Original

The Kalman Filter as Post-Processor for Analog Data–Model Assimilation in Paleoclimate Reconstruction

Eugene Wahl,[a] Eduardo Zorita,[b] and Andrew Hoell[c]

[a] *NOAA National Centers for Environmental Information, Boulder, Colorado*
[b] *Helmholtz Zentrum Hereon, Geesthacht, Germany*
[c] *NOAA Physical Sciences Laboratory, Boulder, Colorado*

ABSTRACT: We present an offline paleo-data assimilation methodology that formally combines the analog assimilation method (AA) and the Kalman filter (KF), utilizing the KF as a postprocessor of the AA output. This methodology can be applied to reconstruct climate fields that are spatially separated from proxy-based reconstructions by using the spatial covariability generated by a climate model. Our method is applied to a set of spatially resolved and spatially consistent climate reconstructions of several variables reflecting different seasons, incorporating the application of methodological variants that have undergone rigorous testing in terms of both improving statistical methodology and physical interpretation. This contrasts with applications primarily based on transfer relationships of annual means of local, single variable or bivariate, climate model priors into paleo proxy states. The gains from adding the KF postprocessor are modest in our test case of reconstructing sea level pressure (SLP) geopotential height fields in the northeast Pacific, utilizing paleoclimatic temperature and moisture reconstructions in western North America. Notably, SLP reconstruction skill is enhanced in the oceanic region south of Alaska that is strongly associated with wet winters in western North America. The results suggest that the AA method is approaching optimality in this test case, driven by the quality of the paleoreconstruction information used to drive the AA process, along with the realism of the climate model employed, to which the KF postprocessing step is added. The derived reconstructions are then used for evaluation of the relationship between winter SLP and precipitation in California over the past ~450 years.

KEYWORDS: North Pacific Ocean; Pressure; Paleoclimate; Kalman filters; Data assimilation

1. Introduction

Quantitative reconstructions of past climates have traditionally been based on statistical calibrations (e.g., Neukom et al. 2019; Wahl and Smerdon 2012; Jones et al. 2009; National Research Council 2006; Cook et al. 1994) and other mathematical methods for relating current and past conditions: for example, selection of analogs in multivariate spaces (Schenk and Zorita 2012) and neural networks (Guiot et al. 2005), or on experimental or in situ relationships between biological or physical characteristics of proxy information and climate variations (cf. Anderson et al. 2013). Proxy climate records, however, are by nature relatively sparse, irregularly distributed over space and time, and sensitive only to some climatic variations. Paleoclimate model simulations forced by known or estimated past boundary conditions provide an important tool to enhance knowledge of climate system history (PAGES 2k Consortium 2019; PAGES Hydro2k Consortium 2017). Mathematical climate simulations provide complete three-dimensional climate trajectories; however, they are subject to modeling deficiencies per se, limitations of resolution due to computational intensity, and uncertainty in the external forcing factors—especially in the paleoclimate context. In parallel with assimilation methods that join model and observational information for forecasting and historical analysis of weather and climate, climatologists (both "neo" and paleo) have recently applied formal methods to join paleo-observations with paleo-model simulations (e.g., Steiger et al. 2018; Hakim et al. 2016; Schenk and Zorita 2012; Bhend et al. 2012; Widmann et al. 2010). The goals of such assimilation in the paleoclimate context are to exploit the different strengths of proxy-based and model-derived information to enhance 1) reconstruction skill, 2) the breadth of climate variables that can be reconstructed, 3) the spatial coverage of reconstructions, and 4) the estimation of reconstruction uncertainty.

Paleo-assimilation methodologies have generally been of two forms: 1) analog assimilation (AA), which seeks to identify analog model states to those exhibited by existing climate reconstruction data (Franke et al. 2011; Schenk and Zorita 2012; Gómez-Navarro et al. 2017; Neukom et al. 2019), and then to extract from the simulations physically consistent associated states of climate variables that are difficult or impossible to reconstruct from proxy information directly (Wahl et al. 2019); or 2) employment of Kalman filter (KF)-type schemes for adjusting "prior" states derived from climate model output by comparing those states to existing proxy or reconstruction information (e.g., Bhend et al. 2012; Hakim et al. 2016; Steiger et al. 2018; Franke et al. 2020). In this paper, we develop and evaluate a methodology for combining these two approaches in the paleo-reconstruction context, employing the KF as a postprocessor coupled to AA output. This approach was reported by Pfister et al. (2020), in application to

---

weather observations in Switzerland, and our application is the first to accomplish it for spatially explicit paleoclimate reconstruction.

An online variant of the AA has been used by Goosse et al. (2010). In this variant, the analogs are not just selected from an existing pool of simulated states, but rather from an ensemble of simulations. This ensemble is started from the conditions most similar to an initial proxy configuration and run for several decades. At this point, the ensemble is compared to an updated proxy configuration and the process is started from this point in time. This method is, however, only feasible with a simplified climate model, as otherwise the computational requirements would be overwhelming. To our knowledge, no AA+KF setup has been developed for this AA variant.

A fundamental motivation of AA in paleoclimatology is to exploit the large amounts of existing expert knowledge, rigorous testing, and spatially explicit character of the paleoreconstruction data assimilated. These characteristics distinguish the AA method from direct KF-based paleo-assimilation methods that rely on relatively simple (i.e., less expertise required) transfer models of climate state priors into raw paleo proxy data, as has been the case in paleo applications of KF-related methodologies that parallel data assimilation in weather forecasting (e.g., Steiger et al. 2018; Hakim et al. 2016; Bhend et al. 2012; cf. Widmann et al. 2010). Critically important in paleo AA is a priori testing of the climate model(s) employed, to identify simulation platforms that realistically represent both the reconstructed variables assimilated, such as temperature, precipitation, and soil moisture, and the desired output variables, such as large-scale temperature, sea level pressure (SLP) or 200-hPa geopotential height (GPH), and wind fields (Wahl et al. 2019; Diaz et al. 2016). Further testing is also done for parameters related to the optimization of analog selection, so that the output skill can be maximized in terms of independent validation against observations or reanalysis output (Wahl et al. 2019; section 2 herein). These tests represent the functional optimization equivalents of the formal rationales behind the KF methodology (Bhend et al. 2012).

However, AA by itself does not take advantage of the potential information available in the difference (or delta) between the AA results and the original proxy-based reconstructions assimilated, known as the "innovation" term in KF terminology (Hakim et al. 2016; Lahoz and Schneider 2014). The KF provides a natural method for doing this. In the context here, the KF acts as a secondary processing step in which the AA outputs for the desired reconstruction values serve as KF priors, and the filter acts as a weighed gain function (≥ 0, ≤ 1) to determine the extent to which the innovations provide useful information for nudging the original AA output. In this application, the transfer model to enable commensurability of the innovation components is a separate mathematical relationship between the original reconstruction variables assimilated in AA and the desired reconstruction variables, determined independently from observations or reanalysis output. This relationship can be relatively simple, such as linear regression (if linearity is reasonable), since it

is in effect a kind of scaling between the AA input and desired output states, all of which are already represented in terms of climatological variables per se.

In addition, the KF and associated ensemble KF (EnsKF; Steiger et al. 2018; Bhend et al. 2012; Whitaker and Hamill 2002) provide a natural and direct formalism for objectively characterizing uncertainty in the combined AA+KF posterior (or "analysis") output. This is a further key advantage of joining the two methods since it broadens the formal estimation of uncertainty directly available in AA from the variability of the selected analog states. In particular, adding the KF postprocessing step allows explicit incorporation of the uncertainty in the original assimilated paleoreconstruction data themselves, via the residual errors derived from the transfer relation described immediately above. These errors help determine the KF gain: reducing the gain in proportion to their size and in turn reducing the nudging of the AA output priors.

Here, we detail our implementation of the AA+KF methodology in the paleoclimate context, utilizing for this purpose an example of AA that has been developed and analyzed in detail in previous work (Wahl et al. 2019; Diaz et al. 2016). After describing the methods (sections 2a and 2b), we report testing and validation of AA+KF in relation to reconstruction of winter [December$_{t-1}$–February$_t$ (DJF)] circulation variables (SLP and 200-hPa GPH) in the northeastern Pacific and adjacent west coast of North America, based on assimilation of spatially explicit paleoreconstructions of temperature and moisture variables from western temperate North America (sections 2a and 3a). We then apply the derived SLP information to evaluation of the relationship between northeast Pacific winter SLP and California (CA) precipitation over the past 450 years (section 3b), paralleling recently reported paleo-evaluation of the relationship between 200-hPa (effectively, jet stream) winds and CA precipitation and fire regimes (Wahl et al. 2019).

Precipitation in CA displays a strongly "Mediterranean" seasonality, with wet winters and dry summers (e.g., Seager et al. 2019). Moisture delivery during the hydrological half-year is caused by extratropical synoptic systems that advect from the North Pacific, whose track and intensity modulate the distribution of precipitation along the western U.S. coastal regions (e.g., Chang et al. 2002). The factors that influence the extratropical cyclones have been found in several studies to be related to the position of the jet stream and the seasonal Aleutian low-pressure cell (Conil and Hall 2006; Haston and Michaelsen 1997; Castello and Shelton 2004; Pandey et al. 1999; Cayan and Roads 1984), and at times these cyclones also entrain subtropical moisture plumes to form "atmospheric river" events (e.g., Dettinger 2013). El Niño–Southern Oscillation additionally has a teleconnected impact on precipitation in this area (e.g., Hoell et al. 2016; Mo and Higgins 1998; Schonher and Nicholson 1989).

Our application focuses on an important aspect of the relationship between SLP and precipitation in CA: the asymmetry observed, notably in Southern CA, between high precipitation and low SLP in the wettest years versus the relationship between low precipitation and high SLP in the driest years. Low winter-average SLP associated with

the wettest precipitation years is generally not as extreme as the precipitation itself, whereas high winter SLP and low precipitation of the driest years show similar extreme excursions. Finally, we summarize and discuss the development and application of the AA+KF method in broader terms and indicate potential future directions for its application (section 4).

2. Methods

a. Essential aspects of the AA method

The AA-derived winter SLP and 200-hPa reconstructions for the northeast Pacific–North American west coast region were initially described and evaluated in Diaz et al. (2016) and Wahl et al. (2019). The essential aspects of this method specific to these reconstructions are summarized below. Additional information regarding the AA method for climate reconstructions can be found in Schenk and Zorita (2012) and Graham et al. (2007).

As employed here, AA combines proxy-based gridded climate reconstructions and the output of a climate simulation to produce a spatiotemporally complete three-dimensional reconstruction of the climate over the period spanned by the proxy records. Input data for the AA algorithm are 1) a set of annually resolved, proxy-based, reconstructed climate fields (e.g., spatially resolved temperature, soil moisture, precipitation) over a certain period and 2) the output of a long climate simulation from a climate model over approximately the same period covered by the proxy records (although this is not a strict condition) and with the same, or nearly identical, variables as the reconstructions. The simulation used here has been forced by the known or estimated external forcing factors over the past centuries, including solar irradiance, atmospheric aerosols and greenhouse gases, and land use. In theory, any climate simulation or gridded observations producing physically consistent fields could be used as well as a source of analogs. It is advisable, however, that the variability in the analog pool is wide enough to represent the variability of the target fields. In this respect, a long, externally forced, atmosphere–ocean coupled simulation may be more suitable. Wahl et al. (2019, see their supplemental information) provide detailed information regarding climate simulation characteristics, evaluation, and selection, which led to usage of the Max Planck Institute (MPI) ESM-P Earth system model (Giorgetta et al. 2013).

For one point in time in the past, denoted as t, the spatially resolved climate reconstructions for t are set as the target pattern. The AA algorithm identifies analogs to that target pattern from the entire climate simulation. The analogs are defined as the time steps in the simulation for which the simulated fields spatially match the reconstruction patterns as closely as possible. Once similar "analogs" are identified, $t_1, \ldots, t_n$, the desired atmospheric and ocean fields (e.g., SLP or 200-hPa fields) for the time steps $t_1, \ldots, t_n$ are extracted and averaged. This average is defined as the AA climate reconstruction for time t.

The analog reconstruction depends on several factors. One is the definition of similarity between the simulated climate fields and the target pattern. A typical definition of "distance" between these patterns is the multivariate Euclidean distance (e.g., the root-mean-square difference between the two fields). Here we adopt this is the metric and apply it to temporally standardized time series (see section 2). We note that the squared chord distance was also evaluated since that multivariate distance metric has been identified as highly useful in some other paleoenvironmental contexts, but it performed similarly to the Euclidian distance and the latter was utilized in these reconstructions (Wahl et al. 2019). The dimensionality of the space in which the search of analogs is conducted is also relevant. If this dimension is too high (i.e., for a spatially extended climate field containing too many independent degrees of freedom), the pool of potential analogs has to be very big to span a very large, high-dimensional volume of the search space. The dimensionality can be reduced by projecting the fields onto a few leading principal components (PCs) resulting from a principal components analysis (PCA). In this fashion, the analog search is conducted in a much smaller dimensional volume and the chances to find suitable analogs are much higher. If the leading PCs span most of the variability, the identified analogs will capture the main features of the target field. Here, we apply this PCA-based preprocessing, which also offers additional advantages for the computation of some of the required matrices in the KF postprocessing (section 2b). Numerically, eigen-space reduction additionally significantly reduces computation time, which is nontrivial for the AA method with sufficient size of the data grids and time steps involved (for both the original reconstruction variables and the potential simulation analogs), as implemented here. At the completion of the AA+KF process, the desired posterior variables are retransformed into their original states

An important hyperparameter in the AA is the number of analogs that are considered to match the target and, therefore, selected as possible realizations of the large-scale climate field. Pseudoproxy experiments (e.g., Gómez-Navarro et al. 2017) clearly indicate that the correlation between reconstruction and target increases when the number of accepted analogs increases from one to a few tens (depending on the specific problem at hand), but that the temporal and spatial variance of the reconstructed fields also diminishes. This result is logical. As the number of potential analogs grows, their averaging filters out noise. On the other hand, the amplitude of extreme anomalies is not well replicated as a result of averaging analogs, not all of which can accurately resemble a particular extreme target field (section 2b). To produce the AA reconstructions in Wahl et al. (2019), we set the number of analogs to $N = 15$. This is a heuristic choice made from the analysis of self-reconstruction tests. In those tests, the gridded temperature, precipitation, and soil moisture reconstructions are replaced by their climate model counterparts and the SLP is reconstructed searching for analogs in the climate model output, excluding the target year which would result in a perfect match. This reconstructed model SLP field is compared to the model SLP field and their root-mean-square error (RMSE) is calculated. The RMSE displays a relatively flat

minimum around $N = 15$, so that this choice is not very critical as long as it is remains within the boundaries 10–25. Here, we also tested the improvement in the AA+KF setting with $N = 30$. With this sample size, the estimation of the matrices involved in the KF step (the prior error covariance **B**, the observations error covariance **R**, and the gain matrix **K**; see section 2b) is arguably more robust, but the AA reconstruction itself already displays diminishing correlation skill in relation to an independent reanalysis target, making the application of the complete AA+KF scheme less meaningful.

The AA reconstruction is, in theory, global since we use the output of a global climate model. However, the information encapsulated in the reconstruction records is regionally limited (here, the western United States and small portions of northern Mexico and southwestern Canada) and may include just one or a few variables. The skill of the AA reconstruction is, therefore, generally limited to the region(s) and to the output variables that are physically associated with the region(s) where the input reconstruction records are located (Wahl et al. 2019). This skill can be assessed by comparing the AA reconstruction with observations in a period of overlap; we note that the AA method as applied in this context does not incorporate instrumental or reanalysis observations of its output variables so that this skill evaluation represents an entirely independent validation (section 3a).

In the studies by Diaz et al. (2016) and Wahl et al. (2019), whose output we employ here, the AA reconstructions of the winter atmospheric circulation in the northeast Pacific incorporated three predictor datasets: 1) tree ring–derived gridded reconstructions of late winter [February–March (FM)] 2-m air temperature in temperate western North America (Wahl et al. 2014); 2) tree ring–derived summer [June–August (JJA)] near-surface soil moisture over the southwestern United States [Cook et al. 2010a; the North American Drought Atlas (NADA)]; and 3) gridded "water year" [WY; from October of the previous year ($t - 1$) through September of the current year (t)] precipitation in CA, based on tree ring–derived streamflow reconstructions (Diaz and Wahl 2015; Wahl et al. 2017). These three variables are physically related to the northeast Pacific/western coastal North America atmospheric circulation (cf. St. George et al. 2010), even at 200 hPa, and potential analogs for them were searched in climate simulations as described in Wahl et al. (2019; see their supplemental information). The search of analogs is conducted simultaneously for all three variables. The gridded reconstructions are concatenated in an augmented field, after a normalization of the respective amplitudes of variation to align the variances of all three fields. As noted, the augmented field is subject to a PCA and only the leading modes resulting from this analysis are retained, generally on the order of five PCs. The climate model data are then projected onto these spatial modes, thus ensuring that the PCA axes are the same for the reconstructions and in the model output. The combination of the three seasonal predictors into an augmented field and their preprocessing by PCA also should result in a more normal distribution of their values, blurring the usually nonnormal distribution of pointwise precipitation at short time scales.

This point is relevant in the ensuing KF postprocessing (section 2b). In principle, the model analogs could also be searched in a long control run of an atmospheric model that is forced by boundary conditions that vary only with the seasonal cycle. However, an externally forced, fully coupled climate model run can arguably provide better coverage of the paleoclimate variability and thus provide better analogs. Since the coupled paleoclimate simulations are long (over 1000 years), there is no compelling reason to not use them.

All reconstruction and simulation variables were regridded to the 2° × 2° NADA grid. The skill of the reconstructions was assessed by comparing the AA reconstructed fields of SLP and 200-hPa GPH with those of the Twentieth Century Reanalysis version 2 (20CR; Compo et al. 2011; Slivinski et al. 2019), also regridded to the NADA grid scale. The comparison was done over the portion of the AA and reanalysis overlap during which the reanalysis has significantly lower internal ensemble variability, 1930–80 (Diaz et al. 2016, see their supplemental information). The skill is good in the northeast Pacific sector and into adjacent western North America, with highest correlations of ~0.65 for the interannual variations of both winter SLP and 200-hPa GPH; relatively high values also extend into the tropical Pacific at the 200-hPa level [Wahl et al. 2019, their Figs. S5b,d; companion values for the "coefficient of efficiency" metric (CE; Cook et al. 1994) are provided in their Figs. S6b,d].

The information contained in the three predictor variables can to some extent overlap. In Wahl et al. (2019) we additionally evaluated the relationships between the predictor variables, finding that the FM temperature and NADA reconstructions are functionally independent (low empirical correlations, even at the grid cell level). The FM winter temperature and CA precipitation reconstructions are also functionally independent in this way. There is high (≥ 0.7) correlation between some of the NADA grid cells and the CA precipitation grid cells. From the standpoint in that study, these relationships were found not to introduce problematic predictor redundancy that reduced skill in independent validation; rather, including those NADA grid cells in fact improved validation spatial skill. In the context of this paper, these relationships do indicate empirical overlap between the CA precipitation reconstruction and some of the NADA grid cells used in the AA reconstructions of circulation. This overlap is also strongly present in the southwestern United States in instrumental data (St. George et al. 2010), representing the fundamental hydroclimate feature that southwestern regional summer soil moisture is strongly correlated with antecedent winter precipitation (including CA as a core region), which itself is strongly related to northeast Pacific circulation conditions. [We note that a wider fall-to-spring (October$_{t-1}$–May$_t$) precipitation seasonality is strongly correlated with DJF precipitation in CA, at $r = 0.85$ (NCEI 2019).] This physical correlation between temperature and precipitation is reflected in the correlation between their respective gridded reconstructions, indicating it is not caused by spurious (nonclimatic) correlations between different dendrochronologies from the same region that are incorporated into the reconstructions.

These AA-derived reconstructions are available via the World Data Service for Paleoclimatology (NOAA-NCEI; https://www.ncdc.noaa.gov/paleo/study/20465 for SLP and https://www.ncdc.noaa.gov/paleo/study/26030 for 200-hPa u and v wind fields).

b. Essential aspects of the KF for use as AA postprocessor

The fundamental formulations of the KF methodology are provided in Eqs. (1) and (2) for the KF expected value (EV) and Eqs. (3)–(6) for the EnsKF (or Ens), following the exposition of Steiger et al. (2018):

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}[\mathbf{y} - \mathcal{H}(\mathbf{x}_b)], \tag{1}$$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}, \tag{2}$$

$$\overline{\mathbf{x}}_a = \overline{\mathbf{x}}_b + \mathbf{K}[\mathbf{y} - \mathbf{H}(\overline{\mathbf{x}}_b)], \tag{3}$$

$$\mathbf{x}'_a = \mathbf{x}'_b + \tilde{\mathbf{K}}\mathbf{H}(\mathbf{x}'_b), \tag{4}$$

$$\tilde{\mathbf{K}} = \mathbf{B}\mathbf{H}^T[(\sqrt{\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}})]^T(\sqrt{\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}} + \sqrt{\mathbf{R}})^{-1}, \tag{5}$$

$$\mathbf{x}_a = \overline{\mathbf{x}}_a + \mathbf{x}'_a. \tag{6}$$

In Eq. (1) $\mathbf{x}_b$ is the prior mean state vector, the spatially explicit output of the AA algorithm described in section 2a above. The delta $[\mathbf{y} - \mathcal{H}(\mathbf{x}_b)]$ is the "innovation" term described in section 1, incorporating the difference between the original reconstruction observations assimilated in the AA process, $\mathbf{y}$, and the AA output priors converted into the state of the original reconstruction observations by the operator $\mathcal{H}$. In our development, $\mathcal{H}$ (which in general can be nonlinear) is formed as the matrix of simple regression slope coefficients, $\mathbf{H}$, of the three original reconstruction variables on the spatial SLP and 200-hPa states as determined from the 20CR, all in their reduced PC form. These regressions and corresponding PCAs are estimated during the optimum period of overlap between the 20CR and the original reconstruction variables as noted in section 2a, 1930–80. The operator $\mathbf{K}$ is the Kalman gain matrix, whose formulation is outlined in Eq. (2). This matrix has the effect of being either a 1:1 multivariate gain operator for the information represented by the innovations, or some reduction of that gain between 1:1 and 0, with 0 representing no added information value from the innovations. This gain is governed by the "ratio" between the inherent uncertainty/variability of the AA output—in this case the covariance matrix of $\mathbf{x}_b$, $\mathbf{B}$, normalized into the terms of the reconstruction observation space by $\mathbf{H}$, and the covariance matrix of the observational errors associated with the $\mathbf{H}$ matrix regressions, $\mathbf{R}$. That is, the smaller the $\mathbf{R}$ error covariance values are relative to those of normalized $\mathbf{B}$, the more $\mathbf{K}$ approaches the 1:1 gain state; the larger the $\mathbf{R}$ values are relative to those of normalized $\mathbf{B}$, the more $\mathbf{K}$ approaches the 0 gain state. In this way, the relative precision of the AA priors vis-à-vis the accuracy and precision of the $\mathbf{H}$ regressions (combining both the linear transfer of the pressure variables to the

original reconstruction $\mathbf{y}$ values and the inherent errors in the reconstructions per se) governs the extent to which the innovation information is considered useful to adjust the first-order AA output; represented in Eq. (1) as $\mathbf{x}_a$, the "analysis" posterior in KF terminology. This is how the action of the KF postprocessing step occurs when coupled to the AA output as its priors, and represents the direct EV operator of the AA+KF output.

Equations (3)–(6) indicate how the ensemble of analogs selected by the AA algorithm is additionally exploited via the KF method to provide a natural formalism for describing the uncertainty in the combined AA+KF output. In the following, we assume that all individual members of the analog ensemble are a priori equally likely, thereby providing an unbiased sampling of the prior probability distribution. This assumption is not strictly correct since the Euclidean distance of each analog to the target is not the same; however, the similarity between the target and each Nth analog, for $N = 1$–15, is a generally a rather smooth, slowly decreasing, function of N within this range. This characteristic makes the assumption tractable in our application. As noted, $N = 15$ represents an empirical optimum when the AA method is tested for its ability to "self-reconstruct" the SLP and 200-hPa fields. A corollary result of this process would be that the most reasonable representation of the uncertainty in the AA reconstruction is provided by this number of selected analogs, discussed further in the following subsection on estimation of the $\mathbf{R}$ and $\mathbf{B}$ matrices.

Here, Eq. (3) is identical to the way Eq. (1) is used: $\mathbf{x}_b$ is simply the AA mean state output noted above and $\mathbf{K}$ is identical to its definition in Eq. (2), thus the $\mathbf{x}_a$ AA+KF posterior output in Eq. (1) is identical to the $\mathbf{x}_a$ posterior in Eq. (3). The terms $\mathbf{x}'_a$ and $\mathbf{x}'_b$ in Eqs. (4) and (6) represent deviations from their respective mean state vectors for all the selected analogs in the AA process and thus have the form of matrices rather than vectors as in Eqs. (1) and (3), and $\mathbf{K}$ as defined in Eq. (5) represents the equivalent of $\mathbf{K}$ when used with the deviation values in Eq. (4). Note that in Eq. (4) only the $\mathbf{x}'_b$ deviations are gain-weighted by $\mathbf{K}$; the innovations with respect to the originally assimilated reconstruction values, $\mathbf{y}$, do not appear since they are already included in the mean state formulation in Eqs. (3) and (6), giving rise to the form of Eq. (5) for $\mathbf{K}$.

ESTIMATION OF THE $\mathbf{R}$ AND $\mathbf{B}$ MATRICES

As noted, the observational error-covariance matrix $\mathbf{R}$ is the defined as the covariance of regression residuals of the linear operator $\mathbf{H}$, both defined in the PCA space. In the AA+KF implementation we have developed, an assumption of zero cross-covariances among observational data errors (cf. Steiger et al. 2018; Bhend et al. 2012) is not tenable since the methods used to construct the gridded reconstruction data assimilated in the AA process allow for nonzero cross-correlated errors. These errors, also defined in the PCA space, can be spatially long-range correlated, but the reduced effective dimensionality ensures that this does not lead to computational instability. Thus, $\mathbf{R}$ is generally a full, nondiagonal covariance matrix, but since it is estimated within a space of reduced dimensionality,

there is no need to apply additional methods (e.g., covariance localizations to suppress spurious correlations) that are otherwise more usual in KF settings (Bhend et al. 2012).

Following from the discussion above, the sample size $N = 15$ to estimate the **B** covariance matrix is recognized as small, but as outlined in section 2a it cannot be much enlarged without inclusion of less suitable analogs (note the effect of setting $N = 30$), thereby worsening the skill of the AA methodology per se, and making the application of the complete AA+KF scheme less meaningful overall. Thus, we conclude it is reasonable to estimate **B** using the evaluated best sample size of analogs, and note that the dimension reduction also increases the relative "ratio" between the sample size employed and the size of the space for which **B** needs to be estimated. From a strictly computational standpoint, by reducing the effective number of degrees of freedom of the state vector to a few leading PCs we also achieve a more stable estimation of **B** and **R** and of the inverse matrices involved in Eqs. (2) and (5), Otherwise, the respective **K** gain operators can become nearly singular, thereby making the inversions unstable or numerically impossible to compute. Note that the use of N analogs to estimate **B** is logically consistent with the use of these analogs to estimate the prior state $\mathbf{x}_a$. Hakim et al. (2016) use the whole simulation output to estimate $\mathbf{x}_a$ and also, consistently, use the whole simulation output to estimate **B**.

We note that these more stable estimations via dimension reduction additionally reduce the impact of the original reconstructions' errors within the formulation of the **B** matrix, and thus help to isolate the **B** matrix from the **R** matrix. This isolation/damping arises numerically because dimension reduction via PCA to a few leading components inherently reduces noise, that is, variation in lower-order dimensions that carries little of the data's overall information content and that more and more represents random variability through the decreasing orders. This outcome of the application of dimension reduction is a well-understood property that is one of the significant reasons for the use of PCA. The original reconstructions assimilated into the AA process have well-behaved (effectively random) residual errors [Wahl et al. 2014; Diaz and Wahl 2015; see the evaluation of the Cook et al. (2010a) reconstruction by the authors], and thus their influence on the selection of analogs in this context will be greatly damped by the dimension reduction, per se.

More generally, the observational errors (in this case representing explained/unexplained variability when the original reconstructions are projected into the space of the output variable, SLP or 200-hPa geopotential, determined from the **H** matrix regressions) could lead to systematic bias in the covariance structure **B** of the AA priors if the original reconstructions exhibit bias from their true EVs. The reconstructions assimilated are effectively unbiased in postcalibration evaluation, as noted. Thus, while for any given time step their deviation from the true reconstruction EVs could have an impact on analog prior selection, the overall effect of such impacts will be nonsystematic vis-à-vis the analog pool over the entire reconstruction period. In turn, the effect on the estimation of **B** from the selected analogs throughout the reconstruction period will also generally be overall unbiased in relation to the assimilated reconstructions, with the following potential exception. At a few time steps, the sets of selected analogs can be overall biased, namely for those time steps in which the true target pattern plus its noise realization render it an extreme pattern. In those cases, the set of selected analogs will tend to be less extreme, in turn leading to the error in the set of AA reconstructed states to not be independent of the errors in the observations. The number of these situations is, however, small. We conclude overall that these considerations justify the usage of the original reconstruction errors in the formulation of the **R** matrix as reasonably independent of the formulation of the **B** matrix.

3. Results

a. Testing and validation of the AA+KF method

Figures 1 and 2 show reconstruction skill, in terms of correlation, obtained using the AA method only and from the expected value (EV) of the joined AA+KF process (SLP in Fig. 1 and 200-hPa GPH in Fig. 2). The rationale of not also using the root-mean-square error as a measure of skill in this context lies in the nature of the AA. AA essentially resamples from a climate model output, so that the reconstruction bias is impacted by the bias of the climate model. This can eventually be corrected by realigning the simulation data toward relevant observations, but this does not add any further intrinsic methodological skill. The reconstruction variance is also dictated by the variance of the simulated data. This can also be simply adjusted to observations, but again without necessarily demonstrating good skill in the reconstruction of the variance. Thus, in the context of the AA, the most critical measure of methodological skill remains the temporal correlation between reconstructions and observations. In contrast to mean bias, a good correlation is indicative that meaningful analogs can indeed be selected, rendering it a relevant measure of reconstruction skill. We note that, since the paleo and model data incorporated in the reconstructions are entirely independent of the SLP observations and model platform used to implement the 20CR reanalysis, the comparison represents an entirely independent validation, logically parallel to an out-of-sample validation in the context of a statistical reconstruction. In both cases, there is modest but clear improvement with the addition of the KF postprocessing step. In the case of SLP (Fig. 1), there is a strengthening of validation success to correlation values 0.7–0.8 precisely in the core region of northeast Pacific anomalous low pressure known to be associated with wet winters in North America (Wahl et al. 2019; see their Fig. S1, right panel) and dry northeasterly winter conditions in Hawaii (Diaz et al. 2016; see their Fig. 1, top panel). This result is important beyond its meaning vis-à-vis skill, as it highlights the physical realism of the entire through-process of the foundational AA method. This physical realism is supported by the higher reconstruction skill of the AA method precisely in the region of the Pacific atmosphere that most importantly drives the temperature and precipitation anomalies in the western United States. This clearly
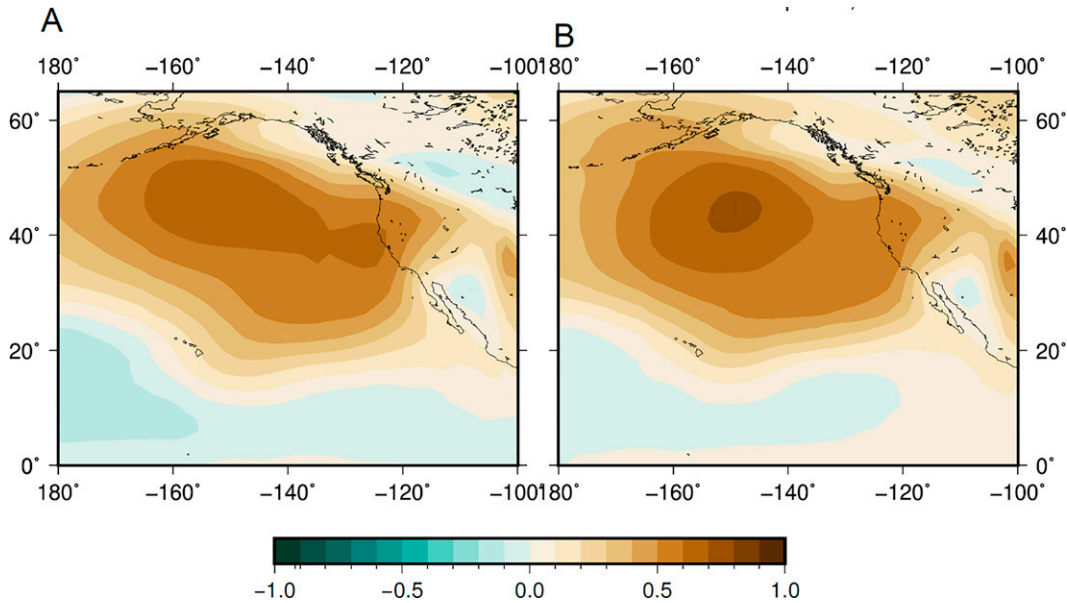
FIG. 1. Spatial correlation between reconstructions of boreal winter (DJF) SLP and corresponding 20CR data, for the (a) AA and (b) AA+KF reconstruction methods outlined in the text. The period of comparison is 1930–80, covering temporal overlap of reconstruction and reanalysis datasets during which the 20CR is considered most reliable based on its internal ensemble variability (Diaz et al. 2016; see the supplemental information therein).

indicates that the selection of simulated analogs is ultimately dictated by realistic physical mechanisms and that these are well represented in the climate simulation, most notably of the climate model used (MPI ESM-P). Second, the KF refinement is also most strongly realized in this core region, also indicating that the improvement brought about by this refinement is also able to identify these physical mechanisms.

There is some contraction of the spatial extent of skill values between 0.6 and 0.7 surrounding the maximum region, but at the same time there is extension of stronger skill values into southwestern Alaska and the nearby Bering Sea region. In the case of 200-hPa GPH (Fig. 2), there is an expansion of the region of higher skill levels in western temperate North America and the nearby Pacific, with the highest correlation values including areas nearer the coastlines of the U.S. states
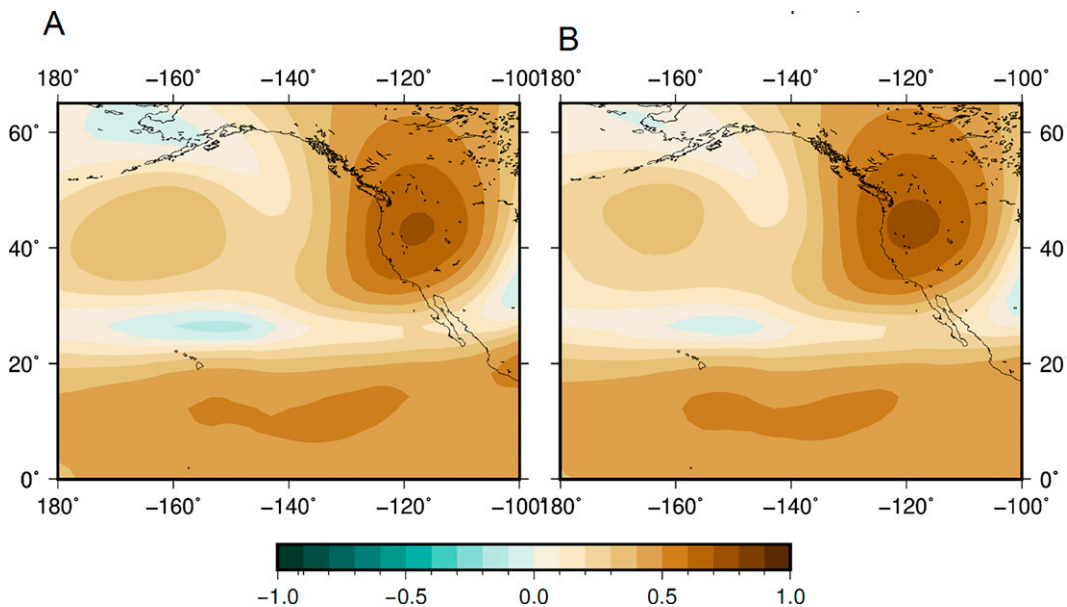

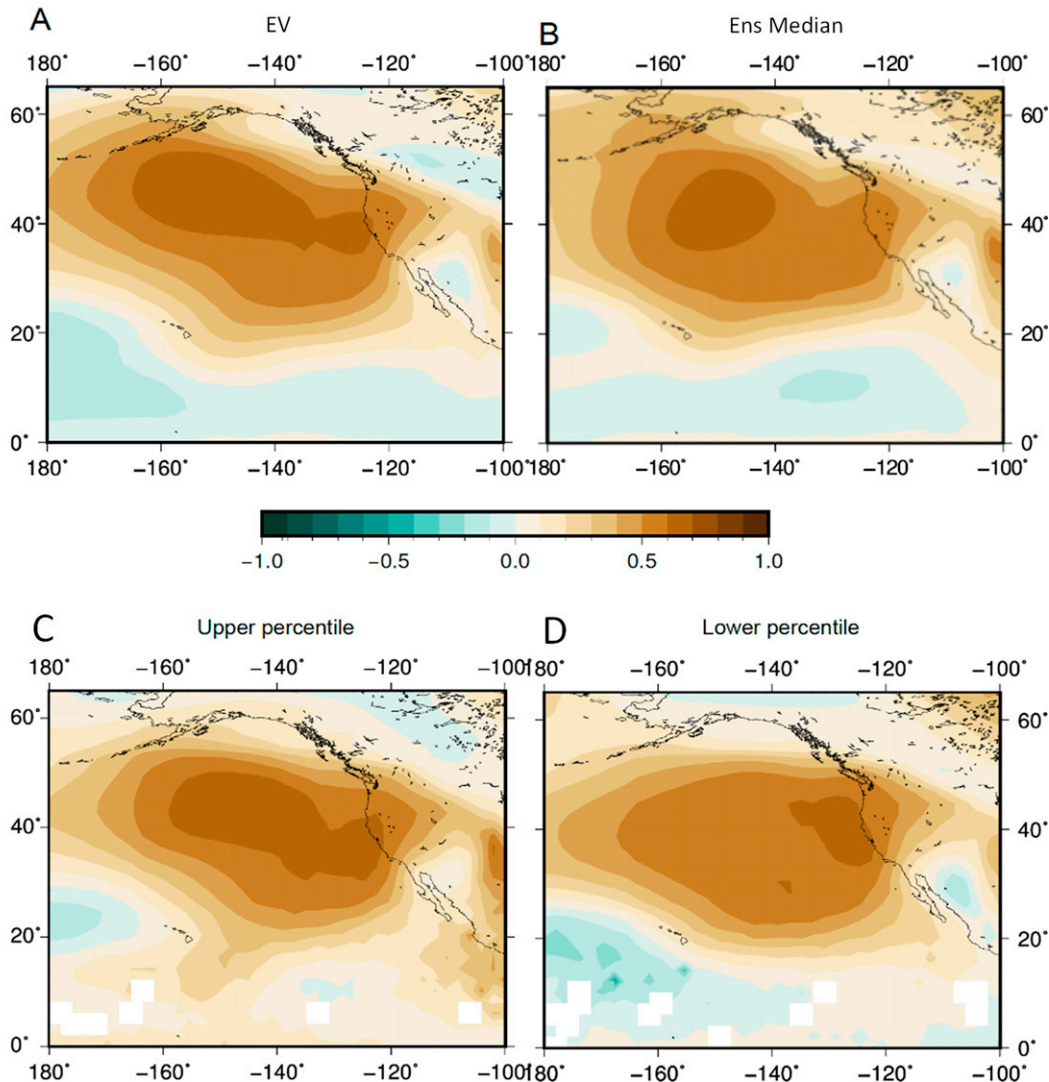FIG. 2. As in Fig. 1, but for DJF 200-hPa GPH fields.

FIG. 3. As in Fig. 1, but for AA+KF Ens (rather than EV) SLP output: (a) as in Fig. 1a, for AA-only output; (b) as in Fig. 1b, but showing AA+KF Ens median output; (c) AA+KF Ens 80th percentile output; and (d) AA+KF Ens 20th percentile output. Noncolored grid cells indicate tropical areas with little DJF variability to estimate upper and lower percentile values. The missing data in (c) and (d) are the result of null temporal variability in a few grid cells in the 20CR reanalysis.

of Washington, Oregon, and Northern CA. As with SLP. We note again the identification of the core regions of known drivers for, for the relationship of northeast Pacific 200-hPa u and v wind fields with CA precipitation and fire indices (Wahl et al. 2019; their Fig. 1), highlighting the physical realism of the entire model and postprocessing chain (climate model, AA method, and AA+KF refinement).The overall closeness of the AA-only and AA+KF results also indicates that the addition of the KF postprocessing component does not introduce spatial instability or other distortions to the overall reconstruction process.

These improvements indicate that the inclusion of the KF postprocessing component with the AA-only results does indeed identify and exploit useful information from the innovation terms, even given the inherent uncertainty in the paleoreconstructions that helps govern the potential KF gain involved. The modest nature of the improvements within this context suggests that the AA method by itself is nearing optimality in this region-specific case. This occurrence is at least to some degree expectable, considering the relatively rich coverage of regional gridded paleo-reconstructions in western North America that are physically suitable for reconstruction of upstream pressure and winds features in the northeast Pacific. Additional considerations for further skill improvement within this region, per se, and the potential for enhancement of the AA method by the KF postprocessor in other regions are taken up in section 4.

The ensemble Kalman filter methodology, encapsulated in Eqs. (5) and (6), provides the median reconstruction and an estimate of the spread of the reconstruction uncertainty,
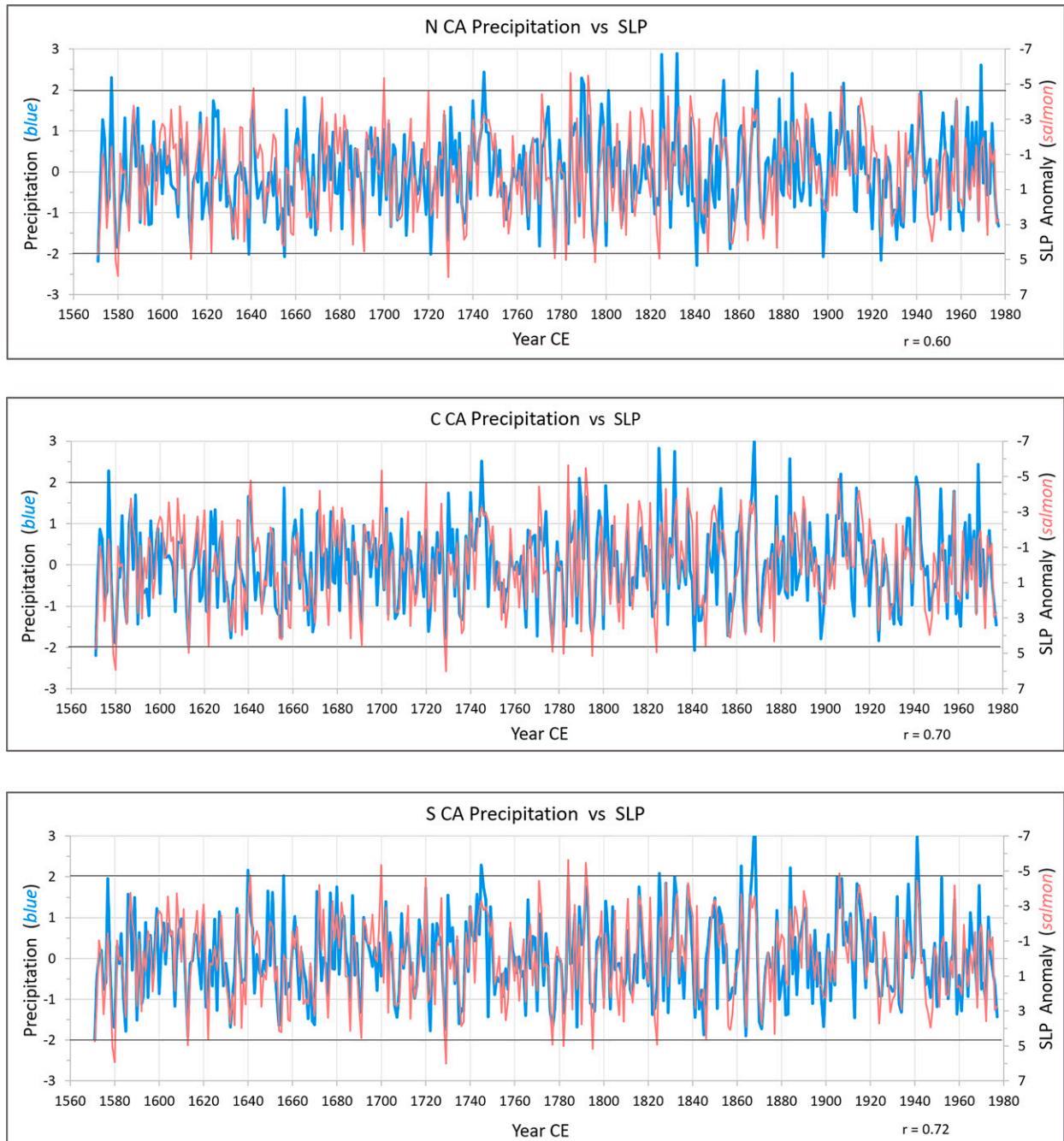
FIG. 4. Reconstructed WY precipitation (blue) for (top) Northern (N), (middle) Central (C), and (bottom) Southern (S) CA regions (Wahl et al. 2020), and DJF SLP index (salmon) for the 32°–45°N, 111°–132°W grid box as developed with the AA+KF method described herein. The time period is 1571–1977 (coverage of precipitation data). Precipitation values are standardized anomalies and SLP values are inverted anomalies in hPa, both relative to full reconstruction period.

which in this case we quantify with the 20th–80th percentiles of the ensemble. This is a natural product of the EnsKF and represents a distinct extension that it provides to the AA-only results. The validation performance of the AA+EnsKF for SLP is shown in Fig. 3. The median results from the ensemble (Fig. 3b) are spatially quite close to those of the EV (Fig. 1b),

although slightly less skillful at the maximum and more like those attainable by the AA method alone (Figs. 1a and 3a). The 20th and 80th percentiles (Figs. 3c,d) show quite reasonable skill decrement/enhancement relative to the median, indicating good utility of the EnsKF for uncertainty estimation in the context of the combined AA+KF method. The overall
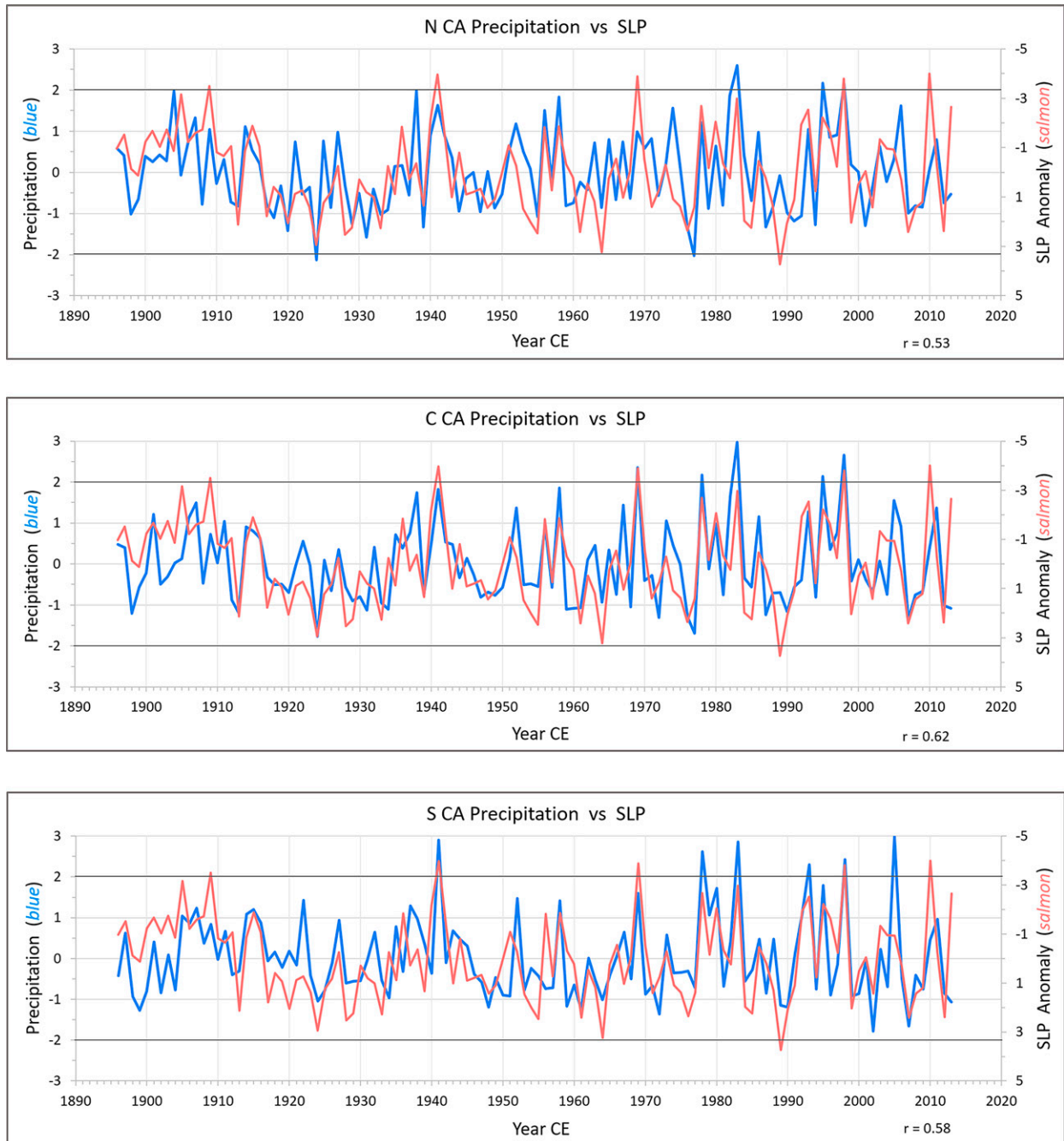
FIG. 5. As in Fig. 4, but for instrumental (precipitation; NCEI 2019) and reanalysis (SLP; 20CR, Compo et al. 2011; Slivinski et al. 2019) data. Time period is 1896–2013.

character of the EnsKF results represents additional confirmation that the joining of the KF as postprocessor with the AA method is both algorithmically and physically reasonable.

b. Northeast Pacific winter SLP and California precipitation over the past 450 years

As an example of real-world employment of the AA+KF reconstructions, we evaluate the relationship between the SLP reconstruction and reconstructions of CA WY precipitation (Fig. 4). The reason for probing this relationship in more detail is, as mentioned in section 1, to illustrate its nonlinear character for extreme hydrological years in that region of North America. Typically, climate reconstruction methods assume a linear relationship between the proxy record and the driving climate conditions. A notable feature of the AA method, and therefore also of the AA-KF implementation, is
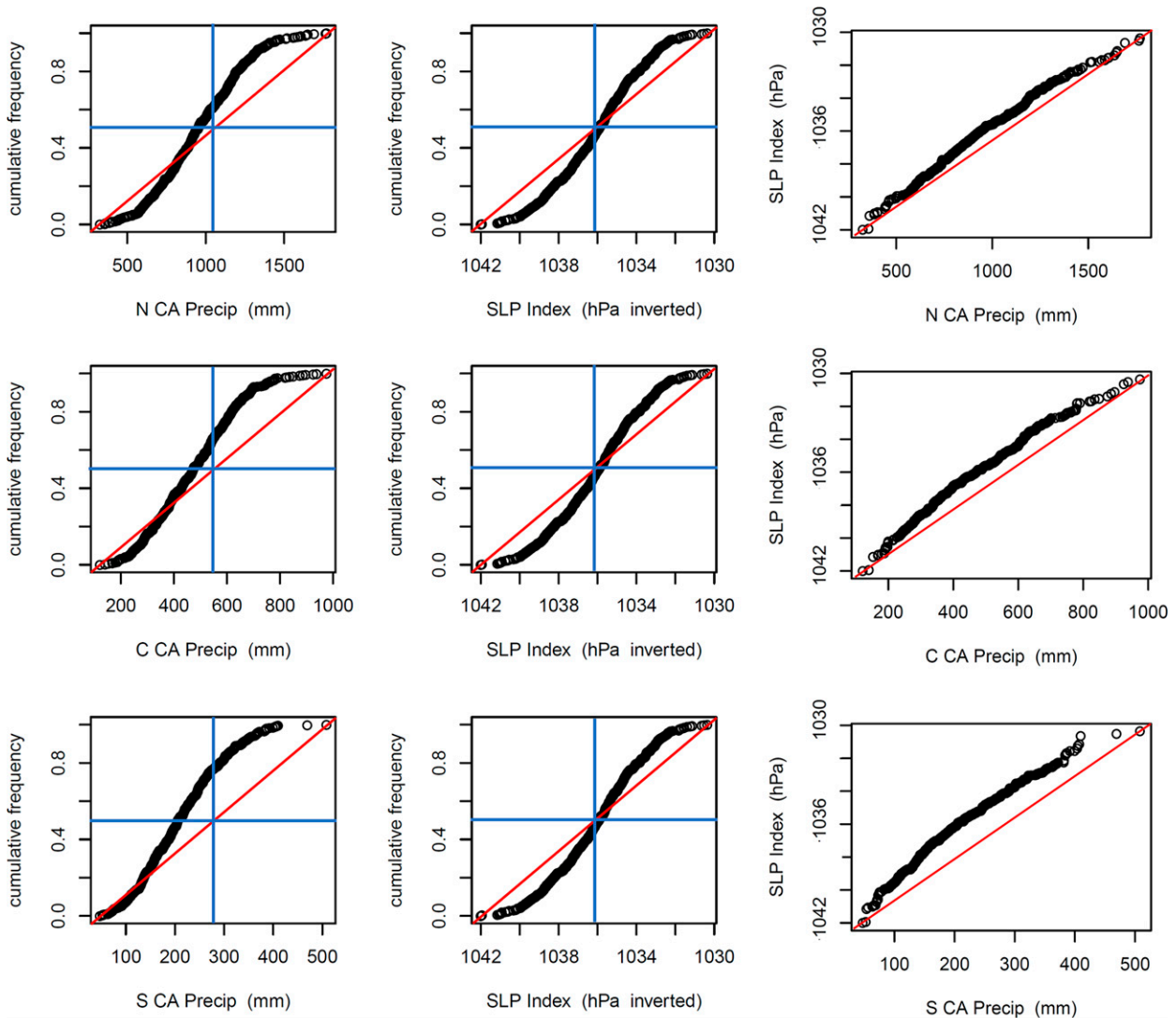
FIG. 6. Empirical CDFs of reconstructed WY (left) precipitation (mm) and (center) SLP (inverted hPa), along with (right) corresponding QQ plots, for the (top) N, (middle) C, and (bottom) S CA regions. The time period is 1571–1977 as in Fig. 4, with the same data sources.

their straightforward capability to capture nonlinear relationships, provided that the pool of analogs sampled from the climate simulations is realistic in this regard. Particularly in Southern CA, the relationship between high precipitation and low SLP in the wettest years versus the relationship between low precipitation and high SLP in the driest years is asymmetric. The reason for this asymmetry lies in the different physical mechanisms that operate in extreme precipitation years. In the wettest years, precipitation is often concentrated within a few days or weeks and is delivered by the passing of a few strong extratropical cyclones or even atmospheric rivers, accompanied by low values of the SLP. Thus, the seasonally averaged SLP is subdued relative to the amplitude of the seasonally accumulated precipitation. By contrast, very dry years are caused by atmospheric blocking, with longer periods of anomalously high SLP that more strongly affect the

seasonally averaged SLP (Fig. 4, bottom; cf. Fig. 5, bottom). Many positive peaks of precipitation (blue line) are not accompanied by corresponding negative peaks of SLP (red line; note the inverted scale), whereas negative peaks of precipitation are more frequently matched or nearly matched by corresponding positive peaks of SLP.

This difference can be viewed more analytically via the empirical cumulative density distributions (called "CDFs" for convenience) of precipitation and SLP, along with the corresponding quantile–quantile (QQ) relationships between these variables (Figs. 6 and 7; see also Fig. A1 in the appendix). The precipitation distribution is skewed dry with a long tail to wet extremes, whereas the SLP distribution is relatively more symmetric/normal in character. Central and particularly Southern CA stand out in this comparison in the reconstruction data: both over the full reconstruction period
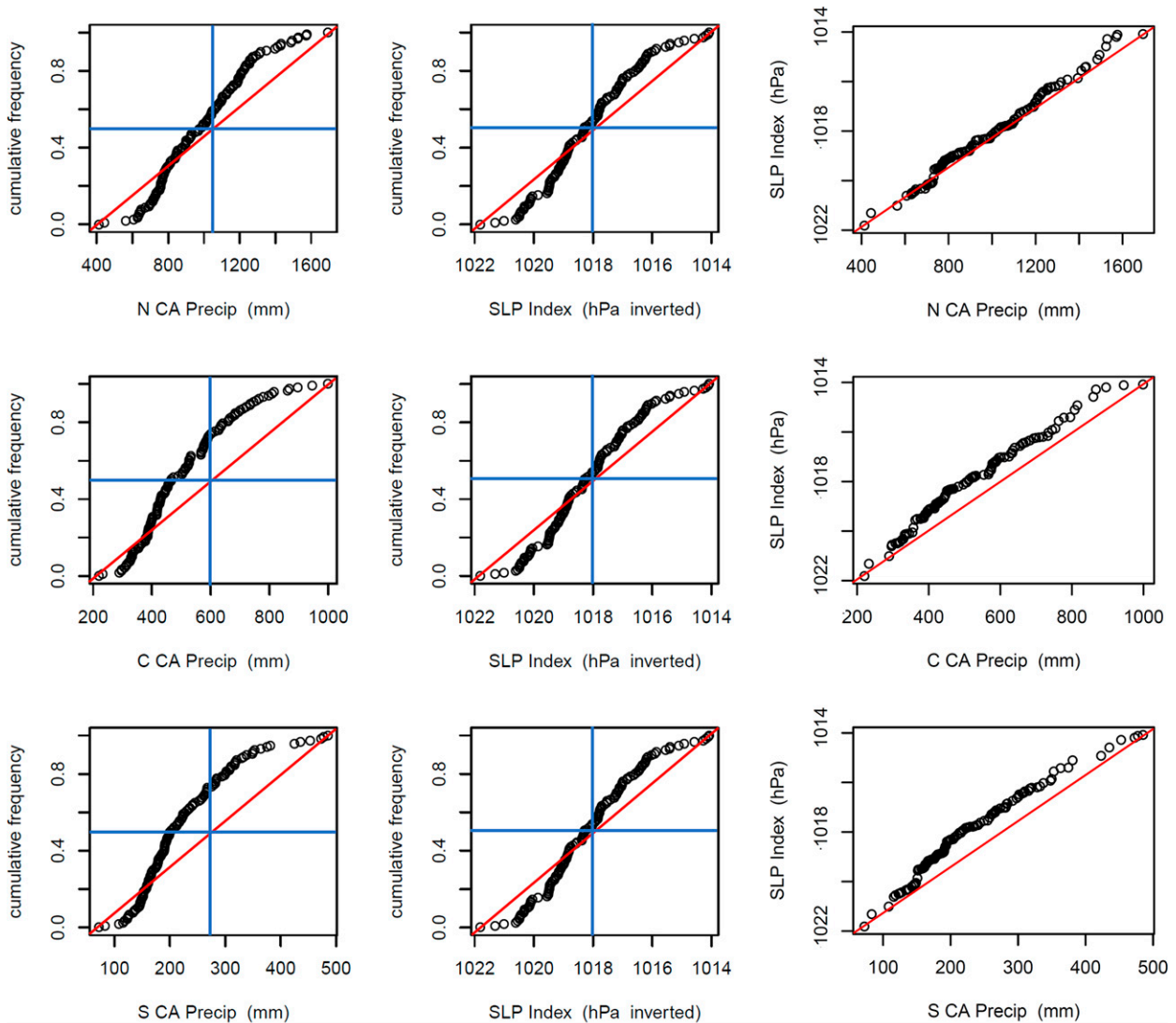
FIG. 7. As in Fig. 6, but for instrumental (precipitation) and reanalysis (SLP) data. The time period is 1896–2013 as in Fig. 5, with the same data sources.

1571–1977 (Fig. 6) and from 1571 to 1895 to be separate of the instrumental data (see Fig. A1 in the appendix). Both central and Southern CA stand out as skewed in the instrumental data from its start in 1896–2013 (the end of the record for the SLP data used) (Fig. 7). We note that apparent differences between the reconstruction and instrumental precipitation CDFs in Figs. 6, 7, and A1, especially for central CA, are in fact related to differences in the scaling of the *x* axes, since the reconstruction data include drier years than experienced in the instrumental record. Referencing all data on identical axes and smoothing shows much smaller differences and highlights the extended extreme dry observations of the reconstruction record (see Fig. A2 in the appendix). The summed absolute differences between the smoothed CDFs are not significant (at any level $p \leq 0.5$) relative to a resampling Monte Carlo analysis of

random CDF summed absolute differences (see text associated with Fig. A2).

More dynamically, evaluation of winter daily SLP values associated with the extreme wet and dry deciles of seasonal winter precipitation since 1950 (Fig. 8) indicates that daily-level low pressure extremes are highly significantly associated with the wettest years, whereas daily high pressures are not as strongly associated with the extreme dry years. Although this characteristic is derived from instrumental/reanalysis data only (we know of no climate proxy information that could be used to constrain reconstruction of daily SLP values), it is reasonable to assume that it has been the case over the reconstruction period as well. As noted, this difference highlights the important role in CA moisture delivery played by individual storms or sequences of them, sometimes associated with entrained subtropical moisture plumes to form atmospheric rivers, whereas the
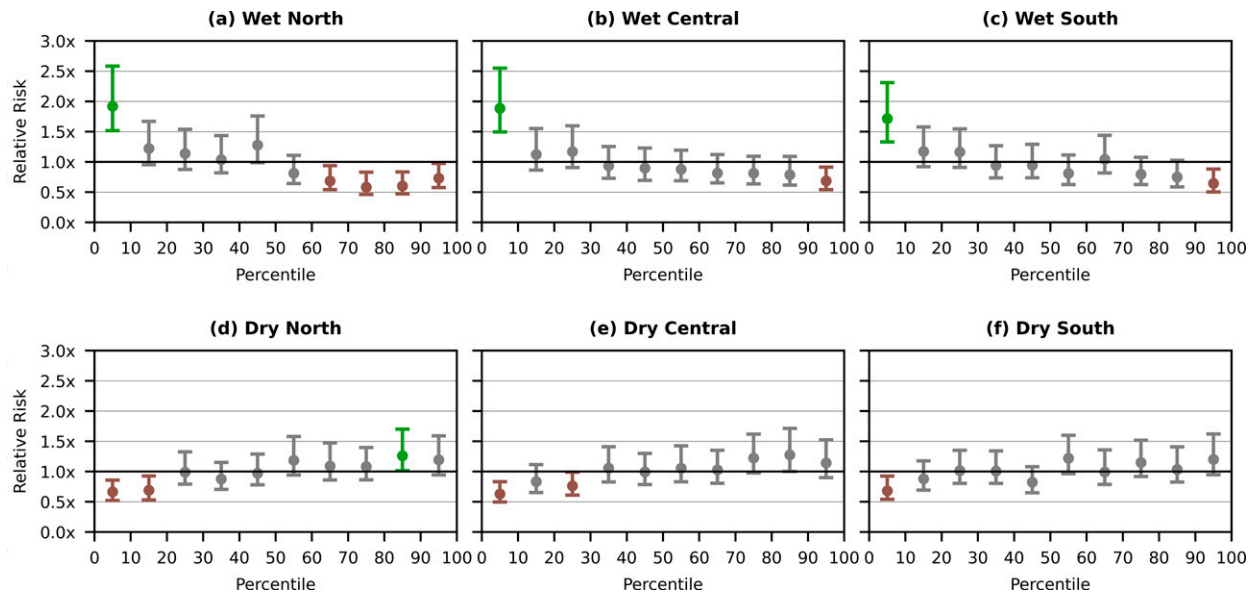
FIG. 8. Relative risk of daily SLP occurrences during (top) wet and (bottom) dry DJF seasons relative to neither wet nor dry winter seasons, for (left) N, (center) C, and (right) S CA. EV (dot) and 95% confidence intervals (whiskers) are estimated from bootstrapped random samples with replacement ($n = 10\,000$) drawn from the collection of neither wet nor dry seasons of a size corresponding to either wet or dry seasons, with relative risk determined for each iteration. Green dot/whisker data indicate relative risk > 1 at 95% confidence; red dot/whisker data indicate relative risk < 1 at 95% confidence.

overall dry characteristic of CA's climate is less influenced by particularly strong high pressure. This enhanced risk profile is an important consideration from the perspectives of water resource development and allocation, emergency management, and infrastructure engineering in one of the world's largest economies and most populated regions in North America (Wahl et al. 2017).

4. Summary and discussion

The work presented here combines the analog assimilation method with a Kalman filter to produce an improved AA climate reconstruction. Our method follows and expands on the application developed by Pfister et al. (2020) to assimilate historical observations (of temperature) that are spatially and seasonally collocated with the model domain. Here, our methodology aims at the assimilation of spatially and seasonally non-collocated proxy reconstructions of several variables (summer soil moisture, winter temperature, and water year precipitation) to reconstruct a different target variable (winter sea level pressure). The results are successful from the standpoint of both mathematical/algorithmic and operational characteristics of the joining, which can be described technically as utilizing the KF as a postprocessor to the fundamental AA output. The joined methodology importantly easily allows direct incorporation of different proxy-based reconstructions (which may also represent different seasonalities), based on the expertise, rigorous testing, and skill-seeking improvement of the existing spatial paleoreconstructions of climatological variables themselves, rather than the need to rely on additional statistical modeling

of the links between proxy observations and KF priors. This exploitation of existing high-quality reconstructions that are closely physically related to the reconstruction target variables is a key motivator of the AA+KF joining.

The skill gains we find from addition of the KF postprocessor component are overall modest in our test case of reconstructing SLP and 200-hPa fields in the northeast Pacific utilizing downstream paleoclimatic temperature and moisture reconstructions in western temperate North America. This can be due to the already good skill of the AA in this particular application (Wahl et al. 2019). The peak correlations between reconstructions and 20CR reanalysis in the northeast Pacific are in the range 0.6–0.7, which, considering the intermediate steps involved—analog selection from a climate model output, and use of gridded tree ring–based reconstructions of different variables—is already a remarkable set in comparison to other reconstruction methods, such as direct calibration to proxy data. We believe that the KF refinement could potentially provide a clearer improvement in cases in which the quality of the proxy-based reconstructions is less. Notably, though, SLP reconstruction skill is specifically enhanced in the oceanic subregion south of central Alaska that is strongly associated with both wet winters in western North America and dry northeasterly winter conditions in Hawaii. Our overall interpretation of these results is that the AA method, per se, is already approaching optimality in the test case situation, driven by the relative richness, quality, and pressure-related characteristics of the paleoreconstruction "predictor" information assimilated and the physical realism of the climate model employed, to which the KF postprocessing step is then added. Whether additional skill
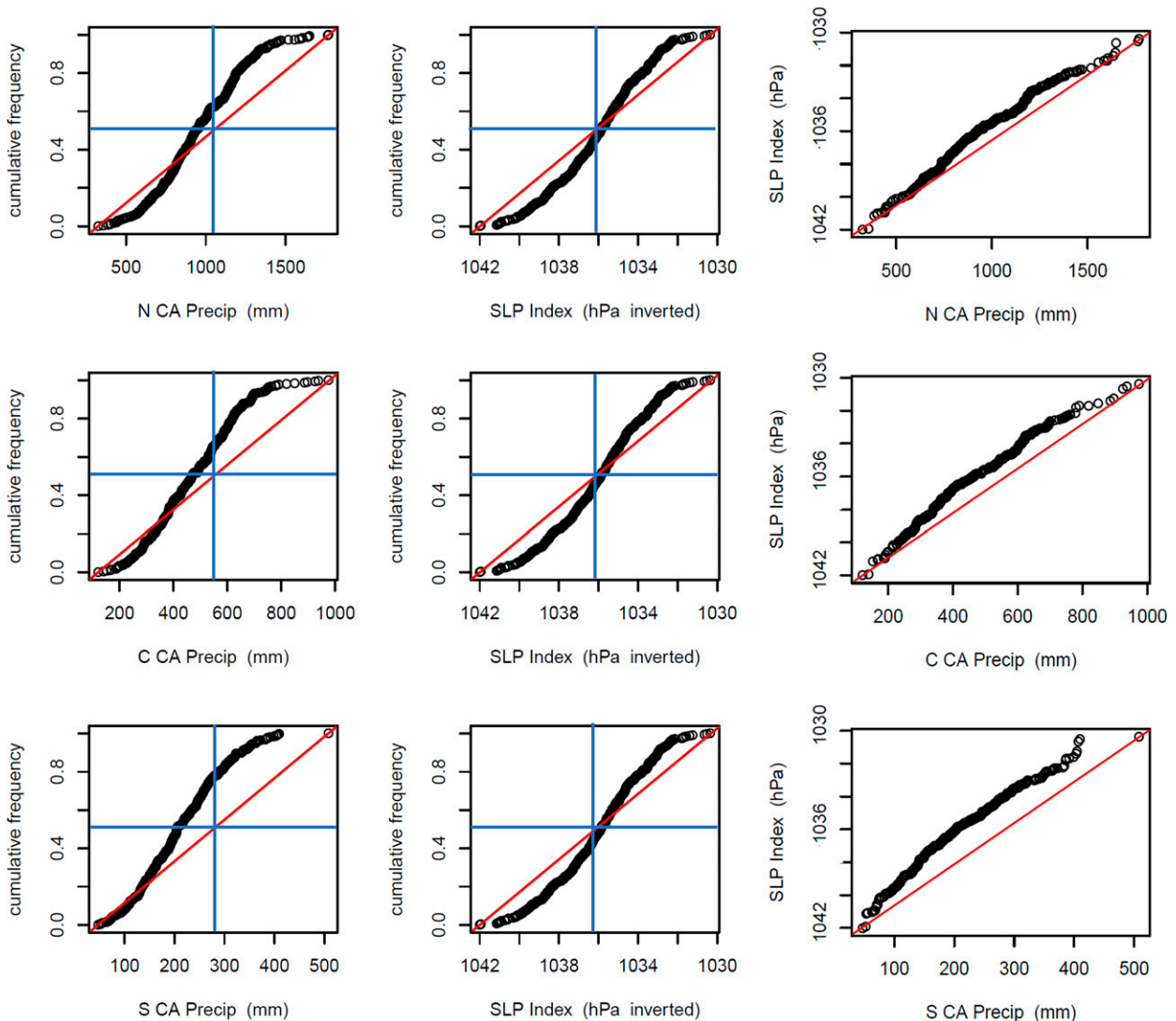
FIG. A1. As in Fig. 6, but for reconstruction data covering 1571–1895.

enhancements to the paleo moisture and temperature pre-
dictor reconstructions we have used could result in greater
skill enhancement for pressure variable reconstruction by
the KF postprocessing component is complex to evaluate.
Given the physical realism of the climate model information
utilized, the AA-component pressure reconstructions would
generally be expected to improve as well and thus the in-
novation terms themselves might be reduced in their po-
tential skill-enhancing impact. At the same time, greater
skill for the paleoreconstruction predictors would gener-
ally reduce the relative effect of the **R** matrix in Eqs. (2)
and (5) and thereby increase the gain characteristic of the
K matrices for incorporation of the innovation information.
Which effect might predominate is not determinate as a gen-
eral characteristic, but rather would need to be evaluated on
a case by case basis, such as that considered here regarding
the western North American predictor reconstructions.

Observing that the expected value (EV) output of the
joined AA+KF methodology slightly outperforms the median
output of the ensemble Kalman filter version, we recommend
experimenting with using the EnsKF results to determine
given percentile levels as ± anomalies relative to the corre-
sponding median values and then adding these to the EV
results. Doing this would retain the best-performance char-
acteristic of the EV results while at the same time incor-
porating the enhanced uncertainty estimation provided by
the EnsKF formalism. Algorithmically, we highlight the
importance and successful use of orthogonal (eigenvector)
space reduction in the process as we have implemented it,
which also helps to ensure both computational stability
and efficiency.

Finally, we note it is possible that the joined AA+KF
method would add additional skill over the AA output itself
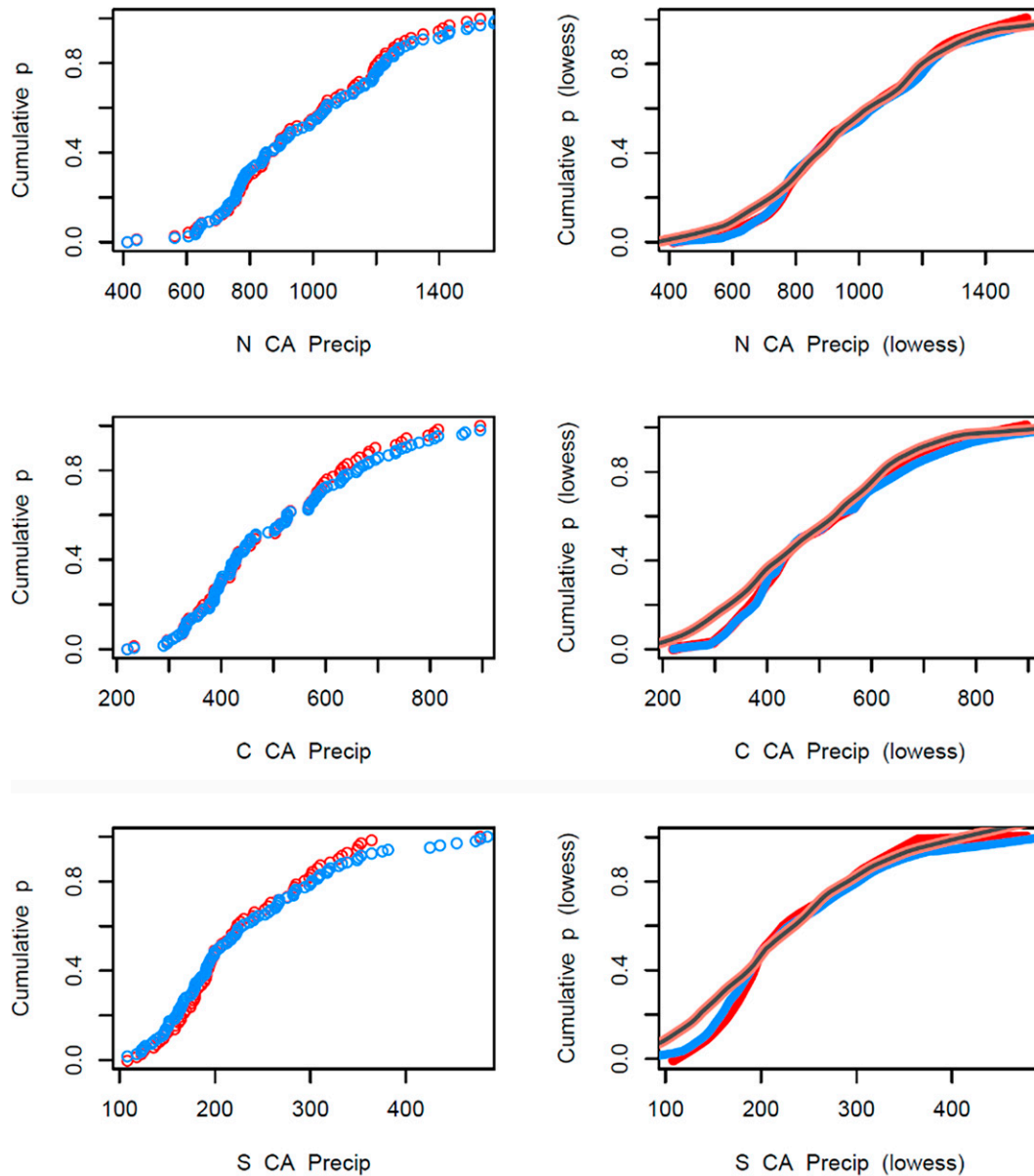in other, significantly different, reconstruction situations, and

FIG. A2. Empirical CDFs of precipitation in N, C, and S CA regions as in Figs. 6 and 7 and Fig. A1, but for (left) instrumental data and (right) smoothed instrumental and reconstruction data using identical horizontal axis scaling for both types of data. Red/blue dots at left and smoothed curves at right indicate instrumental data covering 1906–77 (red) and 1906–2013 (blue) (NCEI 2019); salmon smooth with dark center stripe at right indicates reconstruction data (Wahl et al. 2020) covering 1571–1895. For graphic presentation, lowess curves show ~12-yr smoothing period for instrumental data and ~36-yr smoothing period for reconstruction data.

for other final output variables besides pressure-related ones. Our intent for further research is to apply the methodology to the northeast Atlantic/western coastal Europe region, in that context again with pressure fields as the target, where regional paleoreconstructions of precipitation (Pauling et al. 2006), soil moisture (Cook et al. 2015), and temperature (Luterbacher et al. 2016) are also available. The recent extension of gridded precipitation reconstructions at global scale (Steiger et al. 2018), along with the completion of other continental/subcontinental-scale drought atlases (e.g., Cook et al. 2010b) and regional-scale temperature reconstructions (e.g., Neukom et al. 2011), suggests that exploration of pressure reconstructions for additional parts of the globe will be worthwhile to undertake.

and produced outputs and related figures. EW, AH, and EZ implemented climatological analyses and related figures. EW and EZ wrote the text with input from AH. The work by EZ is part of the project Redmod (reduced complexity models), funded by the Helmholtz Association. Initial work by EW for this project was done at the NOAA National Centers for Environmental Information.

Data availability statement. The AA+KF derived reconstructions of NE Pacific SLP and 200-hPa GPH and *u* and *v* wind fields are available from World Data Service for Paleoclimatology (NOAA/NCEI) upon publication.

APPENDIX

Empirical CDFs of Precipitation and SLP from the Preinstrumental Period with Monte Carlo Estimation of Differences between Reconstruction and Instrumental Precipitation CDFs

Figure A1, corresponding to Fig. 6 in the primary text, shows reconstruction data covering 1571–1895 separately from the instrumental data, which start at 1896.

Figure A2 provides empirical CDFs of precipitation in N, C, and S CA regions as in Figs. 6, 7, and A1, using identical horizontal axis scaling for both types of data.

The summed absolute differences between comparable 36-yr subsamples of the reconstruction and instrumental CDFs in Fig. A2 (all smoothed using ~12-yr lowess) are not significant at any level $p \leq 0.5$, thus inclusive of the commonly used 0.05 and 0.10 levels. A Monte Carlo analysis of summed absolute differences between random 36-yr subsamples of the smoothed reconstruction data ($n = 124\,750$) was used to benchmark significance. This sample size represents all nonrepeating combinations of differences between 500 original random 36-yr subsamples.

The 36-yr subsamples were chosen to enable the summing of absolute differences between the three data periods in a uniform and thereby directly comparable way. For the three periods shown in Fig. A2—72 years for 1906–77 instrumental data (red; to the end of reconstruction period), 108 years for 1906–2013 instrumental data (blue; to the end of SLP reanalysis data), and 324 years for 1571–1895 reconstruction data (salmon with dark center stripe; to the end of preinstrumental reconstruction data)—36 is the largest number of years by which these periods can be divided with resulting whole-number quotients (2, 3, and 9, respectively). Along with identical horizontal axis scaling, standardized-year sampling is necessary as the summation of absolute differences between varying-length CDFs needs to be based on the same number of differences in each case to ensure direct comparability of magnitudes. This is done, for example, by comparing each ninth element of the smoothed reconstruction CDF with each third element of the 1906–2013 instrumental CDF. The smoothing similarly ensures that gaps and relative "roughness" of the empirical CDFs do not bias the difference calculations.

REFERENCES

Anderson, D. M., E. M. Mauk, E. R. Wahl, C. Morrill, A. J. Wagner, D. Easterling, and T. Rutishauser, 2013: Global warming in an independent record of the past 130 years. *Geophys. Res. Lett.*, **40**, 189–193, https://doi.org/10.1029/2012GL054271.

Bhend, J., J. Franke, D. Folini, M. Wild, and S. Brönnimann, 2012: An ensemble-based approach to climate reconstructions. *Climate Past*, **8**, 963–976, https://doi.org/10.5194/cp-8-963-2012

Castello, A. F., and M. L. Shelton, 2004: Winter precipitation on the US Pacific coast and El Niño–Southern Oscillation events. *Int. J. Climatol.*, **24**, 481–497, https://doi.org/10.1002/joc.1011.

Cayan, D. R., and J. O. Roads, 1984: Local relationships between United States west coast precipitation and monthly mean circulation parameters. *Mon. Wea. Rev.*, **112**, 1276–1282, https://doi.org/10.1175/1520-0493(1984)112<1276:LRBUSW>2.0.CO;2.

Chang, E., S. Lee, and K. L. Swanson, 2002: Storm track dynamics. *J. Climate*, **15**, 2163–2183, https://doi.org/10.1175/1520-0442(2002)015<02163:STD>2.0.CO;2.

Compo, G. P., and Coauthors, 2011: The Twentieth Century Reanalysis Project. *Quart. J. Roy. Meteor. Soc.*, **137** (654), 1–28, https://doi.org/10.1002/qj.776.

Conil, S., and A. Hall, 2006: Local regimes of atmospheric variability: A case study of southern California. *J. Climate*, **19**, 4308–4325, https://doi.org/10.1175/JCLI3837.1.

Cook, E. R., K. R. Briffa, and P. D. Jones, 1994: Spatial regression methods in dendroclimatology: A review and comparison of two techniques. *Int. J. Climatol.*, **14**, 379–402, https://doi.org/10.1002/joc.3370140404.

——, R. Seager, R. Heim Jr., R. S. Vose, C. Herweijer, and C. Woodhouse, 2010a: Megadroughts in North America: Placing IPCC projections of hydroclimatic change in a long-term palaeoclimate context. *J. Quat. Sci.*, **25**, 48–61, https://doi.org/10.1002/jqs.1303.

——, K. Anchukaitis, B. Buckley, R. D'Arrigo, G. Jacoby, and W. Wright, 2010b: Asian monsoon failure and megadrought during the last millennium. *Science*, **328**, 486–489, https://doi.org/10.1126/science.1185188.

——, and Coauthors, 2015: Old World megadroughts and pluvials during the Common Era. *Sci. Adv.*, **1**, e1500561, https://doi.org/10.1126/sciadv.1500561.

Dettinger, M. D., 2013: Atmospheric rivers as drought busters on the U.S. west coast. *J. Hydrometeor.*, **14**, 1721–1732, https://doi.org/10.1175/JHM-D-13-02.1.

Diaz, H. F., and E. R. Wahl, 2015: Recent California water year precipitation deficits: A 440-year perspective. *J. Climate*, **28**, 4637–4652, https://doi.org/10.1175/JCLI-D-14-00774.1.

——, ——, E. Zorita, T. G. Giambelluca, and J. Eischeid, 2016: A 480-year reconstruction of Hawaiian Islands rainfall. *J. Climate*, **29**, 5661–5674, https://doi.org/10.1175/JCLI-D-15-0815.1.

Franke, J., J. F. González-Rouco, D. Frank, and N. E. Graham, 2011: 200 years of European temperature variability: Insights from and tests of the proxy surrogate reconstruction analog method. *Climate Dyn.*, **37**, 133–150, https://doi.org/10.1007/s00382-010-0802-6.

——, V. Valler, S. Brönnimann, R. Neukom, and F. Jaume-Santero, 2020: The importance of input data quality and quantity in climate field reconstructions—Results from the assimilation of various tree-ring collections. *Climate Past*, **16**, 1061–1074, https://doi.org/10.5194/cp-16-1061-2020.

Giorgetta, M. A., and Coauthors, 2013: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *J. Adv. Model. Earth Syst.*, **5**, 572–597, https://doi.org/10.1002/jame.20038.

Gómez-Navarro, J. J., E. Zorita, C. C. Raible, and R. Neukom, 2017: Pseudo-proxy tests of the analogue method to reconstruct spatially resolved global temperature during the Common Era. *Climate Past*, **13**, 629–648, https://doi.org/10.5194/cp-13-629-2017.

Goosse, H., E. Crespin, A. de Montety, M. E. Mann, H. Renssen, and A. Timmermann, 2010: Reconstructing surface temperature changes over the past 600 years using climate model simulations with data assimilation. *J. Geophys. Res.*, **115**, D09108, https://doi.org/10.1029/2009JD012737.

Graham, N., and Coauthors, 2007: Tropical Pacific–mid-latitude teleconnections in medieval times. *Climatic Change*, **83**, 241–285, https://doi.org/10.1007/s10584-007-9239-2.

Guiot, J., A. Nicault, C. Rathgeber, J. L. Edouard, F. Guibal, G. Pichard, and C. Till, 2005: Last-millennium summer-temperature variations in western Europe based on proxy data. *Holocene*, **15**, 489–500, https://doi.org/10.1191/0959683605hl819rp.

Hakim, G. J., J. Emile-Geay, E. J. Steig, D. Noone, D. M. Anderson, R. Tardif, N. Steiger, and W. A. Perkins, 2016: The last millennium climate reanalysis project: Framework and first results. *J. Geophys. Res. Atmos.*, **121**, 6745–6764, https://doi.org/10.1002/2016JD024751.

Haston, L., and J. Michaelsen, 1997: Spatial and temporal variability of southern California precipitation over the last 400 yr and relationships to atmospheric circulation patterns. *J. Climate*, **10**, 1836–1852, https://doi.org/10.1175/1520-0442(1997)010<1836:SATVOS>2.0.CO;2.

Hoell, A., M. Hoerling, J. Eischeid, K. Wolter, R. Dole, J. Perlwitz, T. Xu, and L. Cheng, 2016: Does El Niño intensity matter for California precipitation? *Geophys. Res. Lett.*, **43**, 819–825, https://doi.org/10.1002/2015GL067102.

Jones, P. D., and Coauthors, 2009: High-resolution paleoclimatology of the last millennium: A review of current status and future prospects. *Holocene*, **19**, 3–49, https://doi.org/10.1177/0959683608098952

Lahoz, W., and P. Schneider, 2014: Data assimilation: Making sense of Earth observation. *Front. Environ. Sci.*, **2**, 16, https://doi.org/10.3389/fenvs.2014.00016.

Luterbacher, J., and Coauthors, 2016: European summer temperatures since Roman times. *Environ. Res. Lett.*, **11**, 024001, https://doi.org/10.1088/1748-9326/11/2/024001.

Mo, K. C., and R. W. Higgins, 1998: Tropical influences on California precipitation. *J. Climate*, **11**, 412–430, https://doi.org/10.1175/1520-0442(1998)011<0412:TIOCP>2.0.CO;2.

National Research Council (USA), 2006: Surface temperature reconstructions for the last 2,000 years. The National Academies Press, 155 pp.

NCEI, 2019: Climate at a glance (for California water year precipitation). Accessed 24 October 2019 and 19 April 2020, https://www.ncdc.noaa.gov/cag/statewide/time-series/4/pcp/12/9/1895-2019?base_prd=true&begbaseyear=1901&endbaseyear=2000.

Neukom, R., and Coauthors, 2011: Multiproxy summer and winter surface air temperature field reconstructions for southern South America covering the past centuries. *Climate Dyn.*, **37**, 35–51, https://doi.org/10.1007/s00382-010-0793-3.

——, N. Steiger, J.-J. Gómez-Navarro, J. Wang, and J. P. Werner, 2019: No evidence for globally coherent warm and cold periods over the preindustrial Common Era. *Nature*, **571**, 550–554, https://doi.org/10.1038/s41586-019-1401-2.

PAGES Hydro2k Consortium, 2017: Comparing proxy and model estimates of hydroclimate variability and change over the Common Era. *Climate Past*, **13**, 1851–1900, https://doi.org/10.5194/cp-13-1851-2017.

PAGES 2k Consortium, 2019: Consistent multi-decadal variability in global temperature reconstructions and simulations over the Common Era. *Nat. Geosci.*, **12**, 643–649, https://doi.org/10.1038/s41561-019-0400-0.

Pandey, G. R., D. R. Cayan, and K. P. Georgakakos, 1999: Precipitation structure in the Sierra Nevada of California during winter. *J. Geophys. Res.*, **104**, 12 019–12 030, https://doi.org/10.1029/1999JD900103.

Pauling, A., J. Luterbacher, C. Casty, and H. Wanner, 2006: Five hundred years of gridded high-resolution precipitation reconstructions over Europe and the connection to large-scale circulation. *Climate Dyn.*, **26**, 387–405, https://doi.org/10.1007/s00382-005-0090-8.

Pfister, L., S. Brönnimann, M. Schwander, F. A. Isotta, P. Horton, and C. Rohr, 2020: Statistical reconstruction of daily precipitation and temperature fields in Switzerland back to 1864. *Climate Past*, **16**, 663–678, https://doi.org/10.5194/cp-16-663-2020.

Seager, R., T. Osborn, Y. Kushnir, I. Simpson, J. Nakamura, and H. Liu, 2019: Climate variability and change of Mediterranean-type climates. *J. Climate*, **32**, 2887–2915, https://doi.org/10.1175/JCLI-D-18-0472.1.

Schenk, F., and E. Zorita, 2012: Reconstruction of high resolution atmospheric fields for northern Europe using analog-upscaling. *Climate Past*, **8**, 1681–1703, https://doi.org/10.5194/cp-8-1681-2012.

Schonher, T., and S. E. Nicholson, 1989: The relationship between California rainfall and ENSO events. *J. Climate*, **2**, 1258–1269, https://doi.org/10.1175/1520-0442(1989)002<1258:TRBCRA>2.0.CO;2.

Slivinski, L. C., and Coauthors, 2019: Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. *Quart. J. Roy. Meteor. Soc.*, **145**, 2876–2908, https://doi.org/10.1002/qj.3598.

Steiger, N. J., J. E. Smerdon, E. R. Cook, and B. I. Cook, 2018: A reconstruction of global hydroclimate and dynamical variables over the Common Era. *Sci. Data*, **5**, 180086, https://doi.org/10.1038/sdata.2018.86.

St. George, S., D. Meko, and E. Cook, 2010: The seasonality of precipitation signals embedded within the North American Drought Atlas. *Holocene*, **20**, 983–988, https://doi.org/10.1177/0959683610365937.

Wahl, E. R., and J. Smerdon, 2012: Comparative performance of paleoclimate field and index reconstructions derived from climate proxies and noise-only predictors. *Geophys. Res. Lett.*, **39**, L06703, https://doi.org/10.1029/2012GL051086.

——, H. F. Diaz, J. Smerdon, and C. Ammann, 2014: Late winter temperature response to large tropical volcanic eruptions in temperate western North America: Relationship to ENSO phases. *Global Planet. Change*, **122**, 238–250, https://doi.org/10.1016/j.gloplacha.2014.08.005.

——, ——, R. Vose, and W. Gross, 2017: Multi-century evaluation of recovery from strong precipitation deficits in California. *J. Climate*, **30**, 6053–6063, https://doi.org/10.1175/JCLI-D-16-0423.1.

——, E. Zorita, V. Trouet, and A. H. Taylor, 2019: Jet stream dynamics, hydroclimate, and fire in California: 1600–2000

CE. *Proc. Natl. Acad. Sci. USA*, **116**, 5393–5398, https://doi.org/10.1073/pnas.1815292116.

——, A. Hoell, E. Zorita, E. Gille, and H. F. Diaz, 2020: A 450-year perspective on California precipitation. *J. Climate*, **33**, 10221–10237, https://10.1175/JCLI-D-19-0828.1.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924, https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.

Widmann, M., H. Goosse, G. Schrier, R. Schnur, and J. Barkmeijer, 2010: Using data assimilation to study extratropical Northern Hemisphere climate over the last millennium. *Climate Past*, **6**, 627–644, https://doi.org/10.5194/cp-6-627-2010.