.2009.01018.x, 2009.
- 15 Cameron, D., Beven, K. and Tawn, J.: An evaluation of three stochastic rainfall models, *J. Hydrol.*, 228(1–2), 130–149, doi:10.1016/S0022-1694(00)00143-8, 2000.
- Camici, S., Tarpanelli, A., Brocca, L., Melone, F. and Moramarco, T.: Design soil moisture estimation by comparing continuous and storm-based rainfall-runoff modeling: DESIGN SOIL MOISTURE ESTIMATION, *Water Resour. Res.*, 47(5), doi:10.1029/2010WR009298, 2011.
- 20 Ceola, S., Laio, F. and Montanari, A.: Satellite nighttime lights reveal increasing human exposure to floods worldwide, *Geophys. Res. Lett.*, 41(20), 7184–7190, doi:10.1002/2014GL061859, 2014.
- Chen, H., Xu, C.-Y. and Guo, S.: Comparison and evaluation of multiple GCMs, statistical downscaling and hydrological models in the study of climate change impacts on runoff, *J. Hydrol.*, 434–435, 36–45, doi:10.1016/j.jhydrol.2012.02.040, 2012.
- 25 Chen, M., Shi, W., Xie, P., Silva, V. B. S., Kousky, V. E., Wayne Higgins, R. and Janowiak, J. E.: Assessing objective techniques for gauge-based analyses of global daily precipitation, *J. Geophys. Res.*, 113(D4), doi:10.1029/2007JD009132, 2008.
- Cheng, L., AghaKouchak, A., Gilleland, E. and Katz, R. W.: Non-stationary extreme value analysis in a changing climate, *Clim. Change*, 127(2), 353–369, doi:10.1007/s10584-014-1254-5, 2014.
- 30 Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Brekke, L. D., Arnold, J. R., Gochis, D. J. and Rasmussen, R. M.: A unified approach for process-based hydrologic modeling: 1. Modeling concept: A unified approach for process-based hydrologic modeling, *Water Resour. Res.*, 51(4), 2498–2514, doi:10.1002/2015WR017198, 2015a.

- Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Gochis, D. J., Rasmussen, R. M., Tarboton, D. G., Mahat, V., Flerchinger, G. N. and Marks, D. G.: A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies: A unified approach for process-based hydrologic modeling, *Water Resour. Res.*, 51(4), 2515–2542, doi:10.1002/2015WR017200, 2015b.
- 5 Crum, T. D. and Alberty, R. L.: The WSR-88D and the WSR-88D operational support facility., *Bull. Am. Meteorol. Soc.*, 74, 1669–1687, 1993.
- Cudworth, A. G.: *Flood hydrology manual*, US Dept. of the Interior, Bureau of Reclamation, Denver Office., 1989.
- Cunha, L. K., Krajewski, W. F., Mantilla, R. and Cunha, L.: A framework for flood risk assessment under nonstationary conditions or in the absence of historical data, *J. Flood Risk Manag.*, 4(1), 3–22, doi:10.1111/j.1753-318X.2010.01085.x, 10 2011.
- Cunnane, C.: Unbiased plotting positions—a review, *J. Hydrol.*, (37), 205–222, 1978.
- Dawdy, D. R., Griffis, V. W. and Gupta, V. K.: *Regional Flood-Frequency Analysis: How We Got Here and Where We Are Going*, *J. Hydrol. Eng.*, 17(9), 953–959, doi:10.1061/(ASCE)HE.1943-5584.0000584, 2012.
- Douglas, E. M., Vogel, R. M. and Kroll, C. N.: Trends in floods and low flows in the United States: impact of spatial 15 correlation, *J. Hydrol.*, 16, 2000.
- Eagleson, P. S.: Dynamics of flood frequency, *Water Resour. Res.*, 8(4), 878–898, doi:10.1029/WR008i004p00878, 1972.
- El Adlouni, S., Ouarda, T. B. M. J., Zhang, X., Roy, R. and Bobée, B.: Generalized maximum likelihood estimators for the nonstationary generalized extreme value model, *Water Resour. Res.*, 43(3), doi:10.1029/2005WR004545, 2007.
- Enfield, D. B., Mestas-Nuñez, A. M. and Trimble, P. J.: The Atlantic Multidecadal Oscillation and its relation to rainfall and 20 river flows in the continental U.S., *Geophys. Res. Lett.*, 28(10), 2077–2080, doi:10.1029/2000GL012745, 2001.
- England, J. F., Julien, P. Y. and Velleux, M. L.: Physically-based extreme flood frequency with stochastic storm transposition and paleoflood data on large watersheds, *J. Hydrol.*, 510, 228–245, doi:10.1016/j.jhydrol.2013.12.021, 2014.
- Fontaine, T. A. and Potter, K. W.: Estimating probabilities of extreme rainfalls, *J. Hydraul. Eng.*, 115(11), 1562–1575, 1989.
- Foufoula-Georgiou, E.: A probabilistic storm transposition approach for estimating exceedance probabilities of extreme 25 precipitation depths, *Water Resour. Res.*, 25, 799–815, 1989.
- Franchini, M., Helmlinger, K. R., Foufoula-Georgiou, E. and Todini, E.: Stochastic storm transposition coupled with rainfall—runoff modeling for estimation of exceedance probabilities of design floods, *J. Hydrol.*, 175(1–4), 511–532, 1996.
- Franchini, M., Galeati, G. and Lolli, M.: Analytical derivation of the flood frequency curve through partial duration series analysis and a probabilistic representation of the runoff coefficient, *J. Hydrol.*, 303(1–4), 1–15, 30 doi:10.1016/j.jhydrol.2004.07.008, 2005.
- Franks, S. W. and Kuczera, G.: *Flood frequency analysis: Evidence and implications of secular climate variability*, New South Wales, *Water Resour. Res.*, 38(5), 20-1-20–7, doi:10.1029/2001WR000232, 2002.
- Furrer, E. M. and Katz, R. W.: Improving the simulation of extreme precipitation events by stochastic weather generators, *Water Resour. Res.*, 44(12), n/a-n/a, doi:10.1029/2008WR007316, 2008.

- Gilleland, E. and Katz, R. W.: extRemes 2.0: An Extreme Value Analysis Package in *R*, *J. Stat. Softw.*, 72(8), doi:10.18637/jss.v072.i08, 2016.
- Gilroy, K. L. and McCuen, R. H.: A nonstationary flood frequency analysis method to adjust for future climate change and urbanization, *J. Hydrol.*, 414–415, 40–48, doi:10.1016/j.jhydrol.2011.10.009, 2012.
- 5 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(1–2), 80–91, doi:10.1016/j.jhydrol.2009.08.003, 2009.
- Gupta, V. K.: Transposition of Storms for Estimating Flood Probability Distributions, *Colo. State Univ. Hydrol. Pap.*, 59, 35, 1972.
- 10 Haberlandt, U.: A space-time hybrid hourly rainfall model for derived flood frequency analysis, *Hydrol. Earth Syst. Sci.*, 15, 2008.
- Harlin, J. and Kung, C.-S.: Parameter uncertainty and simulation of design floods in Sweden, *J. Hydrol.*, 137(1–4), 209–230, 1992.
- Hirsch, R. M. and Ryberg, K. R.: Has the magnitude of floods across the USA changed with global CO<sub>2</sub> levels?, *Hydrol. Sci. J.*, 57(1), 1–9, doi:10.1080/02626667.2011.621895, 2012.
- 15 Hyndman, D. W.: Impacts of Projected Changes in Climate on Hydrology, in *Global Environmental Change*, edited by B. Freedman, pp. 211–220, Springer Netherlands, Dordrecht., 2014.
- IFC: Hydrologic Assessment of the Turkey River Watershed (DRAFT), Iowa Flood Center, 100 C. Maxwell Stanley Hydraulics Laboratory Iowa City, Iowa 52242., 2014.
- 20 Interagency Advisory Committee on Water Data (IACWD): Guidelines for Determining Flood Flow Frequency, Bulletin 17B, Reston, VA., 1982.
- Jain, S. and Lall, U.: Magnitude and timing of annual maximum floods: Trends and large-scale climatic associations for the Blacksmith Fork River, Utah, *Water Resour. Res.*, 36(12), 3641–3651, doi:10.1029/2000WR900183, 2000.
- Kjeldsen, T. R.: The revitalised FSR/FEH rainfall-runoff method, NERC/Centre for Ecology & Hydrology., 2007.
- 25 Klemeš, V.: Operational testing of hydrological simulation models, *Hydrol. Sci. J.*, 31(1), 13–24, doi:10.1080/02626668609491024, 1986.
- Klemeš, V.: Tall tales about tails of hydrological distributions. I, *J. Hydrol. Eng.*, 5(3), 227–231, 2000a.
- Klemeš, V.: Tall tales about tails of hydrological distributions. II, *J. Hydrol. Eng.*, 5(3), 232–239, 2000b.
- Konrad, C. P. and Booth, D. B.: Hydrologic Trends Associated with Urban Development for Selected Streams in the Puget  
30 Sound Basin, Western Washington, , 48, 2002.
- Lamb, R.: Calibration of a conceptual rainfall-runoff model for flood frequency estimation by continuous simulation, *Water Resour. Res.*, 35(10), 3103–3114, 1999.

- Lamb, R., Faulkner, D., Wass, P. and Cameron, D.: Have applications of continuous rainfall-runoff simulation realized the vision for process-based flood frequency analysis?: Process-based Flood Frequency: Have Applications Realized the Vision?, *Hydrol. Process.*, 30(14), 2463–2481, doi:10.1002/hyp.10882, 2016.
- Li, J., Thyer, M., Lambert, M., Kuczera, G. and Metcalfe, A.: An efficient causative event-based approach for deriving the annual flood frequency distribution, *J. Hydrol.*, 510, 412–423, doi:10.1016/j.jhydrol.2013.12.035, 2014.
- 5 Lin, Y. and Mitchell, K. E.: 1.2 the NCEP stage II/IV hourly precipitation analyses: Development and applications, in 19th Conf. Hydrology, American Meteorological Society, San Diego, CA, USA., 2005.
- Lindström, G., Johansson, B., Persson, M., Gardelin, M. and Bergström, S.: Development and test of the distributed HBV-96 hydrological model, *J. Hydrol.*, 201(1–4), 272–288, 1997.
- 10 Linsley, R. K.: Flood Estimates: How Good Are They?, *Water Resour. Res.*, 22(9S), 1986.
- Luke, A., Vrugt Jasper, A., AghaKouchak, A., Matthew, R. and Sanders, B. F.: Predicting nonstationary flood frequencies: Evidence supports an updated stationarity thesis in the United States, *Water Resour. Res.*, 53(7), 5469–5494, doi:10.1002/2016WR019676, 2017.
- Machado, M. J., Botero, B. A., López, J., Francés, F., Díez-Herrero, A. and Benito, G.: Flood frequency analysis of historical flood data under stationary and non-stationary modelling, *Hydrol. Earth Syst. Sci.*, 19(6), 2561–2576, doi:10.5194/hess-19-2561-2015, 2015.
- 15 Mann, H. B.: Nonparametric tests against trend, *Econometrica*, 13, 245–259, 1945.
- Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P. and Stouffer, R. J.: Stationarity Is Dead: Whither Water Management?, *Science*, 319(5863), 573–574, doi:10.1126/science.1151915, 2008.
- 20 Nathan, R., Jordan, P., Scorch, M., Lang, S., Kuczera, G., Schaefer, M. and Weinmann, E.: Estimating the exceedance probability of extreme rainfalls up to the probable maximum precipitation, *J. Hydrol.*, 543, 706–720, doi:10.1016/j.jhydrol.2016.10.044, 2016.
- Niemeyer, R. J., Link, T. E., Heinse, R. and Seyfried, M. S.: Climate moderates potential shifts in streamflow from changes in pinyon-juniper woodland cover across the western U.S., *Hydrol. Process.*, 31(20), 3489–3503, doi:10.1002/hyp.11264, 2017.
- 25 Ntelekos, A. A., Oppenheimer, M., Smith, J. A. and Miller, A. J.: Urbanization, climate change and flood policy in the United States, *Clim. Change*, 103(3–4), 597–616, doi:10.1007/s10584-009-9789-6, 2010.
- Osuch, M., Romanowicz, R. J. and Booij, M. J.: The influence of parametric uncertainty on the relationships between HBV model parameters and climatic characteristics, *Hydrol. Sci. J.*, 60(7–8), 1299–1316, doi:10.1080/02626667.2014.967694, 2015.
- 30 Ouarda, T. B. M. J., Cunderlik, J. M., St-Hilaire, A., Barbet, M., Bruneau, P. and Bobée, B.: Data-based comparison of seasonality-based regional flood frequency methods, *J. Hydrol.*, 330(1–2), 329–339, doi:10.1016/j.jhydrol.2006.03.023, 2006.

- Park, D. and Markus, M.: Analysis of a changing hydrologic flood regime using the Variable Infiltration Capacity model, *J. Hydrol.*, 515, 267–280, doi:10.1016/j.jhydrol.2014.05.004, 2014.
- Peleg, N., Blumensaat, F., Molnar, P., Faticchi, S. and Burlando, P.: Partitioning the impacts of spatial and climatological rainfall variability in urban drainage modeling, *Hydrol. Earth Syst. Sci.*, 21(3), 1559–1572, doi:10.5194/hess-21-1559-2017, 2017.
- Potter, K. W.: Evidence for nonstationarity as a physical explanation of the Hurst Phenomenon, *Water Resour. Res.*, 12(5), 1047–1052, doi:10.1029/WR012i005p01047, 1976.
- Press, W. H., Ed.: FORTRAN numerical recipes, 2nd ed., Cambridge University Press, Cambridge [England]; New York., 1996.
- Prosdocimi, I., Kjeldsen, T. R. and Miller, J. D.: Detection and attribution of urbanization effect on flood extremes using nonstationary flood-frequency models, *Water Resour. Res.*, 51(6), 4244–4262, doi:10.1002/2015WR017065, 2015.
- Rashid, M. M., Beecham, S. and Chowdhury, R. K.: Simulation of extreme rainfall and projection of future changes using the GLIMCLIM model, *Theor. Appl. Climatol.*, 130(1–2), 453–466, doi:10.1007/s00704-016-1892-9, 2017.
- Rogger, M., Kohl, B., Pirkl, H., Viglione, A., Komma, J., Kirnbauer, R., Merz, R. and Blöschl, G.: Runoff models and flood frequency statistics for design flood estimation in Austria – Do they tell a consistent story?, *J. Hydrol.*, 456–457, 30–43, doi:10.1016/j.jhydrol.2012.05.068, 2012.
- Schilling, K. E. and Libra, R. D.: INCREASED BASEFLOW IN IOWA OVER THE SECOND HALF OF THE 20TH CENTURY, *J. Am. Water Resour. Assoc.*, 39(4), 851–860, doi:10.1111/j.1752-1688.2003.tb04410.x, 2003.
- Schilling, K. E., Gassman, P. W., Kling, C. L., Campbell, T., Jha, M. K., Wolter, C. F. and Arnold, J. G.: The potential for agricultural land use change to reduce flood risk in a large watershed, *Hydrol. Process.*, 28(8), 3314–3325, doi:10.1002/hyp.9865, 2014.
- Seibert, J.: Multi-Criteria calibration of a conceptual runoff model using a genetic algorithm, *Hydrol. Earth Syst. Sci.*, 4, 215–224, 2000.
- Seibert, J.: Reliability of model predictions outside calibration conditions, *Hydrol. Res.*, 34(5), 477–492, 2003.
- Seibert, J. and Vis, M. J. P.: Teaching hydrological modeling with a user-friendly catchment-runoff-model software package, *Hydrol. Earth Syst. Sci.*, 16(9), 3315–3325, doi:10.5194/hess-16-3315-2012, 2012.
- Sen, P. K.: Estimates of the Regression Coefficient Based on Kendall’s Tau, *J. Am. Stat. Assoc.*, 63(324), 1379, doi:10.2307/2285891, 1968.
- Serago, J. M. and Vogel, R. M.: Parsimonious nonstationary flood frequency analysis, *Adv. Water Resour.*, 112, 1–16, doi:10.1016/j.advwatres.2017.11.026, 2018.
- Sivapalan, M. and Samuel, J. M.: Transcending limitations of stationarity and the return period: process-based approach to flood estimation and risk assessment, *Hydrol. Process.*, 23(11), 1671–1675, doi:10.1002/hyp.7292, 2009.
- Smith, J. A., Villarini, G. and Baeck, M. L.: Mixture Distributions and the Hydroclimatology of Extreme Rainfall and Flooding in the Eastern United States, *J. Hydrometeorol.*, 12(2), 294–309, doi:10.1175/2010JHM1242.1, 2011.

- Smith, J. A., Baeck, M. L., Villarini, G., Wright, D. B. and Krajewski, W.: Extreme Flood Response: The June 2008 Flooding in Iowa, *J. Hydrometeorol.*, 14(6), 1810–1825, doi:10.1175/JHM-D-12-0191.1, 2013.
- Stedinger, J. R. and Griffis, V. W.: Getting from here to where? Flood frequency analysis and climate, *JAWRA J. Am. Water Resour. Assoc.*, 47(3), 506–513, 2011.
- 5 Svensson, C. and Jones, D. A.: Review of methods for deriving areal reduction factors: Review of ARF methods, *J. Flood Risk Manag.*, 3(3), 232–245, doi:10.1111/j.1753-318X.2010.01075.x, 2010.
- Viglione, A. and Blöschl, G.: On the role of storm duration in the mapping of rainfall to flood return periods, *Hydrol. Earth Syst. Sci.*, 13(2), 205–216, 2009.
- Viglione, A., Merz, R. and Blöschl, G.: On the role of the runoff coefficient in the mapping of rainfall to flood return periods, *Hydrol. Earth Syst. Sci.*, 13(5), 577–593, 2009.
- 10 Viglione, A., Chirico, G. B., Woods, R. and Blöschl, G.: Generalised synthesis of space–time variability in flood response: An analytical framework, *J. Hydrol.*, 394(1–2), 198–212, doi:10.1016/j.jhydrol.2010.05.047, 2010a.
- Viglione, A., Chirico, G. B., Komma, J., Woods, R., Borga, M. and Blöschl, G.: Quantifying space-time dynamics of flood event types, *J. Hydrol.*, 394(1–2), 213–229, doi:10.1016/j.jhydrol.2010.05.041, 2010b.
- 15 Villarini, G., Serinaldi, F., Smith, J. A. and Krajewski, W. F.: On the stationarity of annual flood peaks in the continental United States during the 20th century, *Water Resour. Res.*, 45(8), doi:10.1029/2008WR007645, 2009.
- Villarini, G., Smith, J. A., Baeck, M. L. and Krajewski, W. F.: Examining Flood Frequency Distributions in the Midwest U.S.<sup>1</sup>, *JAWRA J. Am. Water Resour. Assoc.*, 47(3), 447–463, doi:10.1111/j.1752-1688.2011.00540.x, 2011.
- Vis, M., Knight, R., Pool, S., Wolfe, W. and Seibert, J.: Model Calibration Criteria for Estimating Ecological Flow Characteristics, *Water*, 7(5), 2358–2381, doi:10.3390/w7052358, 2015.
- 20 Wilson, L. L. and Foufoula-Georgiou, E.: Regional rainfall frequency analysis via stochastic storm transposition, *J. Hydraul. Eng.*, 116(7), 859–880, 1990.
- Wright, D. B., Smith, J. A., Villarini, G. and Baeck, M. L.: Estimating the frequency of extreme rainfall using weather radar and stochastic storm transposition, *J. Hydrol.*, 488, 150–165, doi:10.1016/j.jhydrol.2013.03.003, 2013.
- 25 Wright, D. B., Smith, J. A. and Baeck, M. L.: Critical examination of area reduction factors, *J. Hydrol. Eng.*, 19(4), 769–776, 2014a.
- Wright, D. B., Smith, J. A. and Baeck, M. L.: Flood frequency analysis using radar rainfall fields and stochastic storm transposition, *Water Resour. Res.*, 50(2), 1592–1615, doi:10.1002/2013WR014224, 2014b.
- Wright, D. B., Mantilla, R. and Peters-Lidard, C. D.: A remote sensing-based tool for assessing rainfall-driven hazards, *Environ. Model. Softw.*, 90, 34–54, doi:10.1016/j.envsoft.2016.12.006, 2017.
- 30 Zhu, Z., Wright, D. B. and Yu, G.: The Impact of Rainfall Space-Time Structure in Flood Frequency Analysis, *Water Resour. Res.*, 54(11), 8983–8998, doi:10.1029/2018WR023550, 2018.