Understanding Contributions of Paleo-Informed Natural Variability and Climate Changes on Hydroclimate Extremes in the Central Valley Region of California

Rohini S Gupta[1], Scott Steinschneider[1], and Patrick M. Reed[1]

[1]Cornell University

March 13, 2023

Abstract

To aid California's water sector to better manage future climate extremes, we present a method for creating a regional ensemble of plausible daily future climate and streamflow scenarios that represent natural climate variability captured in a network of tree-ring chronologies, and then embed anthropogenic climate change trends within those scenarios. We use 600 years of paleo-reconstructed weather regimes to force a stochastic weather generator, which we develop for five subbasins in the San Joaquin River in the Central Valley region of California. To assess the compound effects of climate change, we create temperature series that reflect scenarios of warming and precipitation series that are scaled to reflect thermodynamically driven shifts in the daily precipitation distribution. We then use these weather scenarios to force hydrologic models for each of the San Joaquin subbasins. The paleo-forced streamflow scenarios highlight periods in the region's past that produce flood and drought extremes that surpass those in the modern record and exhibit large non-stationarity through the reconstruction. Variance decomposition is employed to characterize the contribution of natural variability and climate change to variability in decision-relevant metrics related to floods and drought. Our results show that a large portion of variability in individual subbasin and spatially compounding extreme events can be attributed to natural variability, but that anthropogenic climate changes become more influential at longer planning horizons. The joint importance of climate change and natural variability in shaping extreme floods and droughts is critical to resilient water systems planning and management in the Central Valley region.

Understanding Contributions of Paleo-Informed Natural Variability and Climate Changes on Hydroclimate Extremes in the Central Valley Region of California

Rohini S. Gupta[1], Scott Steinschneider[2], Patrick M. Reed[1]

[1]School of Civil and Environmental Engineering, Cornell University, Ithaca, NY, USA
[2]Department of Biological and Environmental Engineering, Cornell University, Ithaca, NY, USA

Key Points:

- We introduce a framework to create 600-year ensembles of future weather and streamflow for basins in the San Joaquin Valley.
- We discover vast variability and non-stationarity in flood and drought extremes in the region over the past 600 years.
- Variability in extremes is primarily attributed to natural variability, but climate changes are influential under longer planning horizons.

Corresponding author: R.S. Gupta:, `rg727@cornell.edu`

Abstract

To aid California's water sector to better manage future climate extremes, we present a method for creating a regional ensemble of plausible daily future climate and streamflow scenarios that represent natural climate variability captured in a network of tree-ring chronologies, and then embed anthropogenic climate change trends within those scenarios. We use 600 years of paleo-reconstructed weather regimes to force a stochastic weather generator, which we develop for five subbasins in the San Joaquin River in the Central Valley region of California. To assess the compound effects of climate change, we create temperature series that reflect scenarios of warming and precipitation series that are scaled to reflect thermodynamically driven shifts in the daily precipitation distribution. We then use these weather scenarios to force hydrologic models for each of the San Joaquin subbasins. The paleo-forced streamflow scenarios highlight periods in the region's past that produce flood and drought extremes that surpass those in the modern record and exhibit large non-stationarity through the reconstruction. Variance decomposition is employed to characterize the contribution of natural variability and climate change to variability in decision-relevant metrics related to floods and drought. Our results show that a large portion of variability in individual subbasin and spatially compounding extreme events can be attributed to natural variability, but that anthropogenic climate changes become more influential at longer planning horizons. The joint importance of climate change and natural variability in shaping extreme floods and droughts is critical to resilient water systems planning and management in the Central Valley region.

Plain Language Summary

California experiences cycles of floods and droughts that can be driven by both natural variability and climate change. The specific role of these drivers play in influencing extremes is uncertain, but can strongly dictate how to best plan and manage regional water systems for future extremes. To better quantify the role of these drivers, we introduce a framework that utilizes a 600-year tree-ring reconstruction to create long sequences of plausible ensembles of future weather and streamflow for key basins in the San Joaquin Valley. We find that a large portion of variability in extremes can be attributed to natural variability, but that anthropogenic climate changes become more influential at longer planning horizons. Furthermore, our perception of important drivers can be skewed depending on the specific definitions used to analyze floods and droughts,

which can present significant challenges for adaptation planning and infrastructure development tied to hydroclimate indicators. This study also illustrates the vast variability in extremes that the region has experienced over the past 600 years and highlights the pitfalls of using stationary risk measures.

1 Introduction

The recent drought conditions impacting California are occurring within the broader context of two decades of extreme climate variability. Since 2000, California has experienced four periods of drought: (2000-2003, 2007-2009, 2012-2016, and the ongoing drought beginning in the 2020). The former three complete drought periods were all ended by extreme atmospheric river (AR)-driven events. While offering much needed precipitation, these storms often cause widespread flooding and landslides. In 2017, extreme precipitation associated with ARs generated California's wettest winter in a century and caused catastrophic damage to Oroville Dam, which prompted the evacuation of 188,000 people and required nearly $1 billion in repairs (Henn et al., 2020). Since this event, California has ebbed and flowed through wet and dry periods, including experiencing the driest 22-year period in at least 1,200 years (A. P. Williams et al., 2022).

The recent two decades of California climate extremes are in part a manifestation of the extreme natural variability that characterizes the regional climate. Tree ring reconstructions have shown that the region experienced multiple persistent megadroughts over the past two millennia (late 800s, mid-1100s, late 1200s, mid-1400s, and late 1500s), long before anthropogenic influence (Stahle et al., 2000, 2007; A. Williams et al., 2021). Multi-millennial control runs of coupled global climate models (GCMs) have also reproduced megadroughts in the Southwestern U.S. even without any external sea surface temperature (SST) forcing, suggesting that these droughts can develop due to internal climate variability alone (Hunt, 2011). Some (but not all) of this natural drought variability is linked to major modes of atmospheric and oceanic variability, such as the El Niño Southern Oscillation (ENSO) and the Pacific Decadal Oscillation (PDO) (McCabe et al., 2004; Hoerling et al., 2009; Seager et al., 2015; Cook et al., 2016). Interspersed across the past two centuries, California has also experienced several extreme precipitation events (e.g., 1908-1909, 1913-1914, 1940-1941, 1955-1956, 1969, 1986, and 1997); most prominently the Great Flood of 1861-62 that turned the San Joaquin and Sacramento Valleys into an inland sea (M. D. Dettinger & Ingram, 2013). This event notably occurred after a 20-year drought (Null & Hulbert, 2007). Sediment reconstructions in the Klamath Basin suggest that the 1861-1862 megaflood was not an extreme outlier, but rather a 100-200-year event that has been matched in magnitude several times over the last two millennia (e.g., 212, 440, 603, 1029, 1300, 1418, 1605, 1750, and 1810 CE; M. D. Dettinger and Ingram (2013)).

The historic droughts and floods above, independent of anthropogenic-related warming, confirm the strong influence of natural climate variability in California and more broadly across the Western U.S. However, recent studies show that climate change is amplifying the severity of these extremes. Warming due to anthropogenic radiative forcing has intensified recent droughts in the region, primarily through enhanced atmospheric moisture demand and soil moisture depletion (A. P. Williams et al., 2020). As noted above, the recent cumulative drought conditions in California and the rest of the Western U.S. over the past two decades now ranks as the driest 22-year period in at least 1,200 years (A. P. Williams et al., 2022). Similarly, climate change is increasing the risk of extreme precipitation events via an increase in the strength of cool-season AR events associated with a rise in atmospheric moisture content (Kunkel, 2003; Kirchmeier-Young & Zhang, 2020). A recent study by X. Huang and Swain (2022) found that climate change has already doubled the likelihood of AR-driven megastorms similar to that which caused the Great Flood of 1861-62, and that megastorm sequences of increased frequency and larger magnitude are likely with continued warming.

Thus, the present and evolving risks posed by hydrologic extremes in California is defined by the combined influence of natural climate variability and anthropogenic climate change. Yet the degree to which these two factors will control the future frequency and magnitude of extremes remains uncertain (Hamlet & Lettenmaier, 2007; Siler et al., 2019; Bass et al., 2022). From the perspective of water resource decision-makers who are charged with planning and managing large-scale infrastructure to mitigate the impacts of extreme events, this ambiguity presents a significant challenge. If climate change is the dominant factor that will determine the future magnitude, frequency, and duration of extreme events, then resources and attention should be concentrated on identifying and narrowing the uncertainty of the most prominent climate change signals and propagating them into updated design event estimates used for planning. However, if natural variability plays an equal or larger role in determining the properties of hydrologic extremes relevant to engineering design, then research into the plausible range of extremes due to natural variability should be equally prioritized (e.g., see Koutsoyiannis (2021)). A greater role of natural variability also suggests that dynamic and reversible adaptations may be favorable over irreversible investments. It is thus critically important to quantify the relative and joint roles of climate change versus natural variability in shaping the characteristics of hydrologic extremes, to help balance the allocation of attention and resources in a way that best serves the water sector to prepare for future extreme events.

A growing body of work has sought to partition the relative effects of climate change and natural variability, with a focus on climate variables and in the context of multi-model ensemble simulations (Hawkins & Sutton, 2009; Yip et al., 2011; Knutti et al., 2017; Rowell, 2012; Lehner et al., 2020). These studies primarily attribute variability in projected global and regional temperature and precipitation to climate change scenario uncertainty, global climate change model (GCM) uncertainty, and natural variability. Lehner et al. (2020) shows that scenario and model uncertainty are the dominant drivers of global decadal mean annual temperature and precipitation, but that natural variability dominates projections of regional temperatures (in Southern Europe) and precipitation (in the U.S. Pacific Northwest and Sahel region), particularly at shorter (and more decision-relevant) time scales. Fewer studies have explicitly considered the role of natural climate variability when partitioning variance in projections of hydrologic and water systems variables (Kay et al., 2009; Jung et al., 2011; Vidal et al., 2015; Whateley & Brown, 2016; Schlef et al., 2018; Cai et al., 2021). Kay et al. (2009) found that flood frequency and wintertime runoff in Europe are mostly influenced by choice of GCM, although they quantified natural climate variability using a limited number of GCM integrations with different initial conditions. Vidal et al. (2015) found that natural variability highly influences low flows in snow-dominated catchments in the French Alps, and Cai et al. (2021) found that natural variability is a dominant driver of rainy season runoff in Northeastern China. Jung et al. (2011) quantified natural variability using a block bootstrap on the historical record and found it to have the largest impact on the variance of large floods, as compared to GCM structure, emission scenario, land use change scenario, and hydrologic model parameter uncertainty. Similarly, Whateley and Brown (2016) and Schlef et al. (2018) created ensembles of future streamflow projections with a stochastic weather generator and rainfall-runoff model and found that the variance of reservoir storage as well as water system performance measures is mostly driven by natural climate variability, particularly in the first few decades of the projections.

The relative roles of natural variability and climate change on the variance of hydrologic variables of interest often depends on how natural variability is quantified and propagated into an ensemble of projections. In a majority of the climate studies (Hawkins & Sutton, 2009; Yip et al., 2011; Knutti et al., 2017; Rowell, 2012; Lehner et al., 2020) and three hydrologic studies (Kay et al., 2009; Jung et al., 2011; Vidal et al., 2015) referenced above, natural variability was determined using multi-member ensembles of GCMs run with different initial conditions. However, the degree to which initial-condition ensembles can represent true natural climate variability is unclear (Deser et al., 2020). For instance, these models poorly represent regional precipitation and drought persistence (Rocheta et al., 2014; Moon et al., 2018) and underestimate AR moisture flux and frequency (Zhou & Kim, 2018) all of which are important to regional planning and management of water systems. While the recent generation of models in CMIP6 better represents key features of natural climate variability (e.g., blocking; major climate modes) compared to older generations, significant biases remain (Tatebe et al., 2019; Schiemann et al., 2020)

An alternative way to estimate the relative influence of natural variability and climate change on regional hydrologic response is through bottom-up approaches that employ stochastically generated ensembles (Dessai & Hulme, 2004; Wilby & Dessai, 2010; Nazemi & Wheater, 2014). These methods require synthetic generators trained on observed weather or hydrologic records, which can generate large ensembles of scenarios that extrapolate beyond the observation limits of the historical record, maintain physical plausibility, and embed climate changes into the ensemble. The generation and partitioning of variability in the resulting hydroclimate metrics can provide a more robust way to quantify how much variance in regional hydrologic extremes is driven by climate changes versus natural variability. However, the availability of stochastic models to support these analyses is limited, particularly when interested in the variance decomposition of multiple properties of different hydrologic extremes (i.e., magnitude, duration, frequency, and spatial coherence of floods and droughts). Furthermore, the ways in which flood and drought events are defined, and particularly the time horizon (moving window) over which they are defined, can influence how the relative influences of climate variability and change are perceived. As time horizon shortens, it becomes increasingly difficult to identify clear climate change signals amongst the noise of natural climate variability. For example, Lehner et al. (2020) attributed the vast majority of variance in winter precipitation projections over the US Pacific Northwest to natural climate variability, but this was based on a 10-year moving window (i.e., decadal average). It is possible that any climate change impacts on mean winter precipitation, even if present, are not discernable from the noise within such a short moving window. This issue is especially true for the properties of extreme events, because there are so few samples available from which to estimate signal from noise (even with many ensemble members). To date, it remains unclear how the choice of time horizon influences our understanding of the relative roles of natural climate variability and climate change on the uncertainty in hydrologic extremes.

Based on the above knowledge gaps, this study addresses the following questions:

1. What is the relative importance of natural variability and climate change on variability in decision-relevant drought and flood metrics for the Central Valley of California?

2. How does the selected scale of the time horizon used for analyses influence the perceived importance of these drivers?

To answer these questions, we contribute a framework for creating a regionally consistent ensemble of plausible daily future climate and streamflow scenarios that represent natural climate variability captured in a network of tree-ring chronologies, and then embed anthropogenic climate change trends within those scenarios. A key contribution of this study is the use of 600 years of paleo-informed weather regimes (WRs; Gupta et al. (2022)) to force a weather-regime based stochastic generator (Steinschneider et al., 2019; Najibi et al., 2021), which we develop for five watersheds in the San Joaquin River basin. To assess the compound effect of climate change, we create temperature series that reflect projected scenarios of warming and precipitation series that have been scaled to reflect thermodynamically driven shifts in the distribution of daily precipitation. We then use these weather scenarios to force hydrologic models for each basin, generating ensembles of streamflow across the region. Decision relevant hydrologic metrics for characterizing flood and drought conditions are defined and calculated across San Joaquin subbasins and across the paleo-period using time horizons of varying scale (see Appendix B). Variance decomposition is then employed to characterize the relative contributions of natural variability and climate changes as drivers of flood and drought hazards in individual subbasins and for spatially compounding extremes that emerge across groups of subbasins.

2 Data and Methods

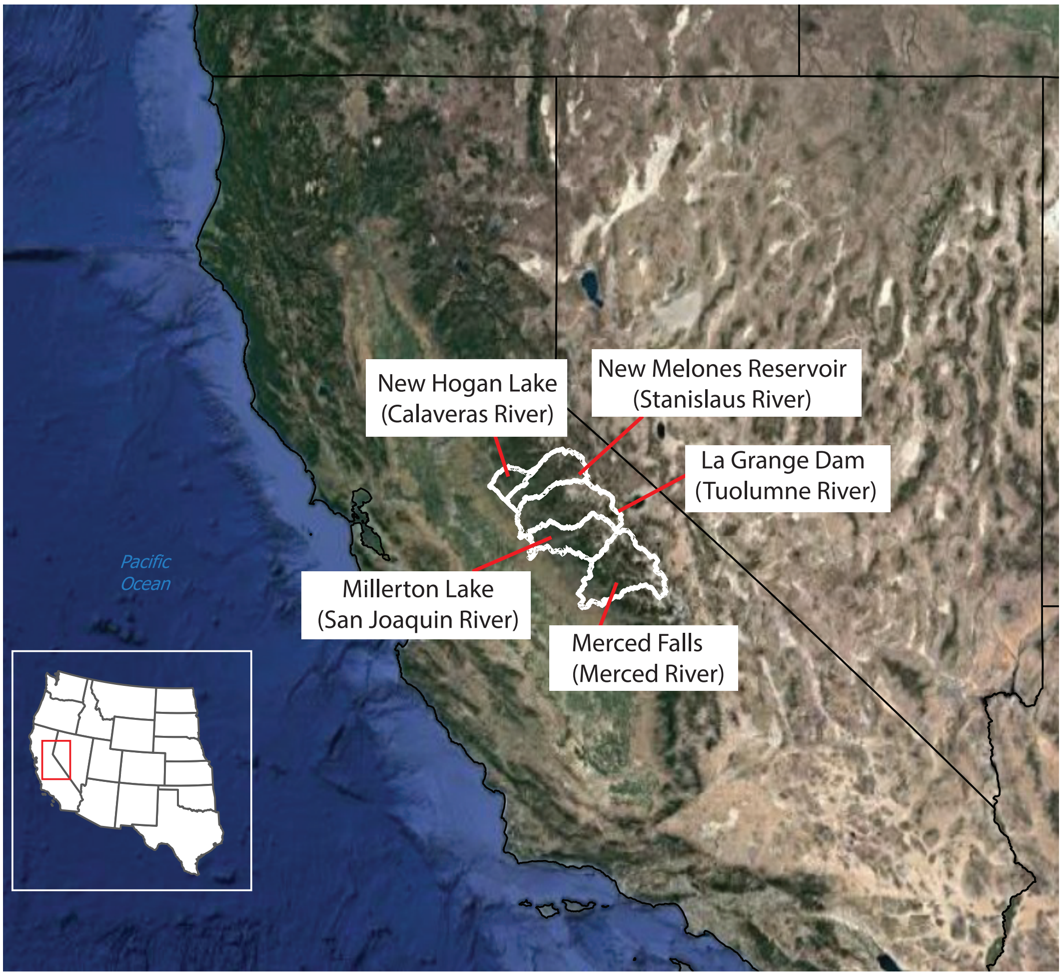

Figure 1. The study area is comprised of five subbasins within the greater San Joaquin River basin.

This study focuses on five subbasins within the San Joaquin River basin (Figure 1): the Tuolumne River, the Merced River, the San Joaquin River, the Stanislaus River, and the Calaveras River. The ultimate goal of this study is to partition the effects of natural climate variability and climate change on different properties of floods and droughts across these watersheds. We contribute a five-step methodology in order to achieve this goal (Figure 2). We first create a novel method to incorporate reconstructed weather regime dynamics (Gupta et al. 2022) into the generation of daily weather through the paleo-period (Section 2.1). Then, we create 600 years of surface weather ensembles across the five subbasins of the San Joaquin conditioned upon these reconstructed dynamics. We also create additional ensembles of surface weather layered with thermodynamic climate changes, such as temperature trends and precipitation scaling (Section 2.2).

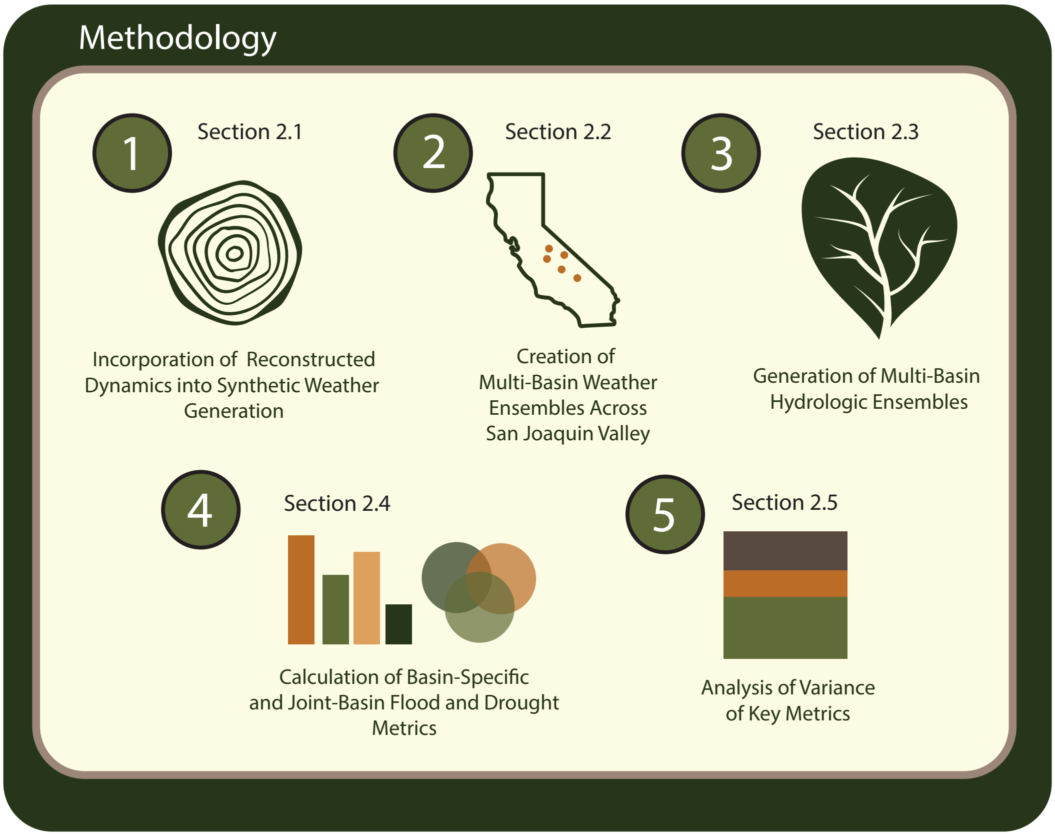

Figure 2. The main methodological components of this study. Reconstructed weather regime dynamics are used to drive weather simulations across the San Joaquin valley that also embed signals of thermodynamic climate change. Ensembles of generated surface weather are used to force hydrologic models to develop streamflow ensembles. Metrics pertaining to flood and drought extremes are calculated throughout the simulation period across basins and ensemble members, and an ANOVA decomposition is used to partition the variance of each metric across the ensemble into variance contributions from thermodynamic climate changes and natural variability.

These ensembles are forced through hydrologic models (SAC-SMA and HYMOD) calibrated for each subbasin to generate ensembles of daily streamflow (Section 2.3). From these streamflow ensembles, we calculate flood and drought metrics, including copula-based metrics to quantify joint flood hazard across basins (Section 2.4). Finally, analysis of variance (ANOVA) is used to partition the contribution of natural variability and the imposed climate changes to variability in the different flood and drought metrics considered (Section 2.5).

2.1 Reconstruction of WR Dynamics

Ensembles of plausible future climate are generated using our extensions of the WR-based stochastic weather generator presented in Steinschneider et al. (2019) and Najibi et al. (2021) to incorporate paleo-reconstructions of WRs. The generator is comprised of a three-step hierarchical structure (Figure 3): (1) identification and simulation of WRs that define large-scale patterns of atmospheric flow across the entire Western U.S., (2) simulation of local weather conditioned on the WRs, and (3) perturbations to the simulated weather reflective of thermodynamic climate change. This study extends step (1) to utilize reconstructed WRs created in Gupta et al. (2022). In that study, a multi-objective optimization and regression-based framework was used to reconstruct the annual frequency of five dominant Western U.S. weather regimes back to 1400 CE based on a gridded, tree-ring based reconstruction of cold season precipitation developed by A. P. Williams et al. (2020) and extended in Borkotoky et al. (2021). Specifically, the first four principal components of annual weather regime occurrence were reconstructed (termed PC_{WR} in Gupta et al. (2022)), which effectively contained all of the information on the annual frequencies of the five WRs. In this study, these principal components are used to force a non-homogeneous hidden Markov model (NHMM), whereby WR states are modeled as a first-order Markov chain with a non-stationary transition probability matrix conditioned on the reconstructed PC_{WR} from Gupta et al. (2022). The NHMM is fit to the first nine principal components of daily 500 hPa geopotential height from NOAA-CIRES-DOE Twentieth Century Reanalysis (V3) dataset (Slivinski et al., 2019) between 180-100°W and 30-60°N (i.e., the Pacific/North American sector) from 1950-2017. The NHMM is forced with the four reconstructed principal components (PC_{WR}) that overlap the same time period, defining a time-varying transition probability matrix shown in Equation 1:

$$P(WR_t = i | WR_t = j, \boldsymbol{X_t} = \boldsymbol{x}) = \frac{\exp(\beta_{0j,i} + \boldsymbol{\beta'_{j,i}}\boldsymbol{x})}{\sum_{k=1}^{K} \exp(\beta_{0j,i} + \boldsymbol{\beta'_{j,i}}\boldsymbol{x})} \quad (1)$$

Here, the transition probability from WR i to WR j at time t is conditioned on $\boldsymbol{X_t'} = [PC_{WR_{1,t}}, PC_{WR_{2,t}}, PC_{WR_{3,t}}, PC_{WR_{4,t}}]$ a vector of daily covariates developed by repeating the annual values of each for each day of the year. These covariates (Level 1 in Figure 3) are used within a multinomial logistic regression with intercepts $\beta_{0j,i}$ and coefficients $\boldsymbol{\beta_{j,i}}$ to define the transition probabilities, with a prime denoting the vector transpose. The fitted multinomial regression can be used to estimate the time-varying transition probabilities and simulate WRs across the entire 600-year period over which reconstructed values of PC_{WR} are available. More information on the NHMM can be found in Section S1. We use this method to create a 50-member ensemble of daily, 600-year weather regime time series (Level 2 in Figure 3; convergence plots of corresponding streamflow available in Figure S1).

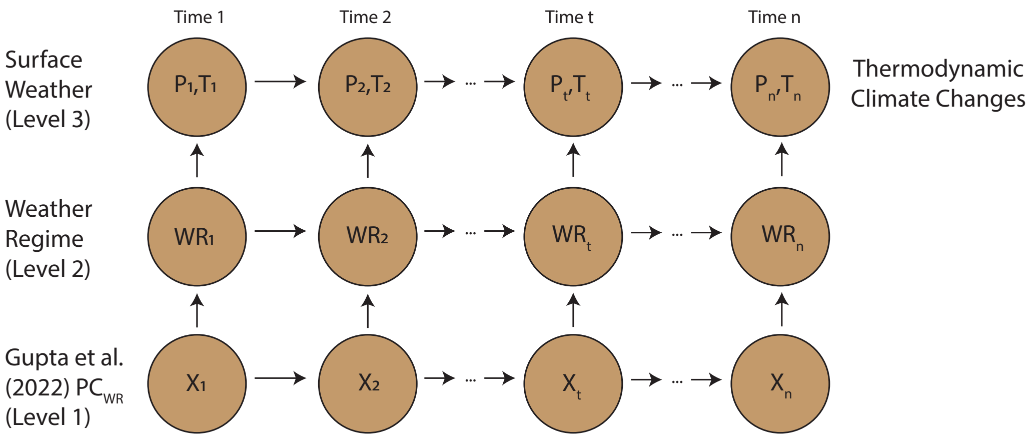

Figure 3. Flow chart of the main components of the weather generator. Boundary forcing variables (here the PC_{WR} from Gupta et al. (2022)) influence the evolution of the discrete weather regime time series. Surface weather is then conditioned upon the weather regime time series. Thermodynamic climate changes (temperature increases and precipitation-temperature scaling) are applied to the surface weather post-generation.

2.2 Generation of Local Surface Weather Conditioned on WRs

Time series of daily surface weather are generated based on the simulated time series of WRs (Level 3 in Figure 3). Here, observed daily precipitation, minimum, and maximum temperature are taken from the 1/16° resolution gridded meteorological dataset of Livneh et al. (2015) for water years (WY) 1950-2013. These historical weather data are block bootstrapped based on the sequence of simulated WRs to create new sequences of weather. For example, if the NHMM simulates a sequence of n consecutive days in WR i, an n-sized block of surface weather is resampled from the historical period that is also in WR i and that meets two other criteria: (1) the chosen historical block falls into a two-week window around the simulated day of the year; and (2) the day prior to the historical block is in the same precipitation state as the simulated day (wet or dry). The weather generator is run simultaneously across the five basins to create internally consistent (i.e., spatially correlated) weather across the region. Given their large synoptic scale, we use the occurrence of an atmospheric river (taken from Gershunov et al. (2017)) to represent a common precipitation state across basins. The process is repeated for each ensemble until a full sequence of 600 years of daily minimum and maximum temperature and precipitation has been generated.

Thermodynamic changes, in the form of shifts in temperature and precipitation scaling with warming, are imposed after surface weather is generated. Step changes in temperature between 0-4°C are added to each grid cell's simulated temperature. Quantile mapping is used to scale the precipitation distribution with warming, whereby the upper tail (99.9$^{\text{th}}$ percentile) of the non-zero precipitation distribution is made more intense, the lower tail of non-zero precipitation is suppressed downward, but the mean of daily precipitation is left unchanged. This scaling reflects an intensification of the precipitation regime and is consistent with GCM-based projections of future precipitation in California (Michaelis et al., 2022). We consider 5 different scaling rates equivalent to $0X$ (0% $°C^{-1}$), $0.5X$ (3.5% $°C^{-1}$), $1X$ (7% $°C^{-1}$), $1.5X$ (10.5% $°C^{-1}$), and $2X$ (14% $°C^{-1}$) the Clausius-Clapeyron (CC) scaling rate, which dictates how the moisture holding capacity of the atmosphere scales with warming. That is, the 99.9th percentile of non-zero precipitation is scaled up by either 0% $°C^{-1}$, 3.5% $°C^{-1}$, 7% $°C^{-1}$, 10.5% $°C^{-1}$, or 14% $°C^{-1}$, while the lower body of the distribution is scaled down accordingly to maintain the same distribution mean. The range of selected scaling rates are derived from observational and model-based studies that most often indicate extreme precipitation-temperature scaling at the $1X$CC rate, but occasionally suggest the possibility for sub-CC ($0X$, $0.5X$) or super-CC ($1.5X$, $2X$) scaling rates due to interactive effects between enhanced specific humidity and storm dynamics (Wasko et al., 2018; Martinkova & Kysely, 2020; Ali et al., 2022; Michaelis et al., 2022; Sun & Wang, 2022).

These scaling rates are combined with the different scenarios of warming, so that precipitation scaling is tied to the imposed temperature scenario and respects the thermodynamic mechanism that drives precipitation change. We consider five scenarios of warming, including 0°C, 1°C, 2°C, 3°C, and 4°C above the climatological average. This range of warming was inferred from an ensemble of CMIP6 mid-century (2015-2050) projections over central California under the SSP2-4.5 scenario, taken from CarbonPlan (see Figure S2; Chegwidden et al. (2022)). All together, we develop 25 different scenarios of climate change (5 temperature scenarios and 5 scaling scenarios), with each scenario containing 50 ensemble members (i.e., 50 stochastic 600-year time series of precipitation and temperature), in addition to a baseline ensemble with no changes imposed. Technical details on the quantile mapping procedure, and other details of the stochastic weather generator, are provided in Steinschneider et al. (2019) and Najibi et al. (2021).

2.3 Generation of Regional Streamflow Through Process-Based Hydrologic Models

Surface weather ensembles are used to simulate daily streamflow ensembles at the mouth of each of the five San Joaquin subbasins using the Sacramento Soil and Moisture Accounting Model (SAC-SMA) (Burnash et al., 1995) coupled with a SNOW-17 model (Anderson, 1976). The models, documented in Wi and Steinschneider (2022), are spatially distributed and utilize a Lohmann routing model Lohmann et al. (1998) to trace runoff from hydrologic response units (HRUs) through each river channel. The SAC-SMA models are calibrated using a pooled calibration approach (Wi et al., 2015) based on the average Nash Sutcliffe Efficiency (NSE) across the five subbasins simultaneously. Calibration and evaluation was based on historical Full Natural Flows (FNF) between WY 1989-2013, acquired from California Data Exchange Center (CDEC) FNF stations that lie within each subbasin: Tuolumne River at La Grange Dam (TLG), Friant Dam on Millerton Lake (MIL), Merced River near Merced Falls (MRC), New Hogan Lake (NHG), and New Melones Reservoir (NML) (G. Huang & Kadir, 2016). The models are calibrated over WY 1989-2003 and then evaluated across WY 2004-2013.

To verify that our streamflow extremes and variance decomposition results are not strongly dependent on the selection of the SAC-SMA model, we also employ the HYMOD conceptual hydrologic model (HYMOD; (Moore, 2007)) specifically in the Tuolumne Basin. Our primary results will be presented using the SAC-SMA model but more detailed analysis of how hydrologic model selection impacts the estimates of flood and drought metrics as well as their partitioning of variance is provided in Section S2. More information about the calibration process and parameter values for all hydrologic models can be found in Wi and Steinschneider (2022).

2.4 Metrics of Hydrologic Extremes

A series of flood and drought metrics, described below, are calculated for each ensemble member and each climate scenario and across two time horizons: 30 and 100 years. As stated in the latest update to the Central Valley Flood Protection Plan (CVFPP), the state of California is actively prioritizing investments in flood management over a 30-year planning horizon (California Department of Water Resources, 2022). A 100-year planning horizon is not actively used in the CVFPP, but it represents a time scale relevant to longer term major infrastructure investments. Further, it allows the exploration of the longer climate time horizon drivers. We partition the variance of each metric between the drivers of climate change and natural climate variability using the ensemble of scenarios described above. Appendix A contains a glossary with commonly used terms that are referred to through the methods. Appendix B contains a summarized list of all of the flood and drought metrics used in this study, including their decision relevance.

2.4.1 Flood Metrics

Flows associated with a 10-year and 100-year return period are used as flood metrics in this study. The 100-year floodplain currently drives larger riverine infrastructure development and flood risk management in California (California Department of Water Resources, 2022). Though not as common for current planning and management in California, the 10-year return period flow captures risk to smaller floodplains and drives smaller investments (California Department of Water Resources, 2006). The decadal and centennial flood are estimated by fitting a generalized extreme value (GEV) distribution to the three-day annual maxima at each CDEC gauged location in the five subbasins. The three-day flood was chosen because it a common metric used in flood risk assessments in California (California Department of Water Resources, 2006; Chung, 2009; Brekke et al., 2009; Maurer, Brekke, & Pruitt, 2010; Maurer, Hidalgo, et al., 2010), and because it better captures the concurrence of flooding across multiple basins (described further in section 2.4.3). For each ensemble member, we fit the GEV distribution for the whole 600-year paleo-period as well as across smaller 30-year and 100-year moving windows.

2.4.2 Drought Metrics

There is no state statutory definition of drought since it can be classified differently across impacted sectors and stakeholders. Historical hydrologic droughts have been traditionally identified based on a combination of metrics that capture the magnitude and duration of water deficit at key reservoirs (California Department of Water Resources, 2015). Since we develop metrics for gauged locations near these reservoirs, we opt to use a more generalized Standardized Streamflow Index (SSI) to quantify hydrologic drought (Vicente-Serrano et al., 2012). To calculate the SSI, daily simulated flows are first aggregated to a monthly time step. We then use a flexible non-parametric empirical method to estimate non-exceedance probabilities using the Gringorten plotting position (see Farahmand and AghaKouchak (2015)). To create the SSI, the associated non-exceedance probabilities are passed through the quantile function of the standard normal distribution, resulting in a series with an assumed mean of zero and standard deviation of one. We then use the SSI index to define three drought metrics, following McKee et al. (1993):

1. Drought Occurrence: The number of months characterized by an SSI value less than -1, divided by the total months in the window over which the metric was calculated. An SSI value of less than -1 captures moderate to severe drought hazard.

2. Drought Intensity: The minimum SSI value in the moving window.

3. Drought Duration: The maximum number of consecutive months with an SSI below -1.5 in the moving window. An SSI value of less than -1.5 captures severe drought hazard.

The SSI index is calculated for each ensemble member and climate change scenario, and the metrics are reported across 30-year and 100-year moving windows.

2.4.3 Copula-Based Flooding Metrics

The San Joaquin basin is a key component in the state's comprehensive water delivery system, and a levee breach due to compounding flooding across subbasins in the region could disrupt deliveries of irrigation water to 3 million acres of farmland in the

Central Valley (Taylor, 2017). Thus, we develop a spatially-compounding flood metric to capture this hazard. As discussed in Zscheischler et al. (2020), spatially compounding flood hazard can be characterized using an n-dimensional Gaussian copula that defines a metric of joint flood hazard across n basins simultaneously. Let $x_{t,1}, \ldots, x_{t,n}$ be the annual maxima of 3-day mean streamflow in each of the n basins in year t. We first fit GEV distributions to the individual three-day annual maxima for each basin ($i = 1, \ldots, n$). The three-day annual maxima in each year t are then transformed to be uniform pseudo-observations, $u_{t,i} = F_{GEV}^{-1}(x_{t,i})$, where F_{GEV}^{-1} is the inverse cdf of the fitted GEV distribution for basin i. These pseudo-observations are used to evaluate the joint CDF of the flood data based on a Gaussian copula:

$$C(u_{t,1}, ..., u_{t,n}) = P(U_1 \leq u_{t,1}, ..., U_n \leq u_{t,n}) = \Phi_n(\phi^{-1}(u_{t,1}), ..., (\phi^{-1}(u_{t,n})|\Sigma) \quad (2)$$

Here, ϕ^{-1} is the inverse CDF of the standard normal distribution and $\Phi_n(\cdot|\Sigma)$ is the multivariate normal CDF with zero mean and correlation matrix Σ, which is set equal to the Spearman rank correlation matrix for three-day annual maxima across basins. Using the fitted copula, we can then calculate the joint probability that multiple subbasins experience flooding above some threshold. For example, consider two subbasins with 100-year flood magnitudes of x_1 and x_2, respectively, inferred from their fitted (GEV) marginal distributions. Then, the probability that both subbasins simultaneously experience floods that exceed the 100-year flood is equal to (Zhang & Singh, 2019):

$$\begin{aligned} P(X_1 > x_1, X_2 > x_2) &= 1 - P(X_1 \leq x_1) - P(X_2 \leq x_2) + P_{1,2}(X_1 \leq x_1, X_2 \leq x_2) = \\ &1 - F_{GEV1}(x_1) - F_{GEV2}(x_2) + \Phi_n(\phi^{-1}(F_{GEV1}(x_1)), \phi^{-1}(F_{GEV2}(x_2))|\Sigma) \quad (3) \end{aligned}$$

Similar calculations are available to evaluate the probability that three or more basins experience flooding above set thresholds. These probabilities can be used directly as a metric of joint flood hazard, and we can partition the variance of this metric between climate changes and natural variability across our ensemble and for 30-year and 100-year moving windows.

2.5 Analysis of Variance in Hydrologic Metrics

We use an ANOVA to partition the variance in the hydrologic flood and drought metrics above into components attributable to different sources of variation. A two-way ANOVA was used to determine the uncertainty in hydrologic metrics attributable to uncertainty in temperature change (T), precipitation scaling rate (P), their interactions, and uncertainty in metrics attributable to natural variability. The temperature change factor has $i = 1, \ldots, 5$ levels (0, 1, 2, 3, 4 °C), and precipitation scaling factor has j=1,. . . ,5 levels (0%, 3.5%, 7%, 10.5%, 14% per °C). For each combination of levels, there are 50 stochastic realizations of the metric of interest. The linear model on which the ANOVA is based is given as:

$$x(i,j,s) = \mu + \alpha(i) + \beta(j) + \gamma(i,j)^{TP} + \varepsilon(i,j,s) \tag{4}$$

Where $x(i,j,s)$ is the hydrologic metric for a given level i and j of factors T and P, respectively, and a given ensemble member s. The grand mean for the metric x across the entire ensemble is μ; $\alpha(i)$ equals the average deviation in x from μ for ensemble members with temperature changes at level i; $\beta(j)$ equals the average deviation in x from μ for ensemble members with precipitation scaling rate at level j; $\gamma(i,j)^{TP}$ is the interaction term between temperature change and precipitation scaling; and $\varepsilon(i,j,s)$ is the error term, which is used here to represent natural variability in the metric not explained by the different climate change factors. The total sum of squares SS_{total} expresses the total variation in the hydrologic metric x, and is comprised of the sum of variation attributable to temperature change (SS_T), precipitation scaling rate (SS_P), their interaction (SS_{Int}), and natural variability (SS_ε):

$$SS_{total} = SS_T + SS_P + SS_{Int} + SS_\varepsilon \tag{5}$$

The fraction of variance attributable to each source is calculated by dividing each component by SS_{total}. This fraction of attributable variance is calculated separately in 30-year and 100-year rolling windows for each of the metrics above.

3 Results

The results of this work are presented as follows. First, Section 3.1 shows a comparison of the variability in the paleo-informed streamflow with events from the available observed historical record. Then, Section 3.2 shows the flood and drought extremes reconstructed for the baseline scenario (i.e., influence of natural variability alone). Section 3.3 demonstrates how the imposed climate changes affect those extremes. Section 3.4 demonstrates the variance partitioning of extremes across climate change and natural variability. A more detailed evaluation of the stochastic weather generator's performance is presented in the Supporting Information (see Figures S3-S7), which demonstrates how well the generator captures characteristics of precipitation and minimum and maximum temperature.

3.1 Paleo-Informed Streamflow Characteristics

Figure 4 demonstrates the broader variability that is attained in the streamflow ensembles when SAC-SMA is forced with paleo-reconstructed weather at the Don Pedro gauge in the Tuolumne Basin. Figure 4a focuses on 7-day flows and the lower tail of the distribution and Figure 4b zooms in on the upper tail distribution of 3-day flows. Each grey line represents sorted flow volumes across 30-year chunks of the paleo-reconstruction across all 50 ensemble members. These volumes are compared with those that come from forcing the generator over the modern period (1987-2013) with historical Livneh precipitation and temperature data (red line). Key events from the observed record are annotated as colored horizontal lines. Overall, the paleo-informed streamflow envelopes and expands upon the historical SAC-SMA model flows by creating instances of wetter 3-day flows and drier 7-day flows. Furthermore, the paleo-ensemble is characterized by drier events than key drought periods from the observed record as demonstrated in Figure 4a. The generator is unable to create 3-day flows that reach the peak of the 1997 New Year's flood period due to underestimation of precipitation associated with this storm that is a known error in the Livneh dataset (Pierce et al., 2021). In turn, models conditioned on the Livneh dataset tend to underestimate the flows associated with this event. However, the inclusion of the paleo-reconstruction allows the generator to create flows that far surpass the magnitude of peak flows associated with the 1995 and 2017 floods. Overall, the expanded envelope of daily scale streamflows enabled by the paleo-reconstruction provide rich context for exploring plausible flood and drought extremes in the Tuolumne

Basin. Figure S9 demonstrates similar results for the rest of the San Joaquin River basins, particularly in capturing drought dynamics. The generator conditioned on the Livneh dataset suffers from the same difficulty of capturing the 1997 flood peak flows; however, in some basins like Merced and Millerton, the paleo-conditioned generator provides extended variability that can help overcome these limitations (Figures S9b,d). New Hogan Lake is the only gauged location in which the Livneh-conditioned model can capture the 1997 flood peak flows, but this is primarily because the associated peak flows were not as extreme in this region relative to other notable flooding events. Of the five basins, capturing dynamics in the Tuolumne is the most challenging; it is also representative of high-elevation basins that exhibit rich snow dynamics. Thus, we proceed through the rest of the results with a focus on the Tuolumne Basin, though corresponding figures for the rest of the basins can be found in the supplement. Section 3.2 further elaborates on the value of the paleo-forced generator and its representation of key flood and drought metrics through the reconstruction.

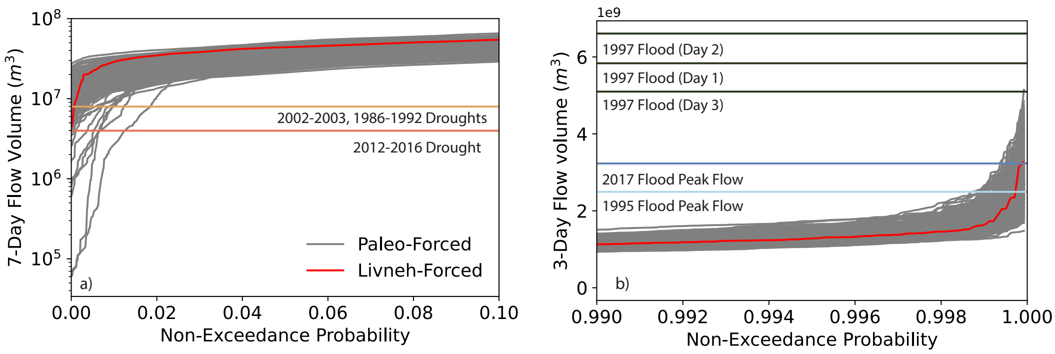

Figure 4. a) 7-day and b) 3-day flow volumes at the Don Pedro gauge in the Tuolumne Basin derived from the paleo-informed streamflow ensembles compared to the Livneh-forced generator over the modern period. Key events from the observed record are shown as colored lines. Each grey line represents sorted volumes for each year in 30-year chunks of the paleo-reconstruction across all 50 ensemble members.

3.2 Reconstruction of Natural Variability in Extremes

3.2.1 Individual Basin Flood Hazards

The individual Tuolumne subbasin flood hazard is quantified based on the 10-year and 100-year flood events associated with 3-day annual maximum flows, calculated using a GEV distribution fit to 3-day maxima in each basin and with two moving windows of length 30 and 100 years. Figures 5a and 5c show these return levels at the Don Pedro gauge in the Tuolumne Basin using a 30-year moving window. The return levels are calculated for all ensemble members of the baseline generator, where the solid line represents the mean return level across the ensemble members and the shading represents the $5^{th}/95^{th}$ percentiles. Figures 5b and 5d are non-exceedance plots of the three-day annual maxima across the extent of the paleo-reconstruction ensemble. The dashed black line represents the three-day annual maxima associated with the 10-year and 100-year return period events as derived from the SAC-SMA model forced with Livneh historical precipitation and temperature that overlaps with the observed record (1987-2013). In order to facilitate the most equivalent comparison between the two datasets, each gray line represents the sorted three-day annual maxima volumes over sets of 30-year segments of the paleo-reconstruction and across all 50 ensemble members.

The return levels in Figures 5a,c both show clear peaks centered around 1600 CE, which highlights a prominent pluvial period in the region's past hydroclimate. This pluvial is represented in the original WR reconstruction from Gupta et al. (2022) and broadly confirmed by other reconstructions (D'Arrigo & Jacoby, 1991; Schimmelmann et al., 1998; Stahle et al., 2007; M. D. Dettinger & Ingram, 2013). M. D. Dettinger and Ingram (2013) have also reconstructed pluvials around 1750-70 CE and 1810-20 CE, and while less pronounced than the 1600s pluvial, both panels a) and c) show increases in three-day annual maxima during these times. When compared to the model-based modern hydrology (dashed black line), both figures suggest that return levels in the most recent 30-year period are lower than those that have been experienced in prior centuries of the paleo-period reconstruction. Panels b) and d) show the modern estimates of the three-day annual maxima for the 10-year and 100-year events respectively, in comparison with the extent of the three-day annual maxima created by the paleo-informed generator. The ensemble from the generator encompasses the modern estimates of the return levels and also provides many instances of more extreme flooding events, which provides additional challenging flood scenarios that can be used to understand the vulnerability of water systems in each of the Central Valley subbasins explored in this study. As shown in Figure S10, the rest of the basins display similar three-day annual maxima dynamics, though the magnitude of the flows differs across all basins and return periods. Lower peak flows tend to be associated with basins that are smaller in area, elevation, and slope (i.e., New

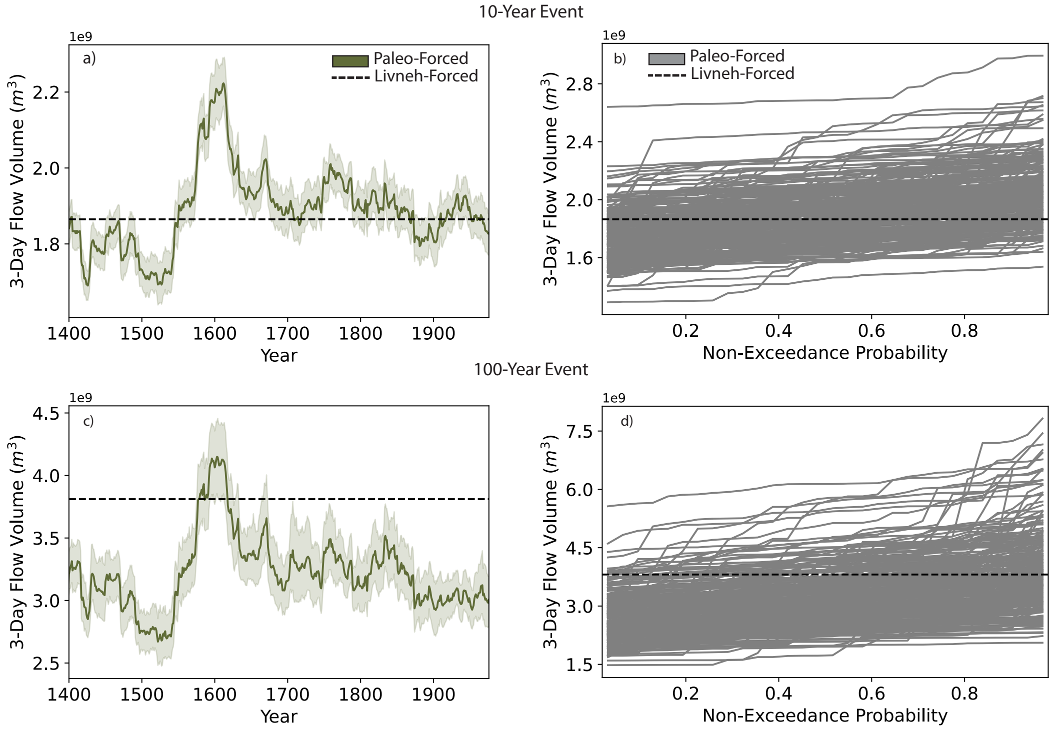

Figure 5. Three-day annual maxima associated with the a) 10-year return period event and c) 100-year return period event for the Don Pedro gauge in the Tuolumne subbasin calculated in 30-year moving windows and across the time period from 1400-2017. The dark green line represents the mean flooding return levels and the shading represents the 5^{th} and 95^{th} percentile confidence bounds. Panels b) and d) are non-exceedance plots of the three-day annual maxima across the extent of the paleo-reconstruction ensemble. Each gray line represents the sorted three-day annual maxima volumes for each year in a 30-year segment of the paleo-reconstruction. The dashed black line represents the three-day annual maxima associated with the 10-year and 100-year return period events as derived from the SAC-SMA-simulated peak flows when forced with Livneh historical data (1987-2013).

Hogan Lake, Table S1). The ensemble member spread also tends to be larger for the more extreme and uncertain 100-year flood event. Panels a) and c) exhibit clear non-stationary tendencies in the representation of the 10-year and 100-year event across the reconstruction that have large implications for hazard characterization. For example, the flow volumes associated with the 10-year event during the 1600s wet period are within range of the 100-year event flows during the 1500s megadrought period. Thus, what may be considered a 10-year flood event in one wet period transitions to be a 100-year event in a dry period. This extent of variability uncovered in the flood metric demonstrates that using only the modern record to define design flood events could severely under-represent flood hazard in the Central Valley region and that defining hazard based off of the 10-year and 100-year flooding events has drastically changed over time.

3.2.2 Individual Basin Drought Hazards

Figures 6a,c,e show the three SSI-based hydrologic drought metrics (occurrence, duration, and severity) calculated across a 30-year moving window for the period of 1400-2017 for the Don Pedro gauge in the Tuolumne River Basin. Figures 6b,d,f are non-exceedance plots, where each line corresponds to the sorted drought metric values derived across the whole reconstructed 617-year record length for each of the 50 ensemble members. The dashed line represents the respective metric values derived from the SAC-SMA model flows forced with Livneh historical precipitation and temperature across the length of the modern record. Similar to the flooding metrics in Section 3.2.1, the drought metrics exhibit clear decadal-scale variability that is also present in the original WR reconstruction from Gupta et al. (2022). For example, Figures 6a,c,e show declines in drought occurrence, severity, and duration during the early 1600s pluvial, while these drought characteristics become more intense during the 1500s megadrought that lasted from the middle of the century to the late 1580s (Stahle et al., 2007). The rest of the San Joaquin subbasins display this key behavior as well (Figure S11). The drought metrics reveal a slight long-term trend toward higher drought occurrence, longer duration, and more intense drought severity through the last three centuries of the reconstruction. This trend could, in part, be driven by key persistent drought periods that occurred in the mid to late 1800s (1856-1865, 1870-1877, and 1890-1896; Herweijer et al. (2006)), the 1900s (the Dust Bowl in the 1930s and drought periods in the 1950s and late 1980s; (Stahle et al., 2007)) and the most recent 20-year drought periods in the 2000s. The black line demonstrates drought occurrence and severity that is on par with the late 1500s megadrought, though exhibiting a slightly shorter duration than a large section of the paleo-reconstruction. The shorter drought duration is likely due to the sporadic periods of wet weather that have characterized the most recent 30-year period, including the early 1980s and late 1990s (M. Dettinger & Cayan, 2014) and periods after each drought instance in the 2000s.

Panels b), d), and f) compare the modern drought metrics to those calculated from the paleo-reconstructed ensembles. The ensembles encompass the modern estimates and

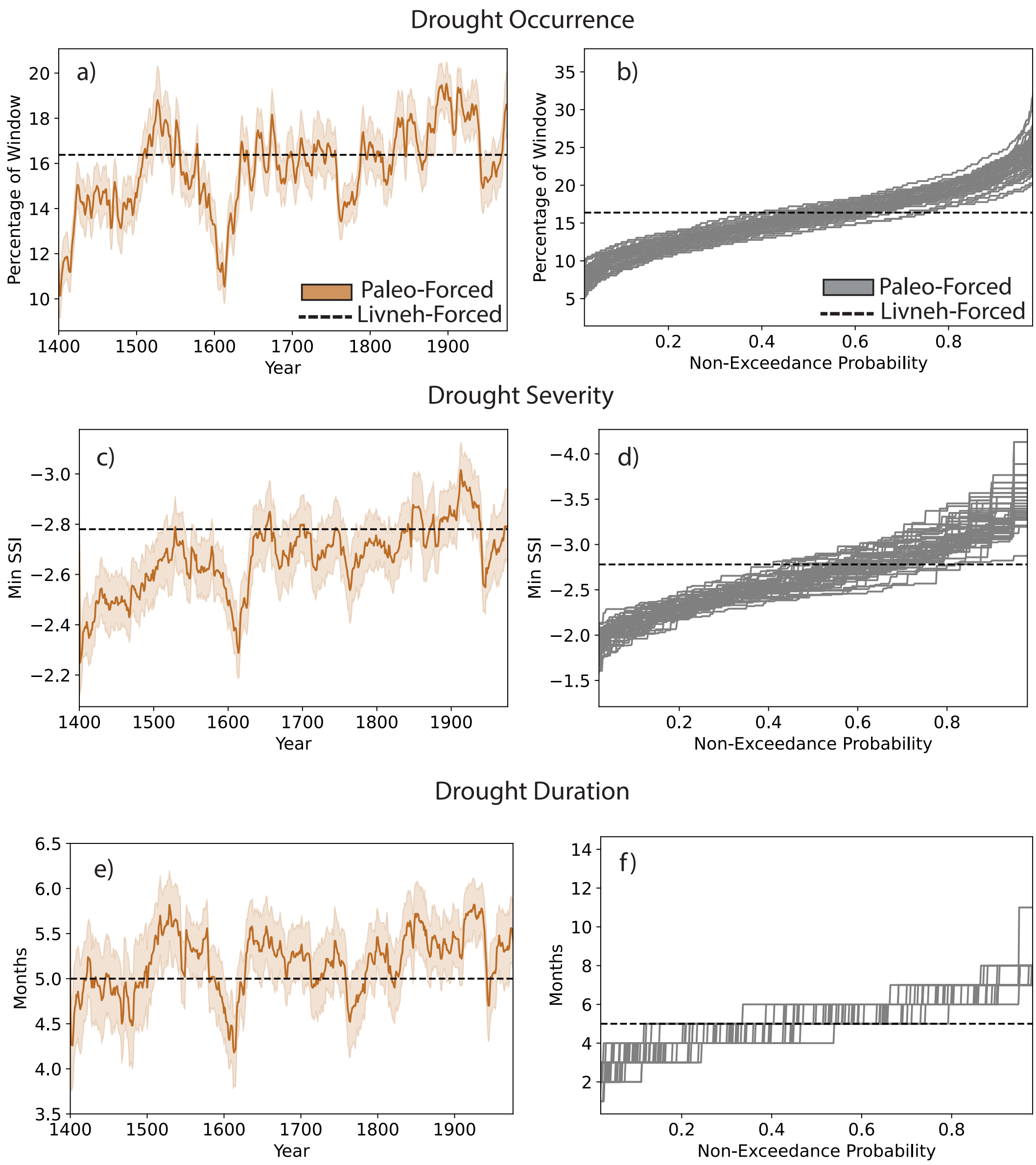

Figure 6. SSI-based hydrologic drought metrics of a) occurrence c) severity, and d) duration for the Don Pedro gauge in the Tuolumne Basin calculated in 30-year moving windows and across the time period from 1400-2017. The dark tan line represents the mean drought metric value and the shading represents the 5^{th} and 95^{th} percentile bounds. Panels b),d), and f) are non-exceedance plots of the three-day annual maxima across the extent of the paleo-reconstruction ensemble. Each gray line represents the sorted three-day annual maxima volumes across the length of the paleo reconstruction. The dashed black line represents the metric values as derived from the SAC-SMA-simulated peak flows associated with the modern record (1987-2013).

also provides many traces that are characterized by more frequent, longer, and severe drought. The plausibility of the Central Valley subbasins confronting drought conditions that extend well beyond those that have been experienced in the modern observed record captured in Livneh forcing data is significant even in the absence of climate change. The traces in panels b), d), and f) emphasize the need to better characterize the subbasin systems vulnerabilities for the challenging drought conditions that are captured within the reconstruction.

3.2.3 Joint Flood Hazard Across Basins

Gaussian copulas were fit to the 3-day annual maxima flows for multiple combinations of basins to characterize joint flood dynamics. The joint probability of flows at Don Pedro in the Tuolumne Basin and at Millerton Lake in the Millerton Basin simultaneously exceeding their respective, GEV-based 100-year flood estimates from the most recent 30-year period from 1987-2017 was calculated for the length of the reconstruction. Figure 7a shows the expected return period associated with those probabilities. Figure 7b includes New Melones Lake into the joint probability estimation. The return periods are calculated using a 30-year moving window across the entire reconstruction. Panels c) and d) are non-exceedance plots of the respective return periods across the extent of the paleo-reconstruction ensemble. The dashed black line represents the return periods for the 10-year and 100-year flood derived from the SAC-SMA model forced with Livneh historical precipitation and temperature. As with the flood metrics, in order to facilitate the most equivalent comparison between the two datasets, each gray line represents the sorted return periods for 30-year segments of the paleo-reconstruction and across all 50 ensemble members.

As demonstrated in Figure 7, there is a strong increase in the likelihood of simultaneously exceeding the recently observed historical estimate of the 100-year event, particularly during the 1600s wet period ($\sim$20% increase in likelihood). That is, the expected frequency of occurrence of simultaneous 100-year flooding events in both the Tuolumne and Millerton jumps to once every 320 years, as compared to once every 405 years in the most recent 30-year period. There is also a significant decline in the likelihood of joint flooding during the late 1500s megadrought. When an additional basin is introduced into the copula-based metric, the overall temporal dynamics are similar (Figure 7b), but the expected return period increases significantly. That is, the likelihood of simultaneously

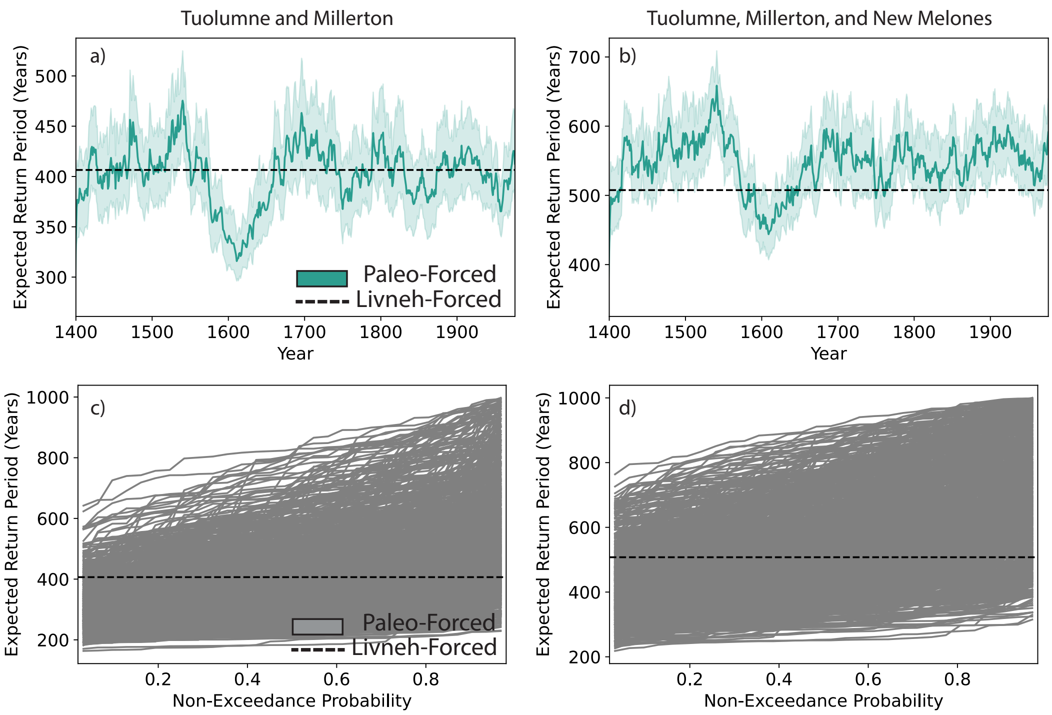

Figure 7. The expected return periods associated with the joint probability of simultaneously exceeding historical 100-year flood flows at a) Don Pedro (Tuolumne Basin) and Millerton Lake (Millerton Basin), and c) including New Melones (Stanislaus Basin) calculated in 30-year moving windows across the time period from 1400-2017. The dark turquoise line represents the average return period respectively across the ensemble, and the shading represents the 5$^{\text{th}}$ and 95$^{\text{th}}$ percentile bounds. Panels b) and d) show the non-exceedance plots for the return periods derived across the whole paleo-reconstruction in 30-year segments. The dashed black line represents the return periods as derived from the SAC-SMA-simulated peak flows associated with the modern record (1987-2013).

exceeding historical flooding thresholds rapidly declines as more basins are considered. During the 1600s wet period, the expected frequency of occurrence of simultaneous 100-year flooding events in the Tuolumne, Millerton, and New Melones jumps to once every 450 years, as compared to once every 507 years in the most recent 30-year period. For both joint flood metrics, the paleo-reconstruction effectively bounds the modern estimation of the return periods, which provides a richer space to characterize joint flood hazards across the subbasins (Figures 7c and 7d).

Similar non-stationary dynamics as observed in the flooding metrics in Section 3.2.1 are apparent in these joint flooding metric as well. Simultaneous flooding in all three basins is rarer and more consequential for water systems planning and management than simultaneous flooding in the Tuolumne and Millerton alone. Figures 7a-b demonstrate that through the paleo-reconstruction, there are periods (like the 1600s wet period) where the likelihood of flooding in the three basins becomes just as common as flooding in the Tuolumne and Millerton alone (around the late 1500s megadrought). The additional variability that the reconstruction provides demonstrates how dramatically the return periods associated with these consequential events changes over time, particularly how these flooding events can become more frequent. Once again, using the modern record to quantify joint hazard across these subbasins could severely underrepresent flood hazards and the magnitude of design events.

3.3 Effects of Thermodynamic Climate Change on Hydrologic Extremes

3.3.1 Changes in Individual Basin Flood Hazard

Figure 8 shows the effect of thermodynamic climate changes on the 100-year, 3-day flood event in the Tuolumne calculated across 30-year moving windows. The flow volumes are represented as deviations from the baseline reconstruction which is shown as a gray dashed line at 0. A modern baseline is placed as a dashed black line and is representative of the difference between the modern and the largest 100-year flood event volume calculated across the reconstruction. Figure 8a shows scenarios where the precipitation scaling rate is kept at 7% $°C^{-1}$ while temperature is increased by 1, 2, and 3 $°C$, while Figure 8b shows scenarios where the temperature trend is maintained at 1°C and the precipitation scaling rate is increased to 0% $°C^{-1}$, 7% $°C^{-1}$, and 14% $°C^{-1}$. Both increasing precipitation scaling rates and temperature trends shift the 100-year flood peak flows upwards, though temperature trends have a stronger impact. For reference, the volume differential between the extreme scenarios in Figure 8a is equivalent to about 100 Oroville Dams worth of water. Conversely, the maximum volume differential associated with the precipitation scaling in Figure 8b is equivalent to 33 Oroville Dams worth of water. The Tuolumne is a snow-dominated basin, and consequently it is not unexpected that the results suggest a greater influence on 100-year flows resulting from increasing temperature rather than increased precipitation scaling. Increased temperature shifts drive increased snowmelt and rain on snow events that promote greater flood volumes.

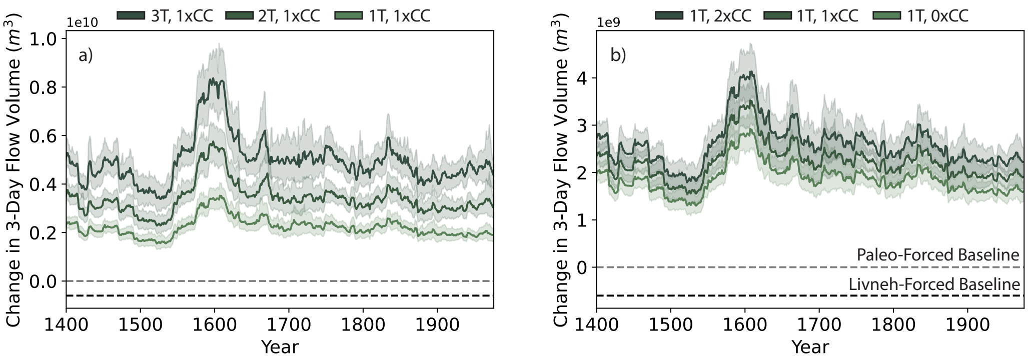

Figure 8. The effect of increasing a) temperature and b) precipitation scaling rates on 100-year, 3-day flood flows at Don Pedro (Tuolumne Basin). The dark green lines represent the increase in mean flooding return levels with respect to the baseline scenario (gray line at 0) and the shading represents the 5th and 95th percentile bounds. A modern baseline (black line) is included as reference and represents the distance from the modern peak flow to the maximum peak flow recorded in the reconstruction.

3.3.2 Changes in Individual Basin Drought Hazards

Figure 9 shows how the same thermodynamic scenarios imposed in Section 3.3.1 influence drought occurrence in the Tuolumne Basin, measured in terms of a change in the percent of the 30-year window that is classified to be in drought conditions with respect to the baseline scenario (gray dashed line at 0). A modern baseline is placed as a dashed black line and is representative of the difference between the modern drought occurrence line from Figure 6a and the worst drought occurrence metric calculated across the reconstruction. An increase in each of the thermodynamic mechanisms tends to increase the percentage of the window classified in drought. A comparison across Figures 9a and 9b show the larger impact of temperature trends on increased drought occurrence (reaching up to 5% or an additional 18 months classified in drought) by way of increased evapotranspiration. Precipitation scaling stretches the daily precipitation distribution which can lead to tail influences that impact the total number of drought months, but has a lower relative influence (reaching up to 1.8% or an additional 6 months classified in drought). For example, there are some instances, particularly in the 1T, 1xCC scenario in Figure 9a that result in values that approach the baseline. This is likely due to the precipitation scaling mechanism causing some months to have an increased SSI above the drought threshold that offsets the temperature increase. However, as the temperature shift further increases, this effect is dominated. Figure S12 shows the same results for drought severity and duration. Overall, there is a greater influence from increasing temperature trends to increasing drought severity and duration. It's worthwhile to note that the impact from both temperature trends and precipitation scaling is relatively small (Figure S12c,d) with respect to increasing consecutive months classified in severe drought and these results are further reflected in Figure 12.

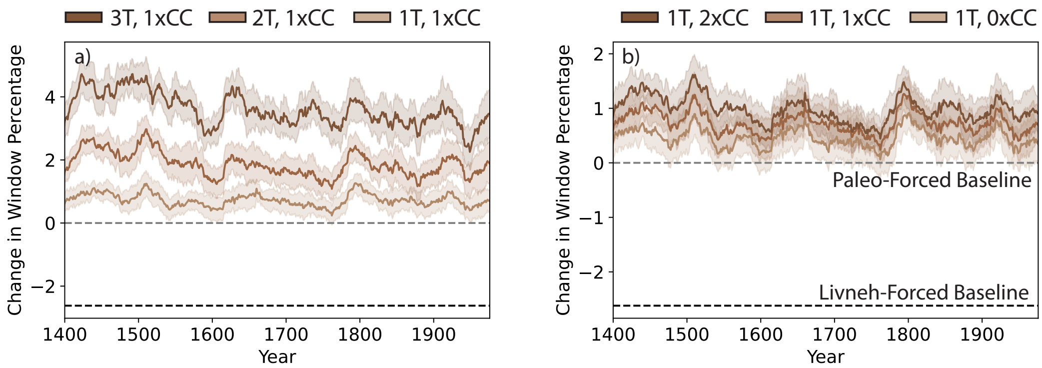

Figure 9. The effect of increasing a) temperature and b) precipitation scaling rates on drought occurrence at Don Pedro (Tuolumne Basin). The dark brown lines represent the increase in the percentage of the 30-year window classified in drought conditions with respect to the baseline scenario (grey line at 0) and the shading represents the 5th and 95th percentile bounds. A modern baseline is included (black line) as a reference and represents the distance from the modern drought occurrence metric to the worst drought occurrence recorded in the reconstruction.

3.3.3 Joint Flood Hazard Across Basins

Figure 10 shows how similar thermodynamic scenarios influence joint flood hazard at Don Pedro (Tuolumne Basin) and Millerton Lake (Millerton Basin), measured in terms of change to return period associated with the 100-year event with respect to the baseline scenario (gray dashed line at 0). As with the prior sections, a modern dashed black baseline is included to represent the difference between the modern return period estimate and the lowest return period calculated across the reconstruction. Much like Figure 8, Figure 10 demonstrates a larger influence from increasing temperature trends on making compound flooding events more likely (Figure 10a). Given that the Tuolumne and Millerton are both snow-dominated basins, temperature trends create similar snowmelt effects that lead to simultaneous flooding events. Precipitation scaling has a relatively reduced, but non-trivial effect (Figure 10b). The greatest influence from precipitation scaling is observed under higher imposed temperature trends (we use a constant 3°C temperature trend in this example). While an increase in precipitation scaling increases the likelihood of flooding in any given basin (Figure 8b), Figure 10b demonstrates that it decreases the likelihood of joint flooding, and makes the events rarer by increasing the return period. Since the imposed precipitation scaling non-linearly adjusts peak flows, it ultimately leads to a decrease in correlation in flows across the two basins and therefore a decrease in joint flooding tendencies.

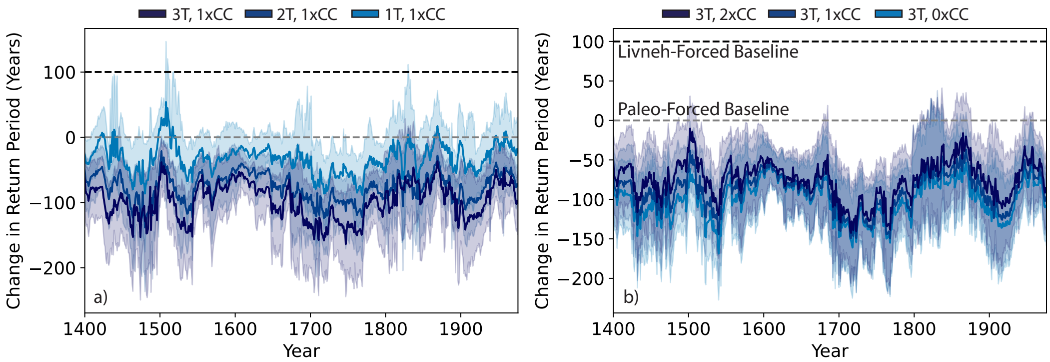

Figure 10. The effect of increasing a) temperature and b) precipitation scaling rates the change in return period associated with simultaneously exceeding historical 100-year-day flood flows at Don Pedro (Tuolumne Basin) and Millerton Lake (Millerton Basin). The dark blue lines represent the change in return period with respect to the baseline scenario (gray line at 0) and the shading represents the 5^{th} and 95^{th} percentile bounds. A modern baseline (black line at 0) is included as reference and represents the distance from the modern return period to the shortest return period recorded in the reconstruction.

3.4 Variance Partitioning of Hydrologic Extremes

The results above show how different metrics of hydrologic extremes vary significantly over time due to natural climate variability as well as different mechanisms of climate change. Below we use variance partitioning to assess the relative importance of these competing factors.

3.4.1 Relative Variance Contributions for Individual Basin Flood Hazard

We conduct an ANOVA to partition the variance of the 10-year and 100-year 3-day floods for each gauged location. Figure 11 shows the results for Don Pedro, while results for the other sites are shown in Figure S13-S16. The columns show the results of the decomposition when flood metrics are derived with a 30-year, 100-year, and 617-year (whole record) time horizon, respectively.

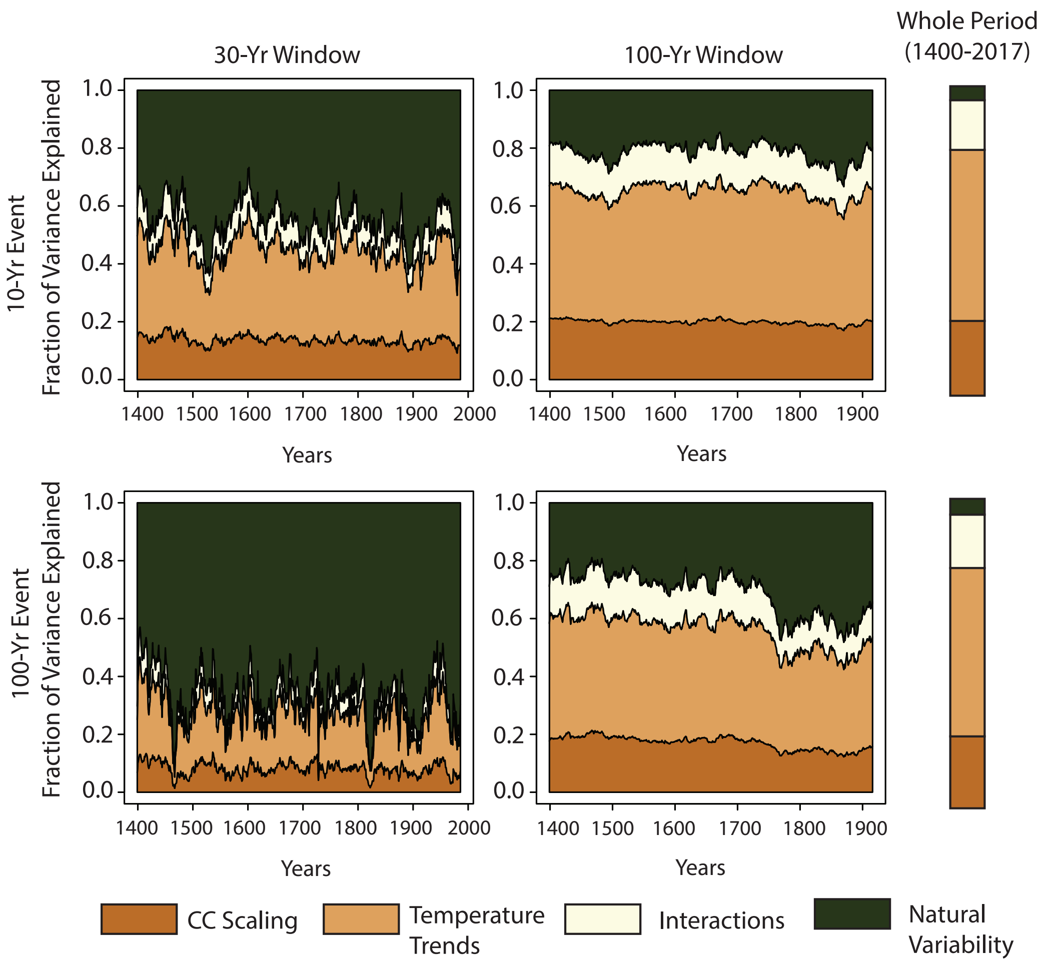

Figure 11. A decomposition of the key drivers of variance in the flood metrics for the Don Pedro gauge in the Tuolumne River Basin for an a,d) 30-year time horizon b,e) 100-year time horizon and c,f) a 617-year time horizon.

Two main insights emerge from Figure 11. First, natural variability is the primary driver of the variance when the flood metrics are calculated using a 30-year time horizon (Figures 11a,d). This is especially true for the 100-year flood, where approximately 70% of the variance in this metric is associated with natural variability. Figure 11d has direct relevance to the design standards actively used to inform California's flood planning and management. However, the influence of natural variability on the spread in flood metrics across the ensemble substantially decreases when the metric is calculated across a 100-year time horizon (Figures 11b,e), and becomes almost negligible when calculated over the entire 617-year period (Figures 11c,f). This suggests that the time horizon over which the flood metrics are calculated highly influences the perception of key drivers. A longer time horizon more clearly captures the effects of longer-term climate change on the variation in the flood metrics, while during shorter windows the variation in flood metrics across the ensemble is more likely to capture noise associated with natural variability. The reasons for this are twofold. First, when the time horizon is large, each ensemble member for a particular climate change scenario contains many annual maxima that are all drawn from the same underlying climate state, helping to converge design event estimates across ensemble members towards similar values. Second, when the time horizon is large, there are more opportunities for climate change signals to influence the distribution of annual maxima flows for all ensemble members under a given climate change scenario, which will help separate the distribution of annual maxima across the different scenarios. Together, these two factors will lead to more variance in the overall ensemble being explained by the climate change scenarios compared to natural variability.

Of the thermodynamic changes, temperature trends are the primary driver of variation in peak flows, followed by precipitation scaling. This result, also seen in Figure 8, suggests that temperature increases that lead to increased snowmelt and rain on snow events influences peak flows in the region more than increases in extreme precipitation due to increased moisture in the atmosphere. The interactions between the two drivers generally accounts for a smaller percentage of the variance, but as the time horizon increases, interactive effects are close to the same magnitude as precipitation scaling (16% vs. 24% for the whole period). This result highlights how the effects of precipitation scaling are dependent on the temperature increase, because precipitation scaling is parameterized as a percentage change in extreme precipitation per °C warming.

Figure S13-S16 show the same results for the remaining four basins. Overall, all basins exhibit similar behavior, where the influence of natural variability decreases with time horizon. Temperature change has a larger impact than precipitation scaling in all basins except for New Hogan Lake (Figure S15). New Hogan Lake is relatively small, has a low elevation, and less snow dominated compared to the other basins (Table S1), and thus sees a greater influence from precipitation scaling on flood variability.

Overall, the results in Figure 11 portray conflicting storylines and complexity for flood planning and management depending on the way the flood metrics are defined. Under current CA planning conditions (represented in Figure 11d), the greater influence of natural variability on individual flood hazard would suggest prioritizing short-term adaptive tools like seasonal forecasts. However, under alternative planning scenarios that may utilize longer time horizons, infrastructure investments look to be more useful to manage hazards from thermodynamic climate changes. Most importantly, water planners will need to engage with both drivers; prioritizing longer horizons of focus could neglect the effects of internal variability in the near term, which as Figure 5 portrays, can lead to magnitudes of peak flows that far surpass those in the modern record. Ultimately, there needs to be consideration of both the exceptional magnitude of internal variability in more immediate decision relevant 30-year timescales while still being cognizant of the longer-term climate changes. Thus, it's important for water resources agencies that utilize dynamic and adaptive planning methods to effectively balance the value, resilience, and potential regrets of near term investments (e.g. Haasnoot et al. (2013); Schlumberger et al. (2022)).

3.4.2 Relative Variance Contributions for Individual Basin Drought Hazards

Figure 12 shows the ANOVA decomposition for drought occurrence, intensity, and duration for 30-year and 100-year moving windows, as well as the entire 617-year period. The variance partitioning for drought occurrence follows a similar pattern to the flood metrics above (Figures 12a-c). For short time horizons of 30 years, about 20-40% of drought occurrence variability across the ensemble is associated with natural variability. However, as the time horizon grows, more variance is partitioned to the climate changes, and for extremely long horizons, almost all of the variance in drought occurrence across the ensemble is associated with climate change. Specifically, temperature change becomes the near-sole driver of drought occurrence variability, likely because of the strong increases in evapotranspiration with warming that drive drought occurrence.

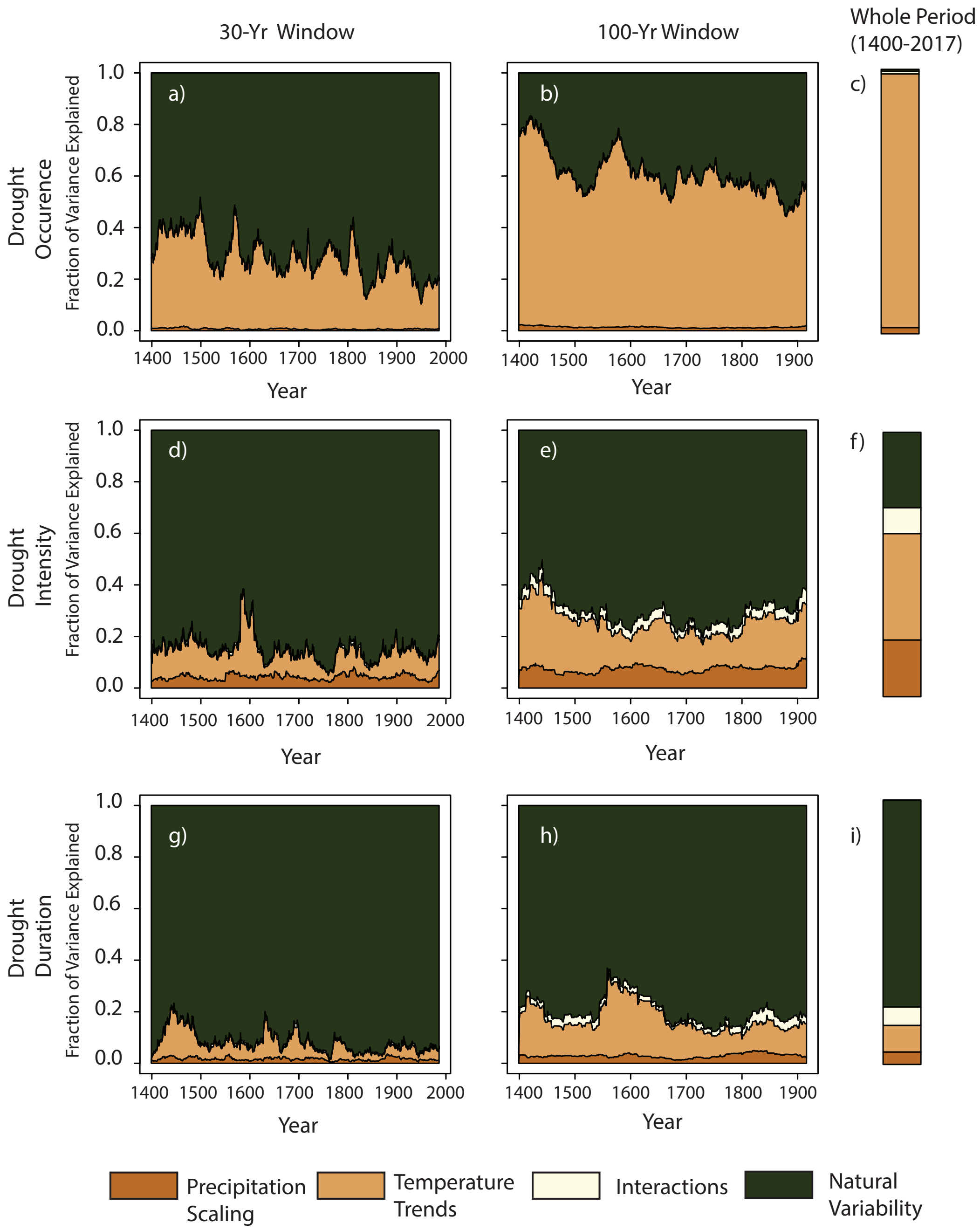

Figure 12. A decomposition of the key drivers of variance in the drought metrics for the Don Pedro gauge in the Tuolumne River Basin for a,d,g) 30-year window b,e,h) 100-year window and c,f,i) a 600-year window.

For drought intensity, we see a similar pattern in variance partitioning between natural variability and climate change factors, but the magnitude and degree of change in the variance partitioning more heavily favors natural variability (Figures 12d-f). For 30-

year windows, natural variability accounts for upwards of 80% of the total variance in drought intensity, and this falls to the (still substantive) value of 28% when the window reaches 617 years. Of the climate changes, temperature trends once again are the main driver, but precipitation scaling and interactive effects also play an important role in drought intensity variability across the ensemble. Given that the mechanism of precipitation scaling stretches the daily precipitation distribution such that large precipitation values become larger and small precipitation values become smaller, we see a more significant influence from this mechanism on drought intensity than in the other metrics.

Unlike the other two drought metrics, drought duration is primarily driven by natural variability, even when the metric is derived across the longest window. Drought duration generally is linked to the length of time in which there is no precipitation. None of the imposed climate changes directly affects this behavior in the same manner that precipitation scaling directly influences drought intensity or temperature trends affect drought occurrence. Temperature increases can somewhat extend drought duration by increasing evapotranspiration at the beginning and end of a drought period (Figure 12h), but ultimately the duration of a drought is dictated by the occurrence of large storms that end the drought, which is primarily driven by natural variability in our climate scenarios. The decomposition results for the remaining four gauged locations are presented in Figure S17-S20. These gauged locations show similar behavior as the Don Pedro gauge. Temperature trends play a large role in influencing drought occurrence, and this influence is particularly large in Merced and New Melones Lake (S17a, S20a). Precipitation scaling plays a small role in drought occurrence, and drought duration is primarily driven by natural variability.

The drivers of drought are more complex than the flood hazard metrics due to the heterogeneity of behavior across the drought metrics. A comparison between Figures 12a,d, and g demonstrate vast differences in drivers (and therefore approaches for managing drought) depending on exactly what characteristic of drought is prioritized in planning. The choice of time horizon further complicates the understanding of the appropriate planning process, especially in the case of drought occurrence (Figures 12a,b). However, drought intensity and drought duration show more stable influence primarily by natural variability and would consequently need a mix of carefully coordinated shorter-term adaptive actions (e.g., water transfers, conservation, and shifts in allocative priorities to higher value uses) that provide flexibility to improve the robustness of longer-term infrastructure investments to extreme variability in Central Valley drought regimes (e.g., improved conveyance, groundwater banking, managed aquifer recharge, and others; Herman et al. (2020); Hamilton et al. (2022)).

3.4.3 Relative Variance Contributions for Joint Flood Hazard

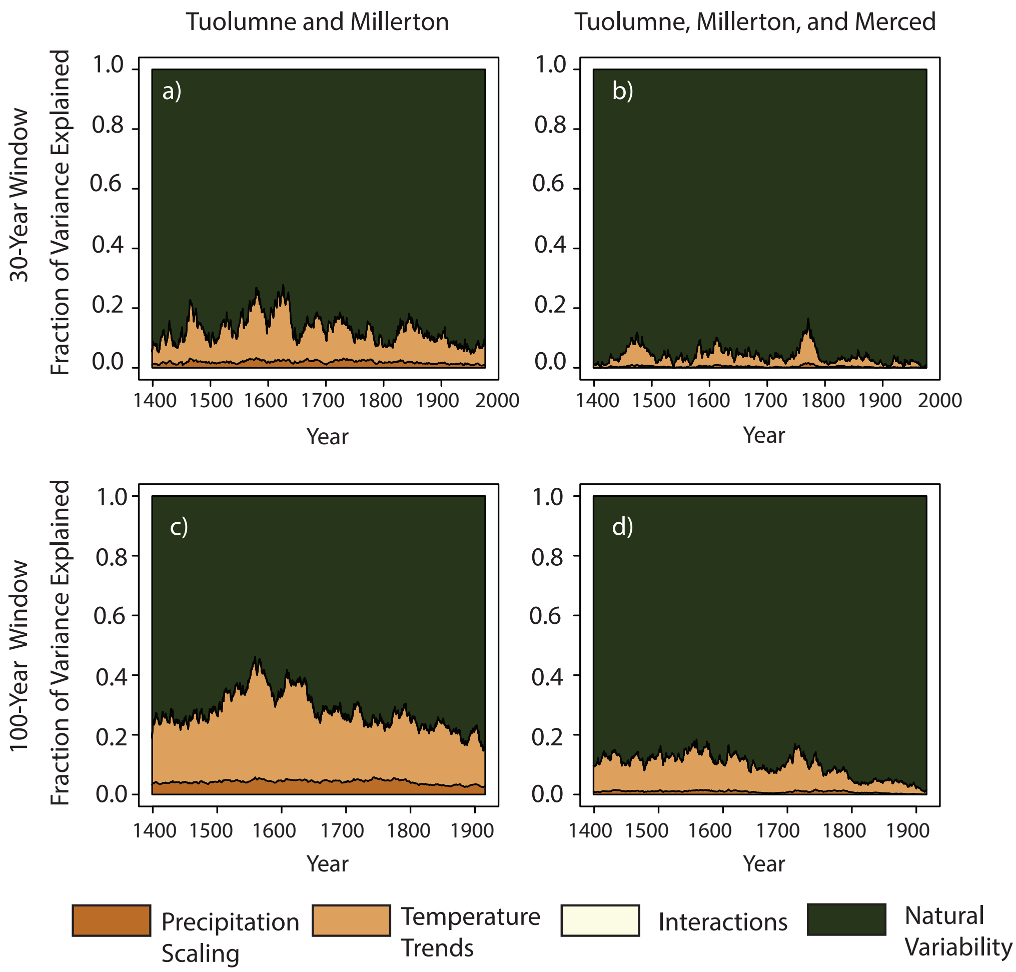

Figure 13. A decomposition of the key drivers of variance in joint flood metrics for a),c) Tuolumne and Millerton and b,d) Tuolumne, Millerton, and Merced.

Figure 13 shows the variance partitioning for the copula-based joint flood hazard metric in two cases: (1) bivariate flood risk in the Tuolumne and Millerton (Figure 13a,c); and (2) trivariate flood hazard in the Tuolumne, Millerton, and Merced (Figure 13b,d), both for the 100-year, 3-day flood. In both cases, the primary driver of joint flood hazard is natural variability. Unlike flood hazard for individual basins (see Figure 11), the contributions of natural variability to the total variance joint flood hazard does not decline substantially with time horizon. Additionally, as more locations are considered when quantifying joint flood hazard, natural variability becomes an even more prominent driver of spatially compounding major flood hazards. These results suggest that the dominating factor that dictates whether basins experience simultaneous large flooding is largely randomness in storm tracks and the associated spatial distribution of extreme precipitation and temperature-driven snowmelt. The thermodynamic climate changes that influence snowmelt or scale up storms do play a role, particularly if the basins are in close proximity (such as the Tuolumne and Millerton in Figures 13a,c). However, as more basins are included, natural variability in the weather during large storms dominates. Figure 13 reveals the inherent challenges of managing for spatially compounding flood hazards in this region. If persistent climate changes are a more dominant factor in driving joint flooding across all basins, then shared investments in canal expansion or rehabilitation across the regions could be used to offset some of this risk. However, since natural variability is the key driver of large flooding, alternative methods of creating unified planning and management strategies again need to be considered, using a mix of carefully coordinated shorter-term adaptive actions that provide flexibility to improve the robustness of longer-term infrastructure investments to the extreme hydro-climatic variability of the Central Valley (Herman et al., 2020; Hamilton et al., 2022).

4 Conclusion

This study contributes a novel framework to better understand the relative role of natural climate variability and climate change in determining the uncertainty in future hydrologic extremes of great importance to water systems planning and management. This framework is complementary to similar approaches based on GCM ensembles, but instead utilizes a large stochastic ensemble of paleo-based weather and hydrologic simulations to capture the plausible range of natural variability in drought and flood dynamics. The impacts of pre-selected mechanisms of climate change, including shifts in temperature and precipitation scaling, are then incorporated into the ensemble. The variance in hydrologic extremes is then partitioned across those climate changes and natural variability in the ensemble.

We first demonstrate the utility of the generator forced with paleodata in capturing and expanding on the dynamics of the modern record, which makes it a particularly useful for facilitating exploratory modeling and further quantification of the robustness of water resources systems to challenging scenarios that have been seen in the region's past hydroclimate. We also highlight the large non-stationarity that exists in the flood and drought metrics through the length of the reconstruction, particularly taking note of consequential 100-year flooding periods that can become as likely as 10-year events in parts of the record (i.e., 10 times more likely). These results have large implications for commonly employed stationary analyses, such as deriving design event estimates from the modern record, to quantify flood risk in this region. Our results suggest that these techniques severely underrepresent hydro-climatic hazards and the magnitude of design events that infrastructure should be built for.

The results of the variance decomposition component of the study highlight the following main conclusions:

- Uncertainty in future flooding within individual basins is largely driven by thermodynamic climate change, especially if evaluated over long time horizons. Flooding within snow-dominated basins is primarily driven by changes in temperature, while lower-elevation basins see a greater influence from precipitation scaling.
- The relative importance of climate change and natural variability on the uncertainty in future drought depends on the drought metric of interest. Changes in temperature drive drought occurrence, while precipitation scaling plays a role in drought intensity. Drought duration is primarily driven by natural variability.
- The uncertainty in simultaneous flood hazard across multiple basins is largely driven by natural variability, and this influence increases as additional basins are considered.
- The perception of the most important driver is highly influenced by the time horizon over which a metric is calculated. Shorter time horizons are less likely to capture how climate change uncertainty influences the uncertainty in hydrologic extremes.

The variance decomposition reveals a complicated path to robust planning and managing for both flood and drought in the region. The results suggest that natural variability and climate change influence both extremes to varying degrees. Furthermore, different characteristics of a single extreme (i.e. drought occurrence and duration) can be influenced by different drivers.

Additionally, if different time horizons are prioritized for planning for extremes, the understanding of the most important drivers of flood and drought hazards also changes. This last facet especially presents a problem for adaptive planning and management. This type of planning triggers management decisions based on the evolution of an observed variable (including hydroclimatic variables like precipitation or streamflow) over a specific horizon. As demonstrated in our study, tracking peak flows over a 30-year or 100-year horizon are both appropriate for longer-term flood management, but prioritizing the latter could neglect the effects of internal variability in the near term while increasing the potential for maladaptive longer-lived capital investments in infrastructure. Thus, it's important for water resources agencies that utilize these dynamic planning methods to effectively balance the value and potential regret of near term investments (Herman et al., 2020; Schlumberger et al., 2022).

One of the most important results of our study is that natural variability plays a very large role in dictating the future uncertainty in key metrics of flood and drought that form the basis of water resources planning; at times much larger than that of prominent climate change signals. This suggests that better quantification of the true range of natural variability in these extremes should be a major priority for the climate and hydrologic research community, and equally important, these efforts should directly inform future planning efforts for water resources systems. However, historically, this has often not been the case, with concerns about climate change often overshadowing the potential impacts of natural variability (see discussions in Koutsoyiannis (2020, 2021)).

Our results show, in particular, the importance of natural variability on spatially compounding flood hazard, which arguably poses a more difficult and complex management problem than addressing hazards in any one basin due to the need for infrastructure coordination across space and time. This highlights the potential value that longer, paleo-based data could bring to the estimation of joint flood hazards. The field of paleoflood hydrology has historically focused on the identification and dating of flood evidence in fluvial sedimentary archives, but incorporating speleothems and botanical archives can substantially increase the comprehensiveness and quality of paleoflood data (Wilhelm et al., 2018). Alluvial archives are also being used in more densely-populated and flood-prone regions (Toonen et al., 2020), and recent studies have shown that incorporation of these data can significantly reduce the uncertainty of extreme flood estimates (Engeland et al., 2020; Reinders & Muñoz, 2021). Methodological advances that can use these new and diverse data sources to constrain joint flood hazard estimates across sites would be particularly helpful, as would guidance on how to appropriately and consistently incorporate paleodata into risk management practices that also consider the effects of climate change. The work of England Jr et al. (2019) that helped incorporate paleodata into U.S. flood frequency guidance (Bulletin 17C) provides inspiration for such an approach.

The results also highlight the significant impact of natural variability on drought uncertainty, especially drought duration and intensity, and the implications stated above for joint flood hazards also extend to drought hazards. There are state-of-the-art techniques currently being applied within the dendrochronology community that can help improve our understanding of the natural range of drought variability. Beyond using tree ring widths, some studies are isolating earlywood and latewood signals for better drought reconstruction (Soulé et al., 2021; Song et al., 2022) or using blue intensity (the intensity of reflectance of the blue channel light from a wood core) to identify more stable climate-growth relationships that inform more robust reconstructions (Akhmetzyanov et al., 2023). Furthermore, better forecasts could provide water managers with more effective ways to navigate drought caused by natural variability. Skillful near-term drought predictions have been achieved by using decadal hindcasts from CMIP6 (Zhu et al., 2020) and Machine learning based approaches, particularly those that can model catchment memory are being used to create skillful seasonal drought predictions (Amanambu et al., 2022; Sutanto & Van Lanen, 2022)

One key limitation of this work is that we only consider a subset of plausible climate change scenarios that are not comprehensive, but rather reflect two mechanisms of change that are likely to occur and to be consequential to the San Joaquin Valley in California. This limitation includes the omission of the possibility that properties of long-term climate variability will itself change in the future under climate change. Another limitation is that we represent natural variability with one statistical model based on historical and paleo data. As others have shown (Koutsoyiannis, 2021), the quantification of natural variability often greatly depends on the statistical model used.

While outside the scope of this study, the framework presented and conclusions drawn here would benefit from a direct comparison against a similar approach using a climate ensemble drawn from a GCM, especially a single model initial-condition large ensemble (SMILE; see Lehner et al. (2020)). In a SMILEs-based framework, projections of precipitation and temperature derived from a single GCM under multiple initial conditions and multiple emission scenarios could be downscaled and propagated through hydrologic models to create a future streamflow ensemble, which could be used for partitioning variance in hydrologic extremes across emission scenarios and natural variability. By comparing results between the framework of this study and a SMILEs-based framework, one could better understand whether and how the relative roles of natural variability and climate change are consistent or depend on methodological choice.

Regardless of method used, the results of this work strongly suggest that large ensembles of natural variability are likely needed to adequately assess future risks to water resources systems that are particularly sensitive to extreme events. In future work, we intend to pair the hydrologic ensembles developed here with a regional, multi-sector model of California's Central Valley (Zeff et al., 2021) to more fully assess the risk that future hydroclimate extremes pose to stakeholders across the system, including groundwater banks and irrigation districts. The ultimate goal of such work is to facilitate a greater understanding of how future extremes lead to heterogeneous shortage and flooding impacts across stakeholders, and to help identify robust adaptation strategies to address these future risks.

Acknowledgments

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1650441. The views expressed in this work represent those of the authors and do not necessarily reflect the views or policies of the National Science Foundation.

Data Availability Statement

Sample input data and code to run the weather generator and hydrologic models, create flood and drought metrics metrics, and create figures can be found at `https://doi.org/10.5281/zenodo.7693324`. Refer to the associated GitHub repository: `https://github.com/rg727/Gupta_WGEN_Partitioning_NatVar_CC_Drivers`

References

Akhmetzyanov, L., Sánchez-Salguero, R., García-González, I., Domínguez-Delmás, M., & Sass-Klaassen, U. (2023). Blue is the fashion in mediterranean pines: New drought signals from tree-ring density in southern europe. *Science of the Total Environment*, *856*, 159291.

Ali, H., Fowler, H. J., Pritchard, D., Lenderink, G., Blenkinsop, S., & Lewis, E. (2022). Towards quantifying the uncertainty in estimating observed scaling rates. *Geophysical Research Letters*, *49*(12), e2022GL099138.

Amanambu, A. C., Mossa, J., & Chen, Y.-H. (2022). Hydrological drought forecasting using a deep transformer model. *Water*, *14*(22), 3611.

Anderson, E. A. (1976). *A point energy and mass balance model of a snow cover.* Stanford University.

Bass, B., Norris, J., Thackeray, C., & Hall, A. (2022). Natural variability has concealed increases in western us flood hazard since the 1970s. *Geophysical Research Letters*, *49*(7), e2021GL097706.

Borkotoky, S. S., Williams, A. P., Cook, E. R., & 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.

Brekke, L. D., Maurer, E. P., Anderson, J. D., Dettinger, M. D., Townsley, E. S., Harrison, A., & Pruitt, T. (2009). Assessing reservoir operations risk under climate change. *Water Resources Research*, *45*(4).

Burnash, R., et al. (1995). The nws river forecast system-catchment modeling. *Computer models of watershed hydrology.*, 311–366.

Cai, W., Zhu, X., Zhao, X., & Zhang, Y. (2021). Estimating the effect of climate internal variability and source of uncertainty in climate-hydrological projections in a representative watershed of northeastern china.

California Department of Water Resources. (2006). Progress on incorporating climate change into planning and management of california's water resources, california department of water resources. *California Department of Water Resources*.

California Department of Water Resources. (2015). California's most significant droughts: Comparing historical and recent conditions. *California Department* *of Water Resources*.

California Department of Water Resources. (2022). Central valley flood protection plan update 2022. *California Department of Water Resources*.

Chegwidden, O. S., Hagen, R., Martin, K., Jones, M., Banihirwe, A., Chiao, C., ... Hamman, J. (2022, October). *Downscaling CMIP6 with multiple statistical methods.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.7145491` (The development of this project was funded, in part, through a grant from the Microsoft AI for Earth program to CarbonPlan.) doi: 10.5281/zenodo.7145491

Chung, F. (2009). *Using future climate projections to support water resources decision making in california.* California Energy Commission.

Cook, B. I., Cook, E. R., Smerdon, J. E., Seager, R., Williams, A. P., Coats, S., ... Díaz, J. V. (2016). North american megadroughts in the common era: Reconstructions and simulations. *Wiley Interdisciplinary Reviews: Climate Change*, *7*(3), 411–432.

D'Arrigo, R. D., & Jacoby, G. C. (1991). A 1000-year record of winter precipitation from northwestern new mexico, usa: a reconstruction from tree-rings and its relation to el niño and the southern oscillation. *The Holocene*, *1*(2), 95–101.

Deser, C., Lehner, F., Rodgers, K. B., Ault, T., Delworth, T. L., DiNezio, P. N., ... others (2020). Insights from earth system model initial-condition large ensembles and future prospects. *Nature Climate Change*, *10*(4), 277–286.

Dessai, S., & Hulme, M. (2004). Does climate adaptation policy need probabilities? *Climate policy*, *4*(2), 107–128.

Dettinger, M., & Cayan, D. R. (2014). Drought and the california delta—a matter of extremes. *San Francisco Estuary and Watershed Science*, *12*(2).

Dettinger, M. D., & Ingram, B. L. (2013). The coming megafloods. *Scientific American*, *308*(1), 64–71.

Engeland, K., Aano, A., Steffensen, I., Støren, E., & Paasche, Ø. (2020). New flood frequency estimates for the largest river in norway based on the combination of short and long time series. *Hydrology and Earth System Sciences*, *24*(11), 5595–5619.

England Jr, J., Cohn, T., Faber, B., Stedinger, J., Thomas Jr, W., Veilleux, A., & Mason Jr, R. (2019). Guidelines for determining flood flow frequency—bulletin

17c (no. 4-b5). *US Geological Survey*.

Farahmand, A., & AghaKouchak, A. (2015). A generalized framework for deriving nonparametric standardized drought indicators. *Advances in Water Resources*, *76*, 140–145.

Gershunov, A., Shulgina, T., Ralph, F. M., Lavers, D. A., & Rutz, J. J. (2017). Assessing the climate-scale variability of atmospheric rivers affecting western north america. *Geophysical Research Letters*, *44*(15), 7900–7908.

Gupta, R. S., Steinschneider, S., & Reed, P. M. (2022). A multi-objective paleo-informed reconstruction of western us weather regimes over the past 600 years. *Climate Dynamics*, 1–20.

Haasnoot, M., Kwakkel, J. H., Walker, W. E., & Ter Maat, J. (2013). Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world. *Global environmental change*, *23*(2), 485–498.

Hamilton, A. L., Zeff, H. B., Characklis, G. W., & Reed, P. M. (2022). Resilient california water portfolios require infrastructure investment partnerships that are viable for all partners. *Earth's Future*, *10*(4), e2021EF002573.

Hamlet, A. F., & Lettenmaier, D. P. (2007). Effects of 20th century warming and climate variability on flood risk in the western us. *Water Resources Research*, *43*(6).

Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, *90*(8), 1095–1108.

Henn, B., Musselman, K. N., Lestak, L., Ralph, F. M., & 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.

Herman, J. D., Quinn, J. D., Steinschneider, S., Giuliani, M., & Fletcher, S. (2020). Climate adaptation as a control problem: Review and perspectives on dynamic water resources planning under uncertainty. *Water Resources Research*, *56*(2), e24389.

Herweijer, C., Seager, R., & Cook, E. R. (2006). North american droughts of the mid to late nineteenth century: a history, simulation and implication for mediaeval drought. *The Holocene*, *16*(2), 159–171.

Hoerling, M., Quan, X.-W., & Eischeid, J. (2009). Distinct causes for two principal us droughts of the 20th century. *Geophysical Research Letters*, *36*(19).

Huang, G., & Kadir, T. (2016). Estimates of natural and unimpaired flows for the central valley of california: Water years 1922–2014. *California Department of Water Resources*.

Huang, X., & Swain, D. L. (2022). Climate change is increasing the risk of a california megaflood. *Science advances*, *8*(31), eabq0995.

Hunt, B. (2011). Global characteristics of pluvial and dry multi-year episodes, with emphasis on megadroughts. *International Journal of Climatology*, *31*(10), 1425–1439.

Jung, I.-W., Chang, H., & Moradkhani, H. (2011). Quantifying uncertainty in urban flooding analysis considering hydro-climatic projection and urban development effects. *Hydrology and Earth System Sciences*, *15*(2), 617–633.

Kay, A., Davies, H., Bell, V., & Jones, R. (2009). Comparison of uncertainty sources for climate change impacts: flood frequency in england. *Climatic change*, *92*(1-2), 41–63.

Kirchmeier-Young, M. C., & Zhang, X. (2020). Human influence has intensified extreme precipitation in north america. *Proceedings of the National Academy of Sciences*, *117*(24), 13308–13313.

Knutti, R., Sedláček, J., Sanderson, B. M., Lorenz, R., Fischer, E. M., & Eyring, V. (2017). A climate model projection weighting scheme accounting for performance and interdependence. *Geophysical Research Letters*, *44*(4), 1909–1918.

Koutsoyiannis, D. (2020). Revisiting the global hydrological cycle: is it intensifying? *Hydrology and Earth System Sciences*, *24*(8), 3899–3932.

Koutsoyiannis, D. (2021). Stochastics of hydroclimatic extremes—a cool look at risk. *National Technical University of Athens: Athens, Greece*.

Kunkel, K. E. (2003). North american trends in extreme precipitation. *Natural hazards*, *29*(2), 291–305.

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., ... Hawkins, E. (2020). Partitioning climate projection uncertainty with multiple large ensembles and cmip5/6. *Earth System Dynamics*, *11*(2), 491–508.

Livneh, B., Bohn, T. J., Pierce, D. W., Munoz-Arriola, F., Nijssen, B., Vose, R., ... Brekke, L. (2015). A spatially comprehensive, hydrometeorological data set for mexico, the us, and southern canada 1950–2013. *Scientific data*, *2*(1), 1–12.

Lohmann, D., Raschke, E., Nijssen, B., & Lettenmaier, D. (1998). Regional scale hydrology: I. formulation of the vic-2l model coupled to a routing model. *Hydrological sciences journal*, *43*(1), 131–141.

Martinkova, M., & Kysely, J. (2020). Overview of observed clausius-clapeyron scaling of extreme precipitation in midlatitudes. *Atmosphere*, *11*(8), 786.

Maurer, E. P., Brekke, L. D., & Pruitt, T. (2010). Contrasting lumped and distributed hydrology models for estimating climate change impacts on california watersheds 1. *JAWRA Journal of the American Water Resources Association*, *46*(5), 1024–1035.

Maurer, E. P., Hidalgo, H. G., Das, T., Dettinger, M. D., & Cayan, D. R. (2010). The utility of daily large-scale climate data in the assessment of climate change impacts on daily streamflow in california. *Hydrology and Earth System Sciences*, *14*(6), 1125–1138.

McCabe, G. J., Palecki, M. A., & Betancourt, J. L. (2004). Pacific and atlantic ocean influences on multidecadal drought frequency in the united states. *Proceedings of the National Academy of Sciences*, *101*(12), 4136–4141.

McKee, T. B., Doesken, N. J., Kleist, J., et al. (1993). The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th conference on applied climatology* (Vol. 17, pp. 179–183).

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*(3), e2021EF002537.

Moon, H., Gudmundsson, L., & Seneviratne, S. I. (2018). Drought persistence errors in global climate models. *Journal of Geophysical Research: Atmospheres*, *123*(7), 3483–3496.

Moore, R. (2007). The pdm rainfall-runoff model. *Hydrology and Earth System Sciences*, *11*(1), 483–499.

Najibi, N., Mukhopadhyay, S., & Steinschneider, S. (2021). Identifying weather regimes for regional-scale stochastic weather generators. *International Journal of Climatology*, *41*(4), 2456–2479.

Nazemi, A., & Wheater, H. S. (2014). Assessing the vulnerability of water supply to changing streamflow conditions. *Eos, Transactions American Geophysical Union*, *95*(32), 288–288.

Null, J., & Hulbert, J. (2007). California washed away: The great flood of 1862. *Weatherwise*, *60*(1), 26–30.

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.

Reinders, J. B., & Muñoz, S. E. (2021). Improvements to flood frequency analysis on alluvial rivers using paleoflood data. *Water Resources Research*, *57*(4), e2020WR028631.

Rocheta, E., Sugiyanto, M., Johnson, F., Evans, J., & Sharma, A. (2014). How well do general circulation models represent low-frequency rainfall variability? *Water Resources Research*, *50*(3), 2108–2123.

Rowell, D. P. (2012). Sources of uncertainty in future changes in local precipitation. *Climate dynamics*, *39*, 1929–1950.

Schiemann, R., Athanasiadis, P., Barriopedro, D., Doblas-Reyes, F., Lohmann, K., Roberts, M. J., … Vidale, P. L. (2020). Northern hemisphere blocking simulation in current climate models: evaluating progress from the climate model intercomparison project phase 5 to 6 and sensitivity to resolution. *Weather and Climate Dynamics*, *1*(1), 277–292.

Schimmelmann, A., Zhao, M., Harvey, C. C., & Lange, C. B. (1998). A large california flood and correlative global climatic events 400 years ago. *Quaternary Research*, *49*(1), 51–61.

Schlef, K. E., Steinschneider, S., & Brown, C. M. (2018). Spatiotemporal impacts of climate and demand on water supply in the apalachicola-chattahoochee-flint basin. *Journal of Water Resources Planning and Management*, *144*(2), 05017020.

Schlumberger, J., Haasnoot, M., de Ruiter, M., & Aerts, J. (2022). Towards a disaster risk management pathways framework for complex and dynamic multi-risk: Dapp-mr. *Available at SSRN 4164233*.

Seager, R., Hoerling, M., Schubert, S., Wang, H., Lyon, B., Kumar, A., … Henderson, N. (2015). Causes of the 2011–14 california drought. *Journal of Climate*, *28*(18), 6997–7024.

Siler, N., Proistosescu, C., & Po-Chedley, S. (2019). Natural variability has slowed the decline in western us snowpack since the 1980s. *Geophysical Research Letters*, *46*(1), 346–355.

Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., ... others (2019). Towards a more reliable historical reanalysis: Improvements for version 3 of the twentieth century reanalysis system. *Quarterly Journal of the Royal Meteorological Society*, *145*(724), 2876–2908.

Song, W., Zhao, B., Mu, C., Ballikaya, P., Cherubini, P., & Wang, X. (2022). Moisture availability influences the formation and characteristics of earlywood of pinus tabuliformis more than latewood in northern china. *Agricultural and Forest Meteorology*, *327*, 109219.

Soulé, P. T., Knapp, P. A., Maxwell, J. T., & Mitchell, T. J. (2021). A comparison of the climate response of longleaf pine (pinus palustris mill.) trees among standardized measures of earlywood, latewood, adjusted latewood, and totalwood radial growth. *Trees*, *35*, 1065–1074.

Stahle, D. W., Cook, E. R., Cleaveland, M. K., Therrell, M. D., Meko, D. M., Grissino-Mayer, H. D., ... Luckman, B. H. (2000). Tree-ring data document 16th century megadrought over north america. *Eos, Transactions American Geophysical Union*, *81*(12), 121–125.

Stahle, D. W., Fye, F. K., Cook, E. R., & Griffin, R. D. (2007). Tree-ring reconstructed megadroughts over north america since ad 1300. *Climatic change*, *83*(1-2), 133.

Steinschneider, S., Ray, P., Rahat, S. H., & 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*(8), 6923–6945.

Sun, X., & Wang, G. (2022). Causes for the negative scaling of extreme precipitation at high temperatures. *Journal of Climate*, *35*(18), 6119–6134.

Sutanto, S. J., & Van Lanen, H. A. (2022). Catchment memory explains hydrological drought forecast performance. *Scientific Reports*, *12*(1), 2689.

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., ... others (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in miroc6. *Geoscientific Model Development*, *12*(7), 2727–2765.

Taylor, M. (2017). Managing floods in california. *Legislative Analyst's Office: The California Legislature's Nonpartisan Fiscal and Policy Advisor*.

Toonen, W. H., Munoz, S. E., Cohen, K. M., & Macklin, M. G. (2020). High-resolution sedimentary paleoflood records in alluvial river environments: a review of recent methodological advances and application to flood hazard assessment. *Palaeohydrology: Traces, Tracks and Trails of Extreme Events*, 213–228.

Vicente-Serrano, S. M., López-Moreno, J. I., Beguería, S., Lorenzo-Lacruz, J., Azorin-Molina, C., & Morán-Tejeda, E. (2012). Accurate computation of a streamflow drought index. *Journal of Hydrologic Engineering*, *17*(2), 318–332.

Vidal, J.-P., Hingray, B., Magand, C., Sauquet, E., & Ducharne, A. (2015). Hierarchy of climate and hydrological uncertainties in transient low flow projections. *Hydrology & Earth System Sciences Discussions*, *12*(12).

Wasko, C., Lu, W. T., & Mehrotra, R. (2018). Relationship of extreme precipitation, dry-bulb temperature, and dew point temperature across australia. *Environmental Research Letters*, *13*(7), 074031.

Whateley, S., & Brown, C. (2016). Assessing the relative effects of emissions, climate means, and variability on large water supply systems. *Geophysical Research Letters*, *43*(21), 11–329.

Wi, S., & Steinschneider, S. (2022). Assessing the physical realism of deep learning hydrologic model projections under climate change. *Water Resources Research*, *58*(9), e2022WR032123.

Wi, S., Yang, Y., Steinschneider, S., Khalil, A., & Brown, C. (2015). Calibration approaches for distributed hydrologic models in poorly gaged basins: implication for streamflow projections under climate change. *Hydrology and Earth System Sciences*, *19*(2), 857–876.

Wilby, R., & Dessai, S. (2010). Robust adaptation to climate change.

Wilhelm, B., Canovas, J. A. B., Aznar, J. P. C., Kämpf, L., Swierczynski, T., Stoffel, M., … Toonen, W. (2018). Recent advances in paleoflood hydrology: From new archives to data compilation and analysis. *Water Security*, *3*, 1–8.

Williams, A., Anchukaitis, K., Woodhouse, C., Meko, D., Cook, B., Bolles, K., &

Cook, E. (2021). Tree rings and observations suggest no stable cycles in sierra nevada cool-season precipitation. *Water Resources Research*, *57*(3), e2020WR028599.

Williams, A. P., Cook, B. I., & Smerdon, J. E. (2022). Rapid intensification of the emerging southwestern north american megadrought in 2020–2021. *Nature Climate Change*, *12*(3), 232–234.

Williams, A. P., Cook, E. R., Smerdon, J. E., Cook, B. I., Abatzoglou, J. T., Bolles, K., ... Livneh, B. (2020). Large contribution from anthropogenic warming to an emerging north american megadrought. *Science*, *368*(6488), 314–318.

Yip, S., Ferro, C. A., Stephenson, D. B., & Hawkins, E. (2011). A simple, coherent framework for partitioning uncertainty in climate predictions. *Journal of climate*, *24*(17), 4634–4643.

Zeff, H. B., Hamilton, A. L., Malek, K., Herman, J. D., Cohen, J. S., Medellin-Azuara, J., ... Characklis, G. W. (2021). California's food-energy-water system: An open source simulation model of adaptive surface and groundwater management in the central valley. *Environmental Modelling & Software*, *141*, 105052.

Zhang, L., & Singh, V. (2019). *Copulas and their properties.* Copulas and Their Applications in Water Resources Engineering. Cambridge ....

Zhou, Y., & Kim, H.-M. (2018). Prediction of atmospheric rivers over the north pacific and its connection to enso in the north american multi-model ensemble (nmme). *Climate Dynamics*, *51*(5-6), 1623–1637.

Zhu, E., Yuan, X., & Wu, P. (2020). Skillful decadal prediction of droughts over large-scale river basins across the globe. *Geophysical Research Letters*, *47*(17), e2020GL089738.

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M., ... others (2020). A typology of compound weather and climate events. *Nature reviews earth & environment*, *1*(7), 333–347.

Appendix A: Glossary of Terms

- **Baseline weather scenario**: The 600-year daily precipitation and temperature scenario that is created by forcing the weather generator with paleo-reconstructed weather regimes. This scenario is comprised of 50 stochastic ensemble members.
- **Baseline streamflow scenario**: The 600-year daily streamflow scenario acquired by driving the hydrologic model with paleo-reconstructed weather (often referred to as 0T, 0CC). This scenario is comprised of 50 stochastic ensemble members.
- **Climate scenario**: A 600-year daily streamflow scenario created by forcing the hydrologic model with a baseline weather scenario that is layered with a set of thermodynamic climate changes.
- **Ensemble member**: Also referred to as a stochastic realization; each climate scenario is comprised of 50 stochastic ensemble members
- **Record length**: The total length of the dataset
  - **Paleo-informed weather and streamflow datasets**: 617 years (1400-2017 CE) at a daily time scale
  - **Observed Livneh climate (temperature and precipitation) dataset**: 63 years (1950-2013 CE) at a daily time scale
  - **Observed CDEC streamflow dataset**: 33 years (1986-2019) at a daily time scale
- **Time horizon**: also referred to as moving window; the length (in years) of the sliding window that passes over the total record length.

Appendix B: Metrics and Time Horizons

Metric	**Description**	**Calculated**	**Justification**	**Citation**
Flood Metric	10-Year Return Period Flow	GEV fit to 3-day maximum flow	Captures risk to smaller floodplains (or nuisance flooding in larger areas) and drives smaller investments.	Progress on Incorporating Climate Change into Planning and Management of California's Water Resources (July 2006)
Flood Metric	100-Year Return Period Flow	GEV fit to 3-day maximum flow	Drives larger riverine infrastructure development and flood risk management. Requires FEMA-mandated insurance.	Central Valley Flood Protection Plan Update 2022 (November 2022)
Drought Metrics	Occurrence, Severity, and Duration	Standardized streamflow-based indices	No state-wide definition. Historical droughts have been identified based on a combination of metrics such as reservoir depth and deficit magnitude and duration.	California's Most Significant Droughts: Comparing historical and recent conditions (February 2015)

Spatially Compounding Flood Metric	Likelihood of simultaneously exceeding historical 10-year and 100-year flow events in n basins	n-dimensional Gaussian copula	Flooding across the San Joaquin system could result in infrastructure failure such as levee breaks and disrupt deliveries of fresh water to 3 million acres of farmland.	Managing Floods in California (March 2017); Zscheischler et al. (2020)
Time Horizon	30-Year	N/A	CA prioritizes investment in flood management over a 30-year planning horizon	Central Valley Flood Protection Plan Update 2022 (November 2022)
Time Horizon	100-Year	N/A	Not actively used in planning and management, but can represent longer-term investments.	N/A

Understanding Contributions of Paleo-Informed Natural Variability and Climate Changes on Hydroclimate Extremes in the Central Valley Region of California

Rohini S. Gupta[1], Scott Steinschneider[2], Patrick M. Reed[1]

[1]School of Civil and Environmental Engineering, Cornell University, Ithaca, NY, USA
[2]Department of Biological and Environmental Engineering, Cornell University, Ithaca, NY, USA

Key Points:

- We introduce a framework to create 600-year ensembles of future weather and streamflow for basins in the San Joaquin Valley.
- We discover vast variability and non-stationarity in flood and drought extremes in the region over the past 600 years.
- Variability in extremes is primarily attributed to natural variability, but climate changes are influential under longer planning horizons.

Corresponding author: R.S. Gupta:, `rg727@cornell.edu`

Abstract

To aid California's water sector to better manage future climate extremes, we present a method for creating a regional ensemble of plausible daily future climate and streamflow scenarios that represent natural climate variability captured in a network of tree-ring chronologies, and then embed anthropogenic climate change trends within those scenarios. We use 600 years of paleo-reconstructed weather regimes to force a stochastic weather generator, which we develop for five subbasins in the San Joaquin River in the Central Valley region of California. To assess the compound effects of climate change, we create temperature series that reflect scenarios of warming and precipitation series that are scaled to reflect thermodynamically driven shifts in the daily precipitation distribution. We then use these weather scenarios to force hydrologic models for each of the San Joaquin subbasins. The paleo-forced streamflow scenarios highlight periods in the region's past that produce flood and drought extremes that surpass those in the modern record and exhibit large non-stationarity through the reconstruction. Variance decomposition is employed to characterize the contribution of natural variability and climate change to variability in decision-relevant metrics related to floods and drought. Our results show that a large portion of variability in individual subbasin and spatially compounding extreme events can be attributed to natural variability, but that anthropogenic climate changes become more influential at longer planning horizons. The joint importance of climate change and natural variability in shaping extreme floods and droughts is critical to resilient water systems planning and management in the Central Valley region.

Plain Language Summary

California experiences cycles of floods and droughts that can be driven by both natural variability and climate change. The specific role of these drivers play in influencing extremes is uncertain, but can strongly dictate how to best plan and manage regional water systems for future extremes. To better quantify the role of these drivers, we introduce a framework that utilizes a 600-year tree-ring reconstruction to create long sequences of plausible ensembles of future weather and streamflow for key basins in the San Joaquin Valley. We find that a large portion of variability in extremes can be attributed to natural variability, but that anthropogenic climate changes become more influential at longer planning horizons. Furthermore, our perception of important drivers can be skewed depending on the specific definitions used to analyze floods and droughts,

which can present significant challenges for adaptation planning and infrastructure development tied to hydroclimate indicators. This study also illustrates the vast variability in extremes that the region has experienced over the past 600 years and highlights the pitfalls of using stationary risk measures.

1 Introduction

The recent drought conditions impacting California are occurring within the broader context of two decades of extreme climate variability. Since 2000, California has experienced four periods of drought: (2000-2003, 2007-2009, 2012-2016, and the ongoing drought beginning in the 2020). The former three complete drought periods were all ended by extreme atmospheric river (AR)-driven events. While offering much needed precipitation, these storms often cause widespread flooding and landslides. In 2017, extreme precipitation associated with ARs generated California's wettest winter in a century and caused catastrophic damage to Oroville Dam, which prompted the evacuation of 188,000 people and required nearly $1 billion in repairs (Henn et al., 2020). Since this event, California has ebbed and flowed through wet and dry periods, including experiencing the driest 22-year period in at least 1,200 years (A. P. Williams et al., 2022).

The recent two decades of California climate extremes are in part a manifestation of the extreme natural variability that characterizes the regional climate. Tree ring reconstructions have shown that the region experienced multiple persistent megadroughts over the past two millennia (late 800s, mid-1100s, late 1200s, mid-1400s, and late 1500s), long before anthropogenic influence (Stahle et al., 2000, 2007; A. Williams et al., 2021). Multi-millennial control runs of coupled global climate models (GCMs) have also reproduced megadroughts in the Southwestern U.S. even without any external sea surface temperature (SST) forcing, suggesting that these droughts can develop due to internal climate variability alone (Hunt, 2011). Some (but not all) of this natural drought variability is linked to major modes of atmospheric and oceanic variability, such as the El Niño Southern Oscillation (ENSO) and the Pacific Decadal Oscillation (PDO) (McCabe et al., 2004; Hoerling et al., 2009; Seager et al., 2015; Cook et al., 2016). Interspersed across the past two centuries, California has also experienced several extreme precipitation events (e.g., 1908-1909, 1913-1914, 1940-1941, 1955-1956, 1969, 1986, and 1997); most prominently the Great Flood of 1861-62 that turned the San Joaquin and Sacramento Valleys into an inland sea (M. D. Dettinger & Ingram, 2013). This event notably occurred after a 20-year drought (Null & Hulbert, 2007). Sediment reconstructions in the Klamath Basin suggest that the 1861-1862 megaflood was not an extreme outlier, but rather a 100-200-year event that has been matched in magnitude several times over the last two millennia (e.g., 212, 440, 603, 1029, 1300, 1418, 1605, 1750, and 1810 CE; M. D. Dettinger and Ingram (2013)).

The historic droughts and floods above, independent of anthropogenic-related warming, confirm the strong influence of natural climate variability in California and more broadly across the Western U.S. However, recent studies show that climate change is amplifying the severity of these extremes. Warming due to anthropogenic radiative forcing has intensified recent droughts in the region, primarily through enhanced atmospheric moisture demand and soil moisture depletion (A. P. Williams et al., 2020). As noted above, the recent cumulative drought conditions in California and the rest of the Western U.S. over the past two decades now ranks as the driest 22-year period in at least 1,200 years (A. P. Williams et al., 2022). Similarly, climate change is increasing the risk of extreme precipitation events via an increase in the strength of cool-season AR events associated with a rise in atmospheric moisture content (Kunkel, 2003; Kirchmeier-Young & Zhang, 2020). A recent study by X. Huang and Swain (2022) found that climate change has already doubled the likelihood of AR-driven megastorms similar to that which caused the Great Flood of 1861-62, and that megastorm sequences of increased frequency and larger magnitude are likely with continued warming.

Thus, the present and evolving risks posed by hydrologic extremes in California is defined by the combined influence of natural climate variability and anthropogenic climate change. Yet the degree to which these two factors will control the future frequency and magnitude of extremes remains uncertain (Hamlet & Lettenmaier, 2007; Siler et al., 2019; Bass et al., 2022). From the perspective of water resource decision-makers who are charged with planning and managing large-scale infrastructure to mitigate the impacts of extreme events, this ambiguity presents a significant challenge. If climate change is the dominant factor that will determine the future magnitude, frequency, and duration of extreme events, then resources and attention should be concentrated on identifying and narrowing the uncertainty of the most prominent climate change signals and propagating them into updated design event estimates used for planning. However, if natural variability plays an equal or larger role in determining the properties of hydrologic extremes relevant to engineering design, then research into the plausible range of extremes due to natural variability should be equally prioritized (e.g., see Koutsoyiannis (2021)). A greater role of natural variability also suggests that dynamic and reversible adaptations may be favorable over irreversible investments. It is thus critically important to quantify the relative and joint roles of climate change versus natural variability in shaping the characteristics of hydrologic extremes, to help balance the allocation of attention and resources in a way that best serves the water sector to prepare for future extreme events.

A growing body of work has sought to partition the relative effects of climate change and natural variability, with a focus on climate variables and in the context of multi-model ensemble simulations (Hawkins & Sutton, 2009; Yip et al., 2011; Knutti et al., 2017; Rowell, 2012; Lehner et al., 2020). These studies primarily attribute variability in projected global and regional temperature and precipitation to climate change scenario uncertainty, global climate change model (GCM) uncertainty, and natural variability. Lehner et al. (2020) shows that scenario and model uncertainty are the dominant drivers of global decadal mean annual temperature and precipitation, but that natural variability dominates projections of regional temperatures (in Southern Europe) and precipitation (in the U.S. Pacific Northwest and Sahel region), particularly at shorter (and more decision-relevant) time scales. Fewer studies have explicitly considered the role of natural climate variability when partitioning variance in projections of hydrologic and water systems variables (Kay et al., 2009; Jung et al., 2011; Vidal et al., 2015; Whateley & Brown, 2016; Schlef et al., 2018; Cai et al., 2021). Kay et al. (2009) found that flood frequency and wintertime runoff in Europe are mostly influenced by choice of GCM, although they quantified natural climate variability using a limited number of GCM integrations with different initial conditions. Vidal et al. (2015) found that natural variability highly influences low flows in snow-dominated catchments in the French Alps, and Cai et al. (2021) found that natural variability is a dominant driver of rainy season runoff in Northeastern China. Jung et al. (2011) quantified natural variability using a block bootstrap on the historical record and found it to have the largest impact on the variance of large floods, as compared to GCM structure, emission scenario, land use change scenario, and hydrologic model parameter uncertainty. Similarly, Whateley and Brown (2016) and Schlef et al. (2018) created ensembles of future streamflow projections with a stochastic weather generator and rainfall-runoff model and found that the variance of reservoir storage as well as water system performance measures is mostly driven by natural climate variability, particularly in the first few decades of the projections.

The relative roles of natural variability and climate change on the variance of hydrologic variables of interest often depends on how natural variability is quantified and propagated into an ensemble of projections. In a majority of the climate studies (Hawkins & Sutton, 2009; Yip et al., 2011; Knutti et al., 2017; Rowell, 2012; Lehner et al., 2020) and three hydrologic studies (Kay et al., 2009; Jung et al., 2011; Vidal et al., 2015) referenced above, natural variability was determined using multi-member ensembles of GCMs run with different initial conditions. However, the degree to which initial-condition ensembles can represent true natural climate variability is unclear (Deser et al., 2020). For instance, these models poorly represent regional precipitation and drought persistence (Rocheta et al., 2014; Moon et al., 2018) and underestimate AR moisture flux and frequency (Zhou & Kim, 2018) all of which are important to regional planning and management of water systems. While the recent generation of models in CMIP6 better represents key features of natural climate variability (e.g., blocking; major climate modes) compared to older generations, significant biases remain (Tatebe et al., 2019; Schiemann et al., 2020)

An alternative way to estimate the relative influence of natural variability and climate change on regional hydrologic response is through bottom-up approaches that employ stochastically generated ensembles (Dessai & Hulme, 2004; Wilby & Dessai, 2010; Nazemi & Wheater, 2014). These methods require synthetic generators trained on observed weather or hydrologic records, which can generate large ensembles of scenarios that extrapolate beyond the observation limits of the historical record, maintain physical plausibility, and embed climate changes into the ensemble. The generation and partitioning of variability in the resulting hydroclimate metrics can provide a more robust way to quantify how much variance in regional hydrologic extremes is driven by climate changes versus natural variability. However, the availability of stochastic models to support these analyses is limited, particularly when interested in the variance decomposition of multiple properties of different hydrologic extremes (i.e., magnitude, duration, frequency, and spatial coherence of floods and droughts). Furthermore, the ways in which flood and drought events are defined, and particularly the time horizon (moving window) over which they are defined, can influence how the relative influences of climate variability and change are perceived. As time horizon shortens, it becomes increasingly difficult to identify clear climate change signals amongst the noise of natural climate variability. For example, Lehner et al. (2020) attributed the vast majority of variance in winter precipitation projections over the US Pacific Northwest to natural climate variability, but this was based on a 10-year moving window (i.e., decadal average). It is possible that any climate change impacts on mean winter precipitation, even if present, are not discernable from the noise within such a short moving window. This issue is especially true for the properties of extreme events, because there are so few samples available from which to estimate signal from noise (even with many ensemble members). To date, it remains unclear how the choice of time horizon influences our understanding of the relative roles of natural climate variability and climate change on the uncertainty in hydrologic extremes.

Based on the above knowledge gaps, this study addresses the following questions:

1. What is the relative importance of natural variability and climate change on variability in decision-relevant drought and flood metrics for the Central Valley of California?

2. How does the selected scale of the time horizon used for analyses influence the perceived importance of these drivers?

To answer these questions, we contribute a framework for creating a regionally consistent ensemble of plausible daily future climate and streamflow scenarios that represent natural climate variability captured in a network of tree-ring chronologies, and then embed anthropogenic climate change trends within those scenarios. A key contribution of this study is the use of 600 years of paleo-informed weather regimes (WRs; Gupta et al. (2022)) to force a weather-regime based stochastic generator (Steinschneider et al., 2019; Najibi et al., 2021), which we develop for five watersheds in the San Joaquin River basin. To assess the compound effect of climate change, we create temperature series that reflect projected scenarios of warming and precipitation series that have been scaled to reflect thermodynamically driven shifts in the distribution of daily precipitation. We then use these weather scenarios to force hydrologic models for each basin, generating ensembles of streamflow across the region. Decision relevant hydrologic metrics for characterizing flood and drought conditions are defined and calculated across San Joaquin subbasins and across the paleo-period using time horizons of varying scale (see Appendix B). Variance decomposition is then employed to characterize the relative contributions of natural variability and climate changes as drivers of flood and drought hazards in individual subbasins and for spatially compounding extremes that emerge across groups of subbasins.

2 Data and Methods

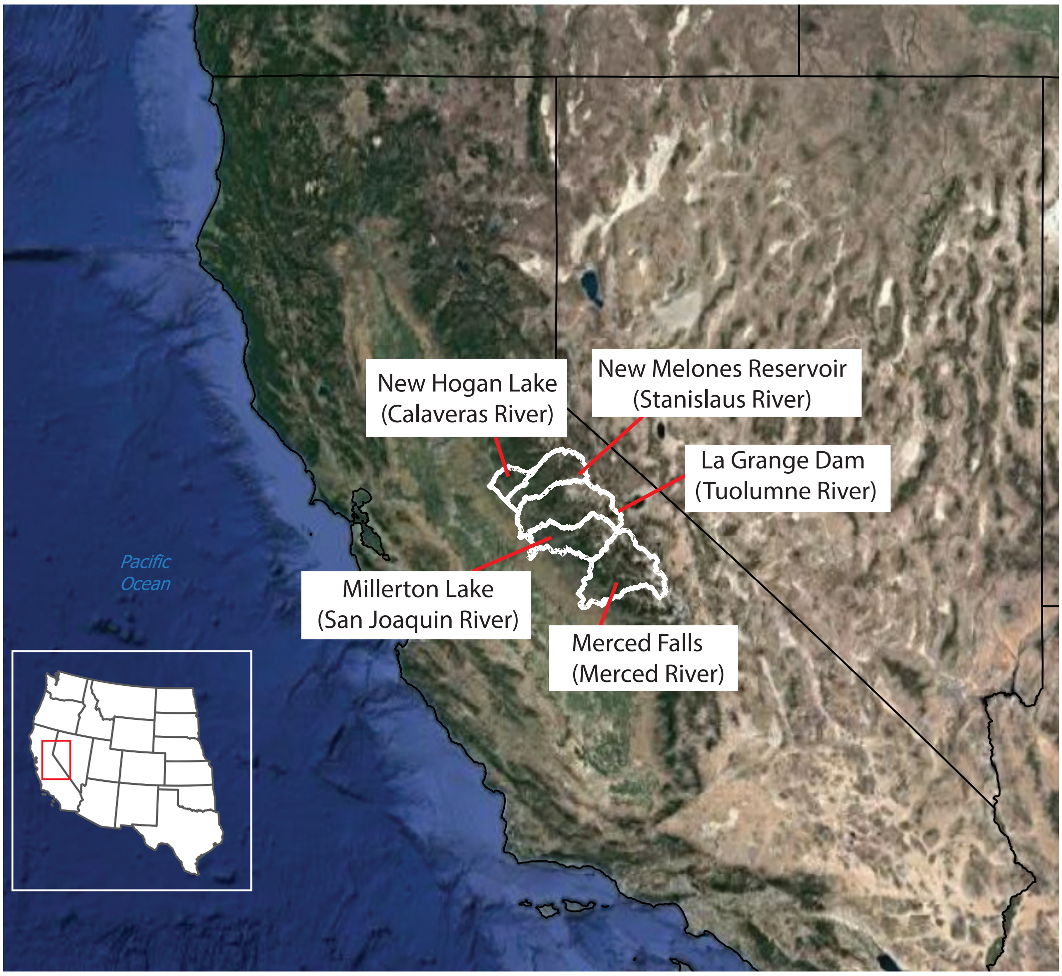

Figure 1. The study area is comprised of five subbasins within the greater San Joaquin River basin.

This study focuses on five subbasins within the San Joaquin River basin (Figure 1): the Tuolumne River, the Merced River, the San Joaquin River, the Stanislaus River, and the Calaveras River. The ultimate goal of this study is to partition the effects of natural climate variability and climate change on different properties of floods and droughts across these watersheds. We contribute a five-step methodology in order to achieve this goal (Figure 2). We first create a novel method to incorporate reconstructed weather regime dynamics (Gupta et al. 2022) into the generation of daily weather through the paleo-period (Section 2.1). Then, we create 600 years of surface weather ensembles across the five subbasins of the San Joaquin conditioned upon these reconstructed dynamics. We also create additional ensembles of surface weather layered with thermodynamic climate changes, such as temperature trends and precipitation scaling (Section 2.2).

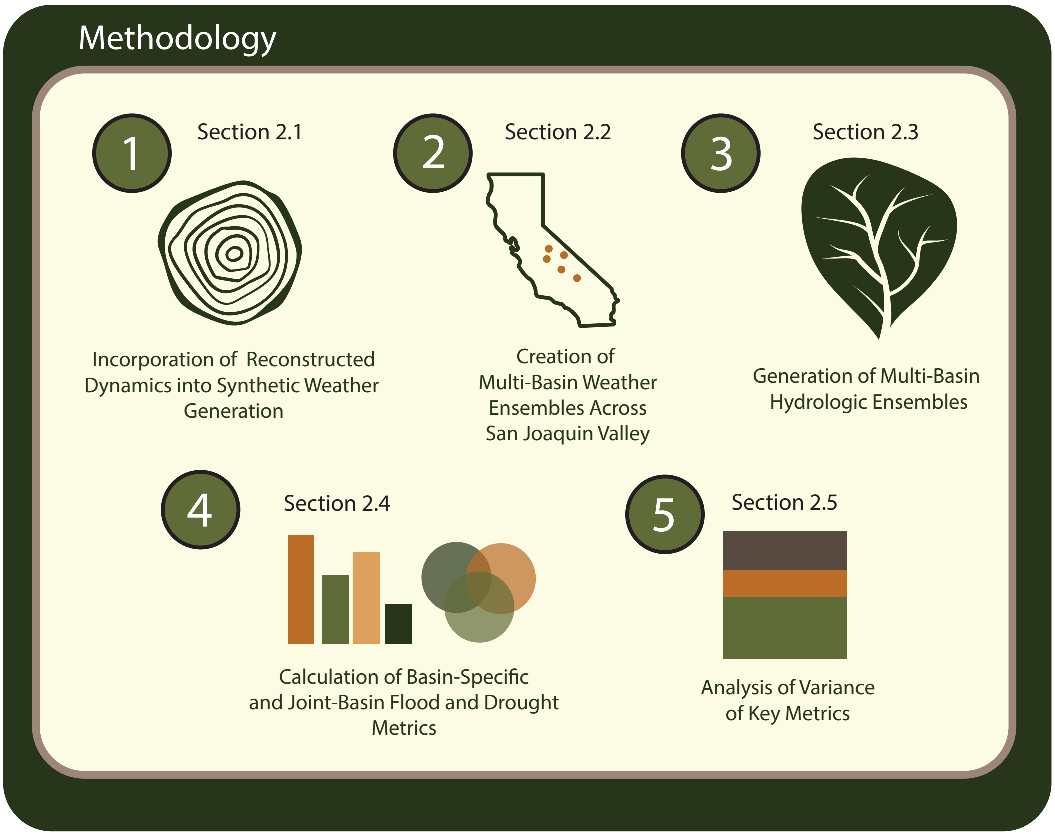

Figure 2. The main methodological components of this study. Reconstructed weather regime dynamics are used to drive weather simulations across the San Joaquin valley that also embed signals of thermodynamic climate change. Ensembles of generated surface weather are used to force hydrologic models to develop streamflow ensembles. Metrics pertaining to flood and drought extremes are calculated throughout the simulation period across basins and ensemble members, and an ANOVA decomposition is used to partition the variance of each metric across the ensemble into variance contributions from thermodynamic climate changes and natural variability.

These ensembles are forced through hydrologic models (SAC-SMA and HYMOD) calibrated for each subbasin to generate ensembles of daily streamflow (Section 2.3). From these streamflow ensembles, we calculate flood and drought metrics, including copula-based metrics to quantify joint flood hazard across basins (Section 2.4). Finally, analysis of variance (ANOVA) is used to partition the contribution of natural variability and the imposed climate changes to variability in the different flood and drought metrics considered (Section 2.5).

2.1 Reconstruction of WR Dynamics

Ensembles of plausible future climate are generated using our extensions of the WR-based stochastic weather generator presented in Steinschneider et al. (2019) and Najibi et al. (2021) to incorporate paleo-reconstructions of WRs. The generator is comprised of a three-step hierarchical structure (Figure 3): (1) identification and simulation of WRs that define large-scale patterns of atmospheric flow across the entire Western U.S., (2) simulation of local weather conditioned on the WRs, and (3) perturbations to the simulated weather reflective of thermodynamic climate change. This study extends step (1) to utilize reconstructed WRs created in Gupta et al. (2022). In that study, a multi-objective optimization and regression-based framework was used to reconstruct the annual frequency of five dominant Western U.S. weather regimes back to 1400 CE based on a gridded, tree-ring based reconstruction of cold season precipitation developed by A. P. Williams et al. (2020) and extended in Borkotoky et al. (2021). Specifically, the first four principal components of annual weather regime occurrence were reconstructed (termed PC_{WR} in Gupta et al. (2022)), which effectively contained all of the information on the annual frequencies of the five WRs. In this study, these principal components are used to force a non-homogeneous hidden Markov model (NHMM), whereby WR states are modeled as a first-order Markov chain with a non-stationary transition probability matrix conditioned on the reconstructed PC_{WR} from Gupta et al. (2022). The NHMM is fit to the first nine principal components of daily 500 hPa geopotential height from NOAA-CIRES-DOE Twentieth Century Reanalysis (V3) dataset (Slivinski et al., 2019) between 180-100°W and 30-60°N (i.e., the Pacific/North American sector) from 1950-2017. The NHMM is forced with the four reconstructed principal components (PC_{WR}) that overlap the same time period, defining a time-varying transition probability matrix shown in Equation 1:

$$P(WR_t = i | WR_t = j, \boldsymbol{X_t} = \boldsymbol{x}) = \frac{\exp(\beta_{0j,i} + \boldsymbol{\beta'_{j,i}}\boldsymbol{x})}{\sum_{k=1}^{K}\exp(\beta_{0j,i} + \boldsymbol{\beta'_{j,i}}\boldsymbol{x})} \tag{1}$$

Here, the transition probability from WR i to WR j at time t is conditioned on $\boldsymbol{X_t'} = [PC_{WR_{1,t}}, PC_{WR_{2,t}}, PC_{WR_{3,t}}, PC_{WR_{4,t}}]$ a vector of daily covariates developed by repeating the annual values of each for each day of the year. These covariates (Level 1 in Figure 3) are used within a multinomial logistic regression with intercepts $\beta_{0j,i}$ and coefficients $\boldsymbol{\beta_{j,i}}$ to define the transition probabilities, with a prime denoting the vector transpose. The fitted multinomial regression can be used to estimate the time-varying transition probabilities and simulate WRs across the entire 600-year period over which reconstructed values of PC_{WR} are available. More information on the NHMM can be found in Section S1. We use this method to create a 50-member ensemble of daily, 600-year weather regime time series (Level 2 in Figure 3; convergence plots of corresponding streamflow available in Figure S1).

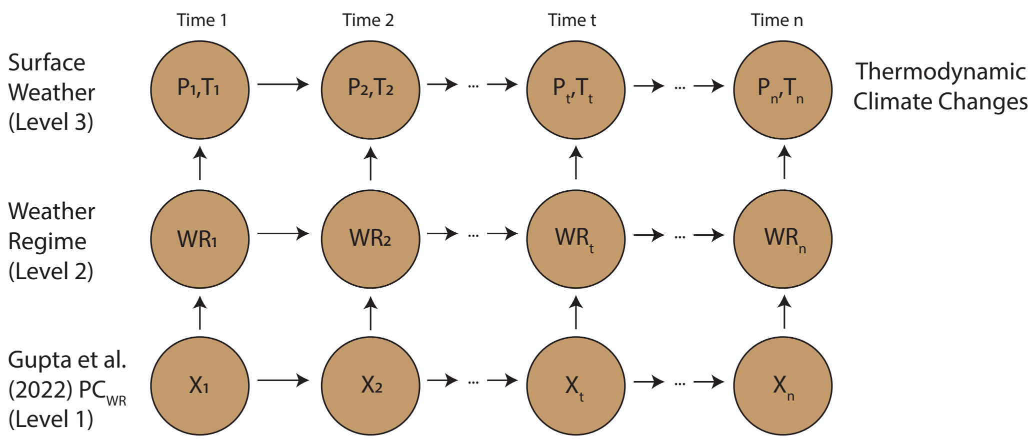

Figure 3. Flow chart of the main components of the weather generator. Boundary forcing variables (here the PC_{WR} from Gupta et al. (2022)) influence the evolution of the discrete weather regime time series. Surface weather is then conditioned upon the weather regime time series. Thermodynamic climate changes (temperature increases and precipitation-temperature scaling) are applied to the surface weather post-generation.

2.2 Generation of Local Surface Weather Conditioned on WRs

Time series of daily surface weather are generated based on the simulated time series of WRs (Level 3 in Figure 3). Here, observed daily precipitation, minimum, and maximum temperature are taken from the 1/16° resolution gridded meteorological dataset of Livneh et al. (2015) for water years (WY) 1950-2013. These historical weather data are block bootstrapped based on the sequence of simulated WRs to create new sequences of weather. For example, if the NHMM simulates a sequence of n consecutive days in WR i, an n-sized block of surface weather is resampled from the historical period that is also in WR i and that meets two other criteria: (1) the chosen historical block falls into a two-week window around the simulated day of the year; and (2) the day prior to the historical block is in the same precipitation state as the simulated day (wet or dry). The weather generator is run simultaneously across the five basins to create internally consistent (i.e., spatially correlated) weather across the region. Given their large synoptic scale, we use the occurrence of an atmospheric river (taken from Gershunov et al. (2017)) to represent a common precipitation state across basins. The process is repeated for each ensemble until a full sequence of 600 years of daily minimum and maximum temperature and precipitation has been generated.

Thermodynamic changes, in the form of shifts in temperature and precipitation scaling with warming, are imposed after surface weather is generated. Step changes in temperature between 0-4°C are added to each grid cell's simulated temperature. Quantile mapping is used to scale the precipitation distribution with warming, whereby the upper tail (99.9^{th} percentile) of the non-zero precipitation distribution is made more intense, the lower tail of non-zero precipitation is suppressed downward, but the mean of daily precipitation is left unchanged. This scaling reflects an intensification of the precipitation regime and is consistent with GCM-based projections of future precipitation in California (Michaelis et al., 2022). We consider 5 different scaling rates equivalent to $0X$ (0% $°C^{-1}$), $0.5X$ (3.5% $°C^{-1}$), $1X$ (7% $°C^{-1}$), $1.5X$ (10.5% $°C^{-1}$), and $2X$ (14% $°C^{-1}$) the Clausius-Clapeyron (CC) scaling rate, which dictates how the moisture holding capacity of the atmosphere scales with warming. That is, the 99.9th percentile of non-zero precipitation is scaled up by either 0% $°C^{-1}$, 3.5% $°C^{-1}$, 7% $°C^{-1}$, 10.5% $°C^{-1}$, or 14% $°C^{-1}$, while the lower body of the distribution is scaled down accordingly to maintain the same distribution mean. The range of selected scaling rates are derived from observational and model-based studies that most often indicate extreme precipitation-temperature scaling at the $1X$CC rate, but occasionally suggest the possibility for sub-CC ($0X$, $0.5X$) or super-CC ($1.5X$, $2X$) scaling rates due to interactive effects between enhanced specific humidity and storm dynamics (Wasko et al., 2018; Martinkova & Kysely, 2020; Ali et al., 2022; Michaelis et al., 2022; Sun & Wang, 2022).

These scaling rates are combined with the different scenarios of warming, so that precipitation scaling is tied to the imposed temperature scenario and respects the thermodynamic mechanism that drives precipitation change. We consider five scenarios of warming, including 0°C, 1°C, 2°C, 3°C, and 4°C above the climatological average. This range of warming was inferred from an ensemble of CMIP6 mid-century (2015-2050) projections over central California under the SSP2-4.5 scenario, taken from CarbonPlan (see Figure S2; Chegwidden et al. (2022)). All together, we develop 25 different scenarios of climate change (5 temperature scenarios and 5 scaling scenarios), with each scenario containing 50 ensemble members (i.e., 50 stochastic 600-year time series of precipitation and temperature), in addition to a baseline ensemble with no changes imposed. Technical details on the quantile mapping procedure, and other details of the stochastic weather generator, are provided in Steinschneider et al. (2019) and Najibi et al. (2021).

2.3 Generation of Regional Streamflow Through Process-Based Hydrologic Models

Surface weather ensembles are used to simulate daily streamflow ensembles at the mouth of each of the five San Joaquin subbasins using the Sacramento Soil and Moisture Accounting Model (SAC-SMA) (Burnash et al., 1995) coupled with a SNOW-17 model (Anderson, 1976). The models, documented in Wi and Steinschneider (2022), are spatially distributed and utilize a Lohmann routing model Lohmann et al. (1998) to trace runoff from hydrologic response units (HRUs) through each river channel. The SAC-SMA models are calibrated using a pooled calibration approach (Wi et al., 2015) based on the average Nash Sutcliffe Efficiency (NSE) across the five subbasins simultaneously. Calibration and evaluation was based on historical Full Natural Flows (FNF) between WY 1989-2013, acquired from California Data Exchange Center (CDEC) FNF stations that lie within each subbasin: Tuolumne River at La Grange Dam (TLG), Friant Dam on Millerton Lake (MIL), Merced River near Merced Falls (MRC), New Hogan Lake (NHG), and New Melones Reservoir (NML) (G. Huang & Kadir, 2016). The models are calibrated over WY 1989-2003 and then evaluated across WY 2004-2013.

To verify that our streamflow extremes and variance decomposition results are not strongly dependent on the selection of the SAC-SMA model, we also employ the HYMOD conceptual hydrologic model (HYMOD; (Moore, 2007)) specifically in the Tuolumne Basin. Our primary results will be presented using the SAC-SMA model but more detailed analysis of how hydrologic model selection impacts the estimates of flood and drought metrics as well as their partitioning of variance is provided in Section S2. More information about the calibration process and parameter values for all hydrologic models can be found in Wi and Steinschneider (2022).

2.4 Metrics of Hydrologic Extremes

A series of flood and drought metrics, described below, are calculated for each ensemble member and each climate scenario and across two time horizons: 30 and 100 years. As stated in the latest update to the Central Valley Flood Protection Plan (CVFPP), the state of California is actively prioritizing investments in flood management over a 30-year planning horizon (California Department of Water Resources, 2022). A 100-year planning horizon is not actively used in the CVFPP, but it represents a time scale relevant to longer term major infrastructure investments. Further, it allows the exploration of the longer climate time horizon drivers. We partition the variance of each metric between the drivers of climate change and natural climate variability using the ensemble of scenarios described above. Appendix A contains a glossary with commonly used terms that are referred to through the methods. Appendix B contains a summarized list of all of the flood and drought metrics used in this study, including their decision relevance.

2.4.1 Flood Metrics

Flows associated with a 10-year and 100-year return period are used as flood metrics in this study. The 100-year floodplain currently drives larger riverine infrastructure development and flood risk management in California (California Department of Water Resources, 2022). Though not as common for current planning and management in California, the 10-year return period flow captures risk to smaller floodplains and drives smaller investments (California Department of Water Resources, 2006). The decadal and centennial flood are estimated by fitting a generalized extreme value (GEV) distribution to the three-day annual maxima at each CDEC gauged location in the five subbasins. The three-day flood was chosen because it a common metric used in flood risk assessments in California (California Department of Water Resources, 2006; Chung, 2009; Brekke et al., 2009; Maurer, Brekke, & Pruitt, 2010; Maurer, Hidalgo, et al., 2010), and because it better captures the concurrence of flooding across multiple basins (described further in section 2.4.3). For each ensemble member, we fit the GEV distribution for the whole 600-year paleo-period as well as across smaller 30-year and 100-year moving windows.

2.4.2 Drought Metrics

There is no state statutory definition of drought since it can be classified differently across impacted sectors and stakeholders. Historical hydrologic droughts have been traditionally identified based on a combination of metrics that capture the magnitude and duration of water deficit at key reservoirs (California Department of Water Resources, 2015). Since we develop metrics for gauged locations near these reservoirs, we opt to use a more generalized Standardized Streamflow Index (SSI) to quantify hydrologic drought (Vicente-Serrano et al., 2012). To calculate the SSI, daily simulated flows are first aggregated to a monthly time step. We then use a flexible non-parametric empirical method to estimate non-exceedance probabilities using the Gringorten plotting position (see Farahmand and AghaKouchak (2015)). To create the SSI, the associated non-exceedance probabilities are passed through the quantile function of the standard normal distribution, resulting in a series with an assumed mean of zero and standard deviation of one. We then use the SSI index to define three drought metrics, following McKee et al. (1993):

1. Drought Occurrence: The number of months characterized by an SSI value less than -1, divided by the total months in the window over which the metric was calculated. An SSI value of less than -1 captures moderate to severe drought hazard.

2. Drought Intensity: The minimum SSI value in the moving window.

3. Drought Duration: The maximum number of consecutive months with an SSI below -1.5 in the moving window. An SSI value of less than -1.5 captures severe drought hazard.

The SSI index is calculated for each ensemble member and climate change scenario, and the metrics are reported across 30-year and 100-year moving windows.

2.4.3 Copula-Based Flooding Metrics

The San Joaquin basin is a key component in the state's comprehensive water delivery system, and a levee breach due to compounding flooding across subbasins in the region could disrupt deliveries of irrigation water to 3 million acres of farmland in the

Central Valley (Taylor, 2017). Thus, we develop a spatially-compounding flood metric to capture this hazard. As discussed in Zscheischler et al. (2020), spatially compounding flood hazard can be characterized using an n-dimensional Gaussian copula that defines a metric of joint flood hazard across n basins simultaneously. Let $x_{t,1}, \ldots, x_{t,n}$ be the annual maxima of 3-day mean streamflow in each of the n basins in year t. We first fit GEV distributions to the individual three-day annual maxima for each basin ($i = 1, \ldots, n$). The three-day annual maxima in each year t are then transformed to be uniform pseudo-observations, $u_{t,i} = F_{GEV}^{-1}(x_{t,i})$, where F_{GEV}^{-1} is the inverse cdf of the fitted GEV distribution for basin i. These pseudo-observations are used to evaluate the joint CDF of the flood data based on a Gaussian copula:

$$C(u_{t,1}, ..., u_{t,n}) = P(U_1 \leq u_{t,1}, ..., U_n \leq u_{t,n}) = \Phi_n(\phi^{-1}(u_{t,1}), ..., (\phi^{-1}(u_{t,n})|\Sigma) \quad (2)$$

Here, ϕ^{-1} is the inverse CDF of the standard normal distribution and $\Phi_n(\cdot|\Sigma)$ is the multivariate normal CDF with zero mean and correlation matrix Σ, which is set equal to the Spearman rank correlation matrix for three-day annual maxima across basins. Using the fitted copula, we can then calculate the joint probability that multiple subbasins experience flooding above some threshold. For example, consider two subbasins with 100-year flood magnitudes of x_1 and x_2, respectively, inferred from their fitted (GEV) marginal distributions. Then, the probability that both subbasins simultaneously experience floods that exceed the 100-year flood is equal to (Zhang & Singh, 2019):

$$\begin{aligned} P(X_1 > x_1, X_2 > x_2) &= 1 - P(X_1 \leq x_1) - P(X_2 \leq x_2) + P_{1,2}(X_1 \leq x_1, X_2 \leq x_2) = \\ &1 - F_{GEV1}(x_1) - F_{GEV2}(x_2) + \Phi_n(\phi^{-1}(F_{GEV1}(x_1)), \phi^{-1}(F_{GEV2}(x_2))|\Sigma) \quad (3) \end{aligned}$$

Similar calculations are available to evaluate the probability that three or more basins experience flooding above set thresholds. These probabilities can be used directly as a metric of joint flood hazard, and we can partition the variance of this metric between climate changes and natural variability across our ensemble and for 30-year and 100-year moving windows.

2.5 Analysis of Variance in Hydrologic Metrics

We use an ANOVA to partition the variance in the hydrologic flood and drought metrics above into components attributable to different sources of variation. A two-way ANOVA was used to determine the uncertainty in hydrologic metrics attributable to uncertainty in temperature change (T), precipitation scaling rate (P), their interactions, and uncertainty in metrics attributable to natural variability. The temperature change factor has $i = 1, \ldots, 5$ levels (0, 1, 2, 3, 4 °C), and precipitation scaling factor has j=1,. . . ,5 levels (0%, 3.5%, 7%, 10.5%, 14% per °C). For each combination of levels, there are 50 stochastic realizations of the metric of interest. The linear model on which the ANOVA is based is given as:

$$x(i,j,s) = \mu + \alpha(i) + \beta(j) + \gamma(i,j)^{TP} + \varepsilon(i,j,s) \tag{4}$$

Where $x(i,j,s)$ is the hydrologic metric for a given level i and j of factors T and P, respectively, and a given ensemble member s. The grand mean for the metric x across the entire ensemble is μ; $\alpha(i)$ equals the average deviation in x from μ for ensemble members with temperature changes at level i; $\beta(j)$ equals the average deviation in x from μ for ensemble members with precipitation scaling rate at level j; $\gamma(i,j)^{TP}$ is the interaction term between temperature change and precipitation scaling; and $\varepsilon(i,j,s)$ is the error term, which is used here to represent natural variability in the metric not explained by the different climate change factors. The total sum of squares SS_{total} expresses the total variation in the hydrologic metric x, and is comprised of the sum of variation attributable to temperature change (SS_T), precipitation scaling rate (SS_P), their interaction (SS_{Int}), and natural variability (SS_ε):

$$SS_{total} = SS_T + SS_P + SS_{Int} + SS_\varepsilon \tag{5}$$

The fraction of variance attributable to each source is calculated by dividing each component by SS_{total}. This fraction of attributable variance is calculated separately in 30-year and 100-year rolling windows for each of the metrics above.

3 Results

The results of this work are presented as follows. First, Section 3.1 shows a comparison of the variability in the paleo-informed streamflow with events from the available observed historical record. Then, Section 3.2 shows the flood and drought extremes reconstructed for the baseline scenario (i.e., influence of natural variability alone). Section 3.3 demonstrates how the imposed climate changes affect those extremes. Section 3.4 demonstrates the variance partitioning of extremes across climate change and natural variability. A more detailed evaluation of the stochastic weather generator's performance is presented in the Supporting Information (see Figures S3-S7), which demonstrates how well the generator captures characteristics of precipitation and minimum and maximum temperature.

3.1 Paleo-Informed Streamflow Characteristics

Figure 4 demonstrates the broader variability that is attained in the streamflow ensembles when SAC-SMA is forced with paleo-reconstructed weather at the Don Pedro gauge in the Tuolumne Basin. Figure 4a focuses on 7-day flows and the lower tail of the distribution and Figure 4b zooms in on the upper tail distribution of 3-day flows. Each grey line represents sorted flow volumes across 30-year chunks of the paleo-reconstruction across all 50 ensemble members. These volumes are compared with those that come from forcing the generator over the modern period (1987-2013) with historical Livneh precipitation and temperature data (red line). Key events from the observed record are annotated as colored horizontal lines. Overall, the paleo-informed streamflow envelopes and expands upon the historical SAC-SMA model flows by creating instances of wetter 3-day flows and drier 7-day flows. Furthermore, the paleo-ensemble is characterized by drier events than key drought periods from the observed record as demonstrated in Figure 4a. The generator is unable to create 3-day flows that reach the peak of the 1997 New Year's flood period due to underestimation of precipitation associated with this storm that is a known error in the Livneh dataset (Pierce et al., 2021). In turn, models conditioned on the Livneh dataset tend to underestimate the flows associated with this event. However, the inclusion of the paleo-reconstruction allows the generator to create flows that far surpass the magnitude of peak flows associated with the 1995 and 2017 floods. Overall, the expanded envelope of daily scale streamflows enabled by the paleo-reconstruction provide rich context for exploring plausible flood and drought extremes in the Tuolumne

Basin. Figure S9 demonstrates similar results for the rest of the San Joaquin River basins, particularly in capturing drought dynamics. The generator conditioned on the Livneh dataset suffers from the same difficulty of capturing the 1997 flood peak flows; however, in some basins like Merced and Millerton, the paleo-conditioned generator provides extended variability that can help overcome these limitations (Figures S9b,d). New Hogan Lake is the only gauged location in which the Livneh-conditioned model can capture the 1997 flood peak flows, but this is primarily because the associated peak flows were not as extreme in this region relative to other notable flooding events. Of the five basins, capturing dynamics in the Tuolumne is the most challenging; it is also representative of high-elevation basins that exhibit rich snow dynamics. Thus, we proceed through the rest of the results with a focus on the Tuolumne Basin, though corresponding figures for the rest of the basins can be found in the supplement. Section 3.2 further elaborates on the value of the paleo-forced generator and its representation of key flood and drought metrics through the reconstruction.

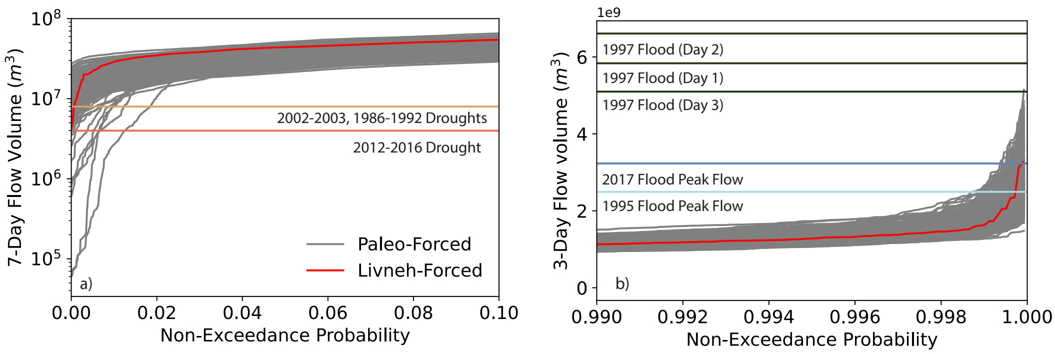

Figure 4. a) 7-day and b) 3-day flow volumes at the Don Pedro gauge in the Tuolumne Basin derived from the paleo-informed streamflow ensembles compared to the Livneh-forced generator over the modern period. Key events from the observed record are shown as colored lines. Each grey line represents sorted volumes for each year in 30-year chunks of the paleo-reconstruction across all 50 ensemble members.

3.2 Reconstruction of Natural Variability in Extremes

3.2.1 Individual Basin Flood Hazards

The individual Tuolumne subbasin flood hazard is quantified based on the 10-year and 100-year flood events associated with 3-day annual maximum flows, calculated using a GEV distribution fit to 3-day maxima in each basin and with two moving windows of length 30 and 100 years. Figures 5a and 5c show these return levels at the Don Pedro gauge in the Tuolumne Basin using a 30-year moving window. The return levels are calculated for all ensemble members of the baseline generator, where the solid line represents the mean return level across the ensemble members and the shading represents the $5^{\text{th}}/95^{\text{th}}$ percentiles. Figures 5b and 5d are non-exceedance plots of the three-day annual maxima across the extent of the paleo-reconstruction ensemble. The dashed black line represents the three-day annual maxima associated with the 10-year and 100-year return period events as derived from the SAC-SMA model forced with Livneh historical precipitation and temperature that overlaps with the observed record (1987-2013). In order to facilitate the most equivalent comparison between the two datasets, each gray line represents the sorted three-day annual maxima volumes over sets of 30-year segments of the paleo-reconstruction and across all 50 ensemble members.

The return levels in Figures 5a,c both show clear peaks centered around 1600 CE, which highlights a prominent pluvial period in the region's past hydroclimate. This pluvial is represented in the original WR reconstruction from Gupta et al. (2022) and broadly confirmed by other reconstructions (D'Arrigo & Jacoby, 1991; Schimmelmann et al., 1998; Stahle et al., 2007; M. D. Dettinger & Ingram, 2013). M. D. Dettinger and Ingram (2013) have also reconstructed pluvials around 1750-70 CE and 1810-20 CE, and while less pronounced than the 1600s pluvial, both panels a) and c) show increases in three-day annual maxima during these times. When compared to the model-based modern hydrology (dashed black line), both figures suggest that return levels in the most recent 30-year period are lower than those that have been experienced in prior centuries of the paleo-period reconstruction. Panels b) and d) show the modern estimates of the three-day annual maxima for the 10-year and 100-year events respectively, in comparison with the extent of the three-day annual maxima created by the paleo-informed generator. The ensemble from the generator encompasses the modern estimates of the return levels and also provides many instances of more extreme flooding events, which provides additional challenging flood scenarios that can be used to understand the vulnerability of water systems in each of the Central Valley subbasins explored in this study. As shown in Figure S10, the rest of the basins display similar three-day annual maxima dynamics, though the magnitude of the flows differs across all basins and return periods. Lower peak flows tend to be associated with basins that are smaller in area, elevation, and slope (i.e., New

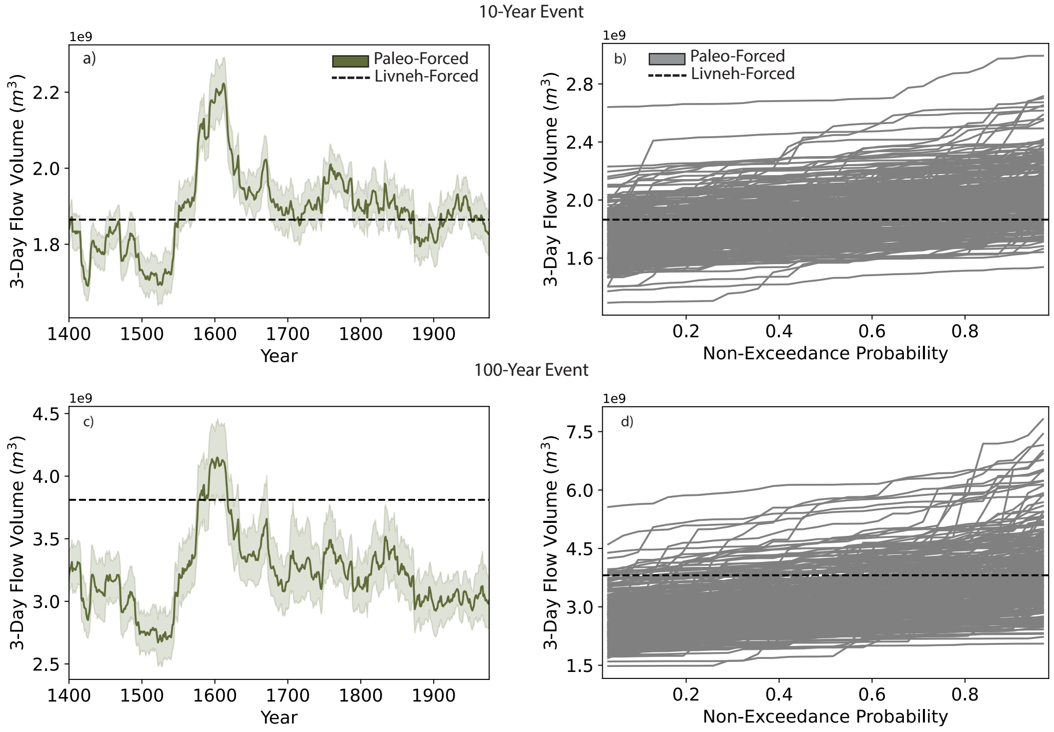

Figure 5. Three-day annual maxima associated with the a) 10-year return period event and c) 100-year return period event for the Don Pedro gauge in the Tuolumne subbasin calculated in 30-year moving windows and across the time period from 1400-2017. The dark green line represents the mean flooding return levels and the shading represents the 5$^{\text{th}}$ and 95$^{\text{th}}$ percentile confidence bounds. Panels b) and d) are non-exceedance plots of the three-day annual maxima across the extent of the paleo-reconstruction ensemble. Each gray line represents the sorted three-day annual maxima volumes for each year in a 30-year segment of the paleo-reconstruction. The dashed black line represents the three-day annual maxima associated with the 10-year and 100-year return period events as derived from the SAC-SMA-simulated peak flows when forced with Livneh historical data (1987-2013).

Hogan Lake, Table S1). The ensemble member spread also tends to be larger for the more extreme and uncertain 100-year flood event. Panels a) and c) exhibit clear non-stationary tendencies in the representation of the 10-year and 100-year event across the reconstruction that have large implications for hazard characterization. For example, the flow volumes associated with the 10-year event during the 1600s wet period are within range of the 100-year event flows during the 1500s megadrought period. Thus, what may be considered a 10-year flood event in one wet period transitions to be a 100-year event in a dry period. This extent of variability uncovered in the flood metric demonstrates that using only the modern record to define design flood events could severely under-represent flood hazard in the Central Valley region and that defining hazard based off of the 10-year and 100-year flooding events has drastically changed over time.

3.2.2 Individual Basin Drought Hazards

Figures 6a,c,e show the three SSI-based hydrologic drought metrics (occurrence, duration, and severity) calculated across a 30-year moving window for the period of 1400-2017 for the Don Pedro gauge in the Tuolumne River Basin. Figures 6b,d,f are non-exceedance plots, where each line corresponds to the sorted drought metric values derived across the whole reconstructed 617-year record length for each of the 50 ensemble members. The dashed line represents the respective metric values derived from the SAC-SMA model flows forced with Livneh historical precipitation and temperature across the length of the modern record. Similar to the flooding metrics in Section 3.2.1, the drought metrics exhibit clear decadal-scale variability that is also present in the original WR reconstruction from Gupta et al. (2022). For example, Figures 6a,c,e show declines in drought occurrence, severity, and duration during the early 1600s pluvial, while these drought characteristics become more intense during the 1500s megadrought that lasted from the middle of the century to the late 1580s (Stahle et al., 2007). The rest of the San Joaquin subbasins display this key behavior as well (Figure S11). The drought metrics reveal a slight long-term trend toward higher drought occurrence, longer duration, and more intense drought severity through the last three centuries of the reconstruction. This trend could, in part, be driven by key persistent drought periods that occurred in the mid to late 1800s (1856-1865, 1870-1877, and 1890-1896; Herweijer et al. (2006)), the 1900s (the Dust Bowl in the 1930s and drought periods in the 1950s and late 1980s; (Stahle et al., 2007)) and the most recent 20-year drought periods in the 2000s. The black line demonstrates drought occurrence and severity that is on par with the late 1500s megadrought, though exhibiting a slightly shorter duration than a large section of the paleo-reconstruction. The shorter drought duration is likely due to the sporadic periods of wet weather that have characterized the most recent 30-year period, including the early 1980s and late 1990s (M. Dettinger & Cayan, 2014) and periods after each drought instance in the 2000s.

Panels b), d), and f) compare the modern drought metrics to those calculated from the paleo-reconstructed ensembles. The ensembles encompass the modern estimates and

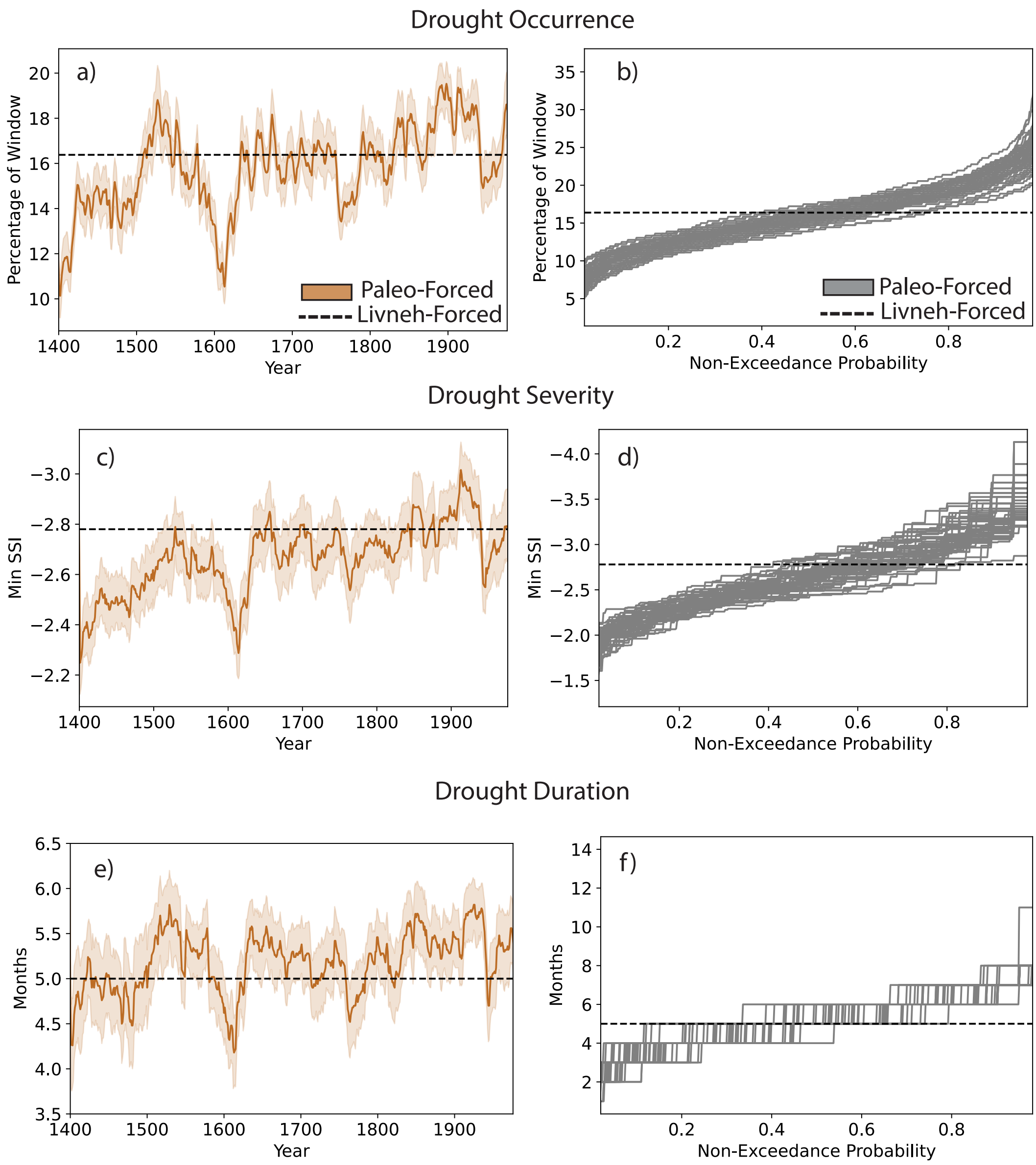

Figure 6. SSI-based hydrologic drought metrics of a) occurrence c) severity, and d) duration for the Don Pedro gauge in the Tuolumne Basin calculated in 30-year moving windows and across the time period from 1400-2017. The dark tan line represents the mean drought metric value and the shading represents the 5^{th} and 95^{th} percentile bounds. Panels b),d), and f) are non-exceedance plots of the three-day annual maxima across the extent of the paleo-reconstruction ensemble. Each gray line represents the sorted three-day annual maxima volumes across the length of the paleo reconstruction. The dashed black line represents the metric values as derived from the SAC-SMA-simulated peak flows associated with the modern record (1987-2013).

also provides many traces that are characterized by more frequent, longer, and severe drought. The plausibility of the Central Valley subbasins confronting drought conditions that extend well beyond those that have been experienced in the modern observed record captured in Livneh forcing data is significant even in the absence of climate change. The traces in panels b), d), and f) emphasize the need to better characterize the subbasin systems vulnerabilities for the challenging drought conditions that are captured within the reconstruction.

3.2.3 Joint Flood Hazard Across Basins

Gaussian copulas were fit to the 3-day annual maxima flows for multiple combinations of basins to characterize joint flood dynamics. The joint probability of flows at Don Pedro in the Tuolumne Basin and at Millerton Lake in the Millerton Basin simultaneously exceeding their respective, GEV-based 100-year flood estimates from the most recent 30-year period from 1987-2017 was calculated for the length of the reconstruction. Figure 7a shows the expected return period associated with those probabilities. Figure 7b includes New Melones Lake into the joint probability estimation. The return periods are calculated using a 30-year moving window across the entire reconstruction. Panels c) and d) are non-exceedance plots of the respective return periods across the extent of the paleo-reconstruction ensemble. The dashed black line represents the return periods for the 10-year and 100-year flood derived from the SAC-SMA model forced with Livneh historical precipitation and temperature. As with the flood metrics, in order to facilitate the most equivalent comparison between the two datasets, each gray line represents the sorted return periods for 30-year segments of the paleo-reconstruction and across all 50 ensemble members.

As demonstrated in Figure 7, there is a strong increase in the likelihood of simultaneously exceeding the recently observed historical estimate of the 100-year event, particularly during the 1600s wet period ($\sim$20% increase in likelihood). That is, the expected frequency of occurrence of simultaneous 100-year flooding events in both the Tuolumne and Millerton jumps to once every 320 years, as compared to once every 405 years in the most recent 30-year period. There is also a significant decline in the likelihood of joint flooding during the late 1500s megadrought. When an additional basin is introduced into the copula-based metric, the overall temporal dynamics are similar (Figure 7b), but the expected return period increases significantly. That is, the likelihood of simultaneously

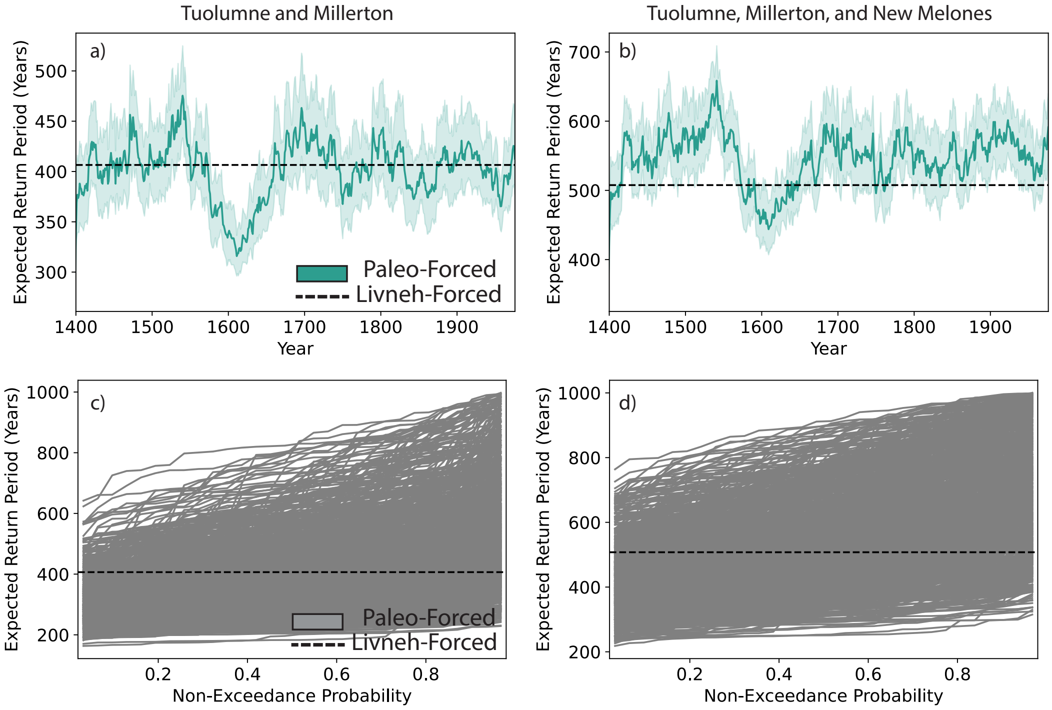

Figure 7. The expected return periods associated with the joint probability of simultaneously exceeding historical 100-year flood flows at a) Don Pedro (Tuolumne Basin) and Millerton Lake (Millerton Basin), and c) including New Melones (Stanislaus Basin) calculated in 30-year moving windows across the time period from 1400-2017. The dark turquoise line represents the average return period respectively across the ensemble, and the shading represents the 5th and 95th percentile bounds. Panels b) and d) show the non-exceedance plots for the return periods derived across the whole paleo-reconstruction in 30-year segments. The dashed black line represents the return periods as derived from the SAC-SMA-simulated peak flows associated with the modern record (1987-2013).

exceeding historical flooding thresholds rapidly declines as more basins are considered. During the 1600s wet period, the expected frequency of occurrence of simultaneous 100-year flooding events in the Tuolumne, Millerton, and New Melones jumps to once every 450 years, as compared to once every 507 years in the most recent 30-year period. For both joint flood metrics, the paleo-reconstruction effectively bounds the modern estimation of the return periods, which provides a richer space to characterize joint flood hazards across the subbasins (Figures 7c and 7d).

Similar non-stationary dynamics as observed in the flooding metrics in Section 3.2.1 are apparent in these joint flooding metric as well. Simultaneous flooding in all three basins is rarer and more consequential for water systems planning and management than simultaneous flooding in the Tuolumne and Millerton alone. Figures 7a-b demonstrate that through the paleo-reconstruction, there are periods (like the 1600s wet period) where the likelihood of flooding in the three basins becomes just as common as flooding in the Tuolumne and Millerton alone (around the late 1500s megadrought). The additional variability that the reconstruction provides demonstrates how dramatically the return periods associated with these consequential events changes over time, particularly how these flooding events can become more frequent. Once again, using the modern record to quantify joint hazard across these subbasins could severely underrepresent flood hazards and the magnitude of design events.

3.3 Effects of Thermodynamic Climate Change on Hydrologic Extremes

3.3.1 Changes in Individual Basin Flood Hazard

Figure 8 shows the effect of thermodynamic climate changes on the 100-year, 3-day flood event in the Tuolumne calculated across 30-year moving windows. The flow volumes are represented as deviations from the baseline reconstruction which is shown as a gray dashed line at 0. A modern baseline is placed as a dashed black line and is representative of the difference between the modern and the largest 100-year flood event volume calculated across the reconstruction. Figure 8a shows scenarios where the precipitation scaling rate is kept at 7% $°C^{-1}$ while temperature is increased by 1, 2, and 3 $°C$, while Figure 8b shows scenarios where the temperature trend is maintained at 1°C and the precipitation scaling rate is increased to 0% $°C^{-1}$, 7% $°C^{-1}$, and 14% $°C^{-1}$. Both increasing precipitation scaling rates and temperature trends shift the 100-year flood peak flows upwards, though temperature trends have a stronger impact. For reference, the volume differential between the extreme scenarios in Figure 8a is equivalent to about 100 Oroville Dams worth of water. Conversely, the maximum volume differential associated with the precipitation scaling in Figure 8b is equivalent to 33 Oroville Dams worth of water. The Tuolumne is a snow-dominated basin, and consequently it is not unexpected that the results suggest a greater influence on 100-year flows resulting from increasing temperature rather than increased precipitation scaling. Increased temperature shifts drive increased snowmelt and rain on snow events that promote greater flood volumes.

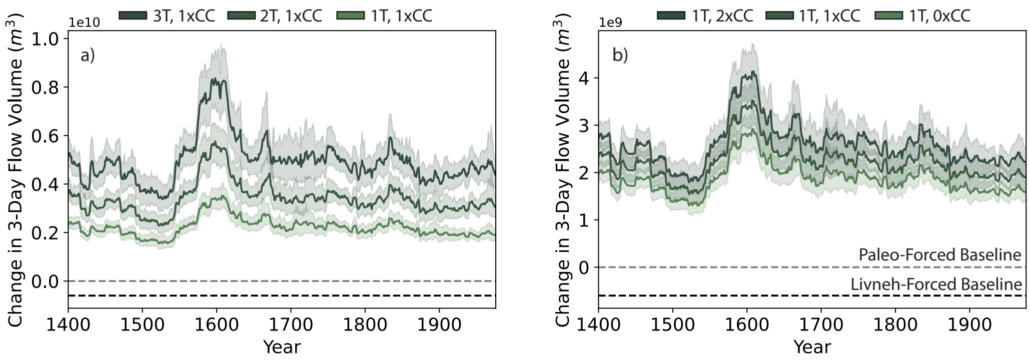

Figure 8. The effect of increasing a) temperature and b) precipitation scaling rates on 100-year, 3-day flood flows at Don Pedro (Tuolumne Basin). The dark green lines represent the increase in mean flooding return levels with respect to the baseline scenario (gray line at 0) and the shading represents the 5^{th} and 95^{th} percentile bounds. A modern baseline (black line) is included as reference and represents the distance from the modern peak flow to the maximum peak flow recorded in the reconstruction.

3.3.2 Changes in Individual Basin Drought Hazards

Figure 9 shows how the same thermodynamic scenarios imposed in Section 3.3.1 influence drought occurrence in the Tuolumne Basin, measured in terms of a change in the percent of the 30-year window that is classified to be in drought conditions with respect to the baseline scenario (gray dashed line at 0). A modern baseline is placed as a dashed black line and is representative of the difference between the modern drought occurrence line from Figure 6a and the worst drought occurrence metric calculated across the reconstruction. An increase in each of the thermodynamic mechanisms tends to increase the percentage of the window classified in drought. A comparison across Figures 9a and 9b show the larger impact of temperature trends on increased drought occurrence (reaching up to 5% or an additional 18 months classified in drought) by way of increased evapotranspiration. Precipitation scaling stretches the daily precipitation distribution which can lead to tail influences that impact the total number of drought months, but has a lower relative influence (reaching up to 1.8% or an additional 6 months classified in drought). For example, there are some instances, particularly in the 1T, 1xCC scenario in Figure 9a that result in values that approach the baseline. This is likely due to the precipitation scaling mechanism causing some months to have an increased SSI above the drought threshold that offsets the temperature increase. However, as the temperature shift further increases, this effect is dominated. Figure S12 shows the same results for drought severity and duration. Overall, there is a greater influence from increasing temperature trends to increasing drought severity and duration. It's worthwhile to note that the impact from both temperature trends and precipitation scaling is relatively small (Figure S12c,d) with respect to increasing consecutive months classified in severe drought and these results are further reflected in Figure 12.

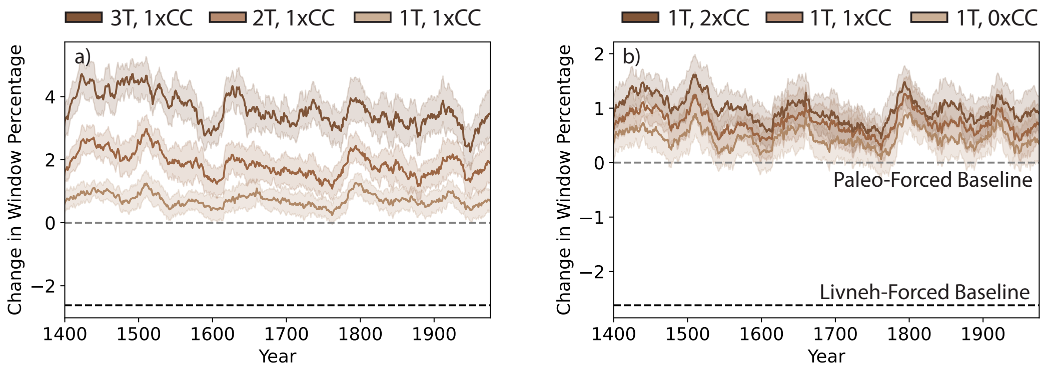

Figure 9. The effect of increasing a) temperature and b) precipitation scaling rates on drought occurrence at Don Pedro (Tuolumne Basin). The dark brown lines represent the increase in the percentage of the 30-year window classified in drought conditions with respect to the baseline scenario (grey line at 0) and the shading represents the 5^{th} and 95^{th} percentile bounds. A modern baseline is included (black line) as a reference and represents the distance from the modern drought occurrence metric to the worst drought occurrence recorded in the reconstruction.

3.3.3 Joint Flood Hazard Across Basins

Figure 10 shows how similar thermodynamic scenarios influence joint flood hazard at Don Pedro (Tuolumne Basin) and Millerton Lake (Millerton Basin), measured in terms of change to return period associated with the 100-year event with respect to the baseline scenario (gray dashed line at 0). As with the prior sections, a modern dashed black baseline is included to represent the difference between the modern return period estimate and the lowest return period calculated across the reconstruction. Much like Figure 8, Figure 10 demonstrates a larger influence from increasing temperature trends on making compound flooding events more likely (Figure 10a). Given that the Tuolumne and Millerton are both snow-dominated basins, temperature trends create similar snowmelt effects that lead to simultaneous flooding events. Precipitation scaling has a relatively reduced, but non-trivial effect (Figure 10b). The greatest influence from precipitation scaling is observed under higher imposed temperature trends (we use a constant 3°C temperature trend in this example). While an increase in precipitation scaling increases the likelihood of flooding in any given basin (Figure 8b), Figure 10b demonstrates that it decreases the likelihood of joint flooding, and makes the events rarer by increasing the return period. Since the imposed precipitation scaling non-linearly adjusts peak flows, it ultimately leads to a decrease in correlation in flows across the two basins and therefore a decrease in joint flooding tendencies.

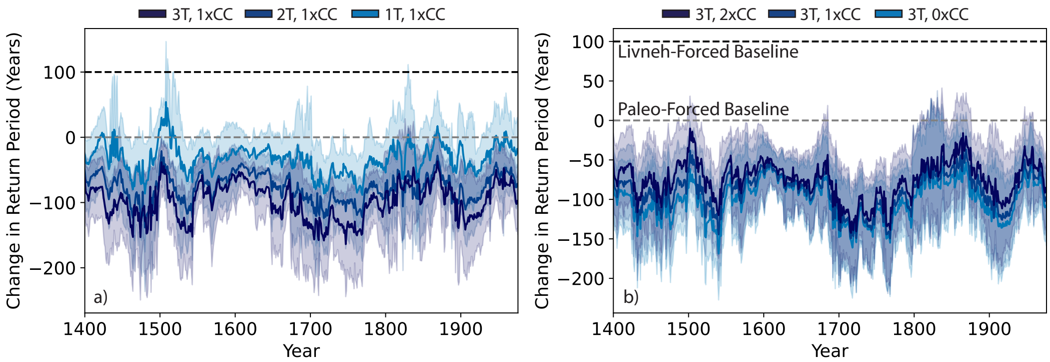

Figure 10. The effect of increasing a) temperature and b) precipitation scaling rates the change in return period associated with simultaneously exceeding historical 100-year-day flood flows at Don Pedro (Tuolumne Basin) and Millerton Lake (Millerton Basin). The dark blue lines represent the change in return period with respect to the baseline scenario (gray line at 0) and the shading represents the 5^{th} and 95^{th} percentile bounds. A modern baseline (black line at 0) is included as reference and represents the distance from the modern return period to the shortest return period recorded in the reconstruction.

3.4 Variance Partitioning of Hydrologic Extremes

The results above show how different metrics of hydrologic extremes vary significantly over time due to natural climate variability as well as different mechanisms of climate change. Below we use variance partitioning to assess the relative importance of these competing factors.

3.4.1 Relative Variance Contributions for Individual Basin Flood Hazard

We conduct an ANOVA to partition the variance of the 10-year and 100-year 3-day floods for each gauged location. Figure 11 shows the results for Don Pedro, while results for the other sites are shown in Figure S13-S16. The columns show the results of the decomposition when flood metrics are derived with a 30-year, 100-year, and 617-year (whole record) time horizon, respectively.

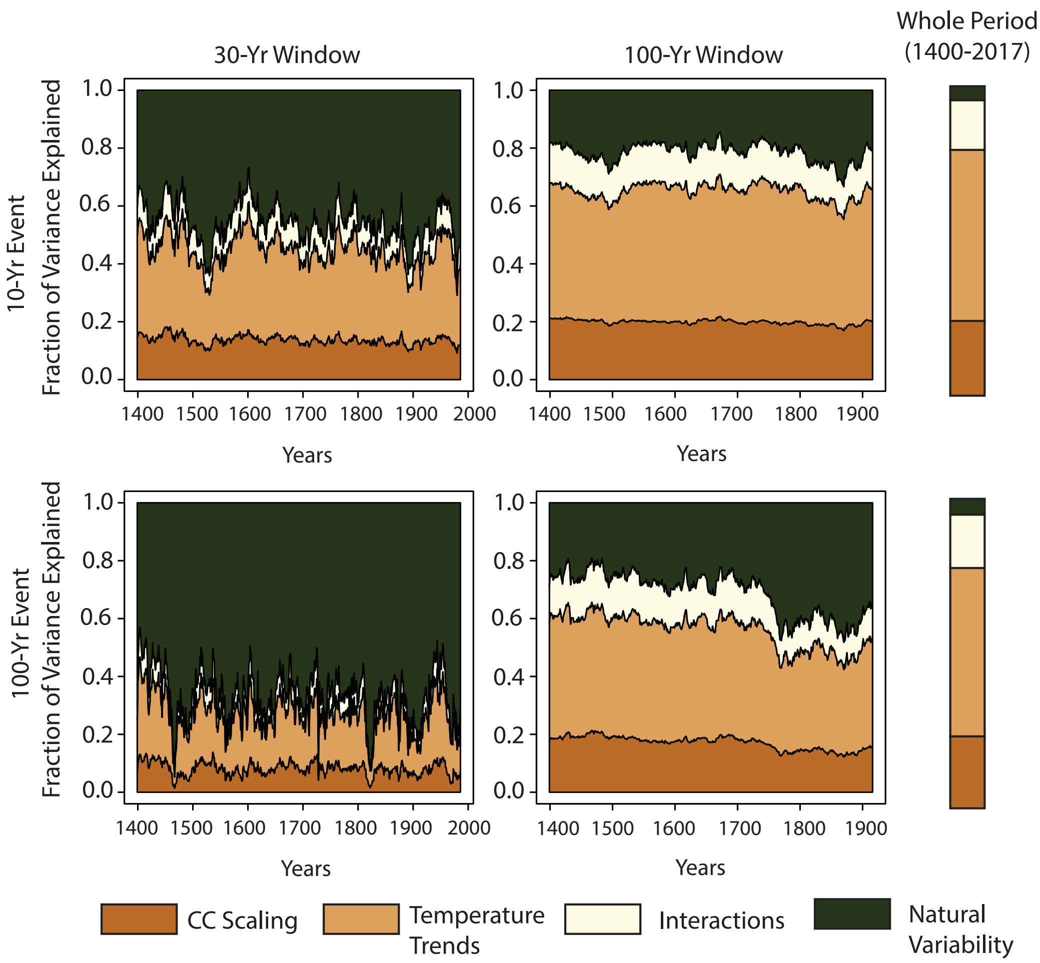

Figure 11. A decomposition of the key drivers of variance in the flood metrics for the Don Pedro gauge in the Tuolumne River Basin for an a,d) 30-year time horizon b,e) 100-year time horizon and c,f) a 617-year time horizon.

Two main insights emerge from Figure 11. First, natural variability is the primary driver of the variance when the flood metrics are calculated using a 30-year time horizon (Figures 11a,d). This is especially true for the 100-year flood, where approximately 70% of the variance in this metric is associated with natural variability. Figure 11d has direct relevance to the design standards actively used to inform California's flood planning and management. However, the influence of natural variability on the spread in flood metrics across the ensemble substantially decreases when the metric is calculated across a 100-year time horizon (Figures 11b,e), and becomes almost negligible when calculated over the entire 617-year period (Figures 11c,f). This suggests that the time horizon over which the flood metrics are calculated highly influences the perception of key drivers. A longer time horizon more clearly captures the effects of longer-term climate change on the variation in the flood metrics, while during shorter windows the variation in flood metrics across the ensemble is more likely to capture noise associated with natural variability. The reasons for this are twofold. First, when the time horizon is large, each ensemble member for a particular climate change scenario contains many annual maxima that are all drawn from the same underlying climate state, helping to converge design event estimates across ensemble members towards similar values. Second, when the time horizon is large, there are more opportunities for climate change signals to influence the distribution of annual maxima flows for all ensemble members under a given climate change scenario, which will help separate the distribution of annual maxima across the different scenarios. Together, these two factors will lead to more variance in the overall ensemble being explained by the climate change scenarios compared to natural variability.

Of the thermodynamic changes, temperature trends are the primary driver of variation in peak flows, followed by precipitation scaling. This result, also seen in Figure 8, suggests that temperature increases that lead to increased snowmelt and rain on snow events influences peak flows in the region more than increases in extreme precipitation due to increased moisture in the atmosphere. The interactions between the two drivers generally accounts for a smaller percentage of the variance, but as the time horizon increases, interactive effects are close to the same magnitude as precipitation scaling (16% vs. 24% for the whole period). This result highlights how the effects of precipitation scaling are dependent on the temperature increase, because precipitation scaling is parameterized as a percentage change in extreme precipitation per °C warming.

Figure S13-S16 show the same results for the remaining four basins. Overall, all basins exhibit similar behavior, where the influence of natural variability decreases with time horizon. Temperature change has a larger impact than precipitation scaling in all basins except for New Hogan Lake (Figure S15). New Hogan Lake is relatively small, has a low elevation, and less snow dominated compared to the other basins (Table S1), and thus sees a greater influence from precipitation scaling on flood variability.

Overall, the results in Figure 11 portray conflicting storylines and complexity for flood planning and management depending on the way the flood metrics are defined. Under current CA planning conditions (represented in Figure 11d), the greater influence of natural variability on individual flood hazard would suggest prioritizing short-term adaptive tools like seasonal forecasts. However, under alternative planning scenarios that may utilize longer time horizons, infrastructure investments look to be more useful to manage hazards from thermodynamic climate changes. Most importantly, water planners will need to engage with both drivers; prioritizing longer horizons of focus could neglect the effects of internal variability in the near term, which as Figure 5 portrays, can lead to magnitudes of peak flows that far surpass those in the modern record. Ultimately, there needs to be consideration of both the exceptional magnitude of internal variability in more immediate decision relevant 30-year timescales while still being cognizant of the longer-term climate changes. Thus, it's important for water resources agencies that utilize dynamic and adaptive planning methods to effectively balance the value, resilience, and potential regrets of near term investments (e.g. Haasnoot et al. (2013); Schlumberger et al. (2022)).

3.4.2 Relative Variance Contributions for Individual Basin Drought Hazards

Figure 12 shows the ANOVA decomposition for drought occurrence, intensity, and duration for 30-year and 100-year moving windows, as well as the entire 617-year period. The variance partitioning for drought occurrence follows a similar pattern to the flood metrics above (Figures 12a-c). For short time horizons of 30 years, about 20-40% of drought occurrence variability across the ensemble is associated with natural variability. However, as the time horizon grows, more variance is partitioned to the climate changes, and for extremely long horizons, almost all of the variance in drought occurrence across the ensemble is associated with climate change. Specifically, temperature change becomes the near-sole driver of drought occurrence variability, likely because of the strong increases in evapotranspiration with warming that drive drought occurrence.

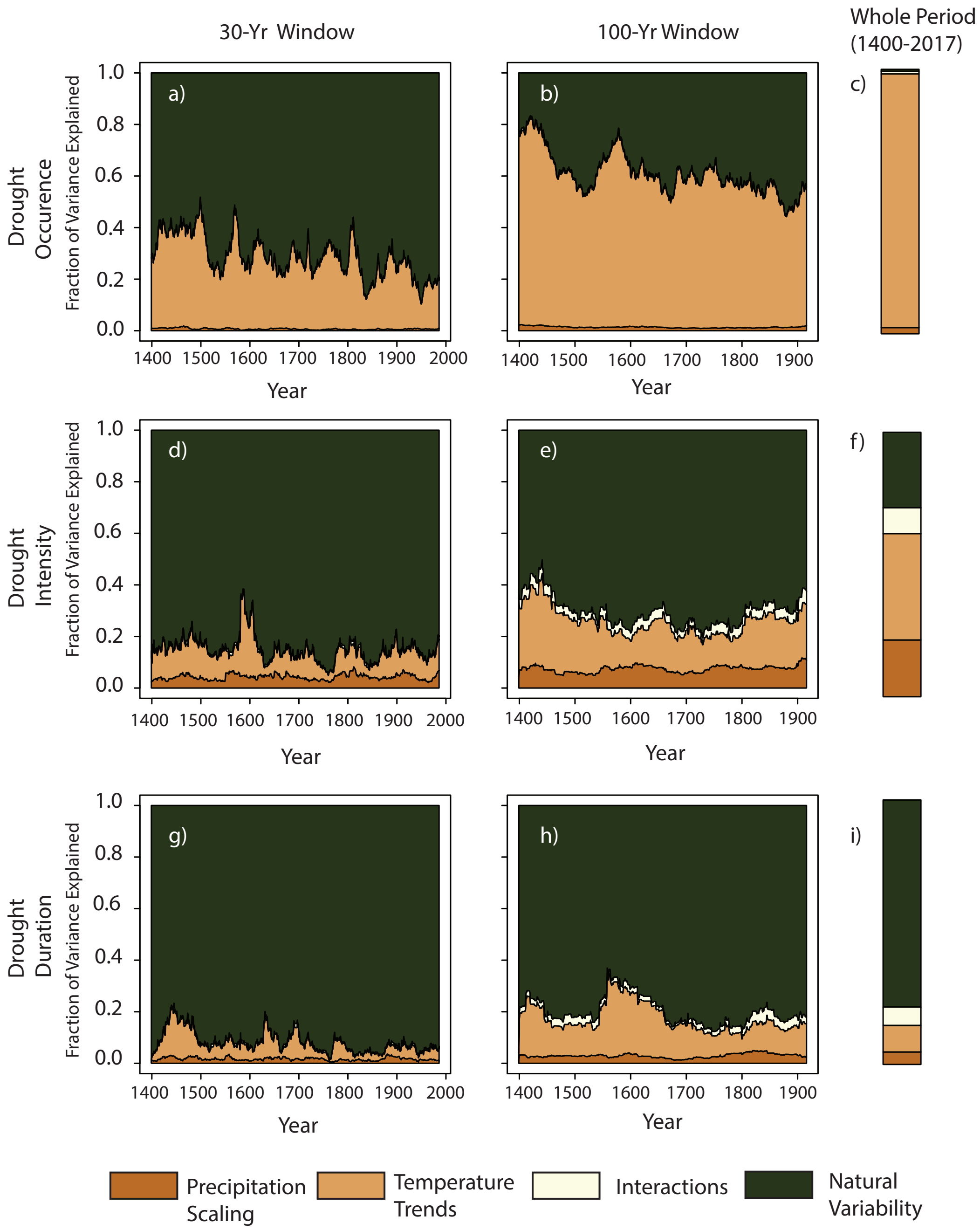

Figure 12. A decomposition of the key drivers of variance in the drought metrics for the Don Pedro gauge in the Tuolumne River Basin for a,d,g) 30-year window b,e,h) 100-year window and c,f,i) a 600-year window.

For drought intensity, we see a similar pattern in variance partitioning between natural variability and climate change factors, but the magnitude and degree of change in the variance partitioning more heavily favors natural variability (Figures 12d-f). For 30-

year windows, natural variability accounts for upwards of 80% of the total variance in drought intensity, and this falls to the (still substantive) value of 28% when the window reaches 617 years. Of the climate changes, temperature trends once again are the main driver, but precipitation scaling and interactive effects also play an important role in drought intensity variability across the ensemble. Given that the mechanism of precipitation scaling stretches the daily precipitation distribution such that large precipitation values become larger and small precipitation values become smaller, we see a more significant influence from this mechanism on drought intensity than in the other metrics.

Unlike the other two drought metrics, drought duration is primarily driven by natural variability, even when the metric is derived across the longest window. Drought duration generally is linked to the length of time in which there is no precipitation. None of the imposed climate changes directly affects this behavior in the same manner that precipitation scaling directly influences drought intensity or temperature trends affect drought occurrence. Temperature increases can somewhat extend drought duration by increasing evapotranspiration at the beginning and end of a drought period (Figure 12h), but ultimately the duration of a drought is dictated by the occurrence of large storms that end the drought, which is primarily driven by natural variability in our climate scenarios. The decomposition results for the remaining four gauged locations are presented in Figure S17-S20. These gauged locations show similar behavior as the Don Pedro gauge. Temperature trends play a large role in influencing drought occurrence, and this influence is particularly large in Merced and New Melones Lake (S17a, S20a). Precipitation scaling plays a small role in drought occurrence, and drought duration is primarily driven by natural variability.

The drivers of drought are more complex than the flood hazard metrics due to the heterogeneity of behavior across the drought metrics. A comparison between Figures 12a,d, and g demonstrate vast differences in drivers (and therefore approaches for managing drought) depending on exactly what characteristic of drought is prioritized in planning. The choice of time horizon further complicates the understanding of the appropriate planning process, especially in the case of drought occurrence (Figures 12a,b). However, drought intensity and drought duration show more stable influence primarily by natural variability and would consequently need a mix of carefully coordinated shorter-term adaptive actions (e.g., water transfers, conservation, and shifts in allocative priorities to higher value uses) that provide flexibility to improve the robustness of longer-term infrastructure investments to extreme variability in Central Valley drought regimes (e.g., improved conveyance, groundwater banking, managed aquifer recharge, and others; Herman et al. (2020); Hamilton et al. (2022)).

3.4.3 Relative Variance Contributions for Joint Flood Hazard

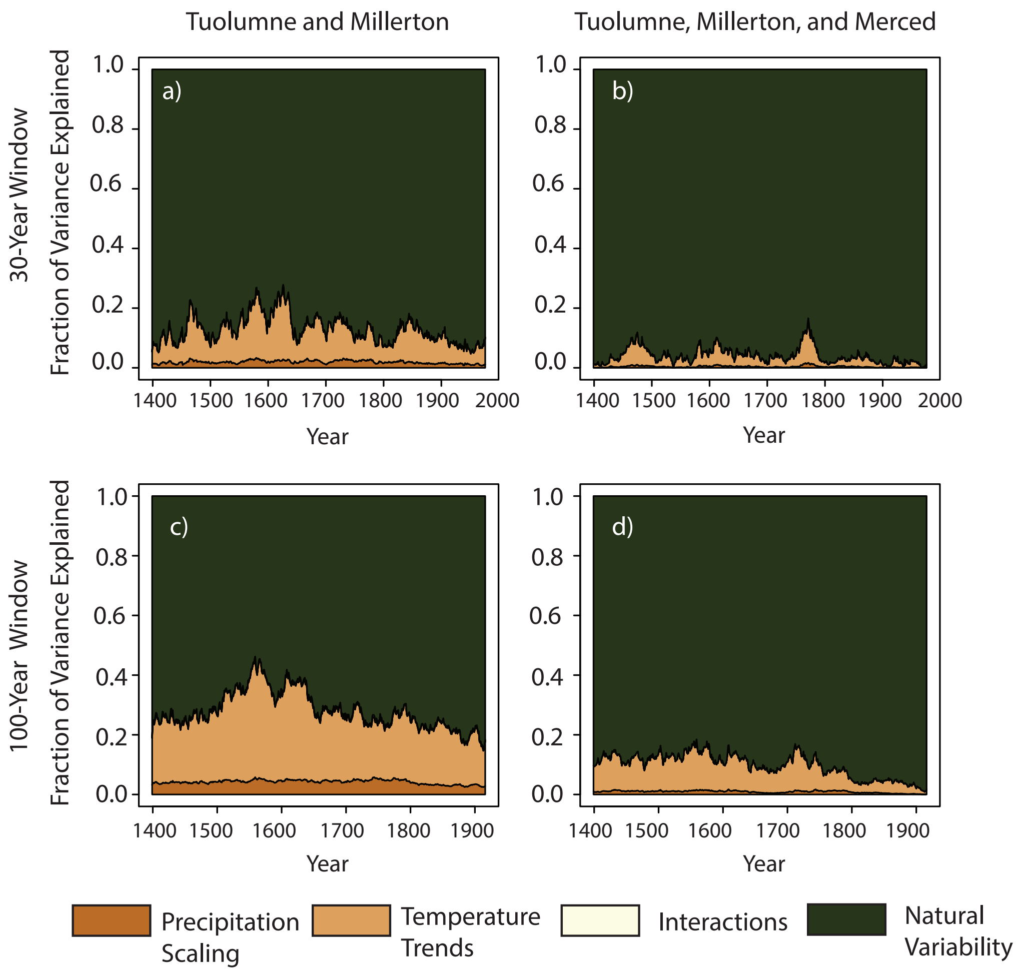

Figure 13. A decomposition of the key drivers of variance in joint flood metrics for a),c) Tuolumne and Millerton and b,d) Tuolumne, Millerton, and Merced.

Figure 13 shows the variance partitioning for the copula-based joint flood hazard metric in two cases: (1) bivariate flood risk in the Tuolumne and Millerton (Figure 13a,c); and (2) trivariate flood hazard in the Tuolumne, Millerton, and Merced (Figure 13b,d), both for the 100-year, 3-day flood. In both cases, the primary driver of joint flood hazard is natural variability. Unlike flood hazard for individual basins (see Figure 11), the contributions of natural variability to the total variance joint flood hazard does not decline substantially with time horizon. Additionally, as more locations are considered when quantifying joint flood hazard, natural variability becomes an even more prominent driver of spatially compounding major flood hazards. These results suggest that the dominating factor that dictates whether basins experience simultaneous large flooding is largely randomness in storm tracks and the associated spatial distribution of extreme precipitation and temperature-driven snowmelt. The thermodynamic climate changes that influence snowmelt or scale up storms do play a role, particularly if the basins are in close proximity (such as the Tuolumne and Millerton in Figures 13a,c). However, as more basins are included, natural variability in the weather during large storms dominates. Figure 13 reveals the inherent challenges of managing for spatially compounding flood hazards in this region. If persistent climate changes are a more dominant factor in driving joint flooding across all basins, then shared investments in canal expansion or rehabilitation across the regions could be used to offset some of this risk. However, since natural variability is the key driver of large flooding, alternative methods of creating unified planning and management strategies again need to be considered, using a mix of carefully coordinated shorter-term adaptive actions that provide flexibility to improve the robustness of longer-term infrastructure investments to the extreme hydro-climatic variability of the Central Valley (Herman et al., 2020; Hamilton et al., 2022).

4 Conclusion

This study contributes a novel framework to better understand the relative role of natural climate variability and climate change in determining the uncertainty in future hydrologic extremes of great importance to water systems planning and management. This framework is complementary to similar approaches based on GCM ensembles, but instead utilizes a large stochastic ensemble of paleo-based weather and hydrologic simulations to capture the plausible range of natural variability in drought and flood dynamics. The impacts of pre-selected mechanisms of climate change, including shifts in temperature and precipitation scaling, are then incorporated into the ensemble. The variance in hydrologic extremes is then partitioned across those climate changes and natural variability in the ensemble.

We first demonstrate the utility of the generator forced with paleodata in capturing and expanding on the dynamics of the modern record, which makes it a particularly useful for facilitating exploratory modeling and further quantification of the robustness of water resources systems to challenging scenarios that have been seen in the region's past hydroclimate. We also highlight the large non-stationarity that exists in the flood and drought metrics through the length of the reconstruction, particularly taking note of consequential 100-year flooding periods that can become as likely as 10-year events in parts of the record (i.e., 10 times more likely). These results have large implications for commonly employed stationary analyses, such as deriving design event estimates from the modern record, to quantify flood risk in this region. Our results suggest that these techniques severely underrepresent hydro-climatic hazards and the magnitude of design events that infrastructure should be built for.

The results of the variance decomposition component of the study highlight the following main conclusions:

- Uncertainty in future flooding within individual basins is largely driven by thermodynamic climate change, especially if evaluated over long time horizons. Flooding within snow-dominated basins is primarily driven by changes in temperature, while lower-elevation basins see a greater influence from precipitation scaling.
- The relative importance of climate change and natural variability on the uncertainty in future drought depends on the drought metric of interest. Changes in temperature drive drought occurrence, while precipitation scaling plays a role in drought intensity. Drought duration is primarily driven by natural variability.
- The uncertainty in simultaneous flood hazard across multiple basins is largely driven by natural variability, and this influence increases as additional basins are considered.
- The perception of the most important driver is highly influenced by the time horizon over which a metric is calculated. Shorter time horizons are less likely to capture how climate change uncertainty influences the uncertainty in hydrologic extremes.

The variance decomposition reveals a complicated path to robust planning and managing for both flood and drought in the region. The results suggest that natural variability and climate change influence both extremes to varying degrees. Furthermore, different characteristics of a single extreme (i.e. drought occurrence and duration) can be influenced by different drivers.

Additionally, if different time horizons are prioritized for planning for extremes, the understanding of the most important drivers of flood and drought hazards also changes. This last facet especially presents a problem for adaptive planning and management. This type of planning triggers management decisions based on the evolution of an observed variable (including hydroclimatic variables like precipitation or streamflow) over a specific horizon. As demonstrated in our study, tracking peak flows over a 30-year or 100-year horizon are both appropriate for longer-term flood management, but prioritizing the latter could neglect the effects of internal variability in the near term while increasing the potential for maladaptive longer-lived capital investments in infrastructure. Thus, it's important for water resources agencies that utilize these dynamic planning methods to effectively balance the value and potential regret of near term investments (Herman et al., 2020; Schlumberger et al., 2022).

One of the most important results of our study is that natural variability plays a very large role in dictating the future uncertainty in key metrics of flood and drought that form the basis of water resources planning; at times much larger than that of prominent climate change signals. This suggests that better quantification of the true range of natural variability in these extremes should be a major priority for the climate and hydrologic research community, and equally important, these efforts should directly inform future planning efforts for water resources systems. However, historically, this has often not been the case, with concerns about climate change often overshadowing the potential impacts of natural variability (see discussions in Koutsoyiannis (2020, 2021)).

Our results show, in particular, the importance of natural variability on spatially compounding flood hazard, which arguably poses a more difficult and complex management problem than addressing hazards in any one basin due to the need for infrastructure coordination across space and time. This highlights the potential value that longer, paleo-based data could bring to the estimation of joint flood hazards. The field of paleoflood hydrology has historically focused on the identification and dating of flood evidence in fluvial sedimentary archives, but incorporating speleothems and botanical archives can substantially increase the comprehensiveness and quality of paleoflood data (Wilhelm et al., 2018). Alluvial archives are also being used in more densely-populated and flood-prone regions (Toonen et al., 2020), and recent studies have shown that incorporation of these data can significantly reduce the uncertainty of extreme flood estimates (Engeland et al., 2020; Reinders & Muñoz, 2021). Methodological advances that can use these new and diverse data sources to constrain joint flood hazard estimates across sites would be particularly helpful, as would guidance on how to appropriately and consistently incorporate paleodata into risk management practices that also consider the effects of climate change. The work of England Jr et al. (2019) that helped incorporate paleodata into U.S. flood frequency guidance (Bulletin 17C) provides inspiration for such an approach.

The results also highlight the significant impact of natural variability on drought uncertainty, especially drought duration and intensity, and the implications stated above for joint flood hazards also extend to drought hazards. There are state-of-the-art techniques currently being applied within the dendrochronology community that can help improve our understanding of the natural range of drought variability. Beyond using tree ring widths, some studies are isolating earlywood and latewood signals for better drought reconstruction (Soulé et al., 2021; Song et al., 2022) or using blue intensity (the intensity of reflectance of the blue channel light from a wood core) to identify more stable climate-growth relationships that inform more robust reconstructions (Akhmetzyanov et al., 2023). Furthermore, better forecasts could provide water managers with more effective ways to navigate drought caused by natural variability. Skillful near-term drought predictions have been achieved by using decadal hindcasts from CMIP6 (Zhu et al., 2020) and Machine learning based approaches, particularly those that can model catchment memory are being used to create skillful seasonal drought predictions (Amanambu et al., 2022; Sutanto & Van Lanen, 2022)

One key limitation of this work is that we only consider a subset of plausible climate change scenarios that are not comprehensive, but rather reflect two mechanisms of change that are likely to occur and to be consequential to the San Joaquin Valley in California. This limitation includes the omission of the possibility that properties of long-term climate variability will itself change in the future under climate change. Another limitation is that we represent natural variability with one statistical model based on historical and paleo data. As others have shown (Koutsoyiannis, 2021), the quantification of natural variability often greatly depends on the statistical model used.

While outside the scope of this study, the framework presented and conclusions drawn here would benefit from a direct comparison against a similar approach using a climate ensemble drawn from a GCM, especially a single model initial-condition large ensemble (SMILE; see Lehner et al. (2020)). In a SMILEs-based framework, projections of precipitation and temperature derived from a single GCM under multiple initial conditions and multiple emission scenarios could be downscaled and propagated through hydrologic models to create a future streamflow ensemble, which could be used for partitioning variance in hydrologic extremes across emission scenarios and natural variability. By comparing results between the framework of this study and a SMILEs-based framework, one could better understand whether and how the relative roles of natural variability and climate change are consistent or depend on methodological choice.

Regardless of method used, the results of this work strongly suggest that large ensembles of natural variability are likely needed to adequately assess future risks to water resources systems that are particularly sensitive to extreme events. In future work, we intend to pair the hydrologic ensembles developed here with a regional, multi-sector model of California's Central Valley (Zeff et al., 2021) to more fully assess the risk that future hydroclimate extremes pose to stakeholders across the system, including groundwater banks and irrigation districts. The ultimate goal of such work is to facilitate a greater understanding of how future extremes lead to heterogeneous shortage and flooding impacts across stakeholders, and to help identify robust adaptation strategies to address these future risks.

Acknowledgments

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1650441. The views expressed in this work represent those of the authors and do not necessarily reflect the views or policies of the National Science Foundation.

Data Availability Statement

Sample input data and code to run the weather generator and hydrologic models, create flood and drought metrics metrics, and create figures can be found at `https://doi.org/10.5281/zenodo.7693324`. Refer to the associated GitHub repository: `https://github.com/rg727/Gupta_WGEN_Partitioning_NatVar_CC_Drivers`

References

Akhmetzyanov, L., Sánchez-Salguero, R., García-González, I., Domínguez-Delmás, M., & Sass-Klaassen, U. (2023). Blue is the fashion in mediterranean pines: New drought signals from tree-ring density in southern europe. *Science of the Total Environment*, *856*, 159291.

Ali, H., Fowler, H. J., Pritchard, D., Lenderink, G., Blenkinsop, S., & Lewis, E. (2022). Towards quantifying the uncertainty in estimating observed scaling rates. *Geophysical Research Letters*, *49*(12), e2022GL099138.

Amanambu, A. C., Mossa, J., & Chen, Y.-H. (2022). Hydrological drought forecasting using a deep transformer model. *Water*, *14*(22), 3611.

Anderson, E. A. (1976). *A point energy and mass balance model of a snow cover.* Stanford University.

Bass, B., Norris, J., Thackeray, C., & Hall, A. (2022). Natural variability has concealed increases in western us flood hazard since the 1970s. *Geophysical Research Letters*, *49*(7), e2021GL097706.

Borkotoky, S. S., Williams, A. P., Cook, E. R., & 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.

Brekke, L. D., Maurer, E. P., Anderson, J. D., Dettinger, M. D., Townsley, E. S., Harrison, A., & Pruitt, T. (2009). Assessing reservoir operations risk under climate change. *Water Resources Research*, *45*(4).

Burnash, R., et al. (1995). The nws river forecast system-catchment modeling. *Computer models of watershed hydrology.*, 311–366.

Cai, W., Zhu, X., Zhao, X., & Zhang, Y. (2021). Estimating the effect of climate internal variability and source of uncertainty in climate-hydrological projections in a representative watershed of northeastern china.

California Department of Water Resources. (2006). Progress on incorporating climate change into planning and management of california's water resources, california department of water resources. *California Department of Water Resources*.

California Department of Water Resources. (2015). California's most significant droughts: Comparing historical and recent conditions. *California Department* *of Water Resources*.

California Department of Water Resources. (2022). Central valley flood protection plan update 2022. *California Department of Water Resources*.

Chegwidden, O. S., Hagen, R., Martin, K., Jones, M., Banihirwe, A., Chiao, C., ... Hamman, J. (2022, October). *Downscaling CMIP6 with multiple statistical methods.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.7145491` (The development of this project was funded, in part, through a grant from the Microsoft AI for Earth program to CarbonPlan.) doi: 10.5281/zenodo.7145491

Chung, F. (2009). *Using future climate projections to support water resources decision making in california*. California Energy Commission.

Cook, B. I., Cook, E. R., Smerdon, J. E., Seager, R., Williams, A. P., Coats, S., ... Díaz, J. V. (2016). North american megadroughts in the common era: Reconstructions and simulations. *Wiley Interdisciplinary Reviews: Climate Change*, *7*(3), 411–432.

D'Arrigo, R. D., & Jacoby, G. C. (1991). A 1000-year record of winter precipitation from northwestern new mexico, usa: a reconstruction from tree-rings and its relation to el niño and the southern oscillation. *The Holocene*, *1*(2), 95–101.

Deser, C., Lehner, F., Rodgers, K. B., Ault, T., Delworth, T. L., DiNezio, P. N., ... others (2020). Insights from earth system model initial-condition large ensembles and future prospects. *Nature Climate Change*, *10*(4), 277–286.

Dessai, S., & Hulme, M. (2004). Does climate adaptation policy need probabilities? *Climate policy*, *4*(2), 107–128.

Dettinger, M., & Cayan, D. R. (2014). Drought and the california delta—a matter of extremes. *San Francisco Estuary and Watershed Science*, *12*(2).

Dettinger, M. D., & Ingram, B. L. (2013). The coming megafloods. *Scientific American*, *308*(1), 64–71.

Engeland, K., Aano, A., Steffensen, I., Støren, E., & Paasche, Ø. (2020). New flood frequency estimates for the largest river in norway based on the combination of short and long time series. *Hydrology and Earth System Sciences*, *24*(11), 5595–5619.

England Jr, J., Cohn, T., Faber, B., Stedinger, J., Thomas Jr, W., Veilleux, A., & Mason Jr, R. (2019). Guidelines for determining flood flow frequency—bulletin

17c (no. 4-b5). *US Geological Survey*.

Farahmand, A., & AghaKouchak, A. (2015). A generalized framework for deriving nonparametric standardized drought indicators. *Advances in Water Resources*, *76*, 140–145.

Gershunov, A., Shulgina, T., Ralph, F. M., Lavers, D. A., & Rutz, J. J. (2017). Assessing the climate-scale variability of atmospheric rivers affecting western north america. *Geophysical Research Letters*, *44*(15), 7900–7908.

Gupta, R. S., Steinschneider, S., & Reed, P. M. (2022). A multi-objective paleo-informed reconstruction of western us weather regimes over the past 600 years. *Climate Dynamics*, 1–20.

Haasnoot, M., Kwakkel, J. H., Walker, W. E., & Ter Maat, J. (2013). Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world. *Global environmental change*, *23*(2), 485–498.

Hamilton, A. L., Zeff, H. B., Characklis, G. W., & Reed, P. M. (2022). Resilient california water portfolios require infrastructure investment partnerships that are viable for all partners. *Earth's Future*, *10*(4), e2021EF002573.

Hamlet, A. F., & Lettenmaier, D. P. (2007). Effects of 20th century warming and climate variability on flood risk in the western us. *Water Resources Research*, *43*(6).

Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, *90*(8), 1095–1108.

Henn, B., Musselman, K. N., Lestak, L., Ralph, F. M., & 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.

Herman, J. D., Quinn, J. D., Steinschneider, S., Giuliani, M., & Fletcher, S. (2020). Climate adaptation as a control problem: Review and perspectives on dynamic water resources planning under uncertainty. *Water Resources Research*, *56*(2), e24389.

Herweijer, C., Seager, R., & Cook, E. R. (2006). North american droughts of the mid to late nineteenth century: a history, simulation and implication for mediaeval drought. *The Holocene*, *16*(2), 159–171.

Hoerling, M., Quan, X.-W., & Eischeid, J. (2009). Distinct causes for two principal us droughts of the 20th century. *Geophysical Research Letters*, *36*(19).

Huang, G., & Kadir, T. (2016). Estimates of natural and unimpaired flows for the central valley of california: Water years 1922–2014. *California Department of Water Resources*.

Huang, X., & Swain, D. L. (2022). Climate change is increasing the risk of a california megaflood. *Science advances*, *8*(31), eabq0995.

Hunt, B. (2011). Global characteristics of pluvial and dry multi-year episodes, with emphasis on megadroughts. *International Journal of Climatology*, *31*(10), 1425–1439.

Jung, I.-W., Chang, H., & Moradkhani, H. (2011). Quantifying uncertainty in urban flooding analysis considering hydro-climatic projection and urban development effects. *Hydrology and Earth System Sciences*, *15*(2), 617–633.

Kay, A., Davies, H., Bell, V., & Jones, R. (2009). Comparison of uncertainty sources for climate change impacts: flood frequency in england. *Climatic change*, *92*(1-2), 41–63.

Kirchmeier-Young, M. C., & Zhang, X. (2020). Human influence has intensified extreme precipitation in north america. *Proceedings of the National Academy of Sciences*, *117*(24), 13308–13313.

Knutti, R., Sedláček, J., Sanderson, B. M., Lorenz, R., Fischer, E. M., & Eyring, V. (2017). A climate model projection weighting scheme accounting for performance and interdependence. *Geophysical Research Letters*, *44*(4), 1909–1918.

Koutsoyiannis, D. (2020). Revisiting the global hydrological cycle: is it intensifying? *Hydrology and Earth System Sciences*, *24*(8), 3899–3932.

Koutsoyiannis, D. (2021). Stochastics of hydroclimatic extremes—a cool look at risk. *National Technical University of Athens: Athens, Greece*.

Kunkel, K. E. (2003). North american trends in extreme precipitation. *Natural hazards*, *29*(2), 291–305.

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., ... Hawkins, E. (2020). Partitioning climate projection uncertainty with multiple large ensembles and cmip5/6. *Earth System Dynamics*, *11*(2), 491–508.

Livneh, B., Bohn, T. J., Pierce, D. W., Munoz-Arriola, F., Nijssen, B., Vose, R., ... Brekke, L. (2015). A spatially comprehensive, hydrometeorological data set for mexico, the us, and southern canada 1950–2013. *Scientific data*, *2*(1), 1–12.

Lohmann, D., Raschke, E., Nijssen, B., & Lettenmaier, D. (1998). Regional scale hydrology: I. formulation of the vic-2l model coupled to a routing model. *Hydrological sciences journal*, *43*(1), 131–141.

Martinkova, M., & Kysely, J. (2020). Overview of observed clausius-clapeyron scaling of extreme precipitation in midlatitudes. *Atmosphere*, *11*(8), 786.

Maurer, E. P., Brekke, L. D., & Pruitt, T. (2010). Contrasting lumped and distributed hydrology models for estimating climate change impacts on california watersheds 1. *JAWRA Journal of the American Water Resources Association*, *46*(5), 1024–1035.

Maurer, E. P., Hidalgo, H. G., Das, T., Dettinger, M. D., & Cayan, D. R. (2010). The utility of daily large-scale climate data in the assessment of climate change impacts on daily streamflow in california. *Hydrology and Earth System Sciences*, *14*(6), 1125–1138.

McCabe, G. J., Palecki, M. A., & Betancourt, J. L. (2004). Pacific and atlantic ocean influences on multidecadal drought frequency in the united states. *Proceedings of the National Academy of Sciences*, *101*(12), 4136–4141.

McKee, T. B., Doesken, N. J., Kleist, J., et al. (1993). The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th conference on applied climatology* (Vol. 17, pp. 179–183).

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*(3), e2021EF002537.

Moon, H., Gudmundsson, L., & Seneviratne, S. I. (2018). Drought persistence errors in global climate models. *Journal of Geophysical Research: Atmospheres*, *123*(7), 3483–3496.

Moore, R. (2007). The pdm rainfall-runoff model. *Hydrology and Earth System Sciences*, *11*(1), 483–499.

Najibi, N., Mukhopadhyay, S., & Steinschneider, S. (2021). Identifying weather regimes for regional-scale stochastic weather generators. *International Journal of Climatology*, *41*(4), 2456–2479.

Nazemi, A., & Wheater, H. S. (2014). Assessing the vulnerability of water supply to changing streamflow conditions. *Eos, Transactions American Geophysical Union*, *95*(32), 288–288.

Null, J., & Hulbert, J. (2007). California washed away: The great flood of 1862. *Weatherwise*, *60*(1), 26–30.

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.

Reinders, J. B., & Muñoz, S. E. (2021). Improvements to flood frequency analysis on alluvial rivers using paleoflood data. *Water Resources Research*, *57*(4), e2020WR028631.

Rocheta, E., Sugiyanto, M., Johnson, F., Evans, J., & Sharma, A. (2014). How well do general circulation models represent low-frequency rainfall variability? *Water Resources Research*, *50*(3), 2108–2123.

Rowell, D. P. (2012). Sources of uncertainty in future changes in local precipitation. *Climate dynamics*, *39*, 1929–1950.

Schiemann, R., Athanasiadis, P., Barriopedro, D., Doblas-Reyes, F., Lohmann, K., Roberts, M. J., … Vidale, P. L. (2020). Northern hemisphere blocking simulation in current climate models: evaluating progress from the climate model intercomparison project phase 5 to 6 and sensitivity to resolution. *Weather and Climate Dynamics*, *1*(1), 277–292.

Schimmelmann, A., Zhao, M., Harvey, C. C., & Lange, C. B. (1998). A large california flood and correlative global climatic events 400 years ago. *Quaternary Research*, *49*(1), 51–61.

Schlef, K. E., Steinschneider, S., & Brown, C. M. (2018). Spatiotemporal impacts of climate and demand on water supply in the apalachicola-chattahoochee-flint basin. *Journal of Water Resources Planning and Management*, *144*(2), 05017020.

Schlumberger, J., Haasnoot, M., de Ruiter, M., & Aerts, J. (2022). Towards a disaster risk management pathways framework for complex and dynamic multi-risk: Dapp-mr. *Available at SSRN 4164233*.

Seager, R., Hoerling, M., Schubert, S., Wang, H., Lyon, B., Kumar, A., … Henderson, N. (2015). Causes of the 2011–14 california drought. *Journal of Climate*, *28*(18), 6997–7024.

Siler, N., Proistosescu, C., & Po-Chedley, S. (2019). Natural variability has slowed the decline in western us snowpack since the 1980s. *Geophysical Research Letters*, *46*(1), 346–355.

Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., … others (2019). Towards a more reliable historical reanalysis: Improvements for version 3 of the twentieth century reanalysis system. *Quarterly Journal of the Royal Meteorological Society*, *145*(724), 2876–2908.

Song, W., Zhao, B., Mu, C., Ballikaya, P., Cherubini, P., & Wang, X. (2022). Moisture availability influences the formation and characteristics of earlywood of pinus tabuliformis more than latewood in northern china. *Agricultural and Forest Meteorology*, *327*, 109219.

Soulé, P. T., Knapp, P. A., Maxwell, J. T., & Mitchell, T. J. (2021). A comparison of the climate response of longleaf pine (pinus palustris mill.) trees among standardized measures of earlywood, latewood, adjusted latewood, and total-wood radial growth. *Trees*, *35*, 1065–1074.

Stahle, D. W., Cook, E. R., Cleaveland, M. K., Therrell, M. D., Meko, D. M., Grissino-Mayer, H. D., … Luckman, B. H. (2000). Tree-ring data document 16th century megadrought over north america. *Eos, Transactions American Geophysical Union*, *81*(12), 121–125.

Stahle, D. W., Fye, F. K., Cook, E. R., & Griffin, R. D. (2007). Tree-ring reconstructed megadroughts over north america since ad 1300. *Climatic change*, *83*(1-2), 133.

Steinschneider, S., Ray, P., Rahat, S. H., & 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*(8), 6923–6945.

Sun, X., & Wang, G. (2022). Causes for the negative scaling of extreme precipitation at high temperatures. *Journal of Climate*, *35*(18), 6119–6134.

Sutanto, S. J., & Van Lanen, H. A. (2022). Catchment memory explains hydrological drought forecast performance. *Scientific Reports*, *12*(1), 2689.

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., … others (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in miroc6. *Geoscientific Model Development*, *12*(7), 2727–2765.

Taylor, M. (2017). Managing floods in california. *Legislative Analyst's Office: The California Legislature's Nonpartisan Fiscal and Policy Advisor*.

Toonen, W. H., Munoz, S. E., Cohen, K. M., & Macklin, M. G. (2020). High-resolution sedimentary paleoflood records in alluvial river environments: a review of recent methodological advances and application to flood hazard assessment. *Palaeohydrology: Traces, Tracks and Trails of Extreme Events*, 213–228.

Vicente-Serrano, S. M., López-Moreno, J. I., Beguería, S., Lorenzo-Lacruz, J., Azorin-Molina, C., & Morán-Tejeda, E. (2012). Accurate computation of a streamflow drought index. *Journal of Hydrologic Engineering*, *17*(2), 318–332.

Vidal, J.-P., Hingray, B., Magand, C., Sauquet, E., & Ducharne, A. (2015). Hierarchy of climate and hydrological uncertainties in transient low flow projections. *Hydrology & Earth System Sciences Discussions*, *12*(12).

Wasko, C., Lu, W. T., & Mehrotra, R. (2018). Relationship of extreme precipitation, dry-bulb temperature, and dew point temperature across australia. *Environmental Research Letters*, *13*(7), 074031.

Whateley, S., & Brown, C. (2016). Assessing the relative effects of emissions, climate means, and variability on large water supply systems. *Geophysical Research Letters*, *43*(21), 11–329.

Wi, S., & Steinschneider, S. (2022). Assessing the physical realism of deep learning hydrologic model projections under climate change. *Water Resources Research*, *58*(9), e2022WR032123.

Wi, S., Yang, Y., Steinschneider, S., Khalil, A., & Brown, C. (2015). Calibration approaches for distributed hydrologic models in poorly gaged basins: implication for streamflow projections under climate change. *Hydrology and Earth System Sciences*, *19*(2), 857–876.

Wilby, R., & Dessai, S. (2010). Robust adaptation to climate change.

Wilhelm, B., Canovas, J. A. B., Aznar, J. P. C., Kämpf, L., Swierczynski, T., Stoffel, M., … Toonen, W. (2018). Recent advances in paleoflood hydrology: From new archives to data compilation and analysis. *Water Security*, *3*, 1–8.

Williams, A., Anchukaitis, K., Woodhouse, C., Meko, D., Cook, B., Bolles, K., &

Cook, E. (2021). Tree rings and observations suggest no stable cycles in sierra nevada cool-season precipitation. *Water Resources Research*, *57*(3), e2020WR028599.

Williams, A. P., Cook, B. I., & Smerdon, J. E. (2022). Rapid intensification of the emerging southwestern north american megadrought in 2020–2021. *Nature Climate Change*, *12*(3), 232–234.

Williams, A. P., Cook, E. R., Smerdon, J. E., Cook, B. I., Abatzoglou, J. T., Bolles, K., ... Livneh, B. (2020). Large contribution from anthropogenic warming to an emerging north american megadrought. *Science*, *368*(6488), 314–318.

Yip, S., Ferro, C. A., Stephenson, D. B., & Hawkins, E. (2011). A simple, coherent framework for partitioning uncertainty in climate predictions. *Journal of climate*, *24*(17), 4634–4643.

Zeff, H. B., Hamilton, A. L., Malek, K., Herman, J. D., Cohen, J. S., Medellin-Azuara, J., ... Characklis, G. W. (2021). California's food-energy-water system: An open source simulation model of adaptive surface and groundwater management in the central valley. *Environmental Modelling & Software*, *141*, 105052.

Zhang, L., & Singh, V. (2019). *Copulas and their properties.* Copulas and Their Applications in Water Resources Engineering. Cambridge ....

Zhou, Y., & Kim, H.-M. (2018). Prediction of atmospheric rivers over the north pacific and its connection to enso in the north american multi-model ensemble (nmme). *Climate Dynamics*, *51*(5-6), 1623–1637.

Zhu, E., Yuan, X., & Wu, P. (2020). Skillful decadal prediction of droughts over large-scale river basins across the globe. *Geophysical Research Letters*, *47*(17), e2020GL089738.

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M., ... others (2020). A typology of compound weather and climate events. *Nature reviews earth & environment*, *1*(7), 333–347.

Appendix A: Glossary of Terms

- **Baseline weather scenario**: The 600-year daily precipitation and temperature scenario that is created by forcing the weather generator with paleo-reconstructed weather regimes. This scenario is comprised of 50 stochastic ensemble members.
- **Baseline streamflow scenario**: The 600-year daily streamflow scenario acquired by driving the hydrologic model with paleo-reconstructed weather (often referred to as 0T, 0CC). This scenario is comprised of 50 stochastic ensemble members.
- **Climate scenario**: A 600-year daily streamflow scenario created by forcing the hydrologic model with a baseline weather scenario that is layered with a set of thermodynamic climate changes.
- **Ensemble member**: Also referred to as a stochastic realization; each climate scenario is comprised of 50 stochastic ensemble members
- **Record length**: The total length of the dataset
  - **Paleo-informed weather and streamflow datasets**: 617 years (1400-2017 CE) at a daily time scale
  - **Observed Livneh climate (temperature and precipitation) dataset**: 63 years (1950-2013 CE) at a daily time scale
  - **Observed CDEC streamflow dataset**: 33 years (1986-2019) at a daily time scale
- **Time horizon**: also referred to as moving window; the length (in years) of the sliding window that passes over the total record length.

Appendix B: Metrics and Time Horizons

Metric	Description	Calculated	Justification	Citation
Flood Metric	10-Year Return Period Flow	GEV fit to 3-day maximum flow	Captures risk to smaller floodplains (or nuisance flooding in larger areas) and drives smaller investments.	Progress on Incorporating Climate Change into Planning and Management of California's Water Resources (July 2006)
Flood Metric	100-Year Return Period Flow	GEV fit to 3-day maximum flow	Drives larger riverine infrastructure development and flood risk management. Requires FEMA-mandated insurance.	Central Valley Flood Protection Plan Update 2022 (November 2022)
Drought Metrics	Occurrence, Severity, and Duration	Standardized streamflow-based indices	No state-wide definition. Historical droughts have been identified based on a combination of metrics such as reservoir depth and deficit magnitude and duration.	California's Most Significant Droughts: Comparing historical and recent conditions (February 2015)

Spatially Compounding Flood Metric	Likelihood of simultaneously exceeding historical 10-year and 100-year flow events in n basins	n-dimensional Gaussian copula	Flooding across the San Joaquin system could result in infrastructure failure such as levee breaks and disrupt deliveries of fresh water to 3 million acres of farmland.	Managing Floods in California (March 2017); Zscheischler et al. (2020)
Time Horizon	30-Year	N/A	CA prioritizes investment in flood management over a 30-year planning horizon	Central Valley Flood Protection Plan Update 2022 (November 2022)
Time Horizon	100-Year	N/A	Not actively used in planning and management, but can represent longer-term investments.	N/A

Supporting Information for
"Understanding Contributions of Paleo-Informed Natural Variability and Climate Changes on Hydroclimate Extremes in the Central Valley Region of California"

Rohini S. Gupta[1], **Scott Steinschneider**[2], **Patrick M. Reed**[1]

[1]School of Civil and Environmental Engineering, Cornell University, Ithaca, NY, USA
[2]Department of Biological and Environmental Engineering, Cornell University, Ithaca, NY, USA

Contents

Tables

Basin	Area (acre-ft)	Elevation (ft)	Slope
Tuolumne	3983	1795	12.6409
Millerton	4338	2156	13.0847
Merced	2784	1647	13.4917
New Hogan Lake	940	639	9.0975
New Melones	2331	1735	12.4246

Table S1. Physical attributes of the five basins as summarized in Wi and Steinschneider (2022).

Corresponding author: R.S. Gupta, `rg727@cornell.edu`

Text S1: Fitting the Paleo-Conditioned Non-Homogeneous Hidden Markov Model

We fit a Non-Homogeneous Hidden Markov Model (NHMM) in order to generate ensembles of plausible daily traces of weather regimes through the 600-year reconstruction. As compared to a Hidden Markov Model which has a stationary transition probability matrix, the NHMM has dynamic transition probability matrices that are conditioned on one or more external covariates that influence transitions between states. In this case, the covariates are the products of the reconstruction which are the first four principal components of weather regime occurrence. More information on the principal components can be found in Gupta et al. (2022). The NHMM is first fit over the instrumental period to the first nine principal components of daily, 500 hPa geopotential height from NOAA-CIRES-DOE Twentieth Century Reanalysis (V3) dataset (Slivinski et al., 2019) between 180-100°W and 30-60°N (i.e., the Pacific/North American sector) from 1950-2017 using the depmixS4 package in R. It is conditioned with the four reconstructed principal components (PC_{WR} from the Gupta et al. (2022) reconstruction) that overlap the same time period. The result from this fitting a time-varying transition probability matrix shown in Equation 1:

$$P(WR_t = i|WR_t = j, X_t = x) = \frac{\exp(\beta_{0j,i} + \beta'_{j,i})}{\sum_{k=1}^{K} \exp(\beta_{0j,i} + \beta'_{j,i})} \tag{1}$$

Here, the transition probability from WR i to WR j at time t is conditioned on $X'_t = [PC_{WR_{1,t}}, PC_{WR_{2,t}}, PC_{WR_{3,t}}, PC_{WR_{4,t}}]$ a vector of daily covariates developed by repeating the annual values of each for each day of the year. These covariates (Level 1 in Figure 3) are used within a multinomial logistic regression with intercepts $\beta_{0j,i}$ and coefficients $\beta_{j,i}$ to define the transition probabilities, with a prime denoting the vector transpose.

Table S2. NHMM Transition Matrix

	WR 1	WR 2	WR 3	WR 4	WR 5
WR 1	$P(1\|1) = \beta_{0,1\|1} + \sum_{k=1}^{4} \beta_{k,1\|1} * PC_{WR_k}$	$P(1\|2) = \beta_{0,1\|2} + \sum_{k=1}^{4} \beta_{k,1\|2} * PC_{WR_k}$	$P(1\|3) = \beta_{0,1\|3} + \sum_{k=1}^{4} \beta_{k,1\|3} * PC_{WR_k}$	$P(1\|4) = \beta_{0,1\|4} + \sum_{k=1}^{4} \beta_{k,1\|4} * PC_{WR_k}$	$P(1\|5) = \beta_{0,1\|5} + \sum_{k=1}^{4} \beta_{k,1\|5} * PC_{WR_k}$
WR 2	$P(2\|1) = \beta_{0,2\|1} + \sum_{k=1}^{4} \beta_{k,2\|1} * PC_{WR_k}$	$P(2\|2) = \beta_{0,2\|2} + \sum_{k=1}^{4} \beta_{k,2\|2} * PC_{WR_k}$	$P(2\|3) = \beta_{0,2\|3} + \sum_{k=1}^{4} \beta_{k,2\|3} * PC_{WR_k}$	$P(2\|4) = \beta_{0,2\|4} + \sum_{k=1}^{4} \beta_{k,2\|4} * PC_{WR_k}$	$P(2\|5) = \beta_{0,2\|5} + \sum_{k=1}^{4} \beta_{k,2\|5} * PC_{WR_k}$
WR 3	$P(3\|1) = \beta_{0,2\|1} + \sum_{k=1}^{4} \beta_{k,3\|1} * PC_{WR_k}$	$P(3\|2) = \beta_{0,3\|2} + \sum_{k=1}^{4} \beta_{k,3\|2} * PC_{WR_k}$	$P(3\|3) = \beta_{0,3\|3} + \sum_{k=1}^{4} \beta_{k,3\|3} * PC_{WR_k}$	$P(3\|4) = \beta_{0,3\|4} + \sum_{k=1}^{4} \beta_{k,3\|4} * PC_{WR_k}$	$P(3\|5) = \beta_{0,3\|5} + \sum_{k=1}^{4} \beta_{k,3\|5} * PC_{WR_k}$
WR 4	$P(4\|1) = \beta_{0,4\|1} + \sum_{k=1}^{4} \beta_{k,4\|1} * PC_{WR_k}$	$P(4\|2) = \beta_{0,4\|2} + \sum_{k=1}^{4} \beta_{k,4\|2} * PC_{WR_k}$	$P(4\|3) = \beta_{0,4\|3} + \sum_{k=1}^{4} \beta_{k,4\|3} * PC_{WR_k}$	$P(4\|4) = \beta_{0,4\|4} + \sum_{k=1}^{4} \beta_{k,4\|4} * PC_{WR_k}$	$P(4\|5) = \beta_{0,4\|5} + \sum_{k=1}^{4} \beta_{k,4\|5} * PC_{WR_k}$
WR 4	$P(5\|1) = \beta_{0,5\|1} + \sum_{k=1}^{4} \beta_{k,5\|1} * PC_{WR_k}$	$P(5\|2) = \beta_{0,5\|2} + \sum_{k=1}^{4} \beta_{k,5\|2} * PC_{WR_k}$	$P(5\|3) = \beta_{0,5\|3} + \sum_{k=1}^{4} \beta_{k,5\|3} * PC_{WR_k}$	$P(5\|4) = \beta_{0,5\|4} + \sum_{k=1}^{4} \beta_{k,5\|4} * PC_{WR_k}$	$P(5\|5) = \beta_{0,5\|5} + \sum_{k=1}^{4} \beta_{k,5\|5} * PC_{WR_k}$

Table S2 shows this symbolic transition matrix. The coefficients of the regression are fit during the instrumental period, and the matrix can vary depending on the value of the daily PC_{WR} from the reconstruction. Thus, a different transition matrix can be developed for each day. From this sequence of daily transition matrices, plausible sequences of daily WRs can be simulated across the entire 600-year period.

Figures

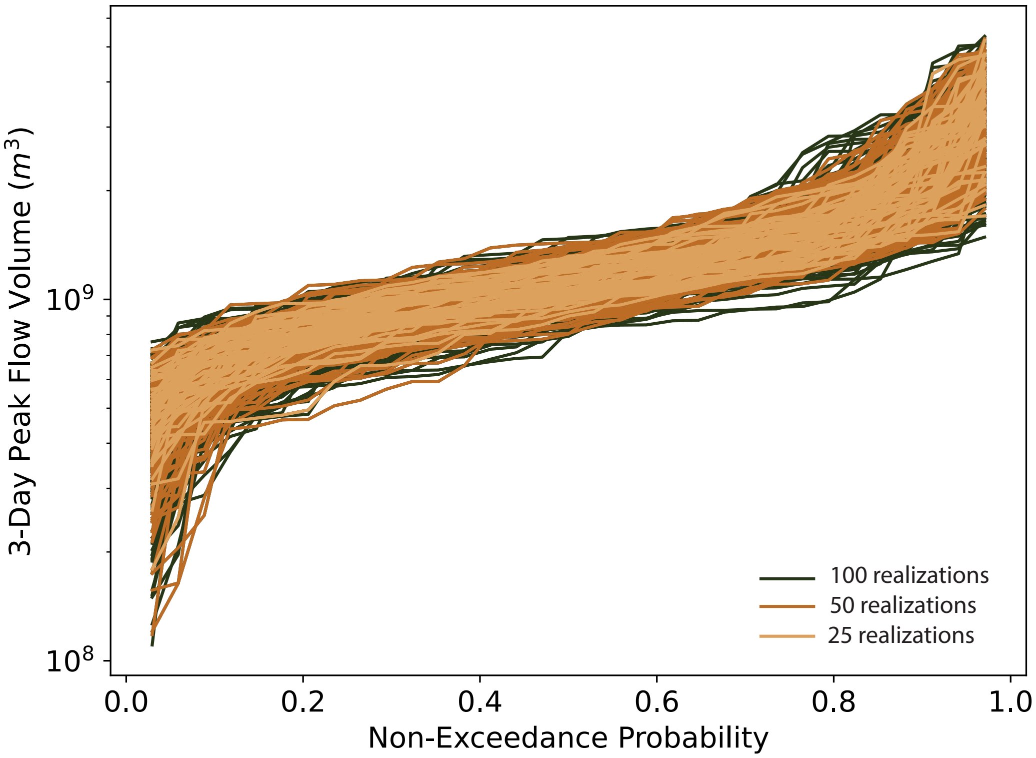

Figure S1. Flow duration curves of 3-day peak flow corresponding to different ensemble sizes. Each line corresponds to 30-year chunks of the 600-year record.

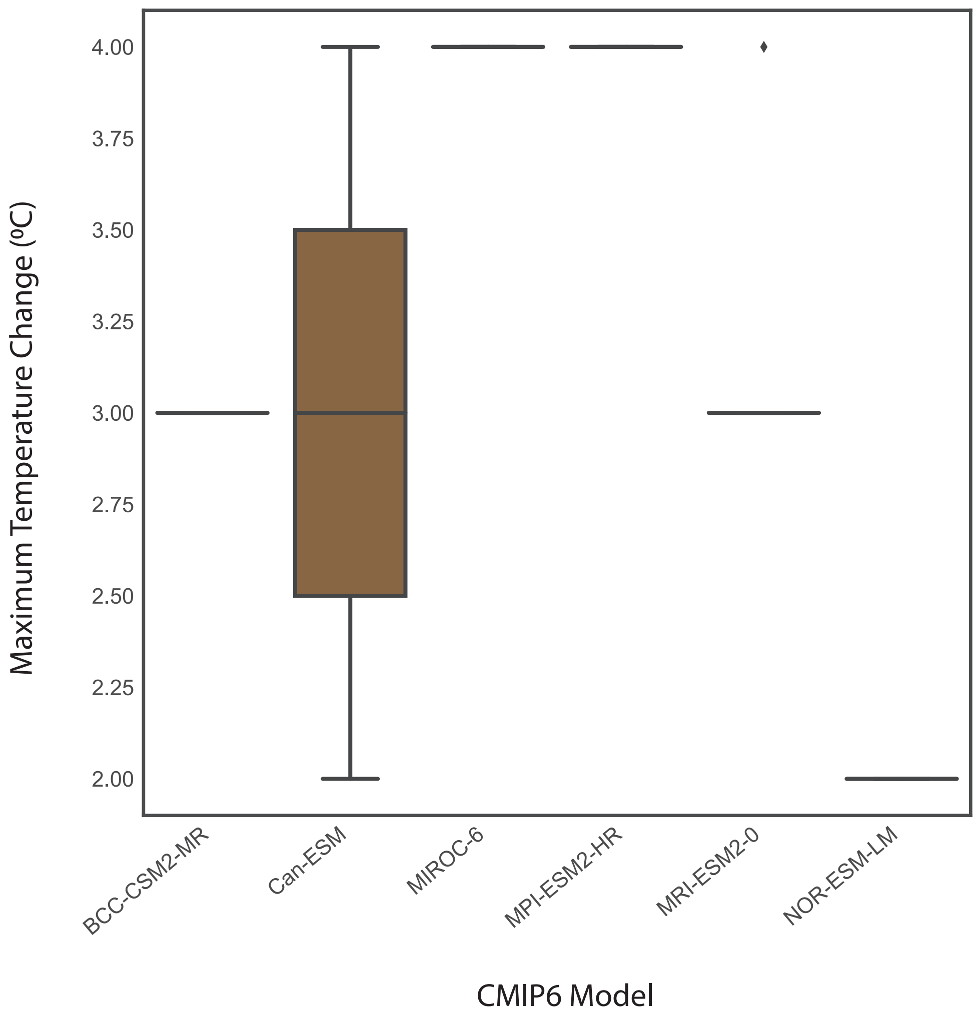

Figure S2. Maximum temperature change dictated by the subset of CMIP6 models downscaled by CarbonPlan (Chegwidden et al., 2022) and for the length of the projection from 2015-2050. All models are run under one initial condition and under multiple downscaling methods.

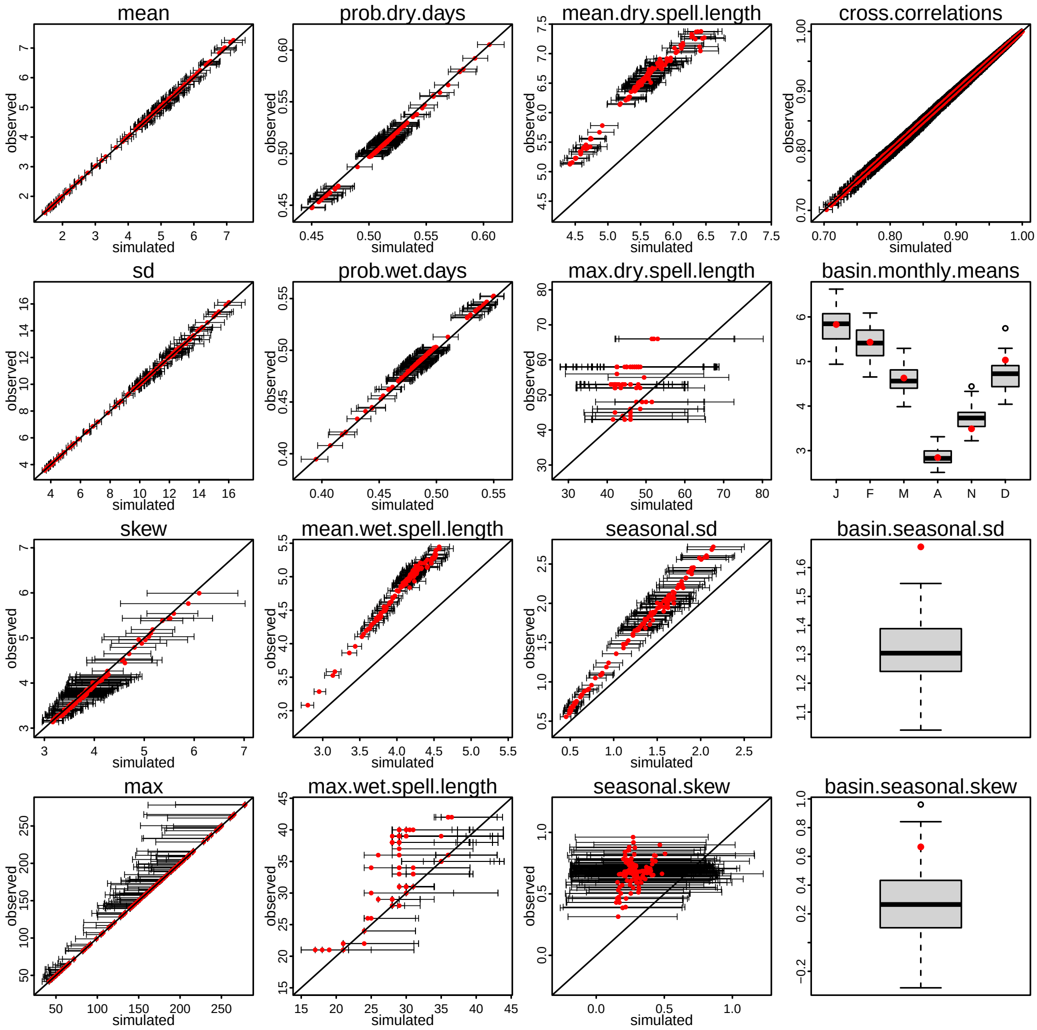

Figure S3. Observed vs. simulated characteristics of daily precipitation in the Tuolumne Basin. For at-site characteristics (180 grid cells), the 95% range for simulated statistics across the 50 ensemble members is shown with whiskers. For basin-averaged statistics, the distribution of simulated statistics is shown as a boxplot along with the observed value.

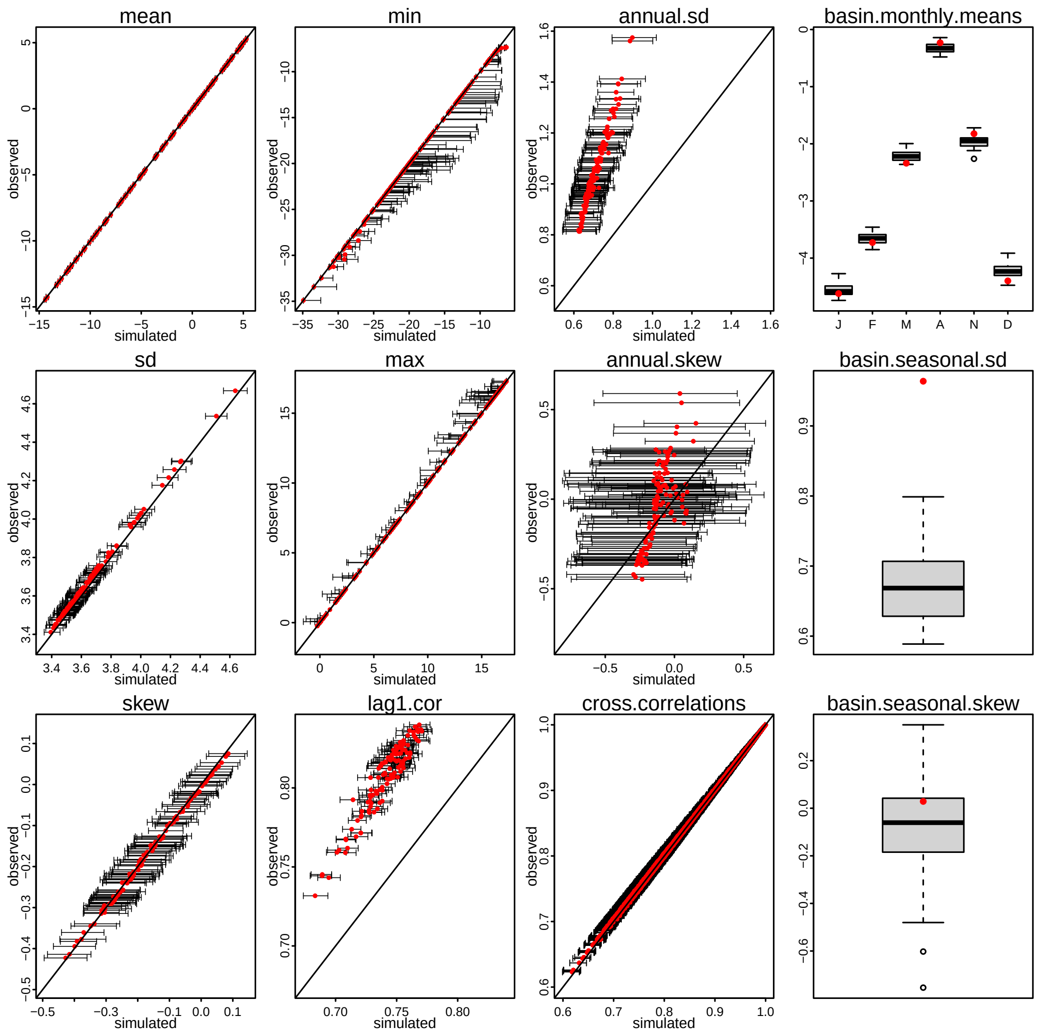

Figure S4. Same as Figure S3 but for observed vs. simulated characteristics of minimum daily temperature in the Tuolumne Basin.

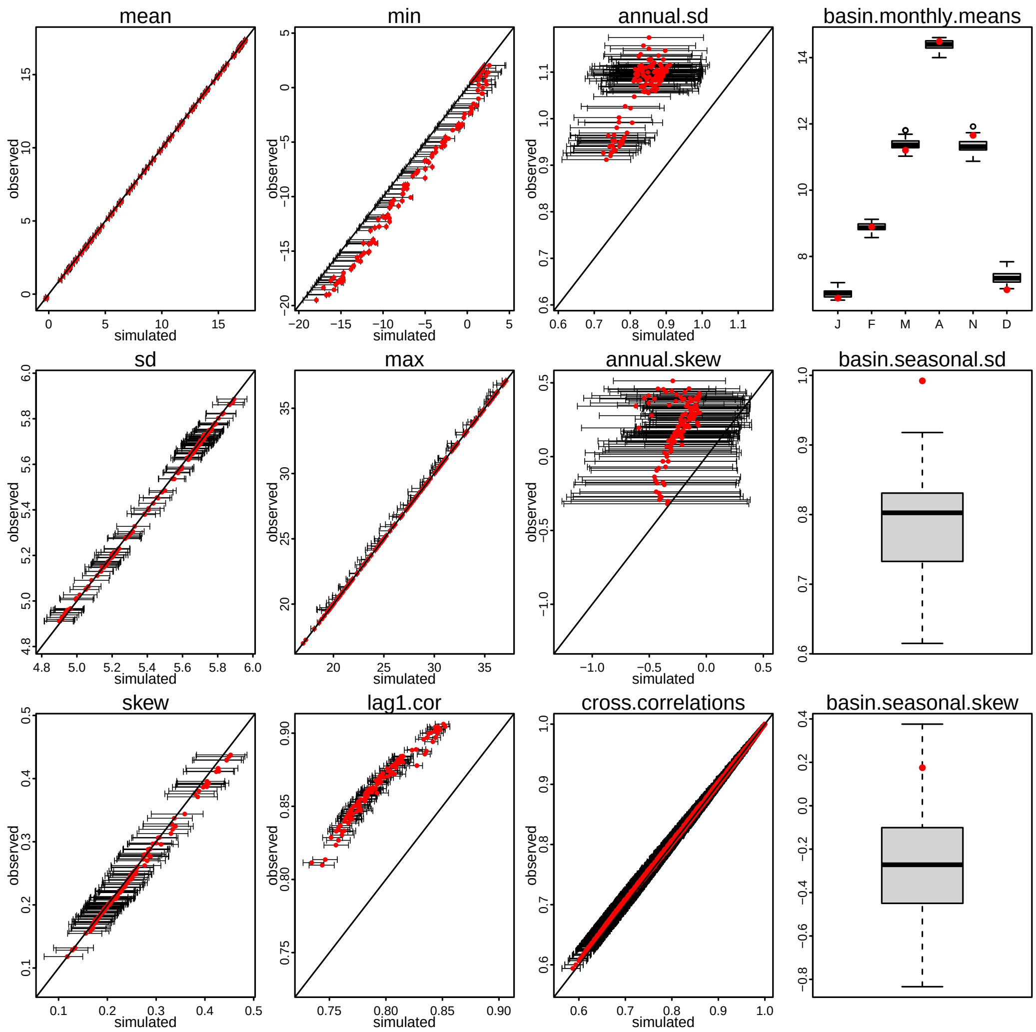

Figure S5. Same as Figure S3 but for observed vs. simulated characteristics of maximum daily temperature in the Tuolumne Basin.

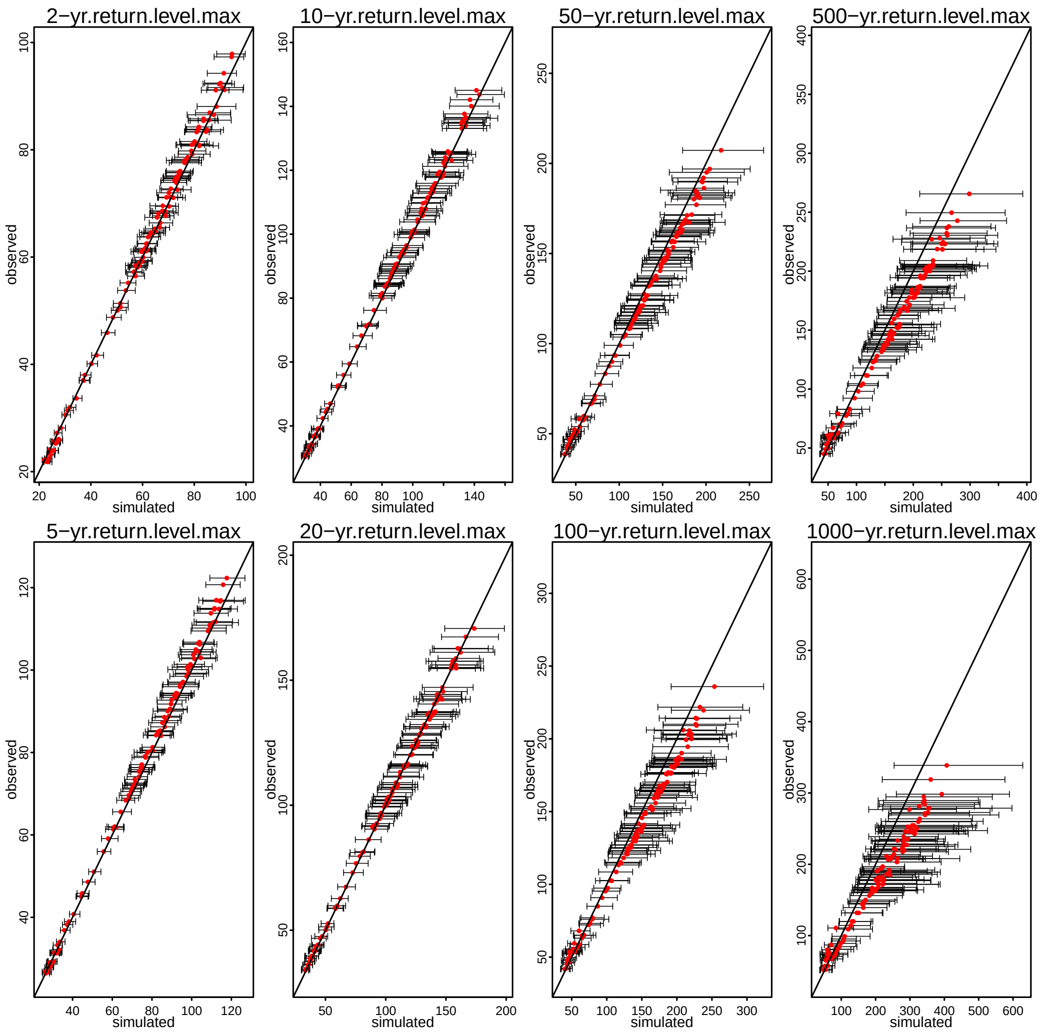

Figure S6. Same as Figure S3 but for observed vs. simulated characteristics of flooding return levels.

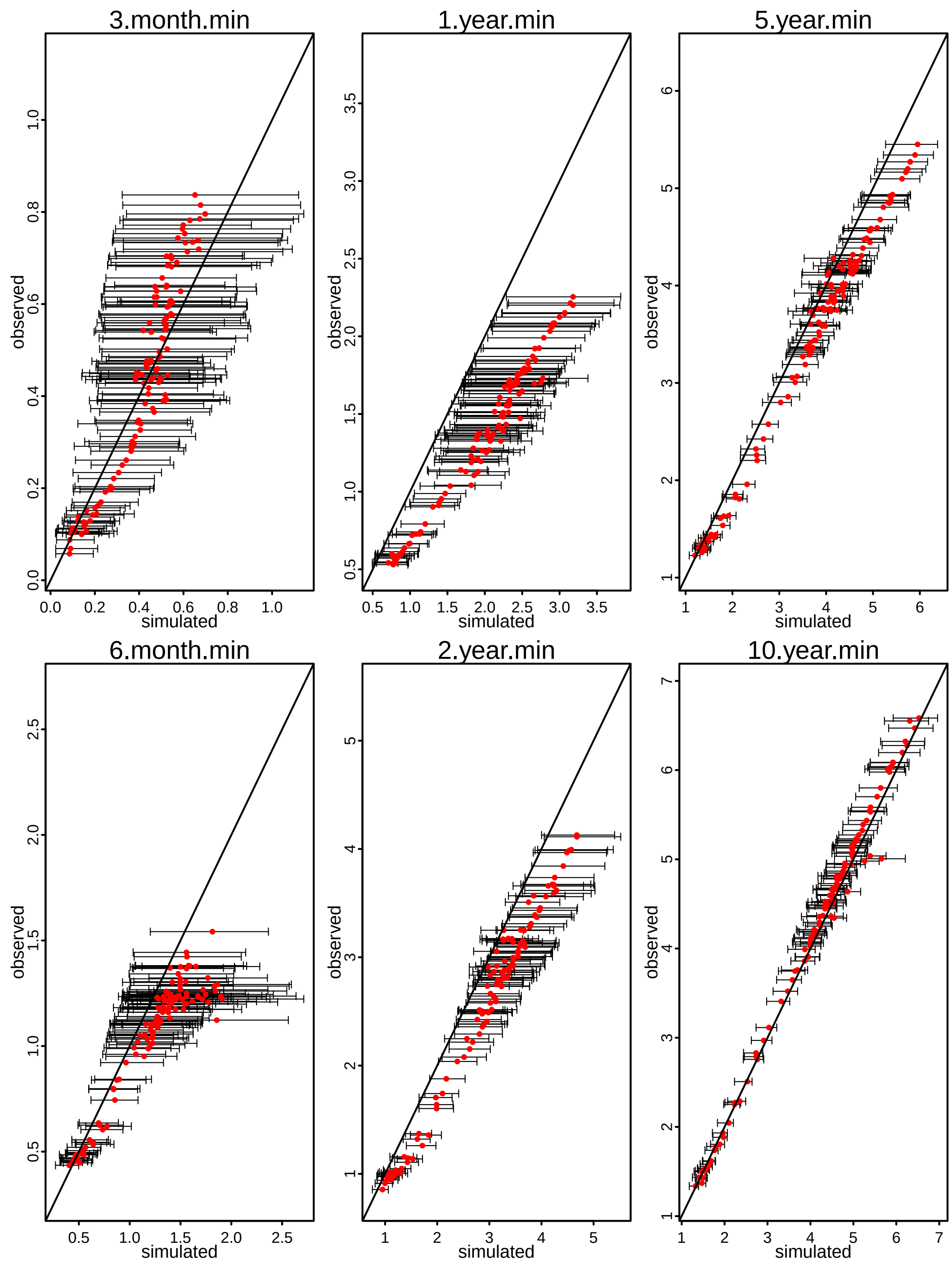

Figure S7. Same as Figure S1 but for observed vs. simulated characteristics of accumulated minimum precipitation totals across varying time periods.

Text S2: Distinctions Across Hydrologic Models

Streamflow ensembles are developed using two hydrologic models: SAC-SMA and HYMOD. Figure S8 shows exceedance plots of 3-day flows associated with each model with a focus on the Don Pedro gauge in the Tuolumne Basin. The flow duration curves exhibit strong differences. The models capture similar peak flow dynamics, but have strong lower-tail distinctions. Particularly, SAC-SMA is capable of producing drier three-day flows. We further conduct a small variance decomposition experiment using SAC-SMA and HYMOD in the Tuolumne Basin. Figures S21 and S22 show the results of the decomposition when choice of hydrologic model is an additional uncertain factor for the flood and drought metrics respectively. Overall, we demonstrate that the choice of model does not impact key drivers of the metrics of interest. Wi and Steinschneider (2022) also demonstrate better out-of-sample performance of SAC-SMA in all five basins over the observed record. Due to these reasons, we opt to continue the study with a single model and choose SAC-SMA.

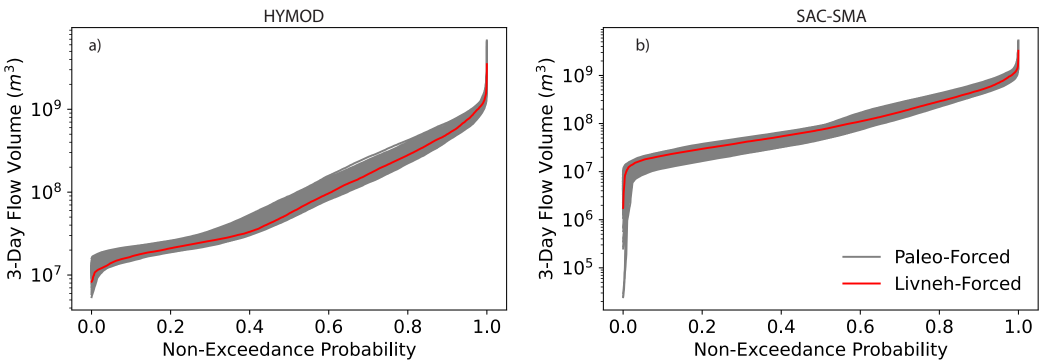

Figure S8. Non-exceedance plots of 3-day flow volumes associated with each model for the Tuolumne Basin. Each grey line represents sorted volumes for each year in 30-year chunks of the paleo-reconstruction across all 50 ensemble members. The red line corresponds to sorted volumes for the 30-year modern record as derived from forcing SAC-SMA-with historic Livneh data from 1987-2013.

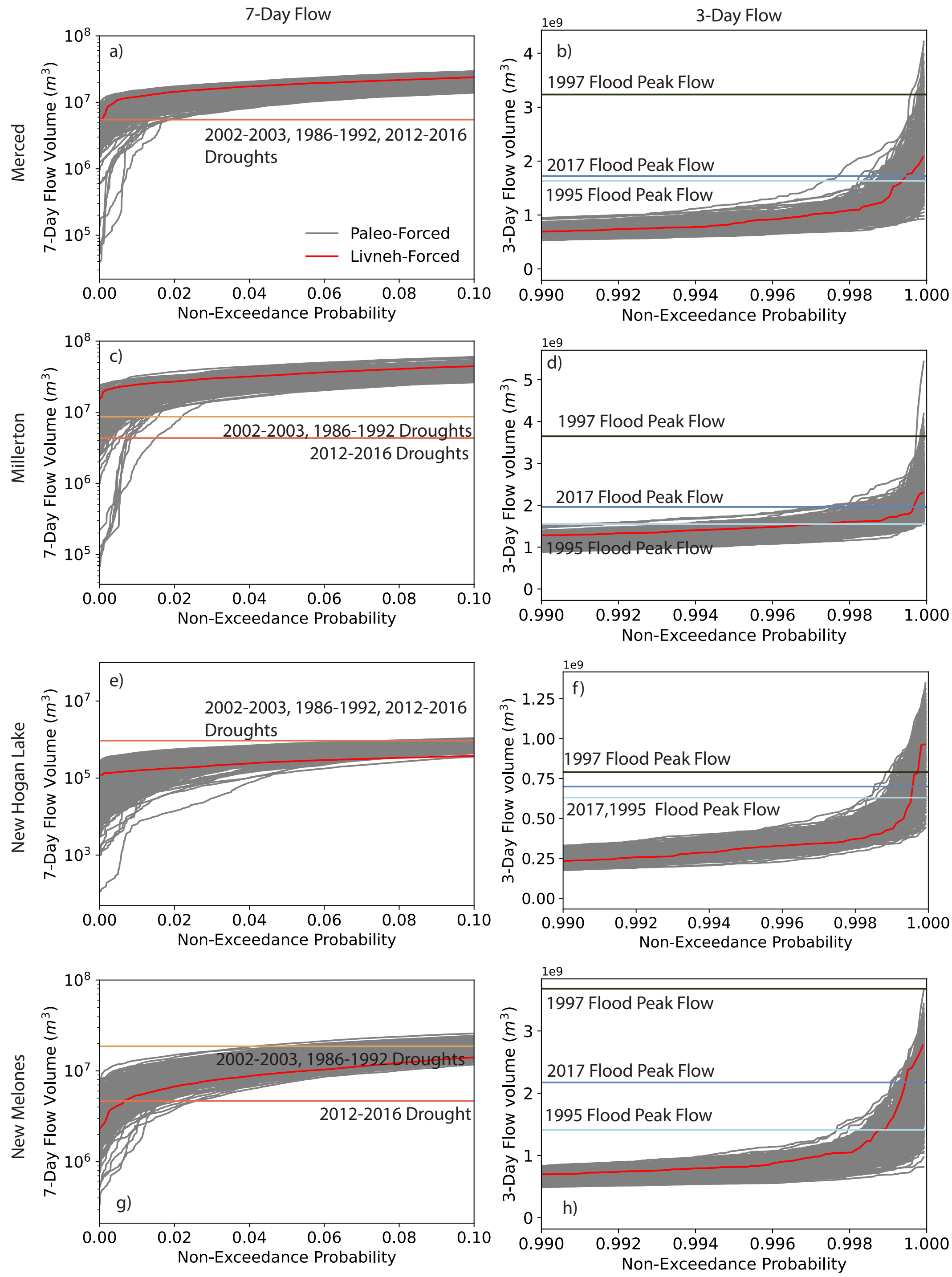

Figure S9. 7-day and 3-day flow volumes at gauged locations across the five basins derived from the paleo-informed streamflow ensembles compared to the Livneh-forced generator over the modern period. Key events from the observed record are shown as colored lines. Each grey line represents sorted volumes for each year in 30-year chunks of the paleo-reconstruction across all 50 ensemble members.

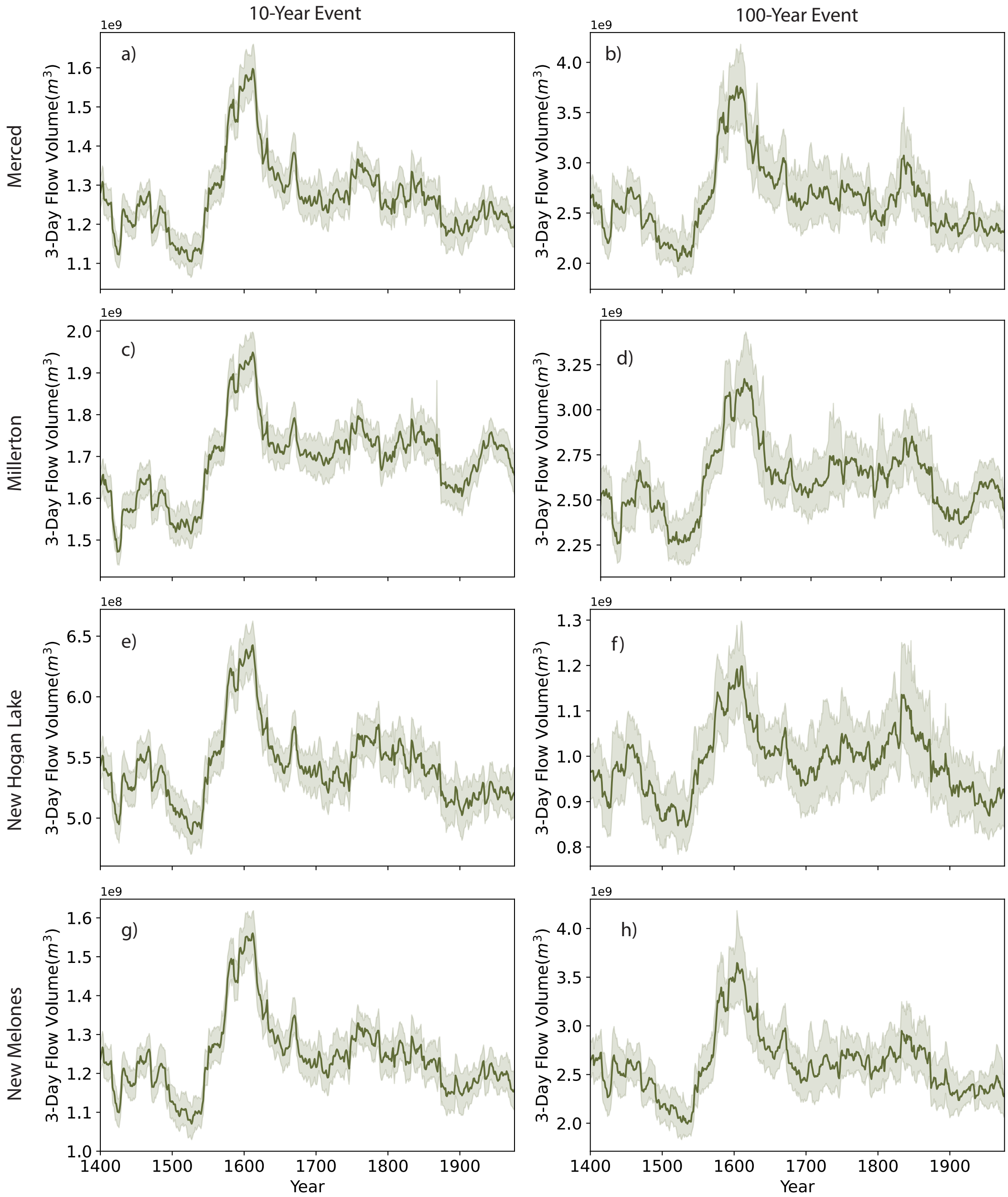

Figure S10. Three-day annual maxima associated with the 10-year return period event and 100-year return period event for remaining four gauged locations. calculated in 30-year moving windows and across the time period from 1400-2017. The dark green line represents the mean flooding return levels and the shading represents the 5$^{\text{th}}$ and 95$^{\text{th}}$ percentile bounds.

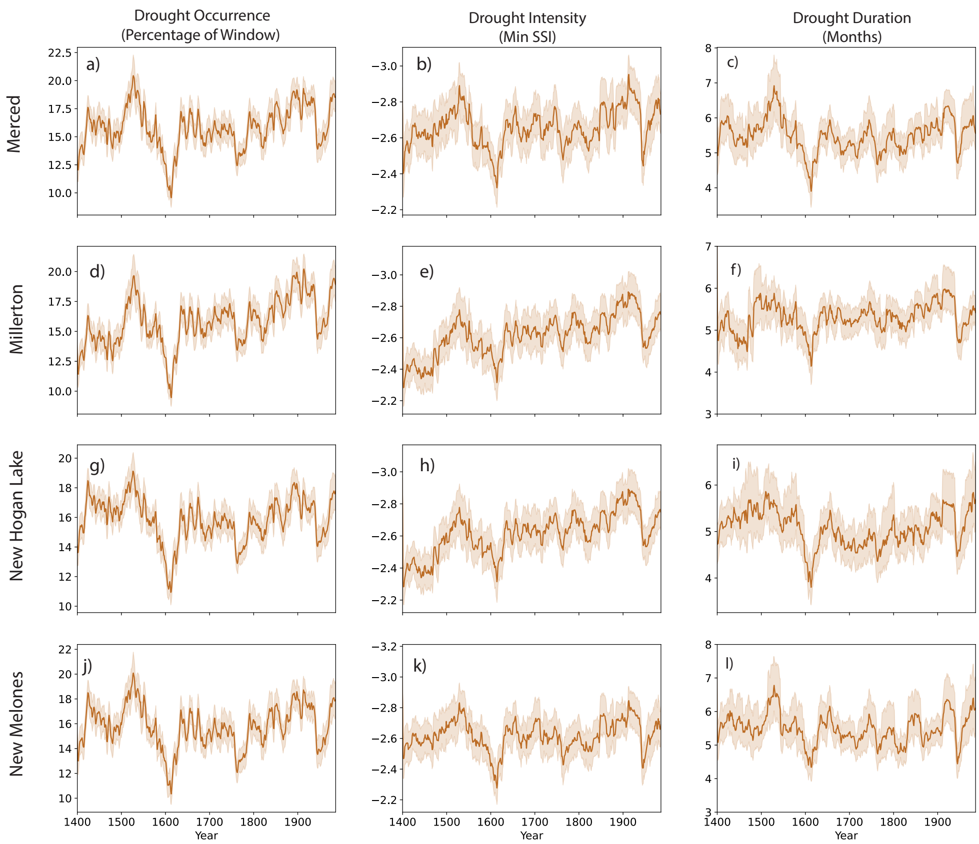

Figure S11. SSI-based hydrologic drought metrics for the remaining four locations calculated in 30-year moving windows and across the time period from 1400-2017. The dark tan line represents the mean drought metric value and the shading represents the 5$^{\text{th}}$ and 95$^{\text{th}}$ percentile bounds.

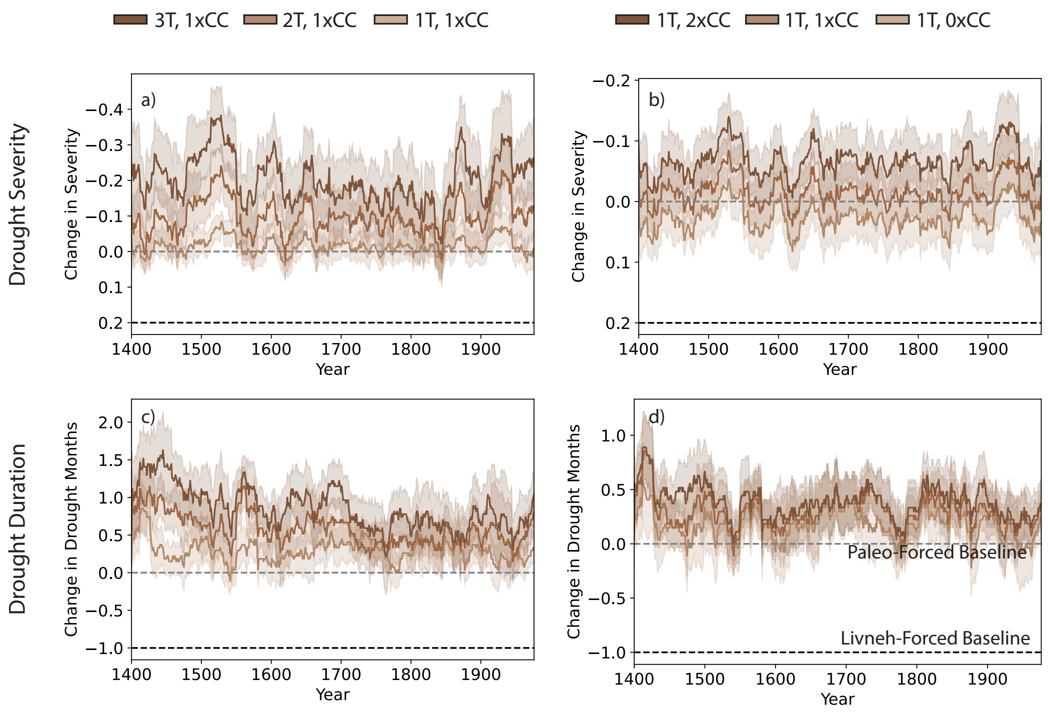

Figure S12. The effect of increasing a,c) temperature and b,d) precipitation scaling rates on drought severity a,b) and drought duration c,d) at Don Pedro (Tuolumne Basin). The dark brown lines represent the increase in the percentage of the 30-year window classified in drought conditions with respect to the baseline scenario (gray dashed line at 0) and the shading represents the 5$^{\text{th}}$ and 95$^{\text{th}}$ percentile bounds.. A modern baseline (black line) is included as reference and represents the distance from the modern metric to the worst drought duration or severity recorded in the reconstruction.

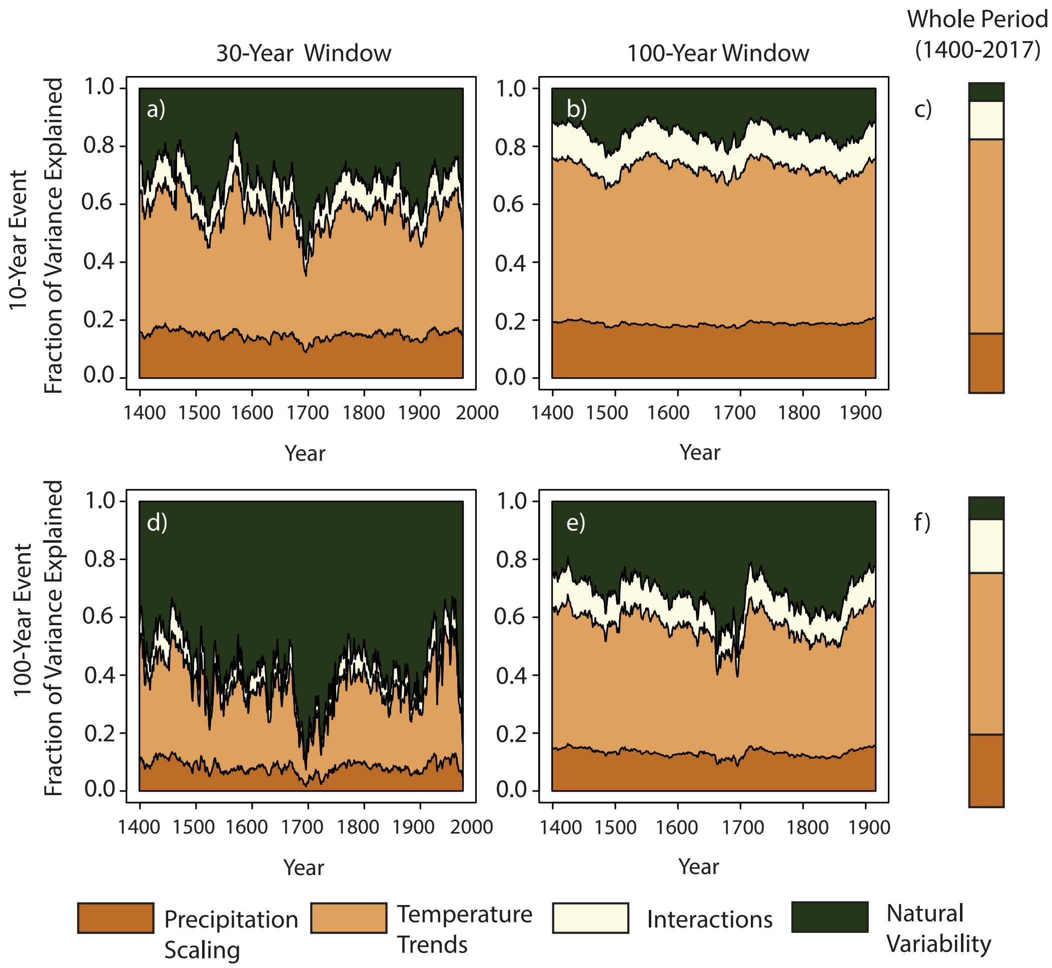

Figure S13. A decomposition of the key drivers of variance in the flood metrics for the Merced Falls gauge in the Merced River Basin for an a,d) 30-year window b,e) 100-year window and c,f) a 600-year window.

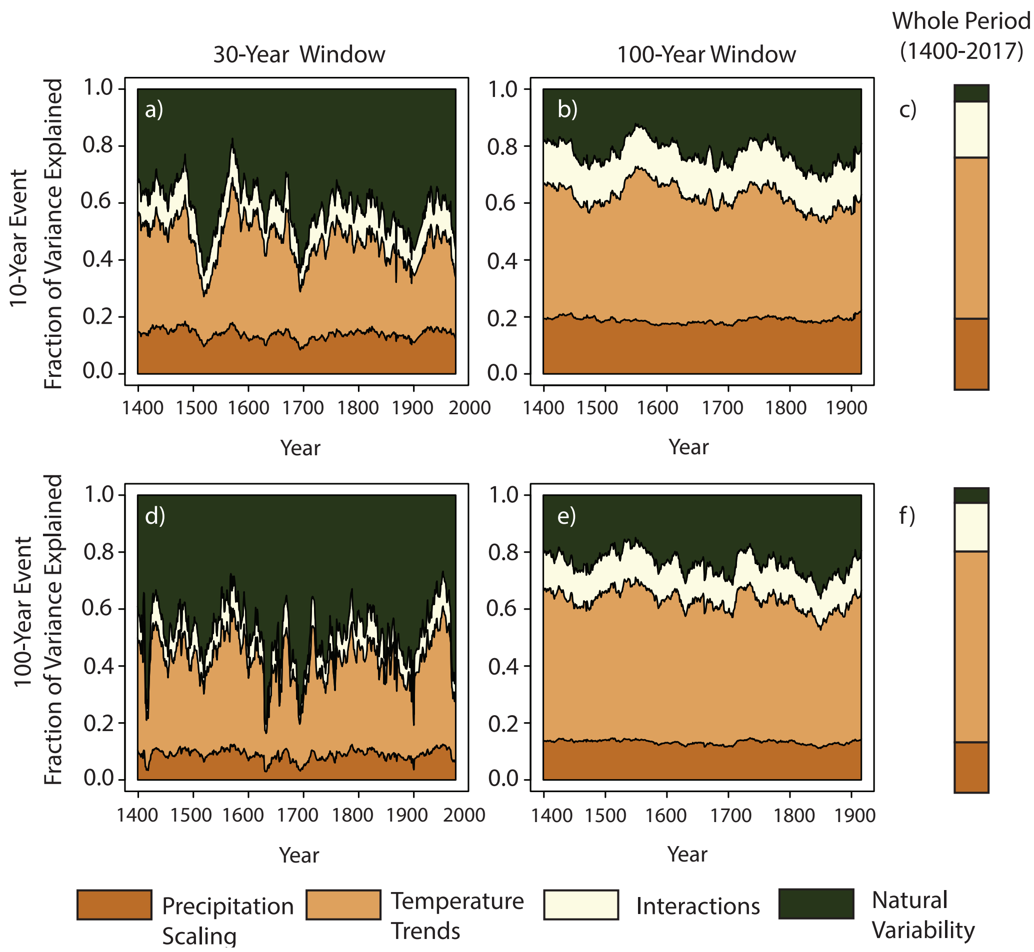

Figure S14. A decomposition of the key drivers of variance in the flood metrics for the Millerton Lake gauge in the San Joaquin Basin for an a,d) 30-year window b,e) 100-year window and c,f) a 600-year window.

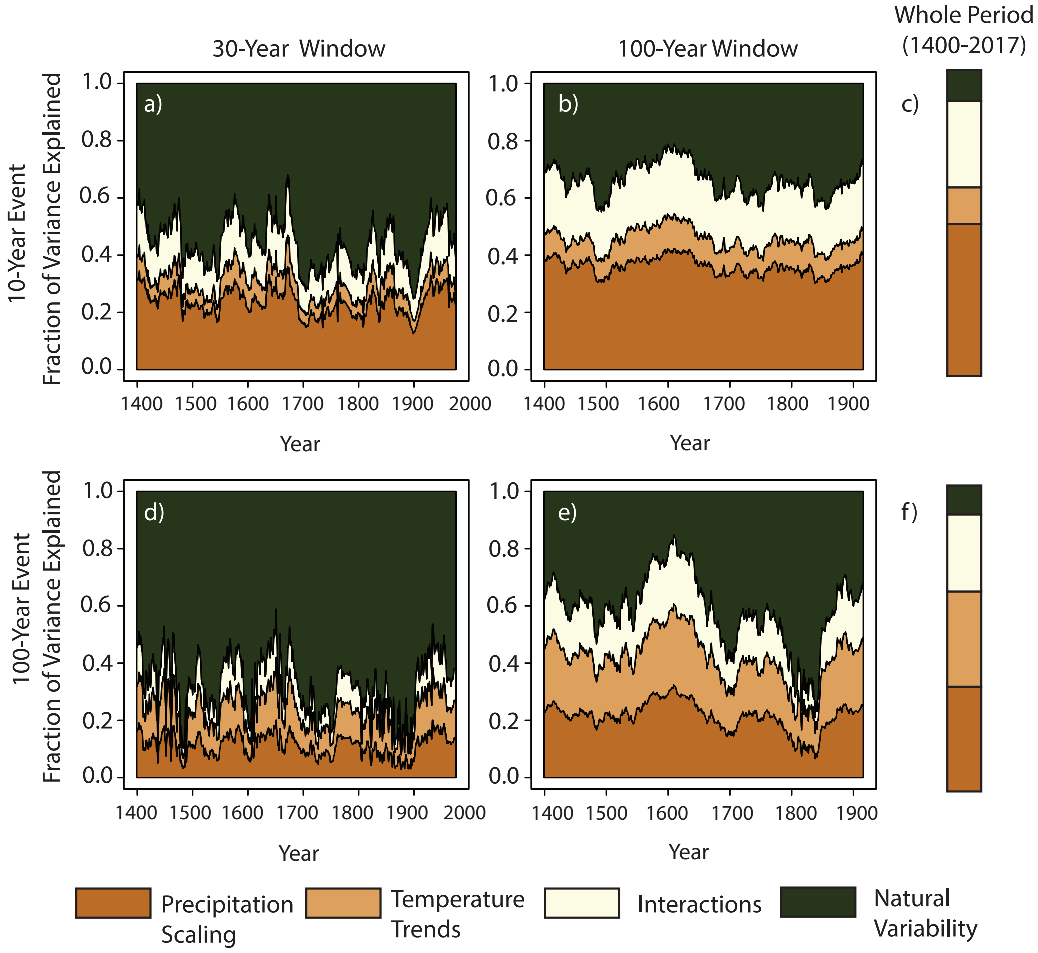

Figure S15. A decomposition of the key drivers of variance in the flood metrics for the New Hogan Lake gauge in the Calaveras River Basin for an a,d) 30-year window b,e) 100-year window and c,f) a 600-year window.

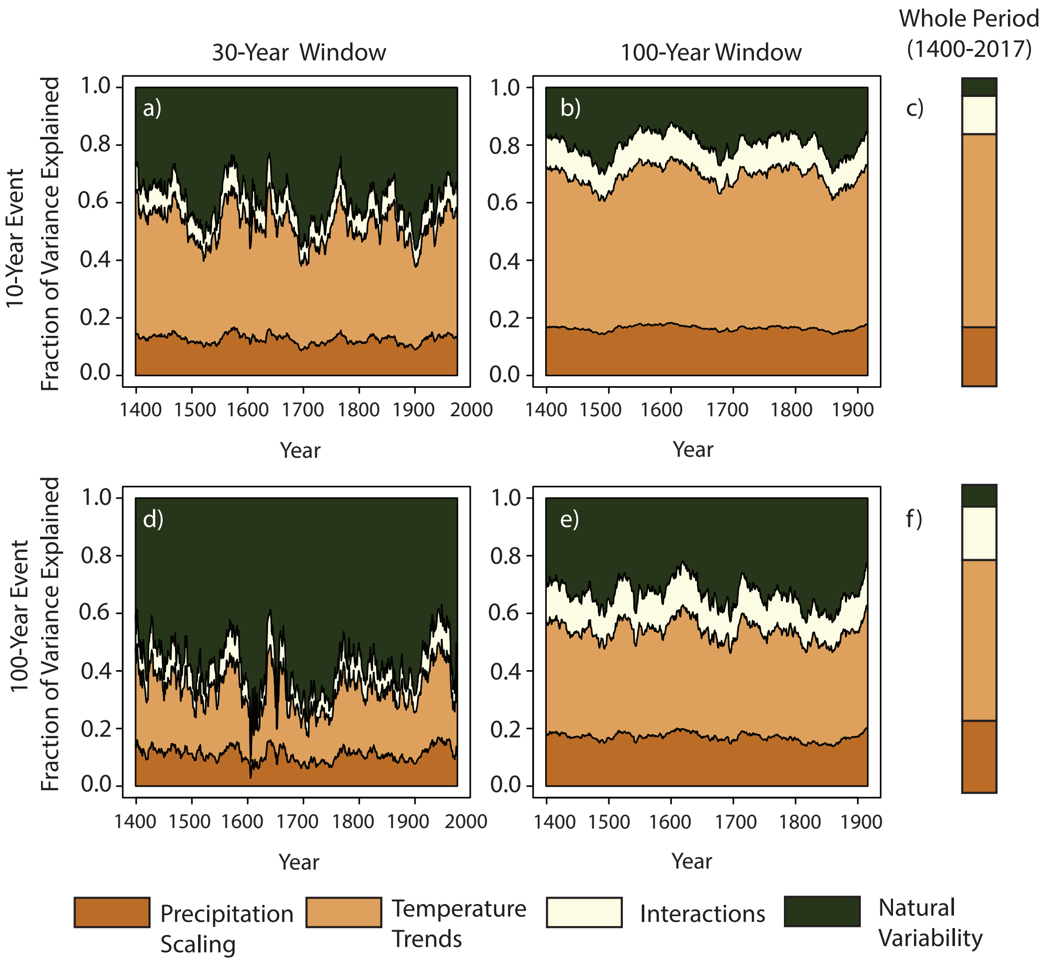

Figure S16. A decomposition of the key drivers of variance in the flood metrics for the New Melones Lake gauge in the Stanislaus River Basin for an a,d) 30-year window b,e) 100-year window and c,f) a 600-year window.

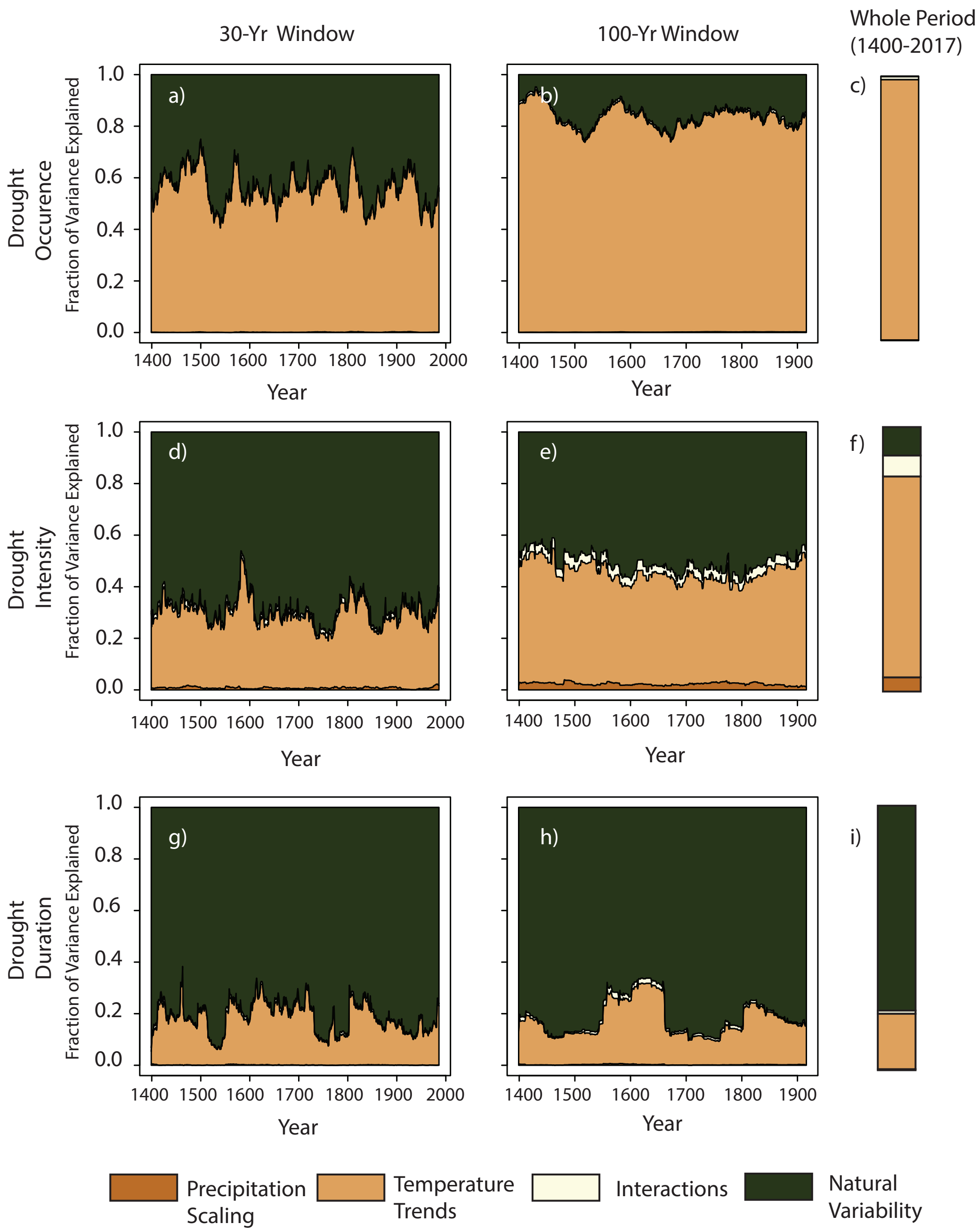

Figure S17. A decomposition of the key drivers of variance in the drought metrics for the Merced Falls gauge in the Merced River Basin for an a,d) 30-year window b,e) 100-year window and c,f) a 600-year window.

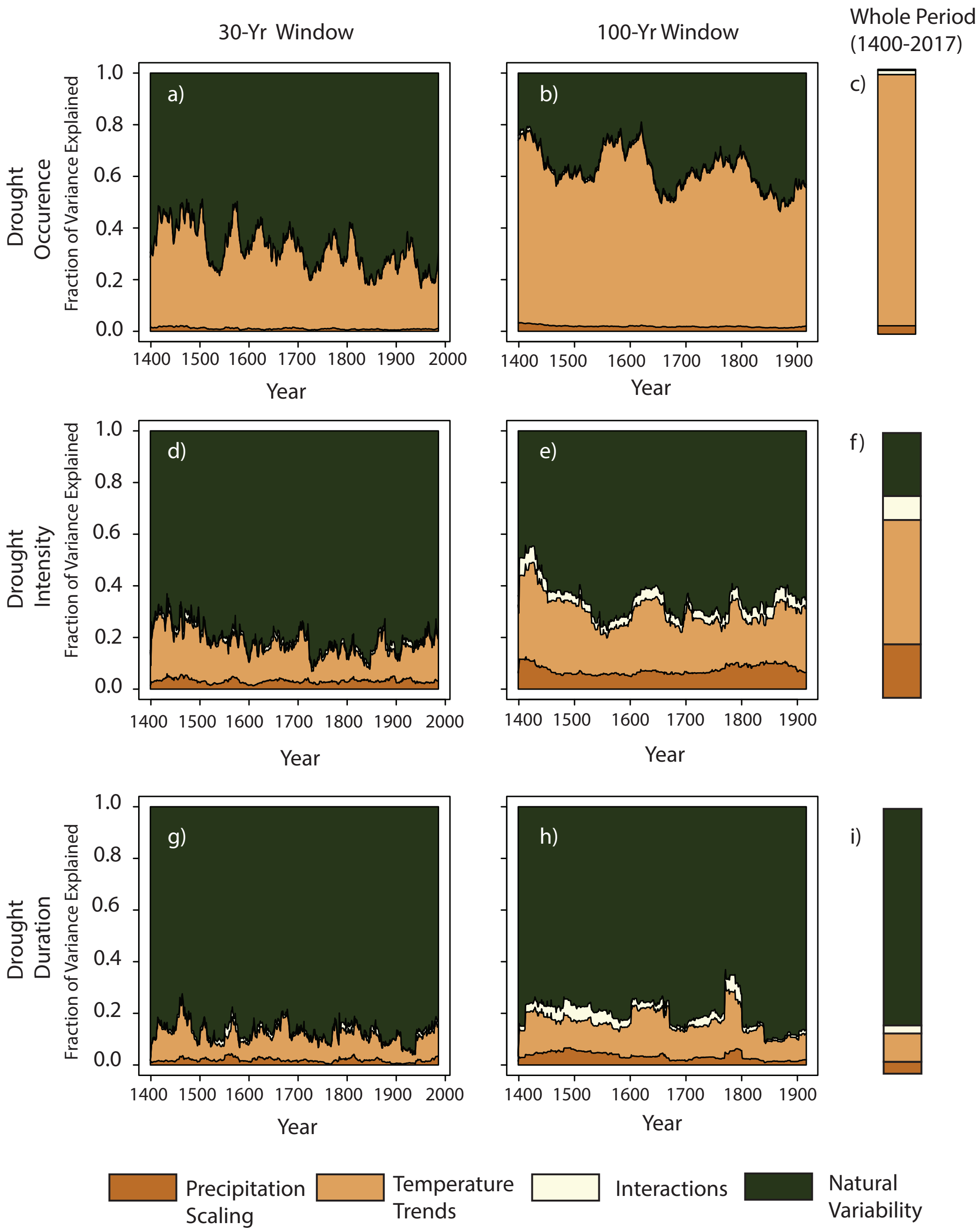

Figure S18. A decomposition of the key drivers of variance in the drought metrics for the Millerton Lake gauge in the San Joaquin Basin for an a,d) 30-year window b,e) 100-year window and c,f) a 600-year window.

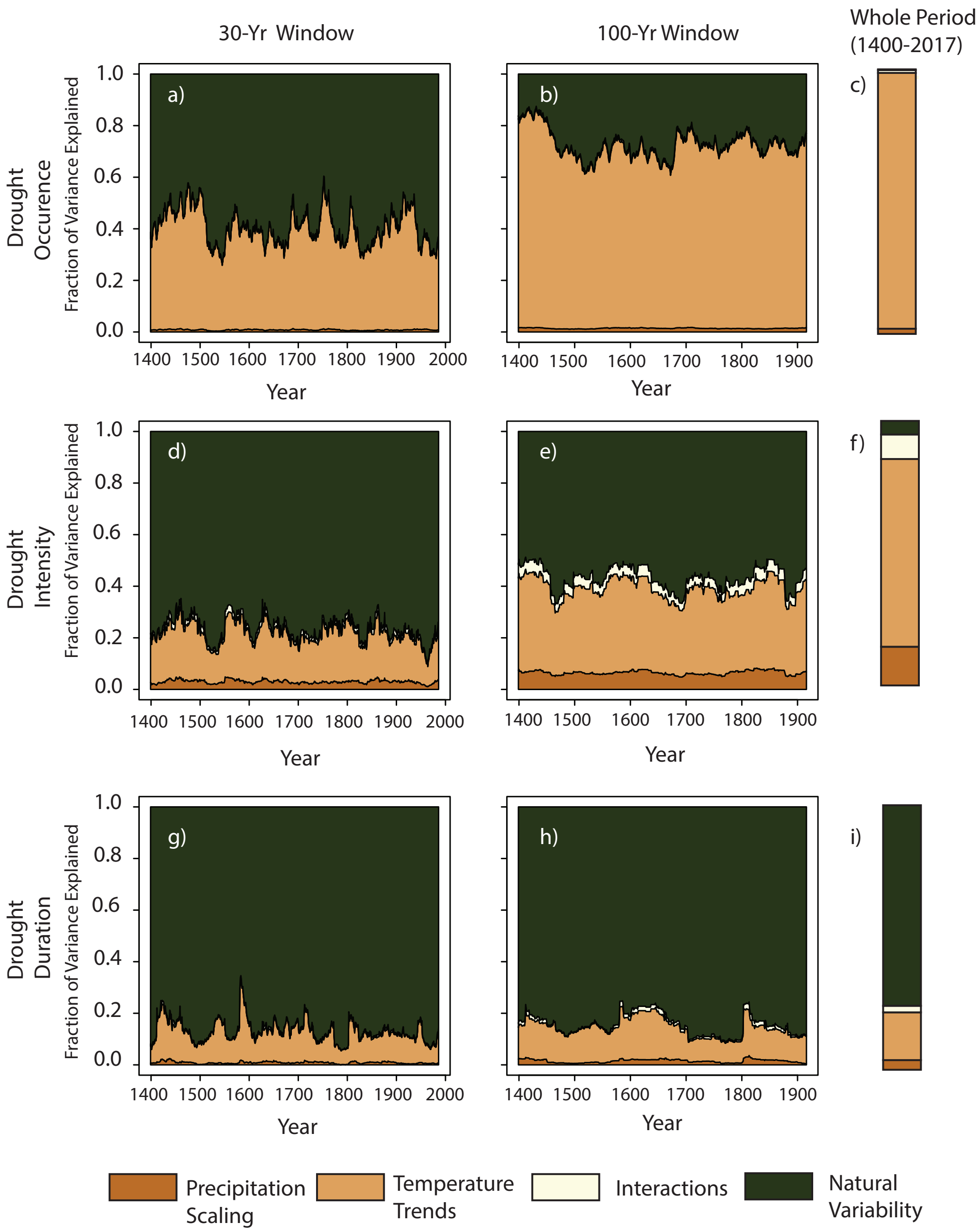

Figure S19. A decomposition of the key drivers of variance in the drought metrics for the New Hogan Lake gauge in the Calaveras River Basin for a,d,g) 30-year window b,e,h) 100-year window and c,f,i) a 600-year window.

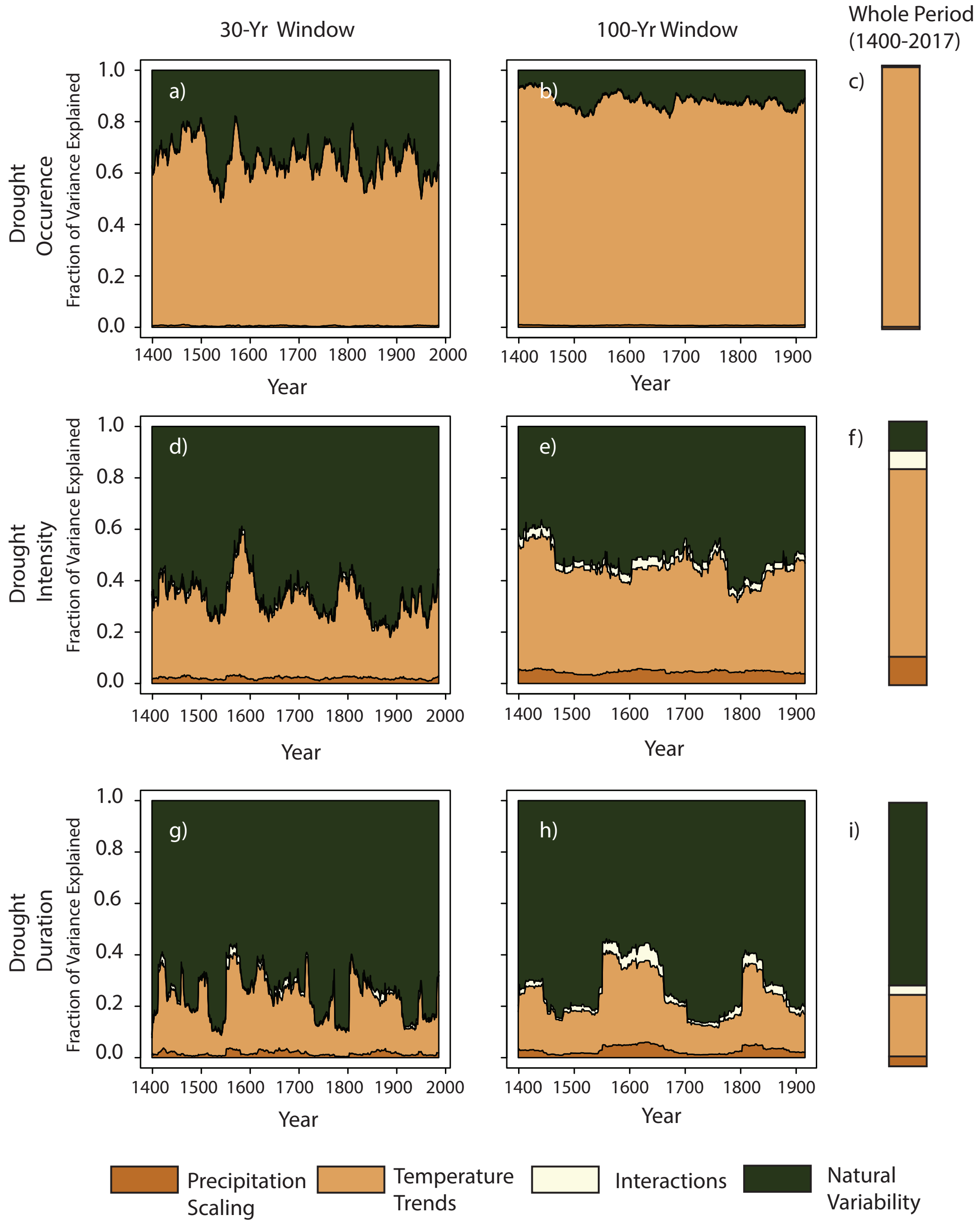

Figure S20. A decomposition of the key drivers of variance in the drought metrics for the New Melones Lake gauge in the Stanislaus River Basin for a,d,g) 30-year window b,e,h) 100-year window and c,f,i) a 600-year window

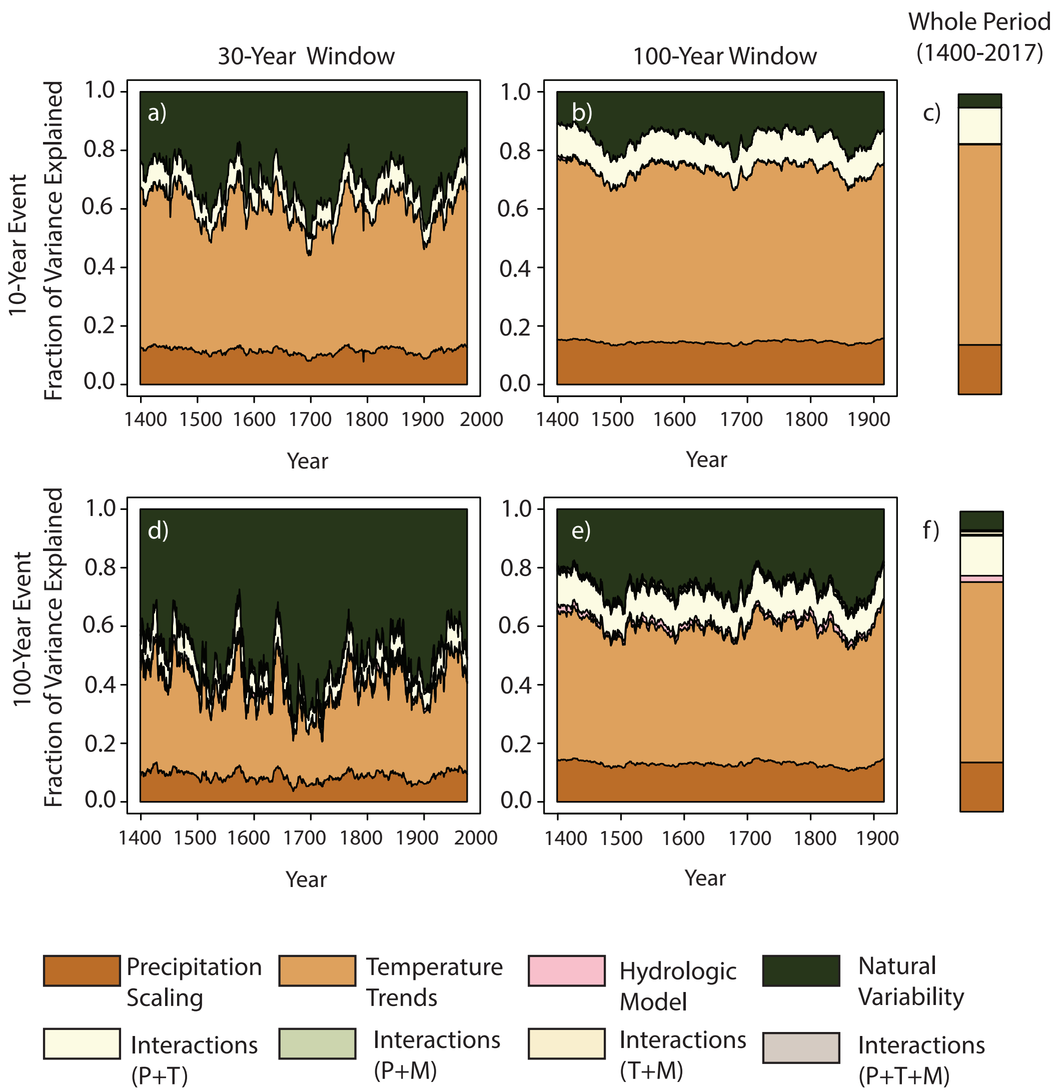

Figure S21. A decomposition of the key drivers of variance in flood metrics for the Tuolumne Basin, with the additional factor of hydrologic model choice.

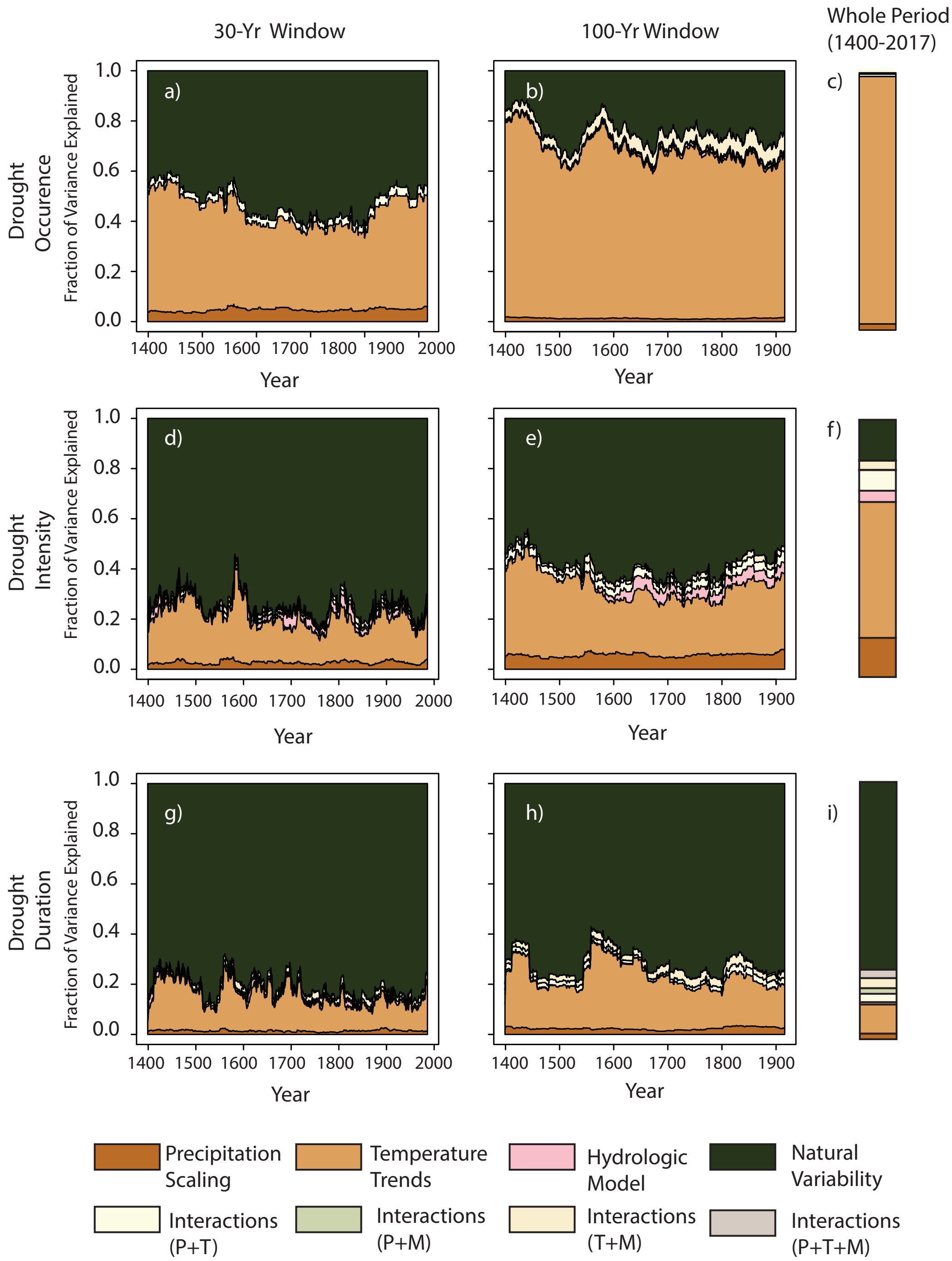

Figure S22. A decomposition of the key drivers of variance in drought metrics for the Tuolumne Basin, with the additional factor of hydrologic model choice.

References

Chegwidden, O. S., Hagen, R., Martin, K., Jones, M., Banihirwe, A., Chiao, C., ... Hamman, J. (2022, October). *Downscaling CMIP6 with multiple statistical methods.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.7145491` (The development of this project was funded, in part, through a grant from the Microsoft AI for Earth program to CarbonPlan.) doi: 10.5281/zenodo.7145491

Gupta, R. S., Steinschneider, S., & Reed, P. M. (2022). A multi-objective paleo-informed reconstruction of western us weather regimes over the past 600 years. *Climate Dynamics*, 1–20.

Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., ... others (2019). Towards a more reliable historical reanalysis: Improvements for version 3 of the twentieth century reanalysis system. *Quarterly Journal of the Royal Meteorological Society*, *145*(724), 2876–2908.

Wi, S., & Steinschneider, S. (2022). Assessing the physical realism of deep learning hydrologic model projections under climate change. *Water Resources Research*, *58*(9), e2022WR032123.