

UCLA

UCLA Previously Published Works

Title

Bias Correction Can Strongly Shape California's Projected Megaflood Risk

Permalink

<https://escholarship.org/uc/item/1729h5nz>

Journal

Journal of Geophysical Research: Atmospheres, 131(16)

ISSN

2169-897X

Authors

Rahimi, Stefan

Huang, Lei

Thackeray, Chad W

et al.

Publication Date

2026-08-28

DOI

10.1029/2026jd046503

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Bias correction choices strongly shape California’s projected megaflood risk

Stefan R. Rahimi-Esfarjani¹, LEI HUANG², Chad W. Thackeray³, Mark Risser⁴, Emilie Gaelle Tarouilly³, Jesse Norris⁵, Alan M. Rhoades⁴, Benjamin Bass³, Alex Hall³, Zachary J Lebo¹, Weichen Liu¹, Flavio Lehner⁶, Tim Corrie¹, and Ankur Dixit⁶

¹University of Wyoming

²Unknown

³University of California Los Angeles

⁴Lawrence Berkeley National Laboratory

⁵UCLA

⁶Cornell University

February 01, 2026

Abstract

A California “megaflood” - a weeks-long cluster of storms with centuries-long return times - would be catastrophic, yet its timing and magnitude under climate change remains uncertain. A central source of uncertainty is whether Earth System Model boundary conditions should be bias corrected before dynamical downscaling. Here, we analyze a large ensemble of downscaled simulations to quantify how this uncertainty shapes California’s megaflood risk. Bias correction is necessary to reproduce observed storm statistics and strongly influences both the megafloods historical and future characteristics. Without correction, storm precipitation variability is inappropriately inflated, delaying the emergence of extreme precipitation signals by decades. Further, megaflood intensity estimates based on statistically-extended observations are at least 10% drier than explicit computations derived from sufficiently-sized storm populations. Our results show that bias correction choices crucially govern projections of rare, high-impact events, with direct implications for infrastructure design/planning and water management in California.

Hosted file

MegaFloodJGRA.docx available at <https://authorea.com/users/742588/articles/1383049-bias-correction-choices-strongly-shape-california-s-projected-megaflood-risk>

Hosted file

SI.docx available at <https://authorea.com/users/742588/articles/1383049-bias-correction-choices-strongly-shape-california-s-projected-megaflood-risk>

Bias correction choices strongly shape California's projected megaflood risk

Stefan Rahimi^{1,2}, Lei Huang², Chad W Thackeray², Mark D Risser³, Emilie Tarouilly⁴, Jesse Norris², Alan Rhoades³, Benjamin Bass², Alex Hall², Zachary J. Lebo⁵, Weichen Liu¹, Flavio Lehner^{6,7}, Tim Corrie¹, and Ankur Dixit⁶

¹University of Wyoming, Department of Atmospheric Science and the School of Computing

²University of California Los Angeles, Department of Atmospheric and Oceanic Sciences

³Lawrence Berkeley National Laboratory, Earth and Environmental Sciences Area

⁴Hydrologic Research Center

⁵University of Oklahoma, School of Meteorology

⁶Cornell University, Department of Earth and Atmospheric Sciences

⁷Polar Bears International

Corresponding author: Stefan Rahimi (srahimi@uwyo.edu)

Key Points:

1. Statistically-extended observations under-predict megaflood intensity by at least 10%.
2. 100- and 200-year storms increase more rapidly over time—both in frequency and magnitude—than storms with shorter recurrence intervals.
3. When passed through dynamic downscaling, biases from ESMs amplify storm sequence variability, delaying emergent climate trends.

Abstract

A California “megaflood” - a weeks-long cluster of storms with centuries-long return times - would be catastrophic, yet its timing and magnitude under climate change remains uncertain. A central source of uncertainty is whether Earth System Model boundary conditions should be bias corrected before dynamical downscaling. Here, we analyze a large ensemble of downscaled simulations to quantify how this uncertainty shapes California's megaflood risk. Bias correction is necessary to reproduce observed storm statistics and strongly influences both the megafloods historical and future characteristics. Without correction, storm precipitation variability is inappropriately inflated, delaying the emergence of extreme precipitation signals by decades. Further, megaflood intensity estimates based on statistically-extended observations are at least 10% drier than explicit computations derived from sufficiently-sized storm populations. Our results show that bias correction choices crucially govern projections of rare, high-impact events, with direct implications for infrastructure design/planning and water management in California.

Plain Language Summary

Infrastructure design and planning are often informed by the best available downscaled climate projections. However, downscaling as a technique can produce low-quality outcomes if errors in driving climate models are not confronted and addressed (i.e., the garbage in, garbage out problem). Since billions in capital may be invested in infrastructure based on such information, it is critical to understand the consequences of climate model bias removal, termed bias correction, on high-impact infrastructure-defining events (e.g., the California megaflood). Here we show that with bias correction, coherent signals in the state's rarest 30-day rainfall events emerge decades sooner than if climate model biases are not removed prior to downscaling. Our results indicate that policymakers and decision-makers may

42 potentially be making investment choices based on inappropriate time horizons in the 21st century when
43 using non-corrected climate data.

1. Introduction

California has historically experienced rare but catastrophic “megafloods”, multi-week storm sequences capable of submerging entire regions, destroying infrastructure, and reshaping the state’s economy. The Great Flood of 1861–62, driven by successive atmospheric rivers, produced nearly 10 feet of rain in some locations over a 43-day period, inundated Sacramento, killed thousands, and destroyed one-third of the state’s taxable property (Dettinger et al., 2012; Huang & Swain, 2022; Ingram, 2013). Such events pose fundamental threats to infrastructure and emergency preparedness frameworks (Porter et al., 2011), making their future behavior under climate change a central concern for California’s resilience planning.

Several specific scenarios have been used to test the resilience of the state’s infrastructure, including probable maximum precipitation (PMP) and the ‘ARkStorm’ (the once per thousand-year Atmospheric River Storm), which serve different purposes. PMP for example is the most extreme individual 3-day event (on the order of ~600 mm [24 inches] in northern CA) thought to be possible at a given location, and is typically estimated for high-hazard dams and other critical infrastructure. ARkStorm scenarios meanwhile are multi-week precipitation sequences and are used to understand vulnerabilities in preparedness over the entire state. The two ArkStorm scenarios that have been developed produced ~965–1600 mm [~38–63 inches] (ARkStorm1.0 (Porter et al., 2011)) and ~1270–1780 mm [~50–70 inches] (ARkStorm2.0; (Huang & Swain, 2022)) at individual locations. Another complementary approach to flood preparedness is flood-frequency analysis to estimate the 100-year (and other return intervals) flood, which is used for most federal infrastructure projects, FEMA floodplain mapping, and state highway design (England Jr. et al., 2018).

Many studies have highlighted that California’s complex terrain is key to modulating storm impacts and its response to warming. This occurs through a variety of mechanisms, including orographic enhancement of precipitation systems by mountain ranges (Houze Jr., 2012), rain-on-snow events at some elevations (Brandt et al., 2022; Heggli et al., 2022), and the control exerted by topography on surface and subsurface water flows. The coarse resolution of Earth System Models (ESMs) is prohibitive for representing these processes, leading to the need for physics-based downscaling.

Dynamical downscaling is typically performed using the Weather Research and Forecasting (WRF; (C. Skamarock et al., 2019; W. C. Skamarock & Klemp, 2008)) model to downscale reanalyses or ESMs to terrain-resolving scales. There are multiple variants of this technique, each with their strengths and weaknesses (Hall et al., 2024). For example, the pseudo-global warming approach modifies historical storms, usually through warming of the boundary conditions, before downscaling. In such simulations, warming was found to intensify (diminish) precipitation amounts for events in which atmospheric rivers are (not) accompanied by a developed extratropical surface cyclone (Patricola & Wehner, 2018). Alternatively, ESM output for the future period can be used directly for WRF downscaling. Using this approach with dynamically downscaled output from the Community Earth System Model’s first Large Ensemble (CESM-LE) for an assortment of AR events, (Huang et al., 2021) found that precipitation intensity from ARs will generally increase (by as much as 10–40% in valleys and in the lee of mountains) and is mostly attributable to thermodynamic increases in water vapor. Similarly using CESM-LE ESMs to drive WRF, The ArkStorm2.0 study (Huang & Swain, 2022) dynamically downscaled a single 30-day historical and one future storm, finding the future storm to be 45% wetter than the historical event. The authors also found that the likelihood of historical megafloods will increase by roughly a factor of 2, 4 and 6 under 1°, 2°, and 4°C of global warming from non-downscaled full CESM-LE simulations.

The above studies provide a strong foundation for California's flood planning efforts. However, several sources of uncertainty have not been accounted for when considering the modeling of extreme storms and future shifts in their characteristics. First, using reanalysis even if downscaled (e.g., (Patricola et al., 2022; Porter et al., 2011)) constrains the dynamical storm behavior to match observed events selected from a short period with the current phasing of weather sequencing (Fischer et al., 2021). Second, most studies to date, including the quantification of ARkStorm2.0 (Huang & Swain, 2022), do not account for ESM-based structural uncertainty; this study also does not quantify the impacts of internal climate variability on the downscaled outcomes (Hawkins & Sutton, 2009). Third, research shows the importance of bias-correction (BC) of the ESM boundary conditions before dynamical downscaling (Bruyère et al., 2014; Holland et al., 2010; Kim et al., 2023a, 2023b; Pontoppidan et al., 2019; Rahimi et al., 2024; Risser et al., 2024). Specifically, (Rahimi et al., 2024) found that ESM bias transmission led to ~40% wet-biased 30-day precipitation events when considering a 270-year historical climatology from dynamically downscaled ESMs. When mean biases in climate-mean temperature, sea-surface temperatures (SSTs), horizontal winds, and moisture from ESMs were removed, biases were reduced to 5-10%. Further, while (Risser et al., 2024) found that BC does not consequently affect climate trends across the western U.S., their results suggested that a deeper analysis of BC impacts for targeted weather events is needed to develop a more complete assessment of its broader applicability as a standard operating procedure in dynamical downscaling.

Here we address the opportunities presented in previous studies by analyzing 40 dynamically downscaled simulations spanning more than 6,000 model years. Our ensemble design enables a systematic comparison of how bias correction choices, structural model diversity, and internal variability shape estimates of California's megaflood risk, providing the most comprehensive assessment to date of methodological influences on projections of rare, catastrophic precipitation events.

2. Downscaling behavior and model bias

We use 30 dynamically downscaled Coupled Model Intercomparison Project Phase 6 (CMIP6) simulations comprising the Western U.S. Dynamically Downscaled Dataset (WUS-D3) spanning the 1980-2100 period. Fifteen members apply gridpoint-specific bias correction to the ESM boundary winds, temperature, humidity, and SSTs (see Materials and Methods; labeled w/BC), and 15 use the same models without bias correction (labeled no-BC), forming two 1,800-year ensembles (3,600 years total) to test sensitivity to ESM choice and bias correction. We add ten WRF-downscaled Community Earth System Model version 2 (CESM2) Large Ensemble members spanning 1850–2100 (WRF-CESM2), also bias-corrected, contributing 2,500 more years to probe internal variability and single-model versus multi-model behavior (6,100 years overall). For observational context, we include a reanalysis-driven simulation using ERA5 boundary conditions (WRF-ERA5; 1950–2023) that reproduces observed weather sequencing. More information on the downscaled data used here can be found in Section S1.1, and a complete listing of downscaled ESMs is provided in Table S1. Model precipitation and related fields are evaluated against gridded observations from PRISM (4 km, 1981–2020) and the Livneh dataset (6 km, 1951–2018) to characterize strengths and weaknesses of the downscaled outputs.

2.1. Effects of downscaling on rx30day

When downscaling ESMs to the kilometer scale, (Rahimi et al., 2024) found that regardless of whether or not BC was applied to the ESM boundary conditions, downscaling led to a wetter representation of the historical rx30day climatology in the California state mean, compared to observations; rx30day is defined to be the climatology of the wettest 30-day precipitation sequence every year. Specifically, an ensemble of no-BC experiments were 30% wetter than PRISM, while w/BC experiments were less than 5% wetter. Lundquist et al. (2019) argue that simulations exhibiting this degree of wet bias relative to gridded products such as PRISM and Livneh may actually be of high quality given the assumptions made in extrapolated precipitation across data sparse regions characterized by mountainous terrain, or known undercatch issues in the observations which are used to build these

gridded datasets. Particularly for intense atmospheric rivers, both issues may lead to the underrepresentation of extreme precipitation events in Livneh and PRISM.

Building on (Rahimi et al., 2024), we take this evaluation a step further by considering rx30day as a function of multiple Return Intervals (Ris): 2-, 10-, 50-, 100-, and 200-years (the latter is defined as a megaflood). Downscaling generally leads to greater intensification of identical 1851-2099 CESM2-LE-simulated rx30day events (ESM-CESM2) with increasing RI, being 1%, 11%, 12%, 16%, and 15% wetter for the 2-year (i.e., median), 10-, 50-, 100-, and 200-year RIs (ESM values of 204, 301, 379, 407, and 439 mm, respectively; Fig. S1, left). By comparison, the strongest rx30day event amongst all w/BC, no-BC, and WRF-CESM2 experiments is simulated between 16 December 1962 and 14 January 1963 (714 mm); this maximum occurs one day earlier in the native ESM (551 mm), representing the 3rd wettest sequence in the entire 100-member CESM2-LE (variant 1011); the most intense storm sequence in the full CESM2-LE of 562 mm occurs in 1981. These storms' occurrence in the historical instead of the future periods underscores the important role that internal climate variability can play in modulating the periodicity of extreme rx30day events. That conclusion is also supported by the findings in (Rhoades et al., 2020), despite statistically robust storm-mean wetting signals in future cumulative precipitation across the western U.S. in their variable resolution downscaling experiments. This is particularly important for interpreting the results of the recently quantified ARkStorm2.0 (Huang & Swain, 2022), whose future storm was chosen to be wetter (~333-year RI in a 2051-2100 climate) than its historical comparative (85-year RI in a 1971-2020 climate).

In WRF-CESM2, rx30day storms are generally wetter than in the raw ESM, but this behavior is not necessarily true in all w/BC experiments. Median, 10-, 50-, 100-, and 200-year storms for the BC experiments are +1%, +1%, -1%, +2%, and +4% different on average than their parent ESMs in the 1981-2099 period (Fig. S1, center). In contrast, a collection of ESMs (MPI-ESM1-2-LR, MPI-ESM1-2-HR, MIROC6, CNRM-ESM2-1, ACCESS-CM2, and EC-Earth3-Veg) depict WRF-to-ESM slopes of less than unity (Fig. S2, top), indicating that these native ESMs are, on average, each wetter than its downscaled counterparts. The strongest storm sequence from native CMIP6 ESMs and w/BC are nearly identical: 602 mm in 2066 (MPI-ESM1-2-HR) and 603 mm in 2089 (UKESM1-0-LL). In no-BC experiments, WRF-simulated storms are 25%, 23%, 19%, 18%, and 22% wetter than identical storms in their parent ESMs for median, 10-, 50-, 100-, and 200-year storms (Fig. S1, right), with all WRF-to-ESM slopes greater than 1 (Fig. S2, bottom). The most intense storm sequence from no-BC occurs in 2096 (690 mm) and is 14.6% wetter than any storm sequence simulated from native CMIP6 ESMs. This systematic wetting of storms by downscaling in no-BC experiments is linked to inherited large-scale ESM biases that favor increased precipitation. When removed, this behavior is muted, and more information on the nature of these biases can be found in (Rahimi et al., 2024).

2.2. Historical biases

We next evaluate historical biases in the WRF-downscaled datasets across time scales using both explicit calculations alongside predictions from Generalized Extreme Value (GEV) Theory (see Sec. S1.2 for specifics on the methodology). WRF-ERA5 is 9.5% and 11.1% wetter than PRISM and Livneh, respectively, for median rx30day storm sequences, and 9.9% and 9.5% wetter for 20-year events (Figure 1a). In the presence of adequate observations, the differences between WRF-ERA5 and observations may be thought of as biases introduced by WRF in downscaling. However, since the largest differences between WRF-ERA5 and observations occur over mountains, where observations are sparse, these differences may be thought of as part model bias and part observational uncertainty (Lundquist et al., 2019). For the downscaled ESMs, BC of ESM boundary conditions results in 2- and 20-year storms that are more realistic compared to reference datasets (Fig. 1a). For median storms, ESM-mean Livneh-relative biases of +50% in the no-BC median rx30day are reduced to +13.9% in w/BC, comparable with WRF-ERA5. And similarly, WRF-CESM2, in which BC is applied, has just a 7.9% bias. For 20-year storms, the bias in no-BC drops from +42% to bias values comparable with that of WRF-ERA5 in w/BC and WRF-CESM2: +12.7% and +10.1%, respectively.

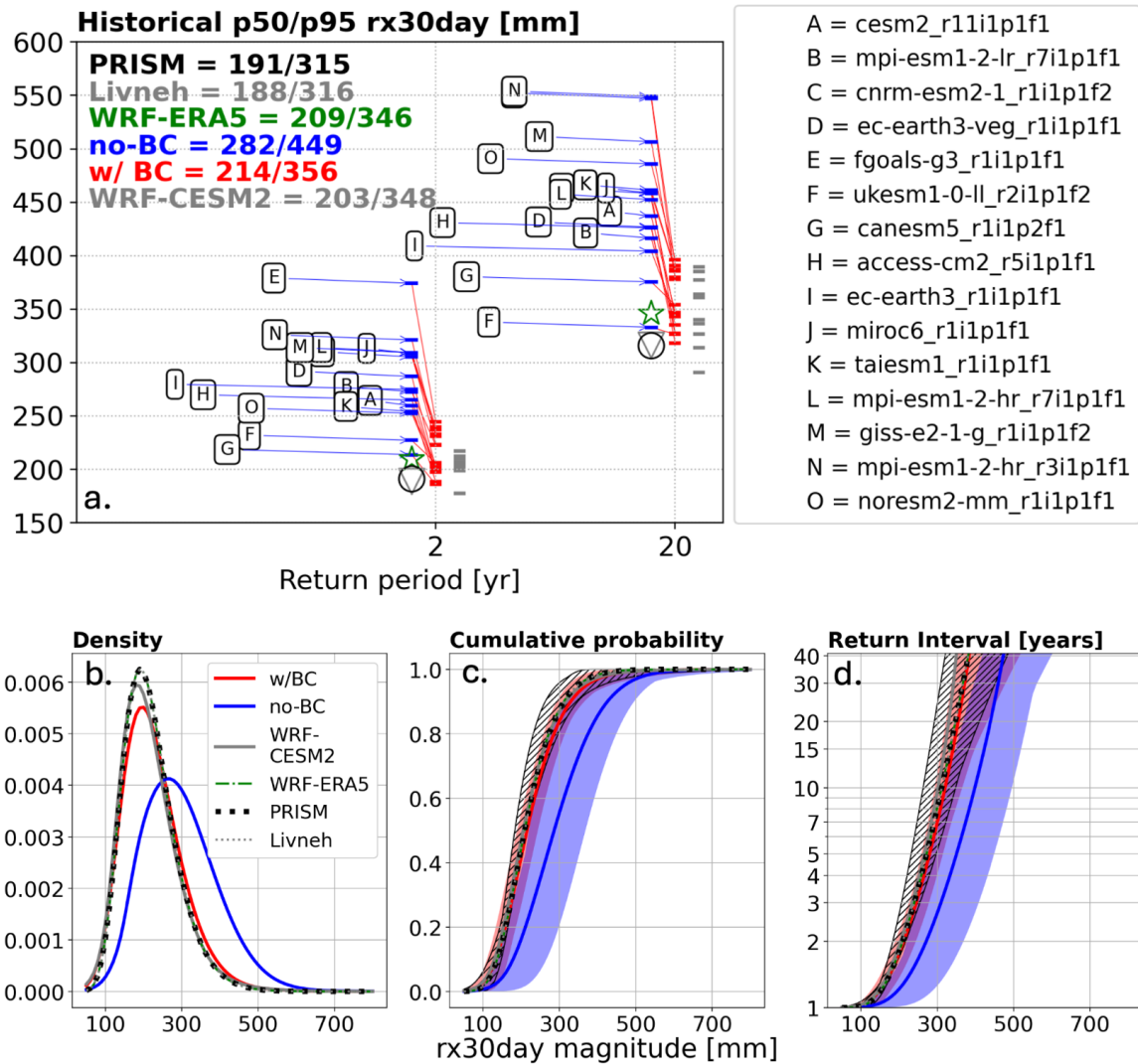


Figure 1. Historical bias profiles: (a) Historical-era (1981-2020) 2- and 20-year rx30day total precipitation magnitudes [mm] for different downscaled ESMs and reference datasets. Non-bias corrected (bias corrected) CMIP6 experiments are displayed as blue (red) horizontal lines and labeled as no-BC (w/BC). Downscaled CESM2-LE (WRF-CESM2) experiments are presented as gray horizontal lines. PRISM, WRF-ERA5, and Livneh values are presented as a black open circle, green open star, and gray open triangle. Twin CMIP6 experiments are connected by red lines. ESM-mean magnitudes for 2- and 20-year return levels are provided in the legend, delineated by a '/'. Panels (b), (c), and (d) present ESM-based probability distribution, cumulative distribution, and return interval estimates from ESM-specific Generalized Extreme Value distribution fits. Blue, red, and hatched shading in (a) depicts the histograms of no-BC, w/BC, and WRF-CESM2, respectively. For (b) and (c), the shading represents the GEV spread produced from multiple ESMs.

Considering the full probability and cumulative densities (Figs. 1b,c), as well as RIs (Fig. 1d) as derived from the GEV approximation, reveals further bias characteristics. WRF-ERA5, as well as the w/BC experiments, closely follow the probability density and cumulative density functions (PDFs and CDFs, respectively) of Livneh and PRISM. Here, ESM-based GEVs are fitted by aggregating 40-year periods from 15 ESMs into 600-year periods to compute the GEV parameters. Without BC, however, an

unrealistically wet-shifted PDF/CDF of rx30day arises, leading to unrealistically low RIs in no-BC experiments. Further, a compressed spread is evident in w/BC (red shading) compared to no-BC (blue shading) experiments. This is also a feature of the 10-ESM WRF-CESM2 experiments (hatched shading). The no-BC experiments meanwhile, have much wider spreads in their PDFs, CDFs, and RI profiles; wider inter-ESM spreads in no-BC-simulated rx30day were found in (Rahimi et al., 2024) owing to differences in precipitation-enhancing mean biases (e.g., temperature and vorticity) transmitted from parent ESMs to WRF. Distributional rx30day profile differences between no-BC and w/BC are also present in other climatological periods (Fig. S3).

2.3. The historical 200-year storm

Presented in Figure 2, GEV-based estimates of the historical megaflood are between 407 and 430 mm, with the PRISM (1981-2024) and Livneh (1951-2018) datasets nearly statistically indistinguishable from each other in terms of their 95th percent confidence intervals (CIs; bars bounding the 2.5th and 97.5th percentiles from 1000 bootstrapped samples). Interestingly, the 1951-2023 WRF-ERA5-based upper bound estimate is much lower than PRISM and Livneh, even though its 200-year return level is virtually identical to PRISM. This could be due to (i) spatial and temporal inhomogeneities in the observational data (i.e., real-world noise) which may drive up the variability in rx30day from which the 200-year RL is defined and (ii) resolutions in PRISM (4-km) and Livneh(6-km) are higher than WRF-ERA5 and downscaled ESM ensembles (9-km).

For ESM experiments, explicitly computed megaflood return levels (RLs) are at least 10% wetter than GEV-extended observations, covering a noticeably smaller range (Fig. 2a); smaller CI ranges are expected given the larger sample sizes. Here, GEV parameters were computed based on aggregating all available realizations for each ESM ensemble over the 1981-2020 time period, resulting in between 400-600 samples from which parameters are derived. Statistical similarity between explicitly calculated and GEV-estimated megaflood intensities is also evident amongst w/BC, no-BC, and WRF-CESM2 experiments, identical WRF-CESM2 storms but in their parent ESM (ESM-CESM2), and all storms from the full 100-member CESM2-LE (Full ESM-CESM2). Finally, using a Gumbel approximation yields megaflood intensities which are 5-10% wetter than when assuming the full GEV (Fig. S4), and its intensity is relatively insensitive to the choice to aggregate all ESM realizations before RL computation (Figs. 2, S4a) instead of computing RLs for each ESM separately using 40-year samples (Fig S4b). In summary, the RL statistics are relatively insensitive to common, alternative methodological choices, lending confidence to the megaflood magnitude in datasets where it cannot be explicitly calculated.

BC, ensemble choice and size, and the use of native ESMs versus downscaled data all affect the estimated megaflood intensity. In no-BC, the explicitly simulated megaflood (562 mm; 22 inches) is, or is nearly, beyond the 95th percent CI of PRISM (Livneh); and its bootstrapped-derived 95th percent CI is beyond the upper bound of any observationally-constrained gridded product (627 mm; 24 inches). Meanwhile, w/BC and WRF-CESM2 exhibit megaflood intensities of 471 and 494 mm (18.5-19.5 inches), respectively, with w/BC covering a narrower statistical uncertainty space and presenting a 5% weaker megaflood than WRF-CESM2. We find that BC of ESMs dries the megaflood by more than 16% when comparing w/BC to no-BC experiments, consistent with rx30day drying found in (Rahimi et al., 2024). Overall, this indicates that the choice to bias correct plays a larger role in modulating historical megaflood intensities than ESM ensemble choices. Additionally, we find that identical storms in the raw ESM ensemble (ESM-CESM2) are wettened by nearly 19% through downscaling, however, this wetting of rx30day events by downscaling may not be universal to all ESMs as indicated in Section 3.1 (Figs. S1, S2). We also find that using all 100 members of the raw CESM2-LE, instead of the 10-member sample that coincides with WRF-CESM2, wettened the megaflood by 9%. Finally, we report that the above calculations for CESM2 make use of output from 1981-2020. When using data from 1851-1980 instead of 1981-2020, the explicitly-computed RLs are virtually unchanged. We cannot however assert that a 400-year sample (40 years X 10 realizations) is large enough to achieve a convergent megaflood intensity, as megaflood intensity is found to fluctuate over time, based on a running 40-year climatology (see Section

3). Despite this, it is clear that BC has at least a comparable or larger influence on megaflood intensity than both downscaling considerations and ESM-ensemble choices.

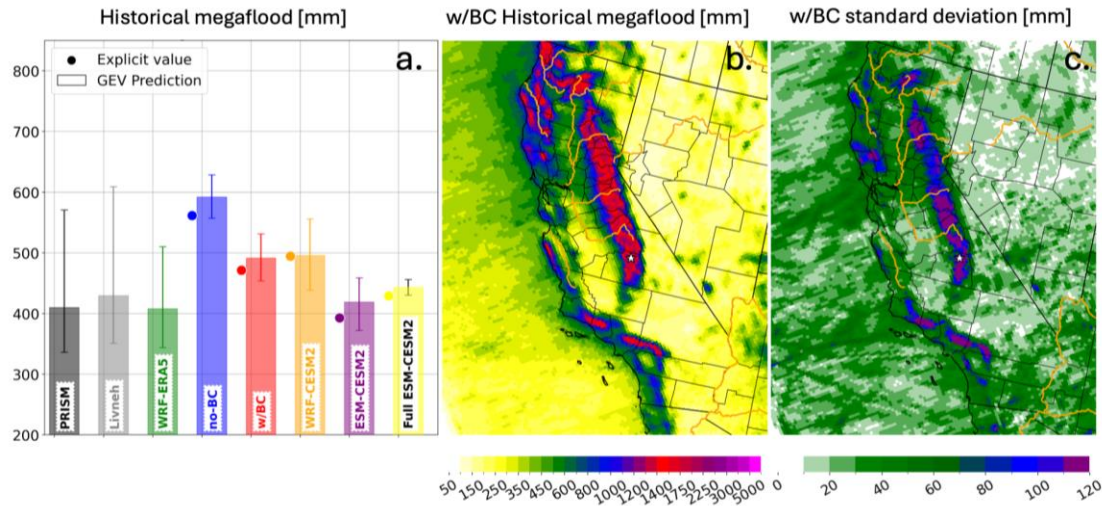


Figure 2. The historical megaflood quantified: Historical (1981-2020) megaflood total precipitation magnitude [mm] across the whole of California computed explicitly (circles) and using Generalized Extreme Value Theory (fat bars). In (a), 200-year return levels (RLs) are computed after aggregating amongst ESM ensembles before distributional parameters are computed. The 95th percent confidence interval for each dataset is estimated using 1000 bootstrapped samples, bounded by each population's 2.5th and 97.5th percentile values, and is presented as whiskers on the bars. Panel (b) shows the explicitly-computed gridcell-specific 200-year RL from w/BC experiments [mm]. Panel (c) shows the standard deviation of the w/BC-simulated megaflood computed using 1000 bootstrapped samples [mm]. The white star shows the location of California's maximum megaflood RL.

We next use our w/BC ensemble to examine megaflood magnitudes at specific locations across California. (Fig. 2b). Here, the megaflood is defined as the local 200-year event, and thus is likely to derive from completely different storms across the state, manifesting in large uncertainties in their magnitudes as quantified by bootstrapped-derived standard deviation values (Fig. 2c). Therefore, the results in Figs. 2a and 2b cannot be directly linked. With this in mind, we present w/BC estimates here given their broader realism in simulating the full rx30day distribution (Fig. 1) relative to no-BC, noting their similarity to WRF-CESM2. The megaflood showcases a modest correlation with the local terrain height: 0.34, 0.35, and 0.41 in WRF-CESM2, w/BC, and no-BC, respectively; correlations in standard deviations and terrain height are similarly modest. The higher correlations in no-BC may be attributed to an inappropriately tight coupling between the orographic forcing and the overlying atmospheric conditions, which are specifically prone to erroneously large upward velocities in the no-BC experiments via ESM biases which are not removed before downscaling. Coastal California's megaflood intensity ranges from 800-1200 mm (31-47 inches), with the coastal ranges of southern California being prone to event intensities between 1500-2000 mm (59-78 inches). The highest elevations of the Sierra Nevada are prone to the largest megaflood intensities across the western U.S. (not shown), with values in excess of 2000 mm (78 inches) in several locations. A California-maximum of 2471 ± 155 mm (97 ± 6.1 inches) is simulated in the southern Sierra Nevada (white star, Fig 2b,c). Meanwhile, the no-BC California maximum is much higher than w/BC: 2861 ± 160 mm (112 ± 6.2 inches), while the maximum in WRF-CESM2 is more similar to w/BC (2529 ± 191 ; 99 ± 7.5 inches). We note a highly similar spatial pattern of megaflood intensity across California between w/BC, no-BC, and WRF-CESM2 experiments (inter-ensemble correlations > 0.98 ; Fig. S5), with no-BC being much wetter relative to w/BC.

3. Future changes in rx30day magnitude and frequency

Next, we turn our attention to examining how California's rx30day magnitude will respond to future regional change. The colorfill in Figure 3 presents rx30day magnitude as a function of rolling 40-year climatology and RI from different ESM ensembles. For a given RI, total precipitation magnitude increases throughout the 21st century. By end-of-century (2061-2099) median rx30day events are projected to increase in magnitude by 22, 20, and 32 mm (7.7, 9.3, and 15.7%) in no-BC, w/BC, and WRF-CESM2, respectively (Fig. 3a-c; see absolute magnitudes in Fig. S5). Meanwhile, the 50-year storm is quite sensitive to BC choices with a projected intensification of 41 (8.1%) in no-BC, but 63 mm (15%) in w/BC and 52 mm (13.8%) in WRF-CESM2. Meanwhile, rarer storms experience even greater fractional increases: the 100-year storm intensifies by 67, 71, and 70 mm (12.2, 16.5, and 16.2%) in no-BC, w/BC and WRF-CESM2, respectively; and the megaflood, i.e., 200-year storm, intensifies by 91, 54, and 77 mm (16.7%, 15.0%, and 11.5%) in the same ensembles. Time series of these changes for specific RIs are shown in Fig. S6. For perspective, a 50 mm increase is approximately 21 km³ (~17 million acre feet) over a 30-day period, representing a volume of water which is ~58% of California's total reservoir storage capacity at the time of writing this paper (<https://engaging-data.com/ca-reservoir-dashboard/>).

The different behaviors between no-BC and w/BC experiments can be explained via summary statistics. Both ensembles exhibit positively skewed rx30day historical distributions, which is much larger (~0.2) in w/BC than no-BC. That is, the distributions of rx30day in w/BC have relatively longer right tails than in no-BC because the latter experiments' inner-quartiles are broader and wet-shifted relative to w/BC as the transmission of ESM bias distorts precipitation statistics in no-BC. This skewness behavior is consistent and systematic throughout the 1981-2099 period (Figs. S3, S7). Before about 2040, skewness gradually decreases in both ensembles, after which time skewness rapidly increases in positivity, eclipsing its historical maxima by 2070 (Fig. S7a). Meanwhile, no-BC experiments exhibit negative Gaussian-relative kurtosis (shorter tails) from the historical period through about 2050 (Fig. S7b), a fundamentally different behavior than the w/BC experiments, which exhibit positive Gaussian-relative kurtosis (longer tails; Fig. S7b). Notable differences in climatology-specific variability trends are present between no-BC and w/BC experiments, as defined by each population's standard deviation (Fig. S7c). While both no-BC and w/BC simulate increasing variability under warming, w/BC shows a more rapid increase than no-BC. Ultimately, these variability, kurtosis, and skewness differences underpin the different behaviors between no-BC and w/BC experiments' RLs during their historical eras. Specifically, ESM mean-state biases are transmitted to WRF in the no-BC experiments, leading to storms that are systematically wetter. This inflates population variability and underskews the PDFs, resulting in downscaled behaviors that mask emergent externally-forced climate trends. By 2040, an anthropogenic warming signal begins to intensify rx30day magnitudes in both sets of experiments (defined by positive skewness trends; Fig. S7a). For this signal to emerge, it must overcome internal variability, which is critically important for modulating storm sequences with longer RIs (i.e., the 100- and 200-year storms). Since the no-BC experiments have unrealistically high variability owing to transmitted mean-state ESM biases, emergent trends in 100- and 200-year storms manifest later in the 21st century. For example, historical 100-year storms in w/BC become 50-year storms by the 2045 climatology (Fig. 3b), while this transition occurs by the 2080 climatology in no-BC (Fig. 3a) experiments. In summary, we contend that trend divergence between no-BC and w/BC is the result of the distribution-wide distortion of rx30day within no-BC experiments associated with ESM bias transmission, leading to masked emergent climate trends, rather than unphysical trend modifications associated with pre-downscaling bias correction choices.

Further divergence in projected rx30day is expressed when examining WRF-CESM2 experiments (Fig. 3c). Different from both w/BC and no-BC, WRF-CESM2 exhibits reduced rx30day intensities from 2010 through about 2040, over which time the 10 underlying CESM2-LE simulations showcase strong regional and global warming trends (not shown). While this feature aligns well between WRF-CESM2 and identical storm sequences from ESM-CESM2 (Figs. 3c,d), Fig. 3e, which depicts the rx30day RLs

from the 100-member CESM2-LE (Full ESM-CESM2), does not show the same decadal-scale variability in RLs. This suggests that internal climate variability as depicted by the selected 10 CESM2 members masks the emergence of a near-future forced response as found in the larger 100 CESM2 member ensemble. Specifically, when considering all 100 members, there are no local minima in RLs centered around 2025 anymore. In fact, around 2025, the historical 200-year storm RI trace begins to monotonically decrease at a rate of ~ 30 years per decade, becoming a 60-year event by 2080.

Finally, since rx30day is wettened by dynamical downscaling in WRF-CESM2, we fit a fourth-order polynomial to the scatter of WRF-CESM2 versus ESM-CESM2 storm magnitudes (Fig. S2), using its coefficients to scale the Full ESM-CESM2 to effectively downscale all storms from the 100-member CESM2-LE (Fig. 3f; Full ESM-CESM2 Scaled). Wetter RLs relative to Full ESM-CESM2 manifest, but the RL patterns between the scaled and unscaled ensembles are nearly identical. This implies smaller relative future changes in the scaled ensemble and has direct implications for determining precipitation scaling factors, which are often derived for probable maximum precipitation studies and applied in statistical downscaling frameworks (Pierce et al., 2014, 2023).

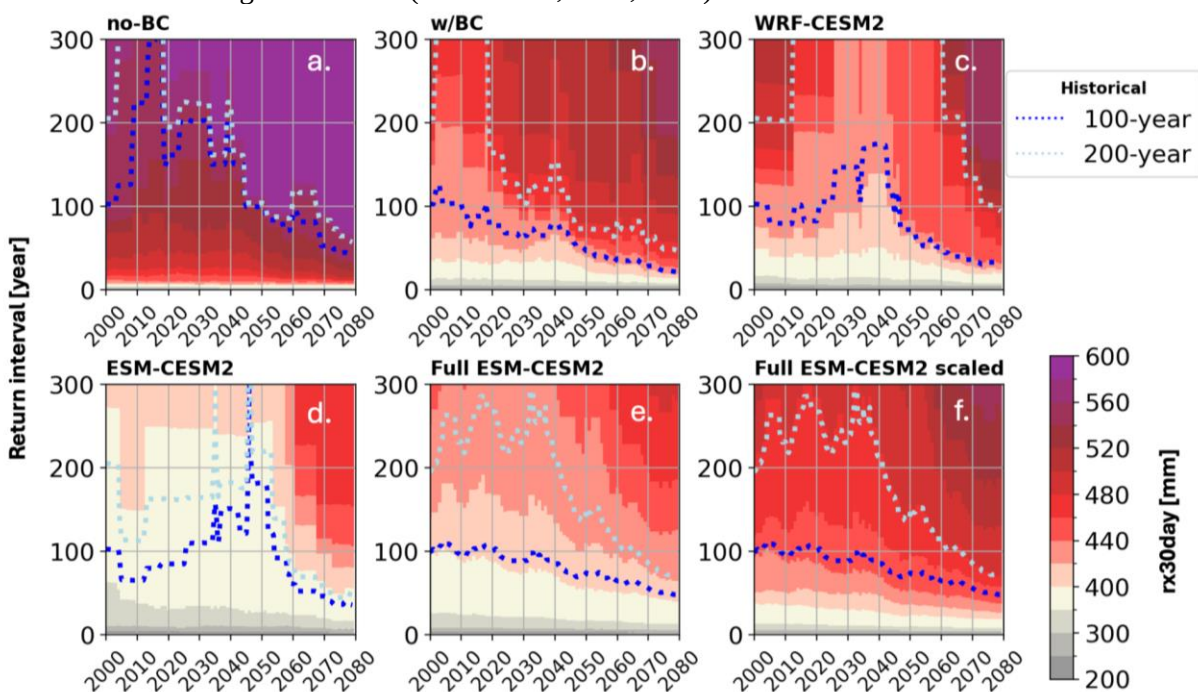


Figure 3. Storm-sequence changes in time: California-mean rx30day magnitude (colors; mm) as a function of year (x-axis) and return period (y-axis). Shown for: (a) no-BC, (b) w/BC, (c) WRF-CESM2, (d) identical storms in the native CESM2-LE (ESM-CESM2), (e) the 100-member CESM2-LE, and (f) the 100-member native CESM2-LE with a fourth-order scaling factor applied, which estimates the wetting effects of downscaling using information derived from WRF-CESM2. Each year on the x-axis represents a 40-year running centered climatology. Line contours represent the 1981-2020 return levels for the 100- and 200-year storms.

Next, we explicitly evaluate how different dynamically downscaled ESM ensembles represent the changing likelihood of rare rx30day events. Figure 4 presents the annual likelihood of 20-, 50-, 100-, and 200-year storms as a function of 40-year climatology amongst different datasets throughout the 21st century. By end-of-century, the historical 20-year storm will occur roughly 72 and 79% more frequently in w/BC and no/BC, respectively. Meanwhile, WRF-CESM2 projects a slightly larger increase of 121%. 50-year storms show an even larger change in their frequency, with end-of-century increases of 123%, 166%, and 125% in no-BC, w/BC, and WRF-CESM2, respectively, effectively reducing the historical 50-year storms' RIs to 22, 19, and 22 years.

The 100-year and 200-year storms generally showcase the greatest jumps in frequency throughout the 21st century while also diverging the most between the ensembles. Specifically, no-BC, w/BC, and WRF-CESM2 predict that the historical 100-year storm will occur 2.5, 4.3, and 3.25 times as often by end-of-century, respectively, reducing the RIs to 40, 23, and 31 years. The historical megaflood, on the other hand, will occur every 60, 50, and 100 years in no-BC, w/BC, and WRF-CESM2, respectively, by end-of-century. Regardless of RI considered, however, amplifications in the frequency of rare rx30day events, from 20 to 200 years, may be mischaracterized or underrepresented when failing to account for ESM biases and structural uncertainties. More broadly, the rarest storms (i.e., 100- and 200-year RIs) are more sensitive to frequency shifts under warming than 20- and 50-year storms, meaning that events with different frequencies respond differently to warming. This overall behavior is also evident when comparing empirical calculations of RIs to GEV-based estimates (Fig. S8).

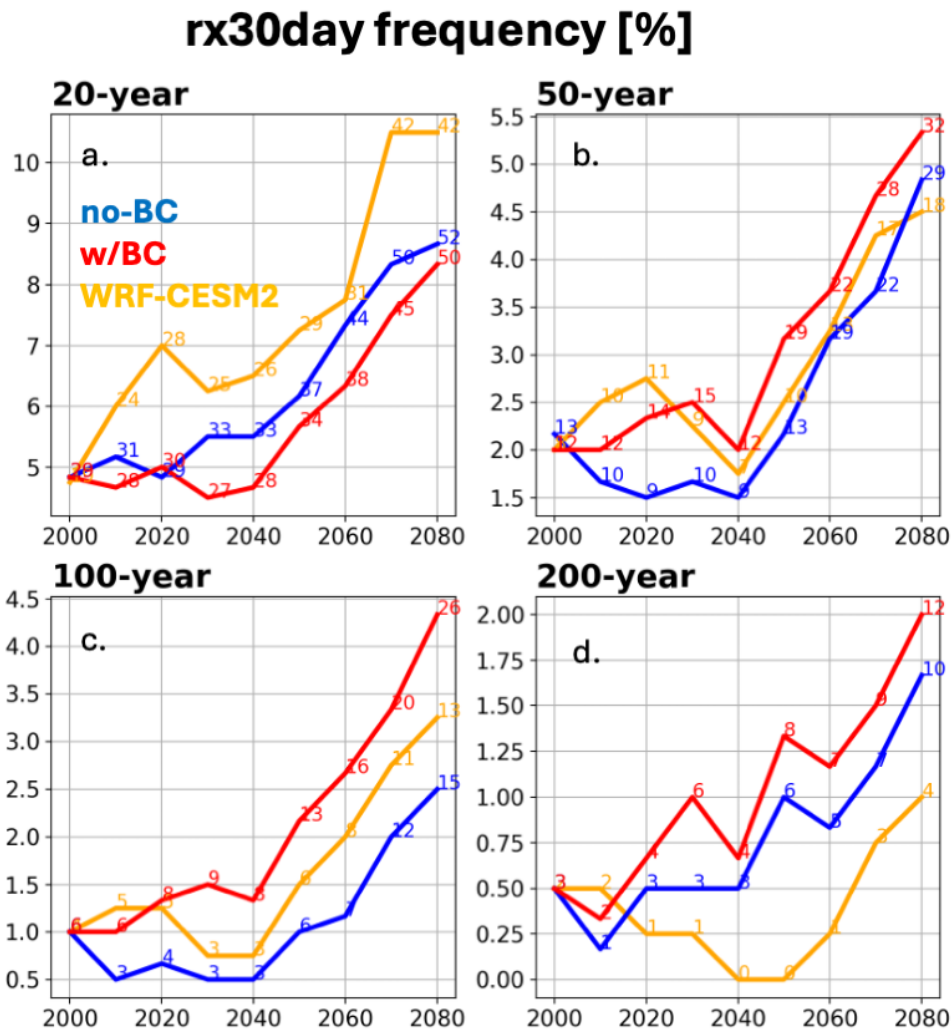


Figure 4. Frequency shifts in storm sequences: Frequency estimates of the historical (1981-2020) 20-year (a), 50-year (b), 100-year (c), and 200-year (d) rx30day events across California from different dynamically downscaled ESM ensembles, including no-BC (blue), w/BC (red), and WRF-CESM2 (orange). The small numbers show the number of storm sequences which exceeded the historical threshold for each considered period. The years labeled on the x-axis represent the midpoint of a 40-year window.

4. Discussion

We use 40 dynamically downscaled simulations to quantify California's historical megaflood, defined as the once-per-200-year total state-wide precipitation magnitude over a 30-day period, and its response to regional warming. We consider the impacts of bias correction (BC), structural uncertainty, and internal variability, as well as the effects of downscaling in quantifying megaflood frequency and magnitude characteristics.

BC of the monthly means of ESM boundary conditions increases the physical realism of historical rx30day across the probability density relative to observations, at least for ESMs that were chosen based on their skillful representation of regional dynamics. The distributions of 30-day storm sequences are unrealistically wetter and exhibit overly strong variability in the simulations without BC. The unrealistically wet behavior was also present in estimations of the historical 200-year storm (the megaflood), with non-bias corrected simulations predicting a megaflood magnitude of 562 mm (~22 inches), a value at the upper edge of or exceeding observational uncertainties derived from Generalized Extreme Value-based (GEV-based) methods. Meanwhile, experiments with BC from the CESM2-LE and CMIP6 ensembles, predicted a 14-to-19% drier megaflood magnitude (471 and 494 mm; 18.5 and 19.5 inches, respectively) relative to simulations without BC. Of note, these explicit calculations of megaflood magnitude were larger than GEV-extended observations, indicating a potential underestimation of megaflood intensity when using statistical methods to estimate the megaflood intensity.

Regarding the impacts of downscaling on storm intensity, we find that megafloods within dynamically downscaled CESM2-LE storms were 19% wetter than identical storms in their parent 10-ESM ensemble. However, even using the full 100-member CESM2-LE fails to capture the magnitude of megafloods predicted from dynamically downscaled experiments within the CESM2-LE; using 4000 versus 400 years of historical native CESM2-LE outputs to calculate the megaflood intensity predicts a value 15% smaller than a value based on 400-years of dynamically downscaled CESM2-LE outputs. While dynamical downscaling tended to wetten rx30day events within the structurally similar CESM2-LE (particularly for the more intense events), this same behavior may not be true for structurally dissimilar CMIP6 ESMs, as 6/15 downscaled simulations tended to be drier than their parent ESMs for rx30day events of increasing magnitude. In summary, the historical megaflood is most sensitive to BC of the ESM boundary conditions and the act of downscaling itself. At least in this study, the effects of internal variability and sample size (using 400 versus 4000 years), and structural ensemble differences (15 single-member ESMs versus a single ESM with many members) are of secondary importance in terms of modulating the 200-year storm's magnitude. On this last point, it was found that considering a 10-member ensemble (ESM-CESM2) versus a 100-member ensemble (Full ESM-CESM2) was important for disentangling decadal-scale internal variability from externally-forced climate trends when considering a single ESM.

Following the SSP3-7.0 trajectory, rx30day events intensify throughout the 21st century, with the rarest events (100- and 200-year RIs) undergoing the largest intensifications and shifts in probability and intensity relative to 20- and 50-year events. By end-of-century, dynamically downscaled estimates project a wetter megaflood by 11.5 to 16.7% depending on ensemble and bias correction choices, while the historical megaflood will occur at least twice as often (every 50-100 years).

Concerningly, we found that trends in historically rare storm sequences are modified when removing climate-mean ESM biases prior to downscaling. For example, the historical 100-year storms in the bias-corrected ensemble become 50-year storms by the 2045 climatology, while this transition occurs by the 2080 climatology in a twin ensemble without BC. This trend divergence potentially signifies that emergent climate signals may be masked by the transmission of ESM biases of different magnitudes, which act to distort the summary statistics of rx30day distributions. For example, it was found that simulations without BC have larger inter-climatological variability during the historical era than their bias-corrected twin ensemble; this relative behavior is echoed for interannual variability, which we directly quantified as a wet bias relative to observations. Under the assumption that the 100-year storm is unrealistically variable, this implies that unrealistic regional climate model behavior is delaying the time-of-emergence for externally forced signals until much later in the century. The 2040-2050 return frequency change and intensification of the rx30day event also has practical implications. Water

infrastructure takes decades to amass the necessary capital and then design, permit, and construct (Cosgrove & Loucks, 2015). It is then expected to be resilient to design storms for decades after its construction. If the hydroclimatic assumptions of those design storms (e.g., rx30day events) rapidly change after the construction of the infrastructure, it will be stress tested more regularly than was anticipated. Therefore, historical assumptions in rx30day events could become less reliable in the next few decades, stressing the need for carefully designed climate model experiments to help ensure that future water infrastructure is designed with appropriate design storms that are more representative of the hydroclimates in which they will exist and operate within.

Taken from a famous scientist, “the more one needs to bias correct, the less useful the practice becomes.” (Dave Pierce, 2024; personal communication) The transmitted ESM biases referenced here are surprisingly similar amongst code bases as presented in (Rahimi et al., 2024) for structurally different CMIP-class climate simulations, leading to a consistent distortion of the statistics of dynamically downscaled precipitation. It is therefore quite possible that the results of previous studies, which focused on extreme precipitation events across California without considering the problem of ESM bias transmission, may have skewed outcomes in terms of their projected hydrometeorological shifts associated with storm sequences under warming, particularly when ESMs were dynamically downscaled to quantify these shifts (Huang et al., 2021; Huang & Swain, 2022). On the other hand, bias correction of the ESM boundary conditions breaks the physical consistency between variables used to drive regional climate models, but the degree to which this is of consequence for dynamical downscaling has yet to be formally tested. We note however, that as applied here, BC is not introducing new uncertainties comparable in their magnitude to traditionally considered sources of climate uncertainty (Risser et al., 2024, 2025). Further, this study highlights the importance of internal variability given that the strongest storm from our 6,100-year ensemble occurred in the 1960s and that even a 10-member ensemble can create multi-decadal trends in return levels that have an opposite sign relative to a larger 100-member ensemble. Future research should focus on a more holistic analysis of the surface hydrometeorology, including runoff, snow, evaporation, and soil moisture, for compound flooding events of differing durations and across different states, to provide estimates of infrastructure-defining weather events on scales of decision relevance to support community partners. This is important as precipitation magnitude alone does not dictate the largest flood events in California history (e.g., the 1997 New Year’s flood event; (Rhoades et al., 2023, 2024)).

Acknowledgments

We acknowledge the immense computational support for this enterprise through the NCAR-Wyoming Supercomputing Center (NWSC) under the auspices of the Computational and Information Systems Laboratory (CISL). We also thank Andrew Newman, Romain Maendly, and the science team at California's Department of Water Resources, who provided important context which shaped the project. We also wish to thank Christine Albano and Daniel Swain of the Desert Research Institute whose feedback also helped to guide the project into its final form. Finally, we acknowledge the TORNERDO consortium which is all powerful.

Funding: SR, BB, AH, and LH acknowledge the financial support from a National Oceanic & Atmospheric Administration MAPP-NIDIS grant, “Towards predicting drought and subsequent water resource challenges at landscape-resolving scales across the western U.S.” (award no. NA230AR4310633). AR and MDR acknowledge the Department of Energy’s “The Calibrated and Systematic Characterization, Attribution and Detection of Extremes (CASCADE)” Science Focus Area (award no. DE-AC02-05CH11231) and the “An Integrated Evaluation of the Simulated Hydroclimate System of the Continental US” (HyperFACETS) project (award no. DE-SC0016605), and WL, TC, and SR acknowledge the National Science Foundation-supported project, “Wyoming Anticipating the Climate-Water Transition (WY-ACT)” (award no. 2149105). SR, JN, AH, and BB also acknowledge funding from the state of California via the California Energy Commission, “Development of Climate

Projections for California and Identification of Priority Projections” (no. EPC-20-006). FL and AD acknowledge funding from ESTCP project “Fit for purpose” (no. W912HQ-23-C-0002).

Open Research

All data are available in the main text or the supplementary materials. The versions of WRF used in this study, a sample Jupyter notebook, their attendant files, and the geography files are archived with Zenodo in an open DOI (<https://zenodo.org/records/10871230>) subject to a Creative Commons License version 4 (Rahimi & Huang, 2024b). Another DOI comprising the WUS-D3 is also available at <https://zenodo.org/records/10635867>. All downscaled data for the Western US Dynamically Downscaled Dataset (WUS-D3), including the full 6-hourly WRF datastream (Tier 1), hourly data for select land-surface variables (Tier 2), and a daily post-processed datastream (Tier 3) are located in an open-data bucket on Amazon S3 and are described at <https://registry.opendata.aws/wrf-cmip6/>. These data are completely open and free to the public. As recommended in the document, these data are most easily downloaded when using Amazon Web Service’s (AWS’) Command Line Interface (CLI) or with wget. An example is presented in the technical access and usage document. Individualized preprocessing codes were developed to create the intermediate binary files for each ESM before ingestion into WRF. Additionally, we provide several ESMs’ binary files for using in WRF downscaling in our AWS S3 bucket. Specifically, all ten downscale-able members of the CESM2 Large Ensemble from 1850-2100 have their 6-hourly binaries archived along with those of TaiESM1 r1i1p1f1, MPI-ESM1-2-hR r3i1p1f1/r7i1p1f1, ACCESS-CM2 r4i1p1f1, MIROC6 r1i1p1f1, NorESM2-MM r1i1p1f1, EC-Earth3-Veg r1i1p1f1, and EC-Earth3 r1i1p1f1 for community use.

Conflict of Interest Disclosure

The authors declare there are no conflicts of interest for this manuscript.

Author contributions

Conceptualization: SR, AH
 Methodology: SR, MDR, AR, CWT, LH
 Investigation: SR, MDR, AR, CWT, BB
 Visualization: SR, AR, MDR
 Supervision: SR, AH, FL, AR, MDR
 Writing—original draft: SR, ET
 Writing—review & editing: SR, MDR, AR, BB, ET, AH, JN, FL, AD, ZJL, WL, TC

References

- Bi, D., Dix, M., Marsland, S., O’Farrell, S., Sullivan, A., Bodman, R., et al. (2020). Configuration and spin-up of ACCESS-CM2, the new generation Australian Community Climate and Earth System Simulator Coupled Model. *Journal of Southern Hemisphere Earth Systems Science*, 70(1), 225–251. <https://doi.org/10.1071/ES19040>
- Brandt, W. T., Haleakala, K., Hatchett, B. J., & Pan, M. (2022). A Review of the Hydrologic Response Mechanisms During Mountain Rain-on-Snow. *Frontiers in Earth Science*, 10. <https://doi.org/10.3389/feart.2022.791760>
- Bruyère, C. L., Done, J. M., Holland, G. J., & Fredrick, S. (2014). Bias corrections of global models for regional climate simulations of high-impact weather. *Climate Dynamics*, 43(7), 1847–1856. <https://doi.org/10.1007/s00382-013-2011-6>
- Cosgrove, W. J., & Loucks, D. P. (2015). Water management: Current and future challenges and research directions. *Water Resources Research*, 51(6), 4823–4839. <https://doi.org/10.1002/2014WR016869>

- Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., et al. (2020). The Community Earth System Model Version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, 12(2), e2019MS001916. <https://doi.org/10.1029/2019MS001916>
- Dettinger, M. D., Martin Ralph, F., Abel, M. R., Das, T., Neiman, P., Cox, D., et al. (2012). Design and quantification of an extreme winter storm scenario for emergency preparedness and planning exercises in California. *Natural Hazards*, 60(3), 1085–1111. <https://doi.org/10.1007/s11069-011-9894-5>
- Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arneth, A., Arsouze, T., et al. (2022, June). *The EC-Earth3 Earth System Model for the Climate Model Intercomparison Project 6*. <https://doi.org/10.5194/gmd-2020-446>
- England Jr., J. F., Cohn, T. A., Faber, B. A., Stedinger, J. R., Jr, W. O. T., Veilleux, A. G., et al. (2018). *Guidelines for determining flood flow frequency — Bulletin 17C* (No. 4-B5). *Techniques and Methods*. U.S. Geological Survey. <https://doi.org/10.3133/tm4B5>
- Fischer, E. M., Sippel, S., & Knutti, R. (2021). Increasing probability of record-shattering climate extremes. *Nature Climate Change*, 11(8), 689–695. <https://doi.org/10.1038/s41558-021-01092-9>
- Gutjahr, O., Putrasahan, D., Lohmann, K., Jungclaus, J. H., von Storch, J.-S., Brüggemann, N., et al. (2019). Max Planck Institute Earth System Model (MPI-ESM1.2) for the High-Resolution Model Intercomparison Project (HighResMIP). *Geoscientific Model Development*, 12(7), 3241–3281. <https://doi.org/10.5194/gmd-12-3241-2019>
- Hall, A., Rahimi, S., Norris, J., Ban, N., Siler, N., Leung, L. R., et al. (2024). An Evaluation of Dynamical Downscaling Methods Used to Project Regional Climate Change. *Journal of Geophysical Research: Atmospheres*, 129(24), e2023JD040591. <https://doi.org/10.1029/2023JD040591>
- Hawkins, E., & Sutton, R. (2009). The Potential to Narrow Uncertainty in Regional Climate Predictions. *Bulletin of the American Meteorological Society*, 90(8), 1095–1108. <https://doi.org/10.1175/2009BAMS2607.1>
- Heggli, A., Hatchett, B., Schwartz, A., Bardsley, T., & Hand, E. (2022). Toward snowpack runoff decision support. *iScience*, 25(5), 104240. <https://doi.org/10.1016/j.isci.2022.104240>
- Holland, G., Done, J., Bruyere, C., Cooper, C. K., & Suzuki, A. (2010). Model Investigations of the Effects of Climate Variability and Change on Future Gulf of Mexico Tropical Cyclone Activity. Presented at the Offshore Technology Conference, OnePetro. <https://doi.org/10.4043/20690-MS>
- Houze Jr., R. A. (2012). Orographic effects on precipitating clouds. *Reviews of Geophysics*, 50(1). <https://doi.org/10.1029/2011RG000365>
- Huang, X., & Swain, D. L. (2022). Climate change is increasing the risk of a California megaflood. *Science Advances*, 8(32), eabq0995. <https://doi.org/10.1126/sciadv.abq0995>
- Huang, X., Swain, D. L., & Hall, A. D. (2021). Future precipitation increase from very high resolution ensemble downscaling of extreme atmospheric river storms in California. *Science Advances*, 6(29), eaba1323. <https://doi.org/10.1126/sciadv.aba1323>
- Ingram, B. L. (2013, January 1). California Megaflood: Lessons from a Forgotten Catastrophe. Retrieved April 12, 2025, from <https://www.scientificamerican.com/article/atmospheric-rivers-california-megaflood-lessons-from-forgotten-catastrophe/>
- Kelley, M., Schmidt, G. A., Nazarenko, L. S., Bauer, S. E., Ruedy, R., Russell, G. L., et al. (2020). GISS-E2.1: Configurations and Climatology. *Journal of Advances in Modeling Earth Systems*, 12(8), e2019MS002025. <https://doi.org/10.1029/2019MS002025>
- Kim, Y., Evans, J. P., & Sharma, A. (2023a). Correcting biases in regional climate model boundary variables for improved simulation of high-impact compound events. *iScience*, 26(9), 107696. <https://doi.org/10.1016/j.isci.2023.107696>
- Kim, Y., Evans, J. P., & Sharma, A. (2023b). Multivariate bias correction of regional climate model boundary conditions. *Climate Dynamics*, 61(7), 3253–3269. <https://doi.org/10.1007/s00382-023-06718-6>

- Li, L., Yu, Y., Tang, Y., Lin, P., Xie, J., Song, M., et al. (2020). The Flexible Global Ocean-Atmosphere-Land System Model Grid-Point Version 3 (FGOALS-g3): Description and Evaluation. *Journal of Advances in Modeling Earth Systems*, 12(9), e2019MS002012. <https://doi.org/10.1029/2019MS002012>
- Lundquist, J., Hughes, M., Gutmann, E., & Kapnick, S. (2019). Our Skill in Modeling Mountain Rain and Snow is Bypassing the Skill of Our Observational Networks. *Bulletin of the American Meteorological Society*, 100(12), 2473–2490. <https://doi.org/10.1175/BAMS-D-19-0001.1>
- Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., et al. (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO₂. *Journal of Advances in Modeling Earth Systems*, 11(4), 998–1038. <https://doi.org/10.1029/2018MS001400>
- Patricola, C. M., & Wehner, M. F. (2018). Anthropogenic influences on major tropical cyclone events. *Nature*, 563(7731), 339–346. <https://doi.org/10.1038/s41586-018-0673-2>
- Patricola, C. M., Wehner, M. F., Bercos-Hickey, E., Maciel, F. V., May, C., Mak, M., et al. (2022). Future changes in extreme precipitation over the San Francisco Bay Area: Dependence on atmospheric river and extratropical cyclone events. *Weather and Climate Extremes*, 36, 100440. <https://doi.org/10.1016/j.wace.2022.100440>
- Pierce, D. W., Cayan, D. R., & Thrasher, B. L. (2014). Statistical Downscaling Using Localized Constructed Analogs (LOCA). *Journal of Hydrometeorology*, 15(6), 2558–2585. <https://doi.org/10.1175/JHM-D-14-0082.1>
- Pierce, D. W., Cayan, D. R., Feldman, D. R., & Risser, M. D. (2023). Future Increases in North American Extreme Precipitation in CMIP6 Downscaled with LOCA. *Journal of Hydrometeorology*, 24(5), 951–975. <https://doi.org/10.1175/JHM-D-22-0194.1>
- Pontoppidan, M., Kolstad, E. W., Sobolowski, S. P., Sorteberg, A., Liu, C., & Rasmussen, R. (2019). Large-scale regional model biases in the extratropical North Atlantic storm track and impacts on downstream precipitation. *Quarterly Journal of the Royal Meteorological Society*, 145(723), 2718–2732. <https://doi.org/10.1002/qj.3588>
- Porter, K., Wein, A., Alpers, C. N., Baez, A., Barnard, P. L., Carter, J., et al. (2011). Overview of the ARkStorm scenario (No. 2010–1312). *Open-File Report*. U.S. Geological Survey. <https://doi.org/10.3133/ofr20101312>
- Rahimi, S., Huang, L., Norris, J., Hall, A., Goldenson, N., Risser, M., et al. (2024). Understanding the Cascade: Removing GCM Biases Improves Dynamically Downscaled Climate Projections. *Geophysical Research Letters*, 51(9), e2023GL106264. <https://doi.org/10.1029/2023GL106264>
- Rhoades, A. M., Jones, A. D., Srivastava, A., Huang, H., O'Brien, T. A., Patricola, C. M., et al. (2020). The Shifting Scales of Western U.S. Landfalling Atmospheric Rivers Under Climate Change. *Geophysical Research Letters*, 47(17), e2020GL089096. <https://doi.org/10.1029/2020GL089096>
- Rhoades, A. M., Zarzycki, C. M., Inda-Diaz, H. A., Ombadi, M., Pasquier, U., Srivastava, A., et al. (2023). Recreating the California New Year's Flood Event of 1997 in a Regionally Refined Earth System Model. *Journal of Advances in Modeling Earth Systems*, 15(10), e2023MS003793. <https://doi.org/10.1029/2023MS003793>
- Rhoades, A. M., Zarzycki, C. M., Hatchett, B. J., Inda-Diaz, H., Rudisill, W., Bass, B., et al. (2024). Anticipating how rain-on-snow events will change through the 21st century: lessons from the 1997 new year's flood event. *Climate Dynamics*, 62(9), 8615–8637. <https://doi.org/10.1007/s00382-024-07351-7>
- Risser, M. D., Rahimi, S., Goldenson, N., Hall, A., Lebo, Z. J., & Feldman, D. R. (2024). Is Bias Correction in Dynamical Downscaling Defensible? *Geophysical Research Letters*, 51(10), e2023GL105979. <https://doi.org/10.1029/2023GL105979>
- Risser, M. D., Feldman, D. R., Boos, W. R., & Rahimi, S. (2025). Upper bounds for 21st-century surface air temperatures in the Western United States. *Environmental Research Letters*, 20(6), 064043. <https://doi.org/10.1088/1748-9326/adda62>

- Rodgers, K. B., Lee, S.-S., Rosenbloom, N., Timmermann, A., Danabasoglu, G., Deser, C., et al. (2021). Ubiquity of human-induced changes in climate variability. *Earth System Dynamics*, 12(4), 1393–1411. <https://doi.org/10.5194/esd-12-1393-2021>
- S  f  rian, R., Nabat, P., Michou, M., Saint-Martin, D., Voldoire, A., Colin, J., et al. (2019). Evaluation of CNRM Earth System Model, CNRM-ESM2-1: Role of Earth System Processes in Present-Day and Future Climate. *Journal of Advances in Modeling Earth Systems*, 11(12), 4182–4227. <https://doi.org/10.1029/2019MS001791>
- Seland,   ., Bentsen, M., Olivi  , D., Toniazzo, T., Gjermundsen, A., Graff, L. S., et al. (2020). Overview of the Norwegian Earth System Model (NorESM2) and key climate response of CMIP6 DECK, historical, and scenario simulations. *Geoscientific Model Development*, 13(12), 6165–6200. <https://doi.org/10.5194/gmd-13-6165-2020>
- Sellar, A. A., Walton, J., Jones, C. G., Wood, R., Abraham, N. L., Andrejczuk, M., et al. (2020). Implementation of U.K. Earth System Models for CMIP6. *Journal of Advances in Modeling Earth Systems*, 12(4), e2019MS001946. <https://doi.org/10.1029/2019MS001946>
- Skamarock, C., Klemp, B., Dudhia, J., Gill, O., Liu, Z., Berner, J., et al. (2019). A Description of the Advanced Research WRF Model Version 4. <https://doi.org/10.5065/1dfh-6p97>
- Skamarock, W. C., & Klemp, J. B. (2008). A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *Journal of Computational Physics*, 227(7), 3465–3485. <https://doi.org/10.1016/j.jcp.2007.01.037>
- Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., et al. (2019). The Canadian Earth System Model version 5 (CanESM5.0.3). *Geoscientific Model Development*, 12(11), 4823–4873. <https://doi.org/10.5194/gmd-12-4823-2019>
- Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., et al. (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geoscientific Model Development*, 12(7), 2727–2765. <https://doi.org/10.5194/gmd-12-2727-2019>
- Wang, Y.-C., Hsu, H.-H., Chen, C.-A., Tseng, W.-L., Hsu, P.-C., Lin, C.-W., et al. (2021). Performance of the Taiwan Earth System Model in Simulating Climate Variability Compared With Observations and CMIP6 Model Simulations. *Journal of Advances in Modeling Earth Systems*, 13(7), e2020MS002353. <https://doi.org/10.1029/2020MS002353>

Table S1 inserted for the purposes of syncing citations in this document:

Model	Variants	Center & Country	Citation
*CESM2	r11i1p1f1	National Center for Atmospheric Research; United States	(Danabasoglu et al., 2020)
*MPI-ESM1-2-LR	r7i1p1f1	Max Planck Institute; Germany	(Mauritsen et al., 2019)
*CNRM-ESM2-1	r1i1p1f2	Centre Europ��en de Recherche et de Formation Avanc��e en Calcul Scientifique; France	(S��f��rian et al., 2019)
*FGOALS-g3	r1i1p1f1	Chinese Academy of Sciences; China	(Li et al., 2020)
*EC-Earth3-Veg	r1i1p1f1	EC-Earth Consortium; Sweden	(D��scher et al., 2019)

			2022)
*ACCESS-CM2	r5i1p1f1	Commonwealth Scientific and Industrial Research Organization; Australia	(Bi et al., 2020)
*CanESM5	r1i1p2f1	Canadian Climate Center; Canada	(Swart et al., 2019)
*UKESM1-0-LL	r2i1p1f2	Met Office Hadley Centre; United Kingdom	(Sellar et al., 2020)
*EC-Earth3	r1i1p1f1	EC-Earth Consortium; Sweden	(Döscher et al., 2022)
*MIROC6	r1i1p1f1	Japan Agency for Marine-Earth Science and Technology; Japan	(Tatebe et al., 2019)
*TaiESM	r1i1p1f1	Research Center for Environmental Changes; Taiwan	(Wang et al., 2021)
*MPI-ESM1-2-HR	r7i1p1f1, r3i1p1f1	Max Planck Institute; Germany	(Gutjahr et al., 2019)
*GISS-E2-1-G	r1i1p1f2	National Aeronautic and Space Administration; United States	(Kelley et al., 2020)
*NorESM2-MM	r1i1p1f1	NorESM Climate Modeling Consortium; Norway	(Seland et al., 2020)
CESM2-LE	1011, 1031, 1051, 1071, 1091, 1111, 1131, 1151, 1171, 1191	National Center for Atmospheric Research; IBS Center for Climate Physics; United States and Republic of Korea	(Rodgers et al., 2021)