1. Introduction

California's Mediterranean climate is characterized by hot and dry summers, while most of its rain falls in winter (Kottek et al., 2006). The seasonal differences in water supply make California vulnerable to long-term droughts and short-term extreme rainfall events, which result in flash floods and mudslides (Hanak et al., 2011). With recent climate models predicting more frequent extreme precipitation in the western United States in the future (Dominguez et al., 2012; Leung et al., 2004; Meehl et al., 2000), year-to-year predictability of California's precipitation characteristics would be helpful in ensuring sustainable water management practices and preparedness for extreme events.

California's geographical diversity and latitudinal extent make the prediction of California's extreme precipitation events challenging. Because precipitation statistics are not spatially homogeneous, extreme events in different regions often have different characteristics and origins. Additionally, the focus on the total annual precipitation in many previous studies (Berg & Hall, 2015; Granger, 1979; Haston & Michaelsen, 1997) has obscured the extent to which the intensity and frequency of extreme events have different impacts and potentially different predictability characteristics.

Most of California's winter precipitation is imported by North Pacific storms. The location and variability of these storms are linked to the underlying Pacific sea surface temperatures (SSTs) and the associated atmospheric teleconnections (Amini & Straus, 2019; Baggett & Lee, 2015; Dong et al., 2018; Liu et al., 2016; Palmer & Mansfield, 1984; Shaw et al., 2016), such as the El Niño-Southern Oscillation (ENSO), which dominates SST variability (Cayan et al., 1999; Mo & Higgins, 1998; Redmond & Koch, 1991; Straus & Shukla, 1997; Trenberth, 2011). Other studies (Brown, 2011; Brown & Comrie, 2004; McCabe & Dettinger, 1999; Sung et al., 2014; Tamaddun et al., 2016) have found that the relationship between ENSO and California's precipitation is additionally modulated by the Pacific Decadal Oscillation (PDO). However, because ENSO and PDO, as traditionally defined, are correlated, their use as independent predictors, for example, of California rainfall characteristics has so far been limited. Wills et al. (2018) recently constructed closely related but
uncorrelated indices PDO* and ENSO* via a low-frequency component analysis (LFCA). Here, we explore their use in predicting California precipitation characteristics.

We perform an analysis of interannual precipitation variability in four geographically different regions of California (Figure 1a). For each region, we analyze precipitation extremes and quantify the contribution of the number of rainy days and their intensity to the annual totals. We further explore the physical links to the North Pacific SSTs and large-scale flow characteristics. We quantify the link between Pacific SSTs and California precipitation extremes using a linear regression model, with ENSO* and PDO* as the predictors.

2. Data and Methods

Daily gridded precipitation data (0.25° × 0.25°) between July 1, 1949, and June 30, 2020, were obtained from CPC US Unified Precipitation data provided by NOAA/OAR/ESRL PSL (https://www.esrl.noaa.gov/psd/data/gridded/data.unified.daily.conus.html). The data were divided into four Californian regions in
Figure 1, namely Southern (S), Northern (N), Central Coastal (CC), and Central Mountainous (CM). The gridded data were averaged over each region, producing four time series upon which we performed the statistical analysis. Each time series was partitioned into water years, defined to stretch from July 1 to June 30 (allowing each entire wet season to be in one water year).

We used correlation and linear regression methods to analyze the contribution of annual precipitation associated with specific events (TP\textsubscript{event}) to the annual total precipitation (TP\textsubscript{sum}) for each water year. TP\textsubscript{event} is the annual sum of daily precipitation that occurred only during rainy days (i.e., days when precipitation exceeds a threshold, which we define as a rainy-day threshold). We then split TP\textsubscript{event} into the average amount of precipitation per day in a water year (intensity, I) and the number of rainy days in the same period (duration, D). Therefore, TP\textsubscript{event} = D × I. TP\textsubscript{sum} is the total annual precipitation from all days; it does not depend on the rainy-day threshold. By contrast, TP\textsubscript{event}, D, and I do depend on the rainy-day threshold.

Existing studies have used different thresholds to define a rainy day. For example, Pierce et al. (2013), Caldwell et al. (2009), and Englehart and Douglas (1985) chose an absolute value of daily precipitation across all their study regions. This can be problematic when comparing climatologically different regions, such as those investigated here. Alternatively, studies on extreme events tend to use percentile-based thresholds (Mass et al., 2011; Polade et al., 2017) to define a rainy day, which here would be the percentage of TP\textsubscript{sum} explained by TP\textsubscript{event}. However, the extreme events identified by this approach may have different origins in the different regions. Here, we present our results both in terms of value-based thresholds (rainy-day threshold) and percentile-based thresholds (TP\textsubscript{event}/TP\textsubscript{sum}). We examine TP\textsubscript{event}, D, and I under different thresholds by varying the rainy-day threshold between 1 and 25 mm day\textsuperscript{-1} with an interval of 0.1 mm day\textsuperscript{-1}. The corresponding percentile-based thresholds are shown in Figure 1b.

Using the approach of Wills et al. (2018), we calculate the mutually uncorrelated ENSO* and PDO* indices for 1948–2020 by applying LFCA (https://github.com/rcjwills/lfca) to the NOAA Extended Reconstruc ted monthly SST data set v5. These are calculated using the first three empirical orthogonal functions of SST over the Pacific Ocean between 45°S and 70°N. The PDO* and ENSO* indices correspond to the second and third low-frequency components, and yield qualitatively comparable patterns to those of the (correlated) traditional PDO and ENSO indices (Wills et al., 2018). We use a multi-linear regression model to cross-correlate ENSO* and PDO* of each month in the preceding year with TP\textsubscript{sum}, TP\textsubscript{event}, I, and D of a water year. We did this for a range of rainy-day thresholds to examine the different predictability of extreme and non-extreme precipitation.

### 3. Results

Figure 1b shows how much of the annual total precipitation TP\textsubscript{sum} in California accumulated between July 1948 and June 2020 is captured by the rainy-day precipitation, TP\textsubscript{event}, as a function of the rainy-day threshold. For reference, we also supply the raw time series of TP\textsubscript{sum} in Figure S1. In the wettest region, N, the precipitation fraction captured by the rainy-day precipitation decays more slowly with the rainy-day threshold than in the other, drier regions, meaning that the N region exhibits a more continuous range of moderate to intense events. Thus, different rainy-day thresholds have a regionally different effect on the type of precipitation events that are being captured in TP\textsubscript{event}.

The interannual variability of TP\textsubscript{sum} is strongly correlated with TP\textsubscript{event} and D, and less strongly with I, across different rainy-day thresholds (Figure 1c). The sensitivity of the correlations to the rainy-day threshold is comparable across all regions and for all statistics. Pearson’s r of TP\textsubscript{sum} with TP\textsubscript{event} decreases from 1 as the threshold increases. Pearson’s r of TP\textsubscript{sum} with D is maximal at a threshold for TP\textsubscript{event}/TP\textsubscript{sum} between 50% and 75%, which is greater than 10 mm day\textsuperscript{-1} in N and less than 10 mm day\textsuperscript{-1} in other regions. Pearson’s r of TP\textsubscript{sum} with I is greatest when the threshold is 1 mm day\textsuperscript{-1}, and it rapidly decreases as the threshold increases. These correlations show that, in all regions, California’s interannual precipitation variability is primarily determined by the number of rainy days, D, as opposed to the intensity on rainy days, I.

Despite the regional similarity in the correlations of the precipitation event statistics with the interannual variability of total water year precipitation (July–June), correlations of the interannual variability with monthly ENSO* and PDO* from the April preceding the water year to March of the water year are markedly
different across the four regions (Figure 2). The correlation of ENSO* and PDO* with total water year precipitation is stronger in S and CC than in N and CM.

In N and CM, the interannual variability of TP<sub>sum</sub> correlates weakly and not statistically significantly with ENSO* and PDO* for all months, except from December to March in CM. There is some significant positive correlation of ENSO* and PDO* with TP<sub>event</sub> for low rainy-day thresholds (below 5 mm day<sup>-1</sup>), and with I for higher rainy-day thresholds (above 15 mm day<sup>-1</sup>).

In S and CC, the correlations of interannual variability of TP<sub>sum</sub>, TP<sub>event</sub>, and D with ENSO* and PDO* for all months are positive and statistically significant, except for April in CC. TP<sub>event</sub> and D defined by lower rainy-day thresholds are more correlated with ENSO* and PDO*, especially from June to March. On the contrary, the interannual variability of I in S and CC is weakly and rarely statistically significantly correlated with ENSO* and PDO*. With D being the main driver of TP<sub>sum</sub> variability, its significant correlation with ENSO* and PDO* suggests potential predictive power of the two SST indices for precipitation in these regions.

The relative predictive skill of ENSO* and PDO* shifts with the lead time (i.e., the month of the previous year from which SST was correlated with precipitation of the current year). The correlation coefficients for ENSO* and PDO* are comparable in spring/early summer. We show explicitly in the Supplementary

---

**Figure 2.** Correlations of ENSO* and PDO* with California rainfall statistics of water years (July–June). Top row: coefficients of multi-linear regression of TP<sub>sum</sub> with monthly ENSO* (red) and PDO* (blue). The Pearson’s r coefficient of the multi-linear regression is shown in gray. The solid line and solid circles represent statistically significant correlations with p < 0.05, where the null hypothesis that zero correlation has a greater absolute Pearson’s r-value than the calculated one was rejected. Dashed lines and open circles denote insignificant correlations. The monthly ENSO* and PDO* indices are evaluated between April preceding the water year and March during the water year. Bottom three rows: Pearson’s r values resulting from correlating TP<sub>event</sub>, D, and I at different rainy-day thresholds (vertical axis) with monthly ENSO* and PDO* (horizontal axis) using multi-linear regression. Stippling indicates significant correlations with p < 0.05. ENSO, El Niño-Southern Oscillation; PDO, Pacific Decadal Oscillation.
Information (Figure S4) that the significantly higher correlation coefficients in spring/summer are attributable to the inclusion of PDO*, especially in S and CC. However, in winter, the ENSO* correlation is much stronger, and the inclusion of PDO* does not improve Pearson’s $r$ substantially in any of the regions.

Finally, we can approximately predict interannual variations of $D$, $T_{\text{event}}$, and $T_{\text{sum}}$ from the interannual variations of ENSO* and PDO*, using the following regression models:

$$D(\tau, t) = b_1(\tau) \text{ENSO}^* (t) + b_2(\tau) \text{PDO}^* (t) + \epsilon(\tau, t) \quad (1a)$$

$$T_{\text{event}}(\tau, t) = D(\tau, t) I(\tau, t) \approx [b_1(\tau) \text{ENSO}^* (t) + b_2(\tau) \text{PDO}^* (t)] I(\tau) \quad (1b)$$

$$T_{\text{sum}}(t) = g(\tau)^{-1} T_{\text{event}}(\tau, t) \approx g(\tau)^{-1} [b_1(\tau) \text{ENSO}^* (t) + b_2(\tau) \text{PDO}^* (t)] I(\tau). \quad (1c)$$

Here, $\tau$ is the rainy-day threshold, $b_1$ and $b_2$ are the multi-linear regression coefficients, and $\epsilon$ is the zero-mean residual noise of the linear regression. The function $g(\tau)$ is the normalized cumulative PDF displayed in Figure 1b. Because of the weak and statistically insignificant correlation of $I$ with ENSO* and PDO*, we use the time-mean precipitation event intensity $I(\tau)$ in our statistical model. The mean event intensity $I(\tau)$ is the multi-year average of $I(\tau, t)$ at a given $\tau$. All of the variables are standardized to a mean 0 and standard deviation 1. The significance intervals were calculated by bootstrapping and training the multi-linear regression model with 70% of the data.

We choose to demonstrate our regression models with ENSO* and PDO* in June because that is when the two SST indices become markedly predictive of the interannual variability of $T_{\text{sum}}$ (Figure 2). By using June SSTs, the model predicts California precipitation (which primarily occurs in winter) approximately 6 months in advance. In all regions, the correlations of the June ENSO* and PDO* with $T_{\text{sum}}$ via $D$ are significant and comparably important in the CC and S regions (Figure 2).

In general, the estimates of $b_1$, $b_2$, and $\epsilon$ produce better predictions if the rainy days are defined using a moderate threshold (e.g., 50%–70% $T_{\text{event}}/T_{\text{sum}}$). For example, in the S region, when $\tau$ is 5.7 mm day$^{-1}$ (threshold at which the correlation of ENSO* and PDO* with $D$ is maximized), Pearson’s $r$ is 0.5, the bootstrapped medians of the coefficients are $b_1 = 0.44$ days year$^{-1}$ and $b_2 = 0.35$ days year$^{-1}$, and $g(\tau) = 0.56$ (indicated by the black marker in Figure 1b). The bootstrapped median of the predicted time series of $T_{\text{sum}}$ in Figure 3 captures 18% of the interannual variability of the observed $T_{\text{sum}}$. We also repeated the same prediction but using the two SST indices from previous July and August. On average, the two SST indices in the previous
summer (June, July, and August) can capture 20% of the interannual variability of the observed TP\textsubscript{sum}. The predictions for the three other regions account for far less of the total annual precipitation (Figure S5) since ENSO* and PDO* only weakly correlate with \(D\) in those regions.

4. Discussion

Despite the regional differences in California’s geography, throughout the state, the number of rainy days is more important than the rainfall intensity in its contribution to the total rainy-day precipitation. That is, interannual precipitation variations are primarily driven by variations in the number of rainy days, suggesting that they are controlled mainly by large-scale weather systems associated with the North Pacific storm track. Wet years have more rainy days because the large-scale circulation in wet years guides more storm track eddies toward California (Branstator, 1995; Higgins & Schubert, 1993; Lau, 1988). This is particularly apparent in the Southern and Central Coastal regions, which are located under the upper-level climatological winter jet (Figure S2). This jet guides storm track eddies, and its latitudinal shifts determine which regions of the West Coast are impacted by those storms and their precipitation.

Directly correlating the latitude of this jet (Archer & Caldeira, 2008) with the number of rainy days (Figure S3) yields negative correlations over all sub-regions, consistent with the notion that the jet is farther south (directed toward California) in California’s wet years. However, the magnitude of these correlations is small, because the jet latitude is a relatively noisy statistic.

The jet location is modulated by the SSTs associated with both ENSO* and PDO*. ENSO* and PDO* are smoother measures of the large-scale climatological changes than the jet latitude, which is why they yield more robust and statistically significant correlations. Time-lagged correlations of multi-linear regression reveal that a substantial fraction (up to 20% of the variance) of Southern California’s rainfall variations is potentially predictable from ENSO* and PDO* of the preceding summer, both with similar importance. Although the amount of total annual precipitation accounted for by both winter ENSO* and winter PDO* is comparable to using both indices from the early summer, ENSO* is a more important predictor than PDO* at shorter lead times, that is, from October to March.

The different correlations of SST with precipitation across the four regions have been explained in previous papers. As Brown and Comrie (2004), Wise (2010), Sung et al. (2014), and others have discussed, Pacific SST variations drive a dipole pattern in precipitation over the western United States. Since the N and CM regions are located in the transition zone of that dipole, their precipitation correlates less well with the SSTs. Instead, these regions are more susceptible to influences of the local geography. This poses a challenge for seasonal to sub-seasonal predictability of precipitation in those regions.

The potential predictability of total precipitation in California has been previously linked to the modulation of the storm track by the Interdecadal Pacific Oscillation (IPO) (Dai, 2013; Dong & Dai, 2015), an SST pattern over the Pacific (Chang & Fu, 2002; Henley et al., 2015; Lau, 1988; Liu & Alexander, 2007; Trenberth & Hurrell, 1994). IPO is often considered to include the SST signals from both ENSO and PDO (Folland et al., 2002; Verdon & Franks, 2006). ENSO* and PDO* together account for 94% of IPO variance. Using IPO alone, or in combination with the other indices, does not provide additional predictability beyond that implied by ENSO* and PDO* (Figure S4).

There are other large-scale atmospheric patterns of variability that regulate the spatial storm track position and SST anomalies, such as the Arctic Oscillation (Strong & Davis, 2008), the North Pacific Oscillation/West Pacific teleconnection (Linkin & Nigam, 2008), and the North Pacific Gyre Oscillation (Liu et al., 2016). We included these other modes of internal variability in our analysis, but they did not add predictive power. Knowledge of ENSO and PDO variability seems to be key for understanding how the large-scale circulation deflects the storm tracks and brings precipitation to Southern California.
5. Conclusions

We discussed the interannual variability of California precipitation and its decomposition into the number of rainy days and precipitation intensity under value-based and percentile-based thresholds. The results show that the frequency of rainy events is the most important determinant of interannual precipitation across all Californian regions, regardless of the rainy-day threshold. This highlights the large-scale impact of the North Pacific storm track, whereby interannual precipitation events reflect the landfall of storm track eddies. The eddy propagation itself is largely driven by variations of the Pacific SSTs on these timescales. We showed that the summer values of ENSO* and PDO*, two uncorrelated indices derived from SSTs, can be used to predict part of the annual precipitation variations in the Southern and Central Coastal regions on sub-seasonal to seasonal timescales. The correlations of June ENSO* and PDO* with TP\textsubscript{event} are directly linked to the correlations with the number of rainy days. The combined predictive power of ENSO* and PDO* persists from summer to winter for precipitation in Southern California, while the relative importance of PDO* increases with lead time. For predictions based on the previous summer, PDO* is as important a predictor as ENSO*. In other parts of California, where the effect of the Pacific SSTs is less pronounced, precipitation is less predictable using either of the two SST indices.

Data Availability Statement

The data used in this study can be downloaded from [https://data.caltech.edu/records/1445](https://data.caltech.edu/records/1445).

References


Acknowledgments

This work was supported by NSF (Grant no. AGS-1760402) and by Eric and Wendy Schmidt (by recommendation of the Schmidt Futures program).

The data used in this study can be downloaded from [https://data.caltech.edu/records/1445](https://data.caltech.edu/records/1445).


