Drivers of Twenty-First-Century U.S. Winter Precipitation Trends in CMIP6 Models: A Storyline-Based Approach

DANIEL F. SCHMIDT[a] AND KEVIN M. GRISE[a]

[a] *Department of Environmental Sciences, University of Virginia, Charlottesville, Virginia*

(Manuscript received 27 January 2021, in final form 28 May 2021)

ABSTRACT: Climate change during the twenty-first century has the potential to substantially alter geographic patterns of precipitation. However, regional precipitation changes can be very difficult to project, and in some regions, global climate models do not even agree on the sign of the precipitation trend. Since some of this uncertainty is due to internal variability rather than model bias, models cannot be used to narrow the possibilities to a single outcome, but they can usefully quantify the range of plausible outcomes and identify the combination of dynamical drivers that would be likely to produce each. This study uses a storylines approach—a type of regression-based analysis—to identify some of the key dynamical drivers that explain the variance in twenty-first-century U.S. winter precipitation trends across CMIP6 models under the SSP3–7.0 emissions scenario. This analysis shows that the spread in precipitation trends is not primarily driven by differences in modeled climate sensitivity. Key drivers include global-mean surface temperature, but also tropical upper-troposphere temperature, El Niño–Southern Oscillation (ENSO), the Pacific–North America (PNA) pattern, and the east Pacific (EP) dipole (a dipole pattern in geopotential heights over North America's Pacific coast). Combinations of these drivers can reinforce or cancel to produce various high- or low-impact scenarios for winter precipitation trends in various regions of the United States. For example, the most extreme winter precipitation trends in the southwestern United States result from opposite trends in ENSO and EP, whereas the wettest winter precipitation trends in the midwestern United States result from a combination of strong global warming and a negative PNA trend.

SIGNIFICANCE STATEMENT: The newest generation of climate models (CMIP6) is now available, but despite some improvements, models still disagree on future precipitation changes over North America. In some ways, this is to be expected: precipitation changes (both in the real world and in the models) depend partly on climate change, but also partly on random natural variability, so they will never be completely predictable. Thus, instead of trying to pin down future precipitation changes to a specific outcome, we show a range of plausible outcomes that policymakers should be prepared for. We also explore the reasons for the differences between model runs. For example, precipitation trends on the U.S. West Coast differ between models partly because precipitation in that region is affected by El Niño events (as well as other factors that we identify), and future changes to El Niño differ between models, even for the same amount of warming.

KEYWORDS: Climate change; North America; El Niño; Pacific-North American pattern/oscillation; Precipitation

1. Introduction

Regional precipitation trends are among the more challenging processes to project in climate change scenarios. Some broadly defined patterns, such as an increase in zonal-mean precipitation in the deep tropics, a decrease in zonal-mean precipitation in the subtropics, and an increase in precipitation at high latitudes, are robust across models and are expected for well-understood theoretical reasons (Held and Soden 2006; Seager et al. 2010; Scheff and Frierson 2012). However, impacts of future precipitation change will be felt not through the zonal mean, but locally, and at local scales the picture is considerably more uncertain, especially over midlatitude land regions such as the contiguous United States (Deser et al. 2012).

The sources of uncertainty in U.S. precipitation trends can be decomposed into three categories: scenario uncertainty, model uncertainty, and internal variability (e.g., Deser et al. 2012). The first two of these could in principle be reduced by better constraints on future emissions and improvements in climate models, respectively. For temperature projections, reductions in scenario and model uncertainty by themselves would be enough to substantially narrow the range of possible trends (Hawkins and Sutton 2009). However, actually achieving this improvement is challenging in practice—the range in equilibrium climate sensitivity in climate models has not narrowed across successive generations of models (Meehl et al. 2020; Zelinka et al. 2020). On the other hand, some improvements have been made in model representations of precipitation (Fiedler et al. 2020).

The third category of uncertainty (internal variability) is an even more challenging problem. Internal variability can be defined as the effectively random variability present in the

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-21-0080.s1.

Corresponding author: Daniel F. Schmidt, dfs2uc@virginia.edu

DOI: 10.1175/JCLI-D-21-0080.1

climate system that is not driven by external forcing. Although it is unforced, internal variability can still produce substantial decadal or centennial trends, and this long-term manifestation of internal variability can be thought of as a sort of ''irreducible uncertainty'' whose distribution cannot be tightened by model improvements. The special challenge of precipitation trends—in contrast to temperature trends, for example—is that precipitation is especially sensitive to internal variability. This sensitivity allows internal variability to mask forced trends in precipitation, both for observations (Hoerling et al. 2010, 2016) and for future projections (Deser et al. 2012; 2014). For some U.S. regions, the forced signal in precipitation may not be distinguishable from internal variability until after 2100 (Giorgi and Bi 2009).

Since the uncertainty in future U.S. precipitation cannot be eliminated, we turn instead to the task of describing the range of plausible outcomes. Internal variability in the climate system can be represented by a large initial-condition ensemble, in which a single model is run many times with identical forcings but with very slight differences in initial conditions [see a recent review by Deser et al. (2020)]. The ensemble mean trend can then be taken to represent the forced response, and the spread among ensemble members can be taken as an estimate of the magnitude of internal variability. Similarly, multimodel ensembles can be used to estimate the range of outcomes due to the combination of two uncertainties: model uncertainty and internal variability. In both cases, the projected outcome can be represented by an ensemble mean with error bars determined by the ensemble spread.

This traditional approach is simple and intuitive, but there are two reasons one may wish to go beyond it. First, the method of identifying an ensemble mean bracketed by error bars representing the spread implicitly assumes (or may be interpreted as assuming) that the spread among ensemble members can be interpreted probabilistically, an assumption that is not necessarily justified (Zappa 2019). From the perspective of communication, another drawback of this method is that representing outcomes in terms of a mean bracketed by error bars naturally draws one's attention to the mean. Thus, despite the best efforts of researchers to communicate uncertainties, this presentation of the results may actually distract attention from them.

Second, we would ideally like to know *why* the models disagree, and which physical drivers in the climate system are responsible for the spread in trends. This question is of scientific interest, but it also may be practically important if some of these drivers are represented more accurately in some models than in others.

Hence, some recent work on climate impacts has moved away from the approach described above, and instead has explicitly focused on exploring multiple outcomes in detail. Several methods have been proposed for doing this (e.g., Lenderink et al. 2014; Hazeleger et al. 2015). These often involve regression-based approaches. That is, the regression of precipitation trends (for example) onto the trends in some dynamical driver across an ensemble of model runs is computed, and the regression slope is used to estimate the sensitivity of precipitation to that driver (Bladé et al. 2012; Manzini et al. 2014; van den Hurk et al. 2014a,b; Deser et al. 2017).

One particularly interesting version of a regression-based method is the so-called storylines approach of Zappa and Shepherd (2017) (see also Shepherd et al. 2018; Shepherd 2019; Zappa 2019). Using CMIP5 models, these authors 1) identified key dynamic or thermodynamic drivers linked to future precipitation and wind speed trends over Europe, 2) determined the sensitivity of the precipitation and wind speed trends to each process, and then 3) constructed synthetic high- and low-impact storylines for the twenty-first century based on combinations of these drivers. The key is that multiple drivers are considered simultaneously, in order to quantify the impacts of various combinations—the ''storylines'' of the method's name. The same method has since been applied to the Southern Hemisphere midlatitudes (Mindlin et al. 2020). This method addresses both of the problems identified above, since it demonstrates the various plausible outcomes without requiring a probabilistic interpretation, and explicitly identifies the processes responsible for the differences between these outcomes.

In this paper, we apply a storylines approach to twenty-first-century precipitation trends for the United States using output from the global climate models that participated in phase 6 of the Coupled Model Intercomparison Project (CMIP6; Eyring et al. 2016), keeping in mind that the key dynamic and thermodynamic drivers relevant for local precipitation over the United States are likely different from those for Europe (Zappa and Shepherd 2017) and the Southern Hemisphere midlatitudes (Mindlin et al. 2020). Specifically, we seek to answer three questions with our analyses: 1) Which drivers are responsible for the differences in U.S. precipitation trends across CMIP6 models and across ensemble members of individual CMIP6 models? 2) What are the wettest and driest plausible storylines that these drivers could combine to produce for each region? 3) What are the relative roles of model uncertainty and internal variability in these precipitation trends?

2. Data and methods

We use monthly mean output from CMIP6 models, using the SSP3–7.0 emissions scenario. This emissions scenario is driven by projected forcings over the period 2015–2100 such that the global average radiative forcing is $7.0\,\mathrm{W\,m^{-2}}$ in the year 2100. We choose this emissions scenario because it was requested to have the largest number of model ensemble members of all the CMIP6 twenty-first-century scenarios (O'Neill et al. 2016).

At the time of download, 185 ensemble members from 34 models were available for the CMIP6 SSP3–7.0 scenario. A few ensemble members have only 40 or 41 years of data, instead of the full 86 years (2015–2100). We exclude these, leaving a total of 176 ensemble members from 32 separate models (see Table S1 in the online supplemental material for a list of the models used). In this paper, we consider only the December–February (DJF) season, and we remove the seasonal cycle by subtracting the mean over all December data from each December, and similarly for other months.

To compare the CMIP6 models results with observations, we also use reanalysis data from the European Centre for Medium-Range Weather Forecasts (ERA5; Hersbach et al. 2019, 2020). Specifically, we use the monthly mean 500-hPa geopotential height, 20-hPa zonal wind, sea surface temperature, and air

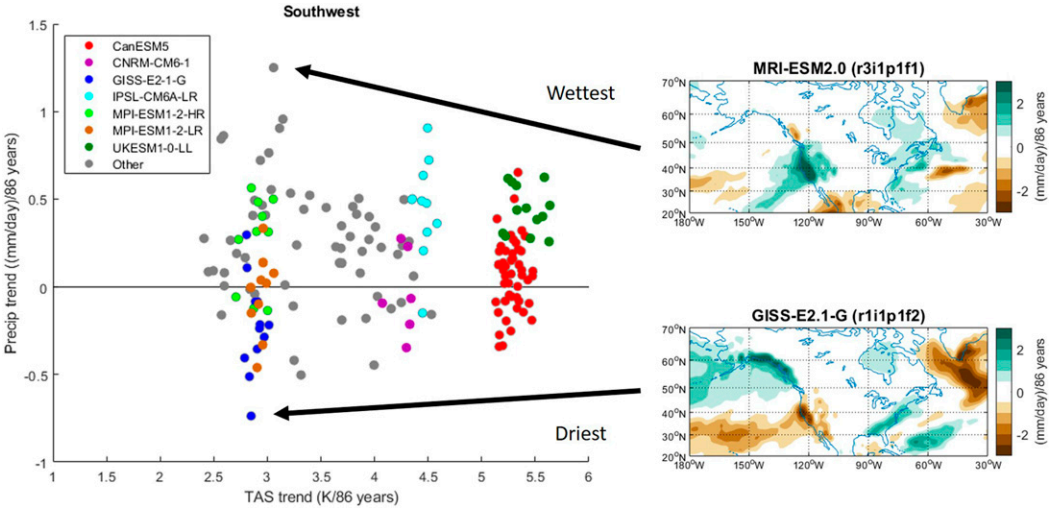

FIG. 1. (left) Scatterplot of DJF-season southwestern U.S. precipitation trends and global-mean surface air temperature (TAS) trends; 176 ensemble members from 32 CMIP6 models are shown. The individual models with the largest ensembles are shown in color. (right) Maps of precipitation trends for the ensemble members with the (top) wettest and (bottom) driest trends for this region.

temperature through the depth of the atmosphere for the period 1979–2020. For the air temperature data, we select only those pressure levels that are also available in CMIP6, for consistency.

We consider modeled precipitation averaged over each of 10 regions of the United States: Alaska (AK), Hawaii (HI), the Northwest (NW), the Southwest (SW), the Northern Plains (NP), the Southern Plains (SP), the Midwest (MW), the Northeast (NE), the Southeast (SE), and the Caribbean (CB). These correspond to the National Climate Assessment (NCA; USGCRP 2018) regions, as shown in Fig. S1 in the online supplemental material. The slight differences between our regions and the original NCA regions are that 1) due to limits of resolution, some small islands are excluded, and 2) for Hawaii and for the Caribbean (i.e., Puerto Rico and the Virgin Islands) we use polygons drawn around the major islands instead of attempting to follow the actual shapes of the individual islands.

At various points in the paper, we consider the component of the twenty-first-century trend in some field y that is uncorrelated with the concurrent trend in global-mean surface temperature (TAS). For example, suppose we are considering the 2015–2100 trend in y for each ensemble member n, denoted $\Delta y(n)$. Let R be the slope of the linear regression of $\Delta y(n)$ onto the 2015–2100 trend in global-mean surface temperature $\Delta\mathrm{TAS}(n)$ across ensemble members. Then, the TAS-congruent trend in y for ensemble member n is $R \times \Delta\mathrm{TAS}(n)$, and the residual trend in y, after TAS removal, is

$$\Delta y_{\mathrm{residual}}(n) = \Delta y(n) - R \times \Delta\mathrm{TAS}(n). \quad (1)$$

3. Identification of key drivers of twenty-first-century U.S. wintertime precipitation trends

a. Motivation

We begin by motivating the importance of the storyline approach. One may suppose that models with higher climate sensitivities will typically also have larger precipitation trends, simply because the amount of water vapor in the atmosphere will increase with rising global temperatures. If global temperature were in fact the only major factor determining the magnitude of twenty-first-century precipitation trends, then we could simply compare models by climate sensitivity, and the storylines approach would be unnecessary. This is not the case, as we illustrate below.

Figure 1 shows the magnitudes of the 2015–2100 DJF precipitation trends for the southwestern United States for each of the 176 CMIP6 ensemble members, as a function of the global-mean surface temperature trend. The ensemble members with the largest and smallest precipitation trends for the Southwest do not even come close to matching those with the largest and smallest global-mean temperature trends. In fact, the largest and smallest precipitation trends correspond to very similar global-mean surface temperature trends (3.1 and 2.8 K, respectively). Even if we ignore these as outliers, the overall relationship between the global-mean surface temperature trend and the Southwest precipitation trend has a correlation of essentially zero. Some other regions of the United States do have somewhat larger correlations than in the Southwest, but even there, the global-mean surface temperature trend typically accounts for less than 25% of the variance in regional-mean precipitation trends across the full ensemble. Hence, global-mean surface temperature is only one of several drivers linked to differences in twenty-first-century U.S. precipitation trends.[1] Our first task is to identify some of these other drivers.

[1] Note, however, that the precipitation trend may still be largely *caused* by warming—perhaps via the influence of warming on other drivers we discuss below—even though it does not *scale* with warming. See section 5.

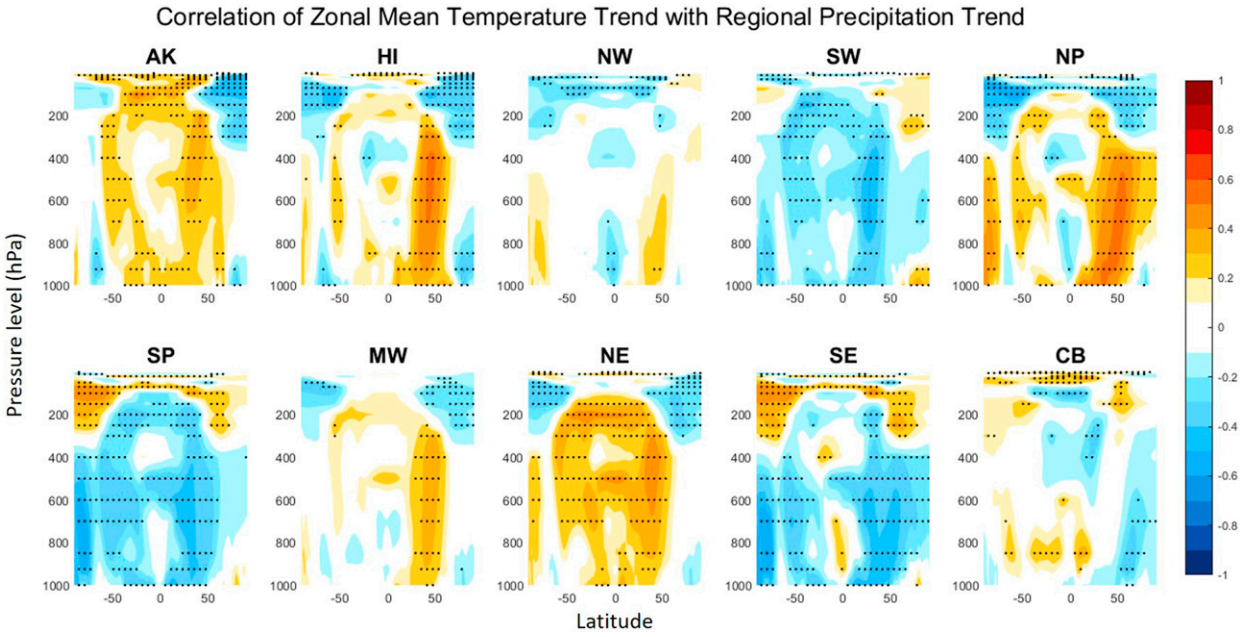

FIG. 2. Correlations of DJF-season zonal-mean air temperature trend at each latitude and pressure level with the regional-mean precipitation trend for each of the 10 regions. The linear regression onto the TAS trend has been removed from all other trends. Stippling indicates correlations that are significant at the $p < 0.05$ level using a two-tailed t test.

b. *Identification of zonal mean drivers*

Based on the results of previous studies, we could search for drivers by simply testing various indices of climate variability (El Niño–Southern Oscillation, the Pacific–North America pattern, etc.) known to significantly impact precipitation variance over the United States. However, this approach could bias the results, since some of the key drivers of twenty-first-century precipitation trends might not be among these well-known modes of variability. To avoid making any such a priori assumptions, we instead plot, for each region, the correlation between the twenty-first-century precipitation trends and the twenty-first-century trends in two key climate variables: zonal-mean temperature in Fig. 2 and 500-hPa geopotential height in Fig. 3. The correlations with zonal-mean zonal wind and sea surface temperature (SST) are also shown in the online supplemental material. To construct these figures, we have removed the components of the trends in all fields that are linearly congruent with global-mean surface temperature [via Eq. (1)], so as to focus on the variance in trends across CMIP6 models that is uncorrelated with global-mean surface temperature, which can be substantial in many cases (as shown in Fig. 1). We can then examine these plots to identify whichever processes stand out, regardless of whether they are already well-known modes of variability. We emphasize that the correlations in these figures are correlations of the *trends* in the variables with the *trends* in regional-mean precipitation, which can differ from correlations based on monthly or interannual variability.

Figure 2 shows the correlation between the twenty-first-century zonal-mean air temperature trend and each region's

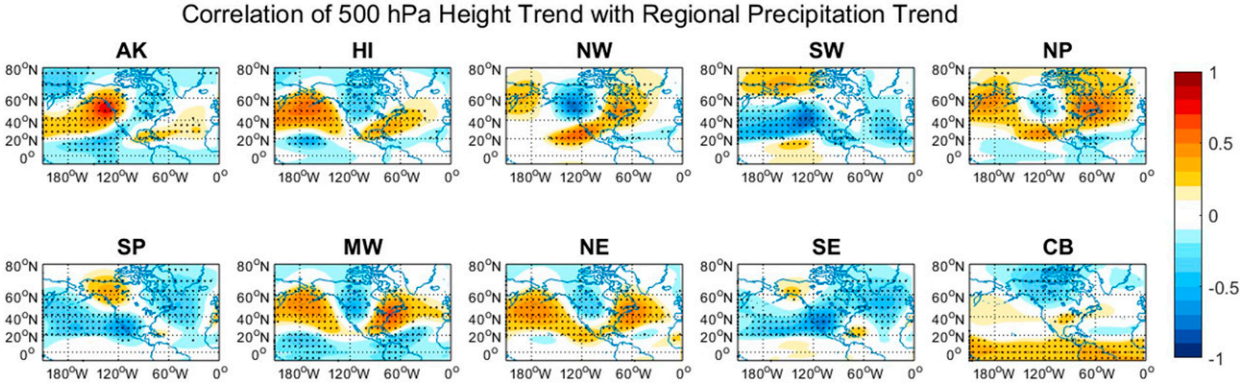

FIG. 3. As in Fig. 2, but for the 500-hPa height trends.

twenty-first-century precipitation trend across all 176 CMIP6 ensemble members. Two features are readily apparent. First, tropical upper-tropospheric warming is positively correlated with precipitation in Alaska and the Northeast, and negatively correlated with precipitation in the Southwest and Southern Plains. That is, individual models or ensemble members with stronger warming in the tropical upper troposphere tend to also have more positive precipitation trends in Alaska and the Northeast, and more negative precipitation trends in the Southwest and Southern Plains.

Second, the correlation maps for several regions (Alaska, Hawaii, Northern Plains, Northeast, and to a lesser extent the Midwest) show a warming of the tropical upper troposphere and a cooling of the polar lower stratosphere, and some other regions show the reverse pattern (Southwest, Southern Plains, Southeast). This changing temperature gradient in the upper troposphere/lower stratosphere is consistent with a change in the strength of the stratospheric polar vortex via thermal wind balance. A similar analysis using zonal-mean zonal wind instead of temperature confirms this more directly (see Fig. S2).

These two features have been studied before: in their analysis of European climate change, Zappa and Shepherd (2017) identified three zonal-mean drivers of twenty-first-century regional climate trends (see also Manzini et al. 2014; Garfinkel et al. 2020): tropical upper-tropospheric temperature, the strength of the stratospheric polar vortex, and Arctic lower-tropospheric temperature. These drivers represent well-known features of the climate change response that vary across models and are known to contribute to substantial uncertainty in future projections of the extratropical storm tracks (Butler et al. 2010; Harvey et al. 2014). The first two (tropical upper-tropospheric temperature and the strength of the stratospheric polar vortex) correspond well to the two features we identified in the correlation maps in Fig. 2. However, we note that there is little evidence of an Arctic lower-tropospheric warming signal in Fig. 2. For now, we include Arctic amplification for the sake of completeness, but as we will show below, its contribution is relatively small [as also found by Zappa and Shepherd (2017)]. Thus, based on Fig. 2 and the results of Zappa and Shepherd (2017), we have identified three candidate drivers that may explain some of the variability in twenty-first-century U.S. winter precipitation trends across CMIP6 ensemble members.

c. *Identification of zonally asymmetric drivers*

One weakness of the analysis in Fig. 2 is that, by construction, it can only help us to discover zonal-mean phenomena connected to precipitation trends. Hence, we show in Fig. 3 the correlations between the twenty-first-century 500-hPa geopotential height trends and each region's twenty-first-century precipitation trend. In this case, two zonally asymmetric patterns emerge.

First, a clear wave pattern is visible in the Hawaii, Midwest, and Northeast panels, which strongly resembles the Pacific–North America (PNA) pattern. It is noteworthy that—according to this figure—this PNA-like pattern has a stronger correlation with precipitation trends in the Northeast than on the west coast. This strong downstream effect of the PNA is consistent with the observational findings of Leathers et al. (1991), Notaro et al. (2006), and Grise et al. (2013).

Second, a wave pattern out of phase with the PNA is clearly visible in the correlation maps for Alaska and the Northwest, and more vaguely apparent in some other regions. This second pattern does not correspond to any well-known mode of 500-hPa geopotential height variability, but it does resemble the second principal component from an analysis of daily zonal wind data by Athanasiadis et al. (2010). We refer to this pattern as the east Pacific dipole (EP).

A notable absence from Fig. 3 is any clear correlation of the regional precipitation trends with the North Atlantic Oscillation (NAO). This absence is confirmed by an analysis of sea level pressure—the more usual variable for defining the NAO (not shown). This is again consistent with the results of Grise et al. (2013), who found that Pacific variability has a greater influence than the NAO on wintertime storm track variability throughout most of North America, with the exception of the Canadian maritime provinces. Other analyses have similarly shown that the strongest impacts of the NAO on precipitation are found in Europe and eastern Canada, rather than in the United States (see Fig. 16 of Hurrell 1995, and Fig. 4c of Hurrell et al. 2003).

A similar analysis of correlations between twenty-first-century SST trends and regional precipitation trends shows a fairly clear El Niño–Southern Oscillation (ENSO) signal in several regions, especially the Southwest, Southern Plains, and Southeast (see Fig. S3). These correlations are to be expected, as these are the same regions in which ENSO has a strong influence on the interannual variability of precipitation according to observations (see, e.g., Fig. 7 of Trenberth and Caron 2000). The sign of the correlation between the ENSO trends and the regional precipitation trends is also consistent with observations, as El Niño events are associated with greater precipitation in these regions.

d. *Definitions of indices for the drivers*

Hence, we have identified six key drivers (other than global-mean temperature itself) whose twenty-first-century trends are potentially linked to U.S. precipitation trends. Three of these are zonally symmetric, and three are zonally asymmetric. Following Zappa and Shepherd (2017) we define indices representing the three zonally symmetric drivers as follows:

1) The tropical upper-tropospheric temperature (TUT) is defined as the zonal-mean air temperature at the 250-hPa level, averaged between 30°S and 30°N. This index captures the amplified tropical upper-tropospheric warming in response to increasing surface temperatures, as the tropical lapse rate is approximately moist adiabatic.
2) The stratospheric polar vortex (SPV) is defined as the zonal-mean zonal wind at the 20-hPa level, averaged between 70° and 80°N. This index captures a weakening of the SPV during the winter season seen in many (but not all) climate models in response to increasing greenhouse gases, which subsequently extends downward and impacts the tropospheric circulation (Karpechko and Manzini 2012; Manzini et al. 2014).
3) Arctic amplification (AA) is defined as the zonal-mean air temperature at the 850-hPa level, averaged between 60° and

90°N. This index captures the amplified Arctic surface temperature warming that occurs in conjunction with increasing global surface temperatures (e.g., Pithan and Mauritsen 2014).

Similar indices were also used by Manzini et al. (2014) and Garfinkel et al. (2020). Indices representing the three zonally asymmetric drivers are defined as follows:

4) ENSO is defined using the Niño-3.4 index: monthly mean SST averaged over the Niño-3.4 region (5°S–5°N, 120°–170°W.
5) The Pacific–North America pattern is defined using the Wallace and Gutzler (1981) station-based index. That is, we use 500-hPa height departures from zonal mean (Z^*) at the following four locations: Z_1^*(20°N, 160°W), Z_2^*(45°N, 165°W), Z_3^*(55°N, 115°W), and Z_4^*(30°N, 85°W). The PNA index is defined as

$$\mathrm{PNA} = \left(\frac{1}{4}\right)(Z_1^* - Z_2^* + Z_3^* - Z_4^*). \tag{2}$$

The PNA can also be defined using the first empirical orthogonal function (EOF) of 500-hPa heights in the region 15°–70°N, 180°–60°W, similar to the definitions used by Chen et al. (2018) and Yeh et al. (2018). The correlation between the trends in the EOF-based and station-based PNA indices across ensemble members is $r = 0.91$, so the two definitions are nearly interchangeable. We use the station-based index for the sake of simplicity.

6) The east Pacific dipole (EP) is defined using a station-based index similar to that for the PNA. The two locations are chosen to be midway between the two northernmost and two southernmost of the PNA locations, respectively: Z_5^*(50°N, 140°W), and Z_6^*(25°N, 122.5°W). The EP index is defined as

$$\mathrm{EP} = \left(\frac{1}{2}\right)(Z_5^* - Z_6^*). \tag{3}$$

The placement of the stations ensures that the pattern of variability captured by the EP will be distinct from the PNA. The EP pattern can also be captured by an EOF analysis using daily zonal wind data (see Athanasiadis et al. 2010), but it is less apparent in the monthly geopotential height data that we use here. Therefore, we use the station-based definition listed above.

In addition to these six indices, we also consider the global-mean TAS, giving a total of seven candidate indices whose differing trends across models may help to explain the differing regional precipitation trends.

The twenty-first-century trends in the other six indices listed above may be closely related to the concurrent trends in the global-mean surface temperature. For example, models with higher climate sensitivity would be expected to have larger tropical upper-tropospheric and Arctic surface temperature warming. Hence, we use Eq. (1) to subtract the regression onto TAS from the other trends. (Zappa and Shepherd 2017 accomplished a similar goal by normalizing their indices by TAS trends.) We then use the residual trends in all six indices in the analyses described below.

In the case that the other indices were strongly correlated with each other, we might need to repeat this process by using Eq. (1) to sequentially remove the influence of each index from the next, in a manner similar to that used in the partial least squares regression method (Wold 1966; Wallace et al. 2012; Black et al. 2017). However, in practice the pairs of indices we will end up using together are already poorly correlated (see Tables S2 and S3).

To double-check that these indices really do capture the patterns from Figs. 2 and 3, we also show the correlations of the twenty-first-century trends in zonal-mean temperature and 500-hPa geopotential height with the trends in the indices themselves (Figs. 4 and 5). From Fig. 4, it is clear that the temperature patterns associated with TUT and SPV do in fact closely resemble the patterns we found in the precipitation correlations from Fig. 2. Figure 5 shows that, likewise, the 500-hPa geopotential height patterns associated with PNA and EP resemble the patterns we found in the precipitation correlations from Fig. 3. Note that there are some similarities between the patterns in Figs. 4 and 5 for some of the indices (e.g., PNA and SPV), suggesting a possible correlation between these indices which we will revisit below.

e. *Characterizing the relative importance of the drivers*

Now that we have identified six candidate indices, plus the TAS trend itself, we can plot the correlations of these indices with the twenty-first-century precipitation trends in each of the 10 NCA regions to summarize the most important drivers of the variance in precipitation trends in each region. These correlations are shown in the bar charts in Fig. 6. Note that we have excluded panels for the SPV and AA indices from this figure (they are shown for reference in Fig. S4). We exclude the SPV index because it is largely redundant with the PNA index, as nonnegligible correlations between SPV trends and PNA trends exist across the CMIP6 model ensemble ($r = -0.53$; Table S2). In fact, after we regress the PNA trends out of the SPV trends using Eq. (1), the residual SPV trends explain little of the precipitation trends (not shown). We exclude the AA index because its trends explain very little of the precipitation trends (Fig. S4), except for modest correlations in the Southeast and Caribbean, the former of which may be due to the anti-correlation between AA and ENSO trends across models ($r = -0.43$; see Table S2). Thus, given the weak roles of SPV and AA, we narrow our list to five indices (TAS, TUT, ENSO, PNA, and EP) for the remainder of this paper.

We also show 90% confidence intervals for the correlation coefficients in Fig. 6. To construct each confidence interval, we randomly select half of the ensemble members and recalculate the correlation for that subsample. We repeat this procedure 10 000 times and use the 5th and 95th percentiles of the resulting spread in subsampled correlations as the endpoints of the confidence interval. We do not apply this procedure to the TAS trends, because in this case the width of each confidence interval is near zero. Note that the caveats from the introduction about interpreting confidence intervals probabilistically apply here as well: while these confidence intervals can be usefully compared with each other, they should not be interpreted as literal probabilities.

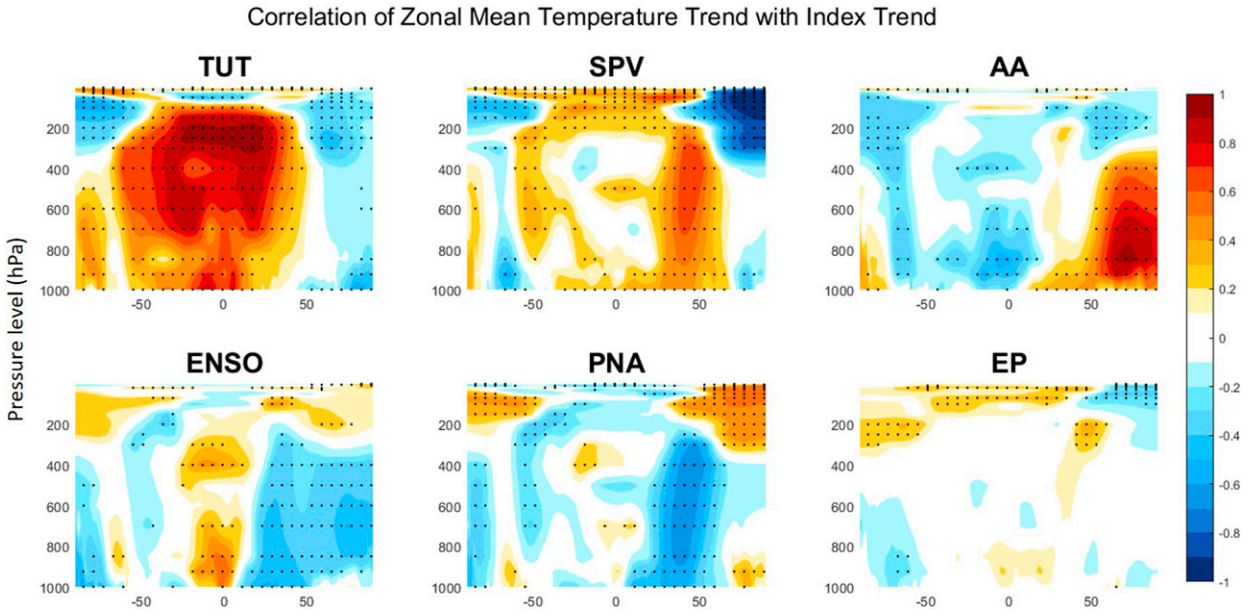

FIG. 4. Correlations of DJF-season zonal-mean air temperature trends with the various index trends, for comparison with Fig. 2. The linear regression onto the TAS trends has been removed from all other trends. Stippling indicates correlations that are significant at the $p < 0.05$ level using a two-tailed t test.

By construction, the results in Fig. 6 summarize those already shown in Figs. 2 and 3, but in a more concise format. For example, according to the bar charts in Fig. 6, the PNA index is most closely linked to precipitation trends in Hawaii, the Midwest, and the Northeast. This is consistent with our discussion of 500-hPa heights from Fig. 3, which showed that these were the same three regions in which the PNA-like pattern was most clearly visible.

From Fig. 6 and Fig. S4, it is immediately apparent that the twenty-first-century trends in the global-mean temperature (green) and the zonally asymmetric indices (ENSO, PNA, EP; red) are generally better correlated with twenty-first-century

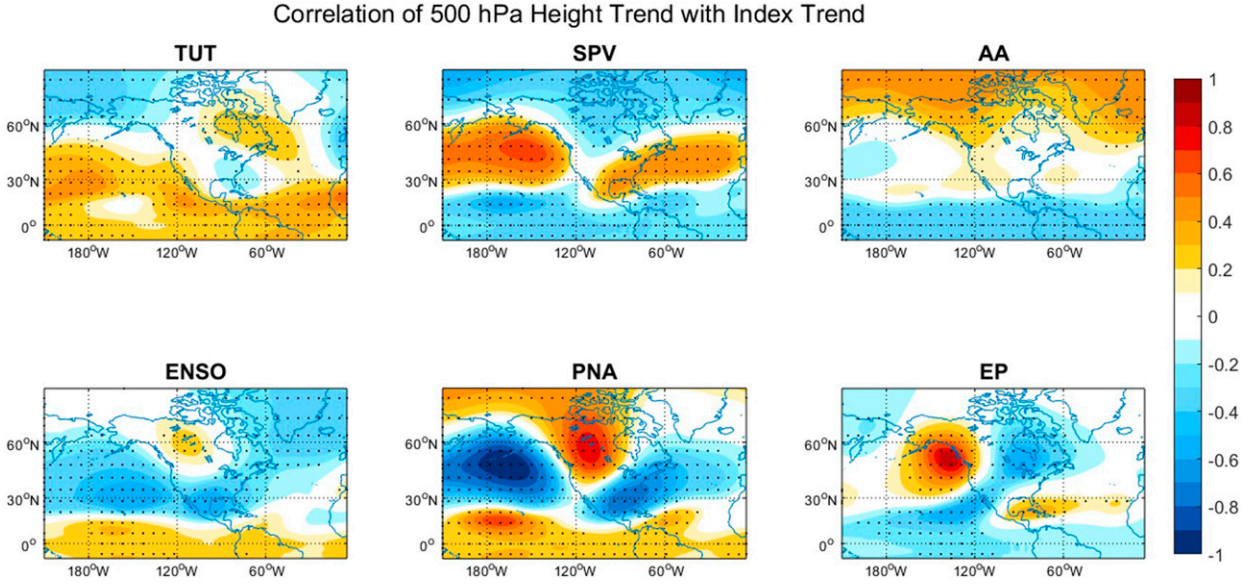

FIG. 5. Correlations of DJF-season 500-hPa height trends with the various index trends, for comparison with Fig. 3. The linear regression onto the TAS trend has been removed from all other trends. Stippling indicates correlations that are significant at the $p < 0.05$ level using a two-tailed t test.

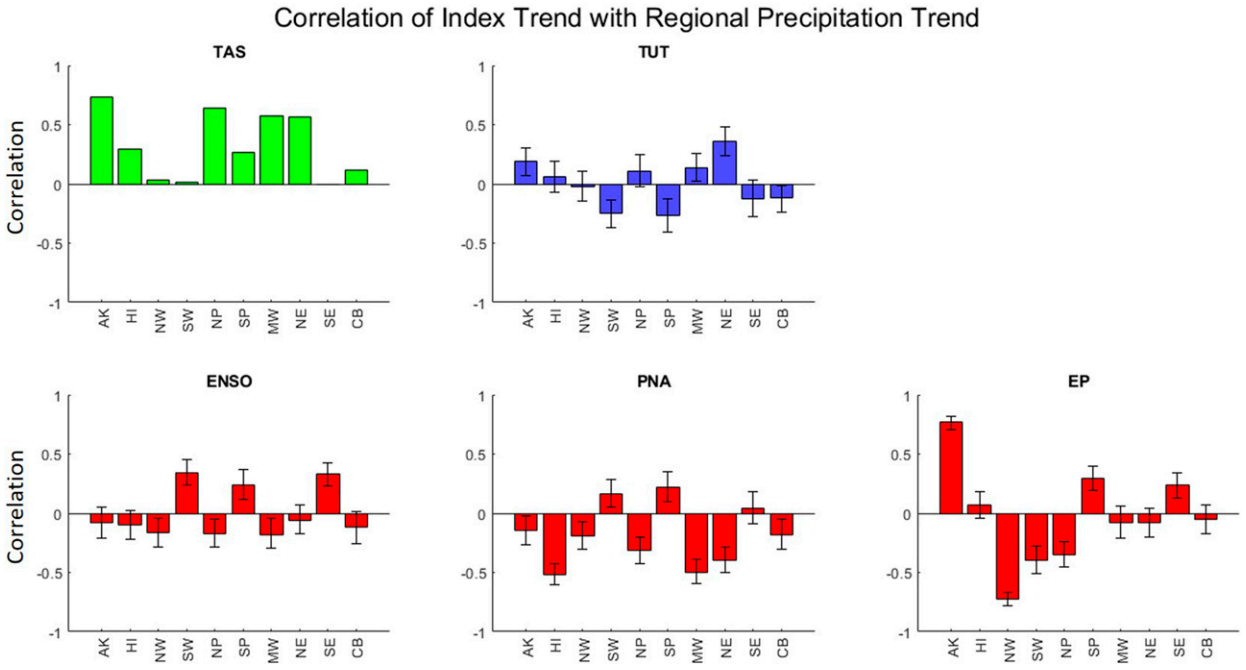

FIG. 6. Correlations between index trends and regional-mean precipitation trends, across all 176 CMIP6 ensemble members. The linear regression onto the TAS trend has been removed from all other trends. Error bars are based on a bootstrapping method described in the text (not applicable to TAS). DJF season only.

regional precipitation trends over the United States than are trends in the zonal-mean indices (TUT, SPV, AA; blue). This is in contrast to—but not in disagreement with—the findings of Zappa and Shepherd (2017) that these zonal-mean indices were quite useful for understanding European precipitation. This contrast illustrates the point made by Garfinkel et al. (2020) that zonal-mean indices are less useful in understanding local precipitation trends over the United States than in many other regions. The results in Fig. 6 emphasize that the storylines approach requires the identification of physical drivers unique to the region of interest. Nonetheless, at least one of the zonal-mean indices of Zappa and Shepherd (2017) does have some explanatory value for the twenty-first-century U.S. precipitation trends. Specifically, trends in the TUT index are significantly correlated with precipitation trends in the northeastern United States ($r = 0.36$; see also Fig. 2).

After excluding SPV and AA, the trends in the remaining indices are generally poorly correlated with each other ($|r| \leq 0.20$), so we will treat them as independent predictors in the storyline analysis in the next section. One exception is the relationship between PNA and ENSO, which—perhaps unsurprisingly—have trends correlated at $r = 0.40$. (See Tables S2 and S3 for additional details on index correlations.) This could be taken as a reason to omit one of the two, but if we remove the influence of either ENSO or the PNA on the other using Eq. (1), the residual remains noticeably correlated with precipitation trends (not shown). Thus, while we will need to treat this particular pair of indices carefully, we will keep both. Also note that while TUT and ENSO are strongly correlated in terms of monthly variability—as should be expected due to the dynamics of ENSO (Table S3)—their trends are not strongly correlated across the full ensemble (Table S2).

Before proceeding to the construction of storylines, several caveats are worth mentioning. If some component of the precipitation trend were driven by ENSO, for example, but was the same across all ensemble members, this analysis would miss it. What we are accounting for here is not the degree to which ENSO is responsible for the mean trend, but the degree to which ENSO explains differences across models and ensemble members.

It is also worth noting that this is a model analysis, and we have argued that differences in *modeled* precipitation trends are due in part to differences in *modeled* trends in drivers such as the PNA. One may question, however, whether the model's representation of these drivers is biased with respect to observations. To assess how well these drivers are represented in models, we examine detrended and deseasonalized month-to-month variability in the models, which can be directly compared with observations. Figures S5–S8 show the correlations of month-to-month variability in 500-hPa height and zonal-mean temperature fields with month-to-month variability in the driver indices in both CMIP6 models and in ERA5 reanalysis data. In other words, we repeat the analysis shown in Figs. 4 and 5, but for month-to-month variability in both observations and models. Overall, the resulting correlation patterns are very similar in observations and models, suggesting that the models possess a reasonable representation of these drivers.

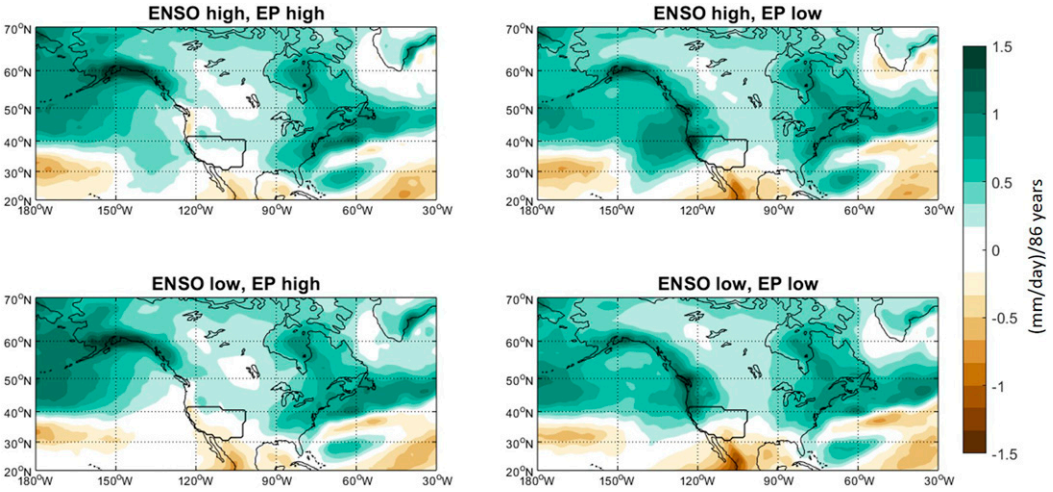

FIG. 7. Storylines for DJF-season precipitation. The storylines are constructed from those indices that are most important in the southwestern United States (outlined), but the resulting precipitation trends are shown across a larger domain.

4. Constructing storylines of twenty-first-century wintertime precipitation change

Having identified five key drivers of twenty-first-century precipitation trends over the United States, we can now construct storylines. Following Zappa and Shepherd (2017), we estimate the combined influence of two or more of the above drivers (as listed in section 3) on twenty-first-century U.S. precipitation trends as follows. If, for example, we consider the ENSO and EP indices, the "high ENSO, high EP" storyline is based on the combination of the 80th-percentile ENSO trend (i.e., the ENSO trend in the 80th-percentile of the full 176-member ensemble) and the 80th-percentile EP trend. The synthetic storyline precipitation trend $\Delta P_{\text{storyline}}$ consists of the ensemble-mean precipitation trend ΔP_{mean} plus the additional precipitation trend due to the high ENSO trend and a second additional precipitation trend due to the high EP trend. We calculate this at each location on the map as follows:

$$\Delta P_{\text{storyline}} = \Delta P_{\text{mean}} + a_{\text{ENSO}}(\Delta\text{ENSO}_{80} - \Delta\text{ENSO}_{\text{mean}}) + a_{\text{EP}}(\Delta\text{EP}_{80} - \Delta\text{EP}_{\text{mean}}). \quad (4)$$

Here a_{ENSO} is the slope of the linear regression of the precipitation trend at this point onto the ENSO trend, ΔENSO_{80} is the 80th-percentile ENSO trend, and $\Delta\text{ENSO}_{\text{mean}}$ is the mean ENSO trend of the full 176-member ensemble. The quantities in the last term for EP are defined similarly. A "high ENSO, low EP" scenario would be defined the same way, except that the EP trend would then be in the 20th percentile. Other combinations are defined similarly. We subtract the linear regression onto TAS from all other trends before performing this analysis [as shown above in Eq. (1)]. We have assumed (as did Zappa and Shepherd in their analysis) that the various drivers we are studying combine linearly. We also assume that the various indices are nearly uncorrelated with one another, an assumption which we verified above in section 3, except for the stronger correlation between the PNA and ENSO.

As a proof of concept, we show example storylines in this section for winter precipitation trends in the Southwest and Midwest, but this method can be readily applied to other regions, other seasons, other emissions scenarios, or variables other than monthly mean precipitation.

Figure 7 shows four storylines of twenty-first-century precipitation trends for the Southwest. Here, for simplicity, we use only the two indices that explain the most variance in this region: ENSO and EP (see Fig. 6). All four storylines show wetting trends at high latitudes and drying trends in the subtropics, consistent with the multimodel-mean trend from CMIP6 models (Fig. S9; see also Fig. 2 of Cook et al. 2020). However, by construction, there are noteworthy differences in precipitation trends in the southwestern United States among the four panels. The largest differences are between the left (higher-than-average EP trend) and right (lower-than-average EP trend) sets of panels, showing the dominant influence of the EP trends on this region. However, the top (higher-than-average ENSO trend) and bottom (lower-than-average ENSO trend) sets of panels do also differ noticeably, demonstrating the smaller but substantial impact of ENSO trends. We can see from the bar charts in Fig. 6 that ENSO and EP have opposite-sign correlations with Southwest precipitation trends, so the strongest impacts should occur when these indices have oppositely signed trends so that their associated precipitation anomalies can reinforce each other. This is indeed the case, and the contrast between the strong wetting trend in the top right panel (high ENSO trend, low EP trend) and the slight drying trend in the bottom-left panel (low ENSO trend, high EP trend) of Fig. 7 is striking.

The differences among the storylines in Fig. 7 are more clearly seen when the ensemble-mean trend is removed (Fig. 8). From this figure, we can see that high EP trends are always associated with a drier-than-average trend on the west coast, but this dry region shifts to the north or south depending on whether we consider the high ENSO or low ENSO scenario, respectively. This is consistent with El Niño events causing

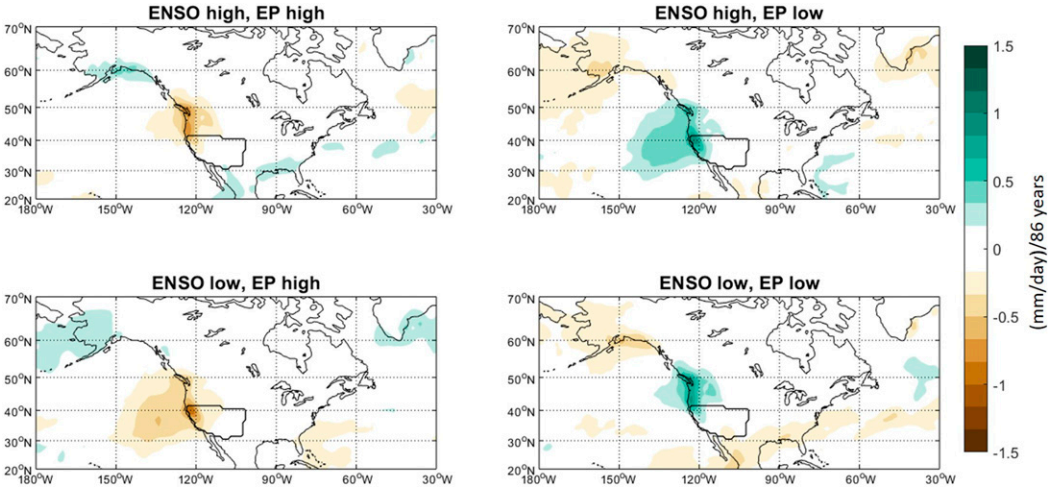

FIG. 8. As in Fig. 7, but with the ensemble-mean precipitation trend subtracted.

wetter-than-average conditions in the Southwest and drier-than-average conditions in the Northwest (Dai and Wigley 2000). Similarly, the low EP scenario is always associated with a wetter-than-average precipitation trend for the west coast, but the wettest region shifts north or south depending on the ENSO scenario. Note that these same storylines could be used to study Northwest precipitation trends, since the same two indices are important there.

Figures 9 and 10 show a second set of storylines, this time optimized for the Midwest, where the key drivers are TAS and PNA (see Fig. 6). In the Midwest, TAS is positively correlated with the precipitation trend, and PNA is negatively correlated with the precipitation trend, so these two drivers again have competing influences. Accordingly, the top-right panel (high TAS trend, low PNA trend) is the wettest of the four storylines, and the bottom-left panel (low TAS trend, high PNA trend) is the driest. Note, however, that none of these storylines involve actual drying in the Midwest, in contrast to the situation for the Southwest (Figs. 7 and 8). Another contrast between the Southwest and Midwest is that in the Southwest EP is clearly the more important of the two drivers, whereas in the Midwest the magnitudes of the TAS and PNA impacts are similar. This can be seen more clearly in Fig. 10, where the impacts of the TAS and PNA trends on Midwest precipitation almost exactly cancel in the top-left and bottom-right panels, giving a precipitation trend nearly the same as the multimodel ensemble mean (seen as zero in Fig. 10, since the mean has been subtracted). While the storylines in Figs. 9 and 10 were constructed with the Midwest in mind, that they are also relevant for Northeast precipitation, except that in the Northeast the TUT trend (not shown in Figs. 9 and 10) is also important.

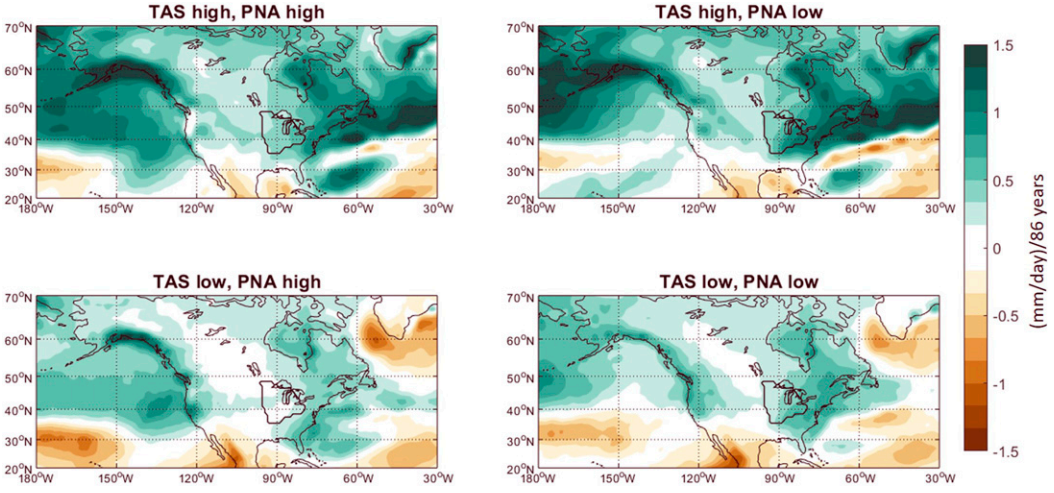

FIG. 9. Storylines for DJF-season precipitation. The storylines are constructed from those indices that are most important in the midwestern region of the United States (outlined), but the resulting precipitation trends are shown across a larger domain.

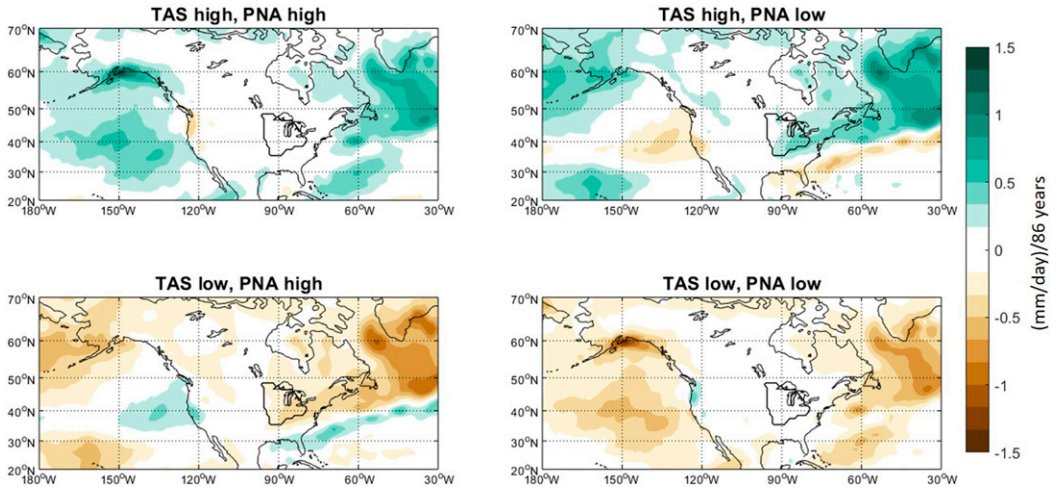

FIG. 10. As in Fig. 9, but with the ensemble-mean precipitation trend subtracted.

One takeaway message from this analysis is the striking difference between the top-right and bottom-left panels of either Fig. 7 or Fig. 9 and the associated wide spread in plausible precipitation outcomes for these regions. A second is the observation that much of this spread is due to the EP (for the Southwest) and the PNA (for the Midwest). The EP and PNA are both standing wave patterns, and such patterns could be modified by climate change, or could simply change randomly with the internal variability of the atmosphere. It therefore becomes important to know whether the EP and PNA are forced by warming. If they are not, then most of the uncertainty in Southwest precipitation trends, and much of the uncertainty in Midwest precipitation trends, as represented in Figs. 7–10, cannot be eliminated. In the next section, we therefore turn our attention to the partitioning of variance into forced and unforced components.

5. Partitioning variance of trends into forced and unforced components

For a single-model ensemble, one can easily decompose trends into forced and unforced components. In this case, the forced trend is the ensemble-mean trend, and the trend due to internal variability for each ensemble member is the deviation from this mean (see, e.g., Deser et al. 2014).

In a multimodel large ensemble, one could similarly divide the trends of some variable into two categories: 1) forced trends, defined by the various single-model ensemble means, and 2) trends due to internal variability, defined by the deviations of trends within single-model ensembles from their respective means. One might naively assume that the forced trends for each model would scale closely with the TAS trend (essentially climate sensitivity) in each model, and that trends due to internal variability would be independent of it.

However, this partitioning misses an important fact: dynamical processes can be forced by increasing greenhouse gases, but still not scale with TAS trends, as noted by Grise and Polvani (2016). That is, we cannot equate *forced* trends with *TAS-congruent* trends. Hence, the more appropriate partitioning of the variance in a multimodel ensemble will in fact need *three* categories: 1) forced trends which scale with the global-mean surface temperature trend, 2) residual forced trends which do not scale in this way, and 3) trends due to internal variability. We define these three categories as follows.

First, we identify the total forced trend (whether it scales with TAS or not) for each model. For a quantity X (e.g., precipitation in the Southwest), the total forced trend for a single model can be estimated by the single-model ensemble mean trend. To set the notation, let $\Delta X(m, n)$ be the trend of the quantity X for ensemble member n of model m. For our dataset, there are 34 models, and each model has anywhere from 1 to 50 ensemble members, so m ranges from 1 to 34, and for each particular value of m, the index n has a range that depends on the size of that single-model ensemble. Let $[\Delta X(m,:)]$ be the mean of ΔX across all ensemble members from model m. This is the total forced trend in X for model m.

Next, we separate the total forced trend $[\Delta X(m,:)]$ into components that do or do not scale with TAS. Let b be the slope of the linear regression of $[\Delta X(m,:)]$ onto $[\Delta\mathrm{TAS}(m,:)]$ and let a be the y intercept of the same regression. (We calculate this as a *weighted* regression, in which each model's mean is repeated according to the number of ensemble members. This is necessary for the partitioning of variance, as explained below.) The regression slope b can be thought of as the sensitivity of the forced X trend to the TAS trend.

We may now define the forced trend that scales with the TAS trend as

$$\Delta X_1(m,n) = a + b \times [\Delta\mathrm{TAS}(m,:)]. \tag{5}$$

The residual forced trend is then simply the full forced trend minus the quantity above:

$$\Delta X_2(m,n) = [\Delta X(m,:)] - \Delta X_1(m,n). \tag{6}$$

Finally, the trend due to internal variability, for each ensemble member of each model, is the original trend $\Delta X(m, n)$ minus both of the forced trends defined above:

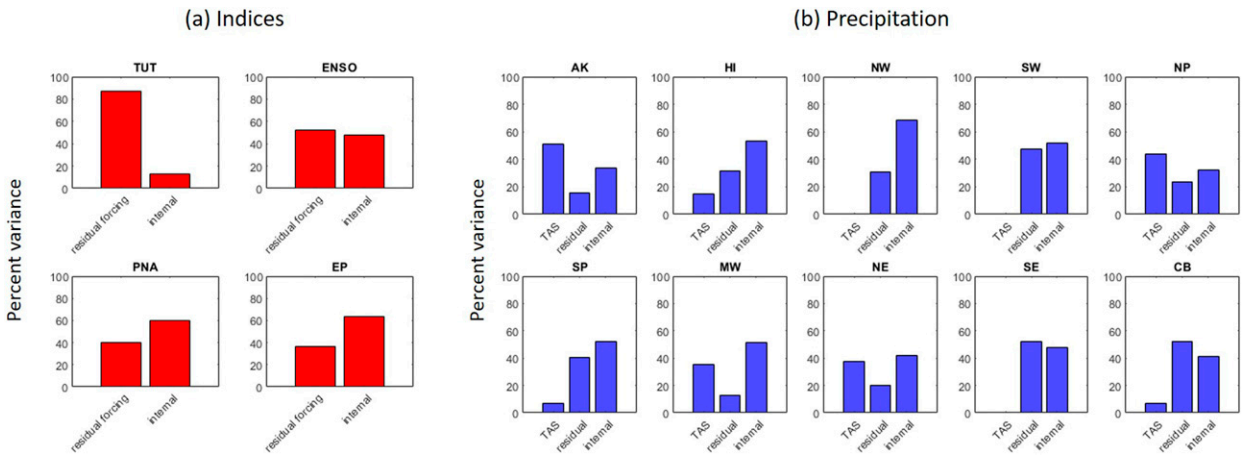

FIG. 11. Partitioning of the variance in (a) index trends and (b) precipitation trends across the ensemble into the categories described in section 5.

$$\Delta X_3(m,n) = \Delta X(m,n) - \Delta X_1(m,n) - \Delta X_2(m,n). \quad (7)$$

These definitions are chosen so that, for each model m and each ensemble member n, the three trends will sum to give the original full trend $\Delta X(m,n)$. That is, $\Delta X_1(m,n) + \Delta X_2(m,n) + \Delta X_3(m,n) = \Delta X(m,n)$ for each m and each n. We have arbitrarily included the y intercept (a) as part of X_1, but this is immaterial since we will only use these three quantities via their variances. Because single-model ensemble means are required for this partitioning, and because models with very few ensemble members have poorly constrained ensemble means, we restrict the analysis—for this section only—to those CMIP6 models with at least five ensemble members each. This leaves a total of 142 ensemble members from 13 models. The results in this section are qualitatively similar for different cutoff values (other than 5) for the minimum number of available ensemble members per model.

Having partitioned the trends, we can now also partition the variances. The variance in ΔX explained by the forcing that scales with the TAS trend is simply the variance of ΔX_1, the residual forced variance is the variance in ΔX_2, and the variance due to internal variability is the variance of ΔX_3. The weighting used here, in which each single-model ensemble mean is repeated in accordance with the number of ensemble members, is necessary for this partitioning to work. If we simply used each single-model ensemble mean once, then the variances due to TAS forcing and residual forcing would each be computed as a variance of 13 individual quantities, whereas the variance due to internal variability would be computed as a variance of 142 individual quantities. In this case, the variances would not add to 100% (not shown).

Figure 11a shows the results of this partitioning for trends in four drivers of twenty-first-century U.S. precipitation trends identified in the previous sections: TUT, ENSO, PNA, and EP. The trends in the fifth driver (TAS) are, by definition, entirely correlated with global-mean surface temperature warming. Note that, for this part of the analysis, the first component (variance explained by TAS forcing) is zero since we always use index trends with the regression to the TAS trend removed (as explained in section 3d). If we did not remove the regression to the TAS trend, most of the variance in the twenty-first-century TUT trends would be associated with TAS warming (not shown), as would be expected since tropical upper-tropospheric temperature is linked by convection to tropical surface temperature, which is in turn linked to the global-mean surface temperature.

Figure 11a shows that, after removing the TAS trend from each index, the variance in TUT trends across models is mostly due to residual forcing, with only a small contribution from internal variability. In contrast, the variance of the twenty-first-century ENSO trends is roughly evenly divided between residual forcing and internal variability, and the variance in the PNA and EP trends is primarily due to internal variability. This last point is significant for the Southwest region, as it shows that the most important process we have identified in that region (Figs. 6–8) is mostly unpredictable.

Figure 11b show the results of the same decomposition, but this time applied to the regional precipitation trends. In this case, we have not removed TAS, so there are still three categories of variance. Results differ between regions, but overall these charts show that each of the three categories of variance is important in some regions, with internal variability being the only category that is reasonably large in all 10 regions. We thus conclude that internal variability is important for both the precipitation trends and the trends in most of our identified driver indices (with the exception of TUT).

For the Southwest in particular, Fig. 11b shows that essentially none of the variance in the twenty-first-century precipitation trends can be attributed to forcing that directly scales with TAS (consistent with Fig. 1). Instead, approximately half the variance is due to forcing that does not scale with TAS, and the remainder is due to internal variability. For the Midwest, forcing that scales with TAS is more important, but internal variability is still the largest contributor to the spread in precipitation trends. Thus, projections of precipitation in both of these two regions will always involve a considerable range of

possible outcomes, regardless of future model improvements. This underscores the need to consider either the full spread of modeled trends or the spread of synthetic trends calculated from a method such as the storylines method (as shown in section 4). The additional benefit of the latter is that it identifies some of the drivers responsible for the model differences.

6. Conclusions

Projections of future U.S. winter precipitation trends differ dramatically among CMIP6 models, and even among ensemble members for the same model. Hence, the use of the multimodel-mean precipitation trend obscures large uncertainties (e.g., through cancellation of opposing trends) and can be a misleading guide for policy makers. Instead, future precipitation trends should be represented in terms of a plurality of possible outcomes. This can be done simply by showing the full range of trends across models. However, the storylines method demonstrated here is another means of quantifying those possibilities, and it has the added benefit of identifying some of the key physical drivers responsible for the differences between possible future outcomes.

We can draw at least two general conclusions from our application of the storylines approach to twenty-first-century U.S. wintertime precipitation trends in CMIP6 models. First, the spread in precipitation trends across models is not exclusively, or even primarily, due to differences in climate sensitivity. Instead, the spread in the precipitation trends is linked to several different dynamic and thermodynamic drivers, which include the global-mean surface temperature trend, but also tropical upper troposphere temperature and teleconnection patterns such as ENSO, PNA, and the east Pacific (EP) dipole. Even when these other physical processes are forced by warming, they have trends that differ across the multimodel ensemble in ways that do not always scale with global-mean temperature.

Second, a considerable amount of internal variability exists, both in the precipitation trends and in the trends of the dynamic or thermodynamic drivers to which they are linked. Unless the models have dramatically, and systematically, overestimated the magnitude of internal variability (which seems unlikely; see Knutson et al. 2013), this means that projected precipitation trends will always have a considerable spread, emphasizing the need to identify a range of plausible outcomes for policy makers.

In closing, we caution against two potential misinterpretations of this method. First, the goal of the storyline method is not to determine which model or ensemble member is more accurate, or to narrow the range of possibilities. Rather, the goal is to quantify the range of possibilities, and to understand the physical processes responsible for the differences. Second, the small values of TAS-forced precipitation trends in some regions, as in the Southwest, should not be construed to suggest that climate change is unimportant in driving precipitation in those regions. Rather, it suggests that the trend due to climate change does not scale with climate sensitivity across models.

This study has served primarily to identify the key drivers of twenty-first-century U.S. wintertime precipitation trends, and to demonstrate the methodology for constructing storylines from them. Follow-up studies can apply this method to focus on specific NCA regions, seasons other than winter, and variables other than precipitation to demonstrate the range of plausible outcomes for each of those contexts.

Acknowledgments. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. We also thank Clara Deser and three anonymous reviewers for helpful comments. This study was supported by NOAA's Climate Program Office's Modeling, Analysis, Predictions, and Projections program (Grant NA19OAR4310293).

Data availability statement. All CMIP6 model output is freely and publicly available from the Lawrence Livermore National Laboratory (https://esgf-node.llnl.gov/search/cmip6/). ERA5 data are available from Copernicus Climate Change Service (Hersbach et al. 2019). Online locations for these data products are provided in the citations to these data sets in the references section below.

REFERENCES

Athanasiadis, P. J., J. M. Wallace, and J. J. Wettstein, 2010: Patterns of wintertime jet stream variability and their relation to the storm tracks. *J. Atmos. Sci.*, **67**, 1361–1381, https://doi.org/10.1175/2009JAS3270.1.

Black, J., N. C. Johnson, S. Baxter, S. B. Feldstein, D. S. Harnos, and M. L. L'Heureux, 2017: The predictors and forecast skill of Northern Hemisphere teleconnection patterns for lead times of 3–4 weeks. *Mon. Wea. Rev.*, **145**, 2855–2877, https://doi.org/10.1175/MWR-D-16-0394.1.

Bladé, I., D. Fortuny, G. J. van Oldenborgh, and B. Liebmann, 2012: The summer North Atlantic Oscillation in CMIP3 models and related uncertainties in projected summer drying in Europe. *J. Geophys. Res.*, **117**, D16104, https://doi.org/10.1029/2012JD017816.

Butler, A. H., D. W. J. Thompson, and R. Heikes, 2010: The steady-state atmospheric circulation response to climate change–like thermal forcings in a simple general circulation model. *J. Climate*, **23**, 3474–3496, https://doi.org/10.1175/2010JCLI3228.1.

Chen, Z., B. Gan, L. Wu, and F. Jia, 2018: Pacific–North American teleconnection and North Pacific Oscillation: Historical simulation and future projection in CMIP5 models. *Climate Dyn.*, **50**, 4379–4403, https://doi.org/10.1007/s00382-017-3881-9.

Cook, B. I., J. S. Mankin, K. Marvel, A. P. Williams, J. E. Smerdon, and K. J. Anchukaitis, 2020: Twenty-first century drought projections in the CMIP6 forcing scenarios. *Earth's Future*, **8**, e2019EF001461, https://doi.org/10.1029/2019EF001461.

Dai, A., and T. M. L. Wigley, 2000: Global patterns of ENSO-induced precipitation. *Geophys. Res. Lett.*, **27**, 1283–1286, https://doi.org/10.1029/1999GL011140.

Deser, C., A. Phillips, V. Bourdette, and H. Teng, 2012: Uncertainty in climate change projections: The role of internal variability. *Climate Dyn.*, **38**, 527–546, https://doi.org/10.1007/s00382-010-0977-x.

——, ——, 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.

——, J. W. Hurrell, and A. S. Phillips, 2017: The role of the North Atlantic Oscillation in European climate projections. *Climate Dyn.*, **49**, 3141–3157, https://doi.org/10.1007/s00382-016-3502-z.

——, and Coauthors, 2020: Insights from Earth system model initial-condition large ensembles and future prospects. *Nat. Climate Change*, **10**, 277–286, https://doi.org/10.1038/s41558-020-0731-2.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, **9**, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016.

Fiedler, S., and Coauthors, 2020: Simulated tropical precipitation assessed across three major phases of the Coupled Model Intercomparison Project (CMIP). *Mon. Wea. Rev.*, **148**, 3653–3680, https://doi.org/10.1175/MWR-D-19-0404.1.

Garfinkel, C. I., O. Adam, E. Morin, Y. Enzel, E. Elbaum, M. Bartov, D. Rostkier-Edelstein, and U. Dayan, 2020: The role of zonally averaged climate change in contributing to intermodel spread in CMIP5 predicted local precipitation changes. *J. Climate*, **33**, 1141–1154, https://doi.org/10.1175/JCLI-D-19-0232.1.

Giorgi, F., and X. Bi, 2009: Time of emergence (TOE) of GHG-forced precipitation change hot-spots. *Geophys. Res. Lett.*, **36**, L06709, https://doi.org/10.1029/2009GL037593.

Grise, K. M., and L. M. Polvani, 2016: Is climate sensitivity related to dynamical sensitivity? *J. Geophys. Res.*, **121**, 5159–5176, https://doi.org/10.1002/2015JD024687.

——, S. Son, and J. R. Gyakum, 2013: Intraseasonal and interannual variability in North American storm tracks and its relationship to equatorial Pacific variability. *Mon. Wea. Rev.*, **141**, 3610–3625, https://doi.org/10.1175/MWR-D-12-00322.1.

Harvey, B. J., L. C. Shaffrey, and T. J. Woollings, 2014: Equator-to-pole temperature differences and the extra-tropical storm track responses of the CMIP5 climate models. *Climate Dyn.*, **43**, 1171–1182, https://doi.org/10.1007/s00382-013-1883-9.

Hawkins, E., and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteor. Soc.*, **90**, 1095–1108, https://doi.org/10.1175/2009BAMS2607.1.

Hazeleger, W., B. van den Hurk, E. Min, G. J. van Oldenborgh, A. C. Petersen, D. A. Stainforth, E. Vasileiadou, and L. A. Smith, 2015: Tales of future weather. *Nat. Climate Change*, **5**, 107–113, https://doi.org/10.1038/nclimate2450.

Held, I. M., and B. J. Soden, 2006: Robust responses of the hydrological cycle to global warming. *J. Climate*, **19**, 5686–5699, https://doi.org/10.1175/JCLI3990.1.

Hersbach, H., and Coauthors, 2019: ERA5 monthly averaged data on pressure levels from 1979 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). Accessed 6 April 2021, https://doi.org/10.24381/cds.6860a573.

——, and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hoerling, M., J. Eischeid, and J. Perlwitz, 2010: Regional precipitation trends: Distinguishing natural variability from anthropogenic forcing. *J. Climate*, **23**, 2131–2145, https://doi.org/10.1175/2009JCLI3420.1.

——, ——, ——, X. Quan, K. Wolter, and L. Cheng, 2016: Characterizing recent trends in U.S. heavy precipitation. *J. Climate*, **29**, 2313–2332, https://doi.org/10.1175/JCLI-D-15-0441.1.

Hurrell, J. W., 1995: Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science*, **269**, 676–679, https://doi.org/10.1126/science.269.5224.676.

——, Y. Kushnir, G. Ottersen, and M. Visbeck, 2003: An overview of the North Atlantic Oscillation. *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*, J. W. Hurrell et al., Eds., American Geophysical Union, 1–35, https://doi.org/10.1029/134GM01.

Karpechko, A. Y., and E. Manzini, 2012: Stratospheric influence on tropospheric climate change in the Northern Hemisphere. *J. Geophys. Res.*, **117**, D05133, https://doi.org/10.1029/2011JD017036.

Knutson, T. R., F. Zeng, and A. T. Wittenberg, 2013: Multimodel assessment of regional surface temperature trends: CMIP3 and CMIP5 twentieth-century simulations. *J. Climate*, **26**, 8709–8743, https://doi.org/10.1175/JCLI-D-12-00567.1.

Leathers, D. J., B. Yarnal, and M. A. Palecki, 1991: The Pacific/North American teleconnection pattern and United States climate. Part I: Regional temperature and precipitation associations. *J. Climate*, **4**, 517–528, https://doi.org/10.1175/1520-0442(1991)004<0517:TPATPA>2.0.CO;2.

Lenderink, G., B. van den Hurk, A. Klein Tank, G. J. van Oldenborgh, E. van Meijgaard, H. de Vries, and J. J. Beersma, 2014: Preparing local climate change scenarios for the Netherlands using resampling of climate model output. *Environ. Res. Lett.*, **9**, 115008, https://doi.org/10.1088/1748-9326/9/11/115008.

Manzini, E., and Coauthors, 2014: Northern winter climate change: Assessment of uncertainty in CMIP5 projections related to stratosphere–troposphere coupling. *J. Geophys. Res.*, **119**, 7979–7998, https://doi.org/10.1002/2013JD021403.

Meehl, G. A., C. A. Senior, V. Eyring, G. Flato, J.-F. Lamarque, R. J. Stouffer, K. E. Taylor, and M. Schlund, 2020: Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models. *Sci. Adv.*, **6**, eaba1981, https://doi.org/10.1126/sciadv.aba1981.

Mindlin, J., T. G. Shepherd, C. S. Vera, M. Osman, G. Zappa, R. W. Lee, and K. I. Hodges, 2020: Storyline description of Southern Hemisphere midlatitude circulation and precipitation response to greenhouse gas forcing. *Climate Dyn.*, **54**, 4399–4421, https://doi.org/10.1007/s00382-020-05234-1.

Notaro, M., W.-C. Wang, and W. Gong, 2006: Model and observational analysis of the northeast U.S. regional climate and its relationship to the PNA and NAO patterns during early winter. *Mon. Wea. Rev.*, **134**, 3479–3505, https://doi.org/10.1175/MWR3234.1.

O'Neill, B. C., and Coauthors, 2016: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geosci. Model Dev.*, **9**, 3461–3482, https://doi.org/10.5194/gmd-9-3461-2016.

Pithan, F., and T. Mauritsen, 2014: Arctic amplification dominated by temperature feedbacks in contemporary climate models. *Nat. Geosci.*, **7**, 181–184, https://doi.org/10.1038/ngeo2071.

Scheff, J., and D. Frierson, 2012: Twenty-first-century multimodel subtropical precipitation declines are mostly midlatitude shifts. *J. Climate*, **25**, 4330–4347, https://doi.org/10.1175/JCLI-D-11-00393.1.

Seager, R., N. Naik, and G. A. Vecchi, 2010: Thermodynamic and dynamic mechanisms for large-scale changes in the hydrological cycle in response to global warming. *J. Climate*, **23**, 4651–4668, https://doi.org/10.1175/2010JCLI3655.1.

Shepherd, T. G., 2019: Storyline approach to the construction of regional climate change information. *Proc. Roy. Soc.*, **475A**, 20190013, https://doi.org/10.1098/rspa.2019.0013.

——, and Coauthors, 2018: Storylines: An alternative approach to representing uncertainty in physical aspects of climate change. *Climatic Change*, **151**, 555–571, https://doi.org/10.1007/s10584-018-2317-9.

Trenberth, K. E., and J. M. Caron, 2000: The Southern Oscillation revisited: Sea level pressures, surface temperatures, and precipitation. *J. Climate*, **13**, 4358–4365, https://doi.org/10.1175/1520-0442(2000)013<4358:TSORSL>2.0.CO;2.

USGCRP, 2018: *Impacts, Risks, and Adaptation in the United States*. Vol. II, *Fourth National Climate Assessment*, D. R. Reidmiller et al., Eds., U.S. Global Change Research Program, https://doi.org/10.7930/NCA4.2018.

van den Hurk, B., G. J. van Oldenborgh, G. Lenderink, W. Hazeleger, R. Haarsma, and H. de Vries, 2014a: Drivers of mean climate change around the Netherlands derived from CMIP5. *Climate Dyn.*, **42**, 1683–1697, https://doi.org/10.1007/s00382-013-1707-y.

——, P. Siegmund, and A. Klein Tank, Eds., 2014b: KNMI'14: Climate change scenarios for the 21st century—A Netherlands perspective. KNMI Tech. Rep. WR 2014-01, 120 pp, http://bibliotheek.knmi.nl/knmipubWR/WR2014-01.pdf.

Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.*, **109**, 784–812, https://doi.org/10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2.

——, Q. Fu, B. V. Smoliak, P. Lin, and C. M. Johanson, 2012: Simulated versus observed patterns of warming over the extratropical Northern Hemisphere continents during the cold season. *Proc. Natl. Acad. Sci. USA*, **109**, 14 337–14 342, https://doi.org/10.1073/pnas.1204875109.

Wold, H., 1966: Estimation of principal components and related models by iterative least squares. *Multivariate Analysis*, P. R. Krishnaiah, Ed., Academic Press, 391–420.

Yeh, S.-W., D.-W. Yi, M.-K. Sung, and Y. H. Kim, 2018: An eastward shift of the North Pacific Oscillation after the mid-1990s and its relationship with ENSO. *Geophys. Res. Lett.*, **45**, 6654–6660, https://doi.org/10.1029/2018GL078671.

Zappa, G., 2019: Regional climate impacts of future changes in the mid-latitude atmospheric circulation: A storyline view. *Curr. Climate Change Rep.*, **5**, 358–371, https://doi.org/10.1007/s40641-019-00146-7.

——, and T. G. Shepherd, 2017: Storylines of atmospheric circulation change for European regional climate impact assessment. *J. Climate*, **30**, 6561–6577, https://doi.org/10.1175/JCLI-D-16-0807.1.

Zelinka, M. D., T. A. Myers, D. T. McCoy, S. Po-Chedley, P. M. Caldwell, P. Ceppi, S. A. Klein, and K. E. Taylor, 2020: Causes of higher climate sensitivity in CMIP6 models. *Geophys. Res. Lett.*, **47**, e2019GL085782, https://doi.org/10.1029/2019GL085782.