The inherent uncertainty of precipitation variability, trends, and extremes due to internal variability, with implications for Western US water resources

Karen A. McKinnon\textsuperscript{1*} and Clara Deser\textsuperscript{2}
\textsuperscript{1}Department of Statistics and Institute of the Environment and Sustainability; University of California, Los Angeles; Los Angeles, CA 90095; kmckinnon@ucla.edu
\textsuperscript{2}National Center for Atmospheric Research; Boulder, CO 80305
Abstract

The approximately century-long instrumental record of precipitation over land reflects a single sampling of internal variability. Thus, the spatiotemporal evolution of the observations is only one realization of ‘what could have occurred’ given the same climate system and boundary conditions, but different initial conditions. While climate models can be used to produce initial-condition large ensembles that explicitly sample different sequences of internal variability, an analogous approach is not possible for the real world. Here, we explore the use of a statistical model for monthly precipitation to generate synthetic ensembles based on a single record. When tested within the context of the NCAR Community Earth System Model version 1 Large Ensemble (CESM1-LE), we find that the synthetic ensemble can closely reproduce the spatiotemporal statistics of variability in winter precipitation over the extended contiguous United States. We additionally create a synthetic ensemble based on the Global Precipitation Climatology Centre (GPCC) dataset, termed the GPCC-synth-LE; comparison of the GPCC-synth-LE with the CESM1-based ensembles reveals differences in the spatial structures and magnitudes of variability, highlighting the advantages of an observationally-based ensemble. We finally use the GPCC-synth-LE to analyze three water resource metrics in the Upper Colorado River Basin: frequency of dry, wet, and whiplash years. Thirty-one year ‘climatologies’ in the GPCC-synth-LE can differ by over 20% in these key water resource metrics due to sampling of internal variability, and individual ensemble members in the GPCC-synth-LE can exhibit large near-monotonic trends over the course of the last century due to sampling of variability alone.
1 Introduction

Precipitation is highly variable both spatially and temporally. Understanding and preparing for this variability has always been critical to human societies, which often rely on consistent water supplies throughout and between years. Given the importance of precipitation and water availability, there has been substantial focus on understanding and predicting changes in precipitation in response to anthropogenic radiative forcing (e.g. Held and Soden, 2006; Trenberth, 2011; Durack et al., 2012; Sarojini et al., 2016; Kooperman et al., 2018). However, the trajectory of the past and future climate system is a function of not only anthropogenic radiative forcing, but also a random sampling of internal variability. The internal variability emerges from processes intrinsic to the coupled climate system, and is not generally predictable after any memory of initial conditions is lost. Because of the highly variable nature of precipitation, the ratio of the externally-forced trend in regional precipitation to internal variability is small over the observational record (e.g., McKinnon and Deser, 2018), and the “time of emergence” – when the forced signal exceeds the noise – is not likely to occur for multiple decades in many regions (Giorgi and Bi, 2009; Mahlstein et al., 2012). Thus, in addition to understanding the forced trend, it is equally critical to properly quantify and model the internal variability of precipitation, which can itself lead to multi-decadal trends.

A dominant source of internal variability is the random fluctuations of the atmospheric circulation. Recent work (Deser et al., 2018) demonstrated how different sampling of this unpredictable component of circulation variability, using either climate model ensembles or statistical resampling of the observations, led to different inferences about the influence of the El Niño-Southern Oscillation (ENSO) on North American precipitation. Indeed, most of the variability in winter US West Coast precipitation has been attributed to internal atmospheric variability rather than tropical or extratropical sea surface temperature (SST) forcing (Dong et al., 2018; Zhang et al., 2021).

On top of this atmospherically-generated noise, which has a large amplitude but minimal
year-to-year memory, precipitation can exhibit lower frequency variations due to ocean influences. These oceanic sources of internal variability, and the atmospheric teleconnections that they induce, are often associated with well-known modes of interannual-to-multidecadal climate fluctuations, particularly ENSO (Ropelewski and Halpert, 1987), the Pacific Decadal Oscillation (PDO; Deser et al., 2004; Wang et al., 2014; Zhang and Delworth, 2015; Newman et al., 2016), and the Atlantic multidecadal variability (AMV; Martin et al., 2014; L’Heureux et al., 2015; Sun et al., 2015; Ruprich-Robert et al., 2018; Simpson et al., 2019).

Cognizant of the important role of internal variability in the climate system – including both atmospheric noise and lower-frequency SST-driven modulations – climate modelers have advanced the concept of “initial-condition large ensembles” for assessing climate variability and change. Such large ensembles consist of a large number of simulations (typically 30-100) with a single fully-coupled climate model under a particular radiative forcing scenario, but with perturbations to the initial conditions. The resulting ensemble spread can be used to characterize the uncertainty in any given climate parameter at any point in time due to unpredictable sampling of internal variability (e.g. Deser et al., 2012; Kay et al., 2015; Maher et al., 2019; Tel et al., 2019). These large ensembles have been used to demonstrate the irreducible uncertainty in climate trends (see Deser et al., 2020, for a recent review), and can be powerful tools for decision-making (Mankin et al., 2020), but do suffer from biases in their simulation of variability and the forced response (McKinnon and Deser, 2018; Suarez-Gutierrez et al., 2020; von Trentini et al., 2020).

A complementary approach to the use of numerical climate simulations is the formulation of statistical models based on the observed record. A common tool used for the purposes of water engineering and planning is the stochastic weather generator (e.g. Wilks and Wilby, 1999), which can be used to simulate time series of a quantity such as precipitation that statistically resembles the observed values at a given location. While these generators have historically focused on high-frequency (daily or subdaily) variability at a single location, they are increasingly designed to better capture decadal variability (Chen et al., 2010) and to in-
corporate the spatial correlation structure of precipitation variability at multiple sites within a single watershed or other entities of comparable size (Kleiber et al., 2012; Steinschneider and Brown, 2013; Verdin et al., 2015; Chen et al., 2018). Nevertheless, these approaches have typically focused on the spatial and temporal scales of weather, rather than climate.

Here, we combine the philosophies of climate model large ensembles and stochastic weather generators into a methodology that allows for the production of synthetic ensembles from the information contained in a single climate record, such as the observations. Our synthetic ensembles preserve both the spatial and temporal correlation structure of precipitation variability on seasonal to multidecadal timescales, in contrast to methods that focus only on the temporal variability at each location separately (e.g. Thompson et al., 2015; Castruccio et al., 2019) so cannot be used to explore variability in large-scale precipitation patterns.

The current work advances the statistical methodology of McKinnon et al. (2017); McKinnon and Deser (2018) and contains novel results based on the synthetic ensembles. The updated statistical methodology focuses on monthly-average precipitation; improves the modeling of the coupled ocean-atmosphere modes including retaining the seasonal cycle of ENSO amplitudes; and contains an automated, rule-based method to choose the block size used in the statistical resampling process. Further, while our prior work focused solely on 50-year trends as a metric for variability, in this work we validate our synthetic ensembles using multiple metrics for internal variability, and use the validation process to explore the more general topic of the challenges of inferring climate statistics from limited data records. We additionally present an analysis of precipitation and water resource metrics for the Upper Colorado River Basin, a major source of water for the Western United States.

The remainder of the paper proceeds as follows. Section 2 provides an overview of the datasets and statistical model used to create the synthetic climate ensembles. Section 3 examines our ability to estimate certain statistical properties of precipitation from a limited record, provides validation for our synthetic ensemble methodology, and compares the
variability in the observations to that in a climate model. Section 4 applies the synthetic
ensemble methodology to precipitation in the Upper Colorado River Basin, with a focus on
three key water resource metrics. We discuss the implications of our findings and conclude
in Section 5.

2 Data sources and statistical model

We first provide details of the climate model output and observational data used in our
analysis, and then describe the statistical model used to create the synthetic ensembles.

2.1 Climate model output and observational data

We use monthly output of precipitation and SST at a nominal spatial resolution of 1 degree
in latitude and longitude from the NCAR Community Earth System Model version 1 Large
Ensemble (CESM1-LE). The CESM1-LE is composed of 40 simulations of CESM1 that that
branch off a parent simulation on January 1, 1920; the spread across the ensemble was
introduced by adding round-off level perturbations (order $10^{-14}$) to the initial atmospheric
temperatures (Kay et al., 2015). The CESM1-LE was forced by the historical forcing scenario
from 1920-2005 (Lamarque et al., 2010) and the RCP8.5 scenario for 2006-2100 (Meinshausen
et al., 2011). We limit our analysis of CESM1-LE to the 1921-2005 period, where we exclude
the first year of the simulations (1920) to reduce any influence of land and atmosphere
initial condition memory, and do not extend beyond the historical period because of different
variability in the forcing related to a lack of episodic volcanic eruptions in the RCP scenarios.

Precipitation observations at one degree resolution are from the Global Precipitation Cli-
matology Centre (GPCC, Schneider et al., 2008), which provides a gridded land-only record
of precipitation from 1891-near present based on in situ measurements. SST observations
used to calculate the time series of ENSO, PDO, and AMV for the observations are from
HadISST (Rayner et al., 2003).
Precipitation for both CESM1-LE and GPCC includes solid and liquid forms, and is normalized to the daily accumulation rate in millimeters (mm) in all of our analyses. For both the CESM1-LE output and the observations, we use the Climate Variability Diagnostics Package (Phillips et al., 2014) to calculate the time series of the modes.

We transform precipitation before estimating the parameters of our statistical model using a Box-Cox power transform (Box and Cox, 1964), which both guarantees that the modeled precipitation amounts remain non-negative, and increases the symmetry and normality of the precipitation distribution. The increased symmetry improves the estimation of the model parameters by reducing the influence of outliers. The parameter $\lambda$ that controls the specific form of the transform is selected via maximization of the log likelihood function for the Gaussian distribution for each gridbox and month independently and has a typical value of 0.34 in CESM1-LE and 0.37 in GPCC. To allow for inter-comparison of model parameters between members of the CESM1-LE, $\lambda$ is estimated using the first member and then used for all members. The few monthly precipitation values that are exactly zero are increased to a very small positive number so that the transform is valid everywhere.

2.2 Statistical model

We first propose to describe monthly-average precipitation over land as a function of four different categories of terms: (1) the mean state, including the annual cycle; (2) the response to external forcing (‘forced component’); (3) the response to large-scale, coupled ocean-atmosphere modes as summarized by the ENSO, PDO, and AMV indices; and (4) the residual ‘climate noise’:

$$P_{i,t} = \beta_0 + \beta_F F_t + \beta_{ENSO} \text{ENSO}_t + \beta_{PDO} \text{PDO}_t + \beta_{AMV} \text{AMV}_t + \epsilon_{i,t}$$ (1)
In our notation, $t$ is time, $m(t)$ is the month, and $i$ is an index of location (gridbox). The $\beta$ coefficients on the right hand side (except $\beta_0$, the mean) describe the monthly-varying spatial pattern of sensitivity of $P$ to anthropogenic forcing, ENSO, PDO, and AMV, respectively.

Following Dai et al. (2015), we initially define the time series used for the estimation of the forced component, $F^t$, as the global mean, ensemble mean time series of near-surface air temperature from the CESM1-LE. By projecting $P$ onto a time series indicative of the evolution of the forced response, the amount of unforced variability aliased onto the forced signal can be reduced compared to assuming a linear trend, although not eliminated. Nevertheless, as will be discussed in Section 3.1, the forced signal in precipitation at a regional level is so small compared to the noise that we find we cannot estimate $\beta^{i,m(t)}_F$ sufficiently well using a single record, so we remove the forced term from our model, and instead use the simpler model,

$$P^{i,t} = \beta_0^{i,m(t)} + \beta^{i,m(t)}_{\text{ENSO}} \text{ENSO}^t + \beta^{i,m(t)}_{\text{PDO}} \text{PDO}^t + \beta^{i,m(t)}_{\text{AMV}} \text{AMV}^t + \epsilon^{i,t} \tag{2}$$

The time series ENSO$^t$ is the standard Niño3.4 SST index. The PDO time series is calculated as the principal component associated with the leading empirical orthogonal function of SST anomalies in the North Pacific, polewards of 20N, where the anomalies are calculated through removing both the climatological annual cycle and global-average SST from each gridpoint. The time series of ENSO and the PDO are highly correlated, and components of the PDO have been suggested to be mid-latitude responses to ENSO (Newman et al., 2016). To more clearly parse the two modes in a statistical sense, we create a version of the PDO time series, PDO$^t_\perp$ that is orthogonal to the standard Niño3.4 time series by removing the projection of the PDO time series onto the Niño3.4 time series. Like the orthogonalized versions of the PDO proposed by Chen and Wallace (2016) and Wills et al. (2018), the SST anomaly pattern associated with PDO$^t_\perp$ is dominated by the classical PDO pattern in the midlatitudes with a negative region of SST anomalies extending from Japan into the central Pacific and a positive region off the west coast of North America, with a much weaker...
tropical component (Supplementary Figure 1). The difference between the SST anomaly pattern of the traditional PDO and that associated with PDO_{⊥,t} strongly resembles an El Niño pattern, and has a pattern correlation with the SST anomaly pattern associated with ENSO_{t} of 0.95 (not shown).

The AMV time series, AMV_{t}, is calculated as the average of SST anomalies in the North Atlantic (0-60N, 80W-0) minus the near-global mean SST (60S-60N) (Trenberth and Shea, 2006). The time series is smoothed with a lowpass Butterworth filter using a forward and backward digital filter and a cutoff frequency of 1/20 year^{-1} in order to isolate the component of the AMV that is primarily ocean-driven (Delworth et al., 2017) and has been shown to have downstream impacts on precipitation (Simpson et al., 2018). The forward-backward Butterworth filter is maximally flat in the passband, does not introduce phase shifts, and performs comparably to other lowpass filters (e.g. Roberts and Roberts, 1978). The AMV_{t} time series is not significantly correlated at the 0.05 level with ENSO_{t} or PDO_{⊥,t} at zero lag. At nonzero lags, there are greater correlations between AMV_{t} and PDO_{⊥,t}, peaking as high as 0.57 when PDO_{⊥,t} leads AMV_{t} by 21 years. However, given the very few degrees of freedom in the AMV_{t} time series, it is difficult to assess whether the relationships are statistically significant and meaningful using the data alone. All three time series (ENSO_{t}, PDO_{⊥,t}, and AMV_{t}) are normalized to have unit standard deviation.

The model is fit using ordinary least squares regression separately for each month due to the known seasonal dependence of the atmospheric teleconnections associated with the modes. The initial estimates of each model parameter are then projected onto the first two harmonics of the annual cycle to enforce the expected smoothness in seasonality. The residual, \( e^{i,t} \), is the ‘climate noise’ that primarily arises from internal atmospheric dynamics, but could also reflect influences from other modes of variability that are uncorrelated with those explicitly considered here. In general, the interannual variance in the residual is greater than that linked to the modes: the fraction of variance in the residual exceeds 0.9 for December-January-February (DJF) in many places across the domain, and still generally
exceeds 0.7 even in the regions with strong teleconnections (Figure 1). As a result, there is considerable scatter around a regression on the mode time series; however, there is not clear evidence of nonlinearity (e.g. Figure 1d, f).

2.3 Generating the synthetic ensemble

In order to move from Eqn. (2), which simply describes the dependence of precipitation on our chosen covariates, to the creation of an ensemble, we must appropriately introduce stochasticity. In the context of Eqn. (2), we view all $\beta$ terms as fixed and reflective of the physics of the climate system. In contrast, we view the time series of each mode ($\text{ENSO}^t$, $\text{PDO}^t$, and $\text{AMV}^t$) as well as the residual ($\epsilon^{i,t}$) as stochastic.

To produce alternative versions of the time series of $\text{ENSO}^t$, $\text{PDO}^t$, and $\text{AMV}^t$, we employ the Iterative Amplitude Adjusted Fourier Transform (IAAFT) method ($\text{Schreiber and Schmitz}, 1996$), which produces synthetic time series that have the same amplitude distributions and power spectra as the originals. We additionally modify the algorithm to retain the seasonal cycle of ENSO amplitudes. An unavoidable constraint of the IAAFT approach is the reliance on the empirical amplitude distribution and power spectra, which themselves are limited by what we have observed. For example, a synthetic ENSO time series would never contain an El Niño event that is larger than the largest observed event, even if such an event could theoretically be produced by the climate system. The resulting synthetic time series of the three modes, by design, do not exhibit coherence with each other, which is consistent with the lack of significant synchronous coherence of the modes in the observational record (Supplementary Figure 2).

To produce alternative realizations of $\epsilon^{i,t}$, we perform a moving block bootstrap in time, where the block size is constrained to be an integer number of years to retain any seasonality in the variability. The block size for each gridbox $i$ and month $m$ is selected using the methods of Wilks (1997) (see their Eqn. (19)), which was developed such that a moving block bootstrap-based significance test on autocorrelated data would reject the null hypothesis.
appropriately. The block size is a function of the autocorrelation and length of the residual
time series. The selected block size for each gridbox is summarized in Supplementary Figure
3, which shows the largest block size selected across months averaged across the full CESM1-
LE, for a single member of the CESM1-LE, for the GPCC observations, and the average
difference between GPCC and individual members of the CESM1-LE. In most regions outside
of North America and Europe, the GPCC-based block size is greater than that estimated
using individual members of the CESM1-LE. The block size for the full spatial field is chosen
as the 97th percentile across gridboxes and months of the selected block sizes, excluding the
dry subtropics (10-30° in both hemispheres) where it is common to have no precipitation for
certain months of the year, leading to estimates of very high autocorrelation. We select a
high percentile rather than the maximum to avoid the undue influence of a small number
of outliers. The block size identified for CESM1-LE is two years, and that for the observed
precipitation from GPCC is four years. The use of a moving block bootstrap in time allows
us to easily retain the complex spatial correlation structure in precipitation without having
to rely on a spatial model with strong parametric assumptions, such as isotropy and spatial
stationarity (e.g. Beusch et al., 2020, for temperature).

Using the subscript ‘synth’ to indicate synthetic data, we can place our alternative ver-
sions of the mode time series and climate noise back into Eqn. (2) to produce new versions
of our precipitation data as follows:

\[
P_{i,t}^{\text{synth}} = \beta_{0}^{i,m} + \beta_{\text{ENSO}}^{i,m} \text{ENSO}_{\text{synth}}^{t} + \beta_{\text{PDO}}^{i,m} \text{PDO}_{\perp}\text{synth}^{t} + \beta_{\text{AMV}}^{i,m} \text{AMV}_{\text{synth}}^{t} + \epsilon_{i,t}^{\text{synth}} \tag{3}
\]

The approach has been applied across the global land masses, although we focus our
subsequent discussion on a subset of North America centered on the contiguous United
States (24-60N, 230-300E) during DJF.
3 How well do we know the statistics of precipitation from a limited data record?

Analysis of climate model large ensembles has highlighted the challenge of estimating the precipitation response to human influence (Deser et al., 2014) and ENSO teleconnections (Deser et al., 2018) from the observational record or a small number of model simulations due to the influence of internal variability independent of the desired signal. In the context of our synthetic ensemble methodology, these topics surface in two ways: first, in the variability of the model parameters of Eqns. (1) and (2), and second, in any dependence of the statistics of precipitation on the specific record on which a synthetic ensemble is based.

We explore both of these questions using 85 years (1921-2005) of model output from the CESM1-LE as our testbed. (see Figure 4 in Deser et al., 2020); this record length is comparable to that of many in situ-based climate datasets. We fit Eqns. (1) and (2) to each of the 40 members of the CESM1-LE, and then use Eqn. (3) to produce 40 synthetic ensemble members from each original member of the CESM1-LE, producing a total of 1600 ensemble members that comprise the CESM1-synth-LE. We then compare the parameters and statistical characteristics of the synthetic ensemble those of the actual CESM1-LE, which is viewed as the truth for the purposes of validation.

Since our ultimate goal is to apply the synthetic methodology to the observations, we also fit Eqn. (2) to the GPCC dataset, using the mode time series based on observed SSTs. The mode time series and residual are randomized 1000 times to produce the GPCC-synth-LE.

We do not account for observational uncertainty (e.g. sensitivity to the choice of data set) in our mode time series, or in the precipitation dataset, in order to focus on uncertainty due to sampling internal variability alone. Similarly, we estimate a single set of regression coefficients based on the mode time series and a fixed period of record, and therefore do not propagate uncertainty in the regression coefficients through the model, although it would be straightforward to do so by performing a bootstrap in time, and repeatedly re-estimating
the parameter values.

### 3.1 Uncertainty in parameter estimation

Through the process of creating the CESM1-synth-LE, we have 40 estimates – one from each original member of the CESM1-LE – of each $\beta$ parameter in the statistical model. The ‘true’ values of the $\beta$ parameters – uncontaminated by sampling of internal variability – should be the same in each ensemble member; we can therefore use the spread across the 40 estimates as a metric for the uncertainty in each parameter due to sampling of internal variability. Figure 2 shows the average estimate of each parameter across the CESM1-LE, the standard deviation of the parameter estimates, and the signal-to-noise ratio (SNR), defined as the absolute value of average divided by standard deviation.

We first demonstrate the challenge of inferring the forced component using Eqn. (1). The SNR of $\beta_F$ is less than one everywhere, and is often close to zero (Figure 2c). Even regions with seemingly large amplitudes for the forced response such as northern Mexico and the US Southwest have a SNR less than one due to the large spread across the ensemble. To more completely visualize the variability in the $\beta_F$ estimates, we plot the map of the parameter estimate for each member of the CESM1-LE (Supplementary Figure 4). The large-scale spatial structures in each panel differ dramatically from each other, with no clear and consistent behavior across the members. Given that the forcing is the same in each model, these differences indicate that it is not possible to accurately estimate the forced component using regression methods for precipitation; rather, the estimates of $\beta_F$ are strongly contaminated by sampling of internal variability. We thus choose to omit the forced component term from our model. This omission does not mean that there is no climate change signal in precipitation, but rather that it is too small compared to the noise to be estimated in the regression model given a single record. All subsequent results do not include estimation of the forced component, i.e. are based on Eqn. (2). Using an alternative method, low-frequency component analysis (Wills et al., 2018), to estimate the forced response does
not increase the SNR (Supplementary Figure 5).

In contrast to the forced component, the across-ensemble spread in estimating the $\beta$ coefficients for each coupled ocean-atmosphere mode is small. For ENSO, the SNR exceeds two across the western and southern parts of the domain where ENSO influences precipitation, and is less than one only in regions with a weak ENSO teleconnection (Figure 2d-f). Similarly, the SNR for the precipitation response to $\text{PDO}_\perp$ is greater than one in most regions where there is a signal, with the exception of interior California (Figure 2g-i). The SNR for the AMV is below one in most places, which reflects the small magnitude of the signal everywhere (Figure 2j-l). The small signal identified for the AMV, and to some extent $\text{PDO}_\perp$, at least within CESM1, is encouraging for our ability to estimate the statistics of precipitation from the observational record, since we have limited samples of these lower-frequency modes of variability.

The high SNR for the $\beta$ coefficients (apart from the forced response) suggests that we can estimate these coefficients within the single observational record. The GPCC-based $\beta$ maps (Figure 3) share many of the same features as CESM1, with a dipole of wetting/drying along the west coast and wetting in the southeast in response to ENSO; a similar but weaker dipole as well as drying in the Midwest in response to $\text{PDO}_\perp$, and a weak response to the AMV.

### 3.2 Spatiotemporal structures in the synthetic ensembles

Having established that the parameters of the statistical model in Eqn. (2) can be well-estimated using a single record, we turn to analysis of the synthetic ensembles. To build intuition for the synthetic ensemble approach, we first show two example fields of January precipitation anomalies from the GPCC dataset and GPCC-synth-LE (Figure 4). The Januaries were arbitrarily chosen to align with low and high January precipitation in the Upper Colorado River Basin and Georgia, respectively, from Figure 1c, e. As can be seen in these examples, the synthetic ensemble approach produces spatially-coherent precipitation fields
that qualitatively reproduce the spatial correlation structures of the original data. The spa-
tial correlation structures are generally retained in the synthetic ensemble due to the spatial
structure of the teleconnection maps (Figure 3) and the use of the block bootstrap approach,
which retains the spatial correlation in the residual, $e^{it}$.

To generalize this evaluation to the full spatiotemporal variability in DJF precipitation,
we perform an empirical orthogonal function (EOF) analysis on DJF-average precipitation
in the CESM1-LE, the CESM1-synth-LE, and the GPCC-synth-LE, where the EOFs are
calculated in each ensemble by appending the members in time (Figure 5). For the EOF
analysis of the synthetic ensembles, we use the 40-member CESM1-synth-LE constructed
from the first member of the CESM1-LE, and similarly only select the first 40 members of
the GPCC-synth-LE, to match the sample size of the original CESM1-LE. We also present
maps of the total variance in each ensemble for comparison.

The total variance in DJF-precipitation in the CESM1-synth-LE is nearly identical to
that in the CESM1-LE (Figure 5a, b). Further, the first four EOFs in the CESM1-LE and
the CESM1-synth-LE are very similar in both their spatial patterns and variance explained,
indicating both that the synthetic ensemble methodology retains the spatiotemporal struc-
ture of the original data, and that fitting the synthetic ensemble with only a single member
is sufficient to capture these structures.

Comparing the GPCC-synth-LE to the CESM1-based ensembles, the total variance in
the GPCC-synth-LE is smaller almost everywhere, although the general structure of high
variance along the West Coast and a secondary maximum in the Southeast is present in both
GPCC and CESM1. The EOFs from the GPCC-synth-LE share many of the large-scale
spatial patterns of those from the CESM1-based ensembles, although some differences are
evident, including a different sign of precipitation anomalies in Texas for EOF1, a different
location of the Eastern negative anomalies in EOF 2, and an opposite phasing of the West
Coast precipitation anomalies in EOF 3. Further, the amplitudes of the GPCC-synth-LE
EOFs over the western half of the continent tend to be weaker than those of CESM1,
consistent with the smaller total variance. The variance explained by the first four EOFs is greater in the CESM1-based ensembles compared to the GPCC-synth-LE (60% versus 53%), and the CESM1-based ensembles have substantially more variance in the first two EOFs than their GPCC-synth-LE counterparts (46-47% versus 32%).

3.3 Characteristics of temporal variance in the synthetic ensembles

The large-scale spatiotemporal variability in precipitation appears to be well-captured by the synthetic ensembles, even when the synthetic ensemble is based on a single record. We now turn to other common metrics of precipitation variability to continue to evaluate the synthetic ensemble approach, and assess the limits of what we can learn from a single record. While there are innumerable metrics to summarize the statistics of precipitation, here we focus on three: high frequency (< 10 year) and low frequency (> 10 year) variability, and the magnitude of extreme (low probability) events. We discuss our findings about low frequency variability in detail below, and refer the reader to the supplement for the other metrics.

Low-frequency variability is calculated as the interannual variance in DJF-average precipitation after using a low-pass forward and backward digital Butterworth filter with a frequency cutoff of 1/10 year⁻¹; the interannual variance calculation excludes the first and last five years to avoid edge effects. Within the context of the CESM1-LE, the best estimate of the low-frequency variability is the average across the low-pass interannual variance calculated for each ensemble member; the spread across the ensemble indicates the uncertainty in this quantity given an 85-year record. Low-frequency variability is maximized along the West Coast, and has a secondary maximum spanning from the Southeast up the Eastern Seaboard (Figure 6a). By definition, the average difference between the estimate of high-frequency variability in any given ensemble member and the ensemble mean estimate is zero (Figure 6b); the spread across ensemble members as measured by the standard deviation has a similar pattern but about a quarter of the magnitude of the ensemble mean esti-
mate (Figure 6c; note the different colorbar scale compared to panel a). To assess how well
the estimate of variability from a single ensemble member – analogous to having a single
observational record – matches the ensemble mean estimate of variability, we show the differ-
ence between the high-frequency variability estimated using a single representative ensemble
member and the ensemble mean in Figure 6d. This single ensemble member shows regions
of both over- and under-estimation of the high-frequency variability as expected, with the
largest differences along the West Coast, a region of high variability.

The above analysis was focused on the CESM1-LE alone, but a similar approach can be
used to validate our synthetic ensemble methodology by asking whether the statistics in the
CESM1-synth-LE are comparable to those from the original CESM1-LE. For each of the 40-
members of the CESM1-synth-LE based on a single member of the CESM1-LE, we estimate
the low-frequency variability as the average across the interannual low-pass variance of the
40 members, giving us 40 estimates – each one based on single member of the CESM1-
LE – of the variability. We benchmark the variability in the CESM1-synth-LE against
the ensemble mean of the CESM1-LE, which is viewed as the ‘truth’. Encouragingly, the
CESM1-synth-LE has similar low-frequency variability to the original CESM1-LE (Figure
6e). While it does exhibit a small positive bias along the West Coast, the spread across
the ensemble is smaller than for the CESM1-LE. Further, the difference between the low-
frequency variability estimated using a single ensemble member from CESM1-LE and the
true low frequency variability is reduced by creating the synthetic ensemble (Figure 6f, g, h).
In other words, the synthetic ensemble produces a slightly biased simulation of low-frequency
variability, but reduces the variance in the estimates when given a single record.

The same conclusions – the synthetic ensemble successfully captures the spatial structures
of variability with a small bias but decreased variance – hold for other metrics, including the
high-frequency variability and the magnitude of the 150-year event (Supplementary Figures
6 and 7).

Since the synthetic ensemble approach allows us to simulate many ‘years’ of data, we also
estimate the magnitude of the 500-year event across the full CESM1-LE, which contains 3400
member-years of data, and compare it to estimates from the CESM1-synth-LE (Figure 7). While the biases and variance of the estimates from the CESM1-synth-LE are, as expected, higher than those for our other metrics, the synthetic ensemble estimate is nearly unbiased in the majority of the domain, indicating that the approach can be used to estimate the probability of very extreme events.

Having validated our estimates of the high- and low-frequency variability, and the 150- and 500-year events, for DJF precipitation in the CESM1-synth-LE, we present each metric based on the GPCC-synth-LE (Figure 8). In general, the GPCC-synth-LE exhibits substantially less variance and smaller extremes in most of the domain, with the exception of the Southeast, where it shows larger variance and extremes than the CESM1-based ensembles (compare the panels in Figure 8 to Supplementary Figure 6a, Figure 6a, Supplementary Figure 7a, and Figure 7a, respectively). Using the 500-year event and the gridbox containing Sacramento, CA as a specific example, the CESM1-LE simulates a value of 14.9 mm/day, and the CESM1-synth-LE simulates a 95% range of 12.5-17.9 mm/day, depending on which member of the CESM1-LE it was constructed from. In contrast, the GPCC-synth-LE suggests that the true magnitude of the event is much smaller, at 7.2 mm/day.

3.4 The relative contribution of the climate noise

All of the results thus far have been based on Eqn. (3), in which precipitation variability is a function of the dependence on the three coupled ocean-atmosphere modes, ENSO, PDO, and AMV, as well as the climate noise. What is the relative role of the climate noise versus the large-scale coupled ocean-atmosphere modes?

To answer this question, we produce ‘noise only’ versions of the CESM1-synth-LE and GPCC-synth-LE, in which the residual climate noise, $\epsilon_{i,t}$, is bootstrapped as before, but the mode time series are not randomized. Thus, the spread across the ensemble reflects different samples of the climate noise alone. We then calculate the ratio of high- and low-frequency
variance in the ‘noise only’ ensemble to that in the full ensemble, where the ensemble-mean is removed from the ‘noise only’ ensemble to remove the influence of the modes, whose time series are identical in all of the synthetic members. In both ensembles, over 90% of the high-frequency variability and over 80% of the low-frequency variability is due to residual climate noise at most gridboxes (Figure 9). While this result may seem counterintuitive given the minimal year-to-year memory in the climate noise, it results from the fact that even white noise has power across all frequencies, and the contribution of the unpredictable atmospheric circulation to precipitation variance is large (see also Dong et al., 2018).

3.5 Validation of 50-year trends

The prior analyses have demonstrated that the variability in the CESM1-synth-LE closely matches the original CESM1-LE. Our final validation focuses on the across-ensemble ranges of 50-year (1956-2005) trends in DJF precipitation in each ensemble. Given that the synthetic ensembles have, by design, a forced trend of zero, how do their trends compare to the CESM1-LE, which does contain a forced component?

That the influence of random sampling of internal variability dominates the forced climate signal for recent 50-year trends is confirmed by comparing the 5th and 95th percentiles of 50-year trends in DJF precipitation (calculated independently at each grid box) in the CESM1-LE (Figure 10a, b) to those in the CESM1-synth-LE (Figure 10c-f). At all locations, the 5th percentile of trends is negative, while the 95th percentile is positive and of roughly equal magnitude, demonstrating the small contribution of the forced component compared to the random sampling of internal variability. Thus, not surprisingly, the 5th and 95th percentile trends in the CESM1-synth-LE are very similar in polarity, spatial structure, and magnitude, as those in the CESM1-LE. The slightly larger amplitude in the synthetic ensemble compared to the original ensemble (Figure 10g, h) is due to the small, but non-zero, forced trend, which is deterministic, being mapped being mapped onto the random internal variability in the synthetic ensemble. The 5th and 95th percentiles of the trends
in the GPCC-synth-LE tend to have smaller magnitudes (Figure 10i, j) than those in the CESM1-synth-LE, consistent with the model’s overestimation of variance in most regions overall (recall Section 3.2).

3.6 Challenges and opportunities of a synthetic ensemble approach

In sum, the evaluation of the CESM1-synth-LE and creation of the GPCC-synth-LE highlight both challenges and opportunities in using a synthetic ensemble approach. The signal-to-noise maps in Figure 2 demonstrate that, with an 85-year record, the impact of the coupled ocean-atmosphere modes on precipitation can be estimated well in regions where their influences are relatively large, but weaker teleconnections may be difficult to ascertain due to the sampling effects of independent internal variability. In contrast, it remains challenging to estimate the forced component of precipitation through standard methods, because it is small compared to the variability. Despite these estimation challenges, the CESM1-synth-LE produces variability that is very similar to that of the original CESM1-LE for two reasons. First, signals that are harder to estimate – the response to the AMV or climate change – are also smaller, so contribute less to estimations of variability. Second, a large fraction of both high- and low-frequency precipitation variability is contained in the climate noise term, whose structure remains largely intact through our block bootstrap resampling methodology. The GPCC-synth-LE exhibits spatiotemporal variability that is distinct from the CESM1-based ensembles, with generally less variance overall, particularly in the Western United States. These differences underscore the utility and importance of producing observationally-based ensembles. Like CESM1-synth-LE, DJF precipitation variance in GPCC-synth-LE is dominated by the climate noise term across a broad range of timescales, with implications for predictability.
4 Application to Western US water resources

Precipitation in the Upper Colorado River Basin (see outline in Figure 10) supplies water to Arizona, California, Colorado, Nevada, New Mexico, Utah, and Wyoming, and can vary substantially from year-to-year based on the particular sampling of internal variability (e.g. Harding et al., 2012). Here, we use the GPCC-synth-LE to examine the effects of sampling uncertainty on the characteristics of precipitation variability in the Upper Colorado River Basin; a validation of the synthetic ensemble approach for this region using the CESM1-LE as a testbed is given in Supplementary Figure 8.

The distribution of DJF precipitation in the Upper Colorado River Basin in the GPCC-synth-LE is similar to that of the actual GPCC record (Figure 11a), although it is clear that the 96-year time series provides incomplete sampling of the distribution, especially at the upper tail. This can be seen clearly in Figure 11b, which shows the 1st to 99th percentiles in actual observed GPCC DJF precipitation against the same percentiles in different members of the GPCC-synth-LE. For the highest quantiles, the GPCC-synth-LE suggests that the range approaches of 0.5 mm/day, meaning that the wettest winter seasons could range in their total precipitation by around 5 cm due to sampling of internal variability alone.

Motivated by prior work on Western water resources (Swain et al., 2018; Persad et al., 2020), we focus on three metrics for the interannual variability in DJF precipitation: the frequency of dry years, the frequency of wet years, and the frequency of pairs of ‘whiplash’ years that alternate from wet-to-dry or dry-to-wet. Wet years are defined as those with precipitation greater than the 80th percentile across the full GPCC record, and dry years are defined as those with precipitation less than the 20th percentile. Frequency is calculated empirically for moving 31-year periods, the standard period used to calculate climate normals. The cutoff for dry (20th percentile) and wet (80th percentile) years are 0.51 mm/day and 0.85 mm/day. By design, the ensemble mean and ensemble spread of each metric over time is constant, reflective of the assumption of no forced trend and stationary variability.
The observed dry, wet, and whiplash frequencies all fall well within the GPCC-synth-LE ensemble during all 31-year periods (Figure 11c, e, g). However, it is clear from comparison to individual members of the GPCC-synth-LE that the variability and secular change in the water resource metrics over the last century could have been substantially greater than what actually occurred (e.g. member 1 shown in Figure 11c-h). In general, the 95% range across the ensemble shows that a given 31-period could have a frequency of dry or wet years ranging from less than 10% to more than 35%, and whiplash years from 0% to nearly 20%. As a result, the change in dry, wet, and whiplash year frequency from the beginning to the end of the record ranges from decreases in excess of 20% to increases in excess of 20%. Even without any climate change signal, Upper Colorado River Basin precipitation has the potential to vary dramatically from one 31-year period to the next, and proper quantification of this variability is necessary for water resource planning.

5 Discussion and conclusions

Understanding regional precipitation variability and trends is critical for the purposes of planning for stable water supplies and floodwater infrastructure. However, precipitation exhibits substantial internal variability that impacts our ability to estimate its statistics and infer the significance and attribution of trends even given an approximately century-long record. In this work, we presented a statistical model that can be used to create a synthetic ensemble of precipitation fields whose spatial and temporal characteristics are consistent with those in the actual observational record in a statistical sense, but whose chronologies differ due to random sampling of internal variability. Such a synthetic “observational large ensemble” can be analyzed in a manner analogous to that of climate model large ensembles to quantify uncertainties related to sampling of internal variability. An effectively unlimited number of ensemble members can be produced with minimal computation time, although there are finite meaningfully different combinations of internal variability.
Testing of the synthetic ensemble methodology using a climate model large ensemble, the CESM1-LE, highlighted the challenge of inferring a forced component in DJF regional precipitation over the last century using a single climate model simulation or observational record, because it remains small compared to the magnitude of internal variability. In the context of the synthetic ensembles, we thus chose to remove the forced component from the statistical model, and produce synthetic ensembles whose variability is a function only of the precipitation response to three large-scale coupled ocean-atmosphere modes (ENSO, PDO, and AMV), and the residual atmospherically-driven climate noise. Nevertheless, the dominant spatiotemporal patterns of variability and the temporal statistics of precipitation in the CESM1-based synthetic ensemble, the CESM1-synth-LE, remained nearly identical to those in the original CESM1-LE. Further, the CESM1-synth-LE closely reproduces the 5th and 95th percentiles of 50-year trends in DJF precipitation simulated by the CESM1-LE, despite the synthetic ensemble having no forced component, due to the much greater importance of internal variability over the forced component at the gridbox scale even for historical multidecadal trends.

The relative contribution of the ocean-atmosphere modes versus the climate noise was also explored using ‘noise only’ versions of the synthetic ensemble, within which all across-ensemble variability was due to resampling of the climate noise only. The large majority of both high- and low-frequency variability was shown to be contained in the climate noise term, rather than linked to one of the three modes.

While our creation of the CESM1-synth-LE was primarily for the purposes of validating the synthetic ensemble methodology, synthetic ensembles could also be produced in order to expand the number of members in a climate model large ensemble. This hybrid approach of expansion of climate model large ensembles with statistical methods may be particularly useful for analyses of extreme events, which may be under-sampled even in an ensemble with tens or hundreds of members. Importantly, the synthetic ensembles are primarily focused on emulation of internal variability; the original climate model large ensembles remain necessary
for estimating the forced trend.

When applied to the observations, the synthetic ensemble offers the benefit of more closely capturing the observed structures of variability than may be possible in a global climate model. For example, the CESM1-based ensembles tend to be overly variable in the western half of the domain as compared to the GPCC-synth-LE, resulting in simulation of extremes that are substantially larger than is realistic (e.g. compare Figures 7a and 8d).

Motivated by this finding, we analyzed additional metrics of precipitation variability for the Upper Colorado River Basin, a region of considerable importance to Western US water resources. Despite the smaller variability in GPCC compared to CESM1, there is still a wide range of uncertainty in 31-year “climatologies” of dry, wet and whiplash years due to random sampling of internal variability alone. Any given 31-year period could have a dry or wet year frequency ranging from 10% to 35% (95% range), and a frequency of whiplash years from 0% to 20%, with the shifts between 31-year averages being largely unpredictable. Further, individual ensemble members can show nearly-monotonic trends in any of the water resource metrics (e.g. Figure 11c, e) due to sampling of internal variability alone.

We next turn to discussion of key assumptions and limitations that should be considered. First, we validated our methodology using CESM1 alone. If CESM1 were to exhibit a vastly different structure of variability than the observations, or a substantially different division between the climate change signal and the internal variability (c.f. the signal-to-noise paradox in climate models, Scaife and Smith, 2018), the validation reported here might not apply to the observational record. In particular, the synthetic ensemble approach will tend to perform less well when there is more low-frequency variability, since it is challenging to estimate this variability from a short record. While there is some evidence that precipitation in the observations has a longer memory than in CESM1— for example, we identify a longer block length for our GPCC-synth-LE than for the CESM1-synth-LE (four years versus two years) – precipitation variability is dominated by high-frequency variations in both the observations and CESM1, so we expect that the CESM1 validation should be suf-
iciently accurate. To perform a more complete validation, one could also produce synthetic
ensembles for all existing large ensembles, as well as validate the observationally-based en-
semble against the single observational record using ensemble forecast verification metrics
such as rank histograms (Suarez-Gutierrez et al., 2020). In addition, it is important to rec-
ognize that there are inherent and unavoidable limitations in validating the methodology as
applied to the observations, because there is no true ensemble of observations to compare
with. Specifically, the estimation of our model parameters themselves will be influenced by
the particular sampling of internal variability that we have observed, and the variability
produced by the synthetic ensemble will not perfectly reproduce the ‘true’ (but unobserved)
variability in the climate system because it is based on a limited record (see Figure 6 and
Supplementary Figures 6 and 7). Nevertheless, our validation of the approach within the
CESM1-LE suggests that, even when the synthetic ensemble is based on a single record, it
still performs well in simulating the full ensemble spread.

Second, we have assumed ergodicity, i.e. that information about the temporal evolution
of the climate system can be used to create an ensemble, as well as stationarity of the vari-
ability and teleconnections over the historical record. Due to the magnitude of interannual
precipitation variability, changes in its variance (Pendergrass et al., 2017) or in the structure
of teleconnections (Van Oldenborgh and Burgers, 2005) are typically not detectable within
the observational record. However, inasmuch as there is reason to believe that precipitation
variability and/or teleconnections will change in a future climate, it would be necessary to
modify the approach before applying it to future projections. For example, multiple climate
model large ensembles project an increase in Upper Colorado River Basin precipitation vari-
ability over the 21st century, despite a lack of agreement about any change in the mean (see
Figure 2 in Deser et al., 2020).

Third, we have assumed that the low-frequency variability in precipitation can be summa-
rized by a linear relationship with the three dominant modes of the climate system, ENSO,
PDO, and AMV, and that the modes behave independently of each other (after orthogonal-
izing the PDO time series with respect to the ENSO time series). Further, we have assumed that the statistics of the modes are stationary in the observational record. Similar to our discussion of precipitation variability and teleconnections above, changes in the modes are not yet detectable in the observational record, and models differ on how they may change in the future (Maher et al., 2018; Beobide-Arsuaga et al., 2021). Nevertheless, inasmuch as it is supported by the observations and/or theory, it would be advantageous to create a more sophisticated model for the coupled ocean-atmosphere modes, including the potential for a change in their statistics in the future. While future modeling work, especially focused on specific regions, may want to incorporate additional predictors, increases in complexity should be justified by significant increases in skill via, e.g. the use of information criteria.

In sum, we have expanded upon prior work building “observational large ensembles” to focus on precipitation variability in North America. We have found that, given an approximately century-long record, the synthetic ensemble can reproduce the variability in important precipitation statistics, including those related to water resources. While our analysis in this and prior work has focused only on temperature, precipitation, and sea level pressure, most physical quantities in the climate system also exhibit substantial spatiotemporal variability that challenges our ability to completely characterize the internal variability in a short observational record. This characterization is key for not only the water resource applications we focused on here, but also for detection and attribution of trends (e.g. Hegerl et al., 1996) and validation of climate models (Deser et al., 2020).

Acknowledgements

The authors thank the four anonymous reviewer for their helpful feedback. C.D. was supported by the National Center for Atmospheric Research (NCAR), which is a major facility sponsored by the National Science Foundation under the Cooperative Agreement 1852977. Analysis was performed on the Casper cluster supported by by NCAR’s Computational and
Data and code availability

Simulations from the NCAR CESM1-LE are publicly available via https://www.cesm.ucar.edu/projects/community-projects/LENS/. GPCC precipitation is publicly available at https://psl.noaa.gov/data/gridded/data.gpcc.html. Code to create the synthetic ensembles is publicly available at https://github.com/karenamckinnon/observational_large_ensemble/. A subset of the GPCC-synth-LE will be made publicly available upon acceptance.
References


Dong, L., L. R. Leung, F. Song, and J. Lu (2018), Roles of SST versus internal atmospheric variability in winter extreme precipitation variability along the US West Coast, *Journal of Climate*, 31(19), 8039–8058.


Maher, N., D. Matei, S. Milinski, and J. Marotzke (2018), ENSO change in climate pro-


Newman, M., M. A. Alexander, T. R. Ault, K. M. Cobb, C. Deser, E. Di Lorenzo, N. J. Man-


Figures
Figure 1: The relative importance of the residual versus the modes in precipitation variability. (a) The ratio of the interannual variance in DJF precipitation in the residual term to that in the residual and modes for CESM1-LE. Variances are calculated for each member and averaged. (b) As in (a), but for GPCC. (c) The time series of transformed January precipitation (gray) from GPCC at a gridbox in the Upper Colorado River Basin (38.5N, 109.5W), the estimated contribution of the modes (blue), a single synthetic time series (orange) and the contribution to that time series from the modes (green). (d) The predicted (transformed) precipitation from the mean value and modes (horizontal axis, blue line) versus the actual transformed precipitation (vertical axis). The scatter around the line is the residual, $\epsilon$. (e, f) As in (c, d) but for a gridbox in Georgia (31.5N, 83.5W) where the relative contribution of the modes is greater. The spatial pattern of the precipitation anomalies in years indicated by purple dots (1972 and 1991) in panels (c) and (e) are shown in Figure 4a, c, respectively.
Figure 2: The mean value (left column), standard deviation (middle column), and signal-to-noise (mean divided by standard deviation, right column) of the DJF model parameters from Eqn. (1) estimated using each member of the CESM1-LE. From top to bottom, parameters are $\beta_{i,\text{DJF}}$, $\beta_{i,\text{ENSO}}$, $\beta_{i,\text{PDO}}$, and $\beta_{i,\text{AMV}}$. Recall that precipitation is transformed via a Box-Cox power transformation before model fitting, so the parameters cannot be interpreted in terms of standard precipitation units. Model parameters are estimated for each month separately, and then averaged to produce the seasonal-average maps.

Figure 3: The DJF-average $\beta$ coefficient maps from fitting Eqn. (2) to the GPCC data.
Figure 4: Example precipitation fields from the GPCC dataset and the GPCC-synth-LE. (a) The observed precipitation anomaly in January 1972, which was relatively dry in the Upper Colorado River Basin. (b) The precipitation anomaly from the first member of the GPCC-synth-LE in January 1972. (c) The observed precipitation anomaly in January 1991, which was relatively wet in Georgia. (d) The precipitation anomaly from the first member of the GPCC-synth-LE in January 1991. The colorbar is nonlinear.
Figure 5: The total variance and first four empirical orthogonal functions (EOFs) of DJF precipitation in each ensemble: CESM1-LE, CESM1-synth-LE, and GPCC-synth-LE. (a)-(c) The total variance across years and members for each ensemble. (d)-(o) The spatial pattern associated with the first four EOFs from each ensemble, where EOFs are taken across years and members. The ensemble name, EOF number, and variance explained by the EOF is in the associated subtitle. The CESM1-synth-LE EOFs are based on the 40 synthetic ensemble members produced from the first member of the CESM1-LE alone, and the GPCC-synth-LE EOFs are similarly based on the first 40 members of that ensemble. All color scales are nonlinear.
Figure 6: Low-frequency variability (frequency < 1/10y) of DJF precipitation in the CESM1-LE and the CESM1-synth-LE. (a) The ensemble mean of the CESM1-LE; (b) the mean bias across ensemble members, calculated as the mean difference between each ensemble member and the CESM1-LE ensemble mean; (c) the standard deviation across ensemble members of the difference between each ensemble member and the CESM1-LE ensemble mean; and (d) the difference between the ensemble member with the average bias and the ensemble mean. Panels (e)-(h) show the same metrics for the CESM1-synth-LE; bias is still assessed against the CESM1-LE ensemble mean shown in panel (a). In all panels, the number in the lower righthand corner is the median value across the gridboxes shown. Note that the color scale in all panels is nonlinear, and that the range in the bottom three rows is one quarter of that in the top row.
Figure 7: The magnitude of the 500-year event in DJF precipitation. (a) The ‘true’ estimate from the full CESM1-LE; (b) the mean bias across ensemble members in the CESM1-synth-LE; (c) the standard deviation of the bias across ensemble members in the CESM1-synth-LE; and (d) the difference between the CESM1-synth-LE ensemble member with the average bias and the ‘truth’ from the CESM1-LE. In all panels, the number in the lower righthand corner is the median value across the gridboxes shown. Note that the color scale in all panels is nonlinear, and that the range in (b)-(d) is one quarter of that in (a).
Figure 8: Maps of (a) high-frequency variability (frequency $> 1/10$y), (b) low-frequency variability (frequency $< 1/10$y), (c) the magnitude of the 150-year event, and (d) the magnitude of the 500-year event in DJF precipitation from the GPCC-synth-LE. In all panels, the number in the lower righthand corner is the median value across the gridboxes shown, and the color scales are nonlinear.
Figure 9: The fraction of the variance from the climate noise term. (a) The ratio of the high-frequency ($> 1/10 y$) variance in the ‘noise only’ CESM1-synth-LE to the high-frequency variance in the full CESM1-LE. (b) As in (a), but for low-frequency ($< 1/10 y$) variance. (c) As in (a), but for the GPCC-synth-LE. (d) As in (b), but for the GPCC-synth-LE. The variance is calculated after the ensemble mean of each ensemble is removed, since the ‘noise only’ ensembles share the same temporal evolution of the modes.
Figure 10: Maps of the 5th and 95th percentiles (at each grid box) in 50-year trends in DJF precipitation during 1956-2005 for (a)-(b) the CESM1-LE; (c)-(d) the CESM1-synth-LE based on the first member of the CESM1-LE; (e)-(f) the full CESM1-synth-LE; (g)-(h) the difference between the CESM1-synth-LE and CESM1-LE; and (i)-(j) the GPCC-synth-LE. The black outline shows the Upper Colorado River Basin, whose areal-average precipitation statistics are shown in Figure 11.
Figure 11: Changes in water resource metrics in the GPCC-synth-LE due to sampling of internal variability. (a) The distribution of Upper Colorado River Basin (see outline in Figure 11) DJF precipitation in the GPCC observational record (blue) and the GPCC-synth-LE (gray). (b) A quantile-quantile plot showing the GPCC observations versus the GPCC-synth-LE. The gray dots show the quantiles in each member of the GPCC-synth-LE, and the blue dots are the average across members. The diagonal line is the one-to-one line. (c) The time series of the frequency of dry years in the GPCC observations (light blue line), the first member of the GPCC-synth-LE (gray line), and the ensemble mean of the GPCC-synth-LE (dark blue line). The light blue shading shows the 5%-95% range across the GPCC-synth-LE. (d) The change in the frequency of dry years from the first (1921-1951) to the last (1988-2018) 31-year period. The vertical lines show the change in the ensemble mean (dark blue), the observations (light blue) and the first member of the GPCC-synth-LE (gray). (e, f) As in (c, d) but for wet years. (g, h) As in (c, d) but for whiplash years.