

Present and Past Sea Surface Temperatures: A Recipe for Better Seasonal Climate Forecasts

MATTHEW B. SWITANEK, JOSEPH J. BARSUGLI, AND MICHAEL SCHEUERER

CIRES, University of Colorado Boulder, and NOAA/Physical Sciences Laboratory, Boulder, Colorado

THOMAS M. HAMILL

NOAA/Physical Sciences Laboratory, Boulder, Colorado

(Manuscript received 3 December 2019, in final form 29 April 2020)

ABSTRACT

Monthly tropical sea surface temperature (SST) data are used as predictors to make statistical forecasts of cold season (November–March) precipitation and temperature for the contiguous United States. Through the use of the combined-lead sea surface temperature (CLSST) model, predictive information is discovered not just in recent SSTs but also from SSTs up to 18 months prior. We find that CLSST cold season forecast anomaly correlation skill is higher than that of the North American Multimodel Ensemble (NMME) and the SEAS5 model from the European Centre for Medium-Range Weather Forecasts (ECMWF) when averaged over the United States for both precipitation and 2-m air temperature. The precipitation forecast skill obtained by CLSST in parts of the Intermountain West is of particular interest because of its implications for water resources. In those regions, CLSST dramatically improves the skill over that of the dynamical model ensembles, which can be attributed to a robust statistical response of precipitation in this region to SST anomalies from the previous year in the tropical Pacific.

1. Introduction

Driven by recent severe droughts across much of the western United States (Dettinger et al. 2015; Prein et al. 2016; Mazdiyasi and AghaKouchak 2015; Cheng et al. 2016), there is unprecedented interest in improving seasonal climate predictions. Dramatic declines in snowpack have been observed over the last century throughout the Intermountain West (Philip et al. 2018; Xiao et al. 2018), while over the last decade water storage in Lake Mead along the Colorado River was at its lowest point since the reservoir was filled (Udall and Overpeck 2018; Sullivan et al. 2019). America's agricultural juggernaut, the state of California, has seen particularly devastating drought conditions (Diffenbaugh et al. 2015; Mao et al. 2015). In California, irrigation accounts for 74% of the total freshwater withdrawals (Dieter et al. 2018) and is used to produce over one-third of the country's vegetables and approximately two-thirds of the country's fruits and nuts. The abundance or scarcity of water resources across the

United States have large-scale implications on economies, agriculture, biodiversity, hydroelectricity, and peoples' own well-being. With such high stakes, it is incumbent upon climate scientists to produce forecasts that are as skillful as possible in order to assist planning that may mitigate the effects of extremes on water resource distribution.

El Niño–Southern Oscillation (ENSO) and its teleconnections have been observed to be the dominant modulator of seasonal-to-interannual precipitation and temperature variability for much of the world (Ropelewski and Halpert 1987) and North America in particular (Redmond and Koch 1991; Cayan et al. 1999; Guo et al. 2017; Kumar and Chen 2017). This has led many in the modeling community to focus on improving dynamical weather models' representation of ENSO and its influence on global climate (Vitart 2014; Zhu et al. 2015). However, recent progress has been slow in improving seasonal climate forecasts. There are substantial challenges associated with numerical prediction of tropical sea surface temperatures (SSTs) and their teleconnections. Numerical predictions of the ocean state in coupled systems can exhibit rapid drift from initial conditions (Capotondi et al. 2006; NAS 2016), and

Corresponding author: Matthew Switaneck, matt.switaneck@noaa.gov

DOI: 10.1175/WAF-D-19-0241.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) (www.ametsoc.org/PUBSReuseLicenses).

these biases can propagate to rest of the globe. The way forward likely requires progress on multiple fronts, including improved observations of the ocean state, development of coupled Earth system data assimilation, and improvements to all components (atmosphere, land, sea ice and ocean) of the climate models used to make seasonal forecasts.

Given the challenges with direct numerical predictions of ENSO and its modulation of climate variability, we consider the potential utility of linear statistical models, both as a stand-in for or complement to the numerical predictions and as a benchmark against which to evaluate them. Numerical modeling systems should be able to take advantage of sources of skill that are implicit in simple linear statistical models, while additionally being able to represent complex and non-linear interactions. Nonetheless, some have suggested that numerical models are nearing a predictability limit for tropical SSTs (Newman and Sardeshmukh 2017), in which case linear statistical forecast models may remain competitive with numerical models for quite some time.

Previous studies have given some context as to the state of statistical versus dynamical model performance (Barnston 1994; Cayan 1996; Cayan et al. 1999; McCabe and Dettinger 2002; Cook et al. 2018; Guo et al. 2017; Lee et al. 2017; Guan et al. 2012; Kumar and Chen 2017; Allen and Luptowitz 2017; Jong et al. 2016; Pan et al. 2019), where statistical seasonal climate forecasts have often been seen to be competitive to dynamical forecasts. Notably, Barnston (1994) used SST leads of a year in a combined linear statistical model. However, that work found predictive skill for temperature but not for precipitation, likely due to the much shorter training period and the sparsity of the precipitation data. Our methodology differs from these previous approaches in the way we make use of, and objectively combine, prior months' SST anomalies. Here we develop and use the combined-lead sea surface temperature (CLSST) model to forecast cold season (November–March) precipitation and 2-m temperature (see section 3). We focus on the cold season because of its importance in influencing snowpack and the water resources of the western United States. The CLSST model forecasts are always made in the October that precedes the cold season, and can then be considered 0.5-month lead-time forecasts. The CLSST model implements a statistical framework that is simple, rigorously validated, efficient and easily replicated. We then compare the skill of the cold season climate forecasts for CLSST, to the October initialized cold season forecasts of the North American Multimodel Ensemble (NMME; Kirtman et al. 2014b), and the European Centre

for Medium-Range Weather Forecasts (ECMWF) seasonal climate model SEAS5 (Johnson et al. 2019b).

2. Data

Sea surface temperature (SST) time series were computed using the NOAA Extended Reconstructed Sea Surface Temperature (ERSST) version 5 (Huang et al. 2017). The SST dataset are monthly averages at a $2^\circ \times 2^\circ$ resolution, and we use it for the period 1899–2018. The PRISM climate dataset (PRISM Climate Group 2019) was used to obtain total precipitation and mean temperature for the contiguous United States. The PRISM data are monthly averages where we further spatially upscaled from $1/24^\circ$ to $1/8^\circ$ using arithmetic averaging. The PRISM data is used for the period 1901–2018.

Accumulated precipitation totals and mean temperature values, from the PRISM dataset, were computed between 1901/02 and 2017/18 (referred to throughout this paper by the years in which the season began: 1901–2017) for the cold season period of November–March. We subsequently calculated areal averages for each hydrologic unit code (HUC) division 4 (Seaber et al. 1987). HUCs use six levels of spatial hierarchy to parse watersheds, represented by numeric codes 2 through 12 (where divisions 2 and 12 delineate the most coarse-scale and finescale resolutions, respectively). There are 202 HUCs in the contiguous United States (CONUS) at division 4, which provides the necessary spatial resolution for many large-scale decisions concerning water resources. Henceforth, we use HUC to refer to this level of spatial resolution.

Historical reforecasts of precipitation and temperature were obtained for the individual models of the North American Multimodel Ensemble (NMME; Kirtman et al. 2014a) in addition to more recent years of real-time forecasts (Kirtman et al. 2014c). The reforecast data and the real-time forecasts correspond to the years 1982–2010 and 2011–present, respectively. We collected these reforecasts and forecasts for individual months with model initialization in October. The ensemble mean of each of the seven contributing NMME models was then calculated for each individual month between November and March, then summed over the cold season, and finally spatially averaged across each HUC. To be consistent with the procedure used to obtain observed winter precipitation and temperature at each HUC, the NMME ensemble mean values were resampled to $1/8^\circ$, prior to averaging, where the 64 finer resolution gridcell anomaly values are simply equal to that of the containing 1° value.

Seasonal forecasts from ECMWF's long-range SEAS5 model were obtained for the years 1993–2016 (Johnson et al. 2019a). Ensemble monthly averages for

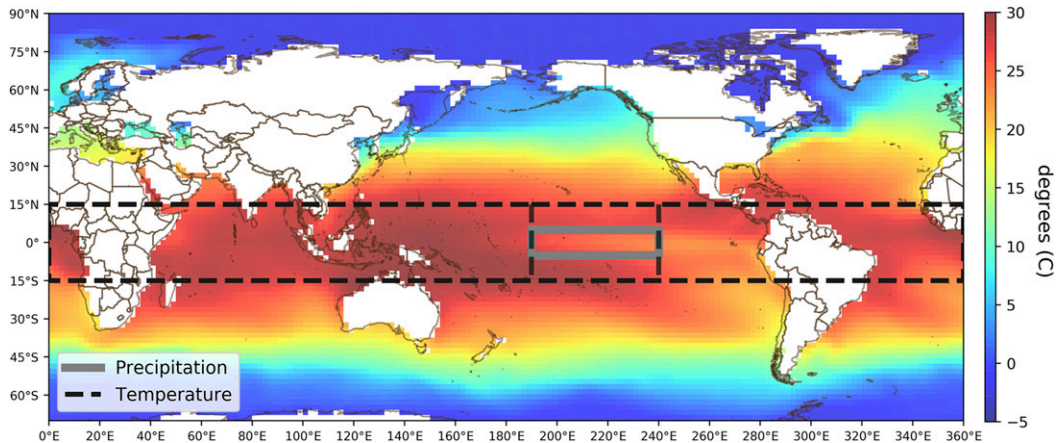


FIG. 1. CLSST predictor regions for precipitation and temperature. The sea surface temperatures shown are for the month of January 2017.

the individual months between November–March were computed with an October model initialization, then summed over the cold season. As with NMME, the data was resampled to $1/8^\circ$ and averaged across the individual HUCs.

Skill metric and statistical significance

As a measure of forecast skill, anomaly correlation skill (ACS) [Eq. (7.60) from Wilks (2006)] is used to compare the performance of the different models. The CLSST model is calibrated in the period 1901–1980 and validated in the period 1981–2017. Comparative ACS values were then evaluated with respect to the validation period 1993–2016 (length of ECMWF-SEAS5 record) for CLSST, NMME, and ECMWF-SEAS5, with an additional comparison made with respect to the period 1982–2017 for CLSST and NMME (length of NMME record).

We establish statistical significance of the ACS by applying the approach advocated by Wilks (2016), which accounts for test multiplicity and field significance (accounting for multiplicity is necessary since all 202 HUCs are considered simultaneously). Following Wilks (2016), we control the false discovery rate (FDR) (Benjamini and Hochberg 1995) at the level $\alpha_{\text{FDR}} = 0.10$ and reject local null hypotheses of zero ACS if the respective p values are smaller than the threshold value p_{FDR}^* [see appendix B and Wilks (2016) for details]. This method to control the FDR and account for test multiplicity is robust to spatial correlation (experiments in Wilks (2016) show that in this case the testing procedure becomes stringent).

3. Methods

a. SST predictors

Retrospective forecasts of precipitation are performed using one SST predictor time series, while temperature

forecasts are made using two predictors (see Fig. 1). The predictor time series used for precipitation is the area average of the SSTs in the Niño-3.4 domain located between 5°S – 5°N and 190°E – 240°E . Temperature uses two area average predictor time series. These are 1) the area average of the SSTs in the domain between 15°S – 15°N and 190°E – 240°E (a broader Niño-3.4 domain), and 2) a longitudinal average tropical SST predictor with its domain between 15°S – 15°N and 0° – 360° . The two predictor regions for temperature were chosen to optimize model performance in the calibration time period.

These area-weighted average SST time series were then standardized as follows and detrended. For each calendar month, m , the entire time series (1901–2017) is standardized by

$$\text{SSTZ}_m = \frac{\text{SST}_m - \mu_{\text{SST}_m}}{\sigma_{\text{SST}_m}}, \quad \text{for } m = 1, \dots, 12, \quad (1)$$

where μ_{SST_m} and σ_{SST_m} are the monthly mean and standard deviations of the time series obtained in the 1901–1980 calibration time period. Next, the time series were detrended with respect to their respective fitted linear least squares trend lines for the years 1901–1980, and separately for each month m . Additionally, we want to avoid too many positive standardized anomalies of our precipitation predictor. We therefore further remove the more recent trend that is contained in the validation period from the precipitation predictor time series. To simulate the making of actual forecasts, we detrended the precipitation predictor time series using only the data that would have been available at the time of the forecasts. Therefore, the value SSTZ_m , corresponding to year 1981 and month m , is detrended with respect to the linear least squares trend fit for years 1901–1981.

We similarly detrend the corresponding normalized SSTZ_m value for year 1982 with respect to the time period 1901–1982. This procedure is further applied for the rest of the years through 2017. In contrast to precipitation, we do not detrend the more recent validation periods of the temperature predictor time series. As a result, our CONUS temperature forecasts will implicitly contain forecast trends due to the trends of the SSTs.

b. CLSST model reforecasts of precipitation and temperature

The precipitation and temperature data, averaged over the HUCs, are both arranged into matrices of size 117 × 202 (117 winter values by 202 hydrologic unit codes). These two matrices are then normalized by subtracting the column means with respect to the 1901–1980 calibration period. The terms **A** and **B** now represent our anomaly precipitation and temperature matrices, respectively. Next, we obtain the covariance matrices of **A** and **B** using only the first 80 rows (that is our calibration period). Then, we separately perform eigen decompositions on our covariance matrices of **A** and **B**. Our principal components (PCs) are subsequently calculated over the entire period 1901–2017 as

$$\mathbf{U}_A = \mathbf{A} \cdot \mathbf{E}_A, \tag{2}$$

where **U_A** are our principal components for precipitation and **E_A** is our corresponding matrix of eigenvectors, columnwise (Wilks 2006). Similarly, for temperature we obtain our PCs for temperature with

$$\mathbf{U}_B = \mathbf{B} \cdot \mathbf{E}_B, \tag{3}$$

For both precipitation and temperature, we keep the leading three PCs that account for 57.0% and 85.0% of the cold season variance, respectively. Temperature forecasts are made using the paradigm: “forecast of opportunity,” where forecasts are made only when certain predictor conditions are met (i.e., predictor magnitude is sufficiently large). In contrast to temperature, precipitation is forecasted regardless of the magnitude of the predictor. Model fitting, which is outlined in the following paragraph, is applied for precipitation to all calibration years and for temperature to the cases in which the average of the two temperature predictors, $\overline{\text{SSTZ}_m}$, has an absolute value greater than 0.5. The parameters governing 1) the SST threshold used in the temperature model and, 2) the choice of how many leading PCs to predict, were determined by optimizing model skill over the calibration period 1901–1980.

A linear least squares model is used to make retrospective forecasts of precipitation and temperature, during October of each year, for the entire period 1901–2017. The model is fit to the calibration period, 1901–1980, and for precipitation takes the following form:

$$y_{l,p} = \beta_{0,l,p} + \beta_{1,l,p}x_{l,p}, \quad \text{for } l = 1, \dots, 18, \quad p = 1, 2, 3, \tag{4}$$

where x_l is our predictor SST time series at lead l and $y_{l,p}$ is the time series of winter season values for the current principal component p at lead l . Leads l range from the preceding September (1 month prior to forecast season) through April (18 months prior). For temperature, which uses two predictor time series and conditioned upon our threshold discussed above, the model fit takes the following form:

$$y_{l,p} = \begin{cases} \beta_{0,l,p} + \beta_{1,l,p}x_{1,l} + \beta_{2,l,p}x_{2,l}, & \text{for } l = 1, \dots, 18, \quad p = 1, 2, 3, \quad \text{if } \left| \overline{\text{SSTZ}_m} \right| > 0.5 \\ 0, & \text{if } \left| \overline{\text{SSTZ}_m} \right| \leq 0.5 \end{cases} \tag{5}$$

Regression coefficients, from Eqs. (4) and (5), are found between each lead-time SST time series and each of the leading PCs. The regression coefficients are obtained from the calibration period, separately for each SST lead, and are subsequently used to forecast the leading PCs for the entire time period. Then, these predicted PCs are transformed back into the original precipitation and temperature data space with

$$\mathbf{Z}_l = \mathbf{Y}_l \cdot \tilde{\mathbf{E}}^T, \tag{6}$$

where **Z_l** is our forecasted precipitation matrix (117 × 202), **Y_l** is our forecasted principal component matrix

(117 × 3), and $\tilde{\mathbf{E}}$ is our matrix of the first three eigenvectors (202 × 3) of either **E_A** for precipitation or **E_B** for temperature. At this point, anomaly correlation coefficients are calculated for each of the 202 HUCs in the calibration time period 1901–80. Now, we have anomaly correlation values for each lead, ranging between 1 and 18, and each HUC, ranging between 1 and 202. As a result of transforming the data and implementing the SST predictor(s) threshold (for temperature), we find that anomaly correlations can be negative in the calibration period. Ultimately, we are going to weight our forecasts as a function of the AC skills in the calibration period. Therefore, we now set any ACS values less than

0.0, obtained in the calibration period, equal to 0.0. As a result, we apply no weight to the forecasts that had negative skill in the calibration period. Retrospective forecasts for each of our HUCs in the validation period, 1981–2017, are then computed as

$$\text{fcst}_{l,h} = \frac{\sum_{l=1}^{l=18} z_{l,h} w_{l,h} d_l}{\sum_{l=1}^{l=18} w_{l,h} d_l}, \quad (7)$$

where d_l is a linear decay function ranging from 1.0 to 0.25 (for $l = 1, \dots, 18$, respectively), w is the anomaly correlation (AC) value in our calibration period, and z is the time series of precipitation or temperature (in our retransformed original data space) for SST lead l and HUC h . Again, the parameter d_l was chosen to optimize model skill in the calibration period.

4. Results

a. A source of improved forecast skill

Figure 2 shows the long-term historical anomaly correlation (AC) values (Wilks 2006) between cold season CONUS precipitation and different preceding months of Niño-3.4 SST anomalies for the period 1901–2017. Hydrologic unit codes [HUCs at division 4 (Seaber et al. 1987), where these units are natural hydrological basin delineations], with statistically significant ACs, are outlined using the standard test for zero AC based on Student's t distribution in combination with the approach advocated by Wilks (2016). We have chosen to perform our analysis of mean areal precipitation and temperature at these natural hydrologic basins, since our results can then be most easily interpreted and applicable to the domain of water resources. Additionally, we check whether serial (temporal) correlation was ever sufficiently large to require adjustments to sample size. Performing tests with null hypothesis of zero lag-1 autocorrelation between AC values across different years for each HUC and each lead time month (again controlling the FDR at $\alpha_{\text{FDR}} = 0.10$), we found no significant serial correlation, and therefore we did not make any sample size adjustments (this was also found to be the case for all subsequent analyses and figures).

Using Niño-3.4 SSTs with a lead time of 1 month (September SSTs correlated with NDJFM precipitation), the typical ENSO response pattern is highlighted. This is illustrated by the southern United States exhibiting a robust positive relationship (wetter than normal during El Niño) with Niño-3.4 SSTs, while parts of the northwestern United States experience an inverse relationship

(drier than normal during El Niño). This pattern becomes less pronounced as the lead time increases between preceding Niño-3.4 SST months and cold season precipitation. Using the preceding April's Niño-3.4 SSTs (Fig. 2f), there are few CONUS HUCs that exhibit statistical significance. However, by further increasing the lead time, we find statistically robust correlations return for the Central Intermountain West (CIW). The CIW here includes Northern California, Oregon, Nevada, Utah, northern Colorado, and southern Idaho. Throughout the CIW, we find a statistically robust pattern emerge with lead times between 9 and 16 months (Figs. 2i–p). This indicates, on average, that the CIW region has historically seen increased/decreased precipitation approximately 1–1.5 years following El Niño/La Niña events. We leveraged this lagged information to build and implement the CLSST model (see section 3).

b. Comparison of cold season forecast skill

Anomaly correlation skill (ACS) values of precipitation forecasts for CLSST, NMME, and ECMWF-SEAS5 can be seen in Fig. 3. The ACS differs from AC insofar that any strongly negative AC values from Fig. 2 can be interpreted as potentially useful predictive information, while ACS (which are used for the duration of the paper) must be positive if the forecasts performed skillfully superior to climatology. ACS is presented for each of the HUCs across CONUS. In contrast to Fig. 2, we now test whether ACS values are significantly positive. As a CONUS average, the CLSST model has modestly higher skill in forecasting winter precipitation than either NMME or ECMWF-SEAS5. The spatial distribution of the precipitation forecast skill of the CLSST generally follows that of the dynamical models, with a few notable exceptions. The NMME and ECMWF-SEAS5 models can be seen to perform better in southern Arizona and Texas. While CLSST exhibits a marked improvement in forecast skill across the CIW region. In the CIW region, Northern California, eastern Oregon, Nevada, northern Utah, northern Colorado, and Idaho all see substantial increases in skill with respect to the dynamical model forecasts. This performance improvement, using CLSST, is observed in both the shorter (1993–2016) and longer (1982–2017) validation periods (corresponding to the length of the reforecast periods of ECMWF-SEAS5 and NMME, respectively).

Figure 4 shows the CLSST model clearly outperforming NMME and ECMWF, on average, in its ability to forecast temperature. The pattern of skill in the longer validation period largely follows that of NMME, while greater differences emerge in the more recent validation period of 1993–2016. The CLSST has greater skill in the southern part of the United States eastward of central

Anomaly Correlation (1901-2017)

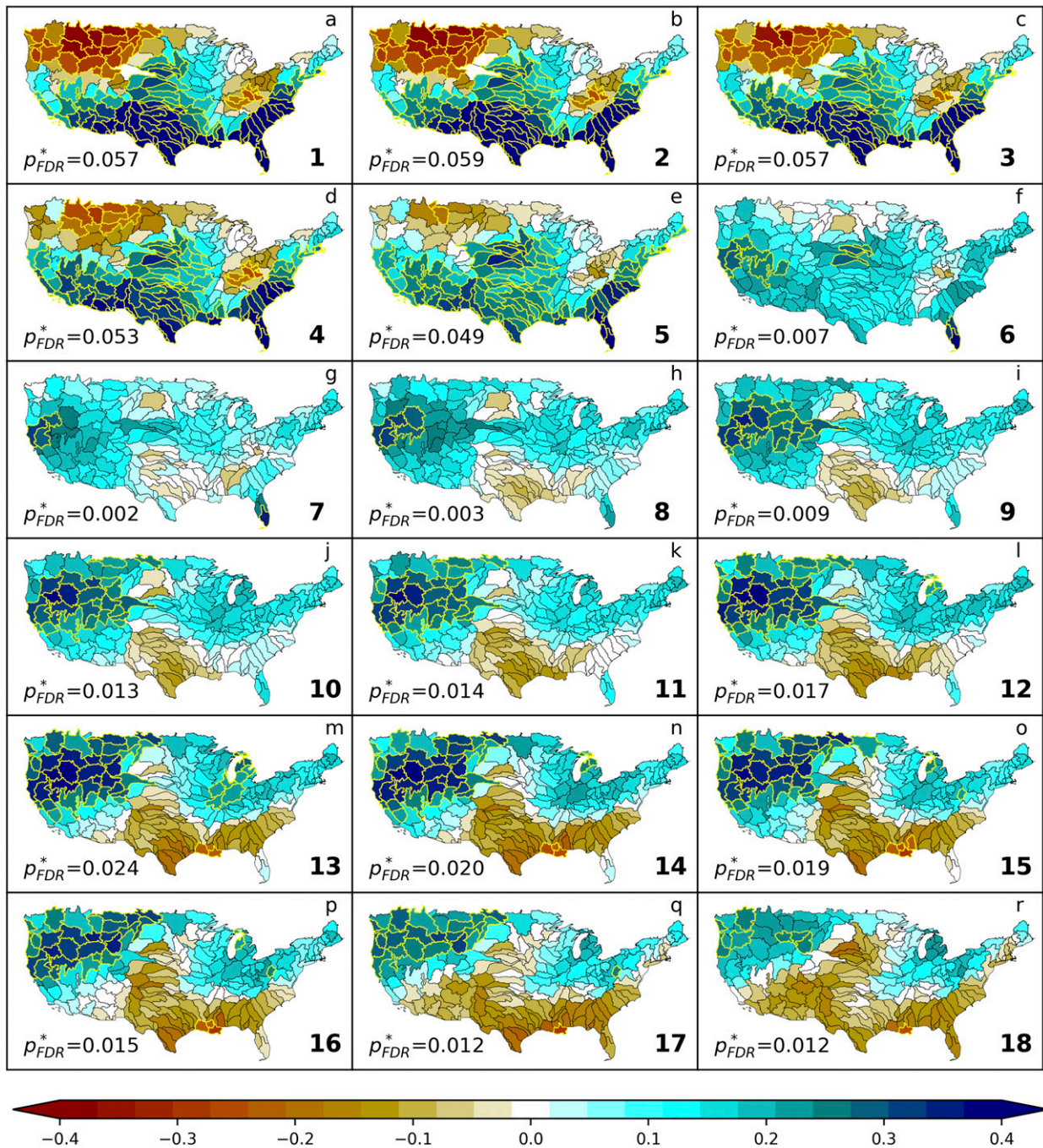


FIG. 2. Anomaly correlation (AC) values at various lags between the standardized and detrended tropical Pacific precipitation predictor time series (Niño-3.4) and each of the 202 hydrologic unit codes (HUCs). (a)–(r) The number of months prior to forecast initialization is shown in the bottom right of each subplot. At 1 month prior the correlation between the preceding September SSTs and cold season (NDJFM) precipitation is shown, 2 months prior is for the preceding August SSTs and cold season precipitation, then further stepping back in time through the April that is 18 months prior to forecast initialization. We assess whether ACs are significantly different from zero (positive or negative) by controlling the false discovery rate (FDR), with $\alpha_{FDR} = 0.10$. The resulting threshold value p_{FDR}^* is shown at the bottom left of each subplot. Statistically significant ACs have the HUCs outlined in yellow.

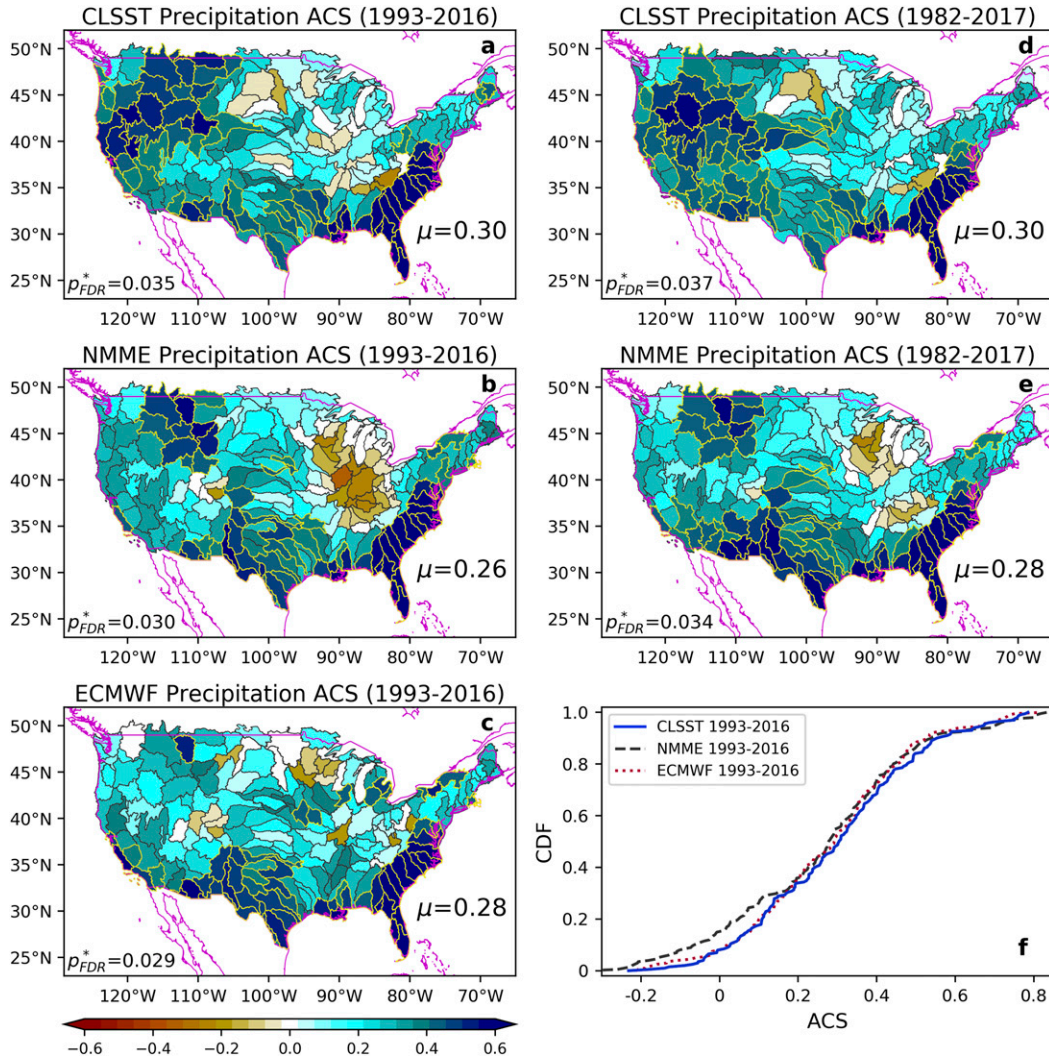


FIG. 3. Anomaly correlation skill (ACS) values for precipitation forecasts. (a)–(c) ACS for the CLSST model, NMME, and ECMWF-SEAS5 corresponding to ECMWF’s period of record. (d),(e) ACS for CLSST and NMME corresponding to NMME’s period of record. (f) The empirical cumulative distribution functions for the ACSs of the 202 major hydrologic units (HUCs) in (a)–(c). HUCs with significantly positive ACSs are outlined in yellow ($\alpha_{FDR} = 0.10$). The resulting threshold value p_{FDR}^* is shown at the bottom left of each subplot, while the CONUS average ACS is shown in the bottom right.

Texas, while NMME and ECMWF have greater skill across the Midwestern states for the shorter validation period. We found the precipitation skill across the three models to be largely unaffected by trends, while trends in temperature are seen to modestly inflate the skills shown in Fig. 4 (see appendix A for details). Note that in Fig. 4, the FDR control at level $\alpha_{FDR} = 0.10$ does not highlight any of the HUCs as statistically significant for both NMME and ECMWF in the shorter period. In other words, we cannot reject the global null hypothesis (i.e., we fail to establish field significance) that the NMME/ECMWF forecasts have zero skill at each HUC. A sensitivity analysis in which we recalculated statistical significance with $\alpha_{FDR} = 0.15$

(see appendix B, Fig. B1) shows that this modest increase in α_{FDR} leads to a much higher threshold value p_{FDR}^* (see appendix B, Fig. B2) and illustrates how minor changes in either the data or the FDR control can have a strong impact on the outcome of the statistical tests. Good statistical practice, however, demands that the control level is fixed a priori, and the supplemental figures should thus be considered only as a diagnostic that helps understand the complete lack of statistical significance in Figs. 4b and 4c.

c. Impact of using multiple lead times

One of the strengths of the CLSST model lies in its objective combination of forecasts made from multiple

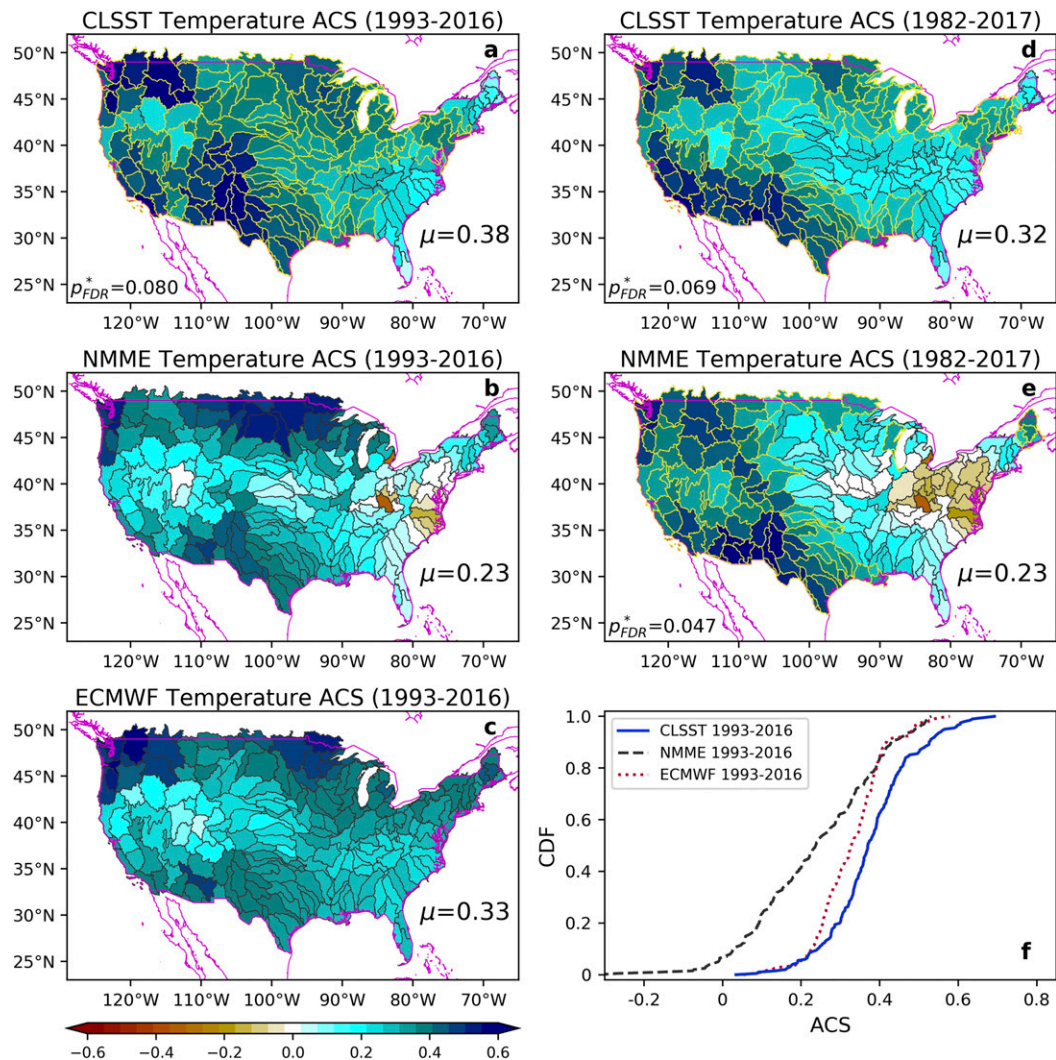


FIG. 4. As in Fig. 3, but for temperature forecasts.

leads (i.e., including SST states up to 18 months prior as predictors). For precipitation, there is only one predictor time series, though predictions are made using each of the 18 months that precede the time of reforecast. By doing this, the methodology exploits the fact that correlative relationships, as a function of lead time, can vary from one region to another. A prime example of this phenomenon can be found in the Sacramento HUC (Fig. 5). SST anomalies, from months approximately 1 year prior to the forecast initial time, are better predictors of winter precipitation in Northern California than those of the most recent SST anomalies. In contrast, the best predictor for southern Florida is the SST anomaly of the most recent month. If one were to use CLSST with only the most recent month's Niño-3.4 anomaly time series alone as a predictor of CONUS precipitation, the AC value for the Sacramento HUC in

the calibration period is approximately zero. In contrast, using CLSST with the Niño-3.4 anomalies from 12 months prior to forecast initialization yields an AC value of 0.34 for Sacramento in the same calibration period. When combining forecasts from the Niño-3.4 anomalies of all the preceding 18 months in the CLSST model, we achieve an AC value of 0.39 in the calibration period and 0.40 in the validation period. The effect of including more lead-time predictors is further illustrated CONUS-wide in the bottom panels of Fig. 5. The CLSST model, when only the most recent available month of SSTs is used, has an average AC (averaged across the 202 HUCs for the validation period) of 0.25. The CLSST model that uses all of the preceding 18 months of available SSTs (i.e., the implementation of the model we use in this study), improves the average AC skill to 0.30. Though the average AC improves when incorporating additional leads, with

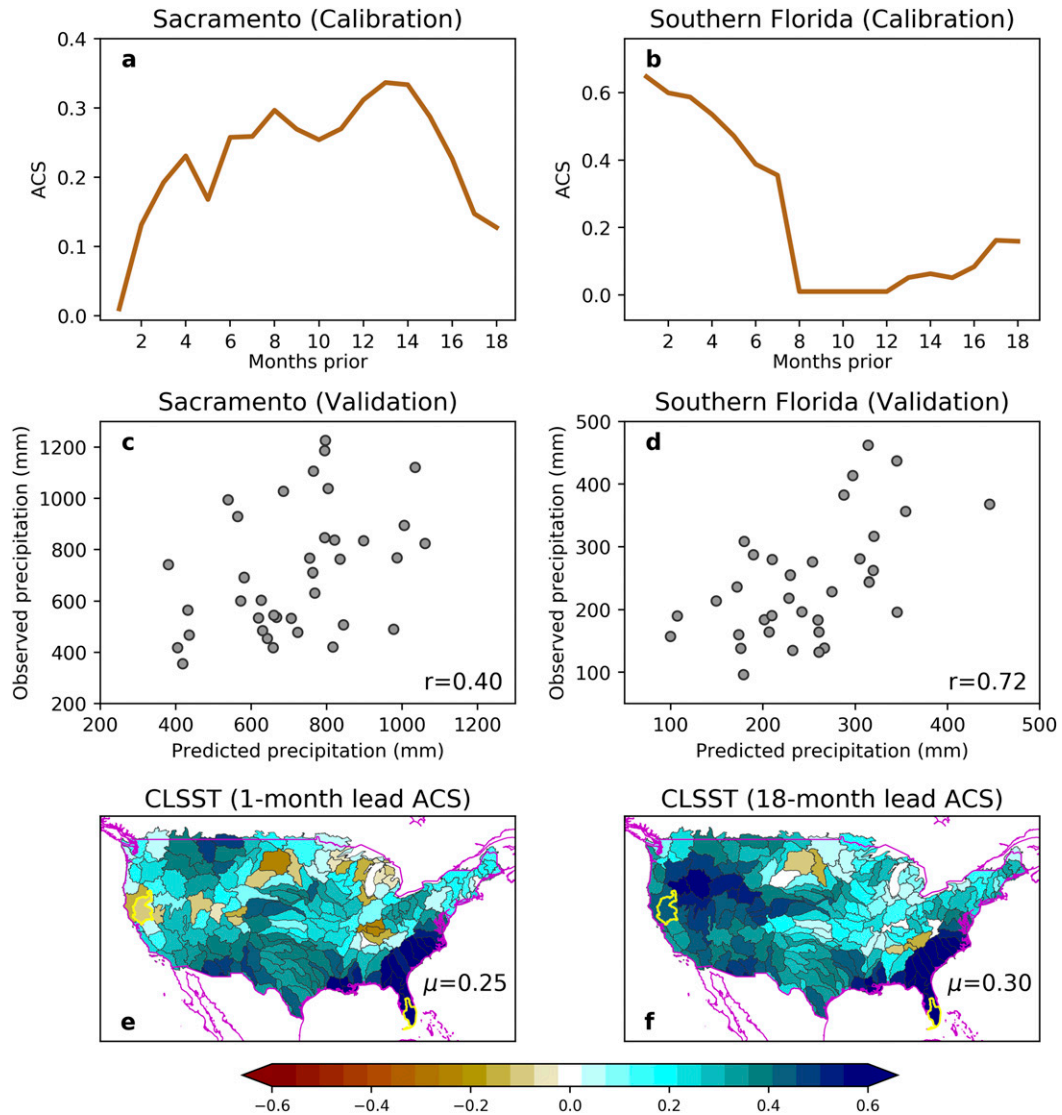


FIG. 5. The impact of the CLSST model on precipitation using multiple preceding months of SSTs as predictors. (a),(b) ACS values, for two example hydrologic units, in the calibration period when using the SST predictor time series from solely the individual months preceding (x axis) the October forecast time. Forecasts of precipitation corresponding to 1 month prior is performed by using only the preceding September SST time series as a predictor, while 2 months prior is using only the preceding August SST time series, and so on. (c),(d) Scatterplots between the 18-month combined predicted values vs observed precipitation in the validation time period (1981–2017). (e),(f) The ACS skill of CLSST as a function of using 1 month or the combined 18 months of preceding SSTs, where the two example hydrologic units shown in (a)–(d) are highlighted in yellow.

particularly notable increases in the CIW, the localized skill is not always improved. Central Montana is a region that sees decreasing skill as additional lags are included. The results for any individual HUC may be influenced by statistical sampling due to meteorological variability. One does not expect the observed spatial footprint of ENSO teleconnections to be temporally invariant even on multidecadal time scales (Deser et al. 2017). However, work here reveals a spatially coherent region of statistically

significant associations that are nearly as strong at 12 months lead as they are at 1 month lead. If one discounts the information at 12 months lead, then one must also consider the one month lead statistically suspect.

d. ENSO transitions and precipitation

The CLSST model points to a robust response of the CIW region to tropical Pacific SSTs approximately

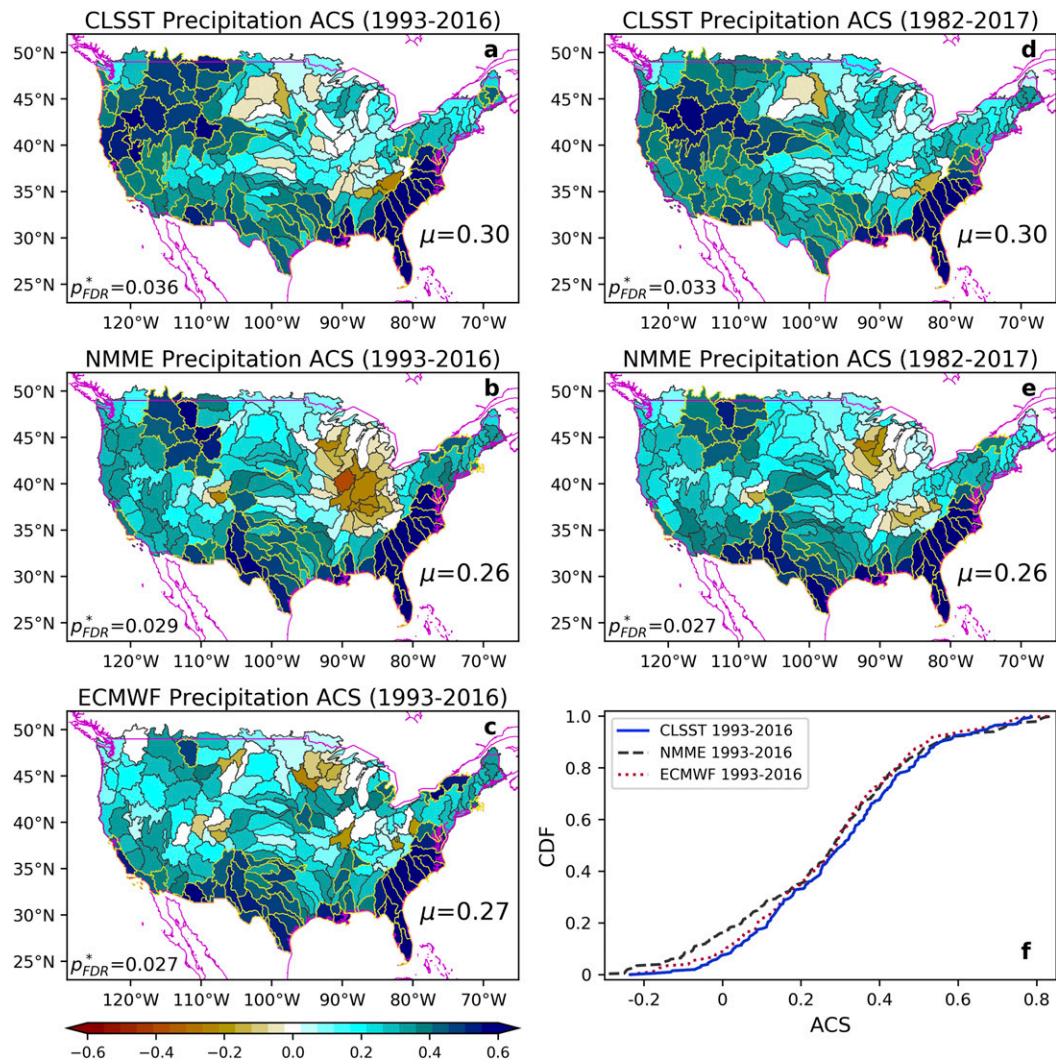


FIG. A1. As in Fig. 3, but for detrended precipitation.

12 months prior to the forecast date. This can be interpreted as an out-of-phase relationship between this region and ENSO. If the current fall and/or winter experiences a stronger El Niño/La Niña, there is a greater likelihood that the CIW will see increased/decreased precipitation in the following winter. However, to a large extent, there is a nonexistent relationship when using only the most recent months tropical Pacific SSTs and cold season precipitation in the CIW.

5. Discussion

We have developed and applied the relatively simple statistical CLSST model to make retrospective forecasts of cold season (NDJFM) precipitation and temperature across the contiguous United States. The model utilizes one tropical Pacific SST time series (Niño-3.4) as a

predictor for precipitation and two SST time series (tropical Pacific and a longitudinal tropical average) as predictors for temperature. The CLSST model is more skillful in anomaly correlation, on average, than either the NMME or the ECMWF-SEAS5 ensemble means. Of particular note, we find substantial improvement in skill for precipitation using CLSST across the central Intermountain West (CIW). The CLSST model exploits predictive information from prior tropical SSTs. The strength of this lagged relationship in the CIW region is maximized at approximately 1 year (Fig. 2) (i.e., SSTs from a year prior provide the best information for anomaly prediction in the upcoming cold season). Furthermore, we applied a rigorous statistical test (i.e., one that accounts for multiplicity and establishes field significance) and found this lagged relationship in the CIW region to be statistically significant. We also find

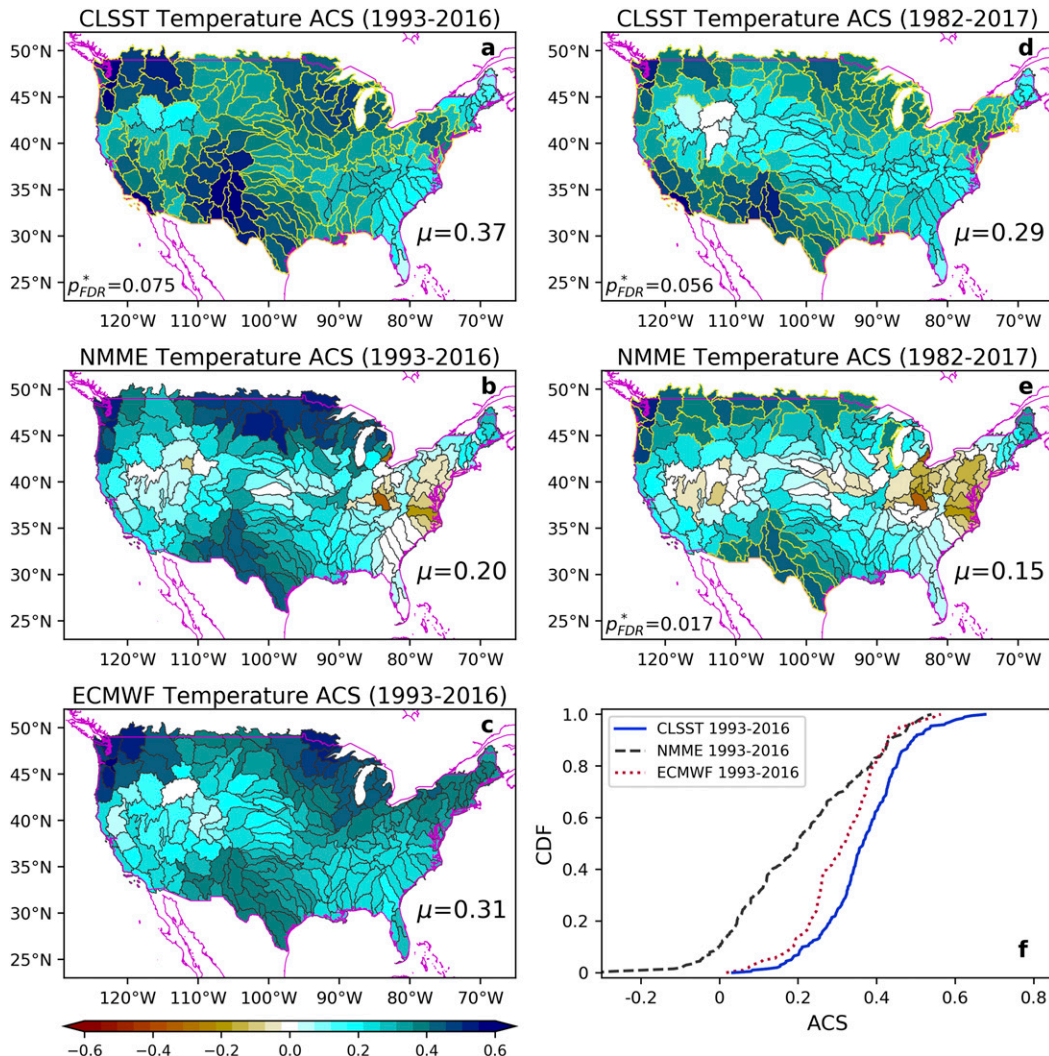


FIG. A2. As in Fig. 4, but for detrended temperature.

that the statistical significance of this correlative relationship for precipitation is not diminished when accounting for trends (in both observed and predicted values) and serial correlation.

In numerical weather and climate prediction, we commonly regard the current estimate of the initial condition as our best estimate of the information content in the system using prior and current observations. The information content of the prior observations is carried through to the current time through a cycle of short-term predictions and updates [see section 6.2.1 in Hamill (2006)]. The lagged predictive effect of SSTs may thus seem at first glance paradoxical. Isn't the current SST field providing sufficient information content for a seasonal statistical prediction? Perhaps this is not so paradoxical. One resolution to this paradox is that there may be predictive information

hidden in the deeper ocean state (Alexander et al. 1999) that the coupled numerical forecast models are not assimilating correctly, or are not taking advantage of due to the rapid development of model biases and forecast errors. Information from the lagged SSTs could be stored in the ocean below the surface but cannot be exploited in a regression model based only on the most recent SSTs. It suggests that either a statistical model that leverages current ocean-state anomalies below the surface would provide improved predictions, as would fully coupled predictions that can overcome many of their systematic errors. Alternatively to the SSTs influence on the deep-ocean state, Niño-3.4 SSTs from 1 year ago could also be a proxy for other more recent anomalies of the climate system. Exploring these possibilities is a subject for our future research.

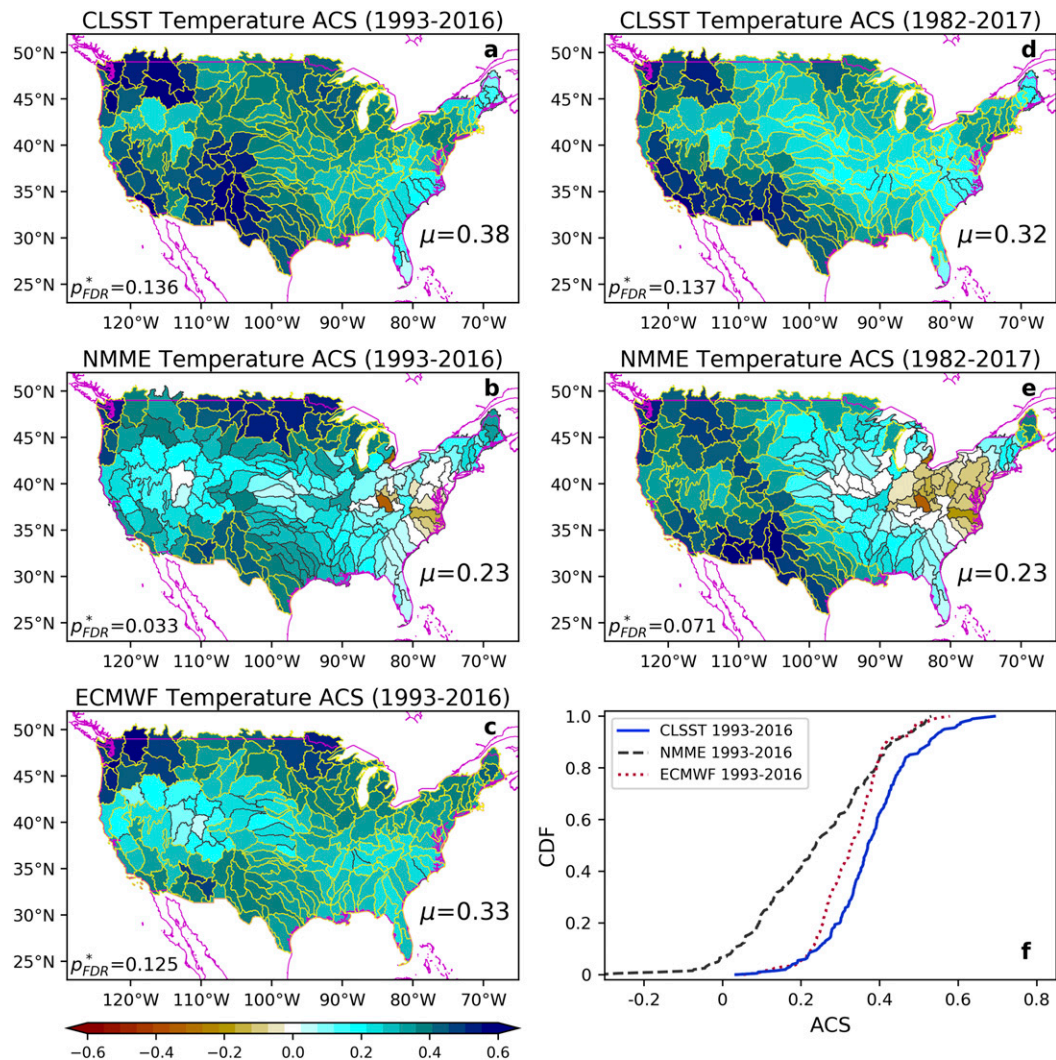


FIG. B1. As in Fig. 4, but showing statistical significance with $\alpha_{FDR} = 0.15$.

Acknowledgments. This study was funded by the California Department of Water Resources through federal Grant 4BM9NCA-P00. The authors do not have any conflicts of interest.

APPENDIX A

The Role of Trends

Within the context of a changing climate, there has been widespread temperature trends across the United States in the period 1982–2017 (which encompasses both of our validation periods). The CLSST is implicitly capturing some amount of observed trending on the land surface due to the trend in the predictor SSTs in the validation period with respect to the calibration period. How much of the skill that we

observe in CLSST, NMME and ECMWF-SEAS5 can be attributed to the ability of the models to forecast interannual variability versus getting the trend correct? To this end, all of the model forecasts and observations of both precipitation and temperature were detrended and ACS were computed again using the detrended data. The detrended ACS skill scores are shown in Figs. A1 and A2. For precipitation, the detrended skills of all three models are nearly identical to the nondetrended skills (Fig. A1). For temperature in the longer verification period, there is on average a 13% reduction in skill for CLSST and a 35% reduction in skill for NMME (Fig. A2). When considering the shorter verification period of 1993–2016, CLSST, NMME, and ECMWF-SEAS5 experience average reductions in skill of 3%, 13%, and 5%, respectively.

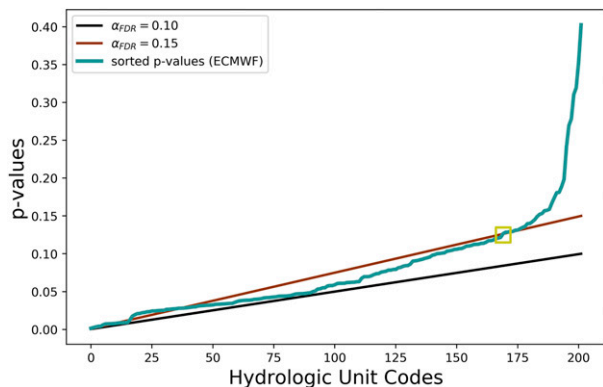


FIG. B2. The blue line shows the sorted p values that correspond to the ACS values of ECMWF from Figs. 4c and B1. The false discovery rates of $\alpha_{\text{FDR}} = 0.10$ and $\alpha_{\text{FDR}} = 0.15$ are shown in black and brown, respectively. The yellow box highlights where the sorted p values cross over the line with $\alpha_{\text{FDR}} = 0.15$ and yields an effective $p_{\text{FDR}}^* = 0.125$.

APPENDIX B

Multiplicity and Statistical Significance

Prior to performing statistical significance tests throughout our paper, we decided upon controlling the false discovery rate at the level $\alpha_{\text{FDR}} = 0.10$, which is a standard choice also used by Wilks (Wilks 2016). For the most part, this led to flagging statistically significant regions that generally appear intuitive. However, choosing a different control level α_{FDR} has the potential to dramatically impact the interpretation. Therefore, to assist readers in making sense of Fig. 4, we reproduce the same plot (Fig. B1) but using $\alpha_{\text{FDR}} = 0.15$. Taking a closer look at the differences in the two figures with the ECMWF model results, Fig. 4c had flagged zero HUCs as statistically significant with an $\alpha_{\text{FDR}} = 0.10$, while 169 HUCs are flagged as statistically significant with an $\alpha_{\text{FDR}} = 0.15$. This can be interpreted as follows: when controlling the FDR at the level $\alpha_{\text{FDR}} = 0.10$ we are not able to define any subset of HUCs where the statistically expected fraction of HUCs falsely flagged as significant is at most 10% (while we expect to correctly reject the null hypothesis for at least 90% of the flagged HUCs, where the null hypothesis is the hypothesis that there is zero ACS). In contrast, when using $\alpha_{\text{FDR}} = 0.15$, we expect that we incorrectly reject the null hypothesis for no more than 15% of the 169 flagged HUCs. This is further illustrated by the implementation of the statistical test used to determine statistical significance for the ECMWF reforecast ACS. In Fig. B2, the blue line shows the sorted p values that correspond to the ACS values of ECMWF from Fig. 4c. The false discovery rates of $\alpha_{\text{FDR}} = 0.10$ and $\alpha_{\text{FDR}} = 0.15$ are also shown. If, and

when, the sorted p values drop below the line controlling the false discovery rate, this is then the effective p_{FDR}^* (in the case of multiple crossings, the largest value is used). With $\alpha_{\text{FDR}} = 0.10$, the sorted p values never drop below the line, while with $\alpha_{\text{FDR}} = 0.15$ the values cross over the line with an effective $p_{\text{FDR}}^* = 0.125$.

REFERENCES

- Alexander, M. A., C. Deser, and M. S. Timlin, 1999: The emergence of SST anomalies in the North Pacific Ocean. *J. Climate*, **12**, 2419–2433, [https://doi.org/10.1175/1520-0442\(1999\)012<2419:TROSAI>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<2419:TROSAI>2.0.CO;2).
- Allen, R. J., and R. Luptowitz, 2017: El Niño-like teleconnection increases California precipitation in response to warming. *Nat. Commun.*, **8**, 16055, <https://doi.org/10.1038/ncomms16055>.
- Barnston, A. G., 1994: Linear statistical short-term climate predictive skill in the Northern Hemisphere. *J. Climate*, **7**, 1513–1564, [https://doi.org/10.1175/1520-0442\(1994\)007<1513:LSSTCP>2.0.CO;2](https://doi.org/10.1175/1520-0442(1994)007<1513:LSSTCP>2.0.CO;2).
- Benjamini, Y., and Y. Hochberg, 1995: Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Stat. Soc.*, **57B**, 289–300, <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- Capotondi, A., A. Wittenberg, and S. Masina, 2006: Spatial and temporal structure of tropical Pacific interannual variability in 20th century coupled simulations. *Ocean Modell.*, **15**, 274–298, <https://doi.org/10.1016/j.ocemod.2006.02.004>.
- Cayan, D. R., 1996: Interannual climate variability and snowpack in the western United States. *J. Climate*, **9**, 928–948, [https://doi.org/10.1175/1520-0442\(1996\)009<0928:ICVASI>2.0.CO;2](https://doi.org/10.1175/1520-0442(1996)009<0928:ICVASI>2.0.CO;2).
- , K. T. Redmond, and L. G. Riddle, 1999: ENSO and hydrologic extremes in the western United States. *J. Climate*, **12**, 2881–2893, [https://doi.org/10.1175/1520-0442\(1999\)012<2881:EAHEIT>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<2881:EAHEIT>2.0.CO;2).
- Cheng, L., M. Hoerling, A. AghaKouchak, B. Livneh, Q. Xiao-Wei, and J. Eischeid, 2016: How has human-induced climate change affected California drought risk? *J. Climate*, **29**, 111–120, <https://doi.org/10.1175/JCLI-D-15-0260.1>.
- Cook, B. I., A. Park Williams, J. S. Mankin, R. Seager, J. E. Smerdon, and D. Singh, 2018: Revisiting the leading drivers of Pacific coastal drought variability in the contiguous United States. *J. Climate*, **31**, 25–43, <https://doi.org/10.1175/JCLI-D-17-0172.1>.
- Deser, C., I. R. Simpson, K. A. McKinnon, and A. S. Phillips, 2017: The Northern Hemisphere extratropical atmospheric circulation response to ENSO: How well do we know it and how do we evaluate models accordingly? *J. Climate*, **30**, 5059–5082, <https://doi.org/10.1175/JCLI-D-16-0844.1>.
- Dettinger, M., B. Udall, and A. Georgakakos, 2015: Western water and climate change. *Ecol. Appl.*, **25**, 2069–2093, <https://doi.org/10.1890/15-0938.1>.
- Dieter, C., M. Maupin, R. Caldwell, M. Harris, T. Ivahnenko, J. Lovelace, N. Barber, and K. Linsey, 2018: Estimated use of water in the United States in 2015. U.S. Geological Survey Circular 1441, 65 pp., <https://doi.org/10.3133/cir1441>.
- Diffenbaugh, N. S., D. L. Swain, and D. Touma, 2015: Anthropogenic warming has increased drought risk in California. *Proc. Natl. Acad. Sci. USA*, **112**, 3931–3936, <https://doi.org/10.1073/pnas.1422385112>.
- Guan, B., D. E. Waliser, N. P. Molotch, E. J. Fetzer, and P. J. Neiman, 2012: Does the Madden–Julian oscillation influence

- wintertime atmospheric rivers and snowpack in the Sierra Nevada? *Mon. Wea. Rev.*, **140**, 325–342, <https://doi.org/10.1175/MWR-D-11-00087.1>.
- Guo, Y., M. Ting, Z. Wen, and D. Lee, 2017: Distinct patterns of tropical Pacific SST anomaly and their impacts on North American climate. *J. Climate*, **30**, 5221–5241, <https://doi.org/10.1175/JCLI-D-16-0488.1>.
- Hamill, T. M., 2006: Ensemble-based atmospheric data assimilation. *Predictability of Weather and Climate*, T. Palmer and R. Hagedorn, Eds., Cambridge University Press, 124–156.
- Huang, B., and Coauthors, 2017: NOAA Extended Reconstructed Sea Surface Temperature (ERSST), version 5. NOAA/National Centers for Environmental Information, accessed 15 January 2019, <https://doi.org/10.7289/V5T72FNM>.
- Johnson, S. J., and Coauthors, 2019a: Seas5 data set. Copernicus Climate Data Store, accessed 21 January 2019, <https://cds.climate.copernicus.eu>.
- , and Coauthors, 2019b: Seas5: The new ECMWF seasonal forecast system. *Geosci. Model Dev.*, **12**, 1087–1117, <https://doi.org/10.5194/gmd-12-1087-2019>.
- Jong, B., M. Ting, and R. Seager, 2016: El Niño's impact on California precipitation: Seasonality, regionality, and El Niño intensity. *Environ. Res. Lett.*, **11**, 054021, <https://doi.org/10.1088/1748-9326/11/5/054021>.
- Kirtman, B. P., and Coauthors, 2014a: Hindcast data set of the North American multimodel ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. NOAA/National Centers for Environmental Prediction, accessed 21 January 2019, <https://ftp.cpc.ncep.noaa.gov/International/nmme>.
- , and Coauthors, 2014b: The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; Phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601, <https://doi.org/10.1175/BAMS-D-12-00050.1>.
- , and Coauthors, 2014c: Real-time forecast data set of the North American multimodel ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. NOAA/National Centers for Environmental Prediction, accessed 21 January 2019, ftp://ftp.cpc.ncep.noaa.gov/NMME/realtime_anom/ENSMEAN.
- Kumar, A., and M. Chen, 2017: What is the variability in US west coast winter precipitation during strong El Niño events? *Climate Dyn.*, **49**, 2789–2802, <https://doi.org/10.1007/s00382-016-3485-9>.
- Lee, S., H. Lopez, E. Chung, P. DiNezio, S. Yeh, and A. T. Wittenberg, 2017: On the fragile relationship between El Niño and California rainfall. *Geophys. Res. Lett.*, **45**, 907–915, <https://doi.org/10.1002/2017GL076197>.
- Mao, Y., B. Nijssen, and D. P. Lettenmaier, 2015: Is climate change implicated in the 2013–2014 California drought? A hydrologic perspective. *Geophys. Res. Lett.*, **42**, 2805–2813, <https://doi.org/10.1002/2015GL063456>.
- Mazdiyasi, O., and A. AghaKouchak, 2015: Substantial increase in concurrent droughts and heatwaves in the United States. *Proc. Natl. Acad. Sci. USA*, **112**, 11 484–11 489, <https://doi.org/10.1073/pnas.1422945112>.
- McCabe, G. J., and M. D. Dettinger, 2002: Primary modes and predictability of year-to-year snowpack variations in the western United States from teleconnections with Pacific Ocean climate. *J. Hydrometeor.*, **3**, 13–25, [https://doi.org/10.1175/1525-7541\(2002\)003<0013:PMAPPOY>2.0.CO;2](https://doi.org/10.1175/1525-7541(2002)003<0013:PMAPPOY>2.0.CO;2).
- NAS, 2016: *Next Generation Earth System Prediction: Strategies for Subseasonal to Seasonal Forecasts*. The National Academies Press, 350 pp., <https://doi.org/10.17226/21873>.
- Newman, M., and P. D. Sardeshmukh, 2017: Are we near the predictability limit of tropical indo-Pacific sea surface temperatures? *Geophys. Res. Lett.*, **44**, 8520–8529, <https://doi.org/10.1002/2017GL074088>.
- Pan, B., K. Hsu, A. AghaKouchak, S. Sorooshian, and W. Higgins, 2019: Precipitation prediction skill for the West Coast United States: From short to extended range. *J. Climate*, **32**, 161–182, <https://doi.org/10.1175/JCLI-D-18-0355.1>.
- Philip, W. M., S. Li, D. P. Lettenmaier, M. Xiao, and R. Engel, 2018: Dramatic declines in snowpack in the western US. *npj Climate Atmos. Sci.*, **1**, 2, <https://doi.org/10.1038/s41612-018-0012-1>.
- Prein, A. F., G. J. Holland, R. M. Rasmussen, M. P. Clark, and M. R. Tye, 2016: Running dry: The US southwest's drift into a drier climate state. *Geophys. Res. Lett.*, **43**, 1272–1279, <https://doi.org/10.1002/2015GL066727>.
- PRISM Climate Group, 2019: Prism gridded climate data. Oregon State University, accessed 15 January 2019, <http://prism.oregonstate.edu>.
- Redmond, K. T., and R. W. Koch, 1991: Surface climate and streamflow variability in the western United States and their relationship to large scale circulation indices. *Water Resour. Res.*, **27**, 2381–2399, <https://doi.org/10.1029/91WR00690>.
- Ropelewski, C. F., and M. S. Halpert, 1987: Global and regional scale precipitation patterns associated with El Niño/Southern Oscillation. *Mon. Wea. Rev.*, **115**, 1606–1626, [https://doi.org/10.1175/1520-0493\(1987\)115<1606:GARSPP>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<1606:GARSPP>2.0.CO;2).
- Seaber, P. R., F. P. Kapinos, and G. L. Knapp, 1987: Hydrologic unit maps. USGS Water-Supply Paper 2294, 66 pp., http://pubs.usgs.gov/wsp/wsp2294/pdf/wsp_2294.pdf.
- Sullivan, A., D. D. White, and M. Hanemann, 2019: Designing collaborative governance: Insights from the drought contingency planning process for the lower Colorado River basin. *Environ. Sci. Policy*, **91**, 39–49, <https://doi.org/10.1016/j.envsci.2018.10.011>.
- Udall, B., and J. Overpeck, 2018: The twenty-first century Colorado River hot drought and implications for the future. *Water Resour. Res.*, **53**, 2404–2418, <https://doi.org/10.1002/2016WR019638>.
- Vitart, F., 2014: Evolution of ECMWF sub-seasonal forecast skill scores. *Quart. J. Roy. Meteor. Soc.*, **140**, 1889–1899, <https://doi.org/10.1002/qj.2256>.
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. International Geophysics Series, Vol. 100, Academic Press, 648 pp.
- , 2016: “The stippling shows statistically significant grid points”: How research results are routinely overstated and overinterpreted, and what to do about it. *Bull. Amer. Meteor. Soc.*, **97**, 2263–2273, <https://doi.org/10.1175/BAMS-D-15-00267.1>.
- Xiao, M., B. Udall, and D. P. Lettenmaier, 2018: On the causes of declining Colorado River streamflows. *Water Resour. Res.*, **54**, 6739–6756, <https://doi.org/10.1029/2018WR023153>.
- Zhu, J., and Coauthors, 2015: ENSO prediction in project Minerva: Sensitivity to atmospheric horizontal resolution and ensemble size. *J. Climate*, **28**, 2080–2095, <https://doi.org/10.1175/JCLI-D-14-00302.1>.