

Internal Variability and Regional Climate Trends in an Observational Large Ensemble

KAREN A. MCKINNON AND CLARA DESER

Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder, Colorado

(Manuscript received 28 December 2017, in final form 10 May 2018)

ABSTRACT


Recent observed climate trends result from a combination of external radiative forcing and internally generated variability. To better contextualize these trends and forecast future ones, it is necessary to properly model the spatiotemporal properties of the internal variability. Here, a statistical model is developed for terrestrial temperature and precipitation, and global sea level pressure, based upon monthly gridded observational datasets that span 1921–2014. The model is used to generate a synthetic ensemble, each member of which has a unique sequence of internal variability but with statistical properties similar to the observational record. This synthetic ensemble is combined with estimates of the externally forced response from climate models to produce an observational large ensemble (OBS-LE). The 1000 members of the OBS-LE display considerable diversity in their 50-yr regional climate trends, indicative of the importance of internal variability on multidecadal time scales. For example, unforced atmospheric circulation trends associated with the northern annular mode can induce winter temperature trends over Eurasia that are comparable in magnitude to the forced trend over the past 50 years. Similarly, the contribution of internal variability to winter precipitation trends is large across most of the globe, leading to substantial regional uncertainties in the amplitude and, in some cases, the sign of the 50-yr trend. The OBS-LE provides a real-world counterpart to initial-condition model ensembles. The approach could be expanded to using paleo-proxy data to simulate longer-term variability.

1. Introduction

Internal variability is expected to be a prominent contributor to uncertainties in projections of regional climate in the coming decades (Räisänen 2001; Hawkins and Sutton 2011; Deser et al. 2012b, 2014; Thompson et al. 2015). The uncertainty emerges in part from the dominant influence of the chaotic atmospheric circulation on regional climate variability, which limits the useful time scale of initial condition forecasts and is associated with a substantial spread of multidecadal trends (Hawkins and Sutton 2009b; Deser et al. 2016; McKinnon et al. 2017). Improved knowledge of the range of possible regional climate trends that could occur because of sampling of different sequences of internal variability is important when developing

adaptation and mitigation strategies (Woodruff 2016), as well as for communicating the range of regional trends that are consistent with a climate change signal (Patt and Dessai 2005; Deser et al. 2012a; Hawkins et al. 2014).

One way to assess the contribution of internal variability to uncertainty in climate trends is through the use of ensembles of climate model simulations. The most straightforward type of ensemble for this purpose is an initial-condition ensemble, which uses a single climate model and forcing scenario, but slightly different initial conditions for each member of the ensemble (Collins and Allen 2002). Most initial condition ensembles are created through perturbing atmospheric initial conditions (e.g., Roeckner et al. 2003; Sterl et al. 2008; Deser et al. 2012b; Kay et al. 2015); these small perturbations quickly propagate such that each ensemble member experiences a different sequence of weather and climate fluctuations. Internal variability can also be assessed using multimodel ensembles (e.g., Tebaldi and Knutti 2007; Polvani and Smith 2013; Swart et al. 2015) such as the ensemble from phase 5 of the Coupled Model Intercomparison Project (CMIP5) (Taylor et al. 2012),

 Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-17-0901.s1>.

Corresponding author: Karen A. McKinnon, karen.mckinnon@post.harvard.edu

although in this case the spread across ensemble members is also due to differences in model physics and effective radiative forcing.

Analysis of model ensembles has shed light on the often surprisingly large influence of internal variability on various climate metrics including multidecadal trends in air temperature, precipitation, sea ice fraction, and sea level rise (e.g., [van Oldenborgh et al. 2013](#); [Hu and Deser 2013](#); [Wettstein and Deser 2014](#); [Swart et al. 2015](#); [Deser et al. 2016](#)), the midlatitude atmospheric response to El Niño–Southern Oscillation ([Deser et al. 2017](#)), the response of the climate system to volcanoes ([McGraw et al. 2016](#); [Lehner et al. 2016](#)), the time of emergence of anthropogenic warming ([Lehner et al. 2017](#)), and the evolution of global mean temperature ([Dai et al. 2015](#)), among many other examples. As such, it is increasingly common to assess climate models based on whether the observations lie within the spread of a model ensemble as has long been done for weather models ([Gneiting and Raftery 2005](#)). The spread across the ensemble is thus interpreted as indicative of the uncertainty in estimating various parameters from the limited observational record ([Deser et al. 2017](#)).

This approach, however, is only valid if the climate model ensemble accurately simulates the spatiotemporal covariance structure of the real world, which is generally not the case ([Ault et al. 2013](#); [Laepple and Huybers 2014](#); [Thompson et al. 2015](#); [McKinnon et al. 2017](#)). An alternative approach is to optimally utilize the observational record, with an eye toward being able to quantify the role of internal variability in various climate processes. One way to do so is to create a statistical model that can be used to generate alternate sequences of weather and climate variability that are consistent with the statistics of the observed record. The variability in the statistical model can then be treated in a manner analogous to the output from an initial condition climate model ensemble.

The choice of statistical model depends on the application at hand. In the simplest case, a time series model can be fit at every grid box ([Thompson et al. 2015](#)) or to a global-mean quantity ([Brown et al. 2015](#)). In this case, only the temporal variability is modeled, and no spatial information is retained. Others have developed more complex models that include spatial information, often via modeling the evolution of dominant patterns of variability ([Navarra et al. 1998](#); [Beltrán et al. 2012](#); [Salazar et al. 2016](#)), among other approaches. [McKinnon et al. \(2017\)](#) used a nonparametric block bootstrap method to model North American winter temperatures, such that the spatial covariance structure was easily retained. Here, we extend on that work by adding a dependence on dominant large-scale modes of

sea surface temperature (SST) variability because their low-frequency variability is not captured through our block bootstrapping methodology. We then apply the updated statistical model to all seasons, land regions, and additional climate variables.

While the focus of [McKinnon et al. \(2017\)](#) and this work is on 50-yr trends, the modeling approach could be used analogously to examine the influence of internal variability on trends over shorter or longer time scales, as well as on other quantities such as the uncertainty in climatologies based on a fixed period of record. The output from the statistical model can also be used to evaluate the representation of internal variability in comprehensive climate models. In this paper, we will assess the variability in past and future trends in temperature and precipitation over land, with a focus on boreal winter and links to the large-scale atmospheric circulation. We begin with a description of the observational datasets and climate model output used to construct our statistical model in [section 2](#), followed by derivation of the statistical model and its validation using the 40-member Community Earth System Model, version 1, Large Ensemble (CESM1-LE) in [section 3](#). The range of 50-yr trends simulated by the statistical model for both the past (1965–2014) and future (2015–64) is discussed in [sections 4–7](#), and the work is summarized in [section 8](#).

2. Data and model output

Temperature data are from the Berkeley Earth Surface Temperature (BEST) dataset at 1° resolution; the dataset is created through spatiotemporal interpolation (kriging) of in situ station measurements ([Rohde et al. 2013](#)). Precipitation data are from the Global Precipitation Climatology Centre (GPCC) at 1° resolution, and are also based on in situ gauge measurements ([Schneider et al. 2017](#)). Sea level pressure (SLP) data are at 2° resolution from the Twentieth Century Reanalysis version 2c, which only assimilates surface pressure observations and uses sea ice concentrations, sea surface temperatures, and time-varying radiative forcing as boundary conditions ([Compo et al. 2011](#)).

The statistical model makes use of monthly indices of three dominant modes of SST variability: Niño-3.4, which is an average of SSTs over 10°N–10°S, 170°–120°W, to represent El Niño–Southern Oscillation (ENSO; https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Data/nino34.long.data); the leading principal component time series of SSTs over the North Pacific to represent the Pacific decadal oscillation (PDO; <http://research.jisao.washington.edu/pdo/PDO.latest.txt>); and the average of SSTs over the North

Atlantic (0° – 80° N) to represent the Atlantic multidecadal oscillation (AMO; <https://www.esrl.noaa.gov/psd/data/correlation/amon.us.long.data>). We note that the global-mean SST has been subtracted prior to defining the PDO and AMO indices.

Ensembles of climate model simulations are used to both validate the methodology and provide an estimate of the forced response to climate change. The primary model ensemble we use is the CESM1-LE, a 40-member initial condition ensemble. Each member of the ensemble was created through small perturbations to the initial atmospheric temperatures on 1 January 1920 of a single parent simulation that begins in 1850 (Kay et al. 2015). These perturbations grow and propagate through the climate system, such that each member experiences a different sequence of internal variability. All simulations are driven by a common radiative forcing, namely the estimated historical forcing until 2005, and the representative concentration pathway 8.5 (RCP8.5) scenario thereafter (Lamarque et al. 2010, 2011; Meinshausen et al. 2011). The forced response is estimated as the mean across the 40 members of the CESM1-LE, under the assumption that 40 members is sufficient to average out internal variability. The time series of ENSO, PDO, and AMO for each model simulation are calculated using the Climate Variability and Diagnostic Package (Phillips et al. 2014). We will also draw on the CMIP5 ensemble (Taylor et al. 2012) as an alternative source of information for the forced response, estimated as the average across the 37 models that provide sufficient data for our analysis. In this calculation, only one ensemble member is taken from each model to avoid biasing the results toward the physics of any given model.

All variables are averaged from monthly values into three-month seasons (DJF, MAM, JJA, and SON). The period 1921–2014 is used to estimate the characteristics of the observed variability based on data and model output availability. The analysis focuses on 50-yr trends in the past (1965–2014) and future (2015–64) over land, excluding Antarctica where there are insufficient data. Unless noted otherwise, results for DJF are presented in the main paper and for JJA in the supplemental material; those for MAM and SON are available by request from the corresponding author.

3. Model and validation

We model temperature, precipitation, and SLP across space and time $X^{i,t}$, where the superscript i indicates spatial location, and the superscript t indicates time. Each variable is a linear combination of its mean state β_0^i , its response to anthropogenic influence $\beta_F^i F^t$, its response to large-scale SST modes $\beta_M^{i,m} M^{m,t}$, and residual variability $\varepsilon^{i,t}$. The

time series F^t represents the temporal structure of anthropogenic influence, and the three time series $M^{m,t}$ represent the evolution of ENSO, PDO, and AMO, where the superscript m indicates each of the three modes.

Combining these terms, we denote the spatiotemporal evolution of temperature, precipitation, and SLP as

$$X^{i,t} = \beta_0^i + \beta_F^i F^t + \sum_{m=1}^3 \beta_M^{i,m} M^{m,t} + \varepsilon^{i,t}. \quad (1)$$

All β coefficients are estimated using ordinary least squares regression on data from 1921 to 2014.

The time series of anthropogenic influence F^t is estimated using the Dai et al. (2015) method, although we have extended the approach to both precipitation and SLP. The method is based upon the assumption that the global-mean ensemble-mean time series of a given variable in a climate model can be used to represent the temporal structure of the true forced response. In our formulation, F^t is the global-mean ensemble-mean temperature, precipitation, or SLP time series calculated from the 40 members of the CESM1-LE, and β_F^i can be interpreted as the grid box-scale sensitivity to anthropogenic influence for each variable, assuming sufficient separability between the forced trend and low-frequency variability.

The $\beta_M^{i,m} M^{m,t}$ terms allow for the explicit modeling of the dependence of temperature, precipitation, and SLP on three large-scale modes of SST variability: ENSO, PDO, and AMO. Note that this term was not in the original formulation of McKinnon et al. (2017) but is included here to address the possible influence of low-frequency SST variability on terrestrial surface climate. Because of some collinearity between the modes, particularly ENSO and the PDO, $M^{m,t}$ is composed of three orthogonalized time series calculated from principal component (PC) analysis of the original observed ENSO, PDO, and AMO time series. Modes 1 and 3 collectively explain the ENSO and PDO time series, whereas mode 2 is almost perfectly correlated with the AMO. The SST pattern associated with each mode is calculated by regressing SSTs from the HadISST, version 1.1, dataset (Rayner et al. 2003) onto the orthogonalized mode time series (Fig. 1 for DJF; see Fig. S1 in the supplemental material for JJA). The orthogonalized modes are qualitatively similar to the first three empirical orthogonal functions (EOFs) of detrended global SSTs (cf. Fig. 4 in Messié and Chavez 2011).

The spatial pattern associated with mode 1 is typical of ENSO, with large anomalies in the eastern and central tropical Pacific, anomalies of the same sign in the eastern North Pacific, and anomalies of the opposite sign in the western and central North Pacific. The SSTs

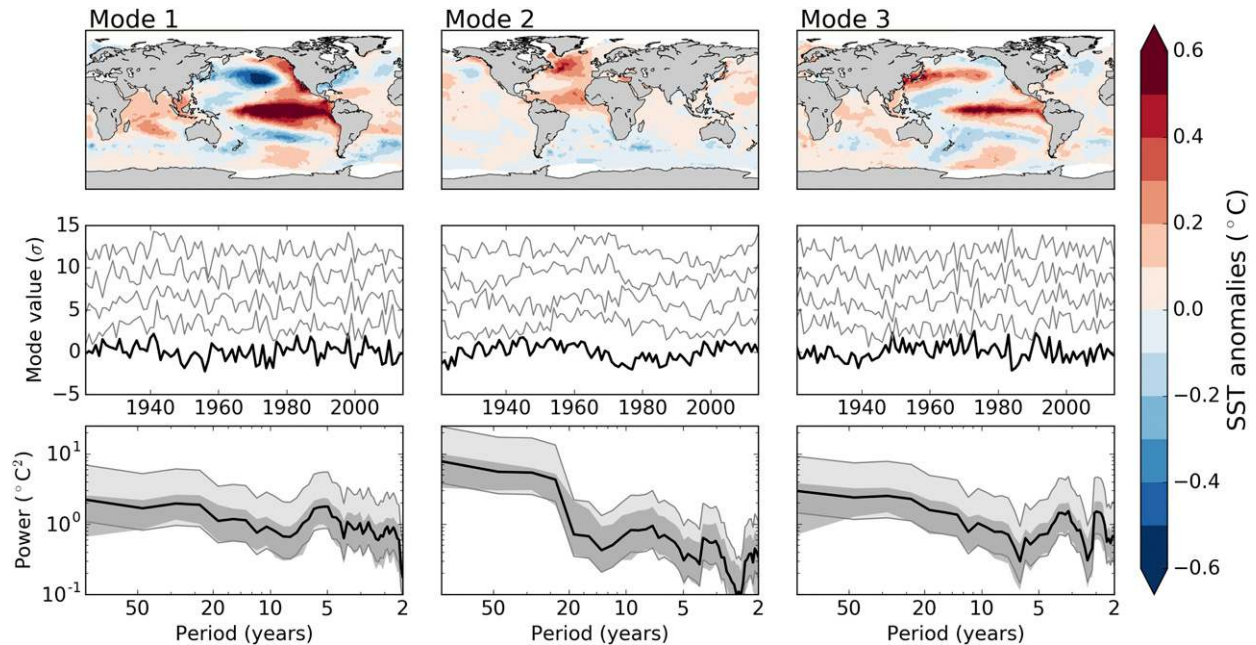


FIG. 1. The spatial pattern and temporal behavior of ENSO, PDO, and AMO, and their surrogates. Modes 1 and 3 collectively explain ENSO and PDO, while mode 2 represents the AMO. (top) The sea surface temperature pattern associated with each mode. (middle) The time series of each mode (thick black line), and four example surrogates (thin gray lines). Each surrogate is offset in the vertical by three units for visual clarity. (bottom) The power spectra of each mode (thick black line) and its 95% confidence interval (light gray shading, outlined by thin black lines). The 95% range of the spectra for the surrogate time series is shown in the dark gray shading. See text for details.

associated with mode 3 also maximize in the eastern and central tropical Pacific, although the meridional extent of the anomalies is smaller, and the western and central North Pacific anomalies are of the same sign. The pattern associated with mode 2 is typical of the AMO, with anomalies of opposite sign in the Atlantic sector of each hemisphere, and maximum anomalies in the subpolar North Atlantic. The spectra of all three modes are red, with the AMO showing the greatest increase in power with decreasing frequency (Fig. 1). The spectra of modes 1 and 3 both show an increase in power in the 2–7-yr band associated with ENSO. Analogous to the interpretation of $\hat{\beta}_F^i$, the $\hat{\beta}_M^i$ values estimated for each grid box are indicative of the local sensitivity to each SST mode for each variable.

The spatial structure of the regression coefficients relating each mode to DJF temperature over land is shown in Fig. 2 (see Fig. S2 in the supplemental material for JJA). There is a significant positive relationship between the first mode and temperature over Canada and Alaska as well as a significant negative relationship to temperature in the southeastern United States, consistent with the known influence of ENSO and the PDO (e.g., Trenberth et al. 1998; Mantua and Hare 2002). There are also weak but significant positive relationships over much of South America, Africa, and Australia. The

relationship between the other two modes and DJF temperature tends to be weaker, with smaller regions of significance. Mode 2 is associated with positive temperature anomalies in northeastern Canada, adjacent to the primary center of action of the AMO SST pattern. Mode 3 is associated with cooler Alaskan temperatures and warmer temperatures east of the Great Lakes, as well as warming around Mongolia. Given the larger and more significant regression coefficients associated with mode 1 compared to the latter two modes, the results would likely be qualitatively similar if it were the only mode included in the statistical model.

To further understand the primary controls on temperature variability in Eq. (1), we examine the power spectra of each term for three representative grid boxes chosen based on the strength of their relationship to the SST modes (Fig. 2). For all three locations, there is a clear time scale separation between the forced component $\beta_F^i F^i$ (red) and the original time series (blue) for periods shorter than approximately 30 yr; that is, the variability associated with the forced component is at least two orders of magnitude smaller than that of the original time series in this frequency band. At the grid box closest to Warsaw, Poland, which is representative of most locations, there is also only a minimal contribution to temperature variability from the sum of the

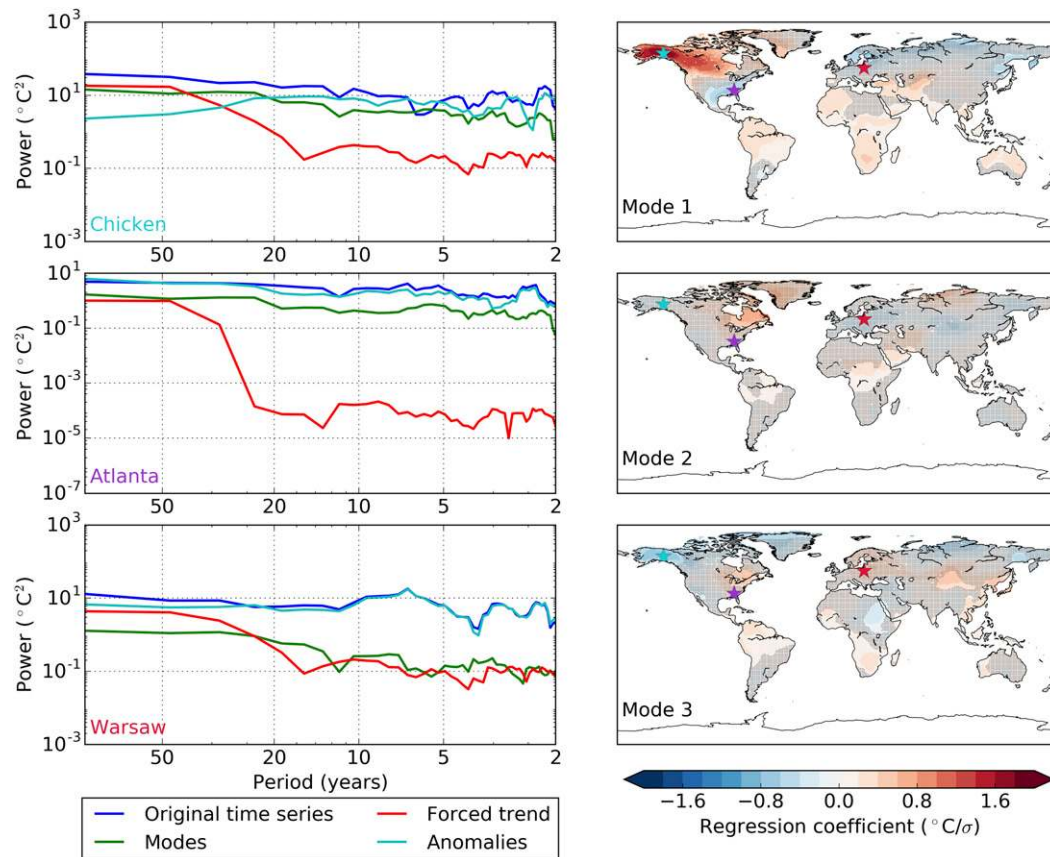


FIG. 2. Contribution of ENSO, PDO, and AMO to interannual DJF temperature variability estimated over 1921–2014. Modes 1 and 3 collectively explain ENSO and the PDO, while mode 2 represents the AMO. (right) The sensitivity of DJF temperature to the three dominant SST modes (see Fig. 1) estimated via linear regression. Stippling indicates an insignificant relationship using a false discovery rate of 10%; p values at each grid box are estimated using a two-sided Student's t test. (left) Power spectra for three representative grid boxes, indicated by the location of the stars in the maps. In all cases, there is a clear separation of time scales between the inferred forced trend (red) and the residual anomalies (teal). In a small number of locations, represented by the grid boxes closest to Chicken, Alaska (teal star), and Atlanta, Georgia (purple star), the contribution of the SST modes to temperature variability (green) is nontrivial. More typically, as represented by the grid box closest to Warsaw, Poland (red star), the fractional contribution of the SST modes to temperature variability is small.

three modes (green), such that the power spectrum of the anomalies $\varepsilon^{i,t}$ (teal) is quite similar to that of the full time series at periods shorter than 20–30 yr. At the grid boxes closest to Chicken, Alaska, and Atlanta, Georgia, which are representative of a small minority of locations, the contribution of the three SST modes to temperature variability is nonnegligible, indicating the importance of explicitly modeling this relationship.

Conceptually similar results are found for DJF precipitation (Fig. 3; see Fig. S3 in the supplemental material for JJA), although the regions where there are significant links to the modes are somewhat different and more numerous. However, even at the grid boxes closest to locations such as Indianapolis, Indiana, and Chihuahua, Mexico, where there are large and

significant regression coefficients, the contribution of the modes to interannual precipitation variability remains small, typically an order of magnitude less than the full time series (see power spectra in Fig. 3). As a result, the spectral behavior of the full time series $X^{i,t}$ and the residual variability $\varepsilon^{i,t}$ are very similar at periods less than 20 yr, beyond which the forced trend becomes important.

After accounting for the forced trend and three SST modes, the null hypothesis of white noise for the residual variability cannot be rejected based on the Ljung–Box test at the 5% level for 97% of the global land area for DJF temperature (95% for precipitation). Similar results are found for JJA (not shown). Thus, the residual anomalies can be interpreted as “climate noise”

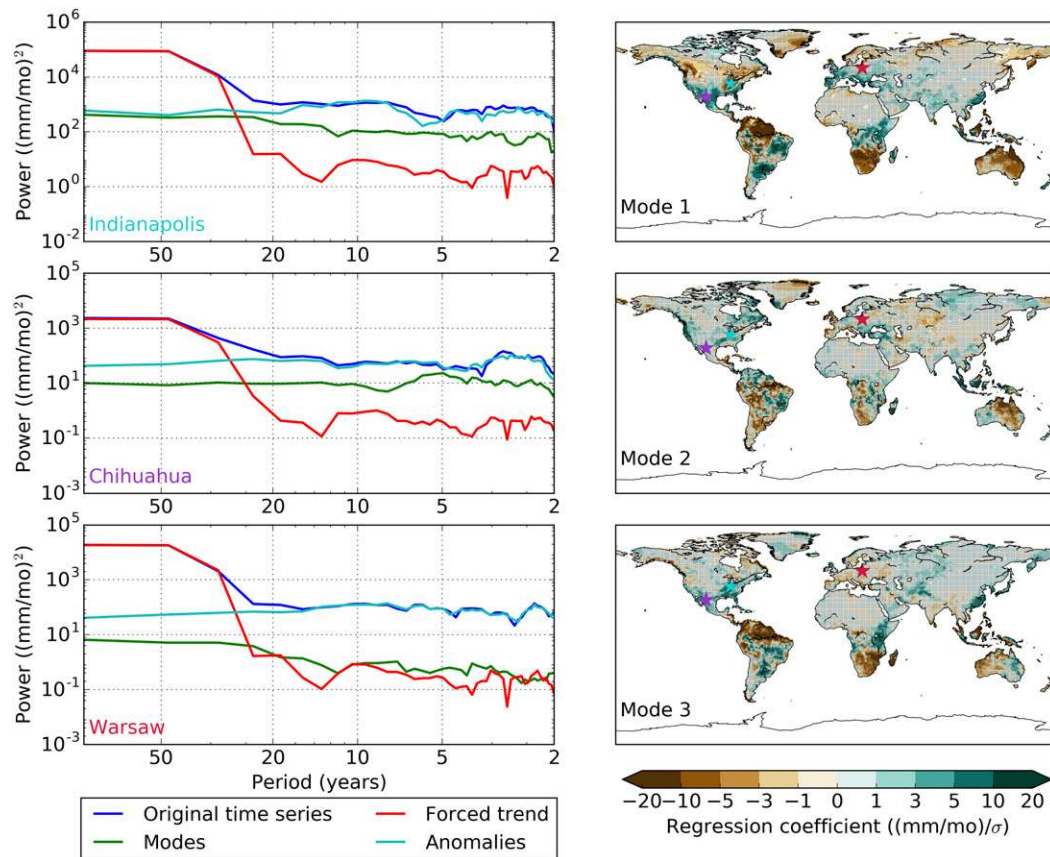


FIG. 3. As in Fig. 2, but for precipitation. Indianapolis, Indiana (teal star), and Chihuahua, Mexico (purple star), are representative of locations with a significant linear relationship between mode 1 and interannual precipitation variability, whereas Warsaw, Poland (red star), is again representative of locations with insignificant linear relationships with the mode time series.

(Madden 1981) resulting from sampling of weather variability, and their spatial structure contains information about the spatial covariance of this variability.

4. Simulating alternative realities

While Eq. (1) can be fit to the observed set of climate variables as illustrated in the previous section, it can also be used to generate a synthetic ensemble of unforced climate histories that could have been observed had a different sequence of internal variability unfolded. To do so, it is necessary to produce counterfactual versions of the time series of the SST modes and residual variability, which will require us to make a number of assumptions that we first examine for their credibility.

First, we assume that the SST modes and residual variability are stationary; that is, they have not shown a response to anthropogenic influence from 1921 to the present and will not in the next 50 years. ENSO and the PDO have not shown any robust changes in their

statistics over the historical record [Deser et al. 2012c; Newman et al. 2016; although Smith et al. (2016) have suggested a potential role of aerosol forcing], nor is there any consensus among climate models that they will change significantly over the next 50 years (Lapp et al. 2012; Chen et al. 2017). Larger changes may occur by the end of the twenty-first century, however. For example, Zhang and Delworth (2016) find that the PDO in their model weakens by approximately 20% and its period shortens from 20 to 12 yr in response to an abrupt doubling of CO_2 , and Zhou et al. (2014) find that, even without a change in ENSO properties, ENSO teleconnections over North America may shift by the end of the century. For now, we assume the simplest model of no change, and note that the methodology could be easily adapted given knowledge of the future statistical properties of either mode. The role of external forcing in the AMO is under active debate, with some studies suggesting a response to aerosol changes during the second half of the twentieth century (Otterå et al. 2010;

Booth et al. 2012; Murphy et al. 2017) and others indicating little sensitivity to anthropogenic radiative forcing (Knight 2009; Zhang et al. 2013). Thus our assumption that the mode statistics are stationary remains an important caveat in our study, although we note that modes 2 and 3 have a much smaller influence on terrestrial temperature and precipitation variability than internal atmospheric circulation anomalies on the time scales considered here (Figs. 2 and 3).

To the best of our knowledge, there is also not evidence for significant observed changes in unforced interannual variability of temperature, precipitation, and SLP. Note that this stands in contrast to *subseasonal* variability, which does appear to have decreased over high-latitude continents, possibly in association with Arctic amplification (Screen 2014; Rhines et al. 2017). To assess whether there may be significant changes in the future, we compare the unforced components of interannual variability simulated by the CESM1-LE during 1965–2014 and 2015–64, estimated as the residual from the ensemble mean. At almost all locations over land, there are no significant changes in interannual variability in either winter or summer (Fig. S4 in the supplemental material). The only exceptions are DJF precipitation over central Africa and the southern Arabian Peninsula and JJA precipitation southeast of the Mediterranean Sea.

Second, we assume that the SST modes do not have initial-value predictability. The presence of initial-value predictability would imply a reduction in the spread of the counterfactual versions of the time series of each mode at the beginning of the time period being modeled. The Fourier phase randomization method we employ, discussed below, will instead produce a set of time series whose spread is constant in time. This is likely a simplification, since there is some evidence that the PDO and AMO exhibit predictability out to about a decade (Hawkins and Sutton 2009a; Branstator et al. 2012; Ding et al. 2016), although others find that the time scales of predictability are more limited (Alexander et al. 2008; Newman et al. 2016). Given the current state of knowledge, we proceed under the assumption of no initial-condition predictability, and note that the model could be easily modified to incorporate predictability if identified at a later date. Furthermore, our focus is on 50-yr trends, well beyond the time scale of any proposed initial-condition predictability.

Given the sufficient credibility of our assumptions, we proceed to generate the synthetic ensemble. The approach taken for the component of variability related to the SST modes is distinct from the approach used for the residual variability because of their different temporal structures as outlined below.

To create alternative SST mode time series, we employ a surrogate data approach to produce an ensemble of time series that have the same mean, variance, and autocovariance of the original data but are otherwise random. The surrogate time series are produced by transforming the original data into the Fourier domain, multiplying its Fourier phases by uniformly distributed random phases, and then transforming back into the time domain (Theiler et al. 1992; Schreiber and Schmitz 2000). The power spectra of the resulting surrogates are largely within the 95% confidence interval of the observed spectrum [estimated as in Percival and Walden (1993)], although the power of the surrogates tend toward smaller values within the confidence interval, especially at low frequencies, likely related to the known whitening effect of the Fourier phase randomization procedure (Fig. 1). This issue could be ameliorated by applying various adaptive iterative methods (Schreiber and Schmitz 2000). By combining these surrogate time series of each mode with the regression coefficients estimated from the observed record $\hat{\beta}_M^{i,m}$, it is possible to produce spatiotemporal patterns of temperature, precipitation, and SLP anomalies that could have occurred given a different sampling of the ENSO, PDO, and AMO time series.

To generate surrogate sets of the residual variability $\varepsilon^{i,t}$ we take advantage of the minimal year-to-year memory in each of our climate variables after explicitly modeling the forced trend and dependence on large-scale SST modes, and perform a block bootstrap procedure. The observed values of the residual variability are grouped into time blocks of 2 yr, and the blocks are randomly resampled with replacement to produce surrogate sets of residual variability. By performing the block bootstrap in time only, the full spatial structure of the anomalies in temperature, precipitation, and SLP is retained. The block length is chosen to be 2 yr as a balance between having time blocks that are sufficiently large compared to the scale of temporal autocorrelation but retaining enough blocks to generate adequate variability between bootstrap samples. Note that, in the case of autocorrelated and short time series, a bootstrapping approach will not tend to satisfy both of these considerations entirely. First, although the residuals are largely indistinguishable from white noise at the 5% level (see end of section 3), weak interannual correlations may still exist. Second, our 94-yr record is unlikely to contain a complete sampling of the possible variability. Both of these effects will lead to small underestimates of the variability in the bootstrapped ensemble that will be quantified below [see McKinnon et al. (2017) for further discussion of the use of the block bootstrap in this context].

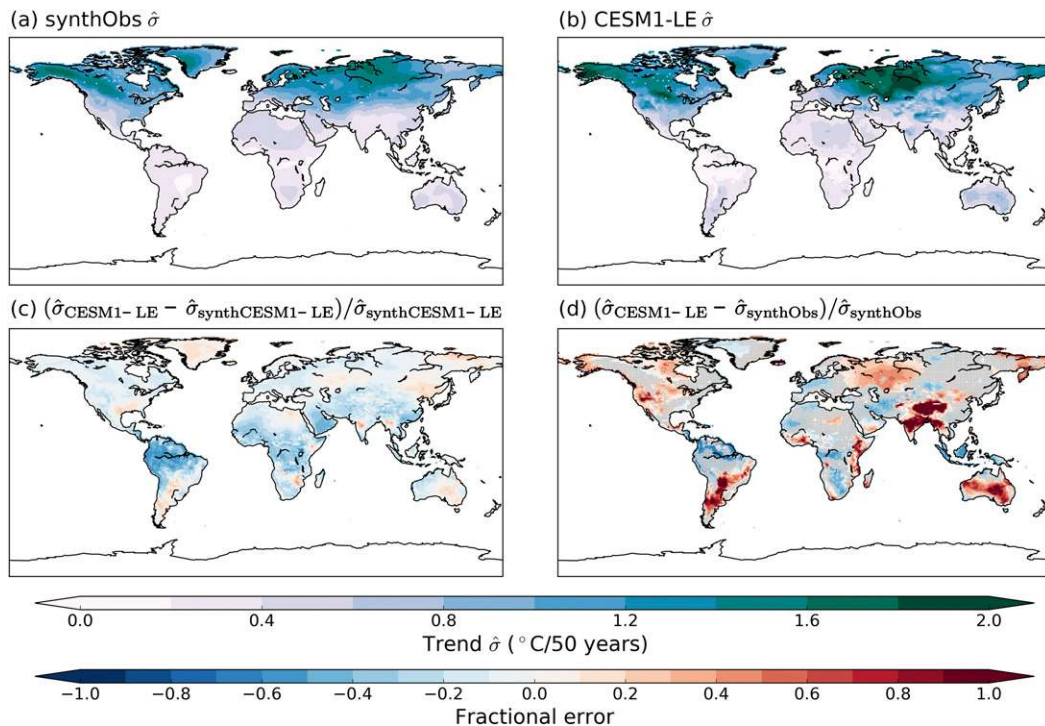


FIG. 4. Trend variability in the synthetic observational ensemble (synthObs), the NCAR CESM1-LE, and validation of the synthetic ensemble methodology for DJF temperature. (a) The standard deviation of 50-yr (1965–2014) linear temperature trends across the 1000 members of the synthObs. (b) The standard deviation of 50-yr linear temperature trends across the 40 members of the CESM1-LE. (c) The fractional error of the synthetic ensemble methodology, estimated as $(\hat{\sigma}_{\text{CESM1-LE}} - \hat{\sigma}_{\text{synthCESM1-LE}}) / \hat{\sigma}_{\text{synthCESM1-LE}}$. (d) The fractional difference between CESM1-LE- and observational-based estimates of the standard deviation of 50-yr linear temperature trends, that is, $[(a) - (b)] / (b)$. Stippling indicates locations where the difference is insignificant using a false discovery rate of 10%; p values are estimated at each grid box as the probability that the difference between the CESM1-LE and the synthetic observational ensemble could occur from applying the same methodology to individual members of the CESM1-LE.

Both the randomization of the mode time series and the block bootstrap are performed 1000 times to produce a synthetic observational ensemble of temperature, precipitation, and SLP data. Because the same surrogate SST mode time series and resampling of years are used for each variable in each ensemble member, it is possible to examine not only the internal variability of a single variable, but also the relationships between the internal variability of different variables.

As in McKinnon et al. (2017), the full process is validated within the context of the CESM1-LE itself. Ideally, the synthetic ensemble produced through the process outlined above would have variability consistent with a true initial condition ensemble. It is possible to test this conjecture by fitting Eq. (1) to individual members of the CESM1-LE, producing a synthetic ensemble based on a single member, and comparing the statistics of the synthetic ensemble to that of the true initial condition ensemble. Consistent with the remainder of our analysis, we compare the spread (as measured by the standard deviation σ) of 50-yr

(1965–2014) trends across the CESM1-LE-based synthetic ensemble $\hat{\sigma}_{\text{synthCESM1-LE}}$ with that simulated by the actual CESM1-LE $\hat{\sigma}_{\text{CESM1-LE}}$. Biases induced by the methodology are quantified by fractional error, calculated as $(\hat{\sigma}_{\text{CESM1-LE}} - \hat{\sigma}_{\text{synthCESM1-LE}}) / \hat{\sigma}_{\text{synthCESM1-LE}}$.

Maps of the fractional error are shown in Figs. 4c and 5c for DJF temperature and precipitation, respectively. Analogous maps for JJA are shown in Figs. S6 and S7 of the supplemental material. In general, the spread of trends is underestimated in the synthetic CESM1-LE as compared to the actual CESM1-LE, as expected based upon the use of the bootstrapping procedure. For both temperature and precipitation during DJF and JJA, the magnitude of the fractional error is less than 11% at the majority of grid boxes, and less than 27% for temperature (23% for precipitation) at 85% of locations. The largest errors (50%–60%) occur for DJF temperatures over Amazonia and JJA temperatures over central Canada. Fractional errors for SLP are generally <20% over the extratropics but reach 40%–50% over the

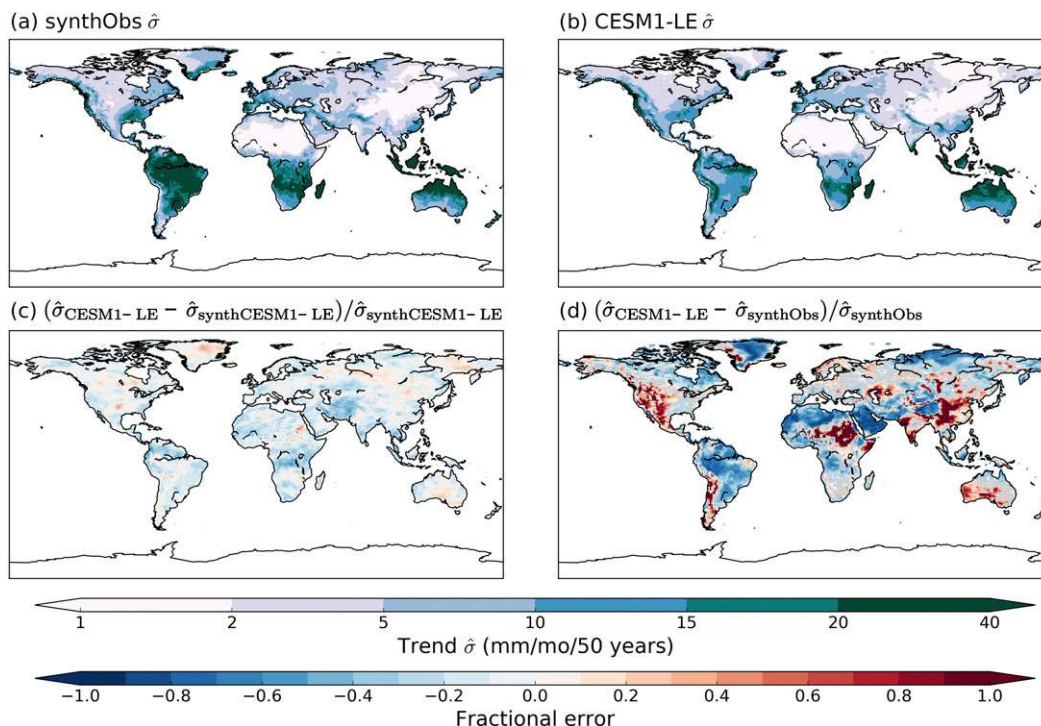


FIG. 5. As in Fig. 4, but for precipitation.

eastern equatorial Pacific and tropical Indian Ocean in both seasons (Figs. S5c and S8c). The large errors in SLP over the tropical oceans are unsurprising, since strong ocean–atmosphere coupling will tend to redden SLP variability in these regions; the red noise behavior appears to persist even after accounting for the three SST modes. Accordingly, the present study emphasizes the extratropics.

For context, the errors induced by the proposed methodology (Figs. 4c and 5c for DJF and Figs. S6 and S7 for JJA) are compared to those associated with using the CESM1-LE in place of the synthetic ensemble. The fractional difference between the two ensembles is calculated analogously as $(\hat{\sigma}_{\text{CESM1-LE}} - \hat{\sigma}_{\text{synthObs}}) / \hat{\sigma}_{\text{synthObs}}$, where $\hat{\sigma}_{\text{synthObs}}$ is the spread in 50-yr trends in the observationally based synthetic ensemble (Figs. 4d and 5d for DJF and Figs. S6 and S7 for JJA). The synthetic ensemble offers a large improvement over the CESM1-LE in simulating a realistic spread of 50-yr trends in temperature and precipitation over land. For example, the CESM1-LE overestimates the spread in 50-yr trends of DJF temperature (Fig. 4d) by more than 35% in the western United States (35° – 50° N, 110° – 120° W) and 45% in eastern Australia (10° – 40° S, 135° – 150° E), while the errors in the synthetic ensemble in the same areas (Fig. 4c) are 10% and 4%, respectively. Note that the North American biases in the CESM1-LE are slightly less than those found in

McKinnon et al. (2017) in regions affected by ENSO, highlighting the importance of including the modes in our statistical model. The biases in the CESM1-LE are even more significant for DJF precipitation (Fig. 5d), with an overestimate in the western United States of nearly 70% and an underestimate in the Amazon region of Brazil (0° – 15° S, 45° – 65° W) of over 50%, compared to errors of magnitude 9% and 11% in the synthetic ensemble (Fig. 5c). Many other regions also show biases in the CESM1-LE exceeding 50% or more for both quantities and seasons. For SLP, the biases in the synthetic ensemble are generally comparable to those in the CESM1-LE, except over parts of the northern continents in JJA where the CESM1-LE underestimates the variability in 50-yr trends by 30%–40% (cf. Figs. S5c,d and S8c,d).

As shown above, the statistical model offers an improvement over the CESM1-LE in its simulation of internal variability. However, unlike a true initial condition ensemble, it cannot inform about the forced component of climatic trends, defined as the “true” response to radiative forcing uncontaminated by sampling of internal variability. We thus replace the mean of the synthetic ensemble with an estimate of the forced component based on climate model ensembles. Our primary estimate of the forced component is the ensemble mean of the CESM1-LE, but we also make use of the ensemble mean of the 37 CMIP5 models. The forced

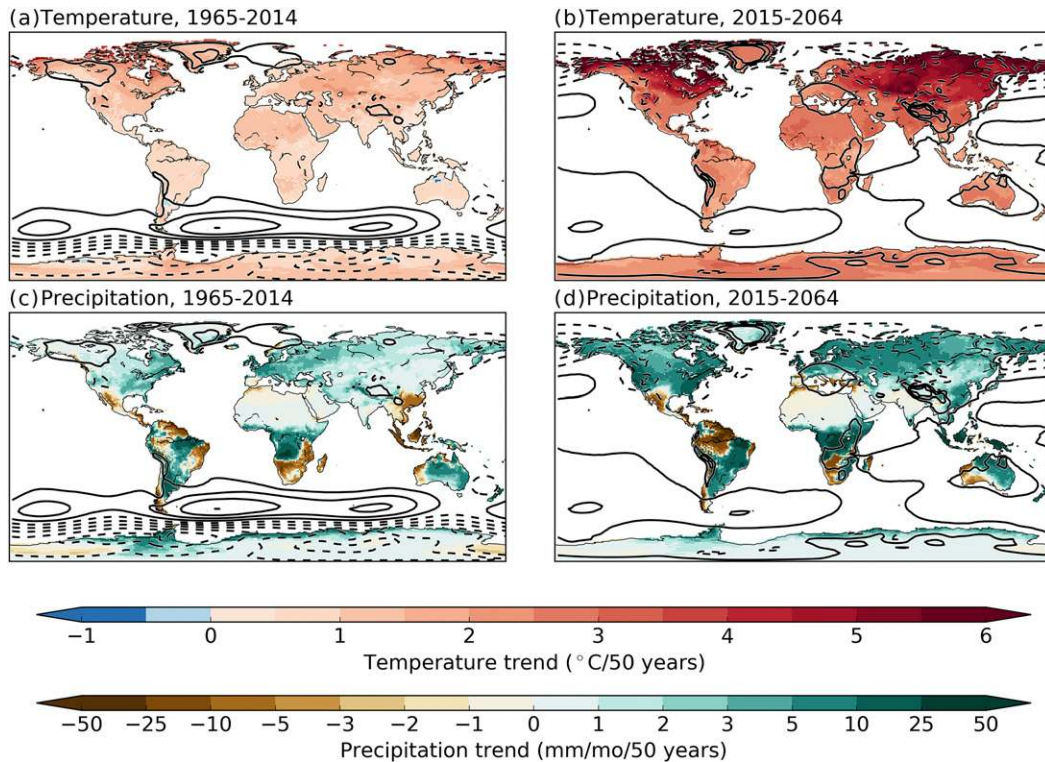


FIG. 6. Ensemble-mean linear trends from the NCAR CESM1-LE for DJF (a),(b) temperature and (c), (d) precipitation for (left) the past (1965–2014) and (right) the next (2015–64) 50 years. Trends in SLP are indicated by contours (dashed contours are negative), with a contour interval of $\pm 1 \text{ hPa } (50 \text{ yr})^{-1}$, beginning with contours of $\pm 0.5 \text{ hPa } (50 \text{ yr})^{-1}$.

component of terrestrial temperature and precipitation trends, in addition to global SLP trends, from 1965 to 2014 based on the CESM1-LE are shown in Fig. 6 for DJF (see Fig. S9 in the supplemental material for JJA), and those based on CMIP5 are in Figs. S10 and S11.

The forced trend in DJF temperature from 1965–2014 based on the CESM1-LE (Fig. 6a) is positive everywhere, with magnitudes generally between 1° and $2^\circ \text{C } (50 \text{ yr})^{-1}$ in the Northern Hemisphere (NH) but reaching $3^\circ \text{C } (50 \text{ yr})^{-1}$ at some Arctic coastal locations, and $<1^\circ \text{C } (50 \text{ yr})^{-1}$ in the Southern Hemisphere (SH). The forced trend in DJF precipitation (Fig. 6c) is generally positive north of 30°N , with maximum values [$5\text{--}10 \text{ mm month}^{-1} (50 \text{ yr})^{-1}$] over western Europe and parts of eastern North America. Remaining land areas show more regional patterns of forced precipitation trends, with pronounced wetting [$>10\text{--}25 \text{ mm month}^{-1} (50 \text{ yr})^{-1}$] throughout tropical southern Africa, much of Australia, and parts of Brazil and Argentina, and comparable amplitude drying in Mexico, the Maritime Continent, southern Africa, and northwestern Australia among other locations. Forced trends in SLP are generally small [magnitudes $<0.5 \text{ hPa } (50 \text{ yr})^{-1}$] with the exception of the SH extratropics in association with

stratospheric ozone depletion (Polvani et al. 2011). The forced trend from the CMIP5 ensemble is similar to that from the CESM1-LE, although some differences are evident such as greater warming at high northern latitudes and larger wetting over western Eurasia and much of North America (Fig. S10 in the supplemental material).

It would also be possible to center the ensemble on observationally based estimates of the forced trend calculated using dynamical adjustment (Smoliak et al. 2015; Deser et al. 2016; Lehner et al. 2017) or statistical methods such as ensemble empirical mode decomposition (Lee and Ouarda 2012), but we do not pursue these directions in the current work. The synthetic ensemble recentered on the forced component from the CESM1-LE is termed the observational large ensemble (OBS-LE).

5. Hemispheric-scale variability in the observational large ensemble

What is the range of possible 50-yr trends that could arise from observed unforced internal variability? To begin to answer this question, we first identify the

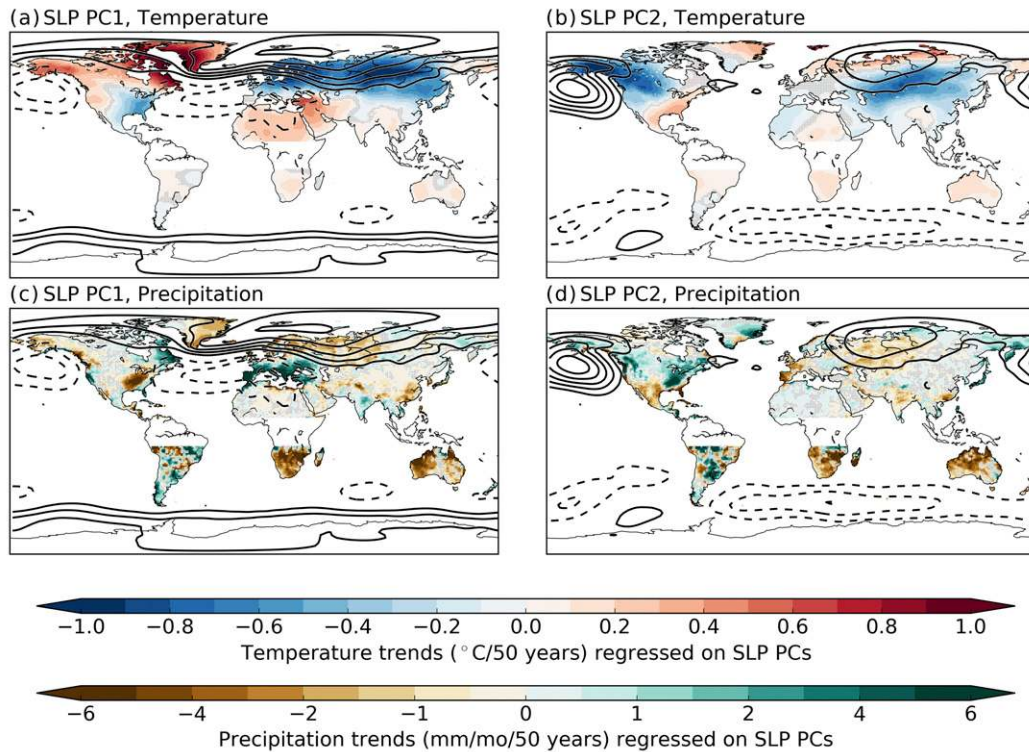


FIG. 7. The regression maps of (a),(b) temperature and (c),(d) precipitation and SLP (contours in all panels) onto the first two normalized PCs of 50-yr (1965–2014) DJF SLP trends across the OBS-LE. Contour interval is ± 0.5 hPa $(50 \text{ yr})^{-1}$ (dashed contours are negative), and the zero contour is suppressed. SLP PCs are calculated separately for each hemisphere (poleward of 10°) but are shown on the same map for conciseness. The fraction of variance explained by the first two PCs is 30% and 18% in the NH (34% and 21% in the SH), respectively.

dominant patterns of variability in 50-yr trends (1965–2014) of SLP across the 1000 members of the OBS-LE, cognizant of the large influence of circulation on temperature and precipitation in the extratropics, especially during winter. The patterns are identified by applying an EOF analysis to the set of 1000 SLP trend maps from the OBS-LE for each hemisphere separately [poleward of 10° in either hemisphere; see Deser et al. (2012b) for a similar calculation]. The associated temperature and precipitation trends are created by regressing the 1000 temperature and precipitation trends at each grid box onto the normalized principal components of the SLP trend EOFs in their respective hemispheres.

The leading SLP trend EOF in the NH, which accounts for 30% of the variance of trends, shows a dipole pattern between the high and middle latitudes that closely resembles the negative polarity of the northern annular mode (NAM); note that the sign of the EOF is arbitrary (Figs. 7a,c). Similarly, the first EOF in the SH captures much of the structure of the southern annular mode (SAM), although the midlatitude anomalies are weaker than typical of the interannual SAM pattern

(Thompson and Wallace 2000). It is well known that interannual variability in the annular modes has an important influence on temperature and precipitation (e.g., Thompson and Wallace 2000, 2001; Hendon et al. 2007), and the same is true for the variability in trends as shown in Fig. 7. In particular, NH SLP trend EOF1 is associated with cooling across most of Eurasia, and a dry–wet precipitation dipole between southern and northern Europe. Other surface climate impacts include cooling over the southeastern United States, warming in Canada, Alaska, Greenland, the western United States, North Africa, and the Middle East, and drying of the eastern United States. Similar NH patterns in SLP, temperature, and precipitation are produced when the SLP trend EOF is calculated using only the Atlantic–Eurasian sector (not shown). The SH SLP trend EOF1 shows climate impacts typical of a negative SAM, including drying of a large fraction of Australia and southern Africa, and wetting over parts of South America.

The second NH SLP trend EOF, which accounts for 18% of the SLP trend variance, resembles the Pacific–North American (PNA) pattern whose primary center of action is located over the North Pacific; an additional

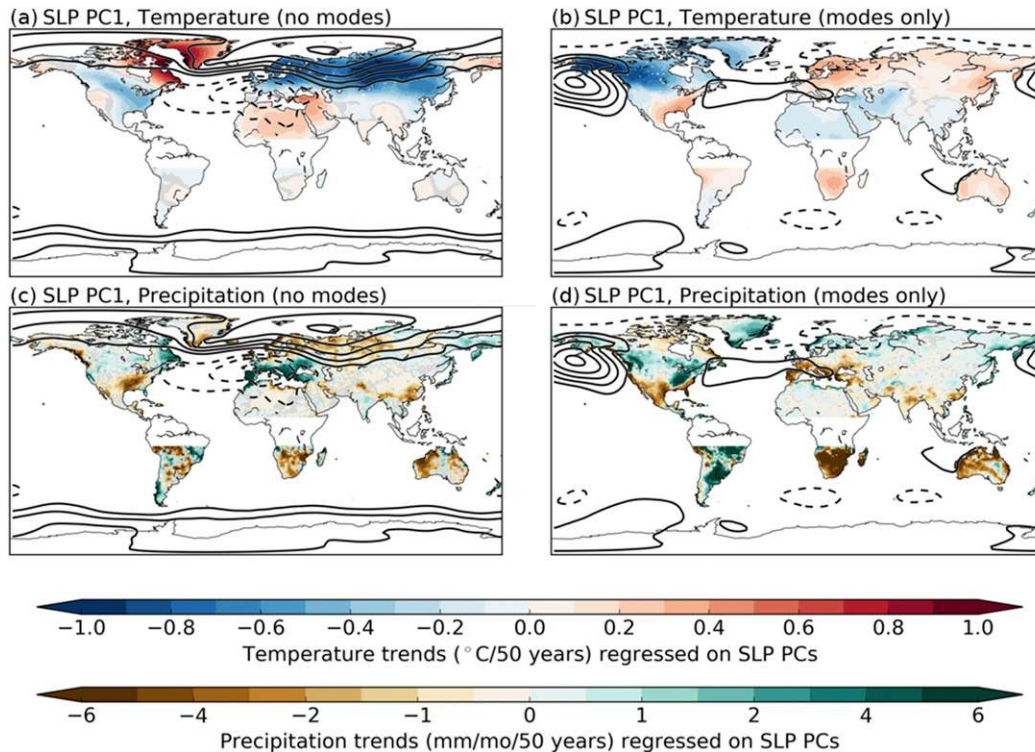


FIG. 8. The regression maps of (a),(b) temperature and (c),(d) precipitation and SLP (contours in all panels) onto the PC1 of 50-yr (1965–2014) SLP trends for two modifications of the OBS-LE. (left) The dominant pattern in SLP and associated temperature and precipitation trends across a modified ensemble created excluding any dependence on the SST modes. The fraction of variance explained by the first EOF is 33% in the NH (37% in the SH). (right) The dominant pattern in SLP and associated temperature and precipitation trends across a modified ensemble created excluding any residual variability beyond the trend and SST modes. The fraction of variance explained by the first EOF is 57% in the NH (62% in the SH). Contour interval is ± 0.5 hPa $(50 \text{ yr})^{-1}$ (dashed contours are negative), and the zero contour is suppressed. SLP PCs are calculated separately for each hemisphere (poleward of 10°) but are shown on the same map for conciseness.

center of action is found over north-central Eurasia (Figs. 7b,d). This SLP trend pattern is associated with large-scale cooling from Alaska to the Great Lakes, and across central Eurasia, as well as weaker-amplitude warming over the southeastern United States and the Arctic coastlines of Eurasia and northeastern Greenland. Precipitation impacts include a drying of the Iberian Peninsula and Mexico, as well as a moistening of the Pacific Northwest and southeastern United States. Much of the structure of the SLP trend EOF2 is reproduced as SLP trend EOF1 by restricting the domain to the PNA sector (not shown). The second SLP trend EOF in the SH is associated with relatively zonal anomalies over the Southern Ocean north of East Antarctica, and a wavelike structure over the Pacific sector. In its positive projection, this EOF is largely associated with a cooling and wetting of Australia and South Africa, and anomalies of mixed sign over South America.

The EOF analysis highlights the substantial magnitudes of 50-yr trends in temperature and precipitation that can occur as a result of unforced variability in the large-scale atmospheric circulation. For many regions, these amplitudes rival those of the forced trends determined from climate model ensembles. For example, a positive (negative) two-standard-deviation departure of NH SLP trend first principal component (PC1) is associated with an average temperature trend over northern Eurasia (50° – 70°N , 10° – 160°E) of -1.3°C ($+1.3^\circ\text{C}$). Given a forced trend of $1.5^\circ\text{C} (50 \text{ yr})^{-1}$ for the same region based on the CESM1-LE [$2.4^\circ\text{C} (50 \text{ yr})^{-1}$ based on CMIP5], it is clear that internal variability related to the atmospheric circulation can have a nonnegligible impact on trends. The relative importance of unforced climate trends assessed from the OBS-LE and forced climate trends evaluated from climate model ensembles is further quantified in section 7.

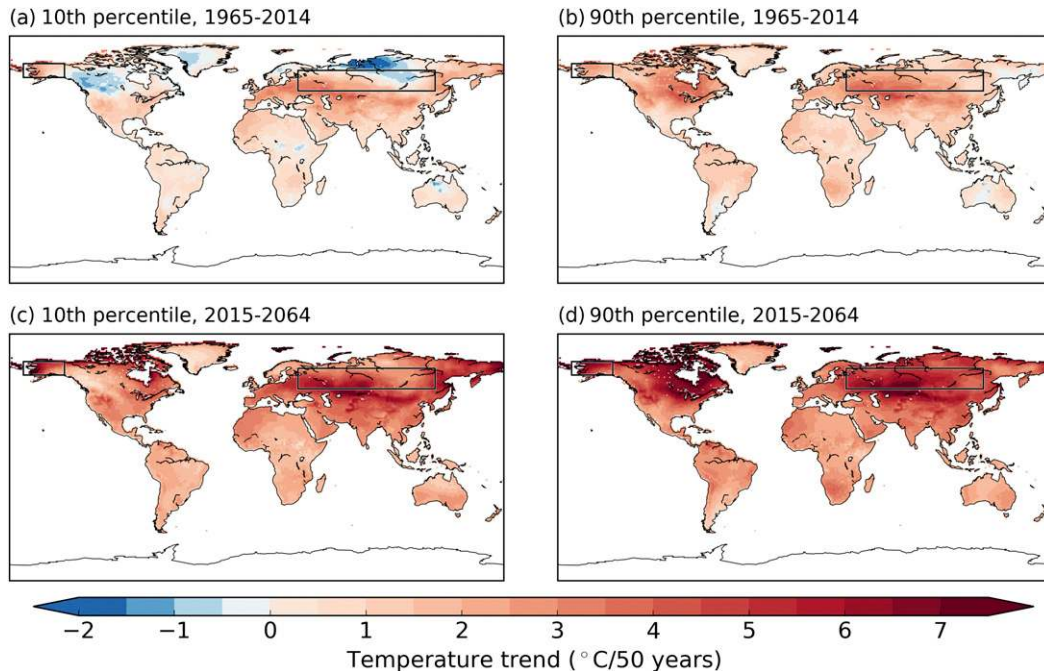


FIG. 9. Range of 50-yr DJF temperature trends from the OBS-LE for the (a),(b) past (1965–2014) and (c), (d) future (2015–2064). (left) The trend maps for the member associated with the 10th percentile of NH extratropical (30° – 90° N) land-average temperature trends. (right) The trend maps for the member associated with the 90th percentile of NH extratropical (30° – 90° N) land-average temperature trends. The 10th and 90th percentile DJF temperature trends for the past (future) are $1.14^{\circ}\text{C} (50\text{ yr})^{-1}$ [3.89° and $4.73^{\circ}\text{C} (50\text{ yr})^{-1}$], respectively. (Boxes indicate the regions used for the histograms in Figs. 11a,b.)

6. Relative roles of SST modes and midlatitude atmospheric variability

The two dominant EOFs of SLP trend variability and their associated climate impacts are, by design, a function of the surrogate ENSO, PDO, and AMO time series and the bootstrapped residual atmospheric variability. It is of interest to understand the relative contributions of each. In the context of the statistical model, this question can be easily answered by producing two alternative versions of the OBS-LE after fitting Eq. (1) to the data: one that removes the contribution of the SST modes via setting the $\beta_M^{i,m}$ to zero, and a second that removes the contribution of the residual variability by setting $\varepsilon^{i,t}$ to zero. Note that this is distinct from the model of McKinnon et al. (2017), which did not include the contribution of the modes when fitting the original statistical model; therefore, variability linearly related to the modes was included in the residual term. To understand the dominant modes of variability in trends in each of these modified versions of the OBS-LE, we again perform an EOF analysis on the trends in SLP, and assess their projection onto trends in temperature and precipitation.

The dominant EOF of the DJF SLP trends in the first modified version of the OBS-LE, which excludes the

contribution from the three SST modes (Figs. 8a,c), is very similar to the first EOF of the full OBS-LE in both hemispheres. The primary difference is the lack of variability around the Aleutian low, and a correspondingly weaker relationship with temperature around Alaska. The SLP, temperature, and precipitation trend variations throughout Eurasia, however, are almost identical to what was found for the complete OBS-LE, indicating that the dominant pattern of Eurasian trend variability is controlled by intrinsic atmospheric variability.

The dominant pattern of variability in SLP trends for the second modified version of the OBS-LE, which excludes the contribution from the residual variability (Figs. 8b,d), is primarily associated with changes in the strength of the Aleutian low—an iconic response to ENSO and the PDO—and in this sense is related the second EOF of the full OBS-LE; however, it lacks the center of action over northern Eurasia and its amplitude over the Eastern Hemisphere part of the Southern Ocean is muted compared to EOF2 of the full OBS-LE. Accordingly, the similarities in the associated temperature and precipitation trends are mainly confined to North America and the southern continents. The strong link between SLP trend variability in this version of the OBS-LE and North American temperature

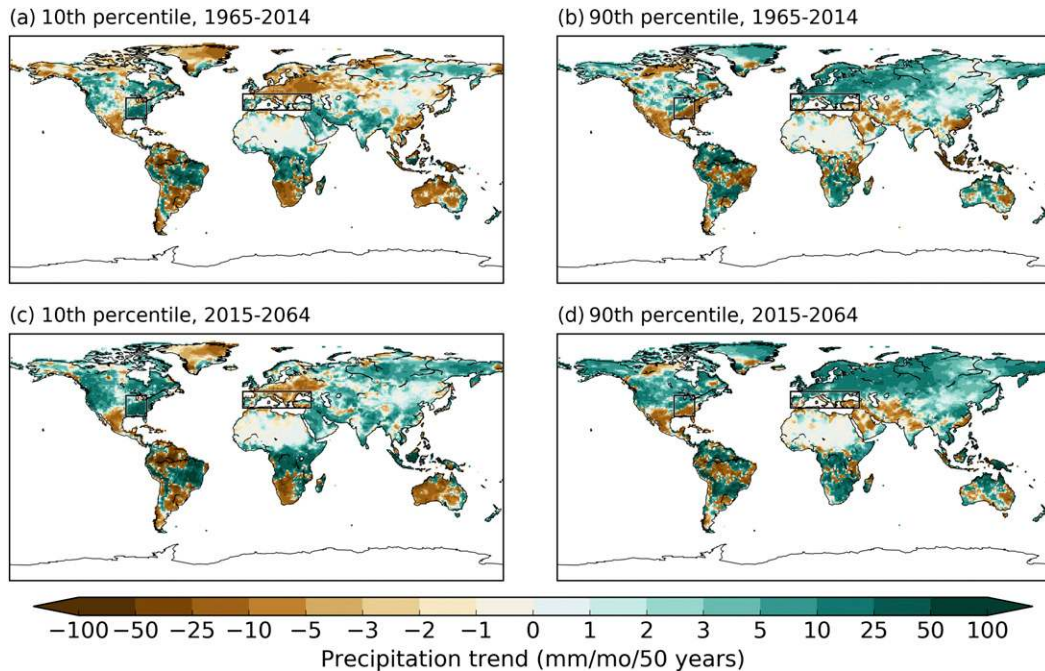


FIG. 10. As in Fig. 9, but for precipitation. The 10th and 90th percentile NH extratropical land DJF precipitation trends for the past (future) are 0.30 and $3.03 \text{ mm month}^{-1} (50 \text{ yr})^{-1}$ [3.78 and $6.51 \text{ mm month}^{-1} (50 \text{ yr})^{-1}$], respectively. (Boxes indicate the regions used for the histograms in Figs. 11c,d.)

and precipitation trends highlights the importance of improving our understanding of whether and how ENSO and the PDO may change in a future climate.

7. Diversity of regional temperature and precipitation trends in the OBS-LE

The prior two sections highlighted the patterns of temperature and precipitation trend variability associated with dominant circulation regimes, thereby providing one lens with which to examine the range of trends possible as a result of internal variations. A complementary view can be found through examining the members of the OBS-LE themselves. To do so in a succinct manner, we rank the members of the OBS-LE according to their terrestrial temperature (or precipitation) trends between 1965 and 2014 averaged over the extratropical NH (30° – 90° N). The patterns of temperature change associated with the 10th and 90th percentiles of the ensemble, as measured by the NH extratropical land trend, are shown in Fig. 9 for DJF (see Fig. S13 in the supplemental material for JJA). For trends over the past 50 years, the 10th percentile ensemble member shows cooling over western Canada and north-central Eurasia of 1° – $2^{\circ}\text{C} (50 \text{ yr})^{-1}$ in magnitude, maximum warming of $3^{\circ}\text{C} (50 \text{ yr})^{-1}$ across central Europe and Asia, and relatively uniform

warming of around $1^{\circ}\text{C} (50 \text{ yr})^{-1}$ elsewhere. In contrast, the 90th percentile ensemble member shows temperature increases almost everywhere, with the greatest warming over northeastern North America and southwestern Russia exceeding 4°C over 50 years (Figs. 9a,b).

Assuming that changes in interannual variability are small between the periods 1965–2014 and 2015–64 (see section 4 and Fig. S4), it is possible to similarly examine the spatial pattern associated with the 10th and 90th percentile of terrestrial extratropical temperature trends by replacing the historical ensemble mean temperature trend with a future projection. Using the RCP8.5 emission scenario and again drawing on the ensemble mean of the CESM1-LE as our forced component, we extend the OBS-LE to 2064. Because this “future OBS-LE” relies on the same variability statistics as our historical OBS-LE, the spread across the ensemble is unchanged. It is, however, of note that—in a future scenario—the ratio of the spread of land temperatures relative to the magnitude of the forced signal is reduced. In other words, as the world continues to warm, the fractional uncertainty in trends caused by internal variability will be reduced because of the larger forced signal. Nevertheless, the imprint of internal atmospheric circulation variability is still apparent in the same regions as before, where relatively

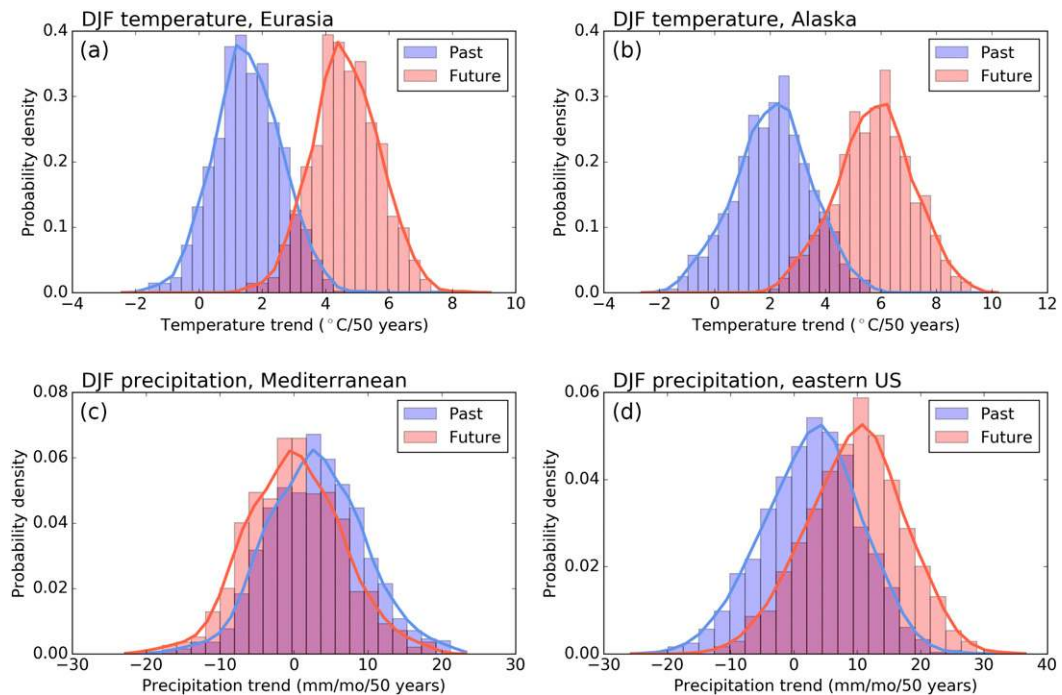


FIG. 11. Histograms of DJF (a),(b) temperature and (c),(d) precipitation trends over the past (blue; 1965–2014) and next (red; 2015–64) 50 years in the OBS-LE. See text and Figs. 9 and 10 for definition of regions.

mutated warming trends [$<1^{\circ}\text{C} (50\text{ yr})^{-1}$] occur in the 10th percentile member compared to trends $>3^{\circ}\text{C} (50\text{ yr})^{-1}$ in the 90th percentile member (Figs. 9c,d).

We next examine the 10th and 90th percentile members of the ensemble as measured by the DJF NH extratropical land precipitation trends (Fig. 10; see Fig. S14 in the supplemental material for JJA). In the 10th percentile member, much of central and eastern Europe exhibits a drying trend, while the 90th percentile member shows wetting. Differences in the sign of the precipitation trends between the 10th and 90th percentile members are also evident in Alaska, the southeastern United States, and parts of the Middle East. Looking into the future (2015–64), internal variability can still affect even the sign of the trend despite the larger forced wetting over the NH continents (recall Fig. 6), for example over much of Europe and the Middle East.

To further illustrate the relative contributions of internal variability and radiative forcing to temperature and precipitation trends, we examine the distributions of the DJF trends for the past and next 50 years in the OBS-LE in a select number of regions: Eurasia, Alaska, the Mediterranean region, and the eastern United States. Boxes outlining the regions are shown in Figs. 9 and 10. The regions were chosen because of their high sensitivity to atmospheric variability based on the EOF analysis summarized in Fig. 7. Temperature trends in both

Eurasia and Alaska (Figs. 11a,b) show a clear shift toward more positive values in the next 50 years, but still exhibit substantial overlap with those of the past 50 years such that a wide range of trends [$1^{\circ}\text{--}5^{\circ}\text{C} (50\text{ yr})^{-1}$ for Eurasia, $2^{\circ}\text{--}6^{\circ}\text{C} (50\text{ yr})^{-1}$ for Alaska] is possible in both the past and the future. Unlike temperature, DJF precipitation trends in the Mediterranean region and eastern United States (Figs. 11c,d) show a smaller radiatively forced component compared to the internal component. Indeed, the Mediterranean region shows almost complete overlap between the distributions of past and future 50-yr trends, with both centered near zero and ranging between approximately $\pm 20\text{ mm month}^{-1} (50\text{ yr})^{-1}$. A slightly larger shift toward positive precipitation trends can be seen in the eastern United States in the future relative to the past, although the overlap as a result of internal variability is again substantial, with trends between -15 and $+25\text{ mm month}^{-1} (50\text{ yr})^{-1}$ being possible in both time periods. The large spread relative to the forced signal for precipitation trends indicates the difficulty of inferring an anthropogenic climate change signal in any single multidecadal record of temperature or precipitation.

To generalize the foregoing results, we calculate the signal-to-noise ratio (SNR) at each grid box, defined as the magnitude of the forced trend divided by the standard deviation of trends across the OBS-LE. We do this

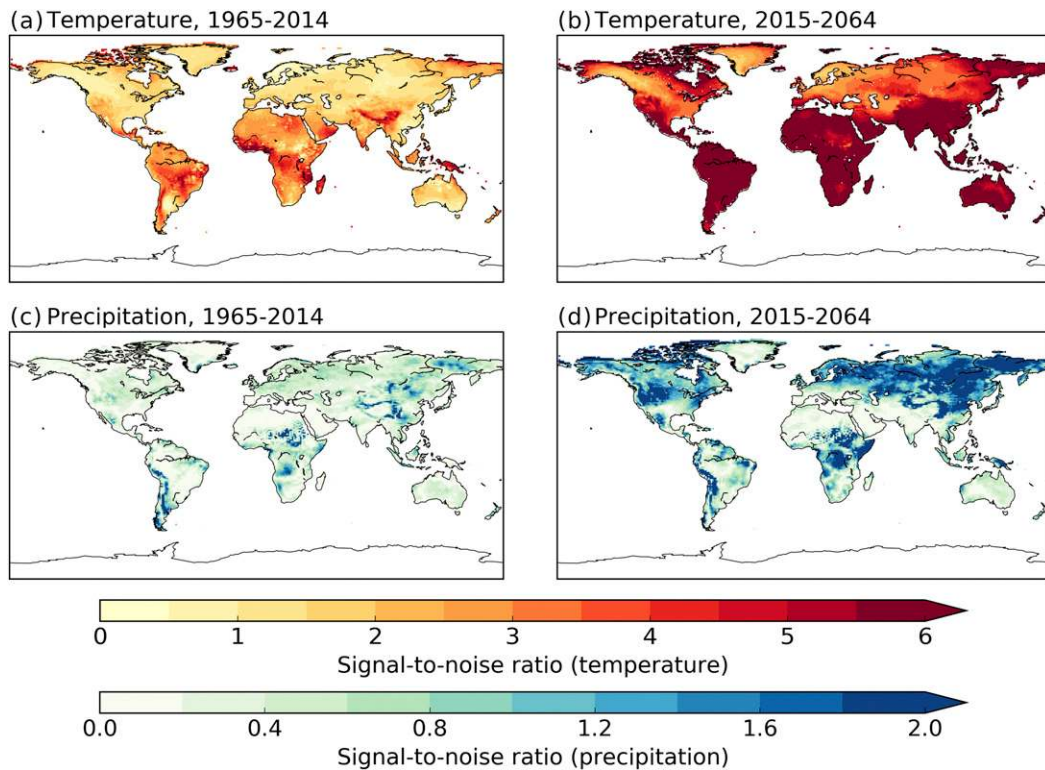


FIG. 12. The SNR in the OBS-LE for 50-yr trends in the (a),(c) past (1965–2014) and (b),(d) future (2015–64) for DJF (top) temperature and (bottom) precipitation. SNR is estimated as the magnitude of the ensemble mean 50-yr trend normalized by the standard deviation of trends across the ensemble.

for temperature and precipitation, and for the past and future (Fig. 12; see Figs. S15–S17 in the supplemental material for JJA and CMIP5 ensemble mean). The larger the ratio, the more the climate change signal is emerging from the climate noise.

As expected because of their relatively low variability, the regions with the highest SNR for DJF temperature are in the tropics. In contrast, about 10% of the global landmasses covering large regions of North America, Eurasia, and Australia exhibit a DJF temperature SNR less than one, indicating that the expected standard deviation of temperature trends from 1965 to 2014 is on par with the forced signal. The average terrestrial DJF temperature SNR is 2.2 for the historical period. The picture is quite different for the future period, when the global warming signal is much larger: the average DJF temperature SNR rises to 6.3, and 95% of land grid boxes have a SNR over 2.5.

Because the spread of 50-yr DJF precipitation trends is large compared to the magnitude of the forced trend, the SNRs are substantially smaller than those for temperature. Specifically, the SNR for precipitation trends over the past 50 years is less than 1 across 93% of land grid boxes and has an average value of 0.4; as such, it is

difficult to identify a climate change signal in DJF precipitation trends at the gridbox level. The signal emerges more clearly for future trends in the NH midlatitudes because the forced signal is larger. Nevertheless, even for DJF precipitation trends over the next 50 years, the global land average SNR is only 1.1. Thus, the forced future trend signal remains less prominent in precipitation compared to temperature.

8. Summary and discussion

We have developed an observational large ensemble that draws upon the observational record to constrain its internal variability, and upon model ensembles to determine its radiatively forced component. The members of the OBS-LE provide information about the magnitude of the contribution of internal variability to important climate metrics such as multidecadal trends.

By design, the variability in the OBS-LE is a function of three dominant SST modes—ENSO, PDO, and AMO—as well as residual variability, primarily associated with internal atmospheric dynamics. In most regions of the world, the contribution of the SST modes to interannual DJF temperature, precipitation, and SLP

variability is small compared to the forced trend and residual variability. The primary exception is North America, where teleconnections link the tropical Pacific to North Pacific atmospheric circulation, with an associated impact on both temperature and precipitation across parts of the continent. In contrast, we find that Eurasian surface climate variability is less strongly linked to the SST modes, and more strongly influenced by atmospheric variability.

Over the past 50 years, the contribution of internal variability to terrestrial climate trends is nonnegligible. Although temperatures are expected to rise with elevated greenhouse gas concentrations, individual members of the OBS-LE can exhibit large swaths of cooling across the extratropics. Assuming the accuracy of the forced trend inferred from the CESM1-LE, the observations can be viewed as one member of the OBS-LE that happened to have anomalously large warming over parts of the NH including western Canada, Greenland, and Scandinavia resulting from sampling of internal variability. These regions of warming led the observations to have a 50-yr trend over the global landmasses that was $0.23^{\circ}\text{C} (50\text{ yr})^{-1}$ greater than the forced trend from the CESM1-LE, but it is important to remember that an alternative reality of more warming or more cooling would both be consistent with the same climate change signal.

Looking to the future, the warming signal becomes sufficiently large that it is almost certain that all terrestrial locations will warm between 2015 and 2064, regardless of internal variability. The same cannot be said of precipitation, for which the SNR barely exceeds unity even in the next 50 years, although the signal of increasing precipitation across the NH midlatitudes does begin to emerge from the noise.

While the focus of this work has been on variability across the OBS-LE, the observationally based ensemble can also be used as a tool for model validation. Because each member of the OBS-LE contains self-consistent spatiotemporal fields of temperature, precipitation, and SLP, it is possible to compare both the variability and covariability in the OBS-LE with that from climate models. As shown in Figs. 4d and 5d, the CESM1-LE has large biases in the spread of 50-yr temperature and precipitation trends over land that should be accounted for before interpreting the model output as a representation of past or future reality. Additional work is necessary to understand the origins of these model biases, but given the close coupling of temperature and precipitation, one avenue of inquiry is the nature and strength of the coupling in the real world versus the model simulations. The members of the OBS-LE generally show the expected (e.g., Trenberth and Shea 2005) negative correlation between temperature and

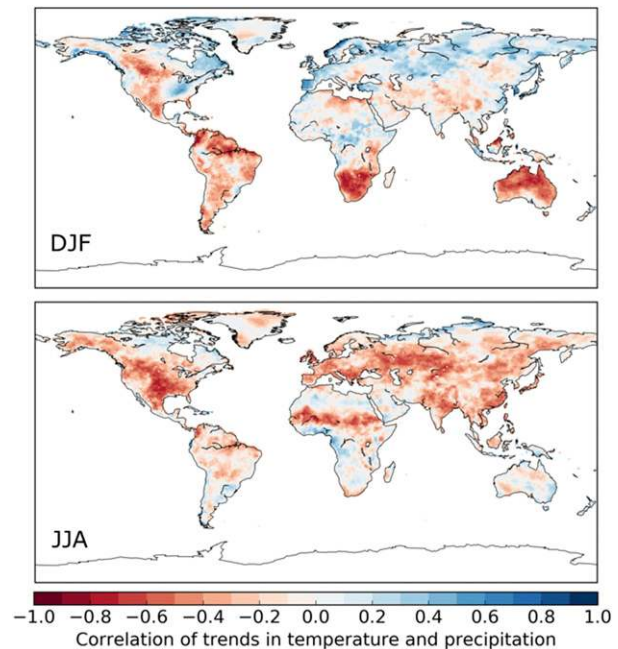


FIG. 13. Local correlation between 50-yr trends in temperature and precipitation for (top) DJF and (bottom) JJA from the OBS-LE.

precipitation trends in the summer hemisphere and positive correlation in the winter hemisphere (Fig. 13). Superimposed upon this pattern, however, are important deviations such as a negative temperature–precipitation correlation during the cold season in Mexico, the interior western United States, and western Canada, and positive correlations during the warm season along the north coast of Canada and Eurasia. Determining whether model ensembles reproduce these structures is important for assessing both model physics and the credibility of future projections.

Neither statistical nor dynamical models alone are sufficient to understand the climate system. Here, we have combined information about interannual variability from the observed world with model-based estimates of the forced response to anthropogenic influence in order to better approximate the role of internal variability in climatic trends. The analysis has been limited to the period of reliable instrumental records, but analogous approaches could be applied to longer records from paleo-proxy data in order to better estimate the longer-term variability in the climate system. In addition, there is scope for expanding the OBS-LE methodology to other quantities of interest such as ocean temperatures, sea ice, and soil moisture that are characterized by longer-term persistence.

Acknowledgments. K. A. M. was supported by the Advanced Study Program at the National Center for Atmospheric Research (NCAR). NCAR is sponsored

by the National Science Foundation. The authors thank Adam Phillips for his assistance with the Climate Variability Diagnostics Package in processing CMIP5 output. Samples from the observational large ensemble (OBS-LE) are available at <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/VYH544>. Code to reproduce the ensemble is available at <https://github.com/karenamckinnon>.

REFERENCES

- Alexander, M. A., L. Matrosova, C. Penland, J. D. Scott, and P. Chang, 2008: Forecasting Pacific SSTs: Linear inverse model predictions of the PDO. *J. Climate*, **21**, 385–402, <https://doi.org/10.1175/2007JCLI1849.1>.
- Ault, T. R., J. E. Cole, J. T. Overpeck, G. T. Pederson, S. St. George, B. Otto-Bliesner, C. A. Woodhouse, and C. Deser, 2013: The continuum of hydroclimate variability in western North America during the last millennium. *J. Climate*, **26**, 5863–5878, <https://doi.org/10.1175/JCLI-D-11-00732.1>.
- Beltrán, F., B. Sansó, R. T. Lemos, and R. Mendelssohn, 2012: Joint projections of North Pacific sea surface temperature from different global climate models. *Environmetrics*, **23**, 451–465, <https://doi.org/10.1002/env.2150>.
- Booth, B. B. B., N. J. Dunstone, P. R. Halloran, T. Andrews, and N. Bellouin, 2012: Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. *Nature*, **484**, 228–232, <https://doi.org/10.1038/nature10946>.
- Branstator, G., H. Teng, G. A. Meehl, M. Kimoto, J. R. Knight, M. Latif, and A. Rosati, 2012: Systematic estimates of initial-value decadal predictability for six AOGCMs. *J. Climate*, **25**, 1827–1846, <https://doi.org/10.1175/JCLI-D-11-00227.1>.
- Brown, P. T., W. Li, E. C. Cordero, and S. A. Mauget, 2015: Comparing the model-simulated global warming signal to observations using empirical estimates of unforced noise. *Sci. Rep.*, **5**, 9957, <https://doi.org/10.1038/srep09957>.
- Chen, C., M. A. Cane, A. T. Wittenberg, and D. Chen, 2017: ENSO in the CMIP5 simulations: Life cycles, diversity, and responses to climate change. *J. Climate*, **30**, 775–801, <https://doi.org/10.1175/JCLI-D-15-0901.1>.
- Collins, M., and M. R. Allen, 2002: Assessing the relative roles of initial and boundary conditions in interannual to decadal climate predictability. *J. Climate*, **15**, 3104–3109, [https://doi.org/10.1175/1520-0442\(2002\)015<3104:ATRROI>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<3104:ATRROI>2.0.CO;2).
- Compo, G. P., and Coauthors, 2011: The Twentieth Century Reanalysis Project. *Quart. J. Roy. Meteor. Soc.*, **137**, 1–28, <https://doi.org/10.1002/qj.776>.
- Dai, A., J. C. Fyfe, S.-P. Xie, and X. Dai, 2015: Decadal modulation of global surface temperature by internal climate variability. *Nat. Climate Change*, **5**, 555–559, <https://doi.org/10.1038/nclimate2605>.
- Deser, C., R. Knutti, S. Solomon, and A. S. Phillips, 2012a: Communication of the role of natural variability in future North American climate. *Nat. Climate Change*, **2**, 775–779, <https://doi.org/10.1038/nclimate1562>.
- , A. Phillips, V. Bourdette, and H. Teng, 2012b: Uncertainty in climate change projections: The role of internal variability. *Climate Dyn.*, **38**, 527–546, <https://doi.org/10.1007/s00382-010-0977-x>.
- , and Coauthors, 2012c: ENSO and Pacific decadal variability in the Community Climate System Model version 4. *J. Climate*, **25**, 2622–2651, <https://doi.org/10.1175/JCLI-D-11-00301.1>.
- , A. S. Phillips, M. A. Alexander, and B. V. Smoliak, 2014: Projecting North American climate over the next 50 years: Uncertainty due to internal variability. *J. Climate*, **27**, 2271–2296, <https://doi.org/10.1175/JCLI-D-13-00451.1>.
- , L. Terray, and A. S. Phillips, 2016: Forced and internal components of winter air temperature trends over North America during the past 50 years: Mechanisms and implications. *J. Climate*, **29**, 2237–2258, <https://doi.org/10.1175/JCLI-D-15-0304.1>.
- , I. R. Simpson, K. A. McKinnon, and A. S. Phillips, 2017: The Northern Hemisphere extratropical atmospheric circulation response to ENSO: How well do we know it and how do we evaluate models accordingly? *J. Climate*, **30**, 5059–5082, <https://doi.org/10.1175/JCLI-D-16-0844.1>.
- Ding, R., J. Li, F. Zheng, J. Feng, and D. Liu, 2016: Estimating the limit of decadal-scale climate predictability using observational data. *Climate Dyn.*, **46**, 1563–1580, <https://doi.org/10.1007/s00382-015-2662-6>.
- Gneiting, T., and A. E. Raftery, 2005: Weather forecasting with ensemble methods. *Science*, **310**, 248–249, <https://doi.org/10.1126/science.1115255>.
- Hawkins, E., and R. Sutton, 2009a: Decadal predictability of the Atlantic Ocean in a coupled GCM: Forecast skill and optimal perturbations using linear inverse modeling. *J. Climate*, **22**, 3960–3978, <https://doi.org/10.1175/2009JCLI2720.1>.
- , and —, 2009b: The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteor. Soc.*, **90**, 1095–1108, <https://doi.org/10.1175/2009BAMS2607.1>.
- , and —, 2011: The potential to narrow uncertainty in projections of regional precipitation change. *Climate Dyn.*, **37**, 407–418, <https://doi.org/10.1007/s00382-010-0810-6>.
- , T. Edwards, and D. McNeill, 2014: Pause for thought. *Nat. Climate Change*, **4**, 154–156, <https://doi.org/10.1038/nclimate2150>.
- Hendon, H. H., D. W. J. Thompson, and M. C. Wheeler, 2007: Australian rainfall and surface temperature variations associated with the Southern Hemisphere annular mode. *J. Climate*, **20**, 2452–2467, <https://doi.org/10.1175/JCLI4134.1>.
- Hu, A., and C. Deser, 2013: Uncertainty in future regional sea level rise due to internal climate variability. *Geophys. Res. Lett.*, **40**, 2768–2772, <https://doi.org/10.1002/grl.50531>.
- Kay, J., and Coauthors, 2015: The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bull. Amer. Meteor. Soc.*, **96**, 1333–1349, <https://doi.org/10.1175/BAMS-D-13-00255.1>.
- Knight, J. R., 2009: The Atlantic multidecadal oscillation inferred from the forced climate response in coupled general circulation models. *J. Climate*, **22**, 1610–1625, <https://doi.org/10.1175/2008JCLI2628.1>.
- Laepple, T., and P. Huybers, 2014: Ocean surface temperature variability: Large model–data differences at decadal and longer periods. *Proc. Natl. Acad. Sci. USA*, **111**, 16682–16687, <https://doi.org/10.1073/pnas.1412077111>.
- Lamarque, J.-F., and Coauthors, 2010: Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: Methodology and application. *Atmos. Chem. Phys.*, **10**, 7017–7039, <https://doi.org/10.5194/acp-10-7017-2010>.
- , G. P. Kyle, M. Meinshausen, K. Riahi, S. J. Smith, D. P. van Vuuren, A. J. Conley, and F. Vitt, 2011: Global and regional evolution of short-lived radiatively-active gases and aerosols in the representative concentration pathways. *Climatic Change*, **109**, 191–212, <https://doi.org/10.1007/s10584-011-0155-0>.

- Lapp, S. L., J.-M. St. Jacques, E. M. Barrow, and D. J. Sauchyn, 2012: GCM projections for the Pacific decadal oscillation under greenhouse forcing for the early 21st century. *Int. J. Climatol.*, **32**, 1423–1442, <https://doi.org/10.1002/joc.2364>.
- Lee, T., and T. B. M. J. Ouarda, 2012: An EMD and PCA hybrid approach for separating noise from signal, and signal in climate change detection. *Int. J. Climatol.*, **32**, 624–634, <https://doi.org/10.1002/joc.2299>.
- Lehner, F., A. P. Schurer, G. C. Hegerl, C. Deser, and T. L. Frölicher, 2016: The importance of ENSO phase during volcanic eruptions for detection and attribution. *Geophys. Res. Lett.*, **43**, 2851–2858, <https://doi.org/10.1002/2016GL067935>.
- , C. Deser, and L. Terray, 2017: Toward a new estimate of “time of emergence” of anthropogenic warming: Insights from dynamical adjustment and a large initial-condition model ensemble. *J. Climate*, **30**, 7739–7756, <https://doi.org/10.1175/JCLI-D-16-0792.1>.
- Madden, R. A., 1981: A quantitative approach to long-range prediction. *J. Geophys. Res.*, **86**, 9817–9825, <https://doi.org/10.1029/JC086iC10p09817>.
- Mantua, N. J., and S. R. Hare, 2002: The Pacific decadal oscillation. *J. Oceanogr.*, **58**, 35–44, <https://doi.org/10.1023/A:1015820616384>.
- McGraw, M. C., E. A. Barnes, and C. Deser, 2016: Reconciling the observed and modeled Southern Hemisphere circulation response to volcanic eruptions. *Geophys. Res. Lett.*, **43**, 7259–7266, <https://doi.org/10.1002/2016GL069835>.
- McKinnon, K. A., A. Poppick, E. Dunn-Sigouin, and C. Deser, 2017: An “observational large ensemble” to compare observed and modeled temperature trend uncertainty due to internal variability. *J. Climate*, **30**, 7585–7598, <https://doi.org/10.1175/JCLI-D-16-0905.1>.
- Meinshausen, M., and Coauthors, 2011: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, **109**, 213–241, <https://doi.org/10.1007/s10584-011-0156-z>.
- Messié, M., and F. Chavez, 2011: Global modes of sea surface temperature variability in relation to regional climate indices. *J. Climate*, **24**, 4314–4331, <https://doi.org/10.1175/2011JCLI3941.1>.
- Murphy, L. N., K. Bellomo, M. Cane, and A. Clement, 2017: The role of historical forcings in simulating the observed Atlantic multidecadal oscillation. *Geophys. Res. Lett.*, **44**, 2472–2480, <https://doi.org/10.1002/2016GL071337>.
- Navarra, A., M. Ward, and N. A. Rayner, 1998: A stochastic model of SST for climate simulation experiments. *Climate Dyn.*, **14**, 473–487, <https://doi.org/10.1007/s003820050235>.
- Newman, M., and Coauthors, 2016: The Pacific decadal oscillation, revisited. *J. Climate*, **29**, 4399–4427, <https://doi.org/10.1175/JCLI-D-15-0508.1>.
- Otterå, O. H., M. Bentsen, H. Drange, and L. Suo, 2010: External forcing as a metronome for Atlantic multidecadal variability. *Nat. Geosci.*, **3**, 688–694, <https://doi.org/10.1038/ngeo955>.
- Patt, A., and S. Dessai, 2005: Communicating uncertainty: Lessons learned and suggestions for climate change assessment. *C. R. Geosci.*, **337**, 425–441, <https://doi.org/10.1016/j.crte.2004.10.004>.
- Percival, D. B., and A. T. Walden, 1993: *Spectral Analysis for Physical Applications*. Cambridge University Press, 612 pp.
- Phillips, A. S., C. Deser, and J. Fasullo, 2014: Evaluating modes of variability in climate models. *Eos, Trans. Amer. Geophys. Union*, **95**, 453–455, <https://doi.org/10.1002/2014EO490002>.
- Polvani, L. M., and K. L. Smith, 2013: Can natural variability explain observed Antarctic sea ice trends? New modeling evidence from CMIP5. *Geophys. Res. Lett.*, **40**, 3195–3199, <https://doi.org/10.1002/grl.50578>.
- , D. W. Waugh, G. J. P. Correa, and S.-W. Son, 2011: Stratospheric ozone depletion: The main driver of twentieth-century atmospheric circulation changes in the Southern Hemisphere. *J. Climate*, **24**, 795–812, <https://doi.org/10.1175/2010JCLI3772.1>.
- Räisänen, J., 2001: CO₂-induced climate change in CMIP2 experiments: Quantification of agreement and role of internal variability. *J. Climate*, **14**, 2088–2104, [https://doi.org/10.1175/1520-0442\(2001\)014<2088:CICCCIC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<2088:CICCCIC>2.0.CO;2).
- Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, <https://doi.org/10.1029/2002JD002670>.
- Rhines, A., K. A. McKinnon, M. P. Tingley, and P. Huybers, 2017: Seasonally resolved distributional trends of North American temperatures show contraction of winter variability. *J. Climate*, **30**, 1139–1157, <https://doi.org/10.1175/JCLI-D-16-0363.1>.
- Roeckner, E., and Coauthors, 2003: The atmospheric general circulation model ECHAM 5. Part I: Model description. Max Planck Institute for Meteorology Rep. 349, 140 pp.
- Rohde, R., and Coauthors, 2013: Berkeley Earth temperature averaging process. *Geoinf. Geostat. Overview*, **1**, 1000103, <https://doi.org/10.4172/2327-4581.1000103>.
- Salazar, E., D. Hammerling, X. Wang, B. Sansó, A. O. Finley, and L. O. Mearns, 2016: Observation-based blended projections from ensembles of regional climate models. *Climatic Change*, **138**, 55–69, <https://doi.org/10.1007/s10584-016-1722-1>.
- Schneider, U., P. Finger, A. Meyer-Christoffer, E. Rustemeier, M. Ziese, and A. Becker, 2017: Evaluating the hydrological cycle over land using the newly-corrected precipitation climatology from the Global Precipitation Climatology Centre (GPCC). *Atmosphere*, **8**, 52, <https://doi.org/10.3390/atmos8030052>.
- Schreiber, T., and A. Schmitz, 2000: Surrogate time series. *Physica D*, **142**, 346–382, [https://doi.org/10.1016/S0167-2789\(00\)00043-9](https://doi.org/10.1016/S0167-2789(00)00043-9).
- Screen, J. A., 2014: Arctic amplification decreases temperature variance in northern mid- to high-latitudes. *Nat. Climate Change*, **4**, 577–582, <https://doi.org/10.1038/nclimate2268>.
- Smith, D. M., and Coauthors, 2016: Role of volcanic and anthropogenic aerosols in the recent global surface warming slowdown. *Nat. Climate Change*, **6**, 936–940, <https://doi.org/10.1038/nclimate3058>.
- Smoliak, B. V., J. M. Wallace, P. Lin, and Q. Fu, 2015: Dynamical adjustment of the Northern Hemisphere surface air temperature field: Methodology and application to observations. *J. Climate*, **28**, 1613–1629, <https://doi.org/10.1175/JCLI-D-14-00111.1>.
- Sterl, A., and Coauthors, 2008: When can we expect extremely high surface temperatures? *Geophys. Res. Lett.*, **35**, L14703, <https://doi.org/10.1029/2008GL034071>.
- Swart, N. C., J. C. Fyfe, E. Hawkins, J. E. Kay, and A. Jahn, 2015: Influence of internal variability on Arctic sea-ice trends. *Nat. Climate Change*, **5**, 86–89, <https://doi.org/10.1038/nclimate2483>.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, <https://doi.org/10.1175/BAMS-D-11-00094.1>.
- Tebaldi, C., and R. Knutti, 2007: The use of the multi-model ensemble in probabilistic climate projections. *Philos. Trans. Roy. Soc. London*, **365A**, 2053–2075, <https://doi.org/10.1098/rsta.2007.2076>.
- Theiler, J., S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, 1992: Testing for nonlinearity in time series: The method of surrogate data. *Physica D*, **58**, 77–94, [https://doi.org/10.1016/0167-2789\(92\)90102-S](https://doi.org/10.1016/0167-2789(92)90102-S).
- Thompson, D. W. J., and J. M. Wallace, 2000: Annular modes in the extratropical circulation. Part I: Month-to-month

- variability. *J. Climate*, **13**, 1000–1016, [https://doi.org/10.1175/1520-0442\(2000\)013<1000:AMITEC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<1000:AMITEC>2.0.CO;2).
- , and —, 2001: Regional climate impacts of the Northern Hemisphere annular mode. *Science*, **293**, 85–89, <https://doi.org/10.1126/science.1058958>.
- , E. A. Barnes, C. Deser, W. E. Foust, and A. S. Phillips, 2015: Quantifying the role of internal climate variability in future climate trends. *J. Climate*, **28**, 6443–6456, <https://doi.org/10.1175/JCLI-D-14-00830.1>.
- Trenberth, K. E., and D. J. Shea, 2005: Relationships between precipitation and surface temperature. *Geophys. Res. Lett.*, **32**, L14703, <https://doi.org/10.1029/2005GL022760>.
- , G. W. Branstator, D. Karoly, A. Kumar, N.-C. Lau, and C. Ropelewski, 1998: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *J. Geophys. Res.*, **103**, 14 291–14 324, <https://doi.org/10.1029/97JC01444>.
- van Oldenborgh, G. J., F. J. Doblas Reyes, S. S. Drijfhout, and E. Hawkins, 2013: Reliability of regional climate model trends. *Environ. Res. Lett.*, **8**, 014055, <https://doi.org/10.1088/1748-9326/8/1/014055>.
- Wettstein, J. J., and C. Deser, 2014: Internal variability in projections of twenty-first-century Arctic sea ice loss: Role of the large-scale atmospheric circulation. *J. Climate*, **27**, 527–550, <https://doi.org/10.1175/JCLI-D-12-00839.1>.
- Woodruff, S. C., 2016: Planning for an unknowable future: Uncertainty in climate change adaptation planning. *Climatic Change*, **139**, 445–459, <https://doi.org/10.1007/s10584-016-1822-y>.
- Zhang, L., and T. L. Delworth, 2016: Simulated response of the Pacific decadal oscillation to climate change. *J. Climate*, **29**, 5999–6018, <https://doi.org/10.1175/JCLI-D-15-0690.1>.
- Zhang, R., and Coauthors, 2013: Have aerosols caused the observed Atlantic multidecadal variability? *J. Atmos. Sci.*, **70**, 1135–1144, <https://doi.org/10.1175/JAS-D-12-0331.1>.
- Zhou, Z.-Q., S.-P. Xie, X.-T. Zheng, Q. Liu, and H. Wang, 2014: Global warming-induced changes in El Niño teleconnections over the North Pacific and North America. *J. Climate*, **27**, 9050–9064, <https://doi.org/10.1175/JCLI-D-14-00254.1>.