A Process-Based Approach to Bottom-Up Climate Risk Assessments: Developing a Statewide, Weather-Regime based Stochastic Weather Generator for California

Final Report

Nasser Najibi
Scott Steinschneider

Biological and Environmental Engineering
Cornell University
Ithaca, NY

August 2023

Table of Contents

Executive Summary

Water resource planners in California must preparefor the increased stressof climate change. However, there is significant uncertainty about how the climate will evolve over the coming decades. While most climate projections agree that California's future will be warmer and precipitation will intensify, the rate of these changes is less clear, as are more nuanced climate changes like shifts in average precipitation or changes to atmospheric circulation that can impact regional weather. Furthermore, water systems in California are very vulnerable to natural swings in climate unrelated to climate change, and the range of this natural variability must also be considered in future planning efforts.

This report documents the development of astochastic weather generator model for California that can create large ensembles of climate traces to support water resources planning under future climate uncertainty. The weather generator is an efficient tool that can quickly create long (1000-year) traces of statewide weather, and allows water managers to flexibly develop and explore climate scenarios associated with different signals of climate change. These includesignals directly related to warming (i.e., thermodynamic changes such as increased temperatures and more intense precipitation linked to an increase in the moisture-holding capacity of warmer air), as well as more nuanced signals related to shifts in atmospheric circulation (i.e., dynamic climate change). This report demonstrates that the weather generator can reproduce the climate across California very well, based on a large suite of performance metrics that include extreme precipitation events and droughts across spatial and temporal scales. The model also compares well against state-of-the-science downscaled global climate model (GCM) simulations. In addition to the model itself, deliverables from this work include a publicly available dataset of 30 unique climate change scenarios, each consisting of 1000 years of simulated climate data (precipitation, maximum temperature, minimum temperature) at a ~6 km resolution across the entire state of California. The 30 scenarios represent a range of plausible climate changes to temperature, average precipitation, and precipitation extremes. A parallel dataset of 30 climate change scenarios, each consisting of 100 years of climate data at the same resolution, is also produced by perturbing the historical meteorological record.

The datasets created in this work highlight how even in the absence of climate change, water managers in California should plan for extreme precipitation events and droughts beyond the worst case from the historical record, because such events are quite plausible due to California's natural climate variability alone. However, when such natural extremes are combined with the effects of anthropogenic climate change,extremes in California become severe and will likely require significant investment in water resources systems to sustain adequate water services across the state.The weather generator and associated datasets should be viewed as a complementary tool to more commonly available downscaled GCM projections. Unlike GCMs, the weather generator is not designed to create scenarios of future climate based on the physical laws of the Earth system and future greenhouse gas emission scenarios. Rather, the model provides away to help translate various signals of climate change from GCMs into traces of weather that are tailored to support water resource planning efforts. In this way, the California weather generator is envisioned as a tool to help promote collaboration between climate scientists and water resource planners across the state.

1. Introduction

Climate change poses a major threat to the sustainability of water systems in California. Over the last two decades, California has experienced four periods of drought (2001-2004, 2007-2009, 2012-2016, 2020-2022), which when taken together rank as the driest 22-year period in at least 1,200 years (Williams et al., 2022). Each of these drought periods were ended by a string of atmospheric rivers (ARs) (Dettinger, 2013; Zechiel and Chiao, 2021), some of which led to record flooding, threatened major infrastructure projects (Henn et al., 2020), and most recently, even drove the re-emergence of the once-dry Tulare Lake.

These extremes are only projected to worsen over the next several decades. Climate change experienced to date has already exacerbated recent droughts via warming temperatures and enhanced drying power of the atmosphere (Williams et al., 2020). Similarly, extreme precipitation during recent, AR-related storms was likely more intense than it otherwise would have been due to warming temperatures and an associated increase in the moisture holding capacity of the atmosphere (Gonzales et al., 2019; Michaelis et al., 2022). These changes are direct thermodynamic responses to a warmer climate, and global climate models (GCMs) project that these types of thermodynamic climate change will accelerate over California and much of the US Southwest into the middle and end of the 21st century (Espinoza et al., 2018; Massoud et al., 2019; Rhoades et al., 2020; Overpeck and Udall, 2020; Huang and Swain, 2022). Other first-order, thermodynamic climate changes that are projected as a direct response to surface warming include greater frequency of extreme heat (Ullrich et al., 2018) and significant changes to seasonal snow accumulation and melt patterns (Berg and Hall, 2017; Ishida et al., 2019; He et al., 2021; Shulgina et al., 2023). Generally, these types of thermodynamic climate changes are consistent across theory, observations, and model projections (Pfahl et al., 2017; Allan et al., 2020), leading to high confidence in the direction of future change, although with residual uncertainty in the ultimate magnitude and rate of change.

Other types of long-term climate changes are also possible, but with a greater degree of uncertainty (Shepherd, 2014; Elbaum et al., 2022). For instance, shifts in atmospheric circulation, or dynamic climate changes, have been observed and projected for California. Zhang et al. (2022) recently showed a 40-year decline (from 1980-2018) in the frequency of a deep trough over Western North America that is associated with strong moisture flux and precipitation in California. This trend in the frequency of one large-scale atmospheric flow pattern appears to explain much of the drying trend observed over California and the Southwestern US during that same period, and an ensemble of historical CMIP6 climate model simulations suggest these observed shifts in atmospheric circulation have been driven by anthropogenically forced climate change. Others have also explored how late 21st century projections of these atmospheric flow patterns influence future climate over the Western US. For instance, Swain et al. (2018) found that changes in the spatial pattern and intensity of atmospheric circulation patterns may be responsible for projected increases in the frequency of dry and wet regimes in California. Shields and Kiehl (2016) focused specifically on the landfall locations of ARs along the West Coast and projected that landfalling ARs will eventually move equatorward during winter. However, others (Gao et al., 2016; Ma et al., 2020) have identified an opposite, poleward movement of ARs in other regions of the world, consistent with the argument that there is substantial regional variability and uncertainty in this type of dynamic climate change for ARs (Payne et al., 2020).

The composition of thermodynamic and dynamic signals of climate change – with some projected to occur with more confidence than others – presents a challenge for water resource planners seeking to ensure adequate water services to people and ecosystems over the next several decades. Conventionally, water resource planners have used downscaled climate scenarios from GCMs for planning purposes. GCMs are an invaluable tool for providing internally consistent scenarios that can be used to examine the possible pathways of climate change under anthropogenic forcing. However, the scenarios produced by GCMs contain a mixture of thermodynamic and dynamic components of climate change (Emori and Brown, 2005; Seager et al., 2010, 2014), along with significant natural variability and a series of biases in hydrologically important variables like precipitation linked to parameterized physics and coarse model resolution. This poses two challenges for water resources planning. First, it is often difficult and time consuming to separate out thermodynamic and dynamic signals of change from natural variability in GCM simulations. However, such separation may be of high interest to water resource planners who want to base their planning efforts on more detectable signals of change linked to thermodynamic mechanisms, but who may be wary of using climate data reflecting dynamic change in which there is less scientific consensus. Second, it is also very challenging to effectively remove biases from GCM simulations. For instance, improved GCM resolution does not necessarily improve precipitation biases linked to atmospheric dynamics (Muñoz et al., 2017; Maher et al., 2018), and when it does, there is necessarily a reduction in computational efficiency that limits ensemble simulations needed for risk-based assessments (Kendon et al., 2018). In addition, statistical corrections to certain types of model bias are far from straightforward, since they can be linked to modeled physical processes that could change under global warming and thus change the bias over time (Stephenson et al., 2012; Maraun et al., 2017).

Given the challenges above, water resource planners can benefit from an efficient alternative to climate scenario generation that can complement downscaled GCM simulations and help investigate water system risk under climate stress. Stochastic weather generators provide one such alternative. Weather generators are statistical models that are parameterized based on existing meteorological records and used to generate large ensembles of simulated daily or hourly weather records that are similar to but not bound by variability in past observations (Richardson, 1981; Wilks and Wilby, 1999; Fowler et al., 2007). For water system applications, weather generators must often develop sequences of multiple weather variables (e.g., precipitation, maximum and minimum temperature) at multiple locations while maintaining realistic persistence and covariance structures associated with transient, multi-day storm events and over longer (seasonal to interannual) timescales. Once fit to historical data, model parameters can be systematically altered to produce new traces of weather that exhibit a wide range of change in their distributional characteristics that may be experienced under climate change, including the intensity and frequency of average and extreme precipitation, heatwaves, and cold spells (Wilks, 2002, 2010, 2012; Acharya et al., 2017; Mukundan et al., 2019).

Stochastic weather generators are an ideal tool to support bottom-up climate vulnerability assessments of water systems (see Figure 1). These vulnerability assessments seek to define the range of climate conditions that lead to critical system vulnerabilities, and therefore are most relevant to the decision-making process of climate adaptation (Brown et al., 2012). Such

frameworks help circumvent the issue of severe uncertainty in more traditional top-down climate impact assessments, in which downscaled GCM projections are used to drive the entire analysis. Bottom-up vulnerability assessments ensure a thorough exploration of system sensitivities to small perturbations in climate that might not be captured when using a relatively small and often biased GCM ensemble. However, these assessments can still use GCM-based information to help define the range of plausible future climate change and their likelihood. Key to these vulnerability assessments is the ability to generate long (100-1000 year) climate traces useful for quantifying both water supply and flood risk, and to do so for many scenarios of potential climate change selected according to a carefully crafted experimental design (Steinschneider et al., 2015). Stochastic weather generators provide exactly this functionality.

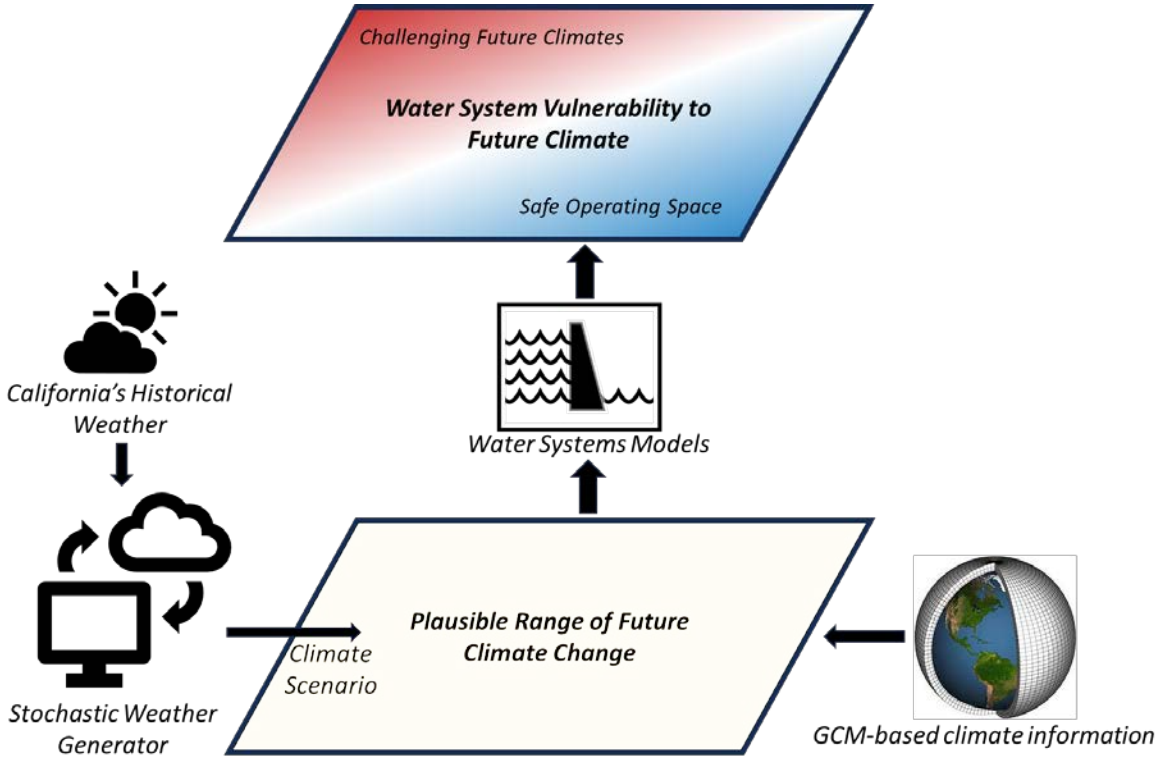

Figure 1. Overview of bottom-up vulnerability assessments for water systems under climate uncertainty. A stochastic weather generator like the model developed in this report can be fit to California's historical weather and then used to generate a large ensemble of future climate scenarios that systematically explore a range of plausible climate changes. This range can be informed by state-of-the-art GCM-based information. The climate scenarios can be used as forcing for models of hydrologic and water infrastructure systems to develop a clear picture of water system vulnerabilities to different types of climate change.

The purpose of the work under this project is to develop a stochastic weather generator and ensembles of future climate scenarios across California to help evaluate the vulnerability of water systems and the robustness of adaptation strategies under climate change across the state. This effort builds on a pilot study in 2018-2019, led by Cornell University and in collaboration with staff at the US Army Corps Hydrologic Engineering Center (HEC), California Department

of Water Resources (DWR), and collaborators at the University of Cincinnati. The pilot study developed a stochastic weather generator tailored for Western US climate and designed to drive bottom-up (or vulnerability-based) climate impact assessments of California water infrastructure while integrating state-of-the-art climate science into the design of future scenarios. The study resulted in the development of a prototype tool (Steinschneider et al., 2019) that was used in a proof-of-concept application for the Tuolumne River basin, which was further refined and expanded to additional basins in California (Najibi et al., 2021; Rahat et al., 2022). The work under this project advances and expands the scope of the model developed in that pilot study through the following objectives:

1. Refine the stochastic weather generator for better reproduction of dry and wet extremes.
2. Expand the scope of the model to generate internally consistent climate scenarios across the entire state of California.
3. Develop an ensemble of publicly available climate scenarios using this modeling framework to support ongoing climate change planning at DWR and other partnering water agencies across the state.

The ultimate goal of this project is to enable closer integration between bottom-up methods being pursued by DWR for climate adaptation and state-of-the-art, top-down climate projections being produced by climate experts across California. DWR has invested in "Decision-Scaling" methods of climate adaptation planning that help bridge vulnerability-based analysis with traditional risk-based assessment methods. The Decision-Scaling approach, which is featured in Phase 3 of the DWR Climate Action Plan (Schwarz et al., 2018; CA DWR, 2019; Ray et al., 2020), relies on climate stress tests driven by long stochastically generated climate scenarios that systematically explore the plausible space of future climate. The climate modeling work under this project provides decision makers at DWR and at other partnering entities across California with a novel scenario generation framework that is consistent with and advances the methods and goals promoted in its Climate Action Plan. The deliverables from this project include:

1. A refined weather generator model with associated code and documentation to support climate scenario development consistent with the Decision-Scaling methodology.
2. A dataset of 30 unique climate change scenarios, each consisting of 1000 years of weather generator simulated climate data (precipitation, maximum temperature, minimum temperature) at a ~6 km resolution across the entire state of California. The 30 scenarios represent a range of plausible climate changes to temperature, average precipitation, and precipitation extremes.
3. A parallel dataset of 30 climate change scenarios, each consisting of 100 years of climate data at a ~6 km resolution across California. These data are developed by directly perturbing the historical meteorological record to reflect plausible future changes in temperature, average precipitation, and precipitation extremes.

In the sections that follow, we describe the data used in the development of the California stochastic weather generator (Section 2), detail the model itself (Section 3), and present results summarizing model validation (Section 4). In Section 5 we introduce a large ensemble of future climate scenarios that are publicly available to support climate impact assessments across the state, and also introduce a preliminary investigation into projected changes of atmospheric

circulation by the end of the 21st century that could form the basis of future scenario generation efforts.

2. Data

2.1. Precipitation and Temperature Records in California

The weather generator can be used with many different continuous climate datasets. In this work, we collected observed daily precipitation (P) [mm] time-series between January 1, 1915 and December 31, 2018 (104 years) from the extreme-preserving gridded daily dataset for the conterminous United States developed by Pierce et al. (2021) (also known as the unsplit Livneh precipitation dataset)[1]. This extreme-preserving dataset, which has a 0.0625°×0.0625° (~ 6-by-6 km) spatial resolution, follows the same gridding method as employed in Livneh et al. (2013) and Livneh et al. (2015), while omitting the time adjustment applied to partition the precipitation gauge observations into a uniform timeframe. Pierce et al. (2021) demonstrated that this time-adjustment mutes extreme precipitation values, and so the unadjusted data are far closer to the observed daily extremes at individual stations compared to Livneh et al. (2013, 2015). Even though the gridding process still reduces precipitation extremes and increases the fraction of wet days, these errors are significantly smaller than those found in the time-adjusted data.

We further scrutinized the precipitation time series and noticed that there were 24 dates with extremely high precipitation intensities (> 10 in or 254 mm) in the summer (June-August, with a large majority in July), while the closest GHCN-d rain gauges showed much lower magnitudes (often ~1/10th of the gridded value). These extreme summertime precipitation intensities were deemed erroneous and were rescaled downward by a factor of 10 to reflect the precipitation intensity of nearby GHCN-d gauges.

We obtained observed daily minimum temperature (Tmin) [°C] and maximum temperature (Tmax) [°C] from the data in Livneh et al. (2013) for the period between January 1, 1915, and December 31, 2015. This dataset was then extended to December 31, 2018 using the PRISM daily dataset (PRISM Climate Group, 2014) to match the timeframe of the precipitation data. A few additional post-processing steps were then employed to prepare the temperature time series before using them in the stochastic weather generator. First, the temperature data in the 1915-2015 timeframe (Livneh et al., 2013) were bias corrected to the monthly PRISM dataset over the entire period. Then, the entire temperature time series (1915-2018) for each grid cell was detrended so that the long-term mean monthly temperatures between 1915-2018 matched those from 1991-2020. This ensured that the entire temperature series reflected warming that has already occurred in recent decades due to climate change. This is important for water resources planning activities that are forward-looking and need to accommodate future climate conditions, rather than past conditions that are unlikely to return. The detrending procedure was conducted as follows: a) calculate monthly averages from the daily temperature data in each year between 1915-2018; b) compute a linear trend of the monthly averages against year, separately by month; c) remove the month-specific trend from the daily data to create a sequence of daily residuals; d) calculate the monthly climatology for 1991-2020 using the monthly averages from step (a); and e) add the daily residuals calculated in step (c) to the monthly climatology calculated in step (d),

[1]An additional 3 years (January 2019-December 2021) of data using PRISM data were subsequently processed and are available for use but have not been included in the data products described in this report.

creating a time series of daily temperatures from 1915-2018 with monthly means from the 1991-2020 period.

All of these data were collected over and slightly outside of the USGS HUC-2 water resources region that spans the entire state of California and some parts of Nevada and Oregon (Region 18). This led to a final dataset consisting of 13,786 grid cells (Figure 2). The final time series of both precipitation and temperature were truncated to the period between 1948-2018 to match the timespan of the atmospheric data used for weather regime classification (described next).

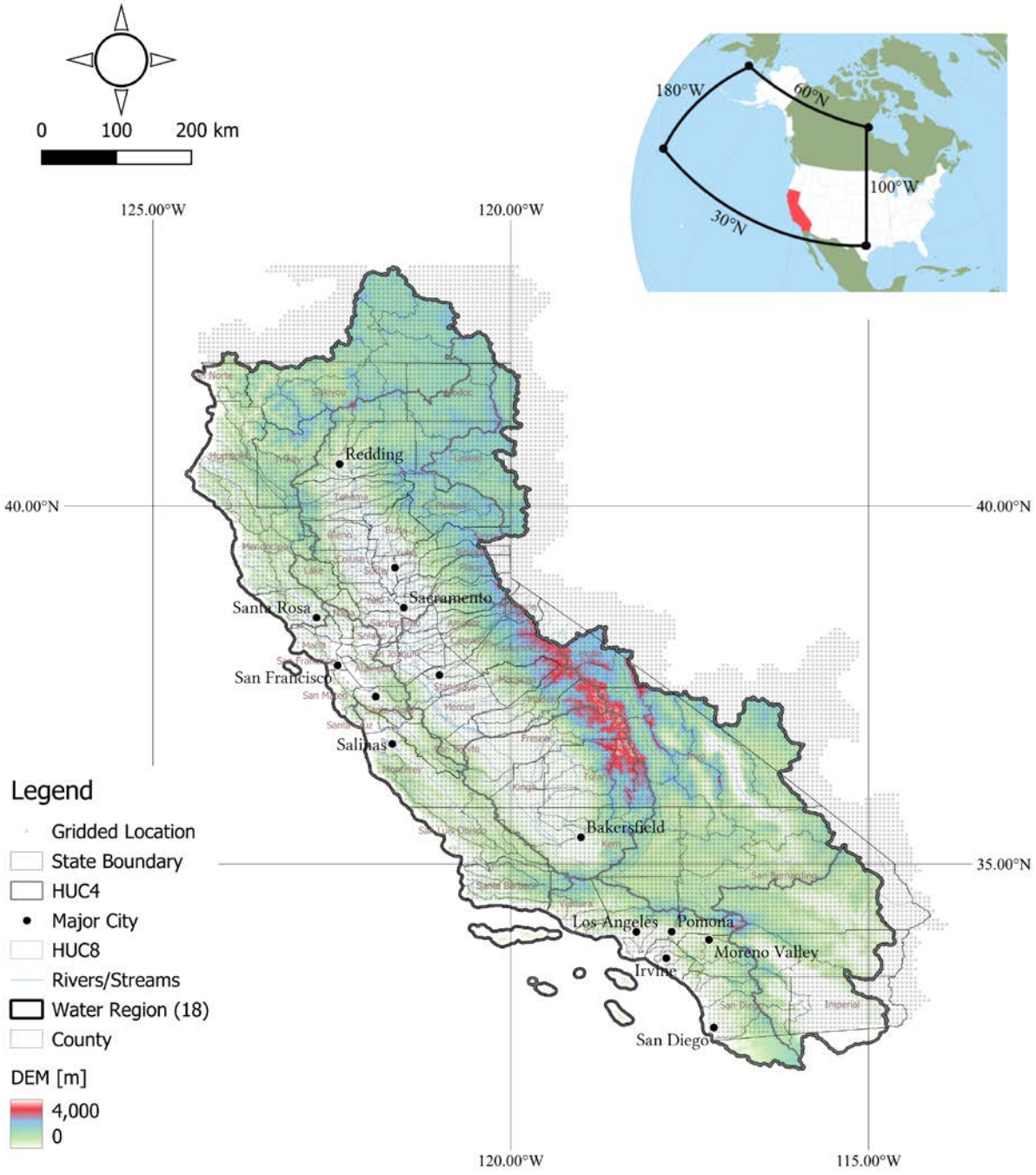

Figure 2. Domain of the HUC-2 water resources region over California (Region 18), along withHUC-4 and HUC-8 subregions. All climate grid cells modeled within the stochastic weather generator are also shown, and extend slightly beyond the HUC-2 water resources region. Inset shows the atmospheric boundary used for weather regime (WR) identification.

2.2. Atmospheric Circulation over the Pacific-North American Sector

We obtained daily gridded (2.5°×2.5°) geopotential heights (GPH) [m] at the 500-hPa level from the National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) reanalysis dataset (NCEP/NCAR Reanalysis 1; Kalnay et al., 1996) between January 1, 1948, and December 31, 2021 (74 years). The gridded data were then extracted for the region between 30°N-60°N and 180°W-100°W, covering much of the Pacific-North American sector (see inset in Figure 2). The GPH data were centered by month to remove their seasonal cycle, producing GPH anomalies (GPHAs).

We also collected the daily occurrences of AR landfalls along the US west coast from the Scripps Institute of Oceanography (SIO)-generated AR catalog (SIO-R1 catalog; Gershunov et al., 2017). This catalog reports individual AR events across western North America from January 1, 1948 to December 31, 2020, detected using integrated vapor transport (IVT) and integrated water vapor (IWV) from the NCEP/NCAR Reanalysis 1. ARs are defined as a 1500 km long structure with IVT and IWV in excess of 250 kg/m/s and 15 mm, respectively, sustained continuously for at least 18 hours. The grid cell corresponding to the maximum IVT along the coastline is considered to be the AR landfall location.

2.3. Annual Standardized Precipitation Index over California

We collected a gridded dataset of a cold-season standardized precipitation index (SPI) (0.5°×0.5°) across California, taken from the study in Borkotoky et al. (2021). Similar to the work presented in Gupta et al. (2022), these data will be used to support a model that can characterize large-scale weather patterns with appropriate inter-annual variance. We gather the SPI data between 1948-2021, which is the same period as the GPHA data described above.

3. Weather-Regime based Stochastic Weather Generator for California

This project advances a semiparametric, multivariate, and multisite stochastic weather generator that was previously developed in the Tuolumne River basin during the cold season (Steinschneider et al. 2019) and further refined across several other basins in California (Najibi et al., 2021; Rahat et al., 2022). The weather generator is designed to separately model dynamic and thermodynamic atmospheric mechanisms of climate variability and change through statistical abstractions of these processes. To capture atmospheric dynamics, the weather generator simulates sequences of weather regimes (WRs). WRs are recurring large-scale atmospheric flow patterns (e.g., upper-level, quasi-stationary blocks and troughs) that appear at fixed geographic locations, persist for days-to-weeks within a season, and organize high-frequency weather systems (Robertson and Ghil, 1999; Robertson et al., 2015). They represent intermediary phenomena in the stochastic continuum of atmospheric perturbations that connect local weather to hemispheric circulation, and provide a parsimonious way of abstracting major patterns of atmospheric circulation into stochastic simulations of weather. More simply, WRs are

patterns in the atmosphere that help link long-term climate to short-term weather. To capture thermodynamic mechanisms of climate change, the weather generator post-processes simulated data to reflect patterns of warming and thermodynamic scaling of precipitation rates with that warming. These properties of the model are represented in a hierarchical structure composed of three primary modules designed to capture natural climate variability and climate change: 1) identification and simulation of WRs that dictate the large-scale atmospheric flow across the Pacific-North American sector; 2) simulation of local weather across California conditioned on the WRs; and 3) perturbations to the simulation schemes in (1) and (2) reflective of thermodynamic and dynamic climate change.

As part of this work, the stochastic weather generator presented in Steinschneider et al. (2019), Najibi et al. (2021), and Rahat et al. (2022) underwent several refinements to improve the representation of droughts and pluvials, extreme multi-site precipitation events, and other climate statistics. Therefore, we present an overview of the new set of algorithms in the sub-sections below (also see Figure 3), and then present the mathematical details of these algorithms in Appendix A.

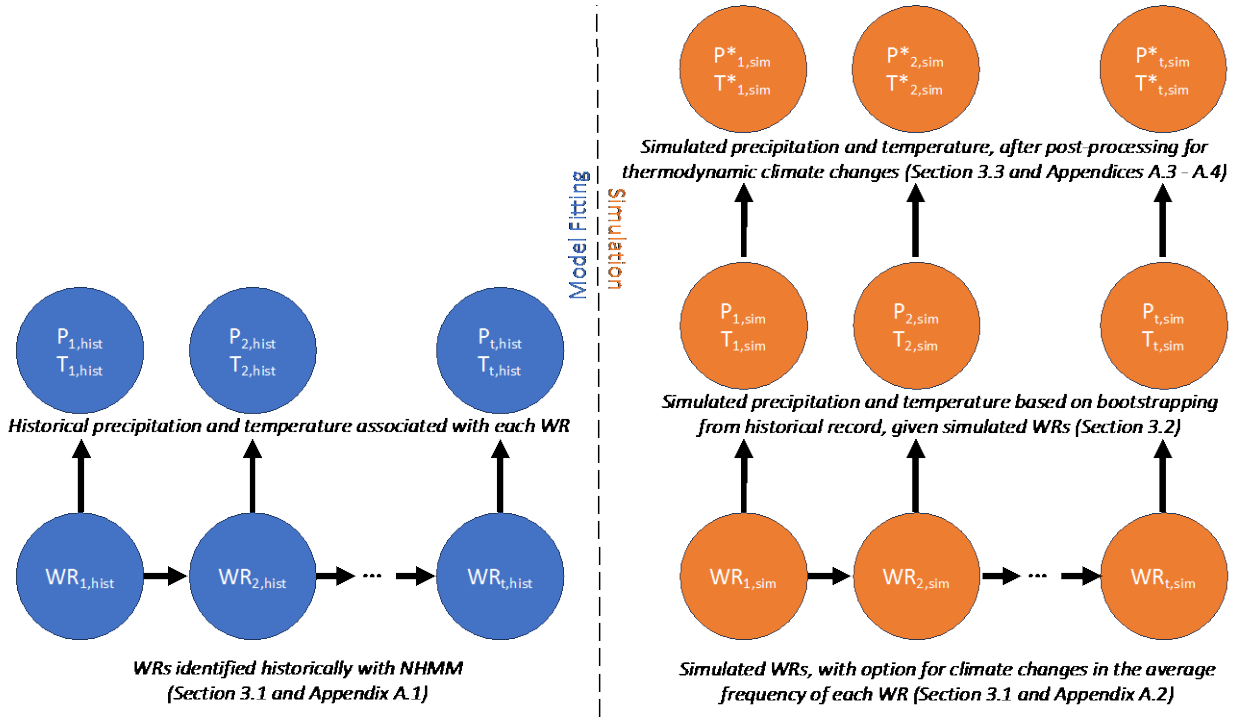

Figure 3. Overview of the stochastic weather generator algorithm. During model fitting, daily weather regimes (WRs) are identified from historical atmospheric circulation data using a Nonhomogeneous Hidden Markov Model (NHMM). Historical daily precipitation (P) and temperature (T) data across the state of California are associated with historically identified WRs. During simulation, the model creates new sequences of WRs, with the option to change their frequency of occurrence as a signal of dynamic climate change. Simulations of daily precipitation and temperature across the state are then generated by bootstrapping values from the historical record, based on the simulated and historical sequences of WRs. Finally, additional thermodynamic climate changes are imposed on the simulations of precipitation and temperature using post-processing methods, creating the final precipitation (P^*) and temperature (T^*) data.

3.1. Weather Regime Identification and Simulation

Following Najibi et al. (2021), we use a Nonhomogeneous Hidden Markov Model (NHMM) to identify WRs. NHMMs are nonlinear statistical models that use latent variables to identify clusters in state-space while simultaneously accounting for the distribution and temporal dynamics of observed data (Rabiner, 1989; Hughes and Guttorp, 1994). In this application, we first divide the 500-hPa GPHAs into two seasons: November-April (cold season) and May-October (warm season). We then project the separated GPHAs onto their first J empirical orthogonal functions (EOFs), where J is chosen using a scree test to ensure that the selected EOFs explain the majority (e.g., > 90%) of the variance in the data (here, J=10). We subsequently evaluate a first-order NHMM on the J PCs of GPHAs to assign each day in the record to one of K separate WRs. This is done separately for the cold and warm season, and K can differ across the two seasons (the selection of K is described further in Section 3.2.2 below). The NHMM is fit using two harmonics as exogenous variables to account for seasonality in the WRs. In addition, the first four PCs of the SPI dataset over California (which explain ~80% of the variance) are also incorporated as exogenous variables to improve the inference of inter-annual variability of cold-season WRs. Importantly, by using the NHMM, days are classified into WRs in a way that explicitly considers WR persistence, which will lead to persistent weather in the weather generator simulations (e.g., long dry spells like those experienced during the Ridiculously Resilient Ridge; Swain et al., 2014). The Expectation-Maximization algorithm (Dempster et al., 1977) with the forward-backward algorithm (Baum and Petrie, 1966; Baum et al., 1970) is used to estimate the parameters of the NHMM. The most probable sequence of hidden states is computed using the Viterbi algorithm (Forney, 1973; Rabiner, 1989). We utilized the R-package 'depmixS4' (Visser and Speekenbrink, 2010) to fit the NHMM. More details and mathematical formulations related to the NHMMs are provided in Appendix A.1.

Future time series simulations of WRs of an arbitrary length could be created through forward simulation of the fitted NHMM, as in Najibi et al. (2021). However, we found that WR simulation using this approach, when coupled with the local weather generation algorithm described below in Section 3.2, underestimated the magnitude of extreme, multi-year droughts and pluvials (i.e., simulations were over-dispersed at inter-annual timescales). One approach commonly used to address this challenge is to incorporate additional covariates in the forward simulation of the NHMM (e.g., an El Niño – Southern Oscillation index or information from tree-ring reconstructions; Steinschneider et al. (2019), Gupta et al. (2022)). However, it is not straightforward to change the frequency of future WRs in a dynamic climate change scenario using this approach. Therefore, we developed a novel non-parametric approach to WR simulation that addresses the issue of overdispersion (suppressed variability) while still allowing for future climate change scenarios with altered WR probabilities. In this approach, we first cluster the historically identified WRs into four-year segments across the historical record of 1948-2019. We then resample four-year segments with equal probability to develop future time series of WRs of an arbitrary length. This approach ensures that inter-annual WR dynamics are almost completely preserved. To simulate WR sequences that reflect a dynamic climate change scenario, we simply resample the four-year segments with pre-specified, non-equal probabilities in a way that results in final WR frequencies that match a pre-set scenario of interest (described in more detail below in Section 3.3 and Appendix A.2).

3.2. Local Weather Generation Conditioned on Weather Regimes

3.2.1. Weather Generation Algorithm

Local weather is simulated by bootstrapping weather data (e.g., daily precipitation, minimum and maximum temperature) based on sequences of simulated WRs. Starting on simulation day t, the vector WR will contain n days of the i^{th} WR (i.e., WR_t through WR_{t+n-1} equal i). Here, n usually varies from a single day to a few weeks, although it can extend longer than 1 month due to the persistence of WRs. To generate weather for those n days, we resample an n^*-day block of historical data that was also classified into the i^{th} WR, based on the absolute difference between the historical and simulated block length (i.e., a historical block with length n^* closer to n will receive a higher probability and will be resampled with a higher likelihood). We also require that any resampled blocks meet two other criteria: 1) the central day of the historical block is within a 3-day window of the day of year for simulation day t; and 2) the day prior to the historical block has the same state of regionally averaged precipitation (i.e., dry (p_{avg}<0.25 mm) or wet (p_{avg}>0.25 mm), where 0.25 mm is trace precipitation). This ensures that the resampled data will preserve the seasonality of local weather and better maintains precipitation persistence across sites. We define regionally averaged precipitation using daily precipitation averaged over the Calaveras, Stanislaus, Tuolumne, Merced, and Upper San Joaquin watersheds (MIL, MRC, NHG, NML, TLG: 425 grids in total), which are centrally located in California. This approach helps ensure that inter-daily precipitation dynamics are preserved in the center of the state, which also helps to preserve these dynamics in the northern and southern ends of the state due to regional coherence in storm tracks that pass over California.

If a historical block happens to be resampled with a longer length than the simulated one (i.e., $n^* > n$), we reduce the length of the resampled block by discarding days from that block randomly from one of its two ends until $n^* = n$. If the length $n^* < n$, then the remaining length $n - n^*$ is used as the basis to resample another block for WR i, and this process is continued until data has been resampled for the entire block of n days. At this point, the WR will change states and the resampling procedure begins again. By using this block bootstrap procedure, the resampled data are more likely to capture the entire life cycle of passing storms (and the resulting space-time structure in weather) over the basin of interest.

The block bootstrap will preserve many of the properties of the marginal and joint distributions of local weather variables, but at the expense of being able to simulate values outside the range of the instrumental record. To address this drawback specifically for heavy precipitation, the weather generator uses a copula-based jittering approach that adds noise to resampled heavy precipitation data as a post-processing step. To do this, we first fit a mixture model to the observed, non-zero precipitation at each site, which uses a gamma distribution for the bulk distribution (by month) up to a threshold and a Generalized Pareto distribution (GPD) to model the tail of the distribution beyond that threshold. For this work, we selected the site-specific 99th percentile of non-zero values as the separating threshold. Then, for heavy precipitation values above the 99th percentile that are resampled in the block bootstrap, we calculate the non-

exceedance probability associated with that event based on the fitted gamma-GPD (only the GPD is used here, since this is only done for resampled precipitation above the threshold). Spatially correlated random noise is then added to these non-exceedance probabilities across sites, which are then mapped back through the gamma-GPD models to develop new heavy precipitation values across locations. This approach is designed such that final values of simulated, jittered precipitation can extend beyond the range of historical precipitation values, but preserve the marginal distribution of precipitation at each site and the rank correlation structure across sites. More mathematical detail on this approach is provided in Appendix A.3.

3.2.2. Selection of the Number of Weather Regimes

The identification and simulation of the WRs (Section 3.1) plays a critical role in the performance of the weather generator. A key parameter of the model is *K*, the number of WRs that should be used in each of the two seasons. To calibrate the value of *K*, we followed the approach in Najibi et al. (2021) and generated a very long trace from the weather generator for the entire state under baseline conditions (i.e., no climate change) and for values of *K* ranging from 2 to 10 WRs separately for the cold season and warm season. This simulation is 1008 years long, although for simplicity we hereafter refer to it as the 1000-year simulation (the use of 1008 years is related to a nuance of the simulation strategy described in Appendix A.2). We then evaluated the distribution of a variety of climate performance measures for a random selection of 100 grid cells across the state of California for different values of *K*. We quantified performance based on percent bias between simulated and observed statistics of interest, which are listed in Table 1 (for precipitation) and Table 2 (for temperature), along with their name and short description of their computation. We select *K* for each season that results in the best weather generator performance across statistics. No one value of *K* (i.e., number of WRs) for either season is likely to maximize performance across all these measures for the entire state of California. Therefore, we selected *K* based on the value that provided the best balance across all statistics over all sampled grid cells.

Table 1. Statistics of precipitation used for model evaluation.

No	Statistic	Description [unit]
1	Mean	Average of precipitation distribution [mm]
2	Standard deviation	Standard deviation of precipitation distribution [mm]
3	Inter-annual standard deviation	Standard deviation of annual total precipitation variation [mm]
4	Seasonality	Monthly total and water-year cumulative precipitation intensity distribution with their mean, median, 10th, and 90th percentiles
5	Extreme 1-day events	Maximum 1-day precipitation intensity (equal/greater 99th site-specific threshold) [mm]
6	Multi -day, -month, and -year maxima	Maximum 7-day, 10-day, 1-month, 3-month, 6-month, and 1-year average precipitation intensity in the entire record [mm]
7	Mean and maximum wet spell length	Average and maximum length of a wet spell (consecutive non-zero precipitation days) [days]
8	Annual pluvials frequency	Frequency of maximum 1, 2, 3, 5, 10, and 30-year rolling average precipitation intensity in the

		entire record [number of water years]
9	Return level events	2-, 5-, 10-, 50-, 100-, 500-, 1000-year return level event, estimated from GEV and GPD-Poisson distributions fit to annual maxima [mm]
10	Annual multi-year droughts frequency	Frequency of minimum 1, 2, 3, 5, 10, and 30-year rolling average precipitation intensity in the entire record [number of water years]
11	Worst multi-year droughts intensity	Minimum 1, 2, …, 10-year precipitation intensity in the entire record [mm]
12	Mean and maximum dry spell length	Average and maximum length of a dry spell (consecutive zero precipitation days) [days]

Table 2. Statistics of temperature used for model evaluation.

No	Statistic	Description [unit]
1	Mean	Average of temperature distribution [°C]
2	Standard deviation	Standard deviation of temperature distribution [°C]
3	Inter-annual standard deviation	Standard deviation of annual average temperature variation [°C]
4	Heat wave frequency	Number of instances with three or more consecutive days with temperature over 32.2°C (90°F) [number of events]
5	Heat stress frequency	Any instances of 3-day rolling mean temperature above 30°C (86°F) [number of events]
6	Mean and maximum of heat wave duration	Average duration and longest duration of heat waves in the record [days]
7	Mean and maximum of heat wave intensity	Average intensity and largest intensity of heat waves in the record [°C]
8	Cold wave frequency	Number of instances with three or more consecutive days with temperature below -7°C (20°F) [number of events]
9	Cold stress frequency	Any instances of 3-day rolling mean temperature below 0°C (32°F) [number of events]
10	Mean and maximum of cold wave duration	Average duration and longest duration of cold waves in the record [days]
11	Mean and maximum of cold wave intensity	Average intensity and largest intensity of cold waves in the record [°C]

3.3. Climate Change Scenarios

In the sub-sections below, we describe the thermodynamic and dynamic scenarios of climate change created through this project, as well as their application to climate data across the state of California.

3.3.1. Thermodynamic and Dynamic Climate Change Scenarios

In this work, we primarily focus on the development of a large set of thermodynamic perturbations composed of different scenarios of warming and intensification of the daily precipitation distribution. There is consensus around the direction of change for these types of

scenarios, although the rate of these changes over the next several decades is less certain. We consider five different scenarios of temperature change, from +1°C to +5°C by 1°C increments (see Figure 4; Table 3). Temperature change is treated simply by adding step changes to baseline daily maximum and minimum temperature data (which have already been detrended to reflect recent warming) uniformly across the entire spatial domain. This range of temperature increase was inferred from an ensemble of climate model projections selected by the CA DWR Climate Change Technical Advisory Group (CA-DWR CCTAG, 2015). Those projections, taken from a subset of ten high-performing GCMs for California from the CMIP5 archive, suggested that +5°C was approximately the maximum amount of warming that could be expected towards the end of the 21st century under the RCP 8.5 emission scenario, based on a multi-model average.

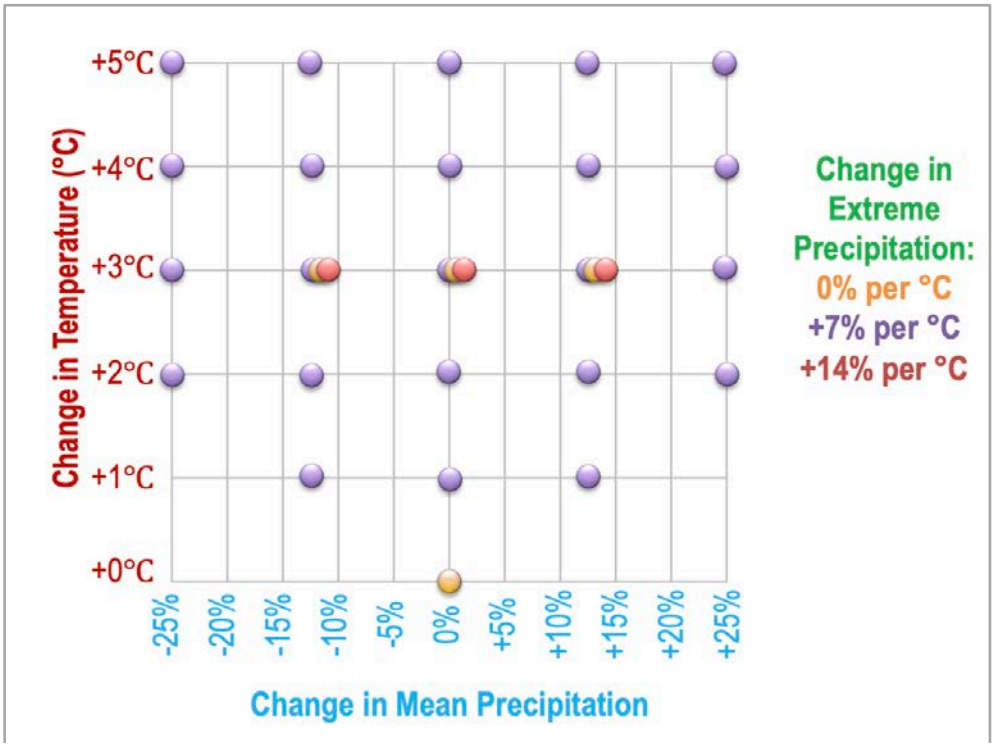

Figure 4. Range of climate change scenarios developed under this project.

Table 3. List of 30 climate scenarios.

Scenario Number	Incremental Temperature Change [°C]	% Change Extreme Precipitation Quantile	% Change Precipitation Mean
1: baseline	0	0	0
2	2	7	-25
3	3	7	-25
4	4	7	-25
5	5	7	-25
6	1	7	-12.5
7	2	7	-12.5
8	3	7	-12.5
9	4	7	-12.5
10	5	7	-12.5
11	1	7	0
12	2	7	0
13	3	7	0
14	4	7	0
15	5	7	0
16	1	7	12.5
17	2	7	12.5
18	3	7	12.5
19	4	7	12.5
20	5	7	12.5
21	2	7	25
22	3	7	25
23	4	7	25
24	5	7	25
25	3	0	-12.5
26	3	0	0
27	3	0	12.5
28	3	14	-12.5
29	3	14	0
30	3	14	12.5

Scenarios of precipitation intensification are modeled by scaling the distribution of daily precipitation in a way that replicates the effects of warming temperatures on precipitation through increases in the moisture holding capacity of the atmosphere. In California, past work

has shown that warming temperatures will lead to an increase in the most intense precipitation events (often associated with ARs) but a decline in the magnitude of smaller precipitation events (Gershunov et al., 2019). This type of change effectively stretches the daily precipitation distribution, making extreme events more extreme and suppressing the magnitude and frequency of lighter precipitation events (see Figure 5 for an illustration).Similar to other studies (Pendergrass and Hartmann, 2014), we mimic this effect by imposing changes to the moments and quantiles of the precipitation distribution as a function of warming. Specifically, we selecta scaling rate for the 99th percentile of non-zero precipitationand adjust thegamma-GPD mixture distribution toimpose that selected scaling rate. For instance, if we assume a scenario with 2°C of warming and a 7% per °C increase in extreme precipitation (which would match the theoretical Clausius-Clapeyron rate of increase in atmospheric water holding capacity with warming; Najibi and Steinschneider (2023)), then the most extreme precipitation events should increase by ~14.4% (1.07^2). We adjust the gamma-GPD models fit to all sites to require this percent change in the far upper tail of the distribution. If mean precipitation is held constant at baseline levels, this change will force smaller precipitation values under the gamma-GPD model to decrease in order to compensate for the increases in extreme events, i.e., the distribution of non-zero precipitation will be stretched, similar to the model-based results of Gershunov et al. (2019).

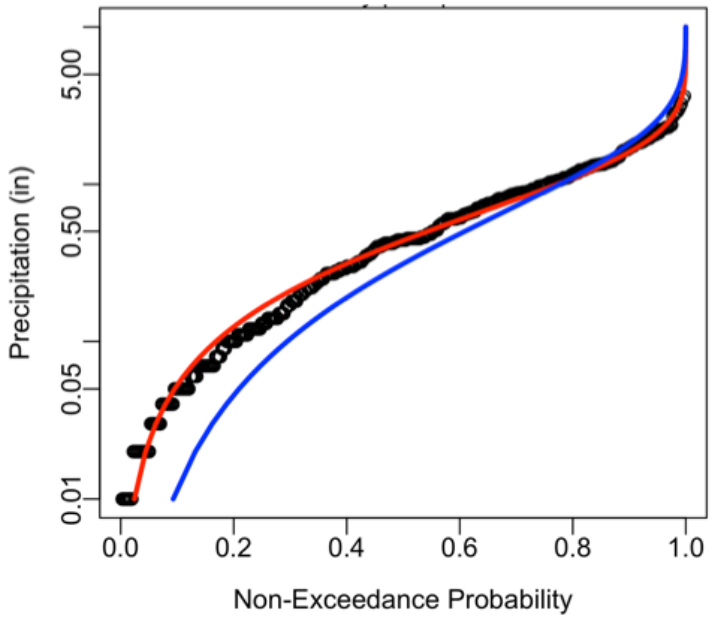

Figure 5. Illustration of thermodynamicallydriven changes to the distribution of daily, non-zero precipitation. The historical distribution of precipitation (black points and red line) changes under warming such that the largest precipitation events (at high non-exceedance probabilities) become even larger (blue line). If average precipitation does not change under warming, then low and moderate precipitation events must scale downward to balance the increases at the upper end of the distribution.

Once new model parameters of the gamma-GPD model are determined for each site and month, daily simulated precipitation is adjusted by first determining the non-exceedance probability, then perturbing the non-exceedance probability using the copula-based jitter model (see Section 3.2 and Appendix A.3), and finally determining a new precipitation value based on the adjustedgamma-GPD distribution. This procedure is repeated for each non-zero precipitation

amount for each site synthesized by the weather generator.The mathematical details of this approach are detailed in Appendix A.4.

We enforce a 7% per °C increase in extreme precipitation for all climate scenarios developed in this work (Figure 4, Table 3), as this rate of extreme precipitation scaling is theoretically the best supported based on the rate at which the atmospheric moisture holding capacity increases with warming. However, as a sensitivity analysis, we also develop a subset of scenarios with 0% per °C and 14% per °C increase in extreme precipitation (only for 3°C of warming), in order to assess how important extreme precipitation scaling is for water systems performance. Importantly, the degree to which extreme precipitation is scaled is directly tied to the temperature trends imposed, thus respecting the underlying thermodynamic mechanism that drives scaling. That is, precipitation scaling in the model is entirely determined after specifying a temperature trend and a scenario of precipitation scaling per degree warming. In this way, emerging hypotheses related to regional warming and precipitation-temperature scaling relationships, which are arguably less uncertain than precipitation changes linked to dynamical processes (Pfahl et al., 2017), can be directly tested with respect to their impact on water systems.

In addition to the thermodynamic climate changes in temperature and extreme precipitation described above, we also develop scenarios of change in mean precipitation. Part of the signal around mean precipitation change may be related to the overall increase in atmospheric moisture linked to warming (a thermodynamic signal), while other components of change in mean precipitation are related to shifts in atmospheric circulation (a dynamic signal). Therefore, changes in mean precipitation reflect a mix of thermodynamic and dynamic climate changes and remain more uncertain than strict thermodynamic changes (warming and precipitation intensification). We consider five different scenarios of mean precipitation change, ranging from -25% to +25% of baseline levels by 12.5% increments. Similar to temperature, this range was based on a subset of end-of-21st-century CMIP5 projections selected by the CA DWR Climate Change Technical Advisory Group. We limit mean precipitation changes to ±12.5% for scenarios with only 1°C of warming to acknowledge the relationship between more intense future warming and larger changes in average precipitation (see Figure 4, Table 3). In addition, we note that the range of projected change under CMIP6 is somewhat narrower (-10% to +15%) than what was projected with CMIP5, based on the IPCC Interactive Atlas (https://interactive-atlas.ipcc.ch/) and a preliminary assessment of CMIP6 projections specifically over California. Altogether, there are 29 separate scenarios of climate change developed in this work, along with a baseline scenario with no changes imposed, leading to a total of 30 scenarios.

Finally, we develop a small set of scenarios of pure dynamical climate change, which are composed of changes to the frequency of different WRs (i.e., shifts in atmospheric circulation). The direction of change for these types of climate change is very uncertain, but the risk of such change for water systems throughout the state is potentially very large. Hence, in this work we seek to develop a limited number of scenarios that reflect dynamic climate change to isolate the potential vulnerabilities that may be associated with these types of change.As a proof-of-concept, we utilize recent trends in WR frequencies over the 72-year historical record (1948-2019) to determine which WRs have changed in frequency most significantly in the observations, and then enforce a scenario of dynamic climate change where these significant changes persist into

the future.We then develop a 1000-year simulation of WRs and associated weather from the stochastic weather generator with these adjusted WR frequencies (see Appendix A.2 for the methodology of this approach), assess the impacts of this scenario on key hydrologic metrics of interest (e.g., frequency and magnitude of drought and extreme precipitation events), and compare these impacts to the impacts observed under of scenarios from Table 3. This comparison will help illustrate how the severity of impacts of dynamic climate change compare to that of thermodynamic climate change scenarios.

3.3.2. Application

We create two separate data products based on the 30 climate change scenarios discussed in Section 3.3.1 and shown in Figure 4 and Table 3. These two datasets are briefly explained below:

a) The first dataset, called the *Historical Climate Change Ensemble*, applies all 30 scenarios of change in Table 3 to the 100-year record (1915-2018) of historical daily precipitation, maximum, and minimum temperature at all grid cells across the state of California. That is, the stochastic weather generator is not used in this first data product, except for the temperature and precipitation scaling algorithms described in Section 3.3.1, which are applied to the 104-year historical record of observations. The Historical Climate Change Ensemble allows water managers to ask questions about the performance of their system if exposed to the same sequences of weather as seen in the historical record, but under shifts in core attributes of the temperature and precipitation distribution that reflect plausible long-term climate change.
b) The second data productis called the *Stochastic Climate Change Ensemble*. Here, we use the full weather generator model described in Sections 3.1-3.3 to generate over 1000 years of daily weather at all grid cells across the state of California, and then apply the 30 separate climate change scenarios in Table 3 to those 1000-year simulations. The Stochastic Climate Change Ensembleprovides a much richer timeseries of weather that can be used to test water system vulnerability to climate. As shown in Section 4 below, the 1000-year records simulated by the stochastic weather generator exhibit plausible droughts and extreme precipitation events that extend beyond the worst seen in the historical record, solely based on natural climate variability captured by the model (i.e., no climate change).*When coupled with the different scenarios of climate change in Table 3, the Stochastic Climate Change Ensemble provides climate time series that can be used to better assess how the combination of natural climate variability and long-term climate change can stress water systems, as compared to the Historical Climate Change Ensemble.*

4. Model Evaluation

4.1. Weather Regime Identification

As discussed in Section 3.2.2, we select the number of WRs (K) for both the cold season and the warm season by evaluating the weather generator's ability to reproduce a variety of weather statistics (see Tables 1 and 2) for values of K ranging from 2 to 10 WRs, separately in each

season. We do not show the model's performance on all statistics above for all combinations of *K* here, but rather focus on the final selected WRs and their interpretation. In Section 4.2 below, we present the performance ofthe final selected model across all metrics in Tables1and 2.

The calibration procedure resulted in the identification of 7 and 3 WRs in the cold and warm seasons, respectively (10 WRs in total). Figure 6 presents the composites of 500-hPa GPHAs for days categorized under each WR, along with their average temporal frequency over a calendar year.WRs 1 to 7 occur in the cold season, while WRs 8 to 10 occur in the warm season. WR1 exhibits a widespread trough centered over the eastern Pacific and off the western US coastline, which is located south of ahigh-pressure anomaly anchored over the Bering Sea. Under WR2, thereis aridge and trough directly over the northwest US and Aleutian Islands, respectively. This pattern is reversed in WR4 and WR7, which differ only in the longitudinal location of the pressure dipole. WR3 exhibits a ridge directly over the northwest US that is slightly more elongated and less intense compared to the ridge in WR2. This ridge is shifted further to the west over the eastern Pacific in WR5. The most notable feature of WR6 is an intense low over the Gulf of Alaska. Finally, the WRs in the summer (WRs 8-10) all exhibit weaker GPHAs, but resemble some of the same spatial patterns as seen in the cold season.

An analysis of California-wide precipitation and temperature anomalies under the WRs in Figure 6 showed climate varies considerably across the state depending on the prevailing WR. For example, conditions across the Central Valley are wettest in the cold season under WR1, followed by WRs 4, 7, and 6. This is consistent with the deep troughs under each of these WRs that direct storm tracks and moisture over the state. In contrast, conditions across the Central Valley are driest in the cold season under WRs 2, 3, and 5 (in that order), which aligns with the ridging under these WRs that blocks moisture flow over California.Precipitation reaches its absolute lowest under WRs 8-10, because these WRs occur in the warm season when a semi-permanent ridge of high pressure expands further north and pushes storm tracks northof the state(note: this is not seen in Figure 6, because seasonality was removed from the GPHAs).

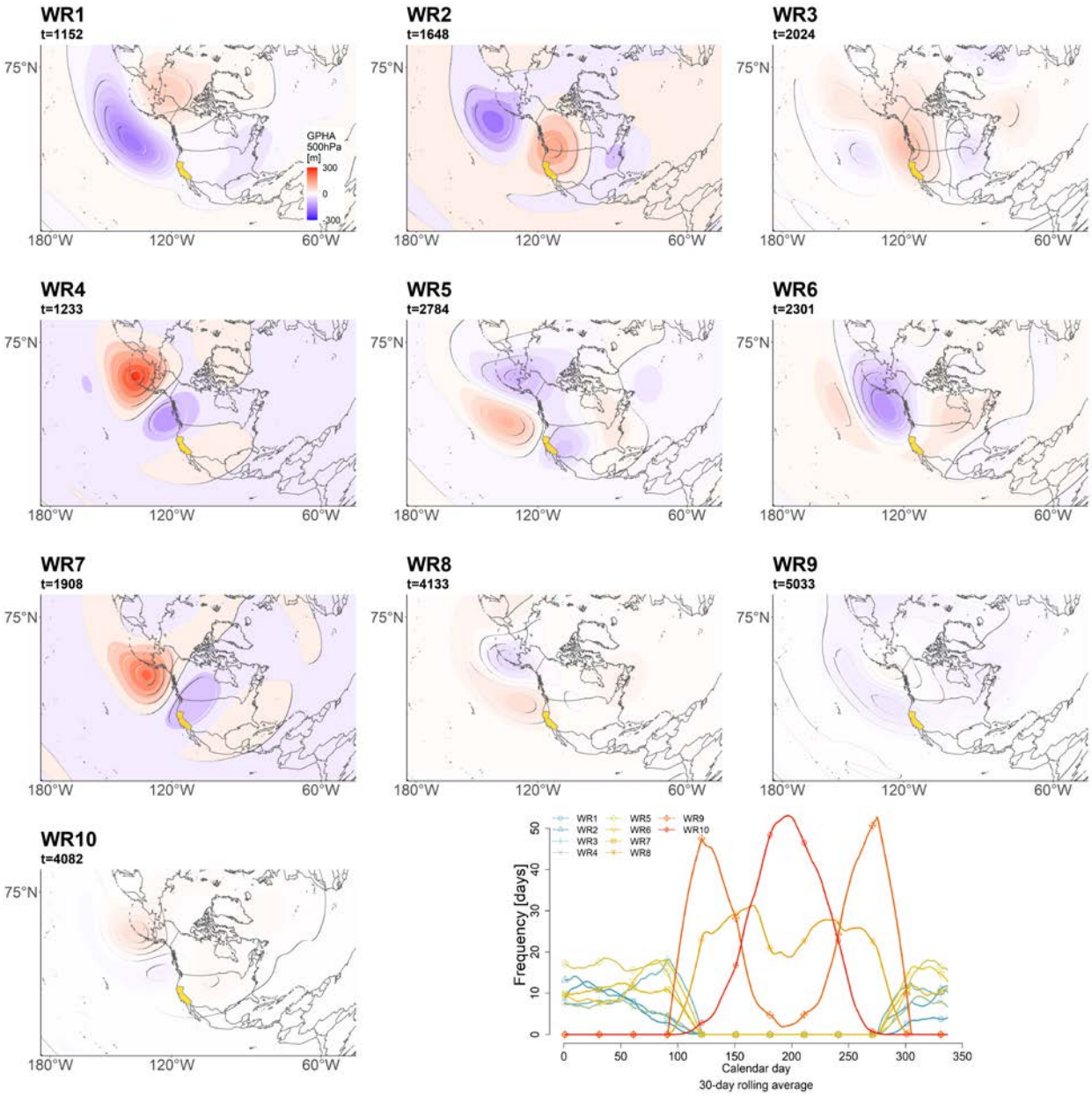

Figure 6. Composites of 500-hPa GPHAs [m] for days categorized under each WR during the cold season (November-April; WRs 1 to 7) and warm season (May-October; WRs 8 to 10). The number of days (*t*) classified under each WRis shown above each composite. California is colored with yellow shading on the map. The temporal frequency of WRs per calendar day based on a 30-day rolling average smooth is also shown (1948-2021).

4.2. Validation of Simulated Weather

Following WR identification, the weather generator was then validated based on the reproduction of weather statistics across the entire state. For the purposes of validation, the precipitation and temperature metrics(Tables 1 and 2) were evaluated at the scale of HUC-4 basins(Figure 7), with some metrics shown based on an average across grid cells within the HUC-4 and other metrics shown for individual grid cells within the HUC-4. Table 4 provides

detailed information about these HUC-4 basins, including their drainage areas and dimensions of the compiled gridded data points. Note that some areas at the boundary of the state are located in HUC-4 regions that largely sit outside of California, and for those locations we evaluated the weather generator in HUC-8 basins that are mostly within California (also shown in Figure 7).For illustration, Figures 8-16below show the validation results for one HUC-4 subregion (HUC-4: 1804 - San Joaquin). All results in these figures are based on a 1000-year baseline simulation of the stochastic weather generator with no climate change. A full list of precipitation and temperature validation statistics for the rest of the HUC-4 subregions in California are available in the Supplementary Material. Results presented for the San Joaquin HUC-4 are very representative of the model's performance for the other HUC-4 regions. Finally, we also present a brief validation of the reproduction of AR landfall frequencies near and along the coast of California.

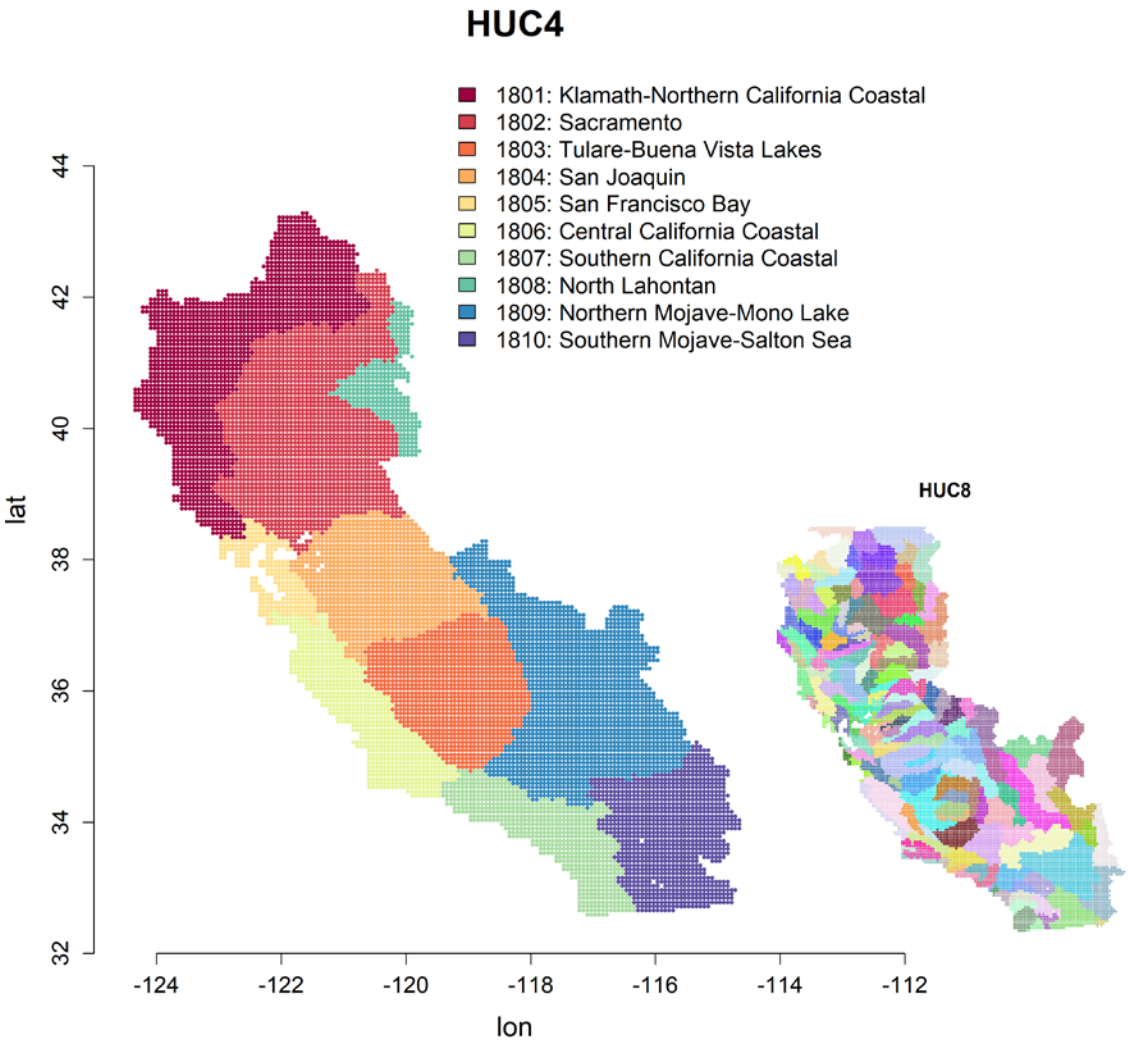

Figure 7. HUC-4 domains (subregion level) across California, with HUC-8 basin boundaries (subbasin level) shown in the inset.

Table 4. Properties of HUC-4 basins across California.

Basin Number	Basin Name	HUC4	State	Area [km^2]	# Gridded Points
1	Klamath-Northern California Coastal	1801	CA, OR	67,762	1758
2	Sacramento	1802	CA, OR	72,013	1941
3	Tulare-Buena Vista Lakes	1803	CA	42,498	1084
4	San Joaquin	1804	CA	40,986	1072
5	San Francisco Bay	1805	CA	13,910	259
6	Central California Coastal	1806	CA	34,287	728
7	Southern California Coastal	1807	CA, MX	35,863	703
8	North Lahontan	1808	CA, NV	11,791	318
9	Northern Mojave-Mono Lake	1809	CA, NV	73,269	1874
10	Southern Mojave-Salton Sea	1810	CA, MX	44,245	1028

4.2.1. Precipitation Validation

Figure 8 shows metrics that quantify the characteristics of daily, monthly,andwater-year precipitation totals, including: the full distribution of HUC-4 basin-scalewater-year precipitation totals (Figure 8a); inter-annual standard deviation of basin-scale water-year precipitation totals (Figure 8b); site-specific mean and standard deviation of thedaily precipitation timeseries at each grid cell within the HUC-4(Figure 8c,d);basin-scale water-year cumulative precipitation totals (Figure 8e);and the distribution ofmonthly precipitation totals (Figure 8f). The results show that the mean and standard deviation of daily precipitation is extremely well preserved (Figure 8c,d), and the full distribution and variability of water-year precipitation total falls well within the range of uncertainty for the observed statistics (Figure 8a,b).The mean, 10th, and 90th percentiles of basin-scale water-year cumulative precipitation totals also follows the observations very closely (Figure 8e), withonly a slight overestimation of the median cumulative precipitation that emerges during the spring.The median and range of monthly precipitation totals also matches the observations very closely (Figure 8f). Overall, these results show that the weather generator produces daily, monthly, and water-year precipitation totals that agree very well with observations, both at the HUC-4 scale and at the scale of individual grid cells.

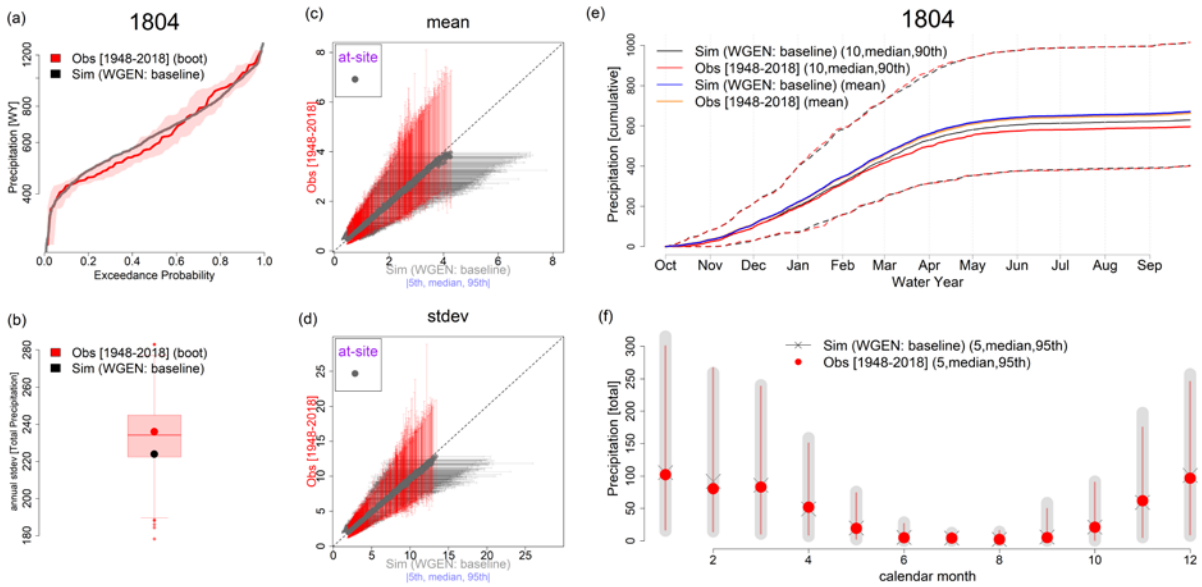

Figure 8. Observed vs. simulated **(a)**distribution of basin-scalewater-year precipitation totals, with the red shaded area representing the uncertainty in the observed distribution using a 95% confidence interval based on bootstrapping from the historical record;**(b)**standard deviation of basin-scale water-year precipitation totals, with the boxplot showing the uncertainty in the observed standard deviation based on bootstrapping from the historical record; **(c)**mean and **(d)**standard deviation of daily precipitation at individual sites (grid cells) within the HUC-4, with dots and whiskers showing the 50th, 5th, and 95th percentiles across individual water years; **(e)**basin-scale cumulative precipitation over the water year, including the mean, median, 10th, and 90th percentiles;**(f)**the distribution of basin-scale monthly total precipitation, including the 5th, 50th, and 95th percentiles. All results are shown for the San Joaquin basin (HUC4: 1804), and weather generator results are based on a baseline simulation of 1000 years with no climate change.

Figure 9presents statistics for extreme precipitation attributes.Figure 9a shows a return period plot of 1-day precipitation extremes at the HUC-4 scale from the observations, based on a GEV distribution fit to annual maxima (red) and a GPD-Poisson model fit to a partial duration series (blue). The black points show annual maxima from the 1000-year stochastic weather generator simulation, sorted and plotted against empirical return periods. The return levels estimated by the weather generator simulation match that of the GEV and GPD-Poisson models fit to the observations very well. Because the weather generator simulates weather using a block-bootstrap, the annual maxima from the weather generator exhibit a jump around a return period of 100 years, which matches the behavior in the observed annual maxima. However, the weather generator is then able to perturb these maxima even higher using the copula-based jittering algorithm (Appendix A.3), and does so in a way that the simulated annual maxima follow the observation-based GEV and GPD-Poisson model estimates for larger return periods (e.g., the 500-year and 1000-year events). We note that this behavior occurs even though the jittering algorithm is conducted at the grid cell (and not HUC-4) scale, and on daily data and not the annual maxima. The reproduction of 500-year and 1000-year extreme precipitation events at the HUC-4 scale is encouraging, as this requires that correlation of extremes be preserved across locations within the basin while still being perturbed upward above the range of the historical data.

Figure 9b shows 1-day precipitation extremes that fall over a high threshold (the 99th percentile) in the observations (y-axis), and also these same events that are bootstrapped in the 1000-year long weather generator simulation. Results here are shown for all grid cells within the HUC-4 region. For each event in the observations, the weather generator often resamples that event multiple times in the 1000-year simulation, which accounts for the scatter and horizontal stratified points in Figure 9b. The results here show that the simulated extremes are centered around and very highly correlated with the historical extremes (Pearson *r* of 0.99, as expected with the bootstrap), but that the simulation adds noise around these extremes with the copula-based jittering algorithm that allows both smaller and larger values than those seen historically. For instance, the maximum 1-day precipitation in the simulation is 517.2 mm, while it is only 429.8 mm in the observations.

Figure 9c shows the magnitude of the maximum precipitation total over the entire time period of the observations and the 1000-year simulation for events of different durations (maximum 7-day, 10-day, 1-month, 3-month, 6-month, 1-year event). The simulations track the observations well, and as expected, are always somewhat larger than the observations because they represent the maximum in a 1000-year (rather than 71-year) series. Finally, Figure 9dshows the distribution of averagewet spelllengths across gridded locations within the HUC-4 region. Across grid cells, average wet spell length matches that in the observations well, with only a slight (<1%) downward bias on average.

We can also evaluate the model's ability to reproduce dynamics within multi-day heavy precipitation events. At most rain gauges in the Western US, these events can be described by a trivariate distribution of the duration, maximum daily intensity, and total precipitation for individual events (similar to the model in Arendarczyk et al. (2018)). Here, events are defined by days with precipitation greater than the local 75^{th} percentile of above-trace (>0.3mm) precipitation. Figure 10 showsthe empirical distribution of event durations (Figure 10a), daily event maxima (Figure 10b), and event totals (Figure 10c), all for the TuolumneRiver watershed within the San Joaquin basin. Results are shown for watershed-average observations (red), as well as for 14 separate 72-year traces from the stochastic weather generator (blue; the 1000-year (technically 1008-year) simulation was separated into 14 equal length segments). The results inFigure 10 show that the weather generator is able to simulate realistic multi-day events in terms of event durations, event totals, and event daily maxima, suggesting that the output of the model is well suited to support the analysis of daily-scale flood risk that often depends on these multi-day events to initiative flood flows at the watershed scale.

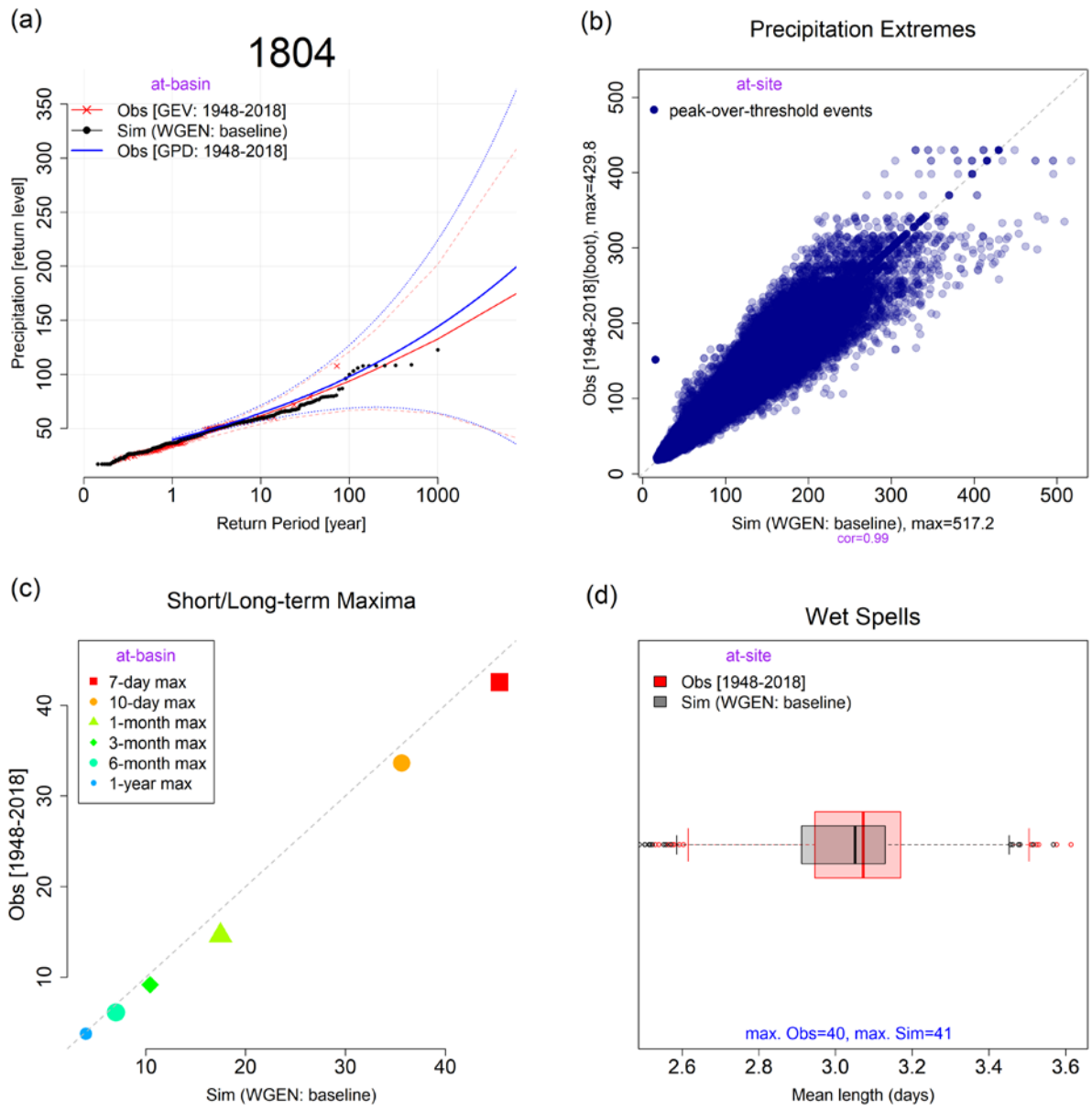

Figure 9.(a) Observed and simulated empirical distribution of HUC-4 scale water-year precipitation maxima, along withGEV andGPD-Poisson based return level estimates and their 95% uncertainty bounds fitted to the observed precipitation;**(b)**the magnitude of precipitation in partial duration seriesfor all grid cells in the HUC-4 region, along with the same events in the weather generator simulation after the values are jittered; **(c)**the maximum magnitude of total precipitation over multiple durationsin the observations and simulations at the HUC-4 scale; and **(d)**the distribution ofaverage wet spell lengths (in days) across sitesin theSan Joaquin (HUC4: 1804). The maximum wet spell length among grid cells for the observations and simulation is shown in blue text.

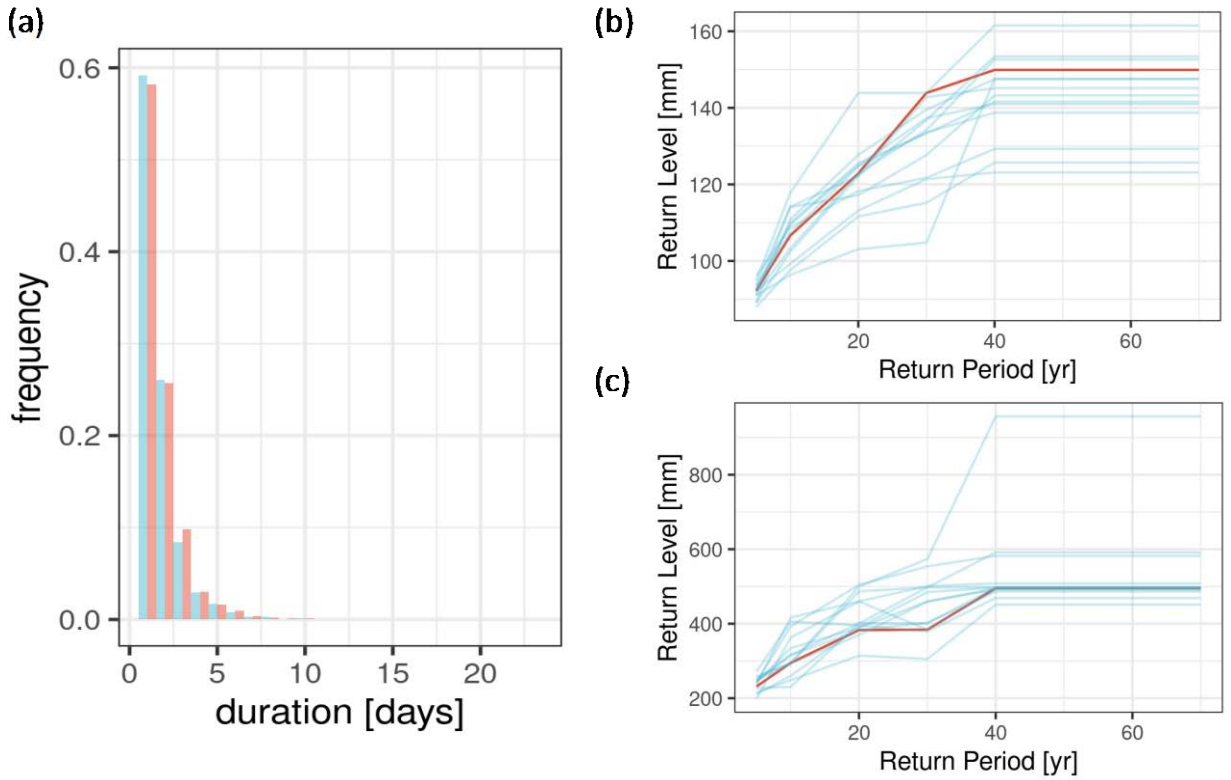

Figure 10.(a) Histogram of event durations from the observations (*red*) and weather generator simulation (*blue*) when precipitation exceeds the 75th percentile of above-trace (>0.3mm) observations. Note that the maximum event duration from the 1000-year weather generator simulation is 22 days. **(b)**Empirical return period plot of event maximum rates from the observations (*red*) and 14 separate 72-year weather generator traces (*blue*). **(c)**Same as (**b**), except for event totals. All results are shown for watershed-average precipitation in the Tuolumne River watershed within the San Joaquin basin (HUC4: 1804).

Figure 11focuses on metrics relevant to HUC-4 scale multi-year droughts and pluvials. Figure 11a shows the average daily precipitation during the worst 1-year, 2-year, through 10-year drought in both the observations and the 1000-year simulation. For all durations, the simulation exhibits a more severe drought-of-record than the historical period, as should be expected given the longer period of record. The weather generator produces average dry spell lengths across gridded locations within the HUC-4 that align very well with the observations (Figure 11b), but is also able to produce maximum dry spells that extend beyond the observations (i.e., maximum dry spell of 248days in the simulation vs. 217 days in the observations).Figure 11c shows the distribution of 1, 2, 3, 5, 10, and 30 water year rolling average precipitation totals, and demonstrates that the frequency of water-year precipitation totals over different multi-year durations is very well preserved in the simulation compared to the observations. However, the simulation is able to generatemulti-water year precipitation totals that extend beyond the observed range for all durations (i.e., more severe worst-case multi-year droughts and pluvials), showing how the model can be used to explore plausible extremes not yet experienced in the observations strictly due to internal climate variability.

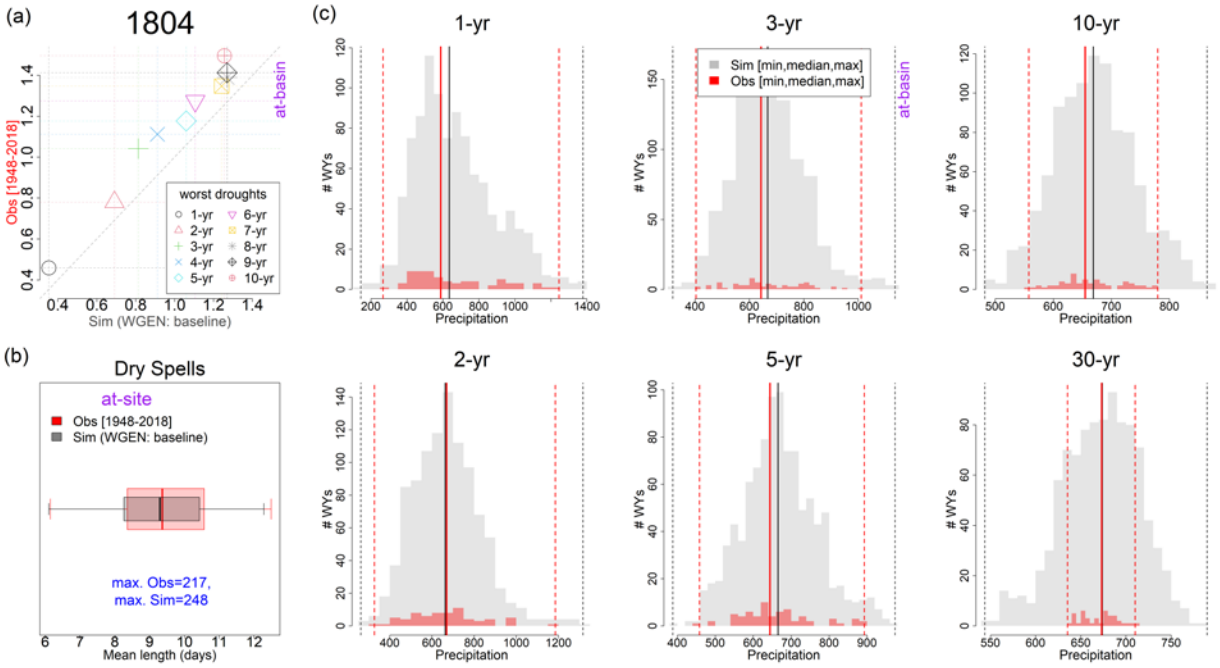

Figure 11.(a) Observed vs. simulated worst multi-year basin scale droughts (1 to 10 years); **(b)**boxplot exhibits the characteristics of at-site dry spells. **(c)**The histogram displays the distribution of basin scalewater-year precipitation totals at 1, 2, 3, 5, 10, and 30-year rolling averages,including the minimum, median, and maximum totals, all forthe San Joaquin basin (HUC4: 1804).

This last point is worth exploring further, as a major benefit of the weather generator is the production of long weather traces that represent plausible but previously unobserved extremes driven solely by climate variability (without any anthropogenically forced climate change). To further demonstrate this point, Figure 12 shows a 24-month standardized precipitation evapotranspiration index (SPEI) for the TuolumneRiver watershed within the San Joaquin basin, which is derived from both precipitation and temperature records (Vicente-Serrano et al., 2010). We observe realistic behavior of the SPEI series from the weather generator as compared to the SPEI calculated from observations (1948-2018). However, in just 150 years of the 1000-year weather generator baseline simulation, the modelcanalso produce prolonged and deep dry periods well beyond what has been observed historically. For example, the weather generator produces a very long drought between simulation year 110-120 and very deep droughts around simulation year 40 and 63 (note that ‘simulation years’ have no correspondence to actual calendar years, as the simulation is a randomly generated process). These results highlight how even in the absence of climate change, water managers in California would benefit fromplanning for extreme droughts beyond the worst case from the historical record.

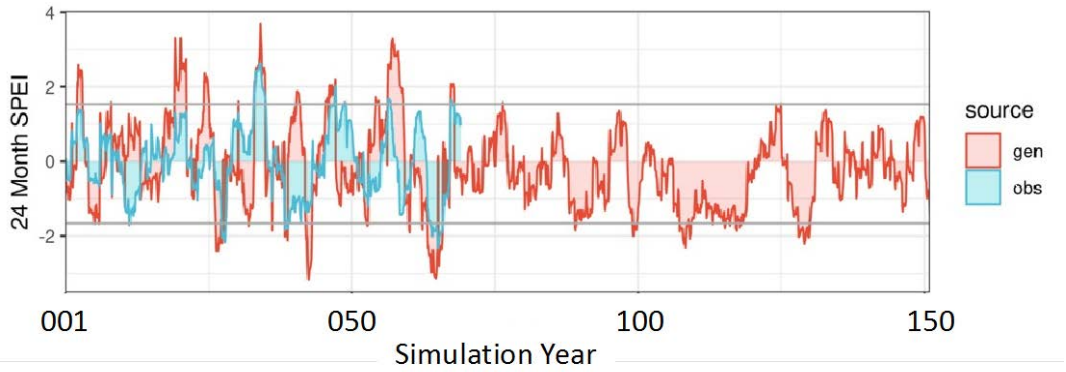

Figure 12. A drought index (24-month SPEI) calculated from watershed-average precipitation and temperature for the observations between 1948-2018 (*blue*) and the first 150 years of the weather generator simulation (*red*) over the Tuolumne River watershed within the San Joaquin basin (HUC4: 1804). Horizontal lines shown observed upper and lower decile limits from the weather generator output, which is an example of a drought/pluvial threshold pair for studying sequences of SPEI.

Yet another way to evaluate the benefit of the weather generator in terms of producing plausible extreme events is to compare the baseline simulation to data from other climate products, like downscaled data from the most recent generation of CMIP6 climate models. We obtained daily precipitation simulated for a historical period of 1950-2014 from 67 individualCMIP6 climate model simulations, downscaled to a 3-km resolution with the new LOCA v.2 downscaling procedure documented inPierce et al. (2023). There are 13 separate models in this dataset, each run with some number of initial conditions that varies by model, with a total of 4355 years of data in the historical period across all 67 simulations. We focus our investigation on basin-average precipitation over the Upper Tuolumne watershed, one of the watersheds in the San-Joaquin basin (HUC4: 1804), and retain a focus on the historical period in the CMIP6 data in order to compare how these model simulations represent extremes due to natural climate variability without climate change. Similar to Figure 9a,Figure 13ashows a return period plot of 1-day precipitation extremes at the watershed scale from the observations, based on a GEV distribution fit to annual maxima (red). The black points show annual maxima from the 1000-year stochastic weather generator simulation, sorted and plotted against empirical return periods. The blue triangles show a similar result for the LOCA v.2 downscaled data, where the annual maxima at the watershed scale are concatenated across the 67 model simulations, sorted, and plotted against empirical return periods. The results show that while the 1000-year weather generator simulation of watershed-scale annual maxima follow the observation-based GEV model very well, the LOCA v.2 data are biased low for return periods greater than 10 years. This bias is especially apparent at the higher return periods.

Figure 13b focuses instead on drought events, and shows the worst 1-, 2-, 3-, 4-, and 5-year drought events in the observed record, the concatenated 4355 historical years from LOCA v.2, and the 1000-year weather generator simulation. Interestingly, both the weather generator and the downscaled climate model data produce worst-case droughts that 1) fall well below the worst-case drought in the observational record, and 2) are of very similar magnitude for all durations. This is particularly surprisingbecause these products were developed independently, where the LOCA v.2 data are derived from physically based GCMs and the weather generator is

a statistical model fit to observations. However, it is worthwhile to note that although the most extreme droughts in both the weather generator simulation and the downscaled (historical) climate model data are of similar magnitudes, it took 4355 years of downscaled GCM data to produce these droughts but only 1000 years of the weather generator simulation. This leads to two important implications: 1) the extreme droughts simulated by the weather generator are physically plausible, because an ensemble of GCM simulations produced droughts of similar magnitude; and 2) those droughts may be much more likely to occur due to natural climate variability than suggested by the GCMs, since the weather generator was able to produce them in less than a quarter of the years required to produce them in the GCMs. We note that we repeated this analysis using only 1000 years from the LOCA v.2 dataset and found worst-case droughts tobe of the same magnitude as the observations (not shown), further suggesting that the GCMs struggle to produce extreme droughts that are driven by natural climate variability.

We repeated a similar comparison of the LOCA v.2 historical data to the baseline weather generator simulation for additional watersheds (e.g., Upper Feather, Upper American, Lake Millerton), and found that in general the 1000-year baseline weather generator simulation was better able to capture basin-scale extreme precipitation frequencies and produced similar extreme droughts compared to the 4355 years of historical LOCA v.2, although with some variation across watersheds(see Supplementary Material).

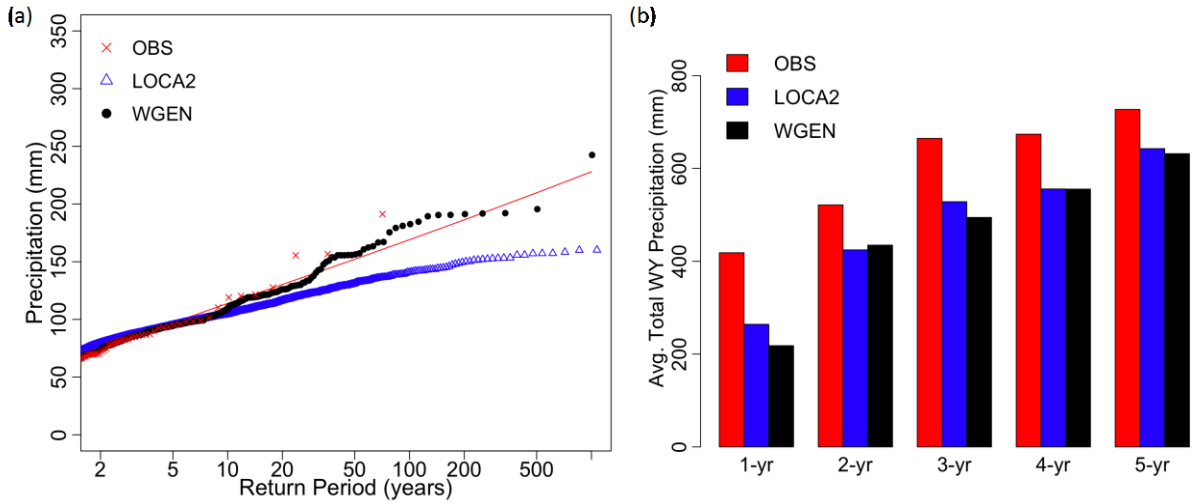

Figure 13.(a) Observed distribution of watershed-scale annual precipitation maxima, along withGEV-based return level estimates fitted to the observed annual maxima (*red*). Also shown are annual maxima from the 1000-year weather generator simulation and 4355 years of historical LOCA v.2 data. **(b)** The worst 1, 2, 3, 4, and 5-year drought events in the observed record, across the historical LOCA v.2 ensemble, and in the 1000-year weather generator simulation. All results are shown for the Upper Tuolumne River basin.

4.2.2. Temperature Validation

Figure 14is similar to Figure 8a-d, but for average temperature. It should be noted that we computed average temperature using simulated daily minimum and maximum temperature at each gridded location. Overall, the distribution of HUC-4 scale average annual temperatures

across water-years is well preserved in the simulation compared to the observations (Figure 14a), although with a small underestimation of the inter-annual variance (Figure 14b). Median daily means and standard deviations across grid cells are very well preserved, as are their range across water years (Figure 14c,d).

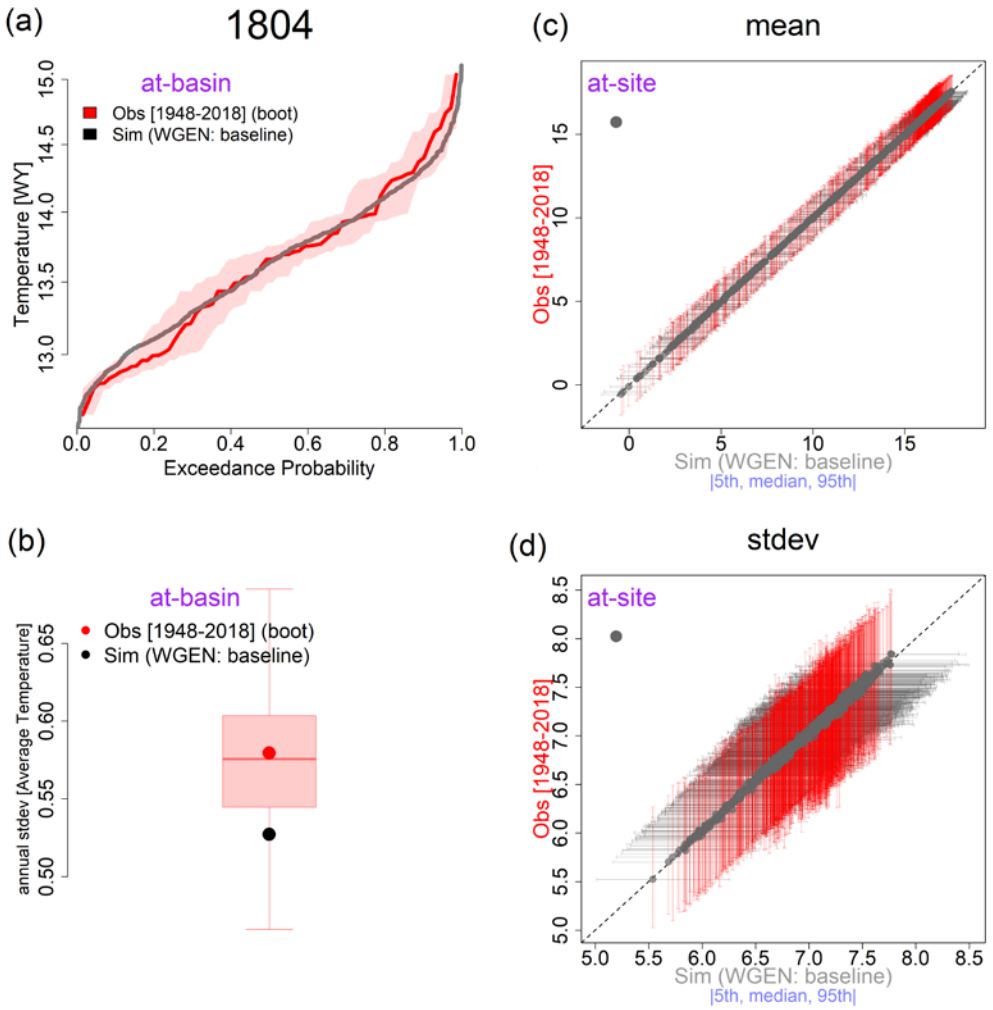

Figure 14. Observed vs. simulated **(a)** distribution of basin-scale water-year average temperature, with the red shaded area representing the uncertainty in the observed distribution using a 95% confidence interval based on bootstrapping from the historical record; **(b)** standard deviation of basin-scale water-year average temperature, with the boxplot showing the uncertainty in the observed standard deviation based on bootstrapping from the historical record; **(c)** mean and **(d)** standard deviation of daily average temperature at individual sites (grid cells) within the San Joaquin (HUC4: 1804), with dots and whiskers showing the 50th, 5th, and 95th percentiles across individual water years. Note that average temperature is calculated based on the minimum and maximum temperature time series.

Figure 15 highlights characteristics related to temperature extremes across grid cells within the HUC-4 region, including heatwaves and coldwaves. The average and maximum duration of heat and cold waves are shown in Figure 15a,b, while the average and maximum intensity are shown in Figure 15c,d. The model performs well for allthese statistics across sites. There is a minor downward bias in the mean duration and mean and maximum intensity of heat waves, but this

bias is quite small. Overall, there is a high level of agreement between the observed and simulated attributes of heat waves and cold waves.

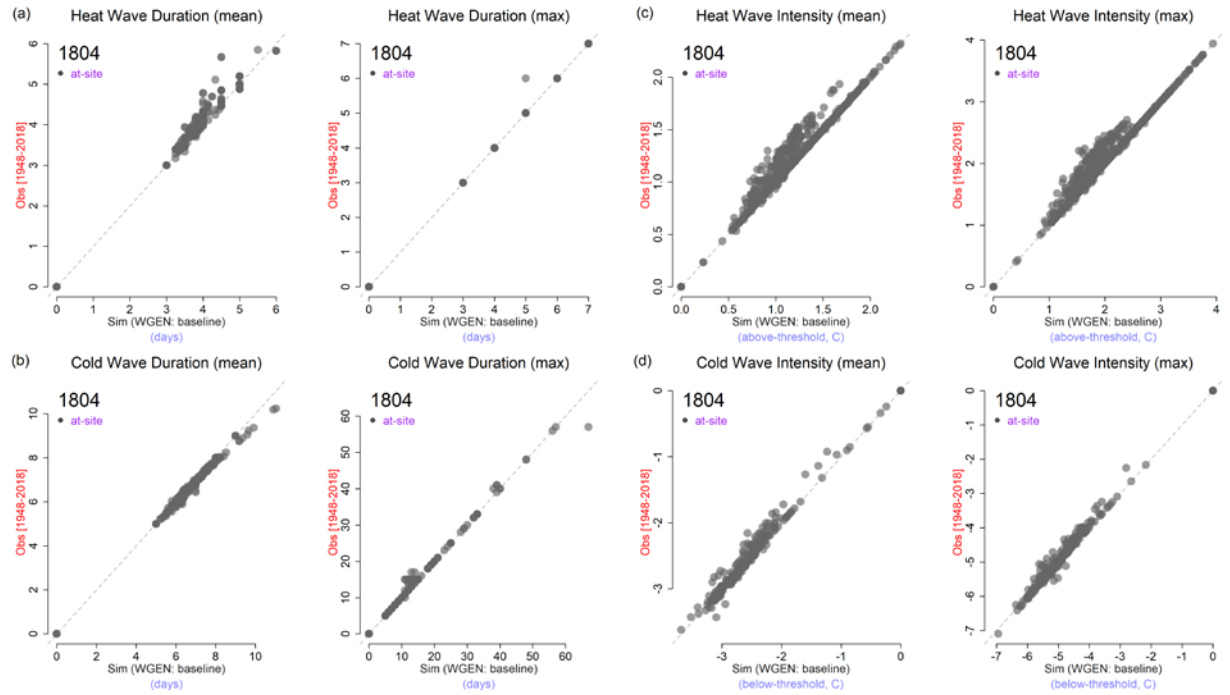

Figure 15. Observed vs. simulated at-site **(a)**duration of heat wave (mean, maximum); **(b)**duration of cold wave (mean, maximum); **(c)** intensity of heat wave (mean, maximum), and **(d)** intensity of cold wave (mean, maximum). Each point represents a gridded location inthe San Joaquin basin (HUC4: 1804).

Finally, Figure 16 shows the spatial distribution of the frequency of heat waves and heat stressevents derived from the observations (Figure 16a,c) and the 1000-year simulation (Figure 16b,d)across the HUC-4 region. A similar result is also shown for cold waves and cold stress events in the observations (Figure 16e,g) and simulation (Figure 16f,h). The distributions are normalized using the site-specific time-series length,so that results are shown as theaverage frequency of heat or cold events per year in each gridded location. As shown in Figure 16, the model is able to reproduce the spatial distribution of these events extremely well as compared to the observations, implying that the stochastic weather generation can fully preserve the spatial organization of multi-day temperature extremes across sitesin the basin.

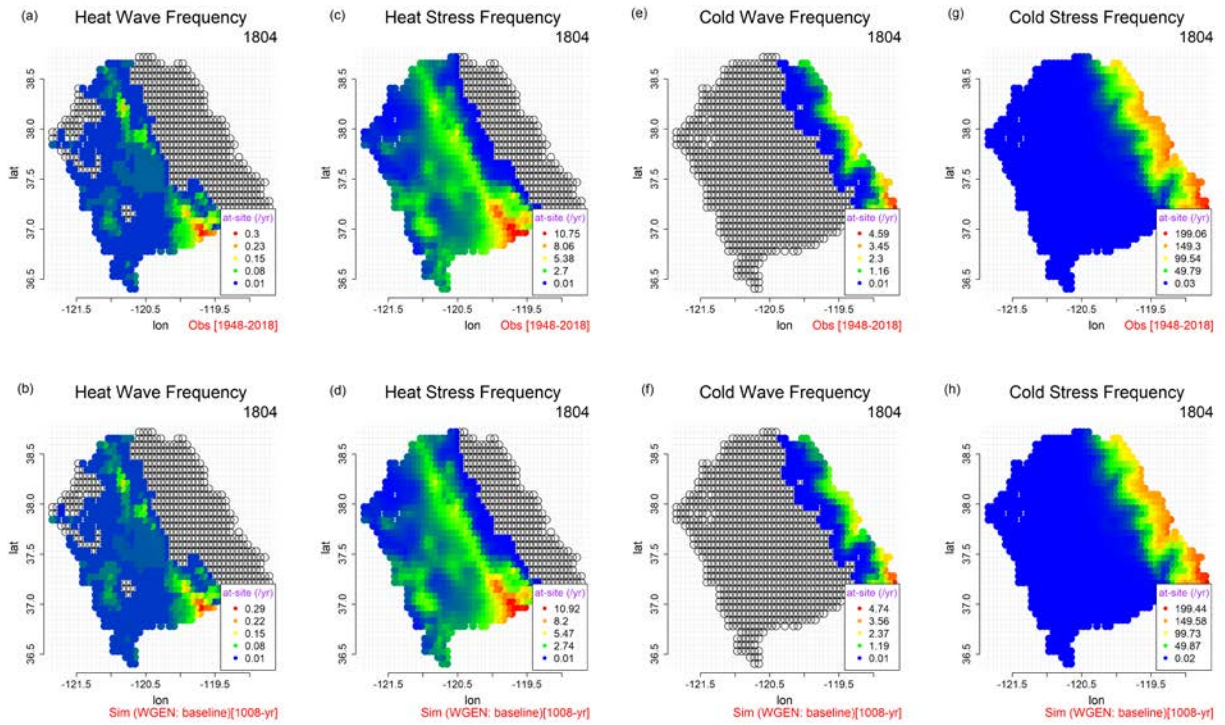

Figure 16. Frequency of **(a)** observed and**(b)** simulated heat waves; frequency of **(c)** observed and**(d)** simulated heat stress events; frequency of **(e)** observed and **(f)** simulated cold waves; frequency of **(g)** observed and **(h)** simulated cold stress events. All results shown for the San Joaquin basin (HUC4: 1804).

4.2.3. AR Landfall Frequency Validation

As a final validation of the 1000-year baseline simulation of the stochastic weather generator with no climate change, we evaluate the frequency of AR landfalls near and along the California coastline (Figure 17). The observed frequency, expressed as the number of events per year for each calendar month, were taken from the historical AR archive in Gershunov et al. (2017). We focus on five locations and observed that for the most northern location (labeled #1 in Figure 17), AR landfalls are most frequent in the earlier part of the cold season (November-December) and become less frequent as the cold season progresses. Conversely, the location furthest south (labeled #5 in Figure 17) shows an increase frequency of ARs later in the cold season (January-March). This general pattern of more ARs later in the cold season is followed across locations moving north to south.

By using the resampled historical dates from the block bootstrap in the stochastic weather generator, we can also derive the monthly frequencies of landfalling ARs in the 1000-year baseline simulation. We find that the weather generator near-perfectly reproduces the monthly frequencies of AR landfalls at all locations. We also note that given the block bootstrap method used to generate weather across the state, it is not possible for the weather generator to (incorrectly) produce AR landfalls at both the southern and northern ends of the domain on the same day.

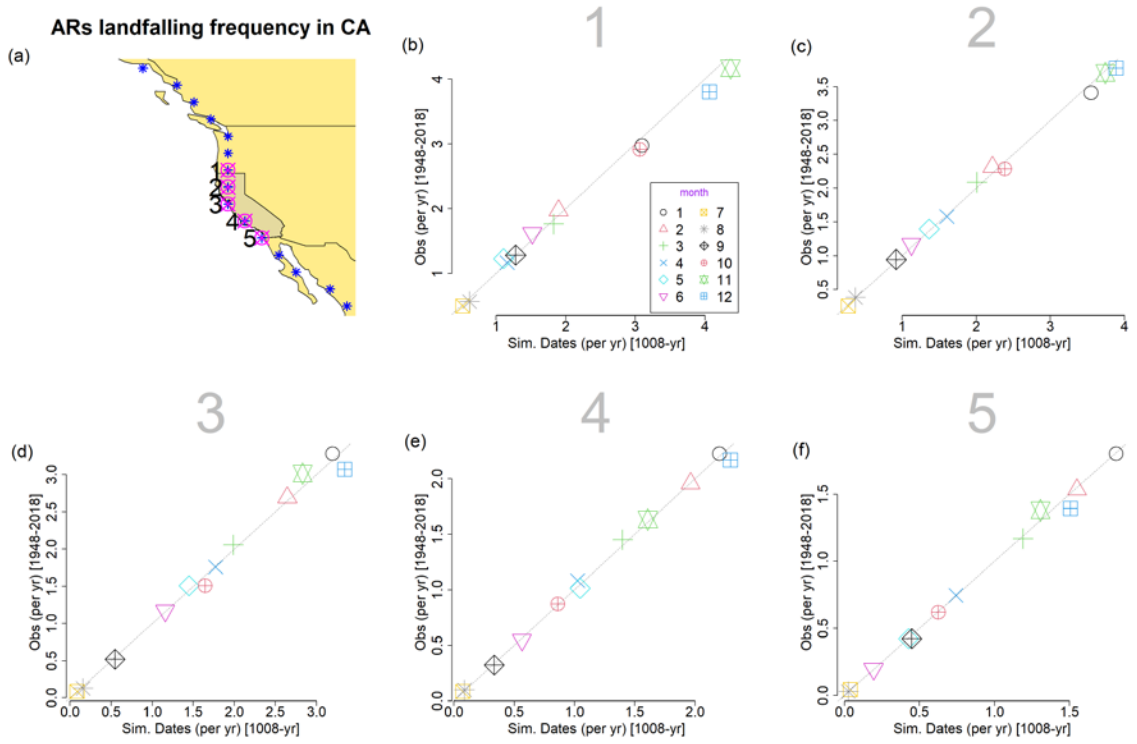

Figure 17.(a) Location of AR landfalls along Western North America (*blue*), with 5 locations near and along the California coastline highlighted (*pink*). **(b-f)** Observed versus weather generator simulated frequency of AR landfalls, expressed as an average number of days per year for each calendar month, across five locations near and along the California coastline.

5. Future Climate Scenarios

In Section 5.1, we demonstrate how extremes (droughts, extreme precipitation events) vary across the climate scenarios delimited in Figure 4 and Table 3 above. In Section 5.2, we briefly assess how these weather generator scenarios compare against projections from the recent LOCA v.2 ensemble of future projections. In Section 5.3, we explore the impacts associated with a small set of dynamic climate change scenarios in which the probabilities of individual WRs are changed based on recent trends.

5.1. Thermodynamic Climate Scenario Impacts

To begin, we first examine the spatial distribution of impacts from a subset of the 30 scenariosin Figure 4 and Table 3. Figure 18 focuses on the largest maximum 1-day precipitation event across the entire record, and shows results both for the Historical Climate Change Ensemble (Figure 18a) and the Stochastic Climate Change Ensemble (Figure 18b). The far left panel in Figure 18a,b shows the magnitude of the 1-day precipitation maximum across the baseline scenario (no climate change) for these two datasets. The only difference is that the Stochastic Climate Change Ensemble is derived from a 1000-year weather generator simulation, while the Historical Climate Change Ensemble baseline is based on the historical record. By comparing the left panels of Figures 18a,b, it is clear that the 1-day maximum in the longer stochastic dataset is

larger than in the historical record (darker shading throughout the state), which is expected given the longer period of record over which extremes can occur.

The middle and right panels of Figure 18a,b show how the largest 1-day precipitation event changes under different climate scenarios. Results in the center panels are shown for scenario #13 from Table 3, in which there is 3°C of warming, an extreme precipitation scaling rate of 7% per °C, and 0% change in mean precipitation. Here, the increase in the 1-day precipitation maxima is largely uniform across the state, around 22.5% above baseline values. This is precisely the scaling one would expect given that 1.07^3=1.225. Conversely, in the right panels of Figure 18a,b, there is almost no change in the 1-day precipitation maxima over baseline levels. Under this scenario (#27 in Table 3, with 3°C of warming, no extreme precipitation scaling, and a 12.5% increase in mean precipitation), mean precipitation increases but extreme precipitation is kept at baseline levels by construction. Therefore, it is the smaller precipitation events under this scenario (rather than the largest) that are shifted upward to account for the change in the mean (not shown here). We note that in both scenarios, the amount of change is largely the same between the Historical and Stochastic Climate Change Ensemble, and it is just the baseline values that differ.

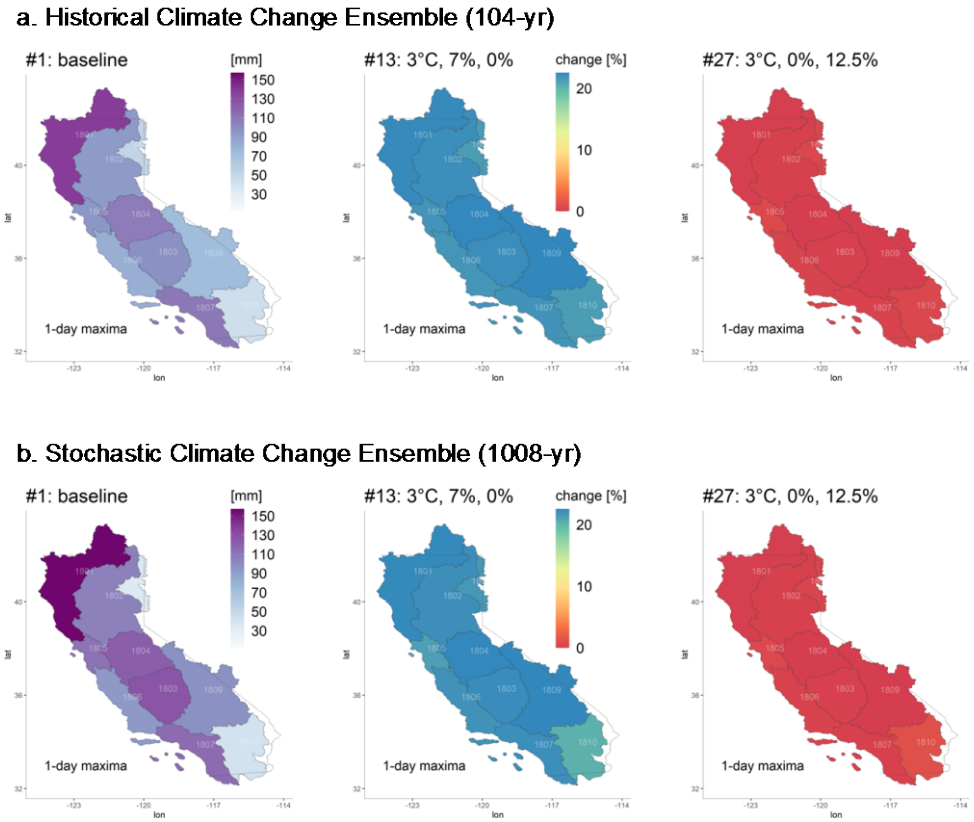

Figure 18. (*left*) The magnitude of the largest 1-day maximum precipitation event at the HUC-4 scale in the baseline scenario (no climate change) of the (**a**) Historical Climate Change Ensemble and the (**b**) Stochastic Climate Change Ensemble. (*middle*) The percent change in the largest 1-day maximum precipitation event from climate scenario #13 (see Table 3) with 3°C of warming, an extreme precipitation scaling rate of 7% per °C, and 0% change in mean precipitation. (*right*) Same as middle panel, but for climate scenario #27 (see Table 3) with 3°C of warming, an extreme precipitation scaling

rate of 0% per °C, and a 12.5% increase in mean precipitation.Note that the color scale is identical for scenarios #13 and #27.

Figure 19 shows similar results to Figure 18 but for the worst 5-year drought on record. Again, the Stochastic Climate Change Ensemble exhibits a deeper worst-case 5-year drought than the Historical Climate Change Ensemble for the baseline scenario (also see in Figure 11 above). Here, we show three climate scenarios in the remaining panels in Figure 19a,b, including scenarios #13, #8, and #18 from Table 3. Under all of these scenarios, temperatures warm by 3°C and extreme precipitation scales at a rate of 7% per °C. The only difference between the scenarios is how mean precipitation changes (0%, -12.5%, and +12.5% of baseline values). Two insights emerge from Figure 19. First, the largest impact on the magnitude of the worst 5-year drought is caused by the change in mean precipitation, with the magnitude of change roughly following the magnitude of change in mean precipitation (on a percentage basis). Second, we note that even under the scenario with no change in mean precipitation (scenario #13), the worst 5-year drought does become slightly worse (by ~5%) simply due to extreme precipitation scaling. In this scenario, light precipitation events must become even lighter to keep mean precipitation unchanged while increasing the magnitude of extreme precipitation events (i.e., the daily precipitation distribution is stretched; see Figure 5). During droughts, there is a very high proportion of dry days and light precipitation days, and so the total precipitation during these long drought events declines because of the downward shift in light precipitation events.

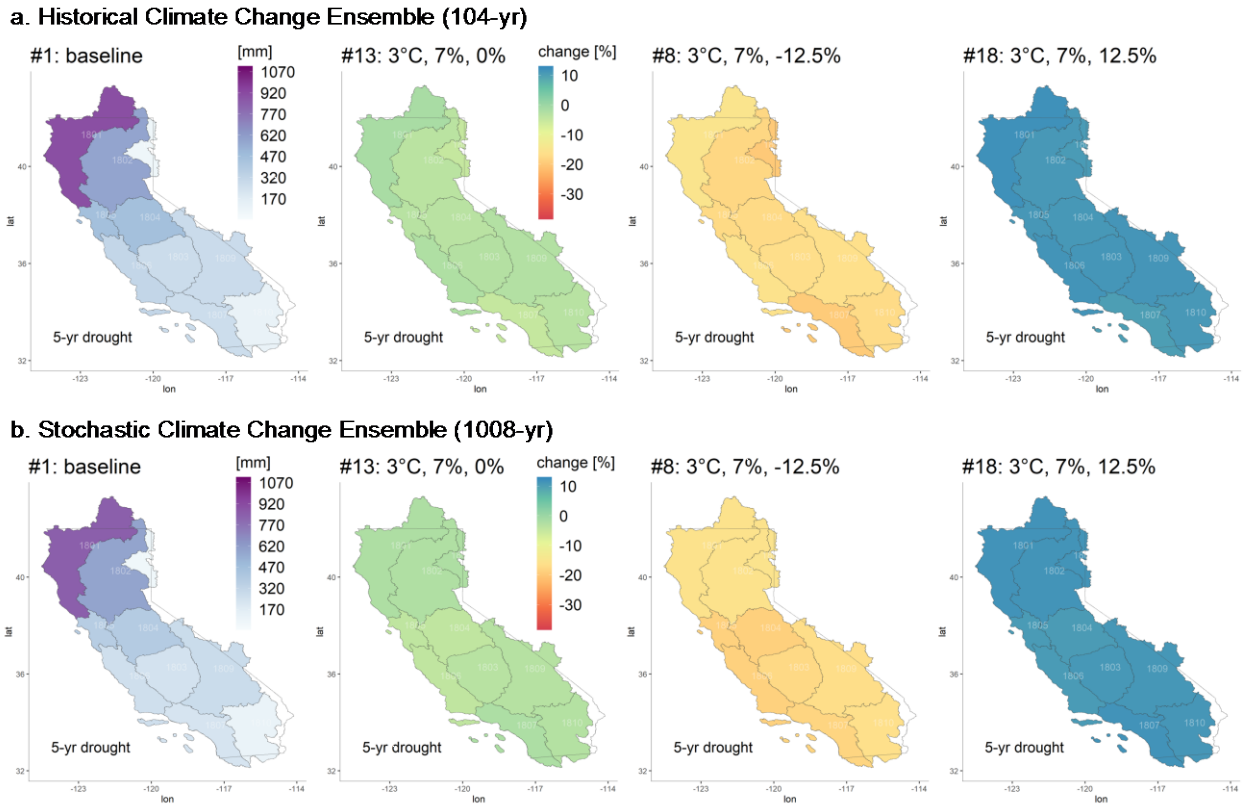

Figure 19. Same as **Figure 18**, but for the worst 5-year drought on record. Results are shown for three climate scenarios from Table 3, including scenarios #13, #8, and #18. In all cases, temperatures warm by 3°C and extreme precipitation scales at a rate of 7% per °C. However, scenarios #13, #8, and #18 differ

by their change in mean precipitation (0%, -12.5%, and +12.5% of baseline values, respectively).Note that the color scale is identical for scenarios #13, #8, and #18.

Similar to Figures18-19for precipitation, Figures 20-21 show the impacts of the thermodynamic climate scenarios onthe spatial distribution of heat wave characteristics, this time focused on one HUC-4 basin(the San Joaquin) for demonstration. Figure 20a,b shows the average frequency of heat waves per year based on the baseline and three climate scenarios (#11, #12, and #13 from Table 3) from theHistorical and Stochastic Climate Change Ensembles. We also show the average and maximum heat wave event duration only based on the Stochastic Climate Change Ensemble across these three scenarios in Figure 21. These selected scenarios have a similar extreme precipitation scaling rate of 7% per °C and 0% changes in mean precipitation. The only difference is how temperature (minimum, maximum) increases incrementally by 1, 2, and 3°C. In Figure 20, there is a consistent increase in the average frequency of heat waveswith warming across the scenarios. Furthermore, a more pronounced increase can be detected in the number of heat wave events over the southern portion of the basin across both datasets. Changes to heat wave frequency with warming are very similar between the Historical and Stochastic Climate Change Ensemble, though the Historical Climate Change Ensembleexhibits a slightly higher baseline average frequency.The average and maximum heat wave event duration presented in Figure 21shows how even 1°C of warming can result in multi-day increases in both average and maximum heat wave duration at individual grid cells. The impacts are especially prominent on the maximum duration of heat waves, which can increase by more than 15 days per year for a scenario with 3°C of warming.

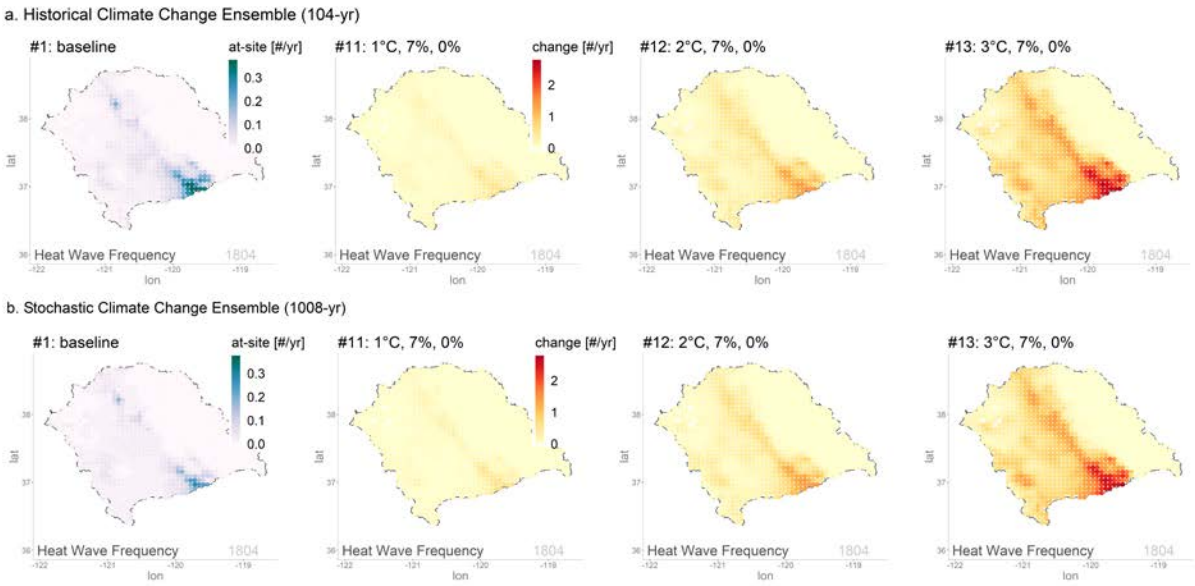

Figure 20. At-site average frequency of heat wave eventper year [number of events per year] in the baseline scenario (no climate change) of the (**a**) Historical Climate Change Ensemble and the (**b**) Stochastic Climate Change Ensemble. The change in the average frequency of heat wave event (number of additional events per year) are shown for three climate scenarios from Table 3, including scenarios #11, #12, and #13 with respectively 1, 2, and 3°C of warming, an extreme precipitation scaling rate of 7% per °C, and 0% change in mean precipitation. Note that the color scale is identical for scenarios #11-13. All results shown for the San Joaquin basin (HUC4: 1804).

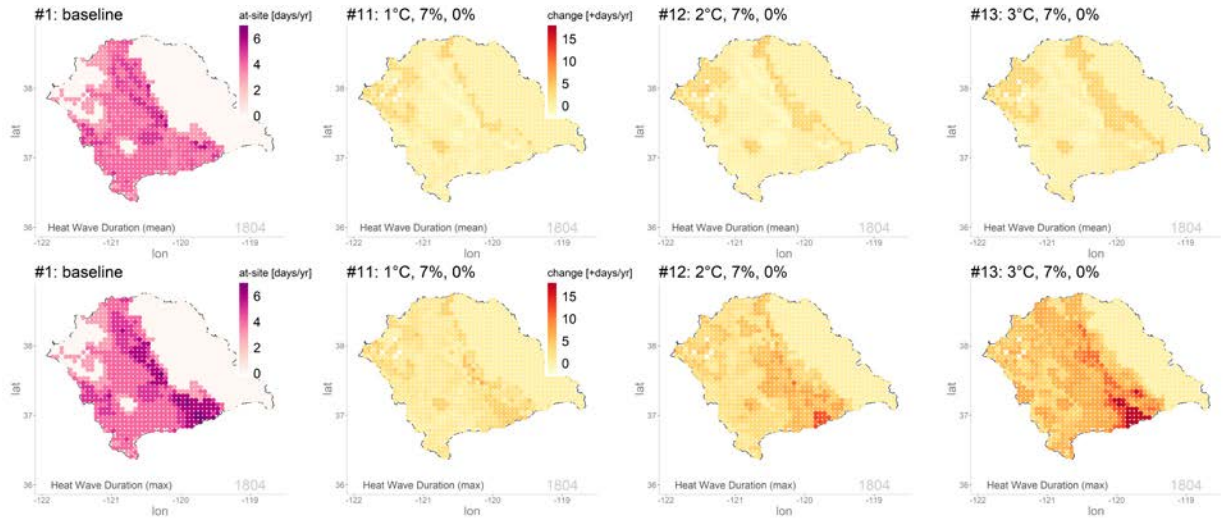

Figure 21. At-site mean (*first row*) and maximum (*second row*) heat wave event durationper year [daysof event per year] in the baseline scenario (no climate change) of the Stochastic Climate Change Ensemble are provided in far-left panels. The change in the average and maximum heat wave event duration (longer event duration per year) are shown for three climate scenarios from Table 3, including scenarios #11, #12, and #13 with respectively 1, 2, and 3°C of warming, an extreme precipitation scaling rate of 7% per °C, and 0% change in mean precipitation. Note that the color scale is identical for scenarios #11-13. All results shown for the San Joaquin basin (HUC4: 1804).

Figures 22-23 show how extreme precipitation and drought events vary across all of the climate scenarios in Figure 4 and Table 3, focusing only on the Stochastic Climate Change Ensemble and one HUC-4 region (the San Joaquin basin (HUC4: 1804)). In Figure 22, we show how the 1-day and the 7-day largest precipitation events change with shifts in temperature and mean precipitation, all for an extreme precipitation scaling rate of 7% per °C. Similar to Figure 18, we find that the 1-day precipitation maxima are largely insensitive to changes in mean precipitation, and instead only respond to changes in temperature that drive the absolute magnitude of extreme precipitation scaling. While temperature changes and extreme precipitation scaling also dominate the rate of increase in 7-day maxima, changes in mean precipitation also have an effect, largely based on how they influence the smaller precipitation values during the 7-day events. For the most intense scenario of change (5°C of warming and a 25% increase in mean precipitation), the 7-day maxima increase by almost 38% over baseline values.

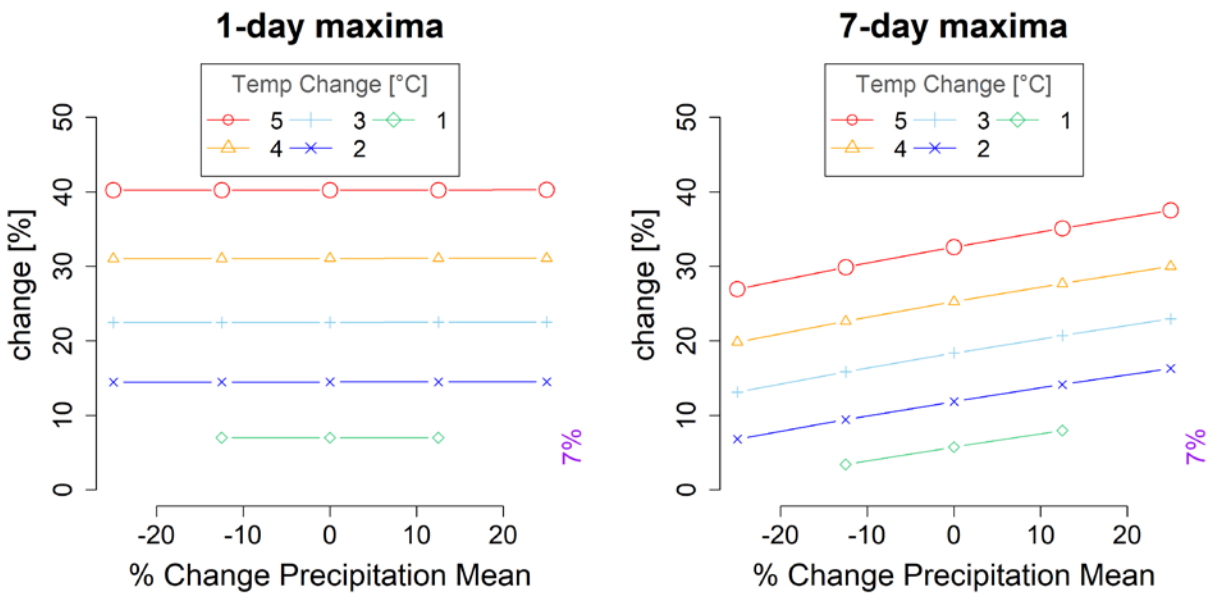

Figure 22. Percent change from baseline for the largest (*left*) 1-day and (*right*) 7-day precipitation maximum in the Stochastic Climate Change Ensemble. Changes are shown for 5 different scenarios of temperature change and 5 different scenarios of precipitation change, all with an extreme precipitation scaling rate of 7% per °C and for the San Joaquin basin (HUC4: 1804).

Figure 23 shows results that are similar to Figure 22, but for the 20-year, 100-year, and 1000-year 1-day precipitation event rather than the largest precipitation event on record. These design events are estimated by fitting a GEV distribution to the annual maxima of the entire 1000-year record of each climate scenario. The overall controls over these design events are similar to that seen in Figure 22, i.e., temperature increases along with extreme precipitation scaling dominate the changes experienced. However, an interesting secondary result also emerges, whereby increases in mean precipitation have a small positive impact on smaller return period events (the 20-year storm), but for the larger return periods (100-year, 1000-year), larger increases in mean precipitation cause a decline in these extremes. This seemingly non-intuitive result is caused by the fact that as mean precipitation decreases but extreme precipitation continues to scale at the same rate with warming, the daily precipitation distribution is forced to stretch further so that increased extremes at the upper end of the distribution are counteracted by larger declines in moderate precipitation events at the middle and lower end of the distribution,thereby leading to an overall reduction inthe mean of the distribution (also see Figure 5). As the distribution stretches further and further, the slope of extreme precipitation events against return period grows, leading to larger estimates of the longest return period events. However, this effect at the longest return periods is really a statistical artifact of the fitting process for the GEV distribution, and it does not imply that the largest simulated extremes in the weather generator dataset are smaller when mean precipitation is increased versus decreased.

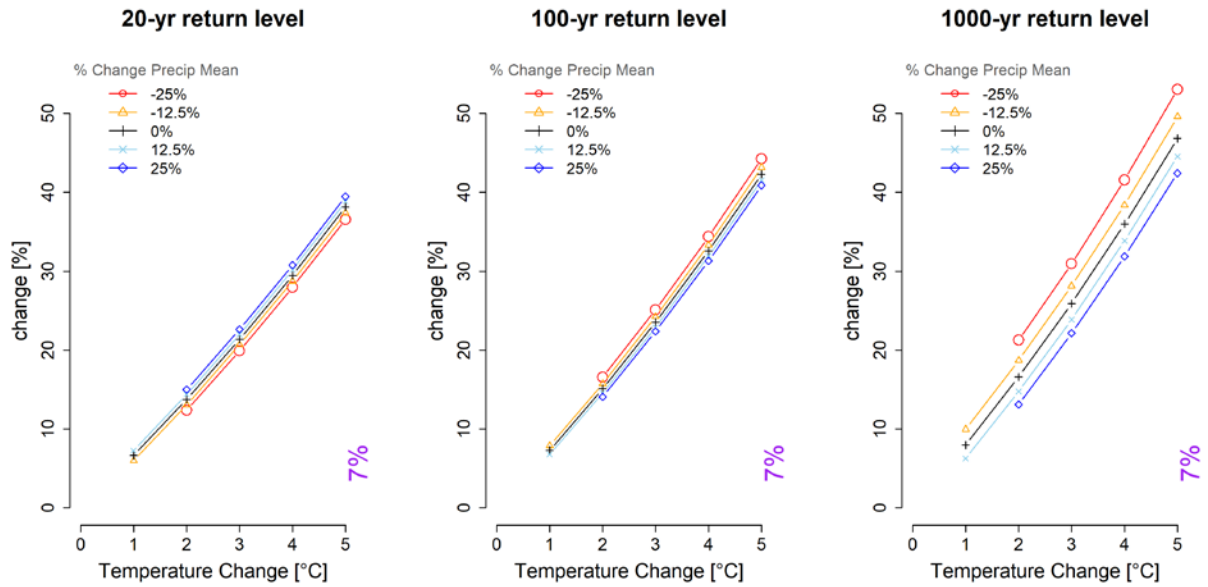

Figure 23. Percent change from baseline for the (*left*) 1-day, 20-year event, (*middle*) 1-day, 100-year event, and (*right*) 1-day, 1000-year event in the Stochastic Climate Change Ensemble. Changes are shown for 5 different scenarios of temperature change and 5 different scenarios of precipitation change, all with an extreme precipitation scaling rate of 7% per °C and for the San Joaquin basin (HUC4: 1804).

We focus on drought-of-record events of different duration in Figure 24. Here, the results show that changes in mean precipitation dominate the change in drought magnitude for all durations, and the relationship is close to 1-to-1 (i.e., a similar percent change in mean precipitation is also seen for the change in drought magnitude). Interestingly, there is also a pronounced temperature effect on the magnitude for 1-year droughts, but this effect weakens for longer durations. This is caused by the same phenomenon described above, where for some change in mean precipitation, extreme precipitation scaling requires that lighter precipitation events be scaled downward (i.e., become even lighter) to compensate for increases in the extremes (also see Figure 5). For 1-year droughts, this effect is more prominent, likely because 1-year droughts can occur with few or no heavy precipitation events, and so the downward scaling is experienced across most or all days in the 1-year drought. However, as drought durations grow longer, this temperature effect is dampened as more heavy and extreme precipitation events that have scaled upward are included in the long-duration drought periods.

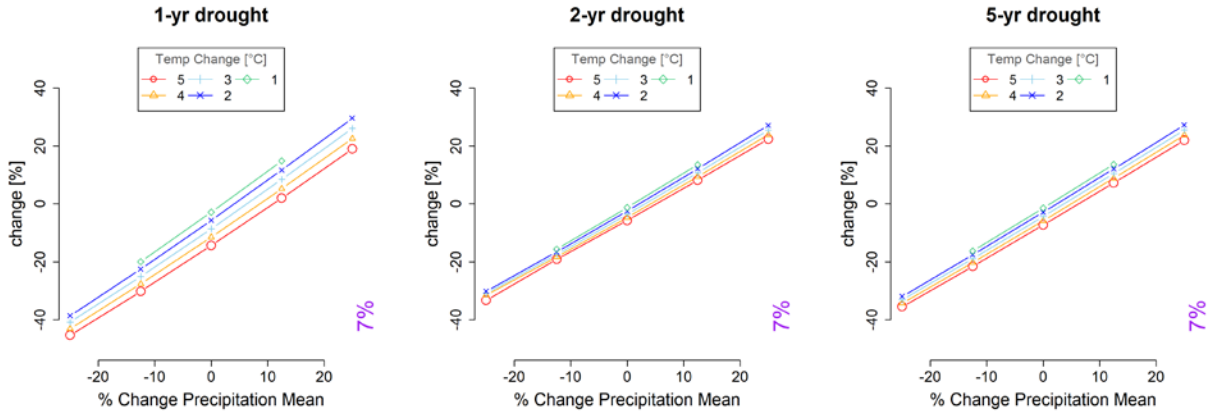

Figure 24. Percent change from baseline for the worst (*left*) 1-year drought, (*middle*) 2-year drought, and (*right*) 5-year drought in the Stochastic Climate Change Ensemble. Changes are shown for 5 different scenarios of temperature change and 5 different scenarios of precipitation change, all with an extreme precipitation scaling rate of 7% per °C and for the San Joaquin basin (HUC4: 1804).

5.2. Comparison against LOCA v.2 Projections

In this section we provide a brief comparison between some of the thermodynamic climate scenarios generated by the weather generator and the ensemble of future projections in the LOCA v.2 archive. We provided a similar comparison above in Figure 13 for the historical period (1950-2014) from the LOCA v.2 archive, but now focus on future projections from LOCA v.2. To facilitate this comparison, we first took each climate model simulation from LOCA v.2 and identified the 30-year period in the future that exhibited an average of 2°C of warming over the baseline period of 1981-2010 in the simulation for a region over the entire Central Valley. Using those 30 years for each simulation, we then found water year 1-day precipitation maxima and concatenated these data across all model simulations (for a total of 3750 years of LOCA v.2 data). Similarly, we found the annual maxima from the 1000-year stochastic weather generator simulation associated with climate scenario #12 in Table 3 (2°C warming, 0% change in mean precipitation, and 7% per °C scaling of extreme precipitation). Similar to Figure 13a,Figure 25a shows a return period plot of 1-day precipitation extremes at the watershed scale from the historical observations, based on a GEV distribution fit to annual maxima (red). The black points show annual maxima from the 1000-year stochastic weather generator simulation under the 2°Cwarming scenario, sorted and plotted against empirical return periods, whilethe blue triangles show a similar result for the LOCA v.2 downscaled data associated with 2°C warming on average. As in Figure 13, this comparison is focused on the Upper Tuolumne River basin.

The results show that under 2°C of warming and 7% per °Cscaling for extreme precipitation,the 1000-year weather generator simulation now produces watershed-scale annual maxima that exceed the observation-based GEV model. Extreme precipitation at the watershed scale that used to be associated with a 100-year return period under the observations becomes associated with a 30-year return period under the weather generator simulation. We also see that the LOCA v.2 data are shifted upwards as compared to their historical distribution (see Figure 13a). However, because LOCA v.2 has watershed-scale annual maxima that are biased low historically, the

future projections from LOCA v.2 centered around 2°C of warminglead to extremes that are aligned with (rather than exceed) the historical observations. That is, the intensification of watershed-scale extreme precipitation in LOCA v.2 under 2°C of warming is counteracted by a low bias under historical conditions, leading to very little change compared to the observed distribution of extremes.

Like Figure 13b, Figure 25b focuses on the worst 1-, 2-, 3-, 4-, and 5-year drought events in the observed record, the LOCA v.2 ensemble associated with 2°C of warming and the 1000-year weather generator simulation associated with climate scenario #12 from Table 3. The results here look very similar to those in Figure 13b. There are some small differences between future (Figure 25b) and baseline (Figure 13b) drought events, with slightly more intense droughts seen in the 2°C warmingscenario for all durationsin both the weather generator simulation and LOCA v.2. Drought intensification is likely linked to the suppression of low and moderate precipitation events under thermodynamic climate change during years that were already dry, as described earlier. For LOCA v.2, declines in the 2-year drought are more severe under 2°C warming compared to the other durations, but this is likely due to natural variability in the ensemble. Overall, the differences in drought intensity between the baseline and future 2°C warming are relatively small and consistent across both climate products.We note that drought here is being defined only on precipitation, and so does not account for potential drying linked to higher temperatures. In addition, it is worthwhile to note that had we examined the 1000-year weather generator simulation associated with a different climate scenario (e.g., -12.5% change in mean precipitation; scenario #7 from Table 3), many of the most extreme droughts seen in Figure 25b for the weather generator would extend beyond the range observed in the LOCA v.2 data.

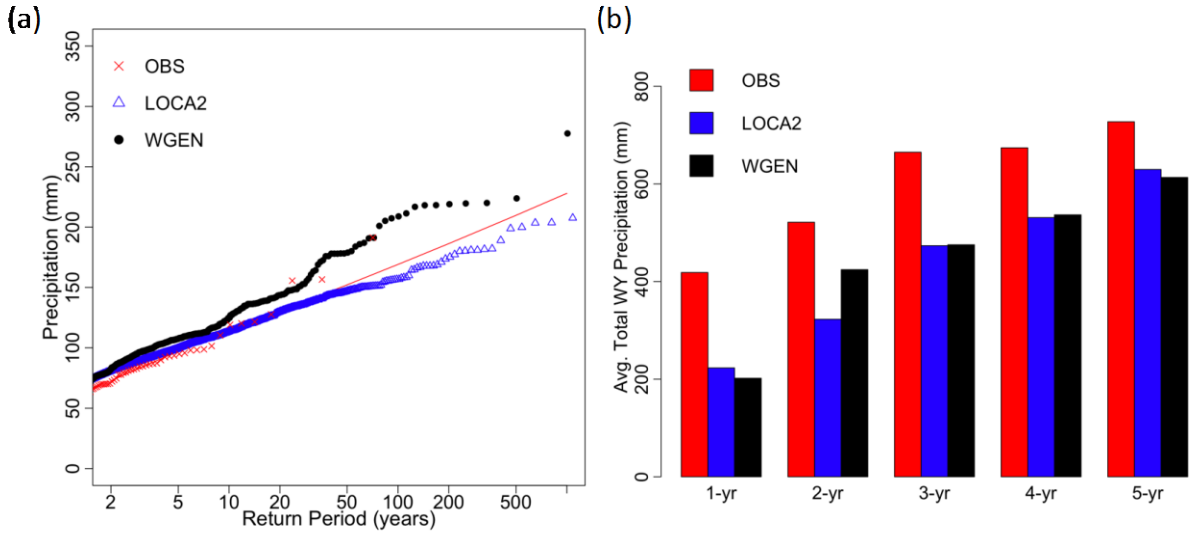

Figure 25.(a) Observed distribution of historical, watershed-scale annual precipitation maxima, along with GEV-based return level estimates fitted to the observed annual maxima (*red*). Also shown are annual maxima from the 1000-year weather generator simulation from scenario #12 (2°C warming, 0% change in mean precipitation, and 7% per °C scaling of extreme precipitation), and annual maxima from the 3750 years of the future LOCA v.2 projections centered around 2°C warming.**(b)** The worst 1, 2, 3, 4, and 5-year drought events in the observed record (*red*), across the ensemble of future LOCA v.2 projections centered around 2°C warming (*blue*), and in the 1000-year weather generator simulation for climate scenario #12 (*black*). All results are shown for the Upper Tuolumne River basin.

Another way to consider the effects of thermodynamic precipitation scaling is to quantify the relative contribution of the upper and lower quantiles of daily precipitation to long-term total precipitation. As shown in Dettinger (2016), historically, the wettest 5% of days have contributed an average of 38% of total precipitation over the Central Valley catchment and represented approximately 85% of the variance in water-year precipitation totals, which yields a precipitation regime that is highly dependent on the few largest storms at annual and interannual timescales. Thermodynamic precipitation scaling acts to amplify this regime, concentrating greater and greater portions of total precipitation into the largest storm events (and less and less into smaller to moderate-sized storms). We explore this property for both the 1000-year weather generator simulation and the LOCA v.2 ensemble following methods applied in Dettinger (2016). First, all non-zero precipitation days (>0.01mm) are extracted, averaged at the watershed basin scale, and split into the wettest 5% of precipitation days (95^{th} percentile and above) and all other precipitation days based on the 50-year historical period 1950-1999 (each LOCA v.2 model-SSP-variant is treated independently). Next, for the weather generator, we break the 1000-year simulations into 20 non-overlapping 50-year blocks and split precipitation days using the historical 95^{th} percentile. For the LOCA v.2 ensemble data, we extract 50-year periods from the projection period where the average temperature change (relative to 1981-2010) has reached incremental warming levels of 1°C, 2°C, 3°C, and 4°C, and split those daily precipitation accordingly. Finally, for each 50-year precipitation block of the weather generator and LOCA v.2 ensemble, the change in precipitation of the wettest 5% of days and all other days is divided by the total historical precipitation, thus yielding the contribution of each quantile to the total precipitation change in a 50-year block.

Figure 26a shows the results of this analysis for weather generator scenario #12 compared to the current condition scenario #1 (i.e., no perturbation). The effect of thermodynamic scaling is visible and mostly equivalent for all 50-year blocks: the contribution of the wettest 5% of days to total precipitation is ≈5% higher (and all other days ≈5% lower) under 2°C warming. For example, in the 50-year block which has ≈13% more total precipitation compared to historical, the wettest 5% of days contribute 9% (14%) without (with) thermodynamic scaling, while all other days contribute 4% (-1%) without (with) thermodynamic scaling. In other words, the scaling of extreme precipitation has made it so that overall wetter periods have become wetter due only to large storm events, while overall drier periods have become drier due mostly to "losing" precipitation from small to moderate sized storms.

Figure 26b shows the results of this analysis for weather generator scenario #12 compared to that of the LOCA v.2 ensemble at the average warming of 2°C. The LOCA v.2 ensemble does not show the same level of extreme precipitation scaling as used in the weather generator, thus changes in contributions from the wettest 5% and all other days in overall wetter and drier 50-year periods are not as severe as those generated in scenario #12. Similarly, Figure 26c compares the weather generator and LOCA v.2 for each incremental warming level 1-4°C, showing that the rate of extreme precipitation scaling is not as high in LOCA v.2 as is assumed in the weather generator scenarios, which reaches approximately twice that of LOCA v.2 at the 4°C warming level.

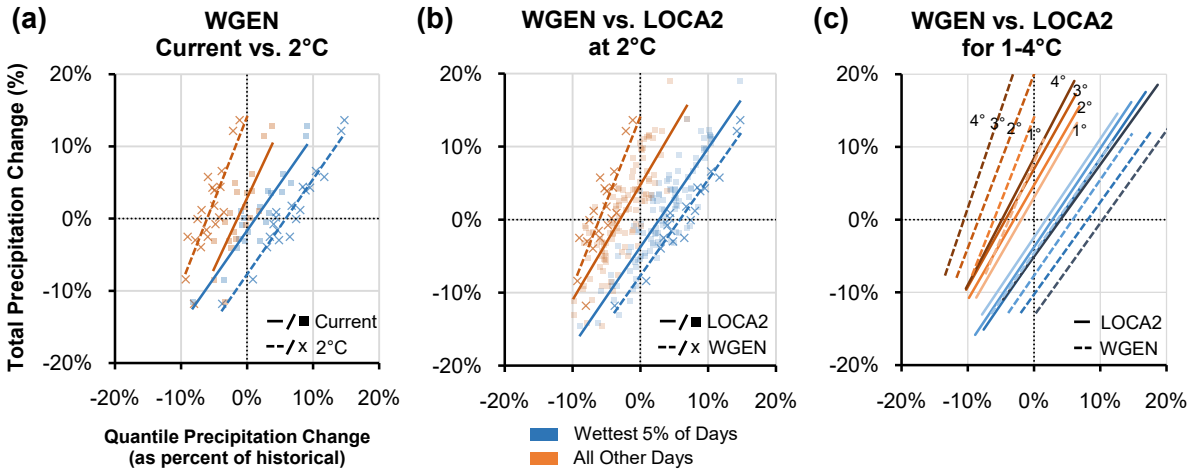

Figure 26. Contribution of daily quantile precipitation changes (normalized by historical precipitation) to total precipitation changes in LOCA v.2 and weather generator, where **(a)** compares the 20, 50-year blocks of the 1000-year weather generator simulation under current climate (scenario #1) to the simulation from scenario #12 (2°C warming, 0% change in mean precipitation, and 7% per °C scaling of extreme precipitation); **(b)** compares the scenario #12 weather generator simulation to the ensemble of LOCA v.2 projections centered around 2°C warming; and **(c)** compares the trends of weather generator simulations (scenario #11, #12, #13, and #14) and the ensemble of LOCA v.2 projections centered around 1°C, 2°C, 3°C, and 4°C warming. All results are shown for the Upper Tuolumne River basin.

These results suggest two important takeaways: 1) thermodynamic scaling as applied in the weather generator produces an effect on overall wetter and drier precipitation regimes where more (less) of total precipitation is derived from large (small to moderate) storms; 2) the rate of thermodynamic scaling (7% per °C) might be considered a conservative assumption relative to what is shown in the LOCA v.2 ensemble. The latter may be attributable to both biases in the LOCA v.2 statistical downscaling product and the presence of reduced rates of extreme precipitation scaling due to underlying and counteracting effects of atmospheric thermodynamics and dynamics (Gu et al., 2023).

5.3. Exploration of Dynamic Climate Changes

The results in Sections 5.1 and 5.2 are based on climate scenarios that were largely developed from a thermodynamic perspective, i.e., changes in temperature and extreme precipitation that are direct consequences of global warming. The changes in mean precipitation could be considered partially thermodynamically driven, although not entirely. In this section, we consider climate changes that are entirely dynamic in nature, in that they are driven only by changes in atmospheric circulation.

As discussed in Section 3.3.1, we select a small set of dynamic climate scenarios based on recent trends in WR frequencies over the 72-year historical record. Figure 27shows these trends for the 10 WRs considered in this work. A red asterisk is used to highlight those WRs with significant trends at the 1% significance level. WRs 3, 8, 9, and 10 all exhibit significant trends at this level. Recall that WRs 1-7 are cold season WRs, while WRs 8-10 occur in the warm season. As the vast majority of precipitation occurs in the cold season, the increased frequency of WR 3 has the largest implications for recent changes in California weather.WR 3 is characterized by a ridge

over western North America and is associated with dry conditions across California. The recent increases in WR 3 are consistent with the 2-decade long drought conditions experienced over California since 2000.

Based on the trends in Figure 27, we consider two dynamic climate change scenarios to explore. In the first (Dynamic Scenario #1), we simulate from the stochastic weather generator after adjusting the long-term mean frequency of WR 3 upwards from its 1948-2019 mean (~7.7% of days) to its end of record average (~10% of days), and adjust the frequency of all other WRs uniformly downward to accommodate this change. This represents an approximate 30% increase in the frequency of days classified as WR 3(see Appendix A.2 for methods on how this change is implemented in the simulation process). Given the association between WR 3 and dry conditions in California, we anticipate this climate change scenario will lead to worsening droughts and possibly a reduction in extreme precipitation events, although the magnitude of the effect is unclear *a priori*. We will juxtaposethe effects of this climate change scenario to one scenario from those listed in Table 3 (scenario #8 with 3°C of warming, extreme precipitation scaling of 7% per °C, a -12.5% change in mean precipitation) as a point of comparison.We will also combine these two scenarios to see their joint impact, i.e., a scenario with 30% more frequent WR 3 days along with 3°C of warming, extreme precipitation scaling of 7% per °C, and a -12.5% change in mean precipitation.

In the second dynamic climate change scenario (Dynamic Scenario #2), we shift the frequency of all WRs (not just WR 3) to their average frequency at the end of the period of record, based on regression estimates for the last year on record from a linear regression against time (i.e., the value estimated by the blue line in Figure 27 for the year 2019). Therefore, recent trends in any WR (significant or not) will be reflected in the weather simulated by the stochastic weather generator.This dynamic climate change scenario implicitly assumes that the atmospheric circulation experienced more recently reflects a “new normal” that will persist over the next several decades. Importantly, we emphasize here that these dynamic climate change scenarios should not be interpreted as confident projections of what will happen to atmospheric circulation under climate change, but rather as “what if” scenarios that allow us to explore the potential impact of these types of change to decision-relevant climate statistics over areas in California. The results of this exploration will help determine how important such dynamic climate changes could be to water systems planning and management throughout the state, which would provide insight into the degree of effort that should be expended in resolving uncertainties around these types of future climate change.

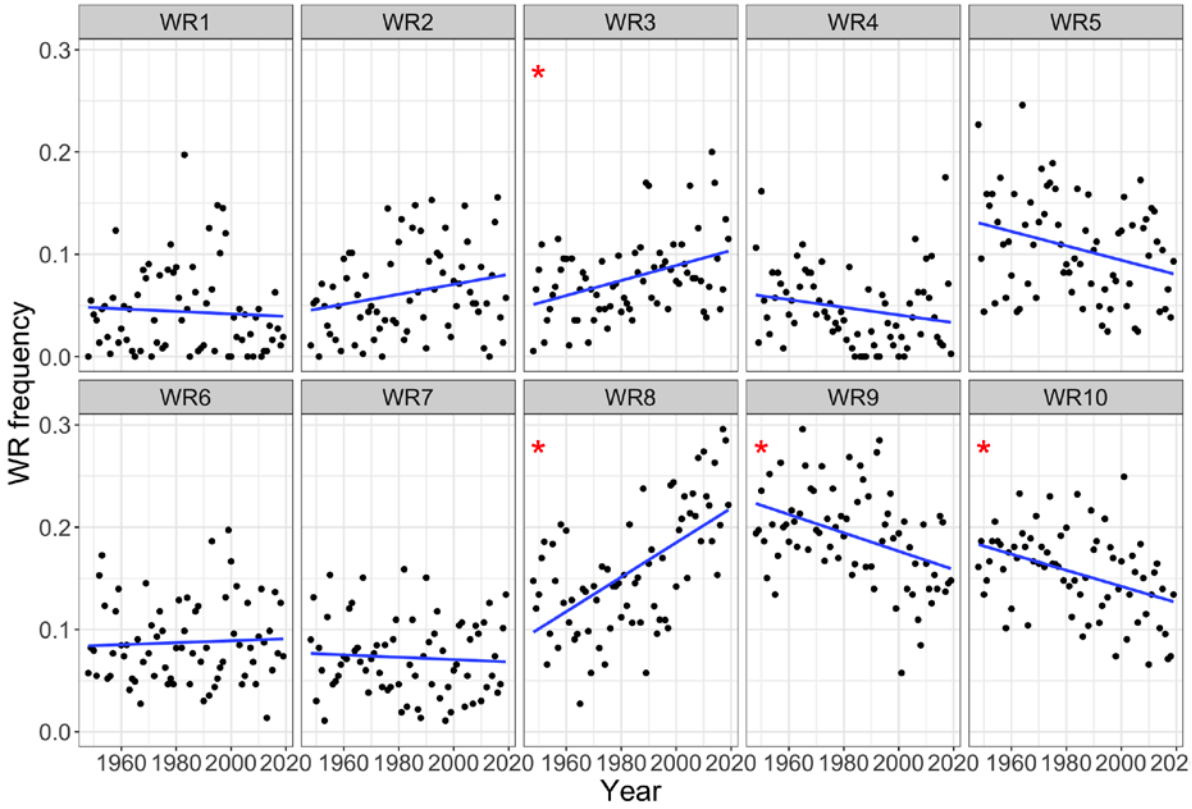

Figure 27. Historical trends in each of the 10 WRs.The blue line indicates a linear regression of WRfrequency against time. Significant trends at the 1% significance level are denoted by a red asterisk.

Figure 28shows the impact of Dynamic Scenario #1 (a 30% increase in the frequency of WR 3) on drought, specifically the magnitude of the worst 5th percentile 2-year, 3-year, and 5-year droughts on record (i.e., the 5th percentile of all 2-year, 3-year, and 5-year water-year precipitation totals across the 1000-year record). Four scenarios are shown, including a baseline (no climate change), Dynamic Scenario #1 (termed 'dynamic change'), the results from Scenario #8 in Table 3 (termed 'thermodynamic change'), and the combined scenario where the changes in Scenario #8 in Table 3 are imposed on the weather generator output forced with a 30% increase in WR 3 frequency (termed 'both'). The results show that for shorter duration droughts (2-year, 3-year), the dynamic climate change scenario has more intense droughts compared to the baseline, but the impact is not as intense as that of the thermodynamic scenario. However, for long-duration droughts (5-year), the impact of the dynamic climate change scenario is very similar compared to that of the thermodynamic scenario. This suggests that for intense long-duration droughts, the importance of a single WR (and the synoptic scale atmospheric flow pattern it represents)becomes increasingly important, likely because the increased frequency of that pattern has more time to manifest in the long-term droughts. When both scenarios are combined, droughts become even more intense; for the 3-year drought, the precipitation total of the drought under the baseline declines by 22% under the combined scenario.

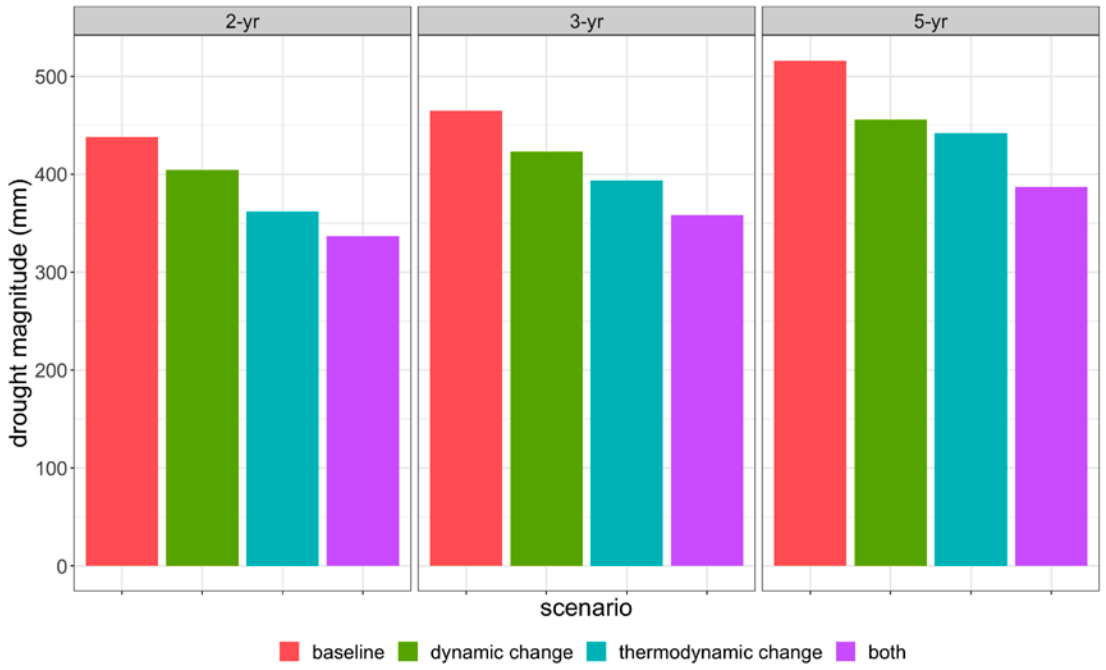

Figure 28. The total precipitation during the worst 5th percentile 2-year, 3-year, and 5-year drought from the baseline scenario (no climate change), Dynamic Scenario #1 ('dynamic change'), Scenario #8 in Table 3 ('thermodynamic change'), and the combined scenario where the changes in Scenario #8 in Table 3 are imposed on Dynamic Scenario #1 ('both'), for the San Joaquin basin (HUC4: 1804).

Figure 29 shows similar results to Figure 28, but for the 20-year and 100-year extreme 1-day precipitation event estimated from a GEV distribution fit to annual maxima. Here, we see that an increase in frequency in WR 3 has a small negative effect on both the 20-year and 100-year storm, as compared to the baseline. However, the magnitude of this effect is small, especially when compared to the increase in these design events under the thermodynamic scenario with extreme precipitation scaling with warming. The combined effect of these two scenarios appears additive, with design events that are slightly smaller than the thermodynamic scenario but still much larger than the baseline.

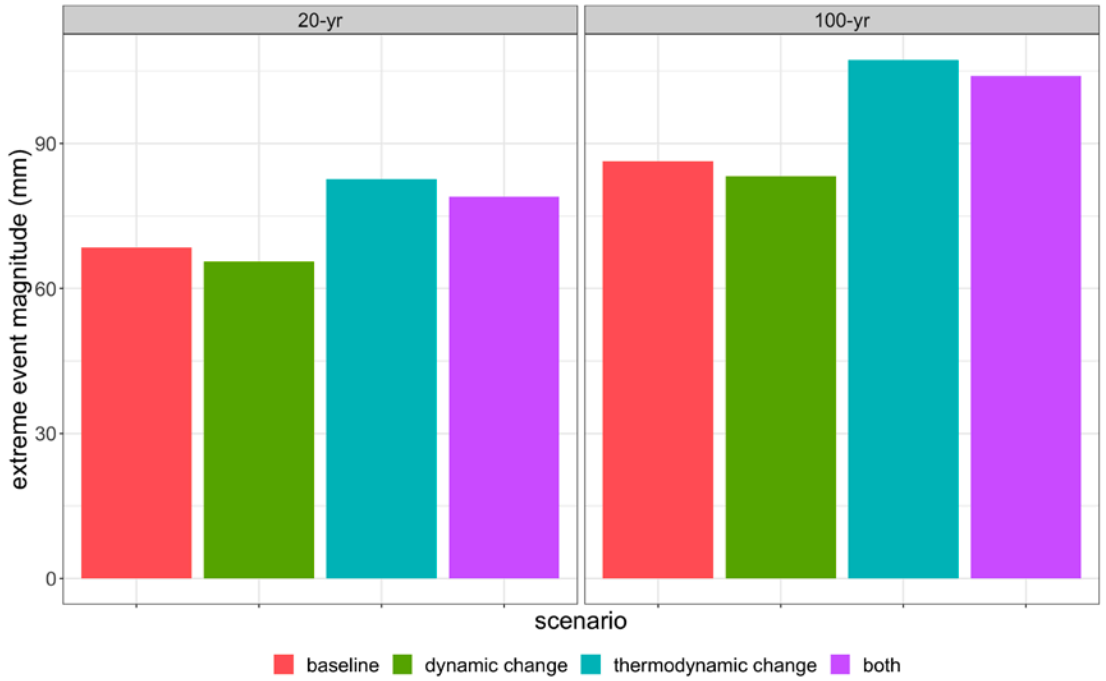

Figure 29. The magnitude of the 20-year and 100-year 1-day extreme precipitation event from the baseline scenario (no climate change), Dynamic Scenario #1 ('dynamic change'), Scenario #8 in Table 3 ('thermodynamic change'), and the combined scenario where the changes in Scenario #8 in Table 3 are imposed on Dynamic Scenario #1 ('both'), for the San Joaquin basin (HUC4: 1804).

Figures30-31show the same results as Figures 28-29, but for Dynamic Scenario #2 (i.e., the most recentfrequencies for all WRs based on regression estimates against time). Similar to Dynamic Scenario #1, the results in Figure 30 show that droughts generally become more intense under Dynamic Scenario #2. However, the effect is not as apparent as compared to that seen in Figure 28, especially for the shorter duration droughts (2-year, 3-year). Conversely, the results in Figure 31 show that the propagation of trends in all WRs has a large negative effect on both the 20-year and 100-year storm, as compared to the baseline. This effect is much stronger than that seen for Dynamic Scenario #1 (Figure 29). Therefore, the results for Dynamic Scenario #2 suggest that by propagating recent trends in all WRs into the weather generator, the resulting simulations are slightly drier than the baseline on average, but reductions in extreme events are significant. The reasons for these differences with Dynamic Scenario #1 are not immediately clear, but they do suggest that shifting frequencies in multiple WRs at once can result in complex dynamics and emergent climate changes in decision-relevant metricsthat are difficult to anticipate *a priori*.

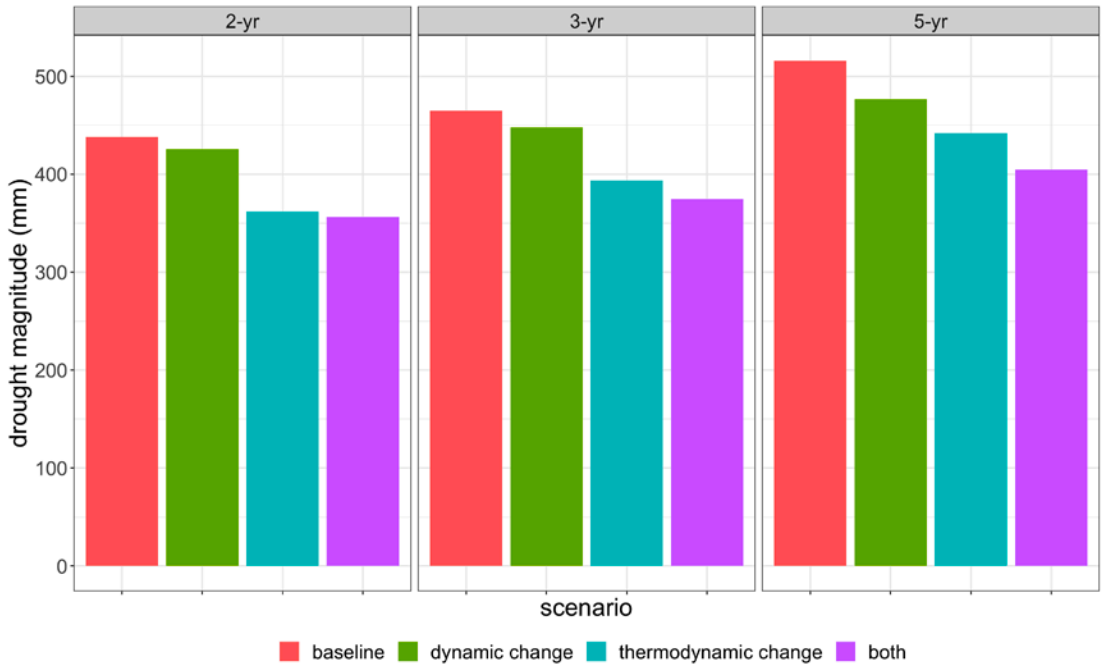

Figure 30. The total precipitation during the worst 5th percentile 2-year, 3-year, and 5-year drought from the baseline scenario (no climate change), Dynamic Scenario #2 ('dynamic change'), Scenario #8 in Table 3 ('thermodynamic change'), and the combined scenario where the changes in Scenario #8 in Table 3 are imposed on Dynamic Scenario #2 ('both'), for the San Joaquin basin (HUC4: 1804).

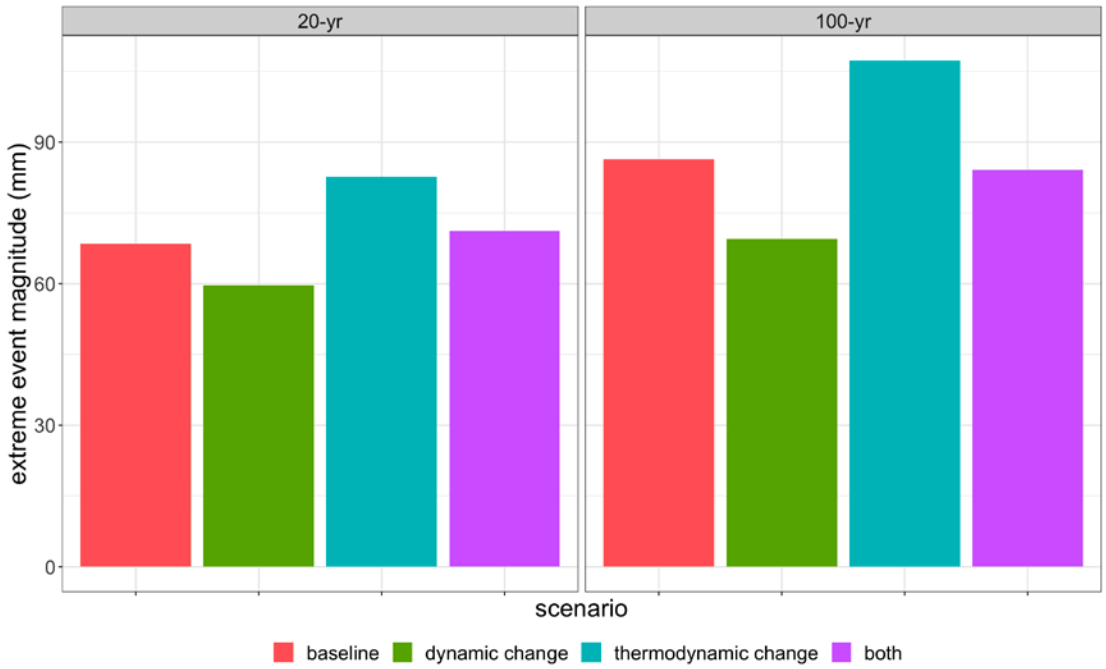

Figure 31. The magnitude of the 20-year and 100-year 1-day extreme precipitation event from the baseline scenario (no climate change), Dynamic Scenario #2 ('dynamic change'), Scenario #8 in Table 3 ('thermodynamic change'), and the combined scenario where the changes in Scenario #8 in Table 3 are imposed on Dynamic Scenario #2 ('both'), for the San Joaquin basin (HUC4: 1804).

6. Discussion and Conclusion

This report presents the development of a stochastic weather generator for the creation of gridded (~6 km) climate change scenarios across the entire state of California. The stochastic weather generator is novel in comparison with other downscaling techniques, because it is designed to distinguish thermodynamic and dynamic mechanisms of climate change, allowing analysts to separately consider those mechanisms they deem most credible for planning purposes. In this work, the weather generator was calibrated and validated across the state of California, and then used to create an ensemble of 1000-year future climate scenarios that primarily reflect thermodynamic climate changes, including temperature increases and the direct scaling of extreme precipitation with temperature, as well as additional changes in mean precipitation. The same scenarios were also applied to the 100-year historical record to produce a complementary dataset. Finally, a limited set of proof-of-concept dynamic climate change scenarios were developed with the stochastic weather generator that represent shifts in the frequency of major atmosphere circulation patterns.

Results in this report showed that the statewide stochastic weather generator is able to simulate long sequences of daily precipitation and minimum and maximum temperature that very accurately mimic the behavior of historical observations at multiple spatial scales (grid cell, basin) and temporal scales (daily, event-based, monthly, annual, inter-annual to decadal). The model reproduces well a large suite of climate statistics at these scales, including moments (averages, variances), spells, both dry and wet extremes, and extreme hot and cold periods. The high level of performance is consistent across the state, allowing the data products to be used by stakeholders throughout California. In addition, simulated weather across the state is correctly correlated across space and between different variables (precipitation, temperature), ensuring that weather generator simulations can be used in hydrologic and water resources analyses that span multiple watersheds across California.

Importantly, the stochastic weather generator can simulate extreme weather conditions, including extreme precipitation across multi-day periods and extreme droughts of varying intensities and durations, that are physically plausible but extend well beyond the range seen in the historical record. These extremes are simulated even in the absence of climate change, highlighting that water resource managers can use the baseline 1000-year weather generator simulation to better understand how natural climate variability could negatively impact their system. When coupled with different scenarios of climate change, weather generator output can then be used to evaluate the combined impacts of natural climate variability and long-term climate change on water system performance.

Despite the very good performance of the model, it is important to highlight the model's limitations and future development needs. It is also important to provide guidance on how to use scenarios from the stochastic weather generator along with other available climate data products to support climate change adaptation efforts in the water sector. Both of these issues are addressed in more detail below.

6.1. Limitations

One benefit of the weather generator is that it can efficiently generate long records of weather data (100's-1000's of years) that are useful for uncovering climate vulnerabilities in water systems. However, it is important to note that the weather generator is still fit to the historical record of observations, and thus uses these observations as a guide when simulating extremes that extend beyond the range of observations. That is, information embedded within the 1000-year simulation from the weather generator should not be considered the same as information within a 1000-year record of observations. Rather, it is a realistic representation of observations that allows for more intense extremes when it generates sequences longer than the observational record.

In addition, there are currently limitations in the way the model can be used to generate scenarios of future climate. For example, when imposing thermodynamic climate changes in which the distribution of daily precipitation is stretched, the weather generator scales up extreme precipitation events and scales down light to moderate events. However, it does not change the frequency of precipitation occurrence (i.e., more zero precipitation days). The increase in zero precipitation days is a prominent thermodynamic signal in climate change projections, and so future adjustments to the model are needed to be able to impose this change directly in scenario development. One possibility for this would be the introduction of an autologistic regression model that can simulate the occurrence of spatially correlated zero and non-zero precipitation. The parameters of this model could then be perturbed directly to create scenarios of more frequent zero precipitation days.

A related limitation in the scenarios generated in this work is the uniform treatment of temperature and precipitation changes across the state. For each scenario generated we do not impose gradients of change across locations, which may exist in a large region like California. We note though that the scenarios created here are such that an analyst could take, for instance, 1 °C warming scenarios for northern locations in the state and 2 °C warming scenarios for southern locations and combine them into a new statewide climate scenario. The timing of storms and resulting weather are identical across these scenarios, since climate changes are imposed as a post-processing step to the statewide weather simulation. Therefore, such a strategy of mixing climate scenarios across locations would preserve the overall space-time distribution of weather over large regions, albeit with an abrupt discontinuity in the long-term climate changes experienced across space.

A broader limitation of the weather generator is that all scenarios generated with the model must be explicitly specified by the analyst. That is, the weather generator will – by design – only create scenarios of future climate that it is directed to create. This is in direct contrast to GCMs, where the governing physics of the earth system and specified boundary conditions (i.e., emission scenarios) lead to plausible scenarios that emerge through internal, interacting processes within the model. This difference is critical in highlighting how the weather generator should (and should not) be used to support climate change adaptation planning, discussed next.

6.2. Guidance for Use

In the context of climate change adaptation planning for water systems, scenarios of future climate created by the weather generator should be viewed as complementary to, rather than in competition with, other downscaled climate projections from GCMs. The weather generator cannot replace GCMs, which are the primary tool for uncovering robust signals of climate change. Rather, the weather generator is designed to provide flexibility in scenario development, so that water resource planners and engineerscan first openly communicatewith the climate science community about which signals within GCM projectionsare most robust or worthy of focused attention. After those signals have been identified, they can be embedded into climate traces from the weathergenerator in a way that is tailored to support planning efforts. For example, water resources planning often requires many models in sequence (e.g., hydrologic, reservoir operations, hydraulic, and groundwater models), many of which can be computationally demanding. The high computational expense often precludes using all data from an entire ensemble of downscaled GCM projections, which even after downscaling and using the entire ensemble can still present challenges in terms of their representation of extremes (see Figure 13). However, with the weather generator an analyst can quicklycreate 1000's of years of weather meant to represent a pre-specified set of climate change signals, and then can select a subset of that trace of an arbitrary length (e.g., 100 years) that 1) is limited in length and so computationally feasible for complex modelling chains; and 2) contains both robust climate change signals and also plausible but challenging extreme events useful for stress testing water system performance. This functionality can support robust climate change adaptation planning in a way that leverages state-of-the-science climate understanding while also accommodating the constraints of typical water resources planning studies.

We envision the process above to be iterative, in which new knowledge derived from climate change projections or emerging observations can quickly be incorporated into new scenarios with the weather generator. For example, in Figure 26, results from the LOCA v.2 ensemble suggest that a smaller scaling rate of extreme precipitation with warming (i.e., less than 7% per °C) may be warranted. This is also supported by a similar result based on a very recent observational study of extremes in California (Najibi and Steinschneider, 2023). The climate scenarios developed in this work (see Table 3) mostly assume a more conservative 7% per °C in extreme precipitation scaling. However, it is straightforward to generate new sets of climate scenarios with a lower scaling rate, based on the knowledge revealed through the analysis of LOCA v.2 and observational records. These new scenarios can be generated quickly (in a matter of hours to days, depending on the spatial scale of the analysis). In this way, the weather generator can be used as a tool to shorten the period between climate knowledge generation and data production to support planning.

For the data developed and presented in this report, stakeholders can utilize both the 100-year historical trace the 1000-year weather generator trace under 30 different climate scenarios (Table 3) to infer the joint impact of both climate change and natural climate variability on their systems.We recommend that stakeholders use these different data products in stages to help developrobust adaptations of water resources infrastructure under climate change. For example, initial adaptation strategies (e.g., new reservoir operational policies; new infrastructure of managed aquifer recharge) could be developed using the baseline 100 historical record to ensure these strategies are able to meet performance requirements under past and recent extreme events. Then, these strategies could be re-evaluated using: 1) the 30 climate change scenarios for the

100-year historical record; 2) the 1000-year baseline weather generator simulation; and 3) the 30 climate change scenarios for the 1000-year weather generator simulation. These evaluations would provide information on how robust a given adaptation strategy is to climate change, natural climate variability, and a combination of the two.
Such information could be used in different ways to alter the initial adaptation strategy. For instance, if a strategy appears vulnerable to the 1000-year baseline weather generator simulation (i.e., natural climate variability) or likely future climate scenarios (e.g., a low degree of future warming projected in most GCMs over the next few decades), this might suggest an immediate need for a more robust strategy. Alternatively, in cases where an adaptation strategy is only vulnerable to the most extreme climate change scenarios or to certain climate change scenarios coupled with the longer (and more extreme) weather generator simulation, the current adaptation strategy may be deemed adequate if coupled with plans for continued climate monitoring and retrofits/adjustments that could be implemented later if needed.

DWR has already begun to use the data products described in this report in two efforts that will be published in late 2023 and mid2024.The 2023 State Water Project (SWP) Delivery Capability Report will deploy the 100-year product to develop response surfaces of important system performance metrics and to develop system risk informed by future climate scenarios for SWP water user planning.The San Joaquin Watershed Studies, which will be finalized by summer 2024, are providing an in-depth, innovative assessment of climate change impacts on the watersector using the 100-year product, includinganassessment of integrated water adaptation solutions to reduce flood risk and replenish depleted aquifers by scaling up Flood-Managed Aquifer Recharge (CA DWR, 2018). Both efforts are designed to account for the uniquewaysthat climate change interacts witheach region's geography, hydrology, socioeconomics, land use patterns, and built infrastructure. The weather generator can supportsuch assessmentsfor any watershed in California by creating climatic perturbations tailored for localized signals of climate variability and change.

Acknowledgements
This work was supported by the California Department of Water Resources, as well as additional support from the Turlock Irrigation District. We would like to acknowledge contributions from Alexander Weyant and Alexander Gershunov at the Scripps Institution of Oceanography, UC San Diego, including their analysis of trivariate precipitation extremes and SPEI presented in this report. We would also like to acknowledge contributions from Wyatt Arnold at the California Department of Water Resources, including his analysis comparing howthe upper and lower quantiles of daily precipitation contribute to long-term total precipitation in the LOCA v.2 and weather generator data.

System Configuration and Data Availability
The weather-regime based stochastic weather generator utilized in this project was initially developed in the R Foundation for Statistical Computing Platform on a Desktop System Type x64-based PC with Intel(R) Core (TM) i7-9700 CPU 3.00 GHz 8 Cores 8 Logical Processors 64.0 GB of RAM (64-bit) R version 4.2.0 (RStudio 2022-04-22 Universal CRT) at Cornell University. The final gridded climate change scenarios for the state were then developed on a high-performance computing system at Cornell (22 nodes with dual 20-Core Intel Xeon Gold 5218R CPUs 2.1 GHz, 192.0 GB of RAM). The gridded climate data across California for each

scenario takes roughly 2-3 hours to generate.The final two datasets developed in this work (the *Historical Climate Change Ensemble* and *Stochastic Climate Change Ensemble*) together fill up ~ 4.4 terabyte (TB) of disk space (0.5 TB, 3.9 TB). The dataset containing two dynamic climate change scenarios occupies an additional ~ 1 TB of data storage. Thesedata have been transferred to the California Department of Water Resources.

APPENDIX A: MATHEMATICAL FORMULATION OF THE STOCHASTIC WEATHER GENERATOR

A.1.Non-homogeneous Hidden Markov Models for Identifying Weather Regimes

We utilize a variant of hidden Markov models(HMMs) to identify weather regimes (WRs) from 500 hPa geopotential height anomalies (GPHAs). An HMMinvolves a finite set of hidden states that transition over time according to the Markovproperty (Markov, 1954). Each hidden state represents a division of the random field's state space. This process is akin to cluster analysis, but with the clusters exhibiting Markovian temporal dynamics. At each time step, the spatial field may correspond to a particular hidden state with a certain probability. By leveraging the transition probabilities, each time period can be assigned to a specific state, optimizing the likelihood of the assignment throughout the data's time span (Rabiner, 1989). Additionally, an external predictor (i.e., a set of exogenous variables, or covariates) can be utilized to influence the transition probabilitiesover time with a specific period. In such cases, the model is referred to as a nonhomogeneous hidden Markov model (NHMM). NHMMs are an extension of HMMs that allow for time-varying transition probabilities, where the transition probabilities between hidden states change over time according to an external predictor or a set of covariates.

Following Najibi et al. (2021), we use an NHMM toinfer WRs from the spatiotemporal evolution of 500 hPa GPHAs, including their persistence, seasonal evolution, and long-term trends. A short description of the mathematical formulation for this approach is provided below.

First, we define the followingnotation:

K: The number of hidden states (i.e., number of WRs).

T: The number of time steps or observations.

$\boldsymbol{\pi}$: The initial state probabilities, aK-dimensional vector where $\pi(i)$ represents the probability of starting in hidden state *i*.

A: The transition probability matrix, a K × K matrix that can vary through time, where A(i,j,t) represents the probability of transitioning from hidden state *i* to hidden state *j* at time *t*.

B: The emission probabilities, a K × T matrix where B(i,t) represents the probability of observing the t-*th* observation given hidden state *i*.

X: The external predictors or covariates, a K × T matrix where each column X(:,t) represents the covariates at time *t*.

The NHMM workflow can be mathematically described inthe following steps:

1. Initialization:
    - Set the initial state probabilities: $\boldsymbol{\pi}(i)$ for i = 1 to *K*.
2. Time-varying transition probabilities:

- Compute the time-varying transition probabilities using the external predictor, or set of covariates: **A**(i,j,t) = Pr(S(t+1) = j | S(t) = i, **X**(:,t)), where S(t) represents the hidden state at time *t*.A multinomial regression frameworkis used here to parameterize the hidden state transitions.
- Two types of exogenous variables (i.e., covariates in **X**) are considered in the NHMM: a) two seasonal harmonics {1+Cos(ψ), 1+Sin(ψ)}, where ψ = $\frac{2\pi t}{365}$represents a periodic signal over the annual cycle; and b) a state-wide wetness index, quantified using the first four principal components from a gridded, water-year standardized precipitation index (SPI) product over California{PC1, PC2, PC3, PC4}.

3. Emission probabilities:
   - Compute the emission probabilities using the observations: **B**(i,t) = P(**O**(t) | S(t) = i), where **O**(t) represents the vector of observations at time *t*.
4. Forward-Backward algorithm:
   - Use the forward-backward algorithm (Baum and Petrie, 1996; Baum et al., 1970) to compute the forward probabilities and backward probabilities (i.e.,model coefficients) for each hidden state *i* and time step *t*.
5. Baum-Welch algorithm:
   - Use the Baum-Welch algorithm (also known as the expectation-maximization (EM) algorithm) (Moon, 1996) to estimate the NHMM parameters **π**, **A**, and **B** based on the forward and backward probabilities.
6. Viterbi algorithm:
   - Use the Viterbi algorithm (Forney, 1973; Rabiner, 1989) to estimate the most probable sequence of hidden states (i.e., historical sequence of WRs).

The Baum-Welch algorithm iteratively updates the NHMM parameters until convergence, maximizing the likelihood of the observed data (i.e., *J* PCs of GPHAs in each season; see Section 3.1 for more details).Note that we only used the first four SPI PCs for identifying the WRs in the cold season, but the harmonics in both warm and cold seasons.

We run the EM algorithm 10 times using different random initializations,and utilize the solutionwith the largest likelihood over all 10 runs to avoid any poor local maxima (Rojo Hernández et al., 2020).We use the R-package 'depmixS4' (Visser and Speekenbrink, 2010) to fit the NHMMs.

A.2. Non-Parametric Simulation of Weather Regimes and Scenarios of Dynamic Climate Change

We developed a novel *non-parametric* approach to WR simulation that addresses the issue of overdispersion in simulated WRs while still allowing for future climate change scenarios with altered WR probabilities. Let $i = 1, \dots, K$ denote the *K* different WRs, which are available as a daily time series over the historical record (1948-2019). Suppose that the historical time series of WRs are clustered into non-overlapping, consecutive segments, where each segment is *D* years long and there are N_D segments in total (in this work, *D=4* and N_D*=18*). In the non-parametric approach for WR simulation, each of the $n=1,\dots,N_D$ segments is given a sampling probability p_n. To simulate a new sequence of daily WRs for an arbitrary number of years, we simply resample

(with replacement) the $n^{th}D$-year segment of daily WRs with probability p_n, until a sufficient number of years has been generated. The final segment can be truncated to ensure a precise number of years of simulated WRs. In this work, WR simulations are set to 1008 years, which corresponds to 14 times the length of the 72-year historical record. By maintaining the simulation length as a whole multiple of the historical record length, we can compare the historical record directly to 14 separate weather generator traces without any differences in sequence length (e.g., as was done in Figure 10).

In the baseline scenario for the weather generator with no dynamic climate change, $p_n = \frac{1}{N_D}$ for $n=1,\ldots,N_D$. That is, each segment is considered equally likely. This is the approach taken for those weather generation scenarios that do not incorporateany dynamic climate changes, i.e., no changes to large-scale circulation patterns.

However, the probabilitiesp_n can be adjusted to alter the frequencies of each of the K WRs in the final simulation, enabling the generation of dynamic climate change scenarios (i.e., scenarios in which the frequencies of different atmospheric flow patterns change compared to their historical frequencies). This is achieved using a linear program. The goal of this model is to identify new sampling probabilities p_n that, when used in the nonparametric simulation approach above, create a sequence of WRs with long-term average frequencies that approach some vector of target frequencies $\theta^* = \{\theta_1^*, \ldots, \theta_K^*\}$ which define a scenario of dynamic climate change.

Let $\theta_{i,n}$ be the average frequency of occurrence (expressed as a probability) of WR i occurring in segment n. This value can be calculated simply by adding up the historical number of daily WR i occurrences in segment n and dividing it by the total number of days in the segment. We develop a series of K equations (for the K WRs) relating the WR frequencies in each segment to the target frequencies θ^* based on the sampling probabilities of each segment, as follows:

$$\begin{aligned} &p_1\theta_{1,1} + \cdots + p_n\theta_{1,n} + \cdots + p_{N_D}\theta_{1,N_D} - \pi_1^+ + \pi_1^- = \theta_1^* \\ &\qquad\qquad\qquad\qquad \vdots \\ &p_1\theta_{K,1} + \cdots + p_n\theta_{K,n} + \cdots + p_{N_D}\theta_{K,N_D} - \pi_K^+ + \pi_K^- = \theta_K^* \end{aligned} \qquad \text{(Eq. A2.1)}$$

Here, $\pi_i^+$and π_i^- are slack variables for the i^{th} WR, which allow the weighted average frequency across the N_D segments to deviate from the target frequency θ_i^*. We also require that the sampling probabilities across segments sum to unity:

$$p_1 + \cdots + p_n + \cdots + p_{N_D} = 1 \qquad \text{(Eq. A2.2)}$$

We seek to minimize the slack variables π_i^+, π_i^- for $i=1,\ldots,K$ in the objective function in order to force the solution to find sampling probabilities $\{p_1, \ldots, p_{N_D}\}$that help achieve the target WR frequencies θ^*:

$$J_1 = \sum_{i=1}^{K} C_\pi(\pi_i^+ + \pi_i^-) \qquad \text{(Eq. A2.3)}$$

Where J_1 is the first component of the objective function to be minimized, and C_π is a cost coefficient. If the model is defined only by the constraints in Eqs. A2.1-A2.2 and the objective

function in Eq. A2.3, there can be a tendency for the solution to force the sampling probabilities for many segments towards zero, and applying non-zero probability to only a handful of segments. This is undesirable, because we do not want to repeat only a few *D*-year segments of daily WRs in our final simulation.Therefore, we further adjust the model to incentivize equal sampling probabilities across segments (i.e., minimize the deviation of p_n from $\frac{1}{N_D}$) as much as is possible while still achieving long-term WR frequencies that approach the target values θ^*. To do this, we introduce a series of additional constraints on the sampling probabilities themselves:

$$p_n - \gamma_n^+ + \gamma_n^- = \frac{1}{N_D} \quad \text{(Eq. A2.4)}$$

Here, $\gamma_i^+$and γ_i^- are slack variables for the n^{th} sampling probability that quantify its deviation from a uniform probability $\frac{1}{N_D}$. Each slack variable is further partitioned into two components that can be minimized in the objective function, allowing for a piecewise linear cost function on these deviations:

$$\begin{aligned} \gamma_n^+ &= \gamma_{n,1}^+ + \gamma_{n,2}^+ \\ \gamma_n^- &= \gamma_{n,1}^- + \gamma_{n,2}^- \end{aligned} \quad \text{(Eq. A2.5)}$$

And:

$$\begin{aligned} &\gamma_{n,1}^+ \le \tau_\gamma \qquad for\ all\ n \in \{1, \dots, N_D\} \\ &\gamma_{n,1}^- \le \tau_\gamma \end{aligned} \quad \text{(Eq. A2.6)}$$

With:

$$J_2 = \sum_{n=1}^{N_D} \left(C_{\gamma_1} \gamma_{n,1}^+ + C_{\gamma_1} \gamma_{n,1}^- + C_{\gamma_2} \gamma_{n,2}^+ + C_{\gamma_2} \gamma_{n,2}^- \right) \quad \text{(Eq. A2.7)}$$

Here, J_2 is the second component of the objective function to be minimized, such that the final objective function is $J = J_1 + J_2$. τ_γ is a user-selected threshold, and C_{γ_1} and $C_{\gamma_2}$are cost coefficients such that $C_{\gamma_2} \gg C_{\gamma_1} \gg C_\pi$. This formulation will seek solutions that drive the sampling probabilities p_n towards a uniform probability $\frac{1}{N_D}$, and will penalize small deviations from $\frac{1}{N_D}$ (i.e., deviations smaller than τ_γ) less than large deviations. Both C_{γ_2} and C_{γ_1} are significantly greater than C_π, which will force the model to prioritize keeping the sampling probabilities close to $\frac{1}{N_D}$ over adjusting the sampling probabilities to achieve the target WR frequencies θ^*. To force the model to achieve WR frequencies that are close to the target, we also constrain the slack variables π_i^+ and π_i^+ with a user-defined threshold (τ_π) to ensure that the weighted average frequency across the N_D segments is within some small distance of the target for all WRs, as below:

$$\begin{matrix} \pi_i^+ \leq \tau_\pi \\ \pi_i^- \leq \tau_\pi \end{matrix} \qquad for\ all\ i \in \{1, \ldots, K\} \tag{Eq. A2.8}$$

The value of τ_π can be set small at first (e.g., 0.0001) and iteratively adjusted upward if the model is initially infeasible. τ_π can be further adjusted upward to strike a balance between achieving the target WR frequencies and retaining sampling probabilities for each segment that are close to a uniform probability. This balance is ultimately subjective and left to the analyst to decide. We recommend that this initial calibration step can be taken with $\tau_\gamma = 0$ (i.e., no piecewise linear cost). Then, the value of τ_γ can be fine-tuned to better balance deviations of sampling probabilities across segments. Our experience suggests that values for τ_γ between 0 and 0.4 provide adequate results.

A.3. Copula-Based Jittering Algorithm

We use a copula-based jittering approach that enables bootstrapped values of daily, heavy precipitation to extend beyond the range of the instrumental record. Let $\tilde{\boldsymbol{p}}_t$ be a vector of simulated precipitation values from the bootstrap at time *t* across all sites. Assume the non-zero, daily precipitation amounts at each site *s* can be modeled by a distribution with cdf $F(p|\boldsymbol{\theta}_s)$. In this study, we assume precipitation follows an extreme value mixture model (Scarrott and MacDonald, 2012), using a gamma distribution for the bulk density (which varies by month) and a Generalized Pareto distribution (GPD) for values in the tail of the distribution. The cdf of this model evaluated for precipitation at site *s* and day *t* is given by:

$$F\left(p_{s,t}\middle|\boldsymbol{\theta}_{s,t}\right) = \begin{cases} \pi_s \dfrac{F_{gamma}\left(p_{s,t}|\alpha_{s,m(t)}, \beta_{s,m(t)}\right)}{F_{gamma}\left(u|\alpha_{s,m(t)}, \beta_{s,m(t)}\right)} & p_{s,t} \leq \zeta_s \\ \pi_s + (1-\pi_s) F_{GPD}\left(p_{s,t}|\sigma_s, \xi_s\right) & p_{s,t} > \zeta_s \end{cases} \tag{Eq. A3.1}$$

Here, ζ_s is a threshold that separates heavy from non-heavy precipitation, F_{gamma} is the cdf of a gamma distribution with parameters $\alpha_{s,m(t)}, \beta_{s,m(t)}$ that vary through time based on calendar month *m(t)*, F_{GPD} is the cdf of a Generalized Pareto distribution with parameters σ_s, ξ_s, and π_s is the probability of precipitation exceeding the threshold ζ_s. The full vector of model parameters is given by $\boldsymbol{\theta}_{s,t}$.

For each simulated day *t* and site *s*, let $\tilde{u}_{s,t} = F\left(\tilde{p}_{s,t}\middle|\boldsymbol{\theta}_{s,t}\right)$ be the non-exceedance probability associated with $\tilde{p}_{s,t}$. We focus specifically on those non-exceedance probabilities associated with heavy precipitation and utilize the conditional non-exceedance probabilities for the GPD given that $\tilde{p}_{s,t} > \zeta_s$:

$$\tilde{u}_{s,t_{GPD}} = \frac{\tilde{u}_{s,t} - \pi_s}{(1-\pi_s)} \tag{Eq. A3.2}$$

The values $\tilde{u}_{s,t_{GPD}}$ are the non-exceedance probabilities for the GPD component of the extreme value mixture model and will range from 0 to 1. At any time *t*, let $\tilde{\boldsymbol{u}}_{t_{GPD}}$ denote the vector of

values $\tilde{u}_{s\in\psi_t,t_{GPD}}$ only for sites $s \in \psi_t$ where $p_{s,t} > \zeta_s$. That is, ψ_t is the subset of sites with heavy precipitation on day *t*, so that $|\psi_t| \leq S$.

We perturb the values in the vector $\tilde{\boldsymbol{u}}_{t_{GPD}}$ to create a new vector of values $\tilde{\boldsymbol{u}}^*_{t_{GPD}}$ that are centered around but are not equal to $\tilde{\boldsymbol{u}}_{t_{GPD}}$. The perturbations are simulated using a Gaussian copula. Let Σ be an *S*×*S* Spearman (rank) correlation matrix for the vector of all daily, observed precipitation across sites, and let $z_{s,t_{GPD}} = \phi^{-1}(\tilde{u}_{s,t_{GPD}})$ be a z-score (i.e., ϕ is the standard normal cdf) for simulated heavy precipitation at time *t* and site *s* in the set ψ_t. Note that for simplicity we drop the notation ψ_t, but emphasize that z-scores at time *t*$(z_{s,t_{GPD}})$ are only calculated for sites *s* with heavy precipitation at time *t*. We create a covariance matrix to simulate new z-scores $\tilde{\boldsymbol{z}}^*_{t_{GPD}}$ from a multivariate normal distribution centered around the original scores $\tilde{\boldsymbol{z}}_{t_{GPD}}$:

$$\tilde{\boldsymbol{z}}^*_{t_{GPD}} \sim MVN(\tilde{\boldsymbol{z}}_{t_{GPD}}, \Omega) \qquad \text{(Eq. A3.3)}$$

With

$$\Lambda = \lambda \mathrm{I}$$

$$\Omega = \Lambda \Sigma \Lambda^T \qquad \text{(Eq. A3.4)}$$

Here, Λ is a diagonal matrix of dimension $|\psi_t| \times |\psi_t|$ with constant diagonal term λ, which is a user-defined parameter. If λ=1, then $\Omega = \Sigma$ and $\tilde{\boldsymbol{z}}^*_{t_{GPD}}$ will deviate substantially from the original values $\tilde{\boldsymbol{z}}_{t_{GPD}}$ (which are based on the bootstrapped precipitation values), but will retain the observed correlation structure across sites. However, as λ is made small, the matrix Ω will have small variances along the diagonal and $\tilde{\boldsymbol{z}}^*_{t_{GPD}}$ will not vary much from $\tilde{\boldsymbol{z}}_{t_{GPD}}$. Then, for each site, the perturbed z-score can be back-transformed to a proposed non-exceedance probability for the GPD, $\tilde{u}^*_{s,t_{GPD}} = \phi(z^*_{s,t_{GPD}})$, and an associated proposed precipitation value $\tilde{p}^*_{s,t_{GPD}} = {F_{GPD}}^{-1}(\tilde{u}^*_{s,t_{GPD}} | \sigma_s, \xi_s)$. The proposed non-exceedance probability will then be selected over the original one based on the following conditional probabilities of observing a different precipitation value given the value that was simulated:

$$\pi = \begin{cases} Pr(P > \tilde{p}^*_{s,t_{GPD}} | P > \tilde{p}_{s,t_{GPD}}) = \frac{Pr(P>\tilde{p}^*_{s,t_{GPD}}, P>\tilde{p}_{s,t_{GPD}})}{Pr(P>\tilde{p}_{s,t_{GPD}})} = \frac{Pr(P>\tilde{p}^*_{s,t_{GPD}})}{P(P>\tilde{p}_{s,t_{GPD}})} = \frac{1-\tilde{u}^*_{s,t_{GPD}}}{1-\tilde{u}_{s,t_{GPD}}} & , \tilde{p}^*_{s,t_{GPD}} > \tilde{p}_{s,t_{GPD}} \\ Pr(P \leq \tilde{p}^*_{s,t_{GPD}} | P \leq \tilde{p}_{s,t_{GPD}}) = \frac{Pr(P\leq\tilde{p}^*_{s,t_{GPD}}, P\leq\tilde{p}_{s,t_{GPD}})}{Pr(P\leq\tilde{p}_{s,t_{GPD}})} = \frac{Pr(P\leq\tilde{p}^*_{s,t_{GPD}})}{Pr(P\leq\tilde{p}_{s,t_{GPD}})} = \frac{\tilde{u}^*_{s,t_{GPD}}}{\tilde{u}_{s,t_{GPD}}} & , \tilde{p}^*_{s,t_{GPD}} \leq \tilde{p}_{s,t_{GPD}} \end{cases}$$

$$\tilde{u}^{final}_{s,t_{GPD}} = \begin{cases} \tilde{u}^*_{s,t_{GPD}} & \pi \leq r_{s,t} \\ \tilde{u}_{s,t_{GPD}} & \pi > r_{s,t} \end{cases} \qquad \text{(Eq. A3.5)}$$

where the random variable *P* is daily precipitation depth and $r_{s,t}$ is a random draw from a uniform distribution between 0 and 1 for site *s* and time *t*. The final heavy precipitation value for each site is then set equal to

$${F_{GPD}}^{-1}(\tilde{u}^{final}_{s,t_{GPD}} | \sigma_s, \xi_s) \qquad \text{(Eq. A3.6)}$$

By virtue of the perturbations embedded in $\tilde{\boldsymbol{z}}^*_{t_{GPD}}$ and thus $\tilde{\boldsymbol{u}}^*_{t_{GPD}}$, the final values of simulated heavy precipitation can extend beyond the range of historical heavy precipitation values, but they preserve the rank correlation structure across sites and the space-time structure captured by the block bootstrap, as long as λ is small. We calibrate λ to balance the reproduction of extreme events beyond the range of the historical record (e.g., estimates of the 500-year and 1000-year storm) while maintaining the general spatial structure of bootstrapped storms. We find that satisfactory results are generally achieved with $\lambda \in (0.1,0.5)$. In this work, we set λ=0.4.

A.4. Thermodynamic Climate Changes to Extreme Precipitation using the GPD-Gamma Extreme Mixture Model

We use quantile mapping to shift and stretch the distribution of daily, non-zero precipitation in a way that replicates the effects of warming temperatures on precipitation through increases in the moisture holding capacity of the atmosphere. In this approach, we must first specify a target value for daily average precipitation and a target scaling for extreme precipitation events under a particular future scenario. For instance, let $\mu^* = (1 + \omega)\mu$ be the desired mean for a future climate scenario such thatμ^* is ω×100% greater than the historic mean μ, and let $(1 + \eta)^{\Delta T}$ be the target scaling rate for extreme precipitation events, where η×100% is the percentage increase in extremes per °C of warming. In this work we assume the same mean change for every calendar month, although a different change for each month is also permissible.

In the weather generator, we resample daily precipitation via a block bootstrap to develop an initial simulation and then adjust heavy precipitation values with a copula-based jittering algorithm (Appendix A.3) and a gamma-GPD extreme mixture model to develop previously unexperienced extreme events. Here, we expand that jittering algorithm and adjust the parameters of the gamma-GPD extreme mixture model to impose the changes to mean and extreme precipitation specified above.

This procedure follows two main steps. First, for all precipitation values that are designed as heavy events (i.e., precipitation $p_{s,t}$ at site *s* and time *t* is greater than the site-specific threshold ζ_s in Eq. A3.1 above), we assume that these values will scale at the rate $(1 + \eta)^{\Delta T}$. Therefore, these heavy precipitation values will first be jittered according to the procedure in Appendix A.3 and then multiplied by $(1 + \eta)^{\Delta T}$.

Second, all other non-heavy precipitation values must be adjusted so that 1) the entire series of both non-heavy and heavy precipitation have a new mean μ^*; and 2) non-heavy precipitation values that approach the threshold ζ_s are also scaled upward so that they approach an asymptotic scaling of $(1 + \eta)^{\Delta T}$. To do this, we infer new parameters for the gamma component of the gamma-GPD extreme mixture model and then impose quantile mapping on resampled, non-heavy precipitation through that gamma component.

Let $\alpha_{s,m}, \beta_{s,m}$ be the parameters of a gamma distribution fit to non-zero, non-heavy precipitation at site *s* in calendar month *m*, let σ_s, ξ_s be the parameters of a GPD distribution fit to heavy precipitation values at site *s* greater than the threshold ζ_s, and let $\pi_{s,m}$ be the probability of precipitation exceeding the threshold ζ_s (which can vary by month). The mean of the GPD

distribution is given by $\mu_{GPD} = \zeta_s + \frac{\sigma_s}{1-\xi_s}$, which becomes $\mu_{GPD^*} = (1+\eta)^{\Delta T}\left(\zeta_s + \frac{\sigma_s}{1-\xi_s}\right)$ after the extreme event scaling is applied in step 1 to all heavy precipitation events. Therefore, if the ultimate goal is to develop daily non-zero precipitation with a new mean of μ^*, then the mean of a newgamma distribution$\mu_{gamma\ *}$needs to take into consideration the new mean of the heavy precipitation under the scaled GPD. Thus, we derive a scaling factor $(1+\delta)$ for the mean of the gamma distribution (μ_{gamma}) fit to the original non-heavy precipitation data as follows:

$$(1+\delta) = \frac{(1+\omega)(1-\pi_{s,m})\mu_{gamma} - [(1+\eta)^{\Delta T} - (1+\omega)]\pi_{s,m}\mu_{GPD\ *}}{(1-\pi_{s,m})\mu_{gamma}} \quad \text{(Eq. A4.1)}$$

$$\mu_{gamma\ *} = (1+\delta)\mu_{gamma} \quad \text{(Eq. A4.2)}$$

Here, $\delta \times 100\%$ represents a required percentage change in the gamma distribution's mean so that, after accounting for the new mean of the heavy precipitation under the scaled GPD (μ_{GPD^*}), the new mean of all non-zero precipitation (μ^*) will be $\omega \times 100\%$ of the original mean μ.

We then identifynew parameters α^*, β^* for the gamma distribution by optimizing a multiplicative factorρas follows:

$$\min_\rho \left(\frac{F^{-1}_{gamma}\ (q|\alpha^*,\beta^*)}{F^{-1}_{gamma}\ (q|\alpha,\beta)} - (1+\eta)^{\Delta T}\right)^2 \quad \text{(Eq. A4.3)}$$

such that

$$\alpha^* = \alpha(1+\delta)\rho$$
$$\beta^* = \beta\rho$$

Here, $\mu_{gamma\ *} = \frac{\alpha^*}{\beta^*} = \frac{\alpha(1+\delta)\rho}{\beta\rho} = (1+\delta)\mu_{gamma}$, guarantying the correct change to the mean of the gamma distribution. The optimization then selects ρ and subsequently α^*, β^* in order to minimize the squared difference between the target change in extreme precipitation ($(1+\eta)^{\Delta T}$) and the ratio of quantile functions of the new and old gamma distributions for some quantile q. We set q equal to some very large quantile (e.g., 0.9999999) so that the optimization tries to find new gamma parameters α^*, β^* that force the tail of the gamma distribution (i.e., the very largest non-heavy precipitation events) to scale at the same rate as the heavy precipitation events.

Once the new gamma parameters are determined for each site and month, daily simulated non-heavy precipitation $\tilde{p}_{s,t}$from the bootstrap is adjusted by first determining the non-exceedance probability $\tilde{u}_{s,t\ gamma} = F_{gamma}\left(\tilde{p}_{s,t}|\alpha,\beta\right)$, and then replacing the resampled precipitation value with a new value derived from the adjustedgamma distribution: $F^{-1}_{gamma}\left(\tilde{u}_{s,t\ gamma}\,\middle|\,\alpha^*,\beta^*\right)$. This procedure is repeated for each nonzero, non-heavy precipitation amount for each site synthesized by the weather generator.

References

Acharya, N., Frei, A., Chen, J., DeCristofaro, L., and Owens, E. M. (2017). Evaluating stochastic precipitation generators for climate change impact studies of New York City's primary water supply. *Journal of Hydrometeorology*, *18*(3), 879-896.

Allan, R.P., Barlow, M., Byrne, M.P., Cherchi, A., Douville, H., Fowler, H.J., Gan, T.Y., Pendergrass, A.G., Rosenfeld, D., Swann, A.L.S., Wilcox, L.J. and Zolina, O. (2020). Advances in understanding large-scale responses of the water cycle to climate change. *Ann. N.Y. Acad. Sci.*, 1472: 49-75. https://doi.org/10.1111/nyas.14337

Arendarczyk, M., Kozubowski, Tomasz. J., Panorska, A. K. (2018). The joint distribution of the sum and maximum of dependent pareto risks. *Journal of Multivariate Analysis*, 167, 136–156. https://doi.org/10.1016/j.jmva.2018.04.002

Baum, L.E., and Petrie, T. (1966), Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics*, 37(6), 1554-1563.

Baum, L.E., and T. Petrie, G. Soules, and N. Weiss, (1970). A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *Ann. Math. Stat.*, 41, 164–171, https://doi.org/10.1214/aoms/1177697196.

Berg, N., and Hall, A. (2017). Anthropogenic warming impacts on California snowpack during drought, *Geophys. Res. Lett.*, 44, 2511– 2518, doi:10.1002/2016GL072104.

Borkotoky, S. S., Williams, A. P., Cook, E. R., and Steinschneider, S. (2021). Reconstructing extreme precipitation in the Sacramento River watershed using tree-ring based proxies of cold- season precipitation. *Water Resources Research*, *57*(4), e2020WR028824.

Brown, C., Ghile, Y., Laverty, M., and Li, K. (2012). Decision scaling: Linking bottom-up vulnerability analysis with climate projections in the water sector, *Water Resour. Res.*, 48, W09537, doi:10.1029/2011WR011212.

CA DWR (2018). FLOOD-MAR: Using Flood Water for Managed Aquifer Recharge to Support Sustainable Water Resource, Technical Report. https://water.ca.gov/-/media/DWR-Website/Web-Pages/Programs/Flood-Management/Flood-MAR/DWR_FloodMAR-White-Paper_a_y20.pdf

CA DWR (2019). Decision Scaling Evaluation of Climate Change Driven Hydrologic Risk to the State Water Project, Technical Report, May 2019. https://water.ca.gov/-/media/DWR-Website/Web-Pages/Programs/All-Programs/Climate-Change-Program/Climate-Action-Plan/Files/CAP-III-Decision-Scaling-Vulnerability-Assessment.pdf

CA-DWR CCTAG (2015). Perspectives and guidance for climate change analysis, Technical Report, August 2015. https://water.ca.gov/-/media/DWR-Website/Web-Pages/Programs/All-Programs/Climate-Change-Program/Climate-Program-Activities/Files/Reports/Perspectives-Guidance-Climate-Change-Analysis.pdf

Dempster, A.P., N. M. Laird, and D. B. Rubin (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. *J. R. Stat. Soc. Ser. B*, 39, 1–22.

Dettinger, M. (2013). Atmospheric Rivers as Drought Busters on the U.S. West Coast. *J. Hydrometeor.*, 14, 1721–1732, https://doi.org/10.1175/JHM-D-13-02.1.

Dettinger, M. (2016). Historical and Future Relations Between Large Storms and Droughts in California. *San Francisco Estuary and Watershed Science* 14 (2).

Elbaum, E., Garfinkel, C. I., Adam, O., Morin, E., Rostkier-Edelstein, D., and Dayan, U. (2022). Uncertainty in projected changes in precipitation minus evaporation: Dominant role of dynamic circulation changes and weak role for thermodynamic changes. *Geophysical Research Letters*, 49, e2022GL097725. https://doi.org/10.1029/2022GL097725

Emori, S., and Brown, S.J. (2005). Dynamic and thermodynamic changes in mean and extreme precipitation under climate change, *Geophysical Research Letters*, 32 (17), L17706, doi:10.1029/2005GL023272.

Espinoza, V., Waliser, D. E., Guan, B., Lavers, D. A., & Ralph, F. M. (2018). Global analysis of climate change projection effects on atmospheric rivers. Geophysical Research Letters, 45, 4299–4308.

Forney, G.D. (1973). The Viterbi algorithm. *Proceedings of the IEEE*, 61(3), 268-278.

Fowler, H.J., Blenkinsop, S., and Tebaldi, C. (2007). Review: Linking climate change modeling to impact studies: recent advances in downscaling techniques for hydrologic modeling, *International Journal of Climatology*, 27, 1547-1578.

Gao, Y., Lu, J. and Leung, L. R. (2016). Uncertainties in projecting future changes in atmospheric rivers and their impacts on heavy precipitation over Europe. *J. Clim.,* 29, 6711–6726.

Gershunov, A., Shulgina, T., Clemesha, R.E.S. et al. (2019). Precipitation regime change in Western North America: The role of Atmospheric Rivers. *Scientific Reports,* 9, 9944. https://doi.org/10.1038/s41598-019-46169-w.

Gershunov, A., Sulghina, T., Ralph, F. M., Lavers, D. A. and Rutz, J. J. (2017). Assessing the climate-scale variability of atmospheric rivers affecting western North America. *Geophysical Research Letters,44*(15), 7900-7908. https://doi.org/10.1002/2017GL074175

Gonzales, K. R., Swain, D. L., Nardi, K. M., Barnes, E. A., and Diffenbaugh, N. S. (2019). Recent warming of landfalling atmospheric rivers along the west coast of the United States. *Journal of Geophysical Research: Atmospheres*, 24, 6810 – 6826.

Gu, L., Yin, J., Gentine, P., Wang, H., Slater, L. J., Sullivan, S. C., Chen, J., Zscheischler, J., and Guo, S. (2023). Large anomalies in future extreme precipitation sensitivity driven by atmospheric dynamics. *Nature Communications*, 14(1), 1-13. https://doi.org/10.1038/s41467-023-39039-7.

Gupta, R., Steinschneider, S., and Reed, P.M. (2022). A Multi-Objective Paleo-Informed Reconstruction of Western U.S. Weather Regimes Over the Past 600 Years, *Climate Dynamics*, 60(1-2), 339-358.

He, M., Anderson, J., Lynn, E., Arnold, W. (2021). Projected Changes in Water Year Types and Hydrological Drought in California's Central Valley in the 21st Century. *Climate*, 9(2):26. https://doi.org/10.3390/cli9020026

Henn, B., Musselman, K. N., Lestak, L., Ralph, F. M., and Molotch, N. P. (2020). Extreme runoff generation from atmospheric river driven snowmelt during the 2017 Oroville Dam spillways incident. *Geophysical Research Letters*, *47*(14), e2020GL088189.

Huang, X., and Swain, D.L. (2022), Climate change is increasing the risk of a California megaflood. Science Advances, 8, eabq0995. doi:10.1126/sciadv.abq0995.

Hughes, J. P., and Guttorp, P. (1994). A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. *Water Resources Research*, 30(5), 1535-1546.

Ishida, K., N. Ohara, Ali Ercan, S. Jang, T. Trinh, M. L. Kavvas, K. Carr, and M. L. Anderson, (2019). Impacts of climate change on snow accumulation and melting processes over mountainous regions in Northern California during the 21st century. *Science of the Total Environment* 685, 104-115.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J. and Zhu, Y., 1996. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77(3), 437-472.Kendon, E.J., N. Ban, N.M. Roberts, H.J. Fowler, M.J. Roberts, S.C. Chan, J.P. Evans, G. Fosser, and J.M. Wilkinson, 2017: Do Convection-Permitting Regional Climate Models Improve Projections of Future Precipitation Change?. *Bull. Amer. Meteor. Soc.*, 98, 79–93, https://doi.org/10.1175/BAMS-D-15-0004.1

Kendon, E.J., N. Ban, N.M. Roberts, H.J. Fowler, M.J. Roberts, S.C. Chan, J.P. Evans, G. Fosser, and J.M. Wilkinson, 2017: Do Convection-Permitting Regional Climate Models Improve Projections of Future Precipitation Change?. *Bull. Amer. Meteor. Soc.*, 98, 79–93, https://doi.org/10.1175/BAMS-D-15-0004.1

Livneh, B., E. A. Rosenberg, C. Lin, B. Nijssen, V. Mishra, K. M. Andreadis, E. P. Maurer, and D. P. Lettenmaier, 2013: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions. *J. Climate*, **26**, 9384–9392, https://doi.org/10.1175/JCLI-D-12-00508.1.

Livneh, B., T.J. Bohn, D.S. Pierce, F. Munoz-Ariola, B. Nijssen, R. Vose, D. Cayan, and L.D. Brekke, 2015: A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and southern Canada 1950-2013, *Nature Scientific Data*, 5:150042, doi:10.1038/sdata.2015.42.

Ma, W., Chen, G., and Guan, B. (2020). Poleward shift of atmospheric rivers in the Southern Hemisphere in recent decades. *Geophysical Research Letters*, 47(21), e2020GL089934.

Maher, P., Vallis, G. K., Sherwood, S. C., Webb, M. J., and Sansom, P.G. (2018). The impact of parameterized convection on climatological precipitation in atmospheric global climate models. *Geophysical Research Letters*, 45, 3728–3736. https://doi.org/10.1002/2017GL076826

Maraun, D., Shepherd, T.G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J.M., Hagemann, S., Richter, I., Soares, P.M., Hall, A. and Mearns, L.O. (2017). Towards process-informed bias correction of climate change simulations. *Nature Climate Change*, 7(11), pp.764-773.

Markov, A. A. (1954). The theory of algorithms. *Trudy Matematicheskogo Instituta Imeni VA Steklova*, *42*, 3-375. (Original title: Teoriya algorifmov, Russian Translation of Works of the Mathematical Institute, Academy of Sciences of the USSR, Available at: http://mi.mathnet.ru/tm1178) .

Massoud, E. C., Espinoza, V., Guan, B., and Waliser, D. E. (2019). Global Climate Model Ensemble Approaches for Future Projections of Atmospheric Rivers. *Earth's Future*, 7(10), 1136-1151. https://doi.org/10.1029/2019EF001249

Michaelis, A.C., Gershunov, A., Weyant, A., Fish, M. A., Shulgina, T., & Ralph, F. M. (2022). Atmospheric river precipitation enhanced by climate change: A case study of the storm that contributed to California's Oroville Dam crisis. *Earth's Future*, 10, e2021EF002537. https://doi.org/10.1029/2021EF002537

Moon, T. K. (1996). The expectation-maximization algorithm. *IEEE Signal Processing Magazine*, *13*(6), 47-60.

Mukundan, R., Acharya, N., Gelda, R.G., Frei, A., and Owens, E.M. (2019). Modeling streamflow sensitivity to climate change in New York City water supply streams using a stochastic weather generator, *Journal of Hydrology: Regional Studies*, 21, 147-158.

Muñoz, Á.G., X. Yang, G.A. Vecchi, A.W. Robertson, and W.F. Cooke (2017). A Weather-Type-Based Cross-Time-Scale Diagnostic Framework for Coupled Circulation Models. *J. Climate*, 30, 8951–8972, https://doi.org/10.1175/JCLI-D-17-0115.1

Najibi, N., and Steinschneider, S. (2023). Extreme precipitation-temperature scaling in California: The role of Atmospheric Rivers, *Geophysical Research Letters*, 50(14), 1–11, e2023GL104606. https://doi.org/10.1029/2023GL104606.

Najibi, N., Mukhopadhyay, S., Steinschneider, S. (2021). Identifying weather regimes for regional-scale stochastic weather generators, *International Journal of Climatology*, 41, 2456–2479. https://doi.org/10.1002/joc.6969.

Overpeck, J.T., and Udall, B. (2020). Climate change and the aridification of North America, *The Proceedings of the National Academy of Sciences*, 117 (22), 11856-11858.

Payne, A.E., Demory, ME., Leung, L.R. et al. (2020). Responses and impacts of atmospheric rivers to climate change. *Nat Rev Earth Environ* 1, 143–157. https://doi.org/10.1038/s43017-020-0030-5

Pendergrass, A.G., and Hartmann, D. L. (2014). The atmospheric energy constraint on global-mean precipitation change. *Journal of Climate*, 27(2), 757-768.

Pfahl S., O'Gorman, P.A., Fischer, E.M. (2017). Understanding the regional pattern of projected future changes in extreme precipitation, *Nature Climate Change*, 7 (6), 423-427.

Pierce, D.W., D. R. Cayan, D. R. Feldman, and M. D. Risser, 2023: Future Increases in North American Extreme Precipitation in CMIP6 Downscaled with LOCA. *J. Hydrometeor.*, 24, 951–975, https://doi.org/10.1175/JHM-D-22-0194.1.

Pierce, D.W., Su, L., Cayan, D. R., Risser, M. D., Livneh, B., & Lettenmaier, D. P. (2021). An extreme-preserving long-term gridded daily precipitation dataset for the conterminous United States. *Journal of Hydrometeorology*, 22(7), 1883-1895.

PRISM Climate Group (2014). Oregon State University, https://prism.oregonstate.edu, data created 4 Feb 2014.

Rabiner, L.R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, 77, 257—286

Rahat, S.H., Steinschneider, S., Kucharski, J., Arnold, W., Olzewski, J., Walker, W., Maendly, R., Wasti, A., and Ray, P. (2022). Characterizing Hydrologic Vulnerability under Non-Stationary Climate and Antecedent Conditions using a Process-Informed Stochastic Weather Generator, *Journal of Water Resources Planning and Management*, 148 (6), https://doi.org/10.1061/(ASCE)WR.1943-5452.0001557.

Ray, P., Wi, S., Schwarz, A., Correa, M., He, M., & Brown, C. (2020). Vulnerability and risk: Climate change and water supply from California's Central Valley water system. *Climatic Change*, 161, 177-199.

Rhoades, A. M., Jones, A. D., Srivastava, A., Huang, H., O'Brien, T. A., Patricola, C. M., et al. (2020). The shifting scales of western U.S. landfalling atmospheric rivers under climate change. *Geophysical Research Letters*, 47, e2020GL089096.

Richardson, C. W. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, *17*(1), 182-190.

Robertson, A.W., and Ghil, M. (1999), Large-scale weather regimes and local climate over the western United States, *Journal of Climate*, 12, 1796-1813.

Robertson, A.W., Y. Kushnir, U. Lall, and J. Nakamura (2015), Weather and climatic drivers of extreme flooding events over the Midwest of the United States. Extreme Events: Observations, Modeling, and Economics, *Geophys. Monogr.*, 214, 113–124.

Rojo Hernández, J.D., Mesa, Ó. J., and Lall, U. (2020). ENSO dynamics, trends, and prediction using machine learning. *Weather and Forecasting*, *35*(5), 2061-2081.

Scarrott, C.J. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT - Statistical Journal,* 10(1), 33-59.

Schwarz, A., Ray, P., Wi, S., Brown, C., He, M., and Correa, M. (2018). Climate change risks faced by the California Central Valley water resource system. *California's fourth climate change assessment. Publication number: CCCA4-EXT-2018-001*. https://www.energy.ca.gov/sites/default/files/2019-12/Water_CCCA4-EXT-2018-001_ada.pdf

Seager, R., Naik, N., and Vecchi, G.A. (2010). Thermodynamic and dynamic mechanisms for large-scale changes in the hydrologic cycle in response to global warming, *Journal of Climate*, 23 (17), 4651-4668.

Seager, R., Neelin, D., Simpson, I., Liu, H., Henderson, N., Shaw, T., Kushnir, Y., and Ting, M. (2014). Dynamical and Thermodynamical Causes of Large-Scale Changes in the Hydrological Cycle over North America in Response to Global Warming, *Journal of Climate*, 27 (20), 7921-7948.

Shepherd, T. (2014). Atmospheric circulation as a source of uncertainty in climate change projections. *Nature Geosci,* 7, 703–708. https://doi.org/10.1038/ngeo2253

Shields, C. A. and Kiehl, J. T. (2016). Atmospheric river landfall-latitude changes in future climate simulations. *Geophysical Research Letters*, 43, 8775-8782. https://doi.org/10.1002/2016GL070470.

Shulgina, T., Gershunov, A., Hatchett, B. J., Guirguis, K., Subramanian, A. C., Margulis, Fang, Y., Cayan, D.R., Pierce, D.W., Dettinger, M., Anderson, M.L., and Ralph, F. M. (2023). Observed and projected changes in snow accumulation and snowline in California's snowy mountains. *Climate Dynamics*, 1-16.

Steinschneider, S., Ray, P., Rahat, S.H., and Kucharski, J. (2019). A weather-regime based stochastic weather generator for climate vulnerability assessments of water systems in the Western United States, *Water Resources Research*, 55. https://doi.org/10.1029/2018WR024446.

Steinschneider, S., McCrary, R., Wi, S., Mulligan, K., Mearns, L., and Brown, C. (2015). Expanded Decision-Scaling Framework to Select Robust Long-Term Water-System Plans under Hydroclimatic Uncertainties. *J. Water Resour. Plann. Manage.*, doi:10.1061/(ASCE)WR.1943-5452.0000536, 04015023.

Stephenson, D. B., Collins, M., Rougier, J. C., and Chandler, R. E. (2012). Statistical problems in the probabilistic prediction of climate change*, Environmetrics*, 23(5), 364–372.

Swain, D.L., B. Langenbrunner, J. D. Neelin, and A. Hall (2018). Increasing precipitation volatility in twenty-first-century California. *Nature Climate Change*, 8 (5), 427–433.

Swain, D.L., M. Tsiang, M. Haugen, D. Singh, A. Charland, B. Rajaratnam, and N. S. Diffenbaugh (2014), The extraordinary California drought of 2013/2014: Character, context, and the role of climate change, *Bull. Am. Meteorol. Soc.*, 95, S3– S7.

Ullrich, P. A., Xu, Z., Rhoades, A. M., Dettinger, M. D., Mount, J. F., Jones, A. D., and Vahmani, P. (2018). California's drought of the future: A midcentury recreation of the exceptional conditions of 2012–2017. *Earth's Future*, 6, 1568–1587.

Vicente-Serrano, S.M., Santiago Beguería, Juan I. López-Moreno, (2010). A Multi-scalar drought index sensitive to global warming: The Standardized Precipitation Evapotranspiration Index - SPEI. *Journal of Climate*, 23: 1696-1718.

Visser, I., and M. Speekenbrink (2010). depmixS4: An R package for hidden markov models. *J. Stat. Softw.*, 36, 1–21, https://doi.org/10.18637/jss.v036.i07.

Wilks, D.S. (2002), Realizations of daily weather in forecast seasonal climate, *J. Hydrometeorol.*, 3, 195–207.

Wilks, D.S. (2010). Use of stochastic weather generators for precipitation downscaling. *Wiley Interdisciplinary Reviews: Climate Change*, *1*(6), 898-907.

Wilks, D.S. (2012). Stochastic weather generators for climate-change downscaling, part II: multivariable and spatially coherent multisite downscaling. *Wiley Interdisciplinary Reviews: Climate Change*, *3*(3), 267-278.

Wilks, D.S. and R.L. Wilby (1999). The weather generation game: A review of stochastic weather models, *Prog. Phys. Geogr.*, 23, 329–357.

Williams, A.P., Cook, B.I. and Smerdon, J.E. (2022). Rapid intensification of the emerging southwestern North American megadrought in 2020–2021. *Nat. Clim. Chang.,* 12, 232–234.

Williams, A.P., Cook, E.R., Smerdon, J.E., Cook, B.I., Abatzoglou, J.T., Bolles, K., Baek, S.H., Badger, A.M. and Livneh, B. (2020). Large contribution from anthropogenic warming to an emerging North American megadrought. *Science*, 368(6488), 314-318.

Zechiel, P.R., and Chiao, S. (2021). Climate Variability of Atmospheric Rivers and Droughts over the West Coast of the United States from 2006 to 2019. *Atmosphere* 12, no. 2: 201. https://doi.org/10.3390/atmos12020201

Zhang, W., Hari, V., S-Y Wang, S., LaPlante, M. D., Garfin, G., Affram, G., and Kumar, R. (2022). Fewer troughs, not more ridges, have led to a drying trend in the western United States. *Geophysical Research Letters*, 49, e2021GL0.