The Upper Tail of Precipitation in Convection-Permitting Regional Climate Models and Their Utility in Nonstationary Rainfall and Flood Frequency Analysis

Guo Yu, Daniel B. Wright, and Zhe Li

1Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI, USA

Abstract
Computational advances have made atmospheric modeling at convection-permitting (≤4 km) grid spacings increasingly feasible. These simulations hold great promise in the projection of climate change impacts including rainfall and flood extremes. The relatively short model runs that are currently feasible, however, inhibit the assessment of the upper tail of rainfall and flood quantiles using conventional statistical methods. Stochastic storm transposition (SST) and process-based flood frequency analysis are two approaches that together can help to mitigate this limitation. SST generates large numbers of extreme rainfall scenarios by temporal resampling and geospatial transposition of rainfall fields from relatively short data sets. Coupling SST with process-based flood frequency analysis enables exploration of flood behavior at a range of spatial and temporal scales. We apply these approaches with outputs of 13-year simulations of regional climate to examine changes in extreme rainfall and flood quantiles up to the 500-year recurrence interval in a medium-sized watershed in the Midwestern United States. Intensification of extreme precipitation across a range of spatial and temporal scales is identified in future climate; changes in flood magnitudes depend on watershed area, with small watersheds exhibiting the greatest increases due to their limited capacity to attenuate flood peaks. Flood seasonality and snowmelt are predicted to be earlier in the year under projected warming, while the most extreme floods continue to occur in early summer. Findings highlight both the potential and limitations of convection-resolving climate models to help understand possible changes in rainfall and flood frequency across watershed scales.

Plain Language Summary
High-resolution “convection-permitting” regional climate model simulations hold great promise in projection of climate change impacts including extreme rainfall and flooding. The relatively short (~10-year) model runs that are currently feasible, however, are insufficient for examining very rare events like 100-year storms and floods. Meanwhile, existing rainfall and flood data sets have a number of shortcomings that make it difficult to understand how floods have and will continue to change. In this study, we use several novel computer modeling methods to help mitigate these limitations. We apply these methods together with detailed simulations of flood hydrology and high-resolution regional climate simulation results to examine current and future extreme rainfall and flooding in an agricultural watershed in northeastern Iowa, in the Midwestern United States. Floods there are projected to become more severe, driven by complex seasonal changes in rainfall, temperature, and snow. The magnitude of these changes depends on upstream watershed area. This work demonstrates how cutting-edge climate and hydrology simulations and methods, together with flood theory and data, can help to predict future changes in flooding.

1. Introduction

Extreme precipitation and the floods they cause represent significant risk to safety and economic security, accounting for approximately 41% of worldwide natural loss events in the last four decades (Munich RE, 2018). They also play valuable roles in alleviating droughts, replenishing water supplies, ensuring food security, and ensuring floodplain biodiversity (e.g., Kam et al., 2013; Pan et al., 2013). Understanding how the likelihood and magnitude of extreme rainfall and flooding will change in future is therefore critical. The procedures used to derive the relationship between the magnitude of rainfall (flood) extremes and the corresponding likelihood are referred to as rainfall (flood) frequency analysis. Likelihood of occurrence is
expressed as annual exceedance probability ($p_e$), or as its inverse, the average recurrence interval or return period.

The statistical rainfall and flood frequency analysis approaches that are commonly used in practice involve fitting a probability distribution to a series of station-based extreme values (i.e., annual maximum rainfall intensities/depths or streamflows) and estimating the desired quantiles (i.e., $p_e$) via extrapolation. However, most statistical frequency analysis approaches assume temporal stationarity (i.e., no change over time; e.g., Cheng & AghaKouchak, 2015; Potter, 1976; Salas & Obeysekera, 2014). This assumption limits their applicability in cases of changes in climate, land use, or other drivers (e.g., Milly et al., 2008; Villarini et al., 2009). In the United States, for example, recent observed changes in rainfall extremes highlight the pressing need to update some of the existing rainfall frequency curves used in hydrologic practice (Wright et al., 2019), while urbanization has substantially modified flood frequency throughout the country (Hodgkins et al., 2019).

Global climate models (GCMs) can provide projections of future climate, but their coarse resolutions (grid spacing > 100 km) make them poorly suited for localized climate change impact studies, in part due to their inability to explicitly represent the fine-scale convection processes that generate extreme precipitation (Prein et al., 2015). The relatively recent emergence of convection-permitting regional climate models (RCMs) offer a source of climate projections at high spatial and temporal resolutions (i.e., subdaily, grid spacing $\leq$ 4 km; e.g., Coppola et al., 2020; Prein et al., 2015). Convection-permitting RCMs can be used directly to study changes in the frequency, intensity, and spatiotemporal rainfall patterns in response to atmospheric warming (e.g., Chang et al., 2016; Feng et al., 2016; Prein et al., 2017; Westra et al., 2014). However, the high computational burden of these RCMs has restricted existing simulation efforts to relatively short time periods (10–20 years). This leaves conventional rainfall and flood frequency analysis approaches, which are predicated on multidecadal data records, incompatible with such simulations.

In this study, we use an alternative framework to examine the upper tail of precipitation from convection-permitting RCMs, and use it for rainfall and flood frequency analyses at a range of spatial and temporal scales. The framework couples a bootstrap resampling technique known as stochastic storm transposition (SST; see Wright et al., 2020, and references therein) with process-based hydrologic modeling. Unlike more conventional statistical approaches, SST can be used with relative short (i.e., 10 to 20 years) gridded precipitation records to perform rainfall frequency analysis (Wright et al., 2013, 2017). When SST is coupled with distributed hydrologic models, they can estimate flood frequencies at multiple scales out to at least 500- to 1,000-year recurrence intervals (Wright et al., 2014, 2017; Yu et al., 2019). This combination of properties suggests that these methods can help to “unlock” the potential of convection-permitting RCMs in rainfall and flood frequency analyses and aid understanding potential future changes in rainfall or flood frequency. In this study, we discuss the potential of this framework and also its limitations and challenges.

The structure of the paper is as follows: Section 2 briefly describes the motivation and main research questions of the study. Section 3 describes the study region, data, RCM outputs, and hydrologic model. The framework of rainfall and flood frequency analyses is outlined in section 4. Results are presented in section 5, and related discussion follows in section 6. Section 7 includes a summary and concluding remarks.

2. Motivation and Research Questions

This study is partly motivated by Sharma et al. (2018), who reviewed the reasons why increases in extreme rainfall are not necessary resulting in corresponding increases in flood magnitudes, at least based on observational evidence. (This fact is not widely known; our experience is that even climate experts assume that flood extremes have been demonstrably worsening due to climate warming.) They conceptualized sensitivity of extreme precipitation and flooding to increases in temperature for different watershed scales (Figure 1).

Sharma et al.’s (2018) conceptual arguments notwithstanding, a great deal of uncertainty exists around the relationships between changes in flood drivers (e.g., rainfall, soil moisture, and snowpack) and flood frequency. This is because the probability distribution of floods requires modeling the joint probabilities of multiple drivers, and thus changes in flood frequency will be the complex product of changes in the joint variabilities of these drivers.
Observation-based investigation of climate-related changes in rainfall and flood properties across scales is limited by the scarcity of observations. Sharma et al. (2018), for example, argued that small watersheds are expected to show positive responses to increases in rainfall intensity (Figure 1). This argument is likely true, but the fact that there are relatively few streamflow gages on lower-order river networks in natural settings leaves us with little observational evidence to test this (Lanfear & Hirsch, 1999; Stokstad, 1999). Observations that would provide direct evidence of the joint probabilities among drivers, such as soil moisture and snowpack, are even more limited than precipitation or streamflow.

Sharma et al. (2018) suggested that future studies should focus on the complexity of relationships among the processes that generate flood extremes. The tools and methods used in this study are able to synthesize observation-based analysis with hydrologic and atmospheric process modeling advances; this synthesis merits deeper consideration for exploration of investigating potential changes in future rainfall and flood extremes and their drivers (see Wright et al., 2020, for an expanded version of this argument). This study specifically focuses on a medium-sized (4,002 km²) watershed in the Midwestern United States and attempts to answer the following questions: (1) How might extreme rainfall and flood frequency in this watershed change in the future? (2) Are these changes scale-dependent? (3) What are the roles of rainfall space-time properties and changing snowpack and soil moisture conditions in driving projected flood frequency? These questions are among the foremost future research foci identified by Sharma et al. (2018). In attempting to answer these questions, we draw attention to some of the key challenges associated with our approach and with using high-resolution RCMs to examine extreme precipitation and flooding more generally.

3. Study Region, Data, and Hydrologic Model

3.1. Study Region

This study focuses on the Turkey River watershed (4,002 km²) located in northeastern Iowa, USA (Figures 2a and 2b). The watershed is oriented northwest-southeast, ranging from 426 m above sea level (masl) in the northwest to 197 masl near the watershed outlet. The stream network is densely distributed among the watershed and there are four U.S. Geological Survey (USGS) stream gages having long-term records within the watershed (Figure 2c). According to the USGS 2016 National Land Cover Dataset, land use in the watershed is 64% cultivated crops, 13% grazing/pasture, 14% forest, 5% developed, and 4% other types including open water and barren land (Figure 2d).

3.2. RCM Simulations

This study uses outputs from two 13-year convection-permitting RCM simulations (Liu et al., 2017) performed by the National Center for Atmospheric Research (NCAR). Each simulation is performed at 4-km grid spacing over the conterminous U.S. using the Weather Research and Forecasting (WRF) model. A “control” simulation (henceforth referred to as CTRL) was nudged to reproduce the mean state and variability of
the current climate (2001–2013) within the Conterminous United States. The control run is forced using boundary conditions from 6-hr, 0.7° ERA-Interim reanalysis (Dee et al., 2011). End-of-the-century (2088–2100) climate simulations were conducted using a pseudo-global warming approach (henceforth referred to as PGW; see Kawase et al., 2009; Schär et al., 1996) in which the ERA-Interim boundary conditions were perturbed by the ensemble mean climate change deltas from fifth Climate Model Intercomparison Project (Taylor et al., 2012) under the high-end emission scenario Representative Concentration Pathway 8.5. In this study, we used eight output variables (Table S1 in the supporting information) from CTRL and PGW simulations as inputs to our process-based rainfall and flood frequency analyses framework, as detailed in section 4.

The CTRL simulation is able to reproduce the annual, seasonal, and subseasonal precipitation and surface temperature climatology in the central United States, except for a summertime (June-September) dry and warm bias (Liu et al., 2017). This bias is caused by weak synoptic-scale gradients attributed to overestimated incoming solar radiation and positive feedback of soil-atmosphere interactions (Prein et al., 2017). The PGW simulation suggests a continental-scale warming of ~3–6°C, and precipitation enhancement in most locations, but a slight decrease in the number of small-scale June-September convective storms in the U.S. Midwest (Liu et al., 2017; Prein, Liu, Ikeda, Trier, et al., 2017).

### 3.3. WRF-Hydro Hydrologic Model

The Weather Research and Forecasting Hydrological modeling system (WRF-Hydro; Gochis et al., 2018) is a distributed modeling structure that has been used in simulating flood responses around the world (e.g., Lin et al., 2018; Senatore et al., 2015; Yucel et al., 2015). Detailed physics options are summarized in Gochis et al. (2018). While there are numerous different possible configurations in WRF-Hydro, the National Water Model (NWM; National Oceanic and Atmospheric Administration, 2016) configuration, which is...
used by National Oceanic and Atmospheric Administration (NOAA) for flood forecasting, was used in this study. The NWM configuration uses two grid resolutions: a 1-km land surface model (LSM; Niu et al., 2011; Yang et al., 2011) grid and a nested 250 m terrain routing grid.

WRF-Hydro calculates energy and moisture fluxes over the LSM grid via the relevant processes (i.e., evapotranspiration, infiltration, percolation, etc.). Ponded water and soil moisture from each LSM grid are disaggregated to the higher-resolution terrain routing grid (Gochis & Chen, 2003), where hillslope dynamics (runoff and runon) occur. Overland flow and baseflow contribute to streamflow once they reach predefined “channel” grid cells. The channel routing module in the NWM configuration of WRF-Hydro uses the Muskingum-Cunge method (Cunge, 1969).

3.4. Ground Observations and Meteorological Forcing Products

Multiple data sets were used for bias-correcting the RCM precipitation and for calibrating and validating WRF-Hydro. These historical (2003–2015; comparable to CTRL period) data sets include NOAA Stage IV multisensor precipitation data (Lin, 2011), North American Land Data Assimilation System version 2 (NLDAS-2; Mitchell, 2004) hourly meteorological forcings, and outputs from the Noah LSM driven by NLDAS-2, referred to here as NLDAS-Noah (Chen et al., 1996; Ek et al., 2003). Daily measurements of soil moisture for 2015–2018 from 22 time domain reflectometry sensors across the watershed were also used to validate the model. Details of these data sets are provided in Table S1.

Four USGS stream gages (Figure 2c) are available in the study region. Their upstream watersheds range in size from 901 to 4,002 km². We henceforth refer to these watersheds by their USGS stream gage site names: Turkey River near Eldorado (USGS 05411850; 1,660 km²), above French Hollow (USGS 05412020; 2,338 km²), at Garber (USGS 05412500; 4,002 km²), and the Volga River at Littleport (USGS 05412400; 901 km²). These gages provide 15-min (or daily when 15-min data are unavailable) discharges and annual peak flows.

4. Methodology

The framework for rainfall and flood frequency analysis presented in this study combines bias correction of RCM precipitation, SST using the RainyDay software (Wright et al., 2017), and both continuous and event-based hydrologic simulation using WRF-Hydro. The approach is illustrated schematically in Figure 3 and summarized in the following subsections. Table S2 summarizes the inputs and outputs used for bias correction as well as for rainfall and flood frequency analyses.

4.1. Bias Correction Using Stochastic Ratio Rescaling

In this study, we attempted to reduce dry bias in RCM June-September rainfall using a stochastic bias correction (BC) method that accounts for differences in the mean, variance, and correlation between Stage IV and CTRL precipitation. The method, referred to as stochastic ratio rescaling, was introduced in Wright and Holman (2019), though for a different purpose; the present study is its first application to RCM bias correction. While the method pairs well with the SST approach used in this study (see section 4.3), we make no claim of its superiority over other BC methods; in section 6, we argue that all BC methods face considerable conceptual limitations. In section 5.2, we use stochastic rescaling ratios not only to correct biases but also to reveal the multiscale complexity of these biases—with important implications for flood applications of convection-permitting model outputs.

Detailed description of this method is provided in Wright and Holman (2019); its application in this study is briefly summarized here. First, the empirical distribution \( \hat{R} \) is estimated directly from the empirical distribution of rescaling ratios.

\[
\hat{R}_k^t = \frac{X_{\text{obs}}^k}{X_{\text{RCM}}^k}, \quad k = 1, 2, ... n
\]

where \( X_{\text{obs}}^k \) and \( X_{\text{RCM}}^k \) are kth-ordered values in the empirical cumulative density functions of Stage IV and RCM extreme precipitation (i.e., storm catalogs as will be described in next subsection) for a specific temporal and spatial resolution. While any suitable continuous distribution could be fitted to \( \hat{R} \). Wright and
Holman (2019) show that the two-parameter lognormal distribution is conceptually attractive to model \( \hat{R} \) since the distribution parameters can be estimated directly from the first two moments of \( X_{\text{obs}} \) and \( X_{\text{RCM}} \) and their covariance. Thus, the fitted distribution \( (R) \) is

\[
R = \ln N(\mu_R, \sigma^2_R) \tag{2}
\]

If \( X_{\text{obs}} \) and \( X_{\text{RCM}} \) are identically distributed, then \( E[R] = 1, \text{Var}[R] = 0 \), and no rescaling is performed. If bias exists, \( \text{Var}[R] \) will be nonzero and \( E[R] \) may differ from 1. We used this ratio both to characterize and correct RCM biases in the June-September rainfall:

\[
\tilde{X}_{\text{RCM}} = rX_{\text{RCM}} \tag{3}
\]

where \( r \) is a randomly sampled variate of \( R \) and \( \tilde{X}_{\text{RCM}} \) is the bias-corrected rainfall. Our decision to bias-correct only June-September rainfall is discussed in section 6.

### 4.2. WRF-Hydro Calibration and Continuous Simulation

WRF-Hydro calibration was carried out for nine sensitive parameters (Cuntz et al., 2016) using in a combination of manual and automated approaches. These parameters are horizontal and vertical soil hydraulic...
conductivities, a snowmelt exponential factor, surface ponding depth, groundwater “bucket” depth, groundwater outflow decay, surface and channel roughness, and a plant canopy wind parameter.

Manual calibration was conducted in a stepwise approach: (1) calibrate parameters that control snowpack processes; (2) calibrate parameters controlling the long-term water balance; (3) calibrate parameters that control infiltration; and (4) calibrate parameters that control the hydrograph shape at the downstream USGS stream gage. Once an acceptable parameter set was identified using the stepwise approach, we used the Dynamically Dimensioned Search (DDS; Tolson & Shoemaker, 2007) algorithm to minimize Kling-Gupta Efficiency (Gupta et al., 2009) for simulated versus observed hydrographs.

The final parameter set was validated with respect to the long-term water balance and hydrographs, flow duration curves, normalized peak flows (i.e., the simulated or observed peaks divided by the 2-year flood), soil moisture, and recession behavior (see Text SI in the supporting information). After calibration and validation, WRF-Hydro was used to perform a 13-year continuous simulation with both observed and RCM-based forcings as inputs. Simulated watershed states (e.g., soil moisture, streamflow, snowpack, and ponded water depth) were saved at daily time steps.

4.3. Stochastic Storm Transposition

SST generates realistic probabilistic rainfall scenarios via temporal resampling and spatial transposition of observed or simulated storms from the surrounding region. By substituting space for time, SST can effectively “lengthen” precipitation records or simulations. One of the main advantages of SST is that the spatio-temporal structures of observed or simulated precipitation are preserved, thus facilitating process-based flood frequency analysis. An overview and various applications of SST are reviewed in Wright et al. (2020).

RainyDay (Wright et al., 2017) is an open-source, Python-based SST software that couples SST with gridded precipitation data. The following steps describe how RainyDay is used in this study; not all of the software’s features are used:

1. We identified a 6° (longitude) by 4° (latitude) transposition domain (Figure 2b) which encompasses the Turkey River watershed. This same domain was also used in Yu et al. (2019) and, importantly for SST, extreme rainfall properties are roughly homogeneous within it. Storm catalogs were created for three durations (1, 24, and 72 hr) and two spatial scales (a single 16 km² Stage IV/RCM grid cell, and the 4,002 km² Turkey River watershed). This led to 18 storm catalogs (i.e., six for Stage IV, six for CTRL, and six for PGW), each composed of the 390 largest precipitation events with respect to that duration and scale, selected from 13-year precipitation data sets.

2. For RCM storm catalogs, any precipitation in June–September was bias corrected using stochastic rescaling ratios $R$ as described in section 4.1. Present-day and future rainfall and flood frequencies derived using bias-corrected RCM precipitation are henceforth referred to as CTRL-BC and PGW-BC, respectively.

3. RainyDay was used to generate a Poisson-distributed $k$ number of storm “arrivals” per year. RainyDay calculates the rate parameter $\lambda$ of this distribution by dividing the total number of events in the storm catalog by the number of years in the record (i.e., $\lambda = 390/13 = 30$ storms year$^{-1}$).

4. RainyDay randomly selects $k$ storms from a single storm catalog, uniformly transposes them within the transposition domain, and computes the resulting precipitation over the Turkey River watershed. The $k$ precipitation values over the watershed comprise a synthetic year of precipitation scenarios. The largest of these, in terms of basin-scale precipitation accumulation, can be understood as “rainfall annual maximum” and was used subsequently for producing rainfall frequency analysis; the largest five events were used for deriving process-based flood frequency analysis.

5. RainyDay was used to repeat Steps 3 and 4, five hundred times to create a single realization of 500 synthetic years of rainfall. This was repeated 100 (20) times for rainfall (flood) frequency analysis to facilitate examination of internal stochastic variability of the SST procedure. Steps 3–5 were repeated for all storm catalogs.

4.4. Event-Based Flood Simulation

The largest five out of the $k$ transposed storm events for each synthetic year (section 4.3) were paired with five sets of watershed initial conditions drawn from the daily continuous simulation (section 4.2), subject to two criteria: (1) all five sets of initial conditions were randomly selected from the same year of the
continuous simulation; (2) their day of year are within ±7 days of the corresponding values of the precipitation. These criteria guarantee both realistic seasonality of floods and interannual variability in watershed conditions. Each paired group constitutes an event-based WRF-Hydro simulation. The largest peak among five events represents a simulated annual peak flow. For each RainyDay-based precipitation realization, we repeated this procedure for all 500 synthetic years to produce 500 annual peak flows and calculated their annual exceedance probability by dividing its (descending) rank by number of synthetic years (500). The 20 realizations provide estimates of variability in flood frequency distributions.

5. Results

5.1. Spatial and Temporal Variability in Precipitation Bias

All June-September precipitation events from both CTRL and Stage IV catalogs were selected to estimate six different distributions of rescaling ratios corresponding to different spatial and temporal scales (Figure 4). Most rescaling ratios and all \( \mu \) are larger than 1.0 for the six scenarios (Figure 4), implying that summertime rainfall events from the CTRL storm catalogs feature a low bias consistent with Liu et al. (2017). For example, 24-hr, grid-scale rescaling ratios (Figure 4b) range from roughly 1.1 to 1.5. These biases depend on the spatial and temporal scales of the storm catalog, however. For example, \( \mu \) at the watershed scale is always larger than that for a single grid cell, and both \( \mu \) and \( \sigma \) increase with storm duration, implying that RCM-simulated convective storms have shorter duration and smaller size than Stage IV; that is, the RCM fails to accurately simulate storm space-time structure.

Randomly drawn variates from the fitted lognormal distributions in Figure 4 can be used to stochastically rescale the June-September CTRL and PGW rainfall to remove dry biases. We selected 24-hr grid-scale and 72-hr watershed-scale rescaling ratio distributions (Figures 4b and 4f, respectively) to bias correct the corresponding CTRL and PGW storm catalogs (henceforth referred to as CTRL-BC and PGW-BC, respectively). The 72-hr watershed-scale storm catalogs using CTRL-BC and PGW-BC are also used for deriving flood frequencies.
5.2. Current and Future Rainfall Frequency Analysis

The stochastic ratio rescaling reduces the dry bias in CTRL storm catalogs, shifting the cumulative distribution functions of bias-corrected catalog toward that of Stage IV (Figure 5a). Then, RainyDay was used with Stage IV, CTRL, and CTRL-BC catalogs to calculate rainfall frequency curves for \( p_e \) ranging from 0.5 to 0.002 for 24-hr, single grid and 72-hr durations, and watershed scales (Figure 5c, respectively). A gage-based rainfall frequency curve from NOAA Atlas 14 (Bonnin et al., 2016) is also plotted in Figure 5b, for comparison purposes.

RainyDay-based frequency analyses for 24-hr, single grid (Figure 5b) and for 72-hr, watershed scales (Figure 5c) show similar patterns: Across the whole range of \( p_e \), Stage IV- and CTRL-BC-based estimates of rainfall frequency closely match while estimates using CTRL rainfall exhibit systematic underestimation. In addition, both Stage IV and CTRL-BC estimates of 24-hr, single grid rainfall frequency well align with the values from Atlas 14. The underestimation using CTRL rainfall is attributable to the aforementioned dry bias. Similarities between results from CTRL-BC and Stage IV/Atlas 14 confirm that both the stochastic ratio rescaling is effective at bias correction and that SST is able to estimate the extreme tail of precipitation from relatively short gridded data sets.

For the 24-hr grid scale, rainfall in PGW-BC catalogs is mostly higher than the one in CTRL-BC catalogs (Figure 5d), and, consequently, the resulting rainfall frequency curves (Figure 5e) show consistent enhancement of rainfall intensity in the future climate across the \( p_e \). A similar pattern of increases is also identified in comparison between CTRL-BC and PGW-BC 72-hr, watershed scale rainfall frequency curves (Figure 5f).

Figure 5. (a) Cumulative distribution functions of 72-hr watershed scale precipitation using Stage IV, CTRL, and CTRL-BC storm catalogs, respectively. Comparison of rainfall frequency curves from Atlas 14 and RainyDay using the Stage IV, CTRL, and CTRL-BC storm catalogs for 24-hr single grid (b) and 72-hr watershed scales (c). (d–f) Same as panels (a)–(c) but for CTRL-BC and PGW-BC storm catalogs. Shaded areas represent the spread among 100 RainyDay realizations. Bars for Atlas 14 show 90% statistical confidence intervals.

10.1029/2020EF001613

Earth's Future
5.3. Evaluation of the Process-Based Flood Frequency Analysis

The validation of the 13-year WRF-Hydro continuous simulation is presented in the supporting information, Text S1. To summarize, WRF-Hydro is able to represent multiple important hydrologic processes and features, including high streamflows at multiple scales, long-term ET, surface runoff, and seasonally varying high flow, soil moisture, and snowpack dynamics, while low flow dynamics are lacking due to model structural shortcomings.

Process-based flood frequency curves using Stage IV precipitation and NLDAS-2 meteorology are shown in Figure 6 for Turkey River at Garber. Results are also shown using CTRL and CTRL-BC. Observed annual peak flows from 1990–2019, a period of elevated flood activity in the watershed (Wright et al., 2017), are shown (Figure 6). CTRL-based flood frequency curves show systematic underestimation except for very rare events ($p_e < 0.005$), consistent with dry biases in CTRL storm catalogs shown in Figure 5a. Stage IV- and CTRL-BC-based results agree well with recent observed flood peaks across all $p_e$, except for very rare event ($p_e < 0.005$), where CTRL-BC shows higher estimates. These higher estimates cannot necessarily be attributed to some inadequacy of CTRL-BC; sampling variability arising from limited data can play a large role in SST, just as in any other type of extreme value (Wright et al., 2014). The differences between CTRL and CTRL-BC results in Figure 6 demonstrate the importance of June-September rainfall in shaping flood frequencies, as CTRL-BC results are closer to both observations and Stage IV-based results.

We examined 100-year simulated peak flows and corresponding top layer (0–40 cm) antecedent volumetric soil moisture values, 72-hr total precipitation depths ($P_{total}$) and maximum hourly precipitation intensities ($P_{max}$), as a function of watershed area using Stage IV- and CTRL-BC-based inputs (Figure 7). CTRL-BC-based 100-year peak flows are higher than the Stage IV-based estimations for small watersheds ($\leq$100 km$^2$) but are similar for larger watershed areas (Figure 7a). The higher 100-year peak flows for small watersheds using CTRL-BC precipitation are mainly due to the higher $P_{max}$ (Figure 7d) rather than antecedent soil moisture (Figure 7b) or $P_{total}$ (Figure 7c). This is because the stochastic ratio rescaling used to remove the dry bias in $P_{total}$ (Figures 7c and 5c), does so at the expense of overcorrecting $P_{max}$ (Figure 7d) due to the scale-dependent bias in CTRL precipitation. The implications of scale-dependent biases in RCM simulations are discussed in section 6.

5.4. Projected Changes in Flood Frequency

5.4.1. Projected Changes in Forcing and Water Balance

We compared watershed-averaged monthly values of eight RCM output variables between CTRL-BC and PGW-BC (Figure S1). All eight outputs (Row 5 of Table S1) are used subsequently as forcings for WRF-Hydro. Incoming longwave radiation, air temperature, and specific humidity show clear increases in PGW-BC relative to CTRL-BC. Monthly precipitation in PGW-BC is slightly higher than in CTRL-BC, with the exception of July and August, which are drier.

We ran 13-year WRF-Hydro continuous simulations with CTRL-BC and PGW-BC forcings. Like the input forcings, simulated monthly soil moisture, ET, and runoff show consistent intensification (Figure S3). The drier and warmer summers in PGW-BC are attributed to the effects of both climate change signals and known model biases (Prein, Liu, Ikeda, Trier, et al., 2017). In addition, simulated SWE in the future shows less accumulation and earlier melt, in line with the increased height of freezing level (Ashley et al., 2020; Prein, Liu, Ikeda, Bullock, et al., 2017) and warmer air temperatures (Figure S4c).

5.4.2. Scale-Dependent Changes in Flood Frequency

Unlike rainfall frequency, which show increases across all $p_e$ in PGW-BC relative to CTRL-BC (Figures 5e and 5f), flood frequency for Turkey River at Garber using PGW-BC yielded significantly larger values for
pe > 0.005, and similar values for rarer pe compared with CTRL-BC results (Figure 8a). Differences in the upper tails of these distributions are likely tied to sampling errors (see section 6). Discrepancies between projected changes in rainfall and flood frequencies can be tied to two related points: (1) the annual rainfall maxima which comprise rainfall frequency curves do not necessarily produce annual peak flows (e.g., Yu et al., 2019), and (2) changes in rainfall extremes are not enough to explain changes in flood frequency. Instead, factors such as snowpack, snowmelt, and soil moisture play important roles.

We also compared changes in 100-year flood magnitudes across the watershed (Figure 8b). PGW-BC-based results showed large increases (>30%) over CTRL-BC in the first- and second-order streams, moderate increases (15–30%) in third- to fourth-order streams, and smaller changes downstream in the main stem of Turkey River (fifth to sixth order). These changes are roughly consistent with the conceptual interpretation in Sharma et al. (2018), and are discussed in section 6.

**5.4.3. Investigation of Changes in Flood Drivers**

PGW-BC and CTRL-BC-based flood seasonalities show a shift toward earlier floods in the future climate, though the dominant flood season remains May-June (Figure 9a). This shift is mainly due to an earlier snowmelt season (Figure 9a), caused by warmer air temperatures in February and March (Figure S2). Although
there are no simulated flood events in December from CTRL-BC-based simulations. December flood events exist in the future climate, due to December precipitation falling as rain in the future climate (Figure S5c).

Large increases (25–50%) were found for mean annual rainfall-driven floods (i.e., no snow present), especially for small watershed scales (≤10 km²; Figure 11b). While the number of simulated snowmelt and rain-on-snow flood events decreased by 35% in PGW-BC, snowmelt rate increased for many of the remaining events, mainly due to higher spring-time temperature and incoming longwave radiation under PGW (Figure S2f). This intensified snowmelt rate leads to increases of roughly 5–30% in the mean annual snowmelt-affected floods (Figure 9b).

We also calculated changes in mean flood magnitude grouped by recurrence interval (2–10, 10–100, and 100–500 years; Figure 10a). For small scales (A ≤ 100 km²), increases vary from 20% to 80% depending on the group and rarer events (e.g., 100–500 years) show the greatest increases, consistent with Sharma et al.’s (2018) theoretical arguments. Small watersheds (A < 100 km²) show greater increases, which can be attributed to less storage capacity and also potentially smaller but more intense convective rain cells in future (e.g., Patricola & Wehner, 2018; Prein, Liu, Ikeda, Trier, et al., 2017). Large watersheds have more soil moisture, surface retention, and channel storage capacity—that can attenuate flood peaks. Rainfall partial coverage (Marco & Valdés, 1998; Zhu et al., 2018) is also likely a factor driving smaller increases in larger watersheds, though we do not examine it here. We also examined associated changes in $P_{\text{total}}$, $P_{\text{max}}$, and top layer antecedent soil moisture (Figure 10b). Changes in antecedent soil moisture are relatively limited, meaning precipitation change is the main driver of increases in 10–100-year flood magnitudes—of ~40% in watersheds smaller than about 100 km², less in larger watersheds.

We reproduced results in Figures 10a and 10b without using precipitation bias correction (Figures 10c and 10d). The results without BC show larger variability in flood magnitude (Figure 10c), attributable to the variability of changing precipitation (Figure 10d); decreases in 100–500-year floods for large watersheds (A > 1,000 km²) are mainly due to the largest...
precipitation event in CTRL storm catalog being larger than the largest event in the PGW catalog; this is not the case when bias correction is used. The general conclusions are nonetheless consistent with or without BC: projected changes in flood magnitude depend on recurrence interval and watershed scale (Figures 10a and 10c) and are driven by changes in $P_{total}$ and $P_{max}$ rather than antecedent soil moisture (Figures 10b and 10d).

To better understand the role of soil moisture in flood response, we plot its simulated annual cycle from the 13-year CTRL-BC and PGW-BC continuous simulations (Figure 11). The 2,500 precipitation events from one PGW-BC process-based flood frequency analysis and the resulting peak flows are also shown (Figures 11a and 11b, respectively). While soil moisture tends to be considerably drier in July-October in PGW-BC than in CTRL-BC, it remains roughly unchanged in May-June, the dominant flood season in the region. While not entirely conclusive, this, together with Figure 10 suggests that projected increases in flood-season rainfall will not be offset by increased soil storage due to drier soils, raising overall levels of flood hazard.

6. Discussion
6.1. Methodological Assumptions
This study examines current and future rainfall and flood frequencies via a combination of convection-permitting RCM outputs, SST, and process-based hydrologic simulation. Here, we highlight three limitations of our study, all of which likely pervade other attempts to use convection-permitting RCM projections for the study of extreme events and their consequences.

First, the high computational burden of convection-permitting RCMs presently precludes multidecadal simulations of multiple ensemble
Table 1

| Simulated/expected changes | Hydrological reasoning | Consistency |
|---------------------------|-----------------------|-------------|
| This study                | Flood quantiles show larger increases in small watersheds than in larger watersheds. | Larger watersheds are less sensitive to changes in extreme precipitation than smaller watersheds due to the greater modulating effects of watershed and channel storage; convective rain cells are more intense but spatially smaller in PGW. | Both studies’ findings and supporting reasons are consistent. |
| Sharma et al. (2018)      | “Discharges increase more for smaller catchments.” | “There is an increased likelihood that a storm will cover the entire catchment and hence lead to soil moisture saturation, with more of the precipitation contributing to the streamflow response.” | Both studies’ findings and supporting reasons are consistent. |
| This study                | For watersheds smaller than 1,000 km², larger increases are identified in rarer flood events. | Peak hourly precipitation intensity (P_max) determines flood magnitudes for small watersheds and increases substantially in future climate scenario. | Both studies’ findings and supporting reasons are consistent. |
| Sharma et al. (2018)      | “The rarer a flood, the more it will increase.” | “The more extreme an event is, the greater is the precipitation intensity and the more likely the catchment is to become saturated, with a greater proportion of the (subsequent) precipitation contributing to streamflow.” | Both studies’ findings and supporting reasons are consistent. |
| This study                | For watersheds larger than 1,000 km², the 100–500 year floods show the smallest increases (or decreases in results without BC). | Event precipitation accumulation (P_total) determines flood magnitudes for larger watersheds, and P_total of watershed scale decreases in future due to sampling error (i.e., only one RCM is used). | Similar findings for both studies but for different reasons. |
| Sharma et al. (2018)      | “Precipitation increases while flow decreases.” | “Changes in precipitation cannot be simply translated to flooding as other factors are dominating.” | Differences between the studies’ findings highlight the key role of seasonality. |
| This study                | Simulated decreases in future soil moisture occur in certain months, and play a relatively minor role in driving flood frequency. | The period of drier soil moisture does not overlap the dominant flood season in the region. | Differences between the studies’ findings highlight the key role of seasonality. |
| Sharma et al. (2018)      | “Drying soil moisture conditions will reduce flood magnitudes.” | “Some precipitation contributes to increasing the soil moisture, with the remaining precipitation contributing to streamflow.” | Differences between the studies’ findings highlight the key role of seasonality. |

Our results reflect this. For example, the largest rainfall event in CTRL has a higher rainfall accumulation than the largest in PGW, which is translated in Figure 10c to an apparent decrease in the very large rainfall and flood quantiles, at large watershed scales. It is likely, however, that this is the product of the relatively short (13-year) simulation periods and the resulting sampling errors in extreme events, rather than a true climate signal. The regional storm catalog approach of SST is only partially able to “make up” for such sampling errors, since the upper tail of the SST-derived rainfall scenarios is driven by the largest one or several events in each storm catalog and their subsequent transposition (Wright et al., 2020). The reliability of this upper bound depends on whether the storm catalog adequately represents the true population of extreme rainstorms. This issue of sampling error is not unique to SST—no matter which analysis technique is chosen, the upper tail of extreme event frequencies is governed by the largest few events in the observed or simulated record (Wright et al., 2014). With computational advances that enable longer convection-permitting RCM simulations, or multiple simulations with multiple GCMs, this sampling error can be alleviated.

Second, the PGW approach used in the RCM simulations perturbs the large-scale thermodynamics (e.g., North American Osculation and planetary waves) but leaves the weather patterns (i.e., mesoscale circulations) entering the boundary unchanged, thus inhibiting the representation of systematic changes in members, including those forced by different GCMs. Ensemble approaches have generally been considered an important part of climate change impact studies, particularly with respect to uncertainty estimation (Alfieri et al., 2015; Wehner, 2013; Wiel et al., 2019). Thus, while our derived rainfall and flood frequency curves (e.g., Figures 5 and 10) display uncertainty associated with internal variability of the SST procedure, they neglect uncertainty in large-scale climate forcings, RCM structural uncertainty, and interannual (i.e., low-frequency) climate variability (e.g., Mehrotra & Sharma, 2007; Rocheta et al., 2017).
storm tracks and their frequency. According to Rasmussen et al. (2011), the PGW approach should be considered as an “imposed” warming experiment rather than a robust prediction of future climate. Therefore, a more complete understanding of changes in future extreme rainfall and floods frequency needs a larger number of RCM runs forced with boundary conditions from individual GCMs. These runs would need to consider biases in the interannual variability of lateral boundary conditions (see Mehrotra & Sharma, 2016, and Rocheta et al., 2017, who provided a bias correction method for addressing these issues).

Third, while bias correction (BC) of climate model outputs is generally considered indispensable in climate change impact studies, Ehret et al. (2012) and others have argued that BC is generally used in an invalid way: it is used to modify one or more variables without consideration of the implications for model consistency (i.e., conservation of mass, energy, etc.) and uncertainties in observation. To date, development and application of BC has been driven by a real or perceived necessity to provide assessments of potential impacts, rather than by validity of the methods themselves (Vannitsem, 2011). With this in mind, we decided to bias-corrected only the most egregious biases (i.e., a previously documented summertime dry bias in Midwest rainfall). Though our results demonstrate its feasibility (Figures 5b, 5c, 6a, and S6), they also suggest that BC for convection-permitting RCMs may be even more problematic than for lower-resolution models, since these biases can vary over relatively small spatial and temporal scales (Figure 4). This issue is nontrivial since flood response to precipitation is highly scale-dependent and has received relatively little attention because earlier coarser-resolution RCMs, with their much larger grid cells, depict little, or no intra-storm structure. It is not obvious how to correct multiscale biases in storm space-time structure, though the work of Chang et al. (2016) provides some possible guidance.

6.2. Comparison of Simulations and Conceptual Arguments

While trends in floods and the individual processes that cause them have been relatively well-studied, how these trends translate into shifts in flood frequency has received much less attention. We believe that recent advances in process-based hydrologic modeling and precipitation measurement and simulation (including weather radar and convection-permitting RCMs) present opportunities for understanding the future of flood frequency, particularly when used in conjunction with more traditional data sources and hydrologic reasoning (Wright et al., 2020).

In section 2, we summarized the causal arguments of Sharma et al. (2018) that helped to motivate this study; Table 1 provides a brief summary of how our findings compare with those arguments. We mostly confirm their reasoning. We found, however, complicated roles of antecedent soil moisture and snowpack dynamics, which illustrates that the future of flood frequency is likely to be complex and, though not studied here, geographically varied.

7. Summary and Conclusions

In this study, we used storm catalogs, SST, and a stochastic bias correction approach to assess the upper tail of precipitation simulated by a convection-permitting RCM. We used these tools to perform rainfall frequency analysis, and, combined with the WRF-Hydro hydrologic model, to create a process-based framework for flood frequency analysis. These frequency analyses were performed using observations as well as using current- and future-climate RCM outputs. This framework facilitates the assessment of extreme rainfall or flood quantiles using relatively short (1–2 decades) RCM runs and observational records. We apply the framework to Turkey River, a 4,002 km² agricultural watershed in the Midwestern United States and changes in the rainfall and flood extremes in watersheds with different climate regimes may vary. Key findings include the following:

1. SST is able to examine the extreme tail of simulated precipitation (e.g., the 100-year recurrence interval and beyond), despite the short (13-year) RCM simulation periods. Combined with stochastic bias correction, SST-based rainfall frequency curves using present-day RCM outputs closely match results from both SST combined with gage-adjusted weather radar and NOAA Atlas 14, a well-known published source (estimates) of rainfall frequencies. Results using outputs from the future climate RCM show extreme rainfall enhancement at both single-grid, 24-hr and watershed, 72-hr scales.

2. Summertime extreme rainfall in the present-day RCM control run exhibits a considerable dry bias due to the underestimated number of convective storms. The most extreme rainfall events in control run and
Stage IV are used to create stochastic rescaling ratios and to reveal the differences in their spatiotemporal structures. While we present a stochastic bias correction technique that can adjust precipitation at a particular space-time scale, this will not necessarily correct bias at other scales, and may indeed exacerbate them. Scale-dependent bias is problematic for many hydrologic applications including flooding, due to the complexity of hydrologic responses to upstream rainfall-runoff processes.

3. We demonstrate a sequential “process-oriented” approach to calibrate and validate the WRF-Hydro model. This included both manual and automated calibration techniques and targeted multiple aspects of model performance, including long-term water balances, hydrographs, peak flows, seasonality of streamflow, soil moisture and SWE, flow duration curve, and recession behavior. With the exception of relatively poor performance for very low flows, the calibrated WRF-Hydro model was able to adequately simulate this range of processes. This lends confidence to our process-based flood frequency analysis approach, which requires accurate simulation of a range of hydrologic processes.

4. Projected increases in rainfall quantiles did not necessarily translate to increases in flood quantiles. Rather, floods across a range of recurrence intervals were predicted to increase, by roughly 40% in small watersheds, with somewhat smaller increases (and in few cases, small decreases) in larger watersheds. These changes were driven primarily by increases in late spring and summertime rainfall extremes, while soil moisture remained relatively unchanged during the flooding season. The importance of snowmelt and rain-on-snow floods is projected to diminish due to lower snow accumulation and earlier springtime melt. Scale dependencies are related to the lower storage capacity of small watersheds to “buffer” rainfall extremes,” the channel attenuation of flood waves at larger watershed scales, and potentially the impact of more intense but spatially smaller convective rain cells in future (e.g., Patricola & Wehner, 2018; Prein, Liu, Ikeda, Trier, et al., 2017).

5. SST combined with process-based flood frequency analysis provides an alternative to conceptual (e.g., Sharma et al., 2018) and observational analysis (see Hodgkins et al., 2017, for a thorough summary) of how rainfall and flood extremes are and will change in a warming climate. This alternative is attractive for several reasons. First, it provides estimates of rainfall and flood frequencies, rather than trends or qualitative statements. Such frequencies are central to infrastructure design, floodplain mapping, and risk management. Second, unlike trend analyses of precipitation or streamflow it allows us to consider the potentially important roles of other hydrometeorologic processes. Third, it allows us to examine future outcomes based on the best available climate modeling capabilities—something that neither observational trend analyses nor conceptual arguments can accomplish.

More decadal-scale convection-permitting RCM simulations using different models and forcings are needed to improve our understanding of climate internal and model variability and their impacts, flooding included, at local scales. Despite the uncertainties and limitations of the presented framework and results, our process-based analysis does further advance understanding of how future rainfall and flood hazards will change and why. It also points to some of the underlying challenges that must be addressed or, at the very least acknowledged, as climate impacts analysts begin to turn to convection-permitting RCMs for answers.

Data Availability Statement
The NCAR’s RCM data are available at this site (https://rda.ucar.edu/datasets/ds612.0). NLDAS-2 forcings and NLDAS-Noah simulation are available at Goddard Earth Sciences Data and Information Services Center (https://disc.gsfc.nasa.gov/datasets?keywords=NLDAS). USGS streamflow data are available at this site (https://maps.waterdata.usgs.gov/mapper/index.html). The results presented in this study are available upon request.

References
Alfieri, L., Feyen, L., Dottori, F., & Bianchi, A. (2015). Ensemble flood risk assessment in Europe under high end climate scenarios. Global Environmental Change, 35, 199–212. https://doi.org/10.1016/j.gloenvcha.2015.09.004
Ashley, W. S., Haberlie, A. M., & Gensini, V. A. (2020). Reduced frequency and size of late-twenty-first-century snowstorms over North America. Nature Climate Change, 10(6), 539–544. https://doi.org/10.1038/s41558-020-0774-4
Bonnin, G. M., Martin, D., Lin, B., Parzybok, T., Yekta, M., & Riley, D. (2016). NOAA atlas 14: Precipitation-frequency atlas of the United States (Vol. 2). Silver Spring, MD: National Weather Service, National Oceanic and Atmospheric Administration.
Chang, W., Stein, M. L., Wang, J., Kotamarthi, V. R., & Moyer, E. J. (2016). Changes in spatiotemporal precipitation patterns in changing climate conditions. Journal of Climate, 29(23), 8355–8376. https://doi.org/10.1175/JCLI-D-15-0844.3
Chen, F., Mitchell, K., Schaake, J., Xue, Y., Pan, H.-L., Koren, V., et al. (1996). Modeling of land surface evaporation by four schemes and comparison with FIFE observations. *Journal of Geophysical Research, 101*(D3), 7255–7268. https://doi.org/10.1029/95JD02165

Cheng, L., & AghaKouchak, A. (2015). Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate. *Scientific Reports, 4*, 7093. https://doi.org/10.1038/srep07093

Coppola, E., Sobolowski, S., Pichelli, E., Raffaele, F., Ahrens, B., Anders, I., et al. (2020). A first-of-its-kind multi-model convection-permitting ensemble for investigating convective phenomena over Europe and the Mediterranean. *Climate Dynamics, 55*(1–2), 3–34. https://doi.org/10.1007/s00382-018-4521-8

Cunge, J. A. (1969). On the subject of a flood propagation computation method (Muskingum method). *Journal of Hydraulic Research, 7*(2), 205–230. https://doi.org/10.1080/00221686909500264

Cunnane, C. (1978). Unbiased plotting positions—A review. *Journal of Hydrology, 37*(3–4), 205–222. https://doi.org/10.1016/0022-1694(78)90017-3

Cuntz, M., Mai, J., Samaniego, L., Clark, M., Wulfmeyer, V., Branch, O., et al. (2016). The impact of standard and hard-coded parameters on the hydrologic fluxes in the Noah-MP land surface model. *Journal of Geophysical Research: Atmospheres, 121*, 10,676–10,700. https://doi.org/10.1002/2016JD025097

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society, 137*(656), 553–597. https://doi.org/10.1002/qj.828

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., & Liebert, J. (2012). HESS opinions “Should we apply bias correction to global and regional climate model data?” *Hydrology and Earth System Sciences, 16*(9), 3391–3404. https://doi.org/10.5194/hess-16-3,391-2012

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grummann, P., Koren, V., et al. (2003). Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscaleEta model. *Journal of Geophysical Research, 108*(D22), 8851. https://doi.org/10.1029/2002JD002396

Feng, Z., Leung, L. R., Hagos, S., Houze, R. A., Burleyson, C. D., & Balaguru, K. (2016). More frequent intense and long-lived storms dominate the springtime trend in central US rainfall. *Nature Communications, 7*(1), 13429. https://doi.org/10.1038/ncomms13429

Goichis, D., Barlage, M., Dugger, A., FitzGerald, K., Karsten, L., McAllister, M., et al. (2018). The WRF-hydro modeling system technical description, (Version 5.0). UCAR/NCAR.

Goichis, D. J., & Chen, F. (2003). Hydrological enhancements to the Community Noah Land Surface Model (No. NCAR/TN-454+STR).

University Corporation for Atmospheric Research. https://doi.org/10.5065/D6DP00X0

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modeling. *Journal of Hydrology, 377*(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2008.09.003

Hodgkins, G. A., Dudley, R. W., Archfield, S. A., & Renard, B. (2019). Effects of climate, regulation, and urbanization on historical flood trends in the United States. *Journal of Hydrology, 573*, 697–709. https://doi.org/10.1016/j.jhydrol.2019.03.102

Hodgkins, G. A., Whitfield, P. H., Burn, D. H., Hannaford, J., Renard, B., Stahl, R., et al. (2017). Climate-driven variability in the occurrence of major floods across North America and Europe. *Journal of Hydrology, 552*, 704–717. https://doi.org/10.1016/j.jhydrol.2017.07.027

Kam, J., Sheffield, J., Yuan, X., & Wood, E. F. (2013). The influence of Atlantic tropical cyclones on drought over the Eastern United States (1980–2007). *Journal of Climate, 26*(10), 3067–3086. https://doi.org/10.1175/JCLI-D-12-00244.1

Kawase, H., Yoshikane, T., Hara, M., Kimura, F., Yasunari, T., Aloikun, B., et al. (2009). Intermodel variability of future changes in the Baiu rainfall estimated by the pseudo global warming downscaling method. *Journal of Geophysical Research, 114*, D24110. https://doi.org/10.1029/2009JD011803

Landear, K. J., & Hirsch, R. M. (1999). USGS study reveals a decline in long-record streamgages. *EOS Transactions of the American Geophysical Union, 80*(50), 605. https://doi.org/10.1029/99EO00406

Lin, P., Hopper, L. J., Yang, Z.-L., Lenz, M., & Zeitler, J. W. (2018). Insights into hydrometeorological factors constraining flood prediction skill during the May and October 2015 Texas Hill Country flood events. *Journal of Hydrometeorology, 19*(8), 1,339–1,361. https://doi.org/10.1175/JHM-D-18-0038.1

Lin, Y. (2011). GCP/IP: O/P surface: Precipitation, NCEP/EMC 4KM gridded data (GRIB) stage IV data, version 1.0. UCAR/NCAR-Earth Observing Laboratory. https://doi.org/10.5065/D6PG1DD

Liu, C., Ikeda, K., Rasmussen, R., Barlage, M., Newman, A. J., Prein, A. F., et al. (2017). Continental-scale convection-permitting modeling of the current and future climate of North America. *Climate Dynamics, 49*(1–2), 71–95. https://doi.org/10.1007/s00382-016-3,327-9

Marco, J. B., & Valdés, J. B. (1998). Partial area coverage distribution for flood frequency analysis in arid regions. *Water Resources Research, 34*(9), 2309–2317. https://doi.org/10.1029/98WR01246

Mehrotra, R., & Shamma, A. (2007). Preserving low-frequency variability in generated daily rainfall sequences. *Journal of Hydrology, 345*(1–2), 102–120. https://doi.org/10.1016/j.jhydrol.2007.08.003

Mehrotra, R., & Shamma, A. (2016). A multivariate quantile-matching bias correction approach with auto- and cross-dependence across multiple time scales: Implications for downscaling. *Journal of Climate, 29*(10), 3519–3539. https://doi.org/10.1175/JCLI-D-15-0356.1

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Stationarity is dead: Whither water management? *Science, 319*(5863), 573–574. https://doi.org/10.1126/science.1151915

Mitchell, K. E. (2004). The multi-parameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research, 116*, D12109. https://doi.org/10.1029/2010JD015139

Pan, M., Yuan, X., & Wood, E. F. (2013). A probabilistic framework for assessing drought recovery. *Geophysical Research Letters, 40*, 3637–3642. https://doi.org/10.1002/grl.50728

Patricola, C. M., & Wehner, M. F. (2018). Anthropogenic influences on major tropical cyclone events. *Nature, 565*(7731), 339–346. https://doi.org/10.1038/s41586-018-0673-2

Potter, K. W. (1976). Evidence for nonstationarity as a physical explanation of the Hurst phenomenon. *Water Resources Research, 12*(5), 1047–1052. https://doi.org/10.1029/WR012i005p01047
Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergens, K., et al. (2015). A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. Reviews of Geophysics, 53(2), 323–361. https://doi.org/10.1002/2014RG000475

Prein, A. F., Liu, C., Ikeda, K., Bullock, R., Rasmussen, R. M., Holland, G. J., & Clark, M. (2017). Simulating North American mesoscale convective systems with a convection-permitting climate model. Climate Dynamics. https://doi.org/10.1007/s00382-017-3993-2

Prein, A. F., Liu, C., Ikeda, K., Trier, S. B., Rasmussen, R. M., Holland, G. J., & Clark, M. P. (2017). Increased rainfall volume from future convective storms in the US. Nature Climate Change, 7(12), 880–884. https://doi.org/10.1038/s41558-017-0007-7

Rasmussen, R., Liu, C., Ikeda, K., Gochis, D., Yates, D., Chen, F., et al. (2011). High-resolution coupled climate runoff simulations of seasonal snowfall over Colorado: A process study of current and warmer climate. Journal of Climate, 24(12), 3015–3046. https://doi.org/10.1175/2010JCLI3985.1

Rocheta, E., Evans, J. P., & Sharma, A. (2017). Can bias correction of regional climate model lateral boundary conditions improve low-frequency rainfall variability? Journal of Climate, 30(24), 9785–9806. https://doi.org/10.1175/JCLI-D-16-0654.1

Salas, J. D., & Obeysekera, J. (2014). Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. Journal of Hydrologic Engineering, 19(3), 554–568. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000820

Schär, C., Frei, C., Lüthi, D., & Davies, H. C. (1996). Surrogate climate-change scenarios for regional climate models. Geophysical Research Letters, 23(6), 669–672. https://doi.org/10.1029/96GL00265

Senatore, A., Mendicino, G., Gochis, D. J., Yu, W., Yates, D. N., & Kunstmann, H. (2015). Fully coupled atmosphere-hydrology simulations for the central Mediterranean: Impact of enhanced hydrological parameterization for short and long time scales. Journal of Advances in Modeling Earth Systems, 7(4), 1693–1715. https://doi.org/10.1175/2015JMA0051001

Sharma, A., Wasko, C., & Lettenmaier, D. P. (2018). If precipitation extremes are increasing, why are not flood peaks? Water Resources Research, 54, 8545–8551. https://doi.org/10.1029/2018WR023749

Stokstad, E. (1999). Scarcity of rain, stream gages threatens forecasts. Science, 285(5431), 1199–1200. https://doi.org/10.1126/science.285.5431.1199

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. Bulletin of the American Meteorological Society, 93(4), 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Tolson, B. A. & Shoemaker, C. A. (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration: Dynamically dimensioned search algorithm. Water Resources Research, 43(1). https://doi.org/10.1029/2005WR004723

Vannitsem, S. (2011). Bias correction and post-processing under climate change. Nonlinear Processes in Geophysics, 18(6), 911–924. https://doi.org/10.5194/npg-18-911-2011

Villarini, G., Serinaldi, F., Smith, J. A., & Krajewski, W. F. (2009). On the stationarity of annual flood peaks in the continental United States during the 20th century. Water Resources Research, 45(8). https://doi.org/10.1029/2008WR007645

Wehner, M. F. (2013). Very extreme seasonal precipitation in the NARCCAP ensemble: model performance and projections. Climate Dynamics, 40(1–2), 59–80. https://doi.org/10.1007/s00382-012-1393-1

Westra, S., Fowler, H. J., Evans, J. F., Alexander, L. V., Berg, P., Johnson, F., et al. (2014). Future changes to the intensity and frequency of short-duration extreme rainfall. Reviews of Geophysics, 52(3), 522–555. https://doi.org/10.1002/2014RG000464

Wiel, K., Wanders, N., Selten, F. M., & Bierkens, M. F. P. (2019). Added value of large ensemble simulations for assessing extreme river flooding: Demonstrations, prospects, and challenges. Water Resources Research, 55(4), 8144–8153. https://doi.org/10.1029/2019WR023823

Wright, D. B., Holman, K. D. (2019). Rescaling transposed extreme rainfall within a heterogeneous region. Journal of Hydrologic Engineering, 24(6), 06019001. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001781

Wright, D. B., Mannita, R., & Peters-Lidard, C. D. (2017). A remote sensing-based tool for assessing rainfall-driven hazards. Environmental Modelling and Software, 90, 34–54. https://doi.org/10.1016/j.envsoft.2016.12.006

Wright, D. B., Smith, J. A., & Baech, M. L. (2014). Flood frequency analysis using radar rainfall fields and stochastic storm transposition. Water Resources Research, 50, 1592–1615. https://doi.org/10.1002/2013WR014224

Wright, D. B., Smith, J. A., Villarini, G., & Baech, M. L. (2013). Estimating the frequency of extreme rainfall using weather radar and stochastic storm transposition. Journal of Hydrology, 488, 150–165. https://doi.org/10.1016/j.jhydrol.2013.03.003

Wright, D. B., Yu, G., & England, J. F. (2020). Six decades of rainfall and flood frequency analysis using stochastic storm transposition: Review, progress, and prospects. Journal of Hydrology, 124, 816. https://doi.org/10.1016/j.jhydrol.2020.124816

Yang, Z.-L., Niu, G.-Y., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins. Journal of Geophysical Research, 116, D12110. https://doi.org/10.1029/2010JD015140

Yu, G., Wright, D. B., Zhu, Z., Smith, C., & Holman, K. D. (2019). Process-based flood frequency analysis in an agricultural watershed exhibiting nonstationary flood seasonality. Hydrology and Earth System Sciences, 23(5), 2225–2243. https://doi.org/10.5194/hess-23-2225-2019

Yucel, I., Onen, A., Yilmaz, K. K., & Gochis, D. J. (2015). Calibration and evaluation of a flood forecasting system: Utility of numerical weather prediction model, data assimilation and satellite-based rainfall. Journal of Hydrology, 523, 49–66. https://doi.org/10.1016/j.jhydrol.2015.01.042

Zhu, Z., Wright, D. B., & Yu, G. (2018). The impact of rainfall space–time structure in flood frequency analysis. Water Resources Research, 54(11), 8983–8998. https://doi.org/10.1002/2018WR023550