# Seasonally Anchored Bias Correction of CMIP5 Hydrological Simulations

Sierks Michael[1], Pierce David[1], Chapman William[2], and Ralph Fred Martin[1]

[1]Scripps Institution of Oceanography
[2]University of California San Diego

November 16, 2022

**Abstract**

Robust and reliable projections of future streamflow are essential to create more resilient water resources, and such projections must first be bias corrected. Standard bias correction techniques are applied over calendar-based time windows and leverage statistical relations between observed and simulated data to adjust a given simulated datapoint. Motivated by a desire to connect the statistical process of bias correction to the underlying dynamics in hydrologic models, we introduce a novel windowing technique for projected streamflow wherein data are windowed based on hydrograph-relative time, rather than Julian day. We refer to this method as 'seasonally anchored'. Four existing bias correction methods, each using both the standard day-of-year and the novel windowing technique, are applied to daily streamflow simulations driven by 10 global climate models across a diverse subset of six watersheds in California to investigate how these methods alter the model climate change signals. Among the methods, only PresRat preserves projected annual streamflow changes, and does so for both windowing techniques. The seasonally anchored window PresRat reduces the ensemble bias by a factor of two compared to quantile mapping (Qmap), cumulative distribution function transform (CDFt), and equidistant quantile matching (EDCDFm) methods. For wet season flows, PresRat with seasonally anchored windowing best preserves the original model change over the entire distribution, particularly at the highest quantiles, and the other three methods show improved performance using the novel windowing method. Concerning temporal shifts in seasonality, PresRat and CDFt preserve the original model signals with both the novel and standard windowing methods.

# Seasonally Anchored Bias Correction of CMIP5 Hydrological Simulations

M. Sierks,[a] D.W. Pierce,[a] W.E. Chapman,[b] F.M. Ralph[a]

[a] *Scripps Institution of Oceanography, UC San Diego, La Jolla, CA, USA*

[b] *National Center for Atmospheric Research, Boulder, CO, USA*

*Corresponding author*: M. Sierks, msierks@ucsd.edu

ABSTRACT

Robust and reliable projections of future streamflow are essential to create more resilient water resources, and such projections must first be bias corrected. Standard bias correction techniques are applied over calendar-based time windows and leverage statistical relations between observed and simulated data to adjust a given simulated datapoint. Motivated by a desire to connect the statistical process of bias correction to the underlying dynamics in hydrologic models, we introduce a novel windowing technique for projected streamflow wherein data are windowed based on hydrograph-relative time, rather than Julian day. We refer to this method as ‘seasonally anchored’. Four existing bias correction methods, each using both the standard day-of-year and the novel windowing technique, are applied to daily streamflow simulations driven by 10 global climate models across a diverse subset of six watersheds in California to investigate how these methods alter the model climate change signals. Among the methods, only PresRat preserves projected annual streamflow changes, and does so for both windowing techniques. The seasonally anchored window PresRat reduces the ensemble bias by a factor of two compared to quantile mapping (Qmap), cumulative distribution function transform (CDFt), and equidistant quantile matching (EDCDFm) methods. For wet season flows, PresRat with seasonally anchored windowing best preserves the original model change over the entire distribution, particularly at the highest quantiles, and the other three methods show improved performance using the novel windowing method. Concerning temporal shifts in seasonality, PresRat and CDFt preserve the original model signals with both the novel and standard windowing methods.

## SIGNIFICANCE STATEMENT

Robust and reliable projections of future streamflow are essential if we are to create more resilient water resources, and such model data must first be bias corrected. We introduce a novel windowing technique to be used in streamflow bias correction methods which improves the preservation of the original model climate change signal. Crucially, these improvements are realized not only for the water year mean signal, which is important as it relates to the total volume of water flowing through the river over the course of the year, but is also true for both low and high streamflow events which have an outsized imprint on California’s hydroclimate, water resources, and ecosystems.

# 1. Introduction

*a. Climate Change Impacts*

Fueled by climate change, rising temperatures and declining snowpacks have revealed that the ways in which the American West has managed water for the past 75 years is insufficient to sustainably meet projected demands (Barnett & Pierce, 2009; Rajagopalan et al., 2009; Udall & Overpeck, 2017). It is equally apparent that river basins, watersheds, and reservoir drainage areas will not be impacted uniformly (Das et al., 2011; Kalra et al., 2008; Mote et al., 2005, 2018). The response of water supply-relevant variables to climate change, such as annual streamflow, total precipitation, or the extent of April 1st snowpack, will be functions of factors like shifting large-scale weather patterns, elevation, topographic aspects, vegetation, and the amplitude of season temperature cycles (Gonzalez et al., 2018; He et al., 2019; Huning & AghaKouchak, 2018; Pierce & Cayan, 2013). As a result, the potential impacts of climate change on water management, riparian health, and associated mitigation or adaptation strategies need to be examined on local scales and on a case-by-case basis.

*b. Downscaling and Bias Correction*

Future climate projections from global climate models (GCMs) are a key tool for estimating likely impacts of climate change on future water availability, but due to limited spatial resolution (typically no finer than 100 km) are insufficient for studying changes at the local-scale of river basins and heterogeneous hydrologic processes (Fowler et al., 2007; Hewitson et al., 2014; Salathé, 2003). Further, biases in the GCMs, arising from model limitations like subgrid parameterizations of cloud microphysics and poorly resolved topography at the native GCM grid scale, can result in distorted projected climate impacts (see discussion in Maraun et al., 2017). Prior to being used in most applications, therefore, climate projections must be downscaled and bias corrected. Implemented through either a 'statistical' or 'dynamical' approach, downscaling techniques interpolate smaller-spatial scale features by combining coarser GCM output with higher-resolution observations, topography, and dynamics to produce projections with resolutions on the order of 10s km. Either as part of the downscaling process or done subsequently, bias correction removes systematic errors in the GCM with the goal of retaining the raw GCM climate change signal.

While some climate change planning projects may be satisfied simply by downscaled and bias corrected GCM output (e.g., temperature, precipitation), many require the use of a land-surface model to produce quantities such as streamflow or soil moisture. Even if downscaled and bias corrected GCM output is used to drive the land surface model, streamflow projections often need 'secondary' bias corrections before they are used for planning due to biases introduced within the land-surface model. Though useful, it should be understood that bias correction is a statistical technique and thus is not able to discern between physical processes responsible for a given data-point or a broader climate change-imposed trend (Maraun et al., 2017).

### 1) Windowing / Bias Correction goals

Although the goal of the windowing and bias correction process is to remove systematic biases while retaining the signal of change from the driving climate model, many bias correction methods alter the model-predicted change for unphysical reasons (Hagemann et al., 2011; Maraun, 2013; Maurer & Pierce, 2014; Pierce et al., 2013). As a result, the application of different bias correction methods to identical datasets will yield varied future projections (Maurer & Pierce, 2014; Teutschbein & Seibert, 2012). Simplified, many bias correction methods establish a correction function that maps a model variable's empirical distribution over a historical period to that of an observed variable's distribution over a historical period. Upon applying this correction function to the model historical data, the bias corrected data is mapped onto the observed variable's distribution thus removing any systematic biases over the historical period. Though the definition of the transfer function varies among methods (Pierce et al. 2015), nearly all methods make assumptions of stationarity, i.e., that model biases occurring in the model's historical period also apply to future model periods. Though commonly adopted, stationarity is not guaranteed, and it is therefore crucial to understand under what conditions it breaks down (e.g., with changes in large-scale circulation).

Because model biases typically vary by season, correction functions are usually developed for specific calendar-based "seasons" (e.g., Thrasher et al. 2012). Specifically, distributions of the observed, historical- and future- model datasets are developed by subsetting in time, either by individual months or by taking some rolling window of fixed width (see Pierce et al., 2015 for a deeper discussion). Although cumulative distribution functions (CDFs) can be generated either empirically (as in the methods examined here) or parametrically, nonparametric methods have yielded higher skill in reducing systemic errors for precipitation (Gudmundsson et al., 2012).

In the context of climate change, for variables whose seasonal cycles are predominately affected by changing amplitudes rather than shifts in seasonality (such as temperature or even precipitation) it may be fair to assume that historical biases for January data can be removed directly from future January data. Consider, though, a variable whose climate change signal is characterized in large part by a temporal shift in its climatology, such as snowfed streamflow. Historically in the western U.S., mountainous rivers experience peak streamflow during the spring (later for higher elevation sites) as the snowpack begins to melt (Serreze et al., 1999). Towards the end of the 21st century, reduced in volume and melting earlier, projected snowpack declines result in peak streamflow shifting significantly earlier into the season (Noah Knowles & Cronkite-Ratcliff, 2018; Udall & Overpeck, 2017). If we were to apply a calendar-fixed window to bias correct streamflow (e.g., comparing data from the month of April, historical to future), it is possible that historical bias corrections of peak or rising-limb streamflow data will be applied to future streamflow data occurring well into the receding limb, thus applying corrections to and from different streamflow regimes and controlled by different physical processes.

*c. Purpose of Paper*

Motivated by the desire to move towards a ‘process-aware’ method of statistical bias correction and the inability of fixed calendar- windowing to account for processes that shift seasonality under future climate scenarios, this paper introduces a new ‘seasonally anchored’ windowing approach that, when applied to existing statistical methods, improves the preservation of original model – used henceforth to refer to hydrologic model output driven by downscaled and bias corrected GCM data – climate change signals in projections of streamflow. We evaluate the performance of several published bias correction methods, using the standard Julian day anchored framework and our seasonally anchored windowing techniques, with respect to their ability to reduce bias while preserving key metrics of climate change from the original model. Specifically, we investigate the preservation of: 1) original model projected changes in water year mean streamflow, 2) original model projected change across all deciles of wet season streamflow, and 3) original model change in seasonality as measured by change in date of peak streamflow.

The paper is structured in the following manner. In Sections 2 and 3, we describe the study domain and the rivers included in this work, and detail the observed and model data sources used to evaluate the various bias correction methods. In Section 4, we detail the seasonally anchored windowing technique and describe the four bias correction methods compared. Section 5 describes the performance of the various methods over both the historical and future climate periods. Lastly, Section 6 summarizes and discusses these results.

## 2. Study Domain

The hydroclimate of California is characterized by distinct wet and dry seasons and is punctuated by high interannual variability (Dettinger et al., 2011). In fact, the presence or absence of just a few storms each year can determine the difference between drought conditions and sufficient water supply (Dettinger & Cayan, 2014). Moreover, interannual variability and dependence on just a few storms per year is expected to increase in the future, with model projections showing fewer wet days but more precipitation on the wet days that occur (Pierce et al., 2013). Because 1) there is a strong latitudinal gradient in the frequency of landfalling winter storms (Payne & Magnusdottir, 2014), and 2) interactions between low-level moisture flux and local orographic forcing is driving mechanism of California precipitation (Neiman et al., 2002), the complex terrain of coastal and inland ranges results in marked spatial heterogeneity in the hydroclimate. California's rivers and streams are as diverse as the landscapes that feed them, with flashy, ephemeral streams in low deserts and snow-fed perennial rivers in the high mountains, the latter of which are responsible for filling some of the nation's largest reservoirs. As projected climate change impacts for riparian environments differ dramatically across watersheds and stream types (Perry et al., 2015 and references therein), the diversity of California's waterways and robustness of its observational network make it an excellent testbed for our bias correction methods.

## 3. Data

Bias correction requires both historical (observed or reanalysis) and model data. Thus, the number of rivers eligible for examination are limited to those with both of the above datasets. Streams were selected based on the following criteria: 1) availability of routed streamflow projections from land-surface models driven by downscaled and bias corrected GCM data, 2) availability of at least 20 years of observational data, and 3) their inclusion enhance the representation of selected streams from along the continuum of rain- versus snow- dominated basins. Upon applying the above criteria, we choose 6 rivers for our case study (Figure 1, Table 1). The selected rivers span hydrologic characteristics of rain-, snow-, and mixed rain-and- snow dominated watersheds, allowing us to test the performance of our seasonally anchored windowing technique for bias correction on streams with and without large projected flow-seasonality changes. We emphasize here that our goal is not to produce an expansive dataset of bias-corrected streamflow, which requires a larger network of streams, though such an exercise will undoubtedly be useful.

Streamflow projections used in this study were obtained from Knowles & Cronkite-Ratcliff, 2018 (see Section 2.2 of Knowles and Cronkite-Ratcliff, 2018 for details). They use Localized Constructed Analogs (LOCA, Pierce et al., 2014) statistically downscaled GCM data to force the Variable Infiltration Capacity (VIC) hydrological model (Liang et al., 1994). Rather than use each of the 31 members in the Coupled Model Intercomparison Project (CMIP5; Taylor et al., 2012) ensemble, we focus our analysis on a subset of 10 GCMs chosen by the California DWR Climate Change Technical Advisory Group as providing passable simulations of the historical California hydroclimate (California Department of Water Resources Climate Change Technical Advisory Group, 2015, models listed in Table 2). Here, we restrict our analysis to the future climate relative concentration pathway (RCP) 8.5 since the climate change signal of shifting seasonality is more easily discerned in higher warming scenarios. However the results found here will apply to other emissions scenarios. Observational data for the six streams span water years (WYs) 1997-2019 and were obtained from the United States Geological Survey (USGS) and California Department of Water Resources (DWR) Data Exchange Center.

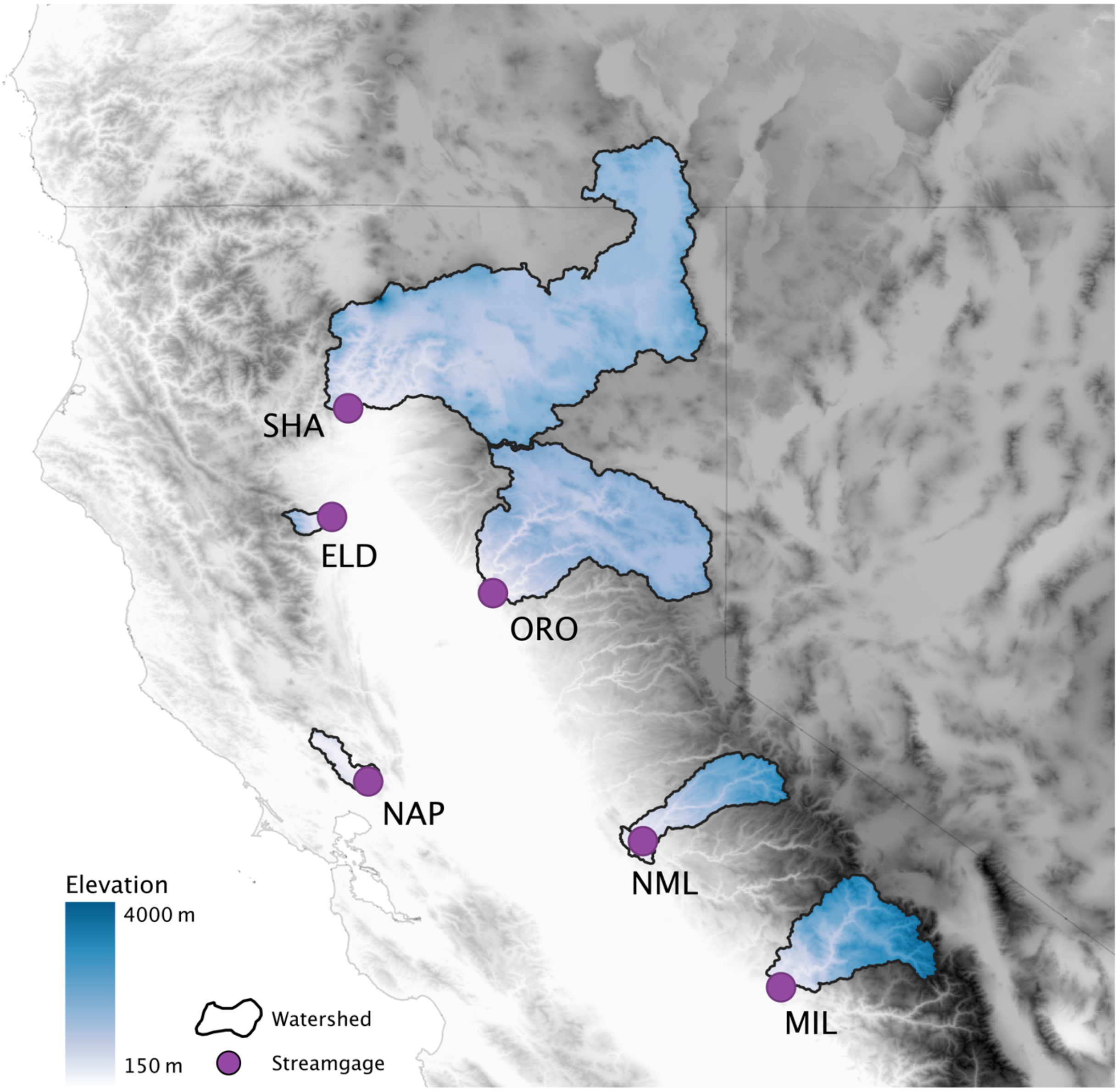



Fig. 1. Map of the study domain with watershed boundaries (black contour), gage locations (purple circle), and three-letter abbreviations for the 6 streams alongside elevation (color shade). Characteristics of each watershed are listed in the accompanying table.

| Stream | Gage ID | Drainage Area (km2) | Min, Max | Mean Elevation (m) | % Streamflow |
|---|---|---|---|---|---|

| | | | Elevation (m) | | Before April 1 (Observed) |
|---|---|---|---|---|---|
| Napa River | USGS, 1145800 | 550 | 7, 1161 | 247 | 87.7 |
| Elder Creek | USGS, 11379500 | 250 | 255, 1959 | 978 | 73.2 |
| Shasta Dam | CDEC, SHA | 18350 | 306, 4113 | 1435 | 62.0 |
| Oroville Dam | CDEC, ORO | 9350 | 240, 2635 | 1545 | 58.0 |
| New Melones Reservoir | CDEC, NML | 2550 | 160, 3381 | 1632 | 43.6 |
| Friant Dam (Millerton) | CDEC, MIL | 4250 | 157, 3954 | 2161 | 30.6 |

Table. 1. Name, gage identification and summary characteristics of each watershed. Streams are listed by their fraction of total water year streamflow occurring before April 1 (proxy used to indicate importance of snowmelt) in descending order with the snowiest watersheds listed in the final rows.

| Model Acronym | Model Source/Institution |
|---|---|
| ACCESS1.0 | Commonwealth Scientific and Industrial Research Organisation (CSIRO) and Bureau of Meteorology, Australia |
| CCSM4 | National Center for Atmospheric Research (NCAR), United States NCAR, United States |

| | |
|---|---|
| CESM1-BGC | National Center for Atmospheric Research (NCAR), United States NCAR, United States |
| CMCC-CMS | Centro Euro-Mediterraneo per I Cambiamenti Climatici |
| CNRM-CM5 | Centre National de Recherches Météorologiques, France |
| CanESM2 | Canadian Centre for Climate Modelling and Analysis, Canada |
| GFDL-CM3 | Geophysical Fluid Dynamics Laboratory (GFDL), Princeton, New Jersey, United States |
| HadGEM2-CC | Met Office Hadley Center, UK |
| HadGEM2-ES | Met Office Hadley Center, UK |
| MIROC5 | Atmosphere and Ocean Research Institute and NIES, Japan |

Table. 2. Selection of 10 GCMs from CMIP5 used in this work along with their originating institutions.

## 4. Methodology

While the following subsections describe the methodology in complete detail, we preface by briefly summarizing the algorithmic approach. First, for a given value that is to be bias corrected, we convert from Julian day to a hydrograph-relative time unit. This is done by locating its position (in time) relative to important climatological features (e.g., day of peak streamflow). Then, we map this point in ‘hydrograph-relative’ time onto all datasets (observed, historical GCM, future GCM) to identify hydrologically similar periods.

### *a. Climatological Hydrograph*

The first step in the new seasonally-anchored approach is to calculate climatological mean hydrographs for the (1) observed, (2) simulated-historical and (3) simulated-future flows. In the present study, the length of climatological periods varies from the observed (n=23 years, 1997-2019), simulated-historical (n=36 years, 1970-2005), and simulated-future flow series (n=31 years, e.g., 2069-2099). The observed period is limited by the period of available record of gaged streamflow. Climatological-average hydrographs are computed at several quantiles (discussed below), but to ease explanation of the process, we explain it for the 70$^{th}$ percentile (P70) first.

For a given Julian day, the 70$^{th}$ percentile streamflow value is estimated from a distribution containing all data from the climatological period within a 31-day centered window (e.g. at Julian day 185, use data from Julian days 170-200) similar to Thrasher et al., 2012. Once done for all days of the year, this array of length 365 is then smoothed by taking the mean of all points within the centered 31-day windows. The resulting smoothed array is the climatological hydrograph at the 70$^{th}$ percentile (Figure 2, top).

*b. Locating Seasonal Milestones*

Once the climatological hydrograph is calculated for a given percentile, milestones along the curve that correspond to key characteristic features of the stream are identified. Similar to previous work by Yarnell et al. (2015) we define our milestones as the Julian days when four climatological-hydrograph transitions occur (detailed below, and shown in Fig. 2): (1) minimum flow, (2) transition from a low-flow period to the rising limb, (3) maximum flow, and (4) transition from the receding limb to a low-flow period.

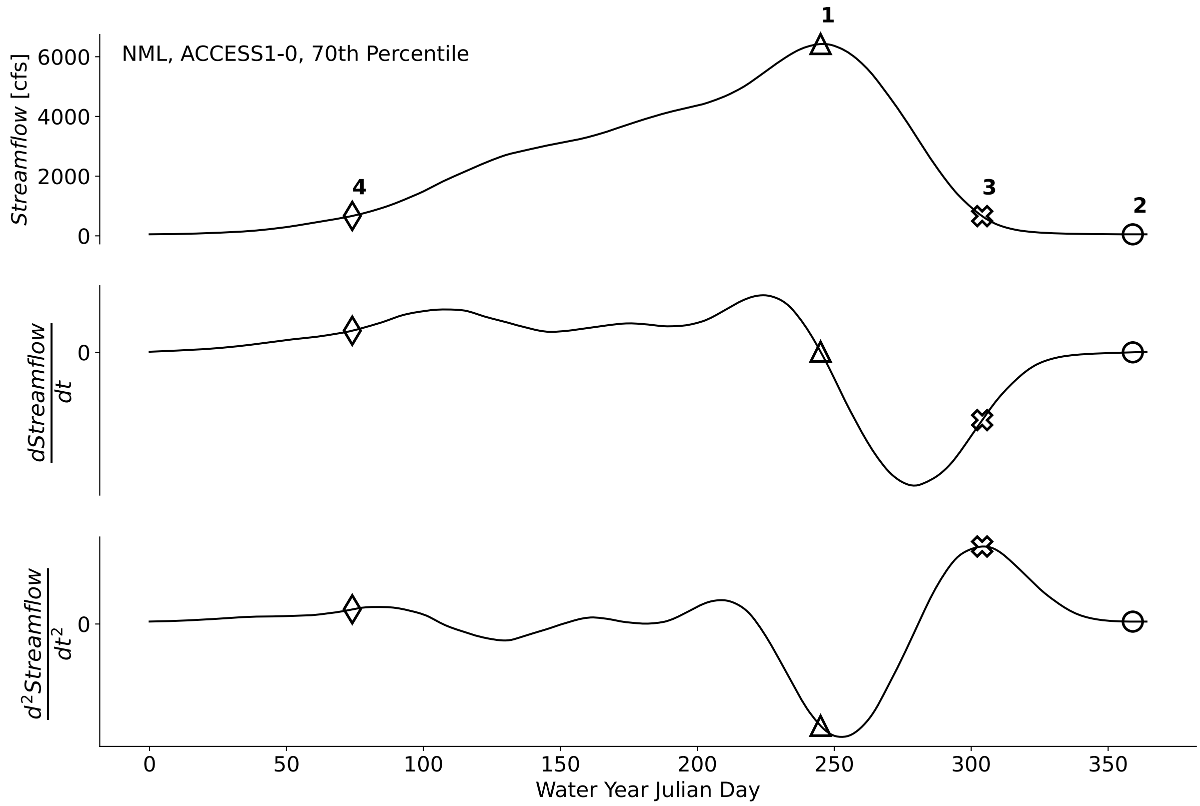



Fig. 2. Visual depiction of the algorithm used to identify climatological milestones for the seasonally anchored windowing method. Climatological data from the New Melones stream and ACCESS1.0 GCM is shown at the $70^{th}$ percentile to illustrate the method. Daily mean climatological streamflow (top), and the first and second derivatives of streamflow with respect to time (middle and bottom respectively) are plotted on against water year Julian day. Numerical annotation is used to indicate the workflow by which the four seasonal milestones are assigned: (1) Peak streamflow (triangle), (2) Minimum streamflow (circle), (3) Start of the dry season/end of receding limb (x), and (4) Start of the wet season/beginning of the rising limb (diamond).

### 1) Maximum and Minimum Streamflow

The milestone of peak flow occurs on the Julian day with the largest climatological streamflow and delineates the change from the rising to the falling limbs of the hydrograph (Fig. 2, triangle). Similarly, the milestone marking the minimum streamflow occurs on the Julian day with the lowest value of climatological streamflow (Fig. 2, circle). This is done for each of the three flow series. Although the vast majority of climatological hydrographs in this study are described by having a single peak, Appendix A describes how the identification of the peak streamflow milestone is handled for 'bimodal' hydrographs.

2) Start of Dry Season

Qualitatively, transitions between distinct streamflow regimes occur when the shape of the hydrograph changes rapidly. We can quantify the shape of the hydrograph by its 1st and 2nd derivatives with respect to time. The first derivative of streamflow ($\frac{dQ}{dt}$, Figure 2 middle) indicates where the stream is rising and falling. If we wish to pinpoint the end of the falling limb, we must only examine days when $\frac{dQ}{dt} < 0$. The second derivative with respect to time ($\frac{d^2Q}{dt^2}$, Figure 2 bottom) is used to identify when the hydrograph has maximum points of curvature, or in the language used above, when the shape of the hydrograph indicates an inflection point. To identify the milestone associated with the end of the falling limb (start of the dry season), we find the Julian day coinciding with the local maximum of $\frac{d^2Q}{dt^2}$ given that $\frac{dQ}{dt} < 0$. This is shown schematically in bottom subplot of Figure 2 by the 'x'. This is done for the observed and simulated-historical datasets. The process for the simulated-future dataset is describes in the following section.

3) Start of Wet Season

For mixed rain- and snow dominated rivers, there is larger variability in the onset of the wet For mixed rain- and snow dominated rivers, there is larger variability in the onset of the wet season than the onset of the dry season (Patterson et al., 2020) as the former is governed by precipitation, which has a high degree of interannual variability (Dettinger et al., 2011), whereas the latter is corresponds to the end of the snowmelt pulse (Stewart et al., 2005), which is driven by lower variability fields such as synoptic temperature advection and solar insolation (Cayan et al., 2001; Mioduszewski et al., 2015; Pederson et al., 2011). Therefore, we don't apply the same method as in the previous section. Rather, we assume the streamflow value associated with the 'start of the dry season' milestone delineates between baseflow and non-baseflow periods. Since the 'start of the wet season' marks the stream's departure from baseflow, the corresponding milestone is set at the Julian day when the climatological hydrograph increases above the streamflow value associated with the 'start of the dry season' milestone. This is shown schematically in top panel of Figure 2 by the diamond marker and is calculated for the observed and simulated-historical datasets.

Motivated by the importance of the low-flow period to ecosystem health (Hill et al., 1991; Petts, 1996; Poff et al., 1997; Richter et al., 1996), we deem this 'threshold value' (evaluated as flow above baseflow) to be characteristic of the stream and assume that the streamflow above baseflow value associated with the low-flow period does not change with time. That is, while baseflow levels may change from historical to future periods, if the historical low-flow period was delineated at 500 cfs above baseflow, then the future climate milestone for the 'start ('end') of the west season' will be located when the stream rises (falls) to 500 cfs above the future climate baseflow. This will yield a consistent comparison of how both the duration of the low-flow period and corresponding streamflow magnitudes change in future climate projections.

*c. Calculating the Mean Milestones Across Quantiles*

The milestone identification process is performed with hydrographs computed at every 5 percentage points between the 40$^{th}$ and 80$^{th}$ percentiles (P40-80). The final milestone locations for the climatological period are computed by taking the mean value of the milestone dates across all individual quantiles. The rationale for choosing to evaluate data between the 40$^{th}$ and 80$^{th}$ percentiles is two-fold. First, the final milestones should represent a broad range of climatological stream conditions, but extreme values—while only a small fraction of total data--can yield outlier milestones that are not representative of the entire distribution. Because of this, quantiles near the distribution tails are excluded from this final step of the milestone process – though they are accounted for during the windowing procedure described in Section 4.d. Second, the asymmetry, relative to the median, of the P40-80 range reflects the differing dynamics of high and low flow climatologies. High quantile flows in the early water year only occur when the synoptic environment is favorable for large precipitation events (typically October at the earliest). As a result, for higher quantile climatological hydrographs, the timing of the start of the wet season is more-or-less constrained to a narrow window of 1-3 months at the beginning of the water year. However, low quantile flows correspond to the absence of large storms and are not constrained in time. For drier years with few storms during the early WY, the start of the wet season can be pushed far into the water year and, if included, skew the mean value taken over all quantiles in P40-80 (see Fig. A2 and A3).

*d. Seasonally Anchored Windowing*

Using the milestones as reference points, we are able to bias correct a day's simulated flow in a way that acknowledges its position in the hydrograph, which ensures that the hydrologic processes at work on that day are accommodated. For example, a flow that occurs on the descending limb of the hydrograph will be bias corrected using the same parameters no matter whether that flow is in, e.g., June in the historical period or in, e.g., April in a future projection (Fig. 3, bottom). This differs from traditional bias correction techniques that are anchored by day-of-year, which means (in cases where the peak flow shifts earlier in the year) the bias correction parameters from the rising limb of the hydrograph in the historical period might be applied to values from the descending limb of the hydrograph in the future period (Fig. 3, top). Because window widths are based on the length of streamflow regime segments (e.g., rising limb), and those segments vary across the observed, historical-, and future GCM climatologies, we do not require the different datasets to have the same window widths. Before the actual bias correction step is done, the empirical distributions of daily streamflow data are cubic hermite spline interpolated to obtain the same length, following Pierce et al., 2015.

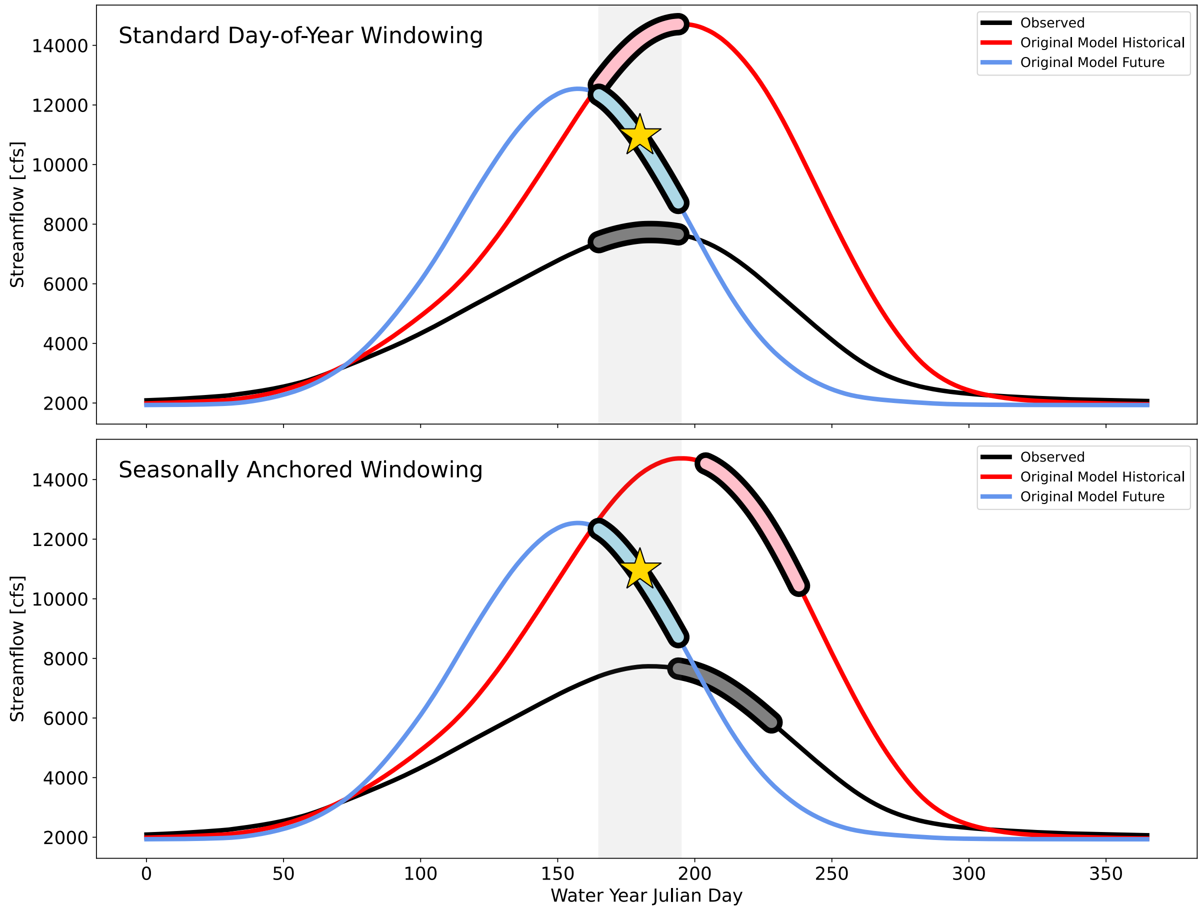



Fig. 3. Schematic highlighting the effect of windowing technique on which segments of the climatological hydrograph are used to bias correct a given model datapoint on the 180th day of the water year (yellow star). Each panel shows the climatological hydrographs in thin lines for the observed (black), historical original model (red), and end-of-century original model (blue) data. The segments of each hydrograph used in bias correction are plotted as bolded lines. Standard day-of-year windowing (top) uses data from each curve that falls within the centered 30-day period (grey shaded region). Seasonally anchored windowing (bottom) uses a centered 30-day window over the climatological period being corrected and then finds equivalent segments of the observed and historical hydrographs to use for bias correction.

### *e. Bias Correction Methods to be Compared*

To illustrate the effect of the windowing technique on bias corrected streamflow data, we examine the performance of 4 different bias correction methods using first the standard DOY windowing, and then seasonally anchored windowing. Similar to (Pierce et al., 2015), we apply the following techniques: PresRat (Pierce et al., 2015), CDF-transform (CDFt, Michelangeli et al., 2009), equidistant CDF matching (EDCDFm, Li et al., 2010), and quantile mapping (Qmap, Panofsky & Brier, 1968; Wood et al., 2002). For the sake of brevity, we do not discuss the details of each bias correction method and instead refer the reader to Section 3 of Pierce et al., 2015.

For the CDFt, EDCDFm, and Qmap methods, we adopt the standard practice of bias correcting data using a 30-day window centered around a given datapoint. The 30-day window enables the methods to represent the seasonal cycle, but such a narrow window is not suitable to correct extreme events. Because extreme values of precipitation can occur at any time during the wet season, the PresRat method from Pierce et al., 2015 does not use a fixed 30-day window like the other 3 methods and instead iteratively bias corrects data using windows of increasing width, providing better correction of extreme values. Rather than applying iterative bias correction in the version of PresRat used in this study, we develop an alternative method to balance the correction of extreme values (requiring a wide window) and the seasonal cycle (requiring a narrow window). Here, we vary the window width based on the 'extremity' of a value. Beginning with a 30-day window, we find the quantile location of the data point being corrected in its climatological distribution. If it falls between the $20^{th}$ and $80^{th}$ percentile, a 30-day window is used. Otherwise the window is expanded by 15-days on either side and the quantile location of the data is found again. If the data falls between the $10^{th}$ and $90^{th}$ percentile, the 60-day window is applied. If not, a 120-day window is used to bias correct the most extreme values. This reflects the fact that extreme events are by nature rare.

Because some methods of bias correction operate on fractional changes between a future and historical model period, they can be sensitive to small errors occurring at low values (Pierce et al., 2015). For this reason, prior to bias correction, we correct for any biases in the model baseflow by adding the difference between the observed and simulated baseflow values to all model data. This greatly improves the efficacy of the bias correction methods at lower flows while having very little impact at higher flows.

## 5. Results and Discussion

The efficacy of a bias correction method can be evaluated by its ability to remove systematic model biases while preserving desired climate change signals from the original model. In this case we choose to consider: 1) mean changes in streamflow magnitude over the entire water year, 2) changes in magnitude evaluated at high, medium, and low quantiles of the distribution, and 3) temporal shifts of seasonality (which may be small in some streams). The combination of bias correction method (QM, CDFt, EDCDFm, PresRat) and windowing technique (standard day-of-year or seasonally anchored) is evaluated by its ability to preserve the 3 quantities listed above. As the magnitude of streamflow varies substantially between California watersheds, working with normalized data allows for a more straightforward comparison of climate change signals. For this reason, we normalize any change between a simulation's future and historical values by a historical baseline (Equation 1) to give a percent change relative to the pre-climate change period.

$$\Delta = \mathbf{100} * \frac{\boldsymbol{Future - Historical}}{\boldsymbol{Historical}} \tag{1}$$

Using a standardized metric (Δ) helps us compare both the magnitude of change across a diverse subset of streams and the performance of bias correction methods in preserving this metric. In the following sections, we evaluate change between the future (also referred to as 'end-of-century') and historical periods defined as water years 2069-2099 and 1970-2005 respectively. Subsequent sections evaluate this change over both the entire water year and over the 'wet season', which we define as the timeframe spanning 1-month prior to the start of the wet milestone to 1-month after the end of the wet season milestone. This covers the period roughly from November to June (although it is subject to both watershed elevation and climatological era).

*a. Validation over the Historical Period*

Applied over the historical period, the PresRat, CDFt, and EDCDFm methods simplify to the quantile mapping method, with the exception of data off the endpoints of the historical distribution (see Section 3 of Pierce et al., 2015 for discussion). Therefore, when using a relatively narrow 30-day window, all methods are effective in correcting the historical GCM data so that it recreates the observed data's seasonal cycle. This is true for both the standard day-of-year windowing technique and the seasonally anchored windowing technique (not shown). Concerning the historical period, the only meaningful difference between the four bias correction methods results from the variable window width used in our PresRat method (Figure 4). Because the window width expands when correcting high-quantile data, large streamflow events occurring early in the wet season can be mapped onto observed streamflow values occurring up to 60 days later (as opposed to 15 days later in the standard method). This can result in slightly elevated mean flows during the transition from dry-to-wet season, driven by streamflow at high quantiles. However, as extreme precipitation and streamflow events can occur at any point between October-April, a variable window based on the quantile of the datum being corrected is more easily justifiable than the common fixed narrow (31-day) window when correcting hydrometeorological variables in the western US.

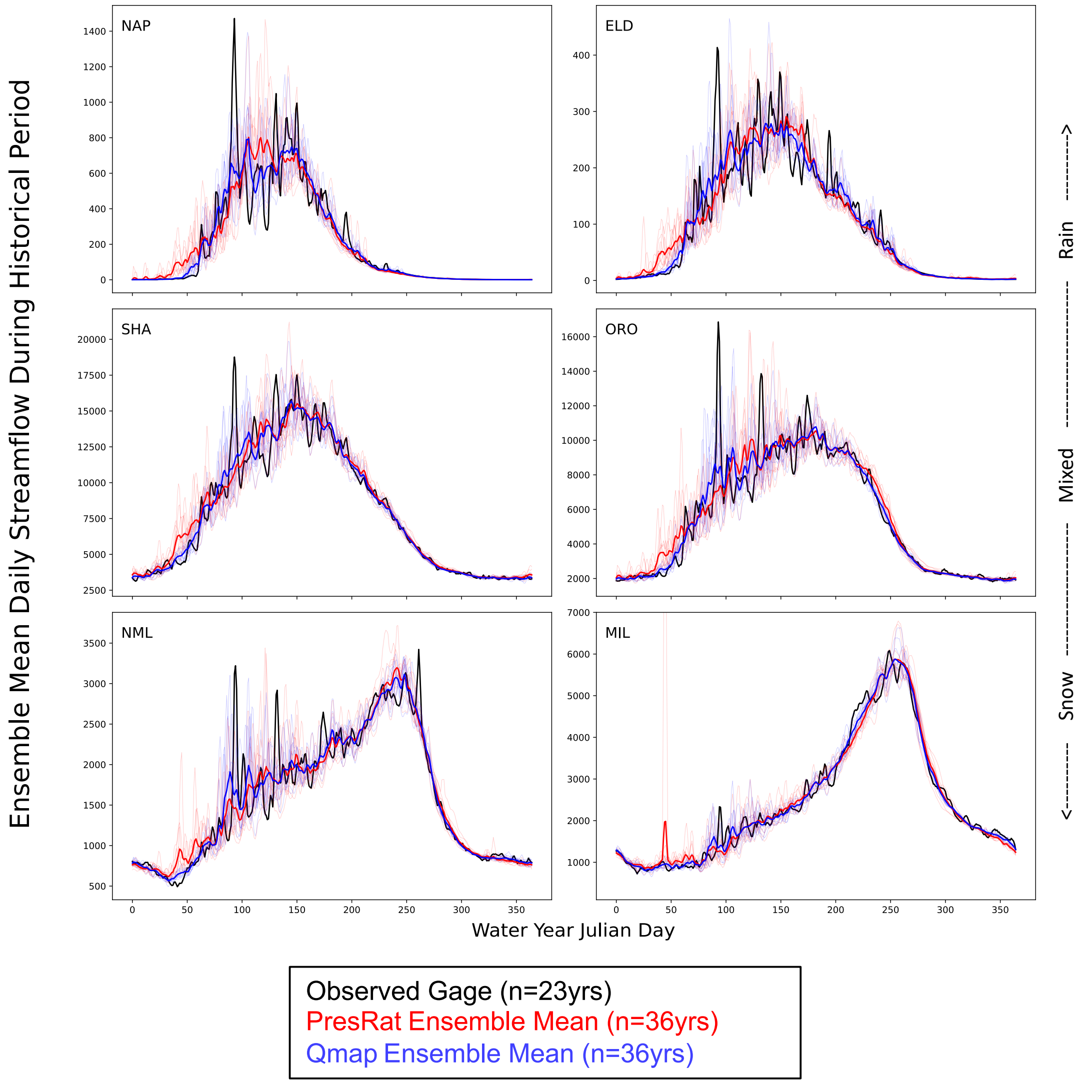



Fig. 4. Daily mean streamflow for the observed (black), and bias corrected data over the historical period for PresRat with seasonally anchored windowing (red) and quantile mapping with day-of-year windowing (blue) for each of the 6 rivers. For the GCM data, the ensemble mean across the 10-GCMs is shown in a bolded line with individual members depicted by

thin lines. Subplots are arranged such that from top-to-bottom, streams transition from rain- to snow dominated watersheds.

*b. Water Year Mean Streamflow*

Before evaluating the ability of each bias correction method to preserve the projected future change in mean water year streamflow, we first examine the original model signal (Figure 5). Across the simulations of 6 streams that were driven by projections from 10 GCMs, the changes in water year mean streamflow differ considerably. Evaluating change in units of percent of historical mean, which removes the influence of differing streamflow magnitudes, we see that CESM1-BGC, CNRM-CM5, and CanESM2 have the largest projected increases in water year mean streamflow with values upwards of 40% for some streams. Although the largest percent increases occur for the streams with the lowest flows, these models suggest higher-flow streams will still see increases beyond 30%. Among the projections based on the 10 GCMs, there is little agreement on the sign or magnitude of the projected change and the projections with very large increases heavily influence the multi-model ensemble mean change.

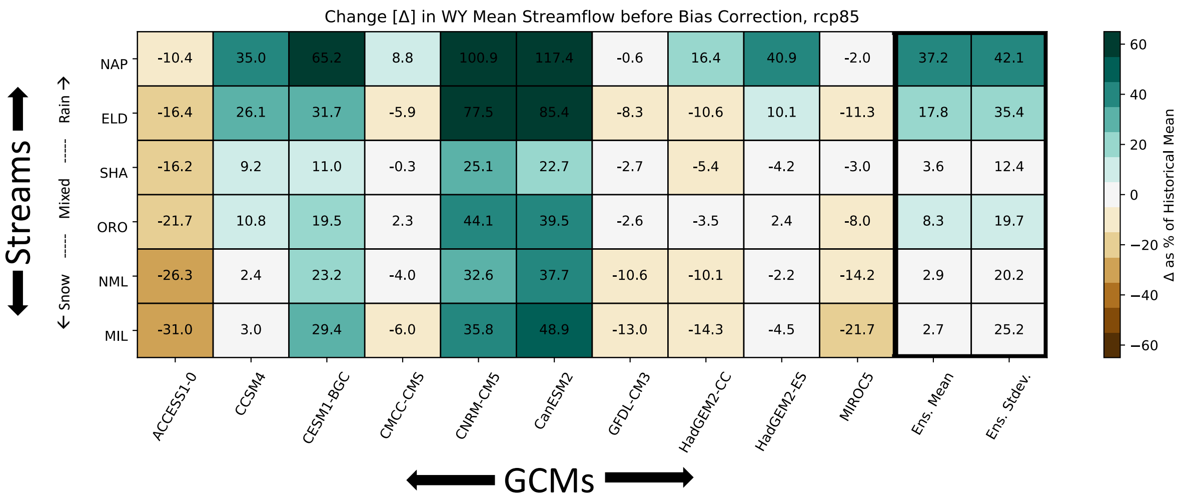



Fig. 5. Original model projected change in water year mean streamflow across the 6 streams (rows) and 10 GCMs (columns) with the 10-member ensemble mean and standard deviation (farthest right columns). Using units of ‘percent change from the historical mean’, increasing streamflow is indicated by green color-shading, and decreasing streamflow by brown. Values under 5% change appear as white.

We use the term 'error' here to refer to the difference between original model projected change and post-bias correction projected change. This quantity is an error in the sense that we expect our bias correction approach to preserve certain key aspects of the model-predicted climate change signal (as listed explicitly above), and is defined in Equation 2 below,

$$Error = \Delta_{Bias\ Corrected} - \Delta_{Raw\ GCM} \qquad (2)$$

wherein Δ is defined by Equation 1. Using the above definition, we now compare the ability of the 4 bias correction methods and 2 windowing techniques in preserving the signal of water year mean change averaged across the 10 GCMs and 6 streams.

The CDFt, EDCDFm, and Qmap bias correction techniques do not intrinsically preserve the water year mean flow change signal from the un-bias corrected projections (Table 3). The PresRat method is the one method that preserves the water year mean flow change, for both windowing methods, with mean and root-mean-square (RMS) errors <1%, and does so by design (see Pierce et al., 2015). The other three methods alter the original model signal to varying degrees. In all cases, applying a seasonally anchored window improves the preservation of the water year mean flow change signal by nearly a factor of 2 compared to the standard day-of-year windowing technique. In addition to improving the mean flow change, the new windowing technique also results in a narrower spread of errors among the models in the PresRat, CDFt, and Qmap methods.

| BC Method | Ens. Mean Error [%] | RMS Error [%] |
|---|---|---|
| PresRat | 0.20 (0.37) | 0.86 (0.91) |
| CDFt | 1.72 (6.18) | 5.57 (8.64) |
| EDCDFm | 4.49 (7.69) | 13.04 (12.96) |
| Qmap | 3.33 (6.14) | 7.64 (10.59) |

Table 3. Summary statistics for each bias correction method (rows) and windowing technique (Standard day-of-year show in parentheses, seasonally anchored shown without) on their

success in preserving the projected mean change in water year mean streamflow across the 6 streams and 10 GCMs.

*c. Wet Season Streamflow by Decile*

The hydroclimate of the western US is dominated by the occurrence (and absence of) extreme precipitation and streamflow events. Therefore, statistical bias correction techniques need to preserve the original-model projected changes at high-flow quantiles, rather than altering the projected change for no physical reason. When viewed at the granularity of a single percentile of flow, the signal of projected changes can be noisy. Therefore, we evaluate here changes at decile levels over the wet season. Again, before evaluating the ability of each bias correction method to preserve this quantity, we first examine the un-bias corrected model signals. Change is calculated with Equation 1 using the mean of all streamflow values within a given decile range for each of the historical and future climate periods. Figure 6 gives a visual depiction of the original model climate change signal at the Shasta gage and the remaining 5 streams can be seen in Appendix B. Although some minor differences exist among the 6 streams and 10 GCM-based histories, there is a near unanimous agreement that the top 10-20% of streamflow values will increase while the middle ~30% of the distribution will decrease. This follows the projected climate change signal in precipitation wherein high-tail events occur with greater frequency (Gershunov et al., 2019). Notably, in the historically most snow dominated watershed, Millerton, larger differences exist between the 10-member ensemble, with 4 models projecting large increases at the high end and 2 projecting decreases. Again, this work does not focus on the impacts or certainty of projected streamflow changes, but rather on the extent to which they are altered by bias correction techniques.

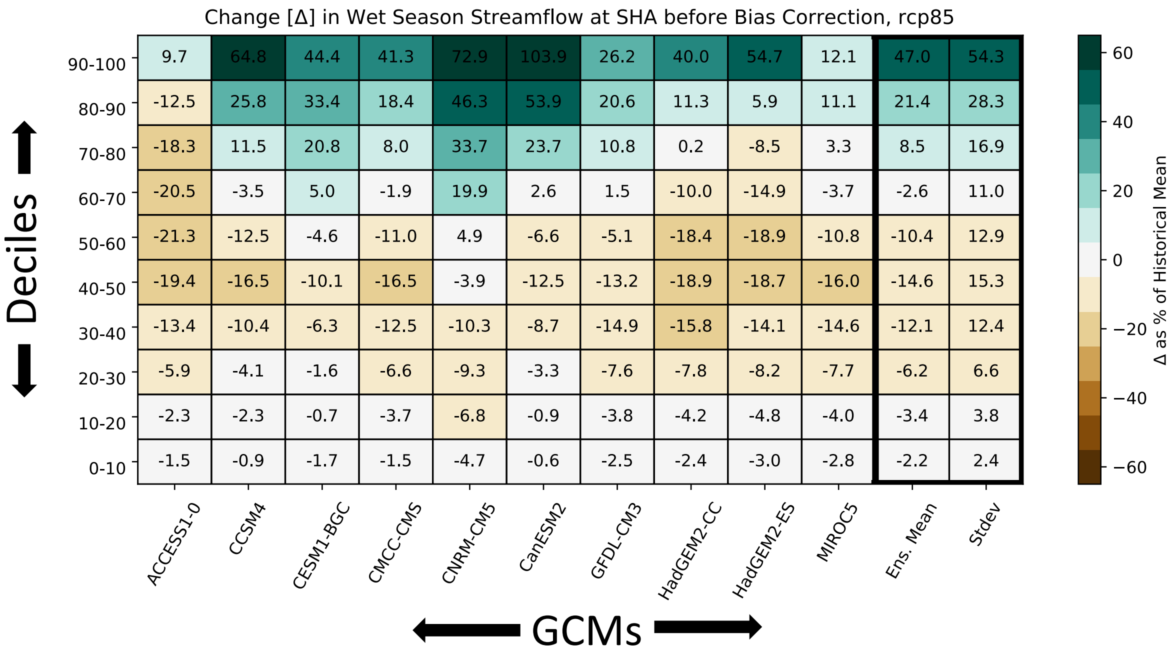



Fig 6. Original model projected change in Shasta streamflow by decile (rows) and GCM (columns) with the 10-member ensemble mean and standard deviation (farthest right columns). Using units of 'percent change from the historical mean', increasing streamflow is indicated by green color-shading, and decreasing streamflow by brown. Values under 5% change appear as white.

Using Equation 2, we evaluate how well each correction method and windowing technique preserves the original projected change at each decile over the wet season. Figure 7 provides a visual representation of these errors for the PresRat method using seasonally anchored windowing. Equivalent plots for the remaining combination of bias correction and windowing methods can be found in Appendix C. Each subplot contains a box-and-whisker plot for a given stream wherein errors for individual GCM projections are shown by grey circles and the mean error across the 10-member ensemble is depicted by an orange line. In Figure 7, with few exceptions, errors in the depiction of future changes in wet season streamflow, by quantile, falls within ± 10% of the original model signal for all the combinations of GCMs and streams. As mentioned earlier, a large fraction of the total streamflow is contained in the top 10-20% of the distribution. To highlight where, in terms of decile, large errors in the bias correction method begets large errors in streamflow, the fraction of total wet season streamflow represented by each decile is plotted on the right y-axis. Though not explicitly shown here, the PresRat method with seasonally anchored windowing has notably smaller biases than other techniques at the top decile of flows.

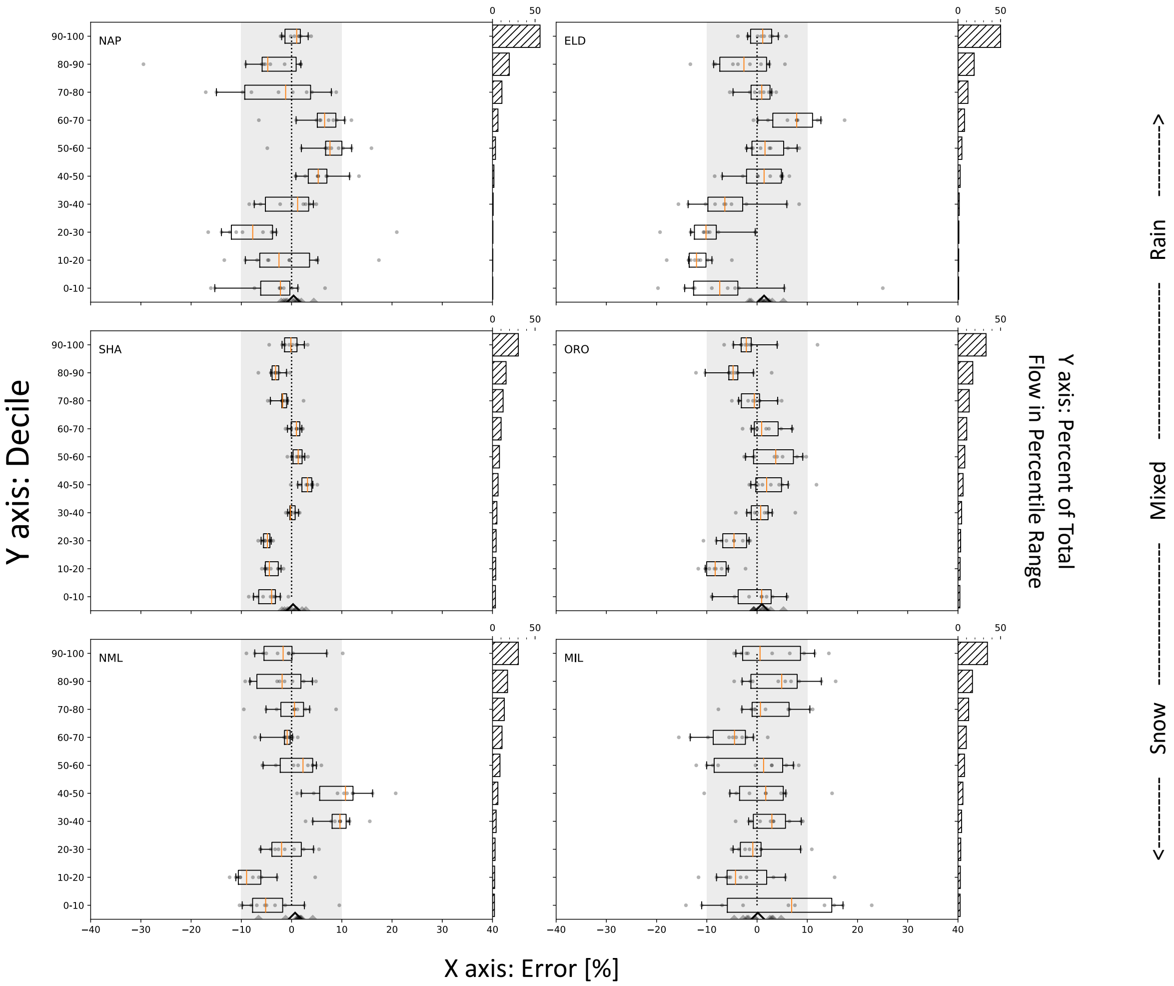



Fig. 7. Error by decile for change in wet season streamflow for PresRat with seasonally anchored windowing. Errors (x-axis) for individual GCMs are depicted by grey circles, the mean error across the 10-member ensemble is depicted by an orange line. The box edges and whiskers represent the middle 5 and 8 GCMs respectively. The error for a single GCM averaged over all deciles is depicted as a small grey triangle on the x-axis and the value for the 10-member ensemble mean is denoted by a large triangle. For reference, the grey shading and dashed black line correspond to ± 10% error and 0% error respectively. For each stream, the right y-axis depicts the historical percentage of total wet season streamflow contained in each decile averaged across the 10-member ensemble. Subplots are organized so that as you move down the rows, streams transition from rain- to snow dominated over the historical period.

To quantitatively assess how well the original signal is preserved, we calculate the root-mean-square error (RMSE) at each decile for each of the 4 correction methods and 2 windowing techniques, using the following equation,

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\Delta_{Bias\ Corrected_i} - \Delta_{Raw\ GCM_i})^2}{N}} \quad (3)$$

where i represents a given driving-GCM. Then, motivated by the importance of large streamflow events to the hydroclimate of the western US, we weight the RMSE at each decile by the percentage of total historical wet season streamflow in the decile, thus emphasizing errors at higher deciles. We refer to this quantity as the flow-weighted RMSE. To equally weight the performance of the bias correction technique on each river, and thus equally sampling the range of snow vs. rain dominated regimes, the flow weighted RMSE at each decile is averaged across the 6 streams.

Figure 8 and Table 4 summarize the above process and depicts the flow weighted RMSE averaged across the 6 streams for all combinations of correction method and windowing technique. At lower deciles, the flow weighted RMSE values are similar across the 4 correction methods. The lines begin to diverge near the 50$^{th}$ percentile with the seasonally anchored PresRat and CDFt methods achieving the best performance between the 50$^{th}$-90$^{th}$ percentiles. For the top 10% of streamflow, where the most impactful of streamflow events exist, PresRat with seasonally anchored windowing best preserves the original model signal of change. If we take the average across all deciles for each line individually, we see that the PresRat method using seasonally anchored windowing not only yields the lowest flow weighted RMSE, but that for each correction method, the seasonally anchored windowing method outperforms the standard day-of-year method (triangle markers).

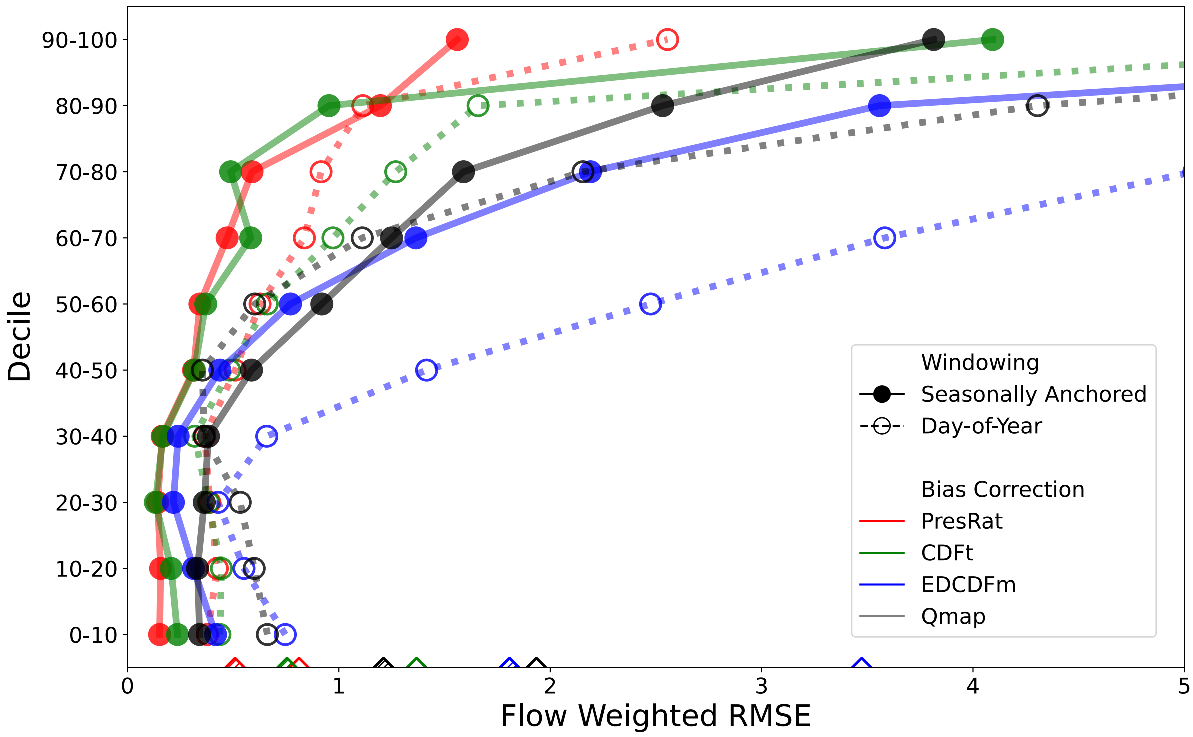



Fig. 8. Flow weighted root-mean-square error (RMSE) in the representation of model-predicted future change in mean flow at each decile averaged across the 6 streams. Solid (dotted) lines represent correction methods using seasonally anchored (standard day-of-year) windowing techniques. For solid (dotted) lines, the value of the flow weighted RMSE averaged across all deciles is indicated by a hatched (unfilled) triangle on the lower x-axis.

| BC Method | Ensemble Mean RMSE Across Deciles [Flow Weighted %] |
|---|---|
| PresRat | 0.51 (0.81) |
| CDFt | 0.75 (1.37) |
| EDCDFm | 1.81 (3.47) |
| Qmap | 1.21 (1.93) |

Table 4. Flow weighted root-mean-square error (RMSE) averaged over all deciles and across all streams. Methods using standard day-of-year windowing are shown in parentheses, and those using seasonally anchored are shown without parentheses.

*d. Temporal Shift in Peak Streamflow*

The previous sections focus on the ability of bias correction techniques to preserve original model projected changes in the magnitude of annual and wet season flow. Because the fingerprint of climate change for mountain rivers is characterized by changes in both magnitude and timing of peak streamflow, we will now evaluate the ability of each correction method and windowing technique to preserve projected shifts in seasonality. Using the climatological milestone associated with the peak streamflow (defined in Section 4), we compare the original model change (measured in days) in peak streamflow timing with the bias corrected changes.

Evaluating the difference in Julian day of peak streamflow between the end-of-century and historical periods, Figure 9 shows the original model change in days on the x-axis and the change from the PresRat with seasonally anchored windowing method on the y-axis for each of the 10 GCM-projected climates and 6 rivers. If the bias correction methods preserve the original model signal exactly, all markers would fall on the dotted black 1-to-1 line. For the overwhelming majority of stream and GCM combinations, we see that the temporal shift is well-preserved. Though the amount of change varies with GCM, it is most strongly related to the how snow-dominated a watershed is over the historical period. The largest signal of change is seen at New Melones (red markers), which loses its historical snowmelt peak entirely. The snowiest basin, Millerton (green markers), doesn't exhibit as large a signal because unlike New Melones, it retains a snowmelt peak in some projections.

Table 5 summarizes the preservation of the projected temporal shifts for the various bias correction methods and windowing techniques. For the PresRat and CDFt methods, which best preserve changes in streamflow magnitude, the seasonally anchored windowing method yields similar error metrics as the standard day-of-year technique. By virtue of locking windows to Julian days, the day-of-year method does a good job at preserving the raw model signal of temporal change. Notably, the seasonally anchored method achieves the same efficacy (for PresRat and CDFt) as the day-of-year method without requiring identical windows for the historical and future datasets.

| BC Method | R2 | Ens. Mean Error [days] | RMSE [days] |
|---|---|---|---|
| PresRat | 0.9 (0.9) | 0.4 (0.1) | 6.8 (6.9) |
| CDFt | 0.9 (0.9) | 0.1 (0.1) | 7.1 (7.0) |
| EDCDFm | 0.6 (0.9) | 3.1 (-0.5) | 15.3 (6.7) |
| Qmap | 0.8 (0.9) | 0.1 (0.4) | 11.1 (7.6) |

Table 5. Summary of the ability of each bias correction method and windowing technique to preserve the original model signal in the temporal shift in climatological peak streamflow. Methods using standard day-of-year windowing are shown in parentheses, and those using seasonally anchored are shown without parentheses.

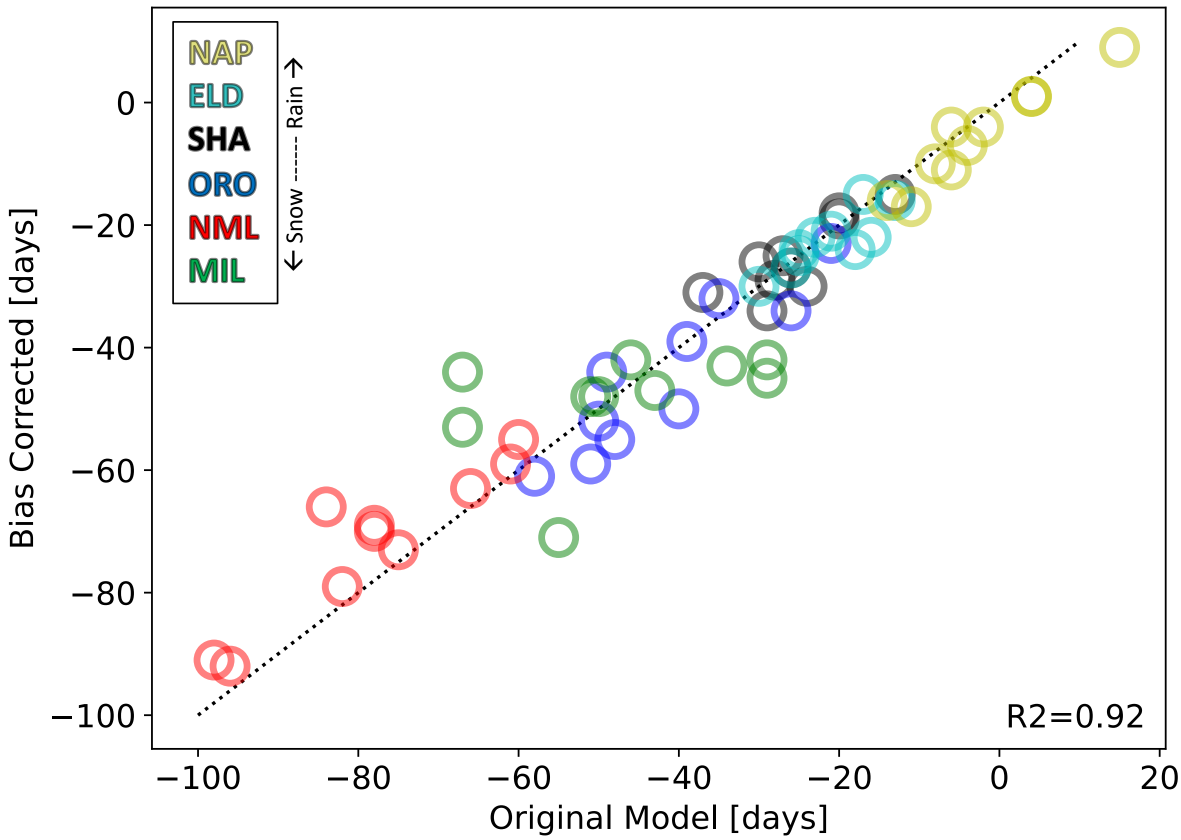



Fig. 9. Change in timing of peak climatological streamflow for the original model change (x-axis) and PresRat with seasonally anchored windowing (y-axis) between the end-of-century

and historical periods. Markers represent the change for each GCM and are color-coded by stream. A dotted black 1-to-1 line is shown for reference.

## 6. Summary and Conclusion

Robust and reliable projections of changes in future streamflow are essential if we are to improve (or maintain) resilient water resources and mitigate damage to riparian ecosystems in the face of climate change. But raw simulations of streamflow are used in applications or impact models only at one's peril, unless first bias corrected. Traditional methods for bias correction operate by comparing future and historical model data from shared ranges of Julian days. However, the physical and environmental process that govern streamflow (or any hydrometeorological variable) are not necessarily fixed to calendar periods, especially in the context of climate change, which will advance melting season earlier in the year and can alter the seasonality of precipitation. In order to better connect the statistical process of bias correction to the underlying processes in hydrologic models, we introduce a novel windowing technique for bias correction of projected streamflows. Data are windowed based on hydrograph-relative time, not calendar day. By locating the temporal position of a given data point undergoing correction in relation to characteristic features of its average hydrograph (e.g., start of rising limb, peak flow, end of falling limb, minimum flow) we window data based on hydrographically-equivalent days across the observed, simulated-historical, and simulated-future periods.

We evaluate the efficacy of several bias correction methods, using both the standard day-of-year and our new seasonally anchored windowing technique, applied to streamflow projections for six California streams that range from rain- to snow dominated watersheds and that are responses to climate projections from a suite of 10 CMIP5 global climate models (GCMs) selected by the California Department of Water Resources as having good representation of the historical California hydroclimate. Based on the importance of individual high-magnitude streamflow events, total water year streamflow, and timing of peak flow on the natural and built environment, we argue that successful bias correction should accomplish three tasks: 1) preserve the water year mean climate change signal of the un-bias corrected flow projections, 2) preserve un-bias corrected changes at all quantiles, and 3) preserve any temporal signal of shifting seasonality in the un-bias corrected flow projections, all while correcting the simulated historical statistics to that of the observed dataset. Evaluated as the percent difference relative to the historical period, we investigate the degree to which the four bias methods and two windowing techniques preserve the un-bias corrected signal of climate change across the study domain.

PresRat is the only bias correction method examined in this work that preserves the original model, shorthand for hydrologic model output driven by downscaled and bias corrected GCM data, signal of water year mean change and does so for both seasonally anchored and standard day-of-year windowing techniques. Although the other 3 methods, CDF transform (CDFt; Michelangeli et al., 2009), Equidistant CDF matching (EDCDFm; Li et al., 2010), and Quantile Mapping (Qmap; Panofsky & Brier, 1968; Wood et al., 2002) do not preserve the original model water year mean changes, even for these methods, seasonally anchored windowing reduces the ensemble mean error by roughly a factor of two while reducing the spread when compared to standard day-of-year windowing.

For an extreme and highly variable hydroclimate, like California, where a large fraction of total water (both streamflow and precipitation) is contained in the top decile of the distribution, it is vital that bias correction does not skew the original model projection signals of change at the highest quantiles. Using the root-mean-square error evaluated at each decile of wet season streamflow to gauge success, we find that 1) PresRat with seasonally anchored windowing best preserves the raw signal at the top decile of flows, 2) PresRat with seasonally anchored windowing best preserves the raw signal averaged over all deciles, and 3) using seasonally anchored windowing improved the performance of each of the four bias correction methods. These findings are true not only for the 10-member ensemble averaged over the six streams, but true for all streams individually. With respect to streamflow magnitude, regardless of their hydrological characteristics (e.g., rain- vs. snow dominated), the seasonally anchored windowing technique was more effective in preserving the original model signals of climate change.

Finally, because any shift in seasonality of snow-fed rivers will have strong impacts on both the natural and built environments, we examine the extent to which bias correction methods alter the original model signals of shifting seasons. We measure the change in seasonality by finding the difference between the Julian dates coinciding with peak streamflow in the simulated-future and simulated-historical periods. While the seasonally anchored windowing technique improved the preservation of original model signals in magnitude, we find that both windowing methods preserve the shift in seasonality equally well (for PresRat and CDFt) with mean bias values < 1 day and root-mean-square error of ~7 days.

In summary, seasonally anchored windowing, as opposed to the standard day-of-year technique, yields better bias corrected projections of future streamflow across a subset of six streams ranging from rain- to snow- dominated ecosystems in California. Without sacrificing any capacity to preserve projected changes in timing of peak streamflow, the seasonally anchored method improves the preservation of magnitude changes in the un-bias corrected flow projections. This is true not only for the water year mean signal, which is important as it relates to the total volume of water flowing through the river over the course of the year, but is also true for both low and high streamflow events which have an outsized imprint on California's hydroclimate, water resources, and ecosystems.

While this work focused largely on wet season streamflow, Figure 10 highlights an important vulnerability of the standard day-of-year windowing concerning late season flows. Since biases from the early receding limb of the historical period hydrograph are applied to near-baseflow streamflow during the end-of-century period (because both occur over the same calendar-based period), the resulting bias corrected projections can represent something very unphysical: streamflow decreases past baseflow during the receding limb before rebounding and then receding once more until it reaches baseflow (dashed lines). In contrast, the future period hydrographs for seasonally anchored methods (solid lines) do not exhibit this unphysical behavior because they apply biases from equivalent stretches of the climatologies. Because this seasonal shift is driven by warming temperatures resulting in diminished snowpack whose peak volume is pushed earlier into the water year, this non-physical feature is most prevalent for historically snow-dominated watersheds.

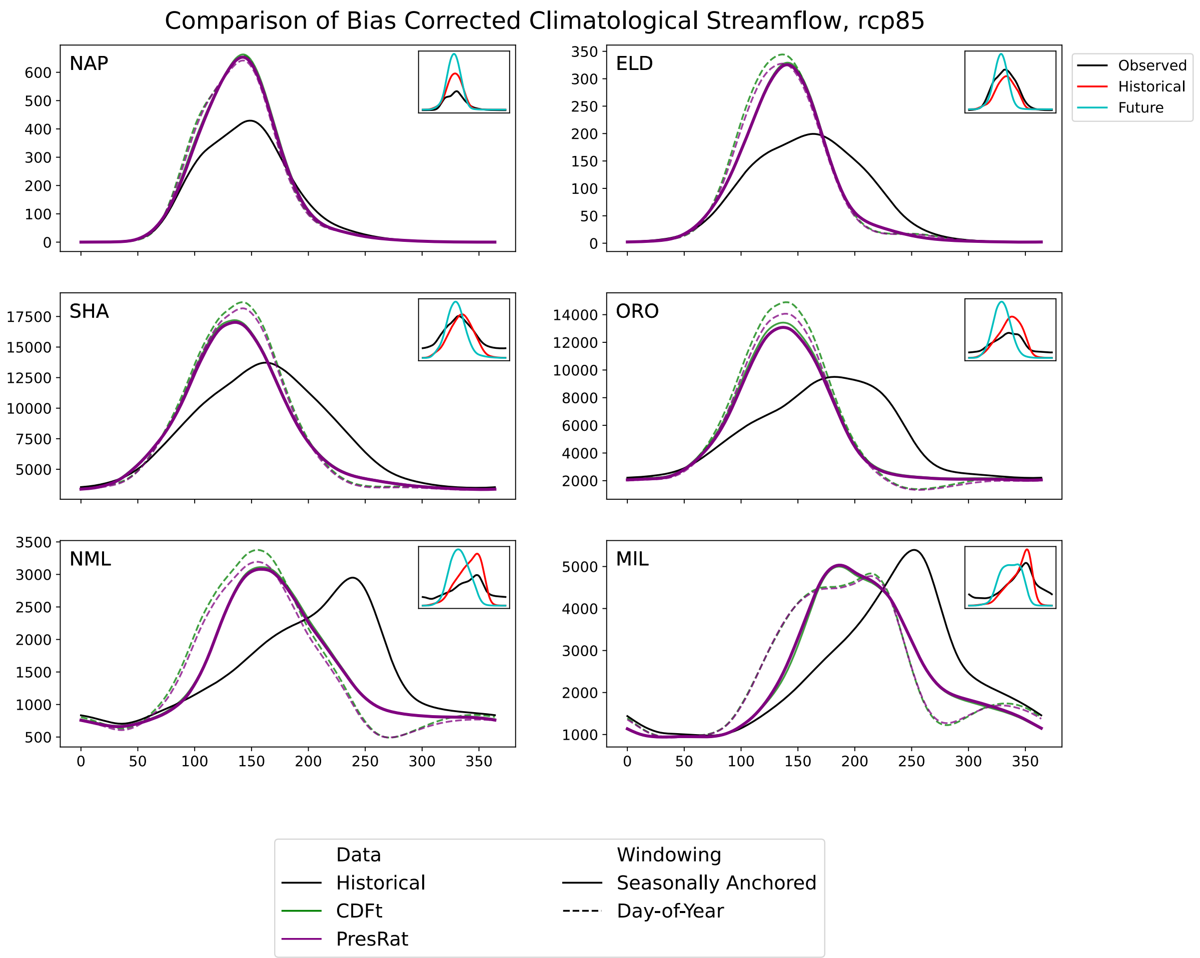



Fig. 10. Ensemble mean climatological hydrographs for raw and bias corrected data at each of the 6 streams. The upper inlay depicts observed (black) and original model hydrographs of the historical (red) and end-of-century (blue) periods for each river. The main subplots show the bias corrected historical hydrograph (black) alongside end-of-century bias corrected hydrographs for PresRat (purple) and CDFt (green) using seasonally anchored (solid) and day-of-year (dashed) windowing methods.

This work 1) demonstrates the inability of bias correction with day-of-year windowing to provide reliable projections of variables whose climate change signal is characterized by changes in both magnitude and seasonality, and 2) introduces a novel windowing method which moves towards 'process informed' bias correction wherein environmental and physical processes, rather than calendar dates, are shared by windowed data. Given the importance of streamflow projections in creating more resilient water resources, it may be pertinent to evaluate the difference in future streamflow projections across a wider range of California rivers using both seasonally anchored and day-of-year windowing methods. Although we conceptualized and applied the method for the purpose of bias correcting streamflow in California, the fundamental technique of windowing based on the position (in time) of a given data point in reference to some climatological milestones could be applied to variables other than streamflow (e.g., snow water equivalent, which is also expected to have a significant seasonal shift in the future). For variables that will have seasonal shifts in the future, it is important to move towards 'process informed' bias correction and away from calendar-based methods that are completely detached from the physical processes governing the systems. Although the method introduced here does not directly tether the statistical process of bias correction to the underlying physics of GCMs or land surface models, it is a step in the appropriate direction.

*Acknowledgments.*

All authors were supported by the CA Department of Water Resources grant number 4600013361. MDS received support from the US Army Corps of Engineers, grant number USACE W912HZ-19-SOI-002. DWP received support from the California Energy Commission grants EPC-16-063, PIR-19-007, and EPC-20-006 and NOAA's RISA California-Nevada Climate Applications Program Award NA17OAR4310284.

*Data Availability Statement.*

The data generated in this study will be available at [currently working on obtaining a DOI for the data].

## APPENDIX

### Appendix A: Identification of Peak Streamflow Milestone for Bimodal Climatological Hydrographs

Section 4.1.b describes the process of selecting peak streamflow milestones. While the vast majority of climatological hydrographs assessed in this study are not bimodal, for some future projections of historically snow-dominated rivers, the climatological hydrograph contains two local maxima (Figure A1). Recall that the purpose of the seasonally anchored windowing technique is to conduct bias correction across data with similar background physical and environmental processes. The peak milestone in the historical period corresponds to streamflow generated from snowmelt. The future period bimodal hydrograph is characterized by what is likely an earlier rain-dominated peak and a later-season snowmelt-dominated peak. Although the rain-dominated peak may be higher in magnitude, we select the later, snowmelt peak for the location of the milestone to better ensure that the data used in the bias correction shares similar physical and environmental processes.

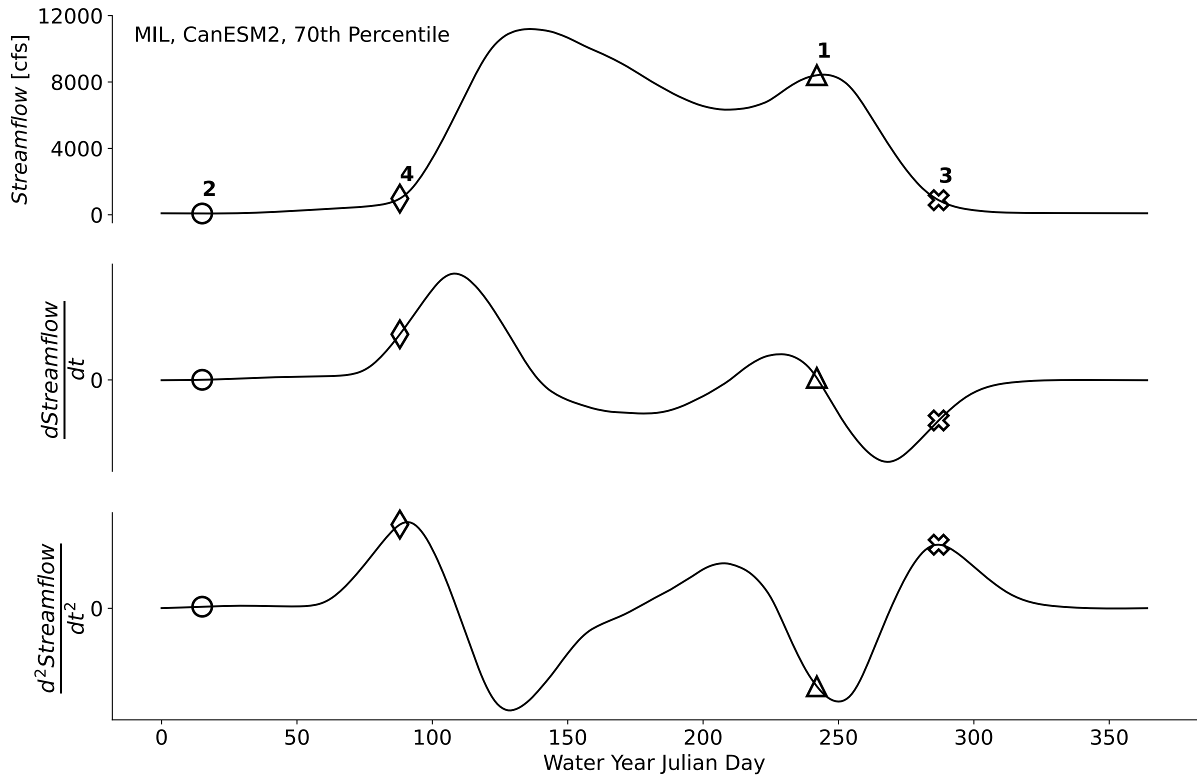



Fig. A1. Visual depiction of the algorithm used to identify climatological milestones for the seasonally anchored windowing method for the special case of bimodal hydrographs. End-of-century climatological data from the Millerton/Friant Dam stream and CanESM2 GCM is shown at the 70$^{th}$ percentile to illustrate the method. Daily mean climatological streamflow (top), and the first and second derivatives of streamflow with respect to time (middle and bottom respectively) are plotted on against water year Julian day. Numerical annotation is used to indicate the workflow by which the four seasonal milestones are assigned: (1) Peak streamflow (triangle), (2) Minimum streamflow (circle), (3) Start of the dry season/end of receding limb (x), and (4) Start of the wet season/beginning of the rising limb (diamond). Note that the peak milestone (triangle), is not associated with the true maximum value of streamflow, but rather with the local maximum during the snowmelt period.

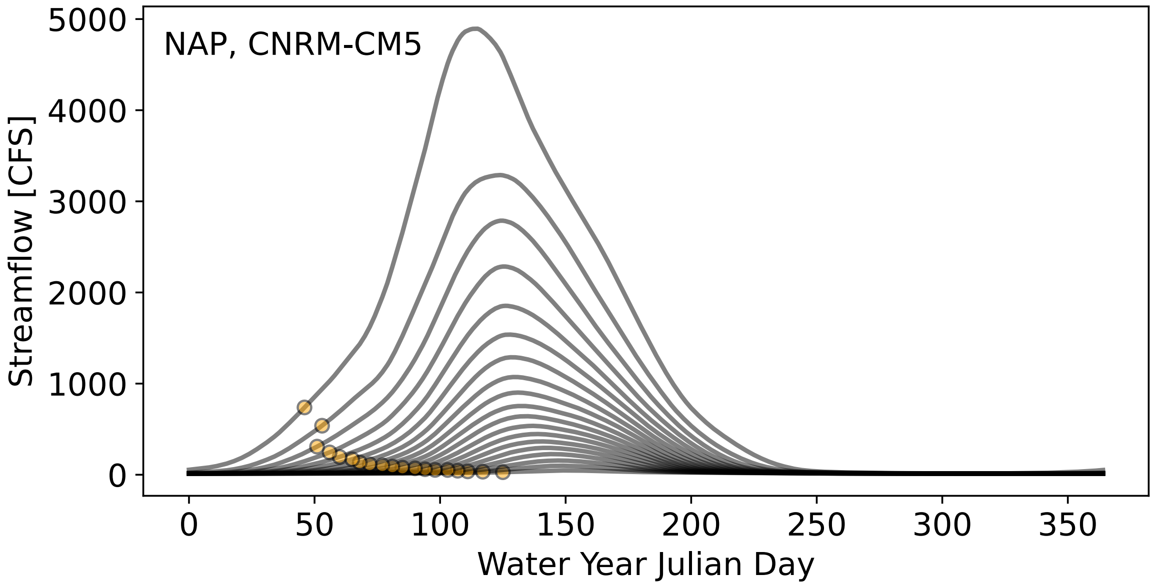



Fig. A2. Climatological hydrographs (lines) for flows ranging from the 5th to 95th percentile by intervals of 5 percentage points. Here, we see how the location of the 'start of wet season' milestone (circle) varies as a function of quantile.

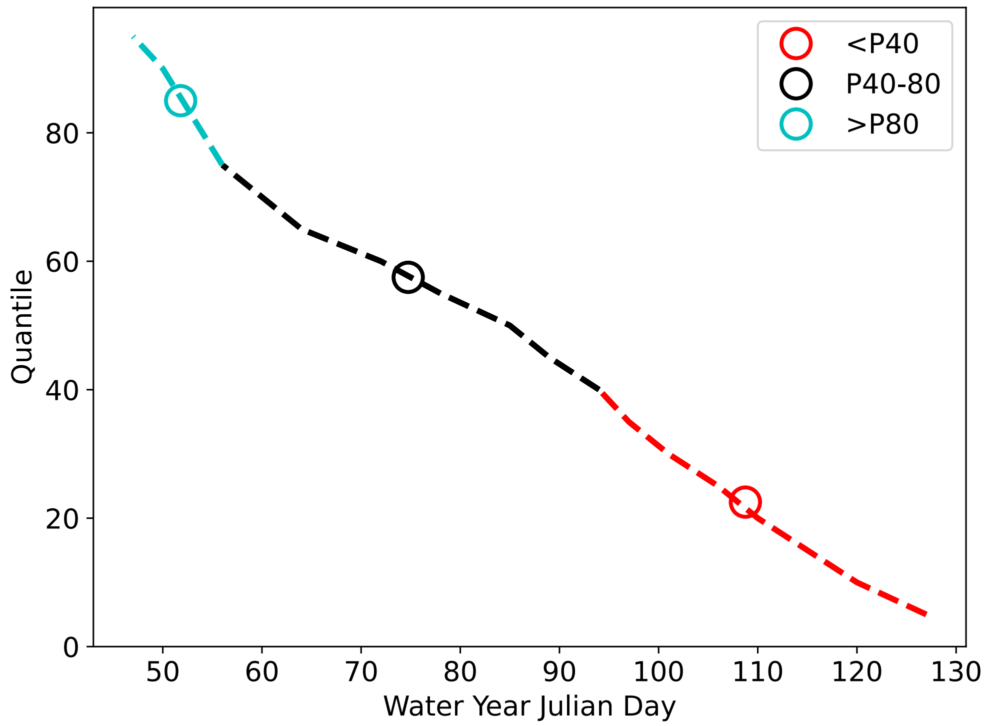


Fig. A3. Day of water year associated with the 'start of wet season' milestone (from Fig. A2) for flows below 40th percentile (red), 40th-80th percentile (black), and above 80th percentile (blue). The mean value over each respective range is indicated by a circle.

## Appendix B: Original model Change in Wet Season Streamflow by Decile

In accompaniment of Figure 6 in section 5.c, Appendix B provides the original model signal in wet season streamflow change by decile for the remaining 5 rivers.

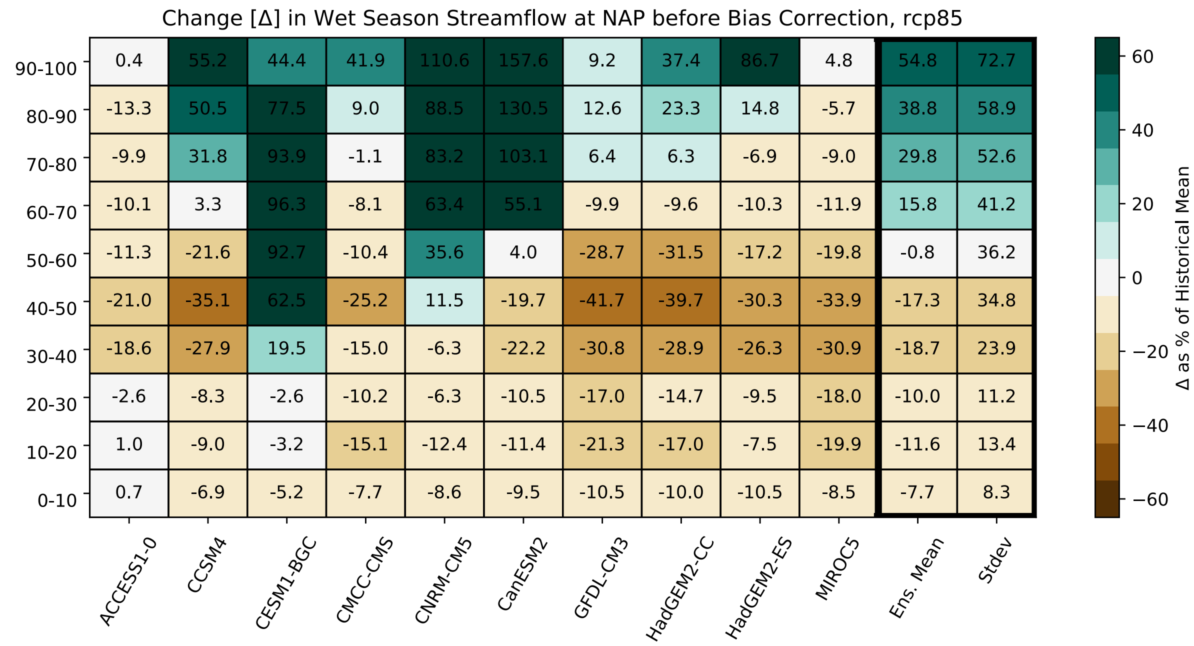



Fig. B1. Original model projected change in Napa River streamflow by decile (rows) and GCM (columns) with the 10-member ensemble mean and standard deviation (farthest right columns). Using units of 'percent change from the historical mean', increasing streamflow is indicated by green color-shading, and decreasing streamflow by brown. Values under 5% change appear as white.

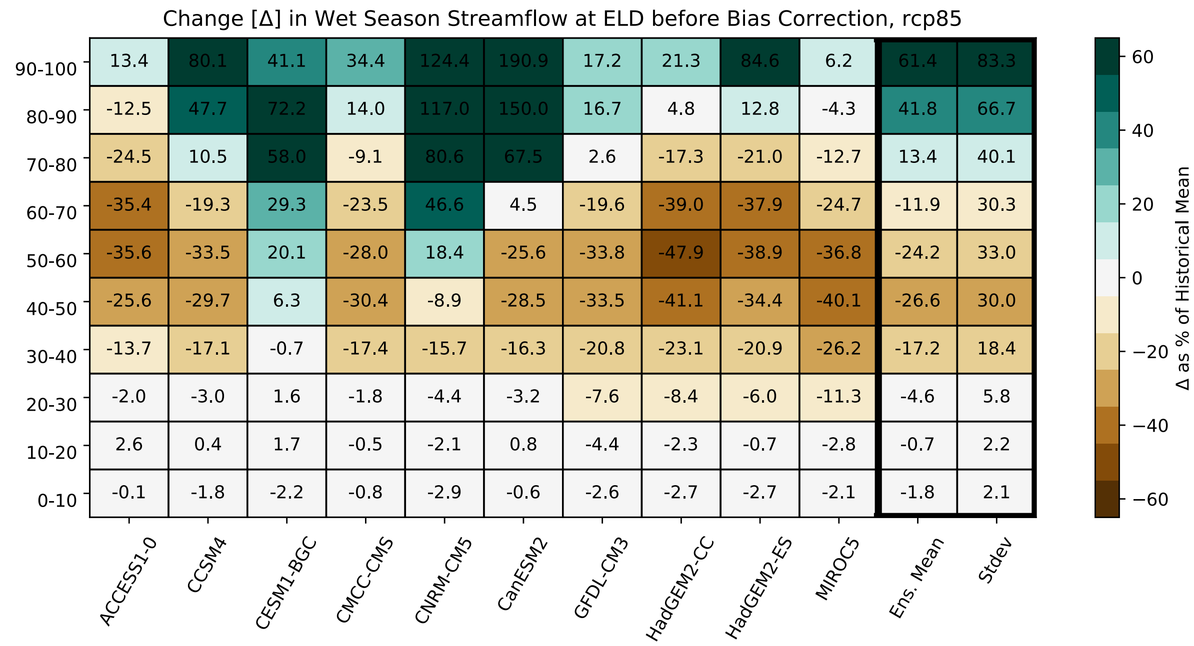



Fig. B2. Original model projected change in Elder Creek streamflow by decile (rows) and GCM (columns) with the 10-member ensemble mean and standard deviation (farthest right columns). Using units of ‘percent change from the historical mean’, increasing streamflow is indicated by green color-shading, and decreasing streamflow by brown. Values under 5% change appear as white.

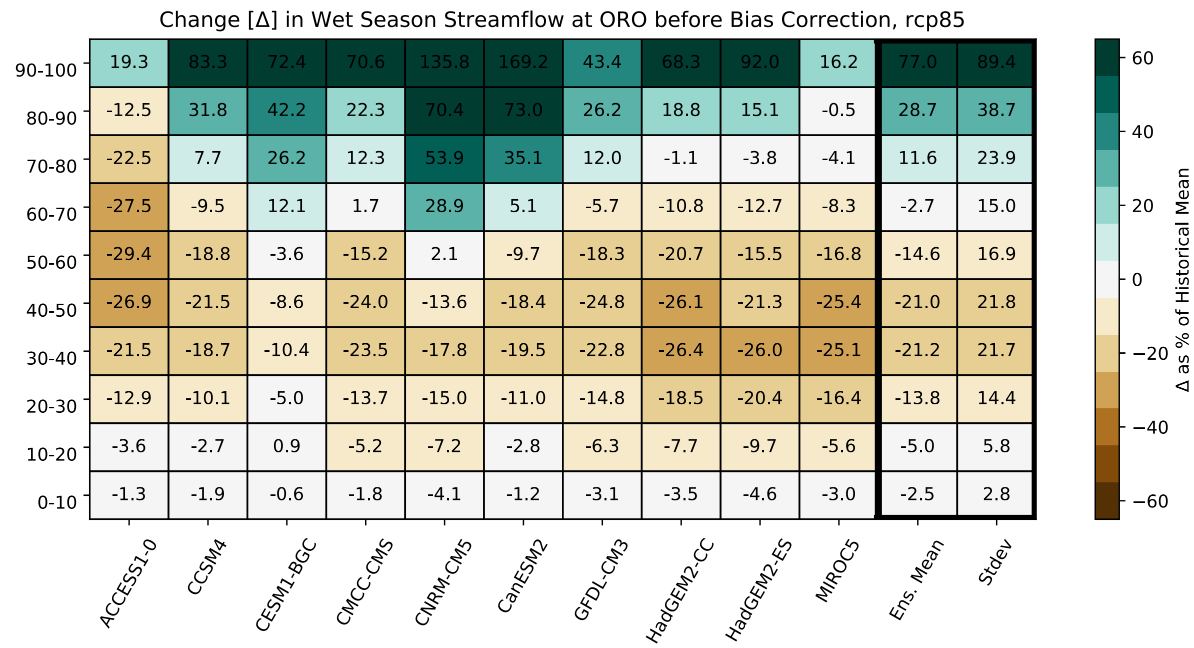


Fig. B3. Original model projected change in Oroville Dam streamflow by decile (rows) and GCM (columns) with the 10-member ensemble mean and standard deviation (farthest right columns). Using units of ‘percent change from the historical mean’, increasing streamflow is indicated by green color-shading, and decreasing streamflow by brown. Values under 5% change appear as white.

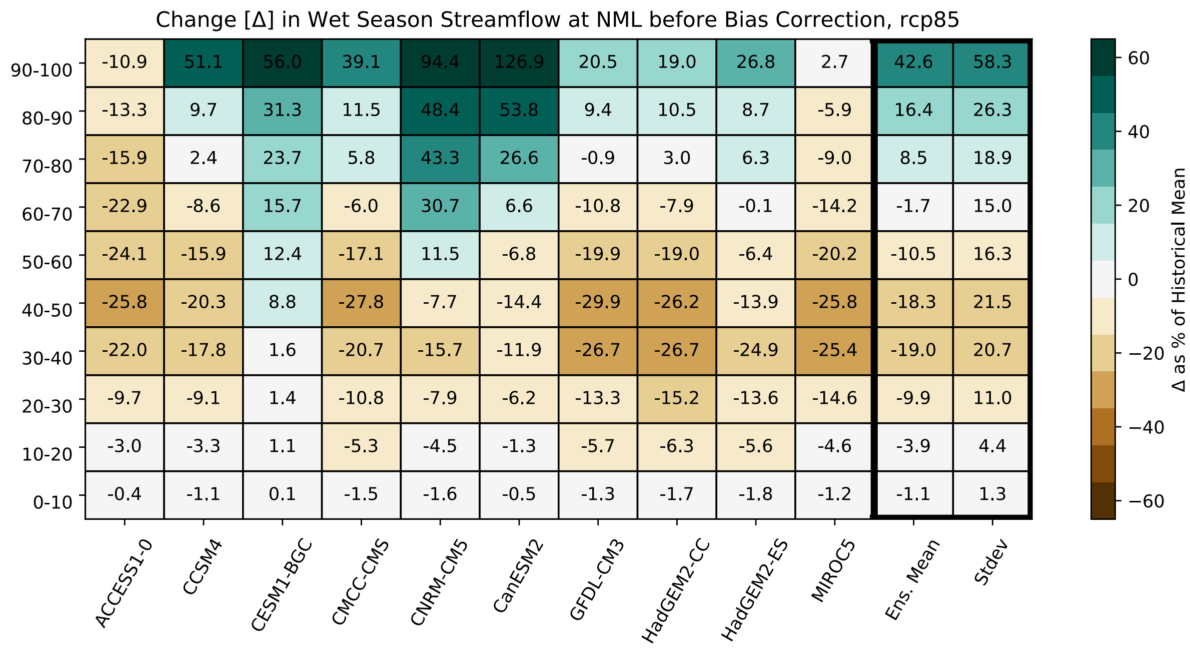



Fig. B4. Original model projected change in New Melones Reservoir streamflow by decile (rows) and GCM (columns) with the 10-member ensemble mean and standard deviation (farthest right columns). Using units of ‘percent change from the historical mean’, increasing streamflow is indicated by green color-shading, and decreasing streamflow by brown. Values under 5% change appear as white.

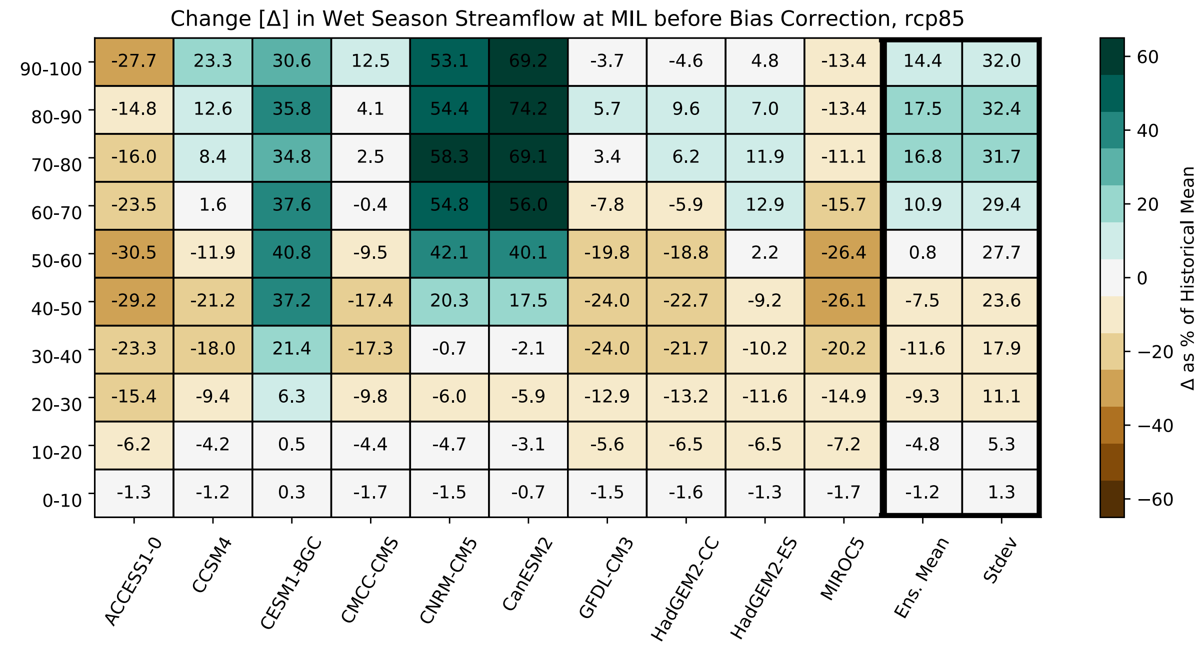



Fig. B5. Original model projected change in Millerton/Friant Dam streamflow by decile (rows) and GCM (columns) with the 10-member ensemble mean and standard deviation (farthest right columns). Using units of 'percent change from the historical mean', increasing streamflow is indicated by green color-shading, and decreasing streamflow by brown. Values under 5% change appear as white.

## Appendix C: Error in Wet Season Streamflow by Decile

In accompaniment of Figure 7 in section 5.c, Appendix C provides the error in wet season streamflow change by decile for the remaining combinations of bias correction and windowing techniques.

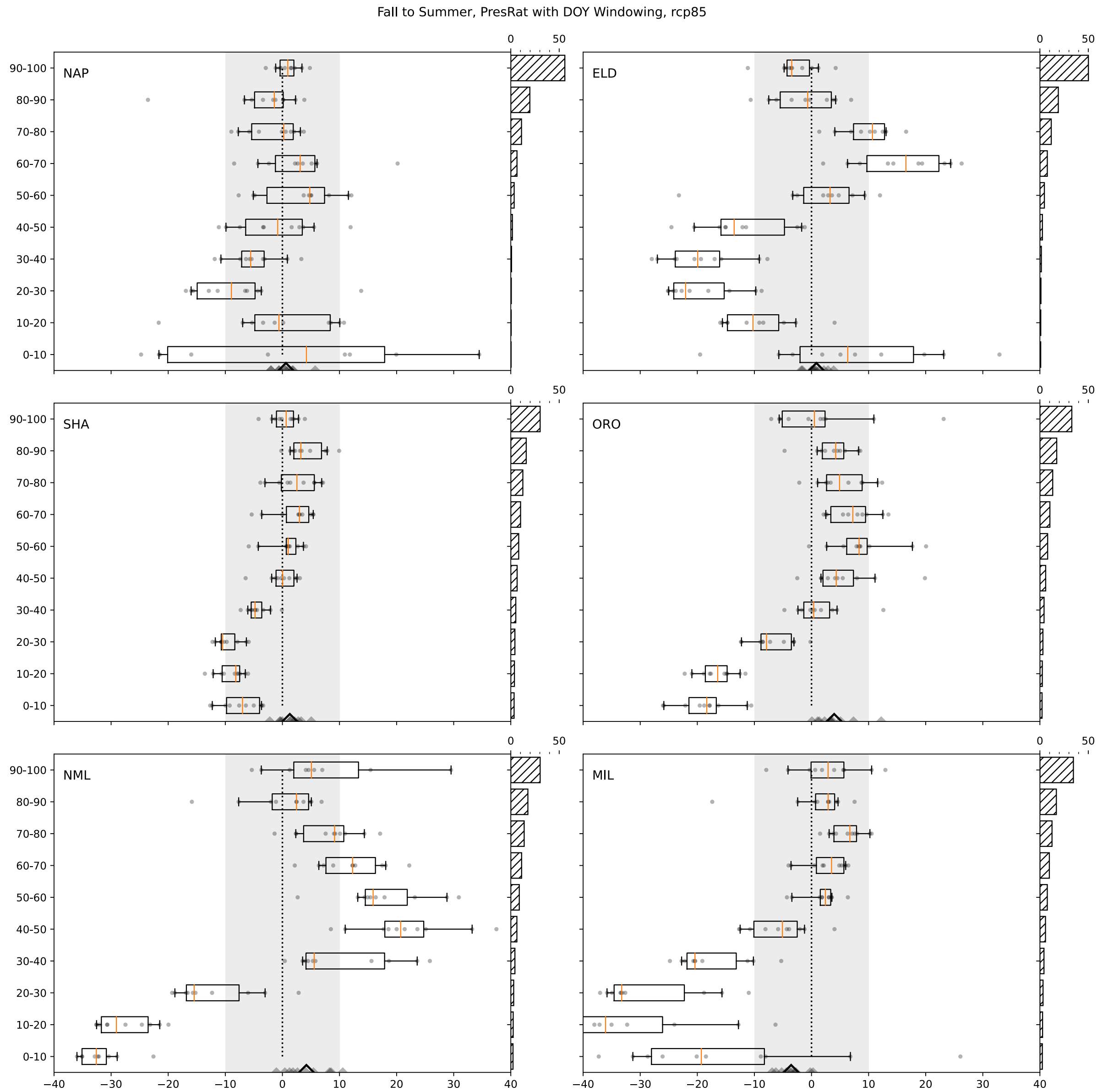



Fig. C1. Error by decile for change in wet season streamflow for PresRat with standard day-of-year windowing. Errors (x-axis) for individual GCMs are depicted by grey circles, the mean error across the 10-member ensemble is depicted by an orange line. The box edges and whiskers represent the middle 5 and 8 GCMs respectively. The error for a single GCM averaged over all deciles is depicted as a small grey triangle on the x-axis and the value for the 10-member ensemble mean is denoted by a large triangle. For reference, the grey shading and dashed black line correspond to ± 10% error and 0% error respectively. For each stream, the right y-axis depicts the historical percentage of total wet season streamflow contained in each decile averaged across the 10-member ensemble. Subplots are organized so that as you move down the rows, streams transition from rain- to snow dominated over the historical period.

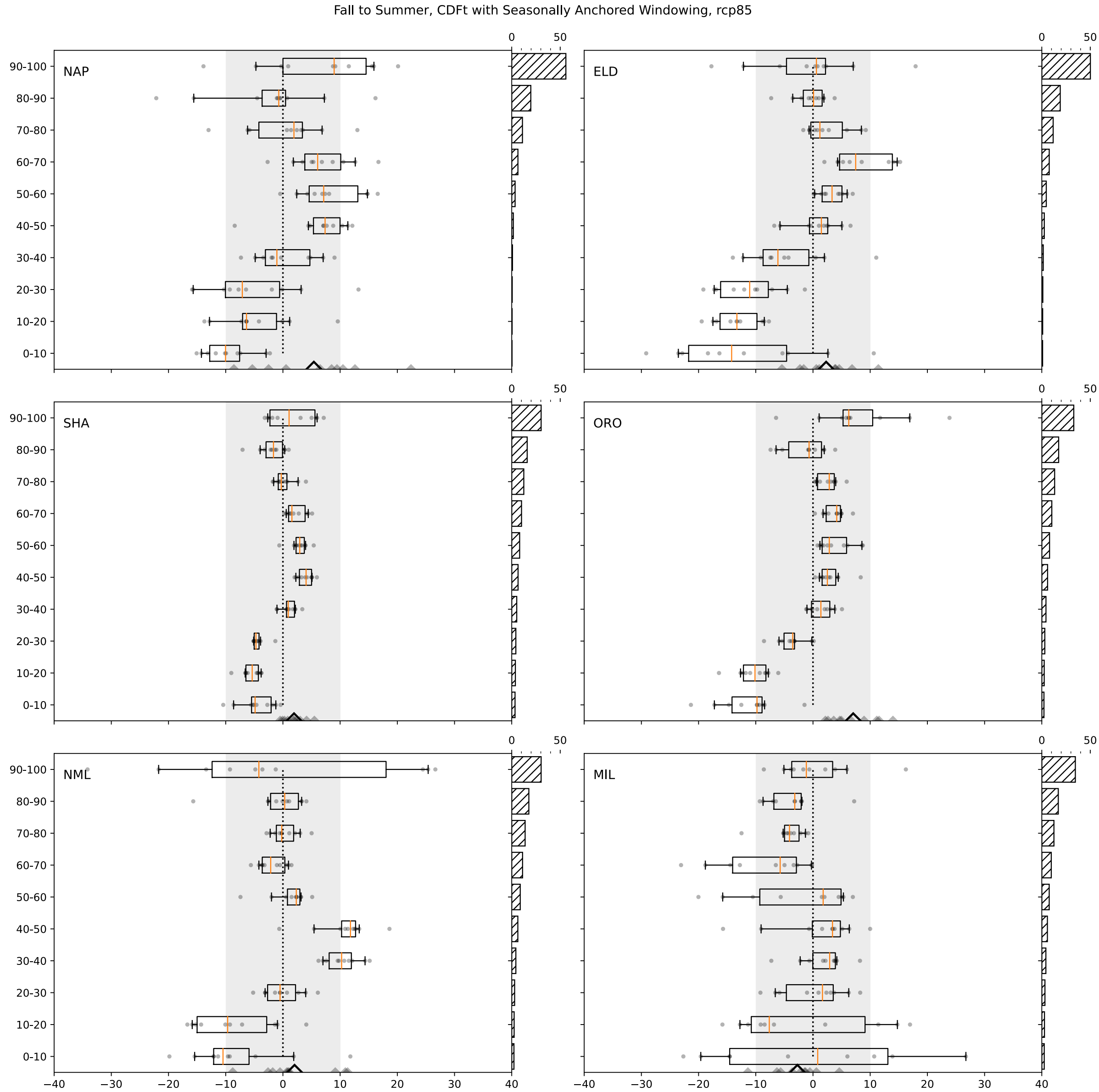



Fig. C2. Error by decile for change in wet season streamflow for CDFt with seasonally anchored windowing. Errors (x-axis) for individual GCMs are depicted by grey circles, the mean error across the 10-member ensemble is depicted by an orange line. The box edges and whiskers represent the middle 5 and 8 GCMs respectively. The error for a single GCM averaged over all deciles is depicted as a small grey triangle on the x-axis and the value for the 10-member ensemble mean is denoted by a large triangle. For reference, the grey shading and dashed black line correspond to ± 10% error and 0% error respectively. For each stream, the right y-axis depicts the historical percentage of total wet season streamflow contained in each decile averaged across the 10-member ensemble. Subplots are organized so that as you move down the rows, streams transition from rain- to snow dominated over the historical period.

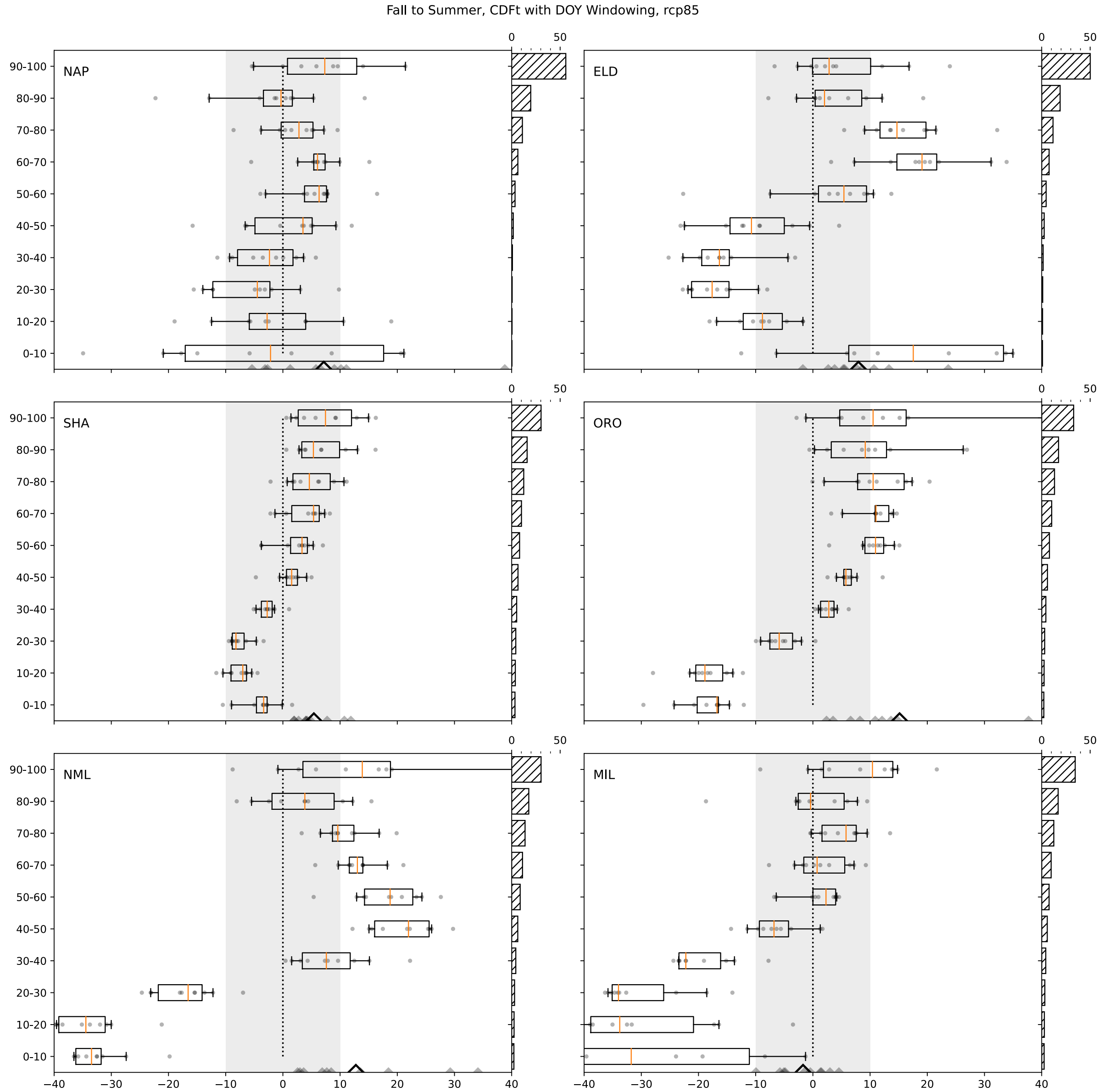
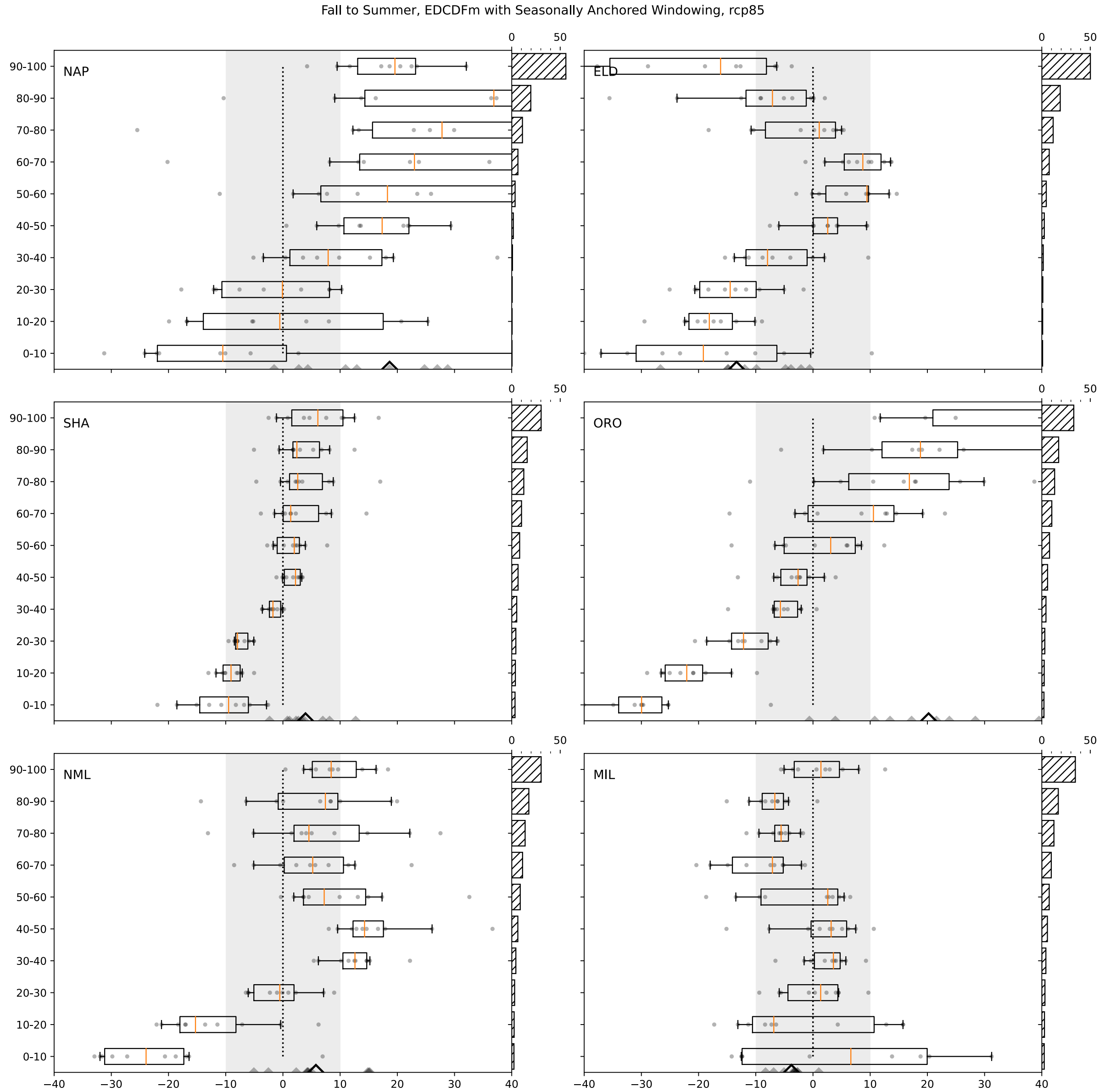



Fig. C3. Error by decile for change in wet season streamflow for CDFt with standard day-of-year windowing. Errors (x-axis) for individual GCMs are depicted by grey circles, the mean error across the 10-member ensemble is depicted by an orange line. The box edges and whiskers represent the middle 5 and 8 GCMs respectively. The error for a single GCM averaged over all deciles is depicted as a small grey triangle on the x-axis and the value for the 10-member ensemble mean is denoted by a large triangle. For reference, the grey shading and dashed black line correspond to ± 10% error and 0% error respectively. For each stream, the right y-axis depicts the historical percentage of total wet season streamflow contained in each decile averaged across the 10-member ensemble. Subplots are organized so that as you move down the rows, streams transition from rain- to snow dominated over the historical period.




846

847 Fig. C4. Error by decile for change in wet season streamflow for EDCDFm with seasonally

848 anchored windowing. Errors (x-axis) for individual GCMs are depicted by grey circles, the

849 mean error across the 10-member ensemble is depicted by an orange line. The box edges and

850 whiskers represent the middle 5 and 8 GCMs respectively. The error for a single GCM

851 averaged over all deciles is depicted as a small grey triangle on the x-axis and the value for

852 the 10-member ensemble mean is denoted by a large triangle. For reference, the grey shading

853 and dashed black line correspond to ± 10% error and 0% error respectively. For each stream,

854 the right y-axis depicts the historical percentage of total wet season streamflow contained in

855 each decile averaged across the 10-member ensemble. Subplots are organized so that as you

856 move down the rows, streams transition from rain- to snow dominated over the historical

857 period.

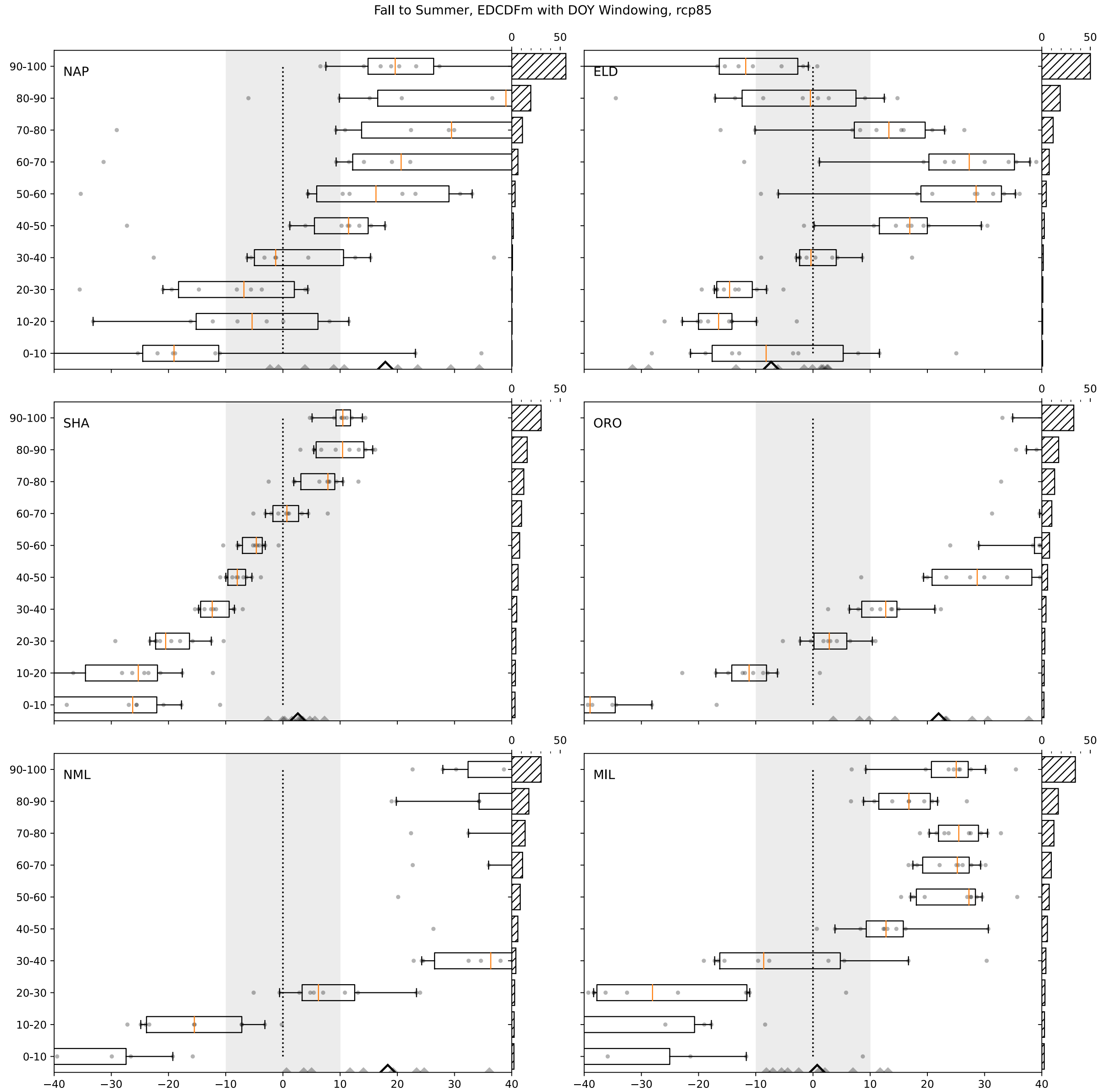



Fig. C5. Error by decile for change in wet season streamflow for EDCDFm with standard day-of-year windowing. Errors (x-axis) for individual GCMs are depicted by grey circles, the mean error across the 10-member ensemble is depicted by an orange line. The box edges and whiskers represent the middle 5 and 8 GCMs respectively. The error for a single GCM averaged over all deciles is depicted as a small grey triangle on the x-axis and the value for the 10-member ensemble mean is denoted by a large triangle. For reference, the grey shading and dashed black line correspond to ± 10% error and 0% error respectively. For each stream, the right y-axis depicts the historical percentage of total wet season streamflow contained in each decile averaged across the 10-member ensemble. Subplots are organized so that as you move down the rows, streams transition from rain- to snow dominated over the historical period.

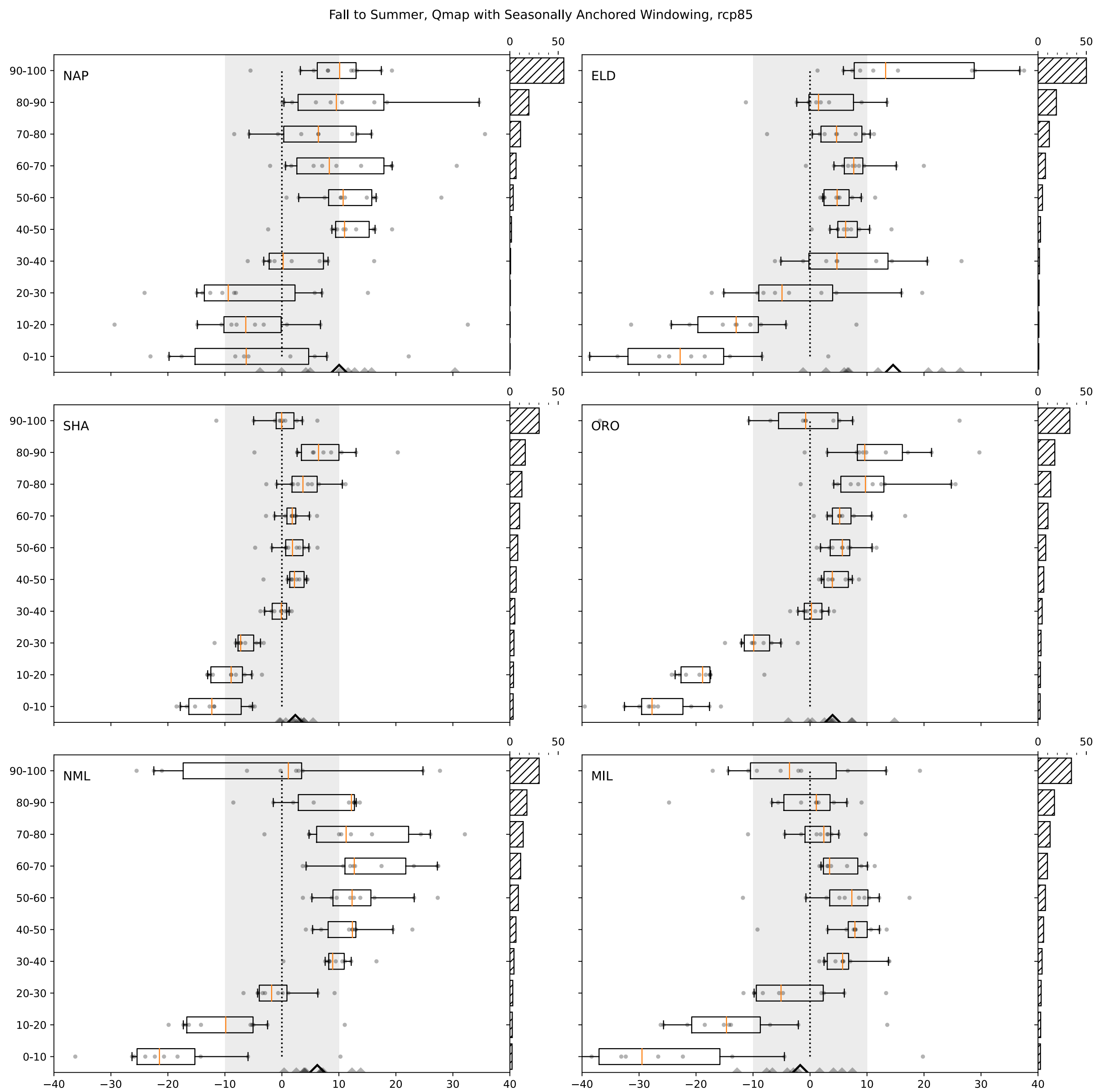



Fig. C6. Error by decile for change in wet season streamflow for Qmap with seasonally anchored windowing. Errors (x-axis) for individual GCMs are depicted by grey circles, the mean error across the 10-member ensemble is depicted by an orange line. The box edges and whiskers represent the middle 5 and 8 GCMs respectively. The error for a single GCM averaged over all deciles is depicted as a small grey triangle on the x-axis and the value for the 10-member ensemble mean is denoted by a large triangle. For reference, the grey shading and dashed black line correspond to ± 10% error and 0% error respectively. For each stream, the right y-axis depicts the historical percentage of total wet season streamflow contained in each decile averaged across the 10-member ensemble. Subplots are organized so that as you move down the rows, streams transition from rain- to snow dominated over the historical period.

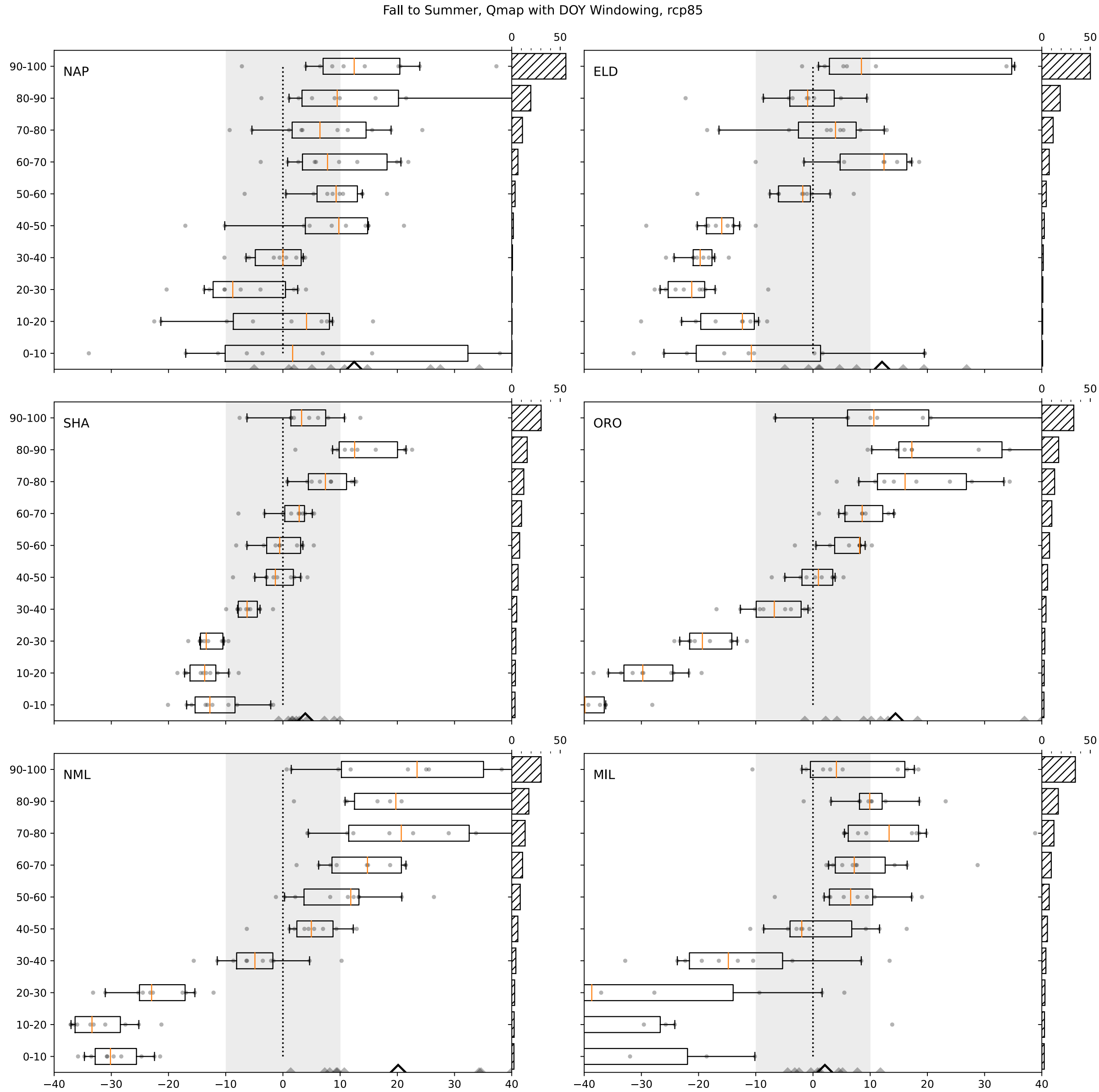



Fig. C7. Error by decile for change in wet season streamflow for Qmap with standard day-of-year windowing. Errors (x-axis) for individual GCMs are depicted by grey circles, the mean error across the 10-member ensemble is depicted by an orange line. The box edges and whiskers represent the middle 5 and 8 GCMs respectively. The error for a single GCM averaged over all deciles is depicted as a small grey triangle on the x-axis and the value for the 10-member ensemble mean is denoted by a large triangle. For reference, the grey shading and dashed black line correspond to ± 10% error and 0% error respectively. For each stream, the right y-axis depicts the historical percentage of total wet season streamflow contained in each decile averaged across the 10-member ensemble. Subplots are organized so that as you move down the rows, streams transition from rain- to snow dominated over the historical period.

## REFERENCES


Barnett, T. P., & Pierce, D. W. (2009). Sustainable water deliveries from the Colorado River in a changing climate. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(18), 7334–7338. https://doi.org/10.1073/pnas.0812762106

California Department of Water Resources Climate Change Technical Advisory Group. (2015). *Perspectives and Guidance for Climate Change Analysis*. Retrieved from http://www.water.ca.gov/climatechange/docs/2015/Perspectives_Guidance_Climate_Change_Analysis.pdf

Cayan, D. R., Kammerdiener, S. A., Dettinger, M. D., Caprio, J. M., & Peterson, D. H. (2001). Changes in the Onset of Spring in the Western United States. *Bulletin of the American Meteorological Society*, *82*(3), 399–415. https://doi.org/10.1175/1520-0477(2001)082<0399:CITOOS>2.3.CO;2

Das, T., Pierce, D. W., Cayan, D. R., Vano, J. A., & Lettenmaier, D. P. (2011). The importance of warm season warming to western U.S. streamflow changes. *Geophysical Research Letters*, *38*(23), 1–5. https://doi.org/10.1029/2011GL049660

Dettinger, M. D., Ralph, F. M., Das, T., Neiman, P. J., & Cayan, D. R. (2011). Atmospheric Rivers, Floods and the Water Resources of California. *Water*, *3*(2), 445–478. https://doi.org/10.3390/w3020445

Fowler, H. J., Blenkinsop, S., & Tebaldi, C. (2007). Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. *International Journal of Climatology*, *27*(12), 1547–1578. https://doi.org/10.1002/joc.1556

Gershunov, A., Shulgina, T., Clemesha, R. E. S., Guirguis, K., Pierce, D. W., Dettinger, M. D., … Ralph, F. M. (2019). Precipitation regime change in Western North America: The role of Atmospheric Rivers. *Scientific Reports*, *9*(1), 1–11. https://doi.org/10.1038/s41598-019-46169-w

Gonzalez, P. (U. S. N. P. S., Garfin, G. M. (University of A., Breshears, D. D., Brooks, K. M., Brown, H. E., Elias, E. H., … Udal, B. H. (2018). Fourth National Climate Assessment: Southwest. *Impacts, Risks, and Adaptation in the United States: Fourth*

*National Climate Assessment*, *II*, 1101–1184. https://doi.org/10.7930/NCA4.2018.CH25

Gudmundsson, L., Bremnes, J. B., Haugen, J. E., & Engen-Skaugen, T. (2012). Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods. *Hydrology and Earth System Sciences*, *16*(9), 3383–3390. https://doi.org/10.5194/hess-16-3383-2012

Hagemann, S., Chen, C., Haerter, J. O., Heinke, J., Gerten, D., & Piani, C. (2011). Impact of a statistical bias correction on the projected hydrological changes obtained from three GCMs and two hydrology models. *Journal of Hydrometeorology*, *12*(4), 556–578. https://doi.org/10.1175/2011JHM1336.1

He, M., Anderson, M., Schwarz, A., Das, T., Lynn, E., Anderson, J., … Arnold, W. (2019). Potential changes in runoff of California's major water supply watersheds in the 21st century. *Water (Switzerland)*, *11*(8). https://doi.org/10.3390/w11081651

Hewitson, B. C., Daron, J., Crane, R. G., Zermoglio, M. F., & Jack, C. (2014). Interrogating empirical-statistical downscaling. *Climatic Change*, *122*(4), 539–554. https://doi.org/10.1007/s10584-013-1021-z

Hill, M. T., Platts, W. S., & Beschta, R. L. (1991). Ecological and geomorphological concepts for instream and out-of- channel flow requirements. *Rivers*, *2*(3), 198–210.

Huning, L. S., & AghaKouchak, A. (2018). Mountain snowpack response to different levels of warming. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(43), 10932–10937. https://doi.org/10.1073/pnas.1805953115

Kalra, A., Piechota, T. C., Davies, R., & Tootle, G. A. (2008). Changes in U.S. Streamflow and Western U.S. Snowpack. *Journal of Hydrologic Engineering*, *13*(3), 156–163. https://doi.org/10.1061/(asce)1084-0699(2008)13:3(156)

Knowles, N., Cronkite-Ratcliff, C., Pierce, D. W., & Cayan, D. R. (2018). Responses of Unimpaired Flows, Storage, and Managed Flows to Scenarios of Climate Change in the San Francisco Bay-Delta Watershed. *Water Resources Research*, *54*(10), 7631–7650. https://doi.org/10.1029/2018WR022852

Knowles, Noah, & Cronkite-Ratcliff, C. (2018). *Modeling Managed Flows in the Sacramento/San Joaquin Watershed, California, Under Scenarios of Future Change for CASCaDE2*. 1–38. https://doi.org/https://doi.org/10.3133/ofr20181101

Li, H., Sheffield, J., & Wood, E. F. (2010). Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *Journal of Geophysical Research Atmospheres*, *115*(10). https://doi.org/10.1029/2009JD012882

Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research*. https://doi.org/10.1029/94jd00483

Maraun, D. (2013). Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *Journal of Climate*, *26*(6), 2137–2143. https://doi.org/10.1175/JCLI-D-12-00821.1

Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J. M., … Mearns, L. O. (2017). Towards process-informed bias correction of climate change simulations. *Nature Climate Change*, *7*(11), 764–773. https://doi.org/10.1038/nclimate3418

Maurer, E. P., & Pierce, D. W. (2014). Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean. *Hydrology and Earth System Sciences*, *18*(3), 915–925. https://doi.org/10.5194/hess-18-915-2014

Michelangeli, P. A., Vrac, M., & Loukos, H. (2009). Probabilistic downscaling approaches: Application to wind cumulative distribution functions. *Geophysical Research Letters*, *36*(11), 2–7. https://doi.org/10.1029/2009GL038401

Mioduszewski, J. R., Rennermalm, A. K., Robinson, D. A., & Wang, L. (2015). Controls on spatial and temporal variability in Northern Hemisphere terrestrial snow melt timing, 1979-2012. *Journal of Climate*, *28*(6), 2136–2153. https://doi.org/10.1175/JCLI-D-14-00558.1

Mote, P. W., Hamlet, A. F., Clark, M. P., & Lettenmaier, D. P. (2005). DECLINING MOUNTAIN SNOWPACK IN WESTERN NORTH AMERICA*. *Bulletin of the American Meteorological Society*, *86*(1), 39–50. https://doi.org/10.1175/BAMS-86-1-39

Mote, P. W., Li, S., Lettenmaier, D. P., Xiao, M., & Engel, R. (2018). Dramatic declines in snowpack in the western US. *Npj Climate and Atmospheric Science*, *1*(1). https://doi.org/10.1038/s41612-018-0012-1

Panofsky, H. A., & Brier, G. W. (1968). Some Applications of Statistics to Meteorology. In *College of Mineral Industries, The Pennsylvania State University*.

Patterson, N. K., Lane, B. A., Sandoval-Solis, S., Pasternack, G. B., Yarnell, S. M., & Qiu, Y. (2020). A hydrologic feature detection algorithm to quantify seasonal components of flow regimes. *Journal of Hydrology*, *585*(August 2019), 124787. https://doi.org/10.1016/j.jhydrol.2020.124787

Pederson, G. T., Gray, S. T., Ault, T., Marsh, W., Fagre, D. B., Bunn, A. G., … Graumlich, L. J. (2011). Climatic controls on the snowmelt hydrology of the northern Rocky Mountains. *Journal of Climate*, *24*(6), 1666–1687. https://doi.org/10.1175/2010JCLI3729.1

Petts, G. E. (1996). Water Allocation to Protect River Ecosystems. *Regulated Rivers: Research & Management*, *12*(4–5), 353–365. https://doi.org/10.1002/(SICI)1099-1646(199607)12:4/5<353::AID-RRR425>3.0.CO;2-6

Pierce, D. W., & Cayan, D. R. (2013). The uneven response of different snow measures to human-induced climate warming. *Journal of Climate*, *26*(12), 4148–4167. https://doi.org/10.1175/JCLI-D-12-00534.1

Pierce, D. W., Cayan, D. R., Maurer, E. P., Abatzoglou, J. T., & Hegewisch, K. C. (2015). Improved bias correction techniques for hydrological simulations of climate change. *Journal of Hydrometeorology*, *16*(6), 2421–2442. https://doi.org/10.1175/JHM-D-14-0236.1

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., Das, T., Cayan, D. R., Maurer, E. P., Miller, N. L., Bao, Y., … Tyree, M. (2013). Probabilistic estimates of future changes in California temperature and precipitation using statistical and dynamical downscaling. *Climate Dynamics*, *40*(3–4), 839–856. https://doi.org/10.1007/s00382-012-1337-9

Poff, N. L. R., Allan, J. D., Bain, M. B., Karr, J. R., Prestegaard, K. L., Richter, B. D., … Stromberg, J. C. (1997). The natural flow regime: A paradigm for river conservation and restoration. *BioScience*, *47*(11), 769–784. https://doi.org/10.2307/1313099

Rajagopalan, B., Nowak, K., Prairie, J., Hoerling, M., Harding, B., Barsugli, J., … Udall, B. (2009). Water supply risk on the Colorado River: Can management mitigate? *Water Resources Research*, *45*(8), 1–7. https://doi.org/10.1029/2008WR007652

Richter, B. D., Baumgartner, J. V., Powell, J., & Braun, D. P. (1996). A Method for Assessing Hydrologic Alteration within Ecosystems. *Conservation Biology*, *10*(4), 1163–1174. https://doi.org/10.1046/j.1523-1739.1996.10041163.x

Salathé, E. P. (2003). Comparison of various precipitation downscaling methods for the simulation of streamflow in a rainshadow river basin. *International Journal of Climatology*, *23*(8), 887–901. https://doi.org/10.1002/joc.922

Serreze, M. C., Clark, M. P., Armstrong, R. L., McGinnis, D. A., & Pulwarty, R. S. (1999). Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data. *Water Resources Research*, *35*(7), 2145–2160. https://doi.org/10.1029/1999WR900090

Stewart, I. T., Cayan, D. R., & Dettinger, M. D. (2005). Changes toward earlier streamflow timing across western North America. *Journal of Climate*, *18*(8), 1136–1155. https://doi.org/10.1175/JCLI3321.1

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

Teutschbein, C., & Seibert, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of Hydrology*, *456–457*, 12–29. https://doi.org/10.1016/j.jhydrol.2012.05.052

Thrasher, B., Maurer, E. P., McKellar, C., & Duffy, P. B. (2012). Technical Note: Bias correcting climate model simulated daily temperature extremes with quantile mapping. *Hydrology and Earth System Sciences*, *16*(9), 3309–3314. https://doi.org/10.5194/hess-16-3309-2012

Udall, B., & Overpeck, J. (2017). The twenty-first century Colorado River hot drought and implications for the future. *Water Resources Research*, *53*(3), 2404–2418. https://doi.org/10.1002/2016WR019638

Wood, A. W., Maurer, E. P., Kumar, A., & Lettenmaier, D. P. (2002). Long-range experimental hydrologic forecasting for the eastern United States. *Journal of Geophysical Research: Atmospheres*, *107*(20), ACL 6-1-ACL 6-15. https://doi.org/10.1029/2001JD000659

Yarnell, S. M., Petts, G. E., Schmidt, J. C., Whipple, A. A., Beller, E. E., Dahm, C. N., … Viers, J. H. (2015). Functional Flows in Modified Riverscapes: Hydrographs, Habitats and Opportunities. *BioScience*, *65*(10), 963–972. https://doi.org/10.1093/biosci/biv102