Skillful Seasonal Prediction of North American Summertime Heat Extremes

Liwei Jia,[a,b] Thomas L. Delworth,[a] Sarah Kapnick,[a] Xiaosong Yang,[a] Nathaniel C. Johnson,[a,c]
William Cooke,[a] Feiyu Lu,[a,c] Matthew Harrison,[a] Anthony Rosati,[a,b] Fanrong Zeng,[a]
Colleen McHugh,[a,d] Andrew T. Wittenberg,[a] Liping Zhang,[a,b] Hiroyuki Murakami,[a,b] and
Kai-Chih Tseng[a,c]

[a] Geophysical Fluid Dynamics Laboratory, NOAA, Princeton, New Jersey
[b] University Corporation for Atmospheric Research, Boulder, Colorado
[c] Department of Geosciences, Princeton University, Princeton, New Jersey
[d] Science Applications International Corporation, Reston, Virginia

ABSTRACT: This study shows that the frequency of North American summertime (June–August) heat extremes is skillfully predicted several months in advance in the newly developed Geophysical Fluid Dynamics Laboratory (GFDL) Seamless System for Prediction and Earth System Research (SPEAR) seasonal forecast system. Using a statistical optimization method, the average predictability time, we identify three large-scale components of the frequency of North American summer heat extremes that are predictable with significant correlation skill. One component, which is related to a secular warming trend, shows a continent-wide increase in the frequency of summer heat extremes and is highly predictable at least 9 months in advance. This trend component is likely a response to external radiative forcing. The second component is largely driven by the sea surface temperatures in the North Pacific and North Atlantic and is significantly correlated with the central U.S. soil moisture. The second component shows largest loadings over the central United States and is significantly predictable 9 months in advance. The third component, which is related to the central Pacific El Niño, displays a dipole structure over North America and is predictable up to 4 months in advance. Potential implications for advancing seasonal predictions of North American summertime heat extremes are discussed.

KEYWORDS: North America; Extreme events; Seasonal forecasting

1. Introduction

Observational evidence has revealed an overall increase in the number of hot days at the global scale since the middle of the twentieth century (Seneviratne et al. 2012). The Intergovernmental Panel on Climate Change (IPCC) reported that it is very likely that the length, frequency, and/or intensity of heat extremes will increase over most land areas through the twenty-first century (Seneviratne et al. 2012). Heat extremes, such as heatwaves, have profound societal, economic, and ecological impacts. They can burden health and emergency services, increase stress on water resources and transportation, and disrupt energy systems, resulting in power shortages. According to the World Health Organization, from 1998 to 2017, more than 166 000 people died due to heatwaves, including more than 70 000 who died during the 2003 heatwave in Europe. The 2003 European heatwave also caused forest fires (Fischer et al. 2007) and decreased agricultural production. It has been estimated that more than 800 deaths can be attributed to the 1995 mid-July heatwave over the central United States (Changnon et al. 1996). Predicting heat extremes on seasonal time scales is crucial in developing early warning systems to improve societal preparedness.

Most of the earlier studies on the predictability and prediction of heat extremes are on medium-range to subseasonal time scales (Vitart and Robertson 2018; Mandal et al. 2019; Hudson et al. 2011; White et al. 2014; Vitart 2005; Teng et al.

2013; McKinnon et al. 2016). Little progress has been made thus far on forecasting heat extremes on seasonal and longer time scales, because it has been historically challenging to predict extremes on such long time scales. We emphasize that prediction of extreme events on seasonal and longer time scales is important for planners and decision makers in management and response. Although predicting individual extreme events on seasonal time scales is extremely challenging, predicting the statistics of the extreme events on seasonal time scales may be possible. So far, there have been few attempts to assess the skill of extremes on or beyond seasonal scales (Hanlon et al. 2013; Hamilton et al. 2012; Pepler et al. 2015). For example, a study by Hamilton et al. (2012) showed skill in predicting the number of daily temperature extremes in the Northern Hemisphere land area over 3-month periods at 1-month lead time, and attributed the skill to the model's ability to predict El Niño–Southern Oscillation (ENSO) and climate change, as well as the initialization of soil moisture. However, there is still a lack of studies quantifying how predictable the extremes are, and distinguishing the predictability sources for different types of extremes. To fill this gap, this study examines the seasonal prediction skill of the frequency of North American (area north of 23°N) summertime heat extremes at various lead times from 0 to 9 months, and explores the potential sources of the prediction skill. We apply a statistical optimization technique to identify predictable components of North American summertime heat extremes, measured by the frequency of summertime hot days (HDs), in the newly developed Geophysical Fluid

Dynamics Laboratory (GFDL) Seamless System for Prediction and Earth System Research (SPEAR) seasonal forecast system, and reveal that the year-to-year variations in the frequency of HDs are skillfully predictable several months in advance. The SPEAR seasonal forecast system will be described briefly in section 2. The results are presented in section 3. This paper concludes with a summary and discussion.

2. Model, data, and methodology

a. SPEAR retrospective forecasts

SPEAR is the newly developed next-generation GFDL modeling system for seasonal to multidecadal prediction and projection (Delworth et al. 2020). It is a coupled ocean–atmosphere–land–sea ice model. The atmosphere and land components are the GFDL AM4-LM4 model (Zhao et al. 2018); the ocean and sea ice components are the MOM6 and SIS2 (Adcroft et al. 2019). This study utilizes the medium-resolution SPEAR version with an atmospheric and land resolution of 50 km and 33 atmospheric vertical levels. For computational speed SPEAR uses a coarse ocean resolution of approximately 1.0° with tropical refinement to 0.3°. Comprehensive details are described in Delworth et al. (2020) and Lu et al. (2020).

To evaluate prediction skill in the SPEAR prediction system, an extensive set of reforecasts (also called hindcasts) were conducted. For each month, from January 1992 through December 2019, a 15-member ensemble of reforecasts was conducted. Each reforecast was of 12-month duration and was initialized using reanalysis from the first day of each month. The oceanic initial conditions of the reforecasts are from a 30-member ocean analysis, produced by an ocean data assimilation (ODA) run in the coupled SPEAR model. Ocean tendency adjustment (OTA) is also used in both the analysis and forecast to reduce model biases (Lu et al. 2020). The atmospheric, land, and sea ice initial conditions for the reforecasts were obtained from a 5-member ensemble of SPEAR restoring simulations in which atmospheric temperature, winds, and moisture were damped back toward values from the Climate Forecast System Reanalysis (CFSR; Saha et al. 2010), and the sea surface temperature was restored to the Optimum Interpolation Sea Surface Temperature (OISST; Reynolds et al. 2002). The 15-member ensemble of reforecasts is generated by applying the 5 restoring members to the first 5 ODA members, the same 5 restoring members to ODA members 6 through 10, and the same 5 restoring members to ODA members 11 through 15. SPEAR has shown significant skill in the prediction of the Niño-3.4 index, temperature over land, midlatitude baroclinic waves, Antarctic sea ice, and atmospheric rivers over western North America (Lu et al. 2020; Zhang et al. 2021; Bushuk et al. 2021; Tseng et al. 2021).

We used 15 members of SPEAR historical simulations to estimate the externally forced pattern of North American summer heat extremes. The 15 members of simulations are initialized from conditions in the SPEAR 1850 control simulation that are 20 years apart at years 101, 121, and every 20 years thereafter until year 381. The radiative forcing agents in the historical simulations are prescribed before year 2014, whereas projections for the Shared Socioeconomic Pathway 5–8.5 (Riahi et al. 2017) are applied after 2014.

b. Verification data

The European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 data (Hersbach et al. 2020), including hourly maximum 2-m temperature, monthly soil moisture (from surface to 289-cm depth), geopotential height at 500 hPa, and SST, are used as the verification data. The daily maximum temperature is calculated based on the hourly maximum temperature. The ERA5 data are on a 0.25° horizontal resolution. These reanalysis data are referred to as observations hereafter. We use the observed Pacific decadal oscillation (PDO) index (Mantua et al. 1997) and Atlantic multidecadal oscillation (AMO) index (Enfield et al. 2001) to identify their relationship with predictable components of North American summer heat extremes. The observed PDO index derived from the ERA5 SST is calculated as the leading principal component from an empirical orthogonal function (EOF) analysis of monthly North Pacific SST anomalies north of 20°N (Mantua and Hare 2002). The AMO index was downloaded from https://psl.noaa.gov/data/correlation/amon.us.data. It is defined as the area weighted average SST over the North Atlantic (0°–70°N). The observed Niño-4 index (averaged SST anomalies over the central equatorial Pacific: 5°S–5°N, 160°E–150°W) was downloaded from https://psl.noaa.gov/gcos_wgsp/Timeseries/Data/nino4.long.anom.data.

c. Average predictability time analysis

To identify predictable components of the frequency of North American summer HDs, we employ the average predictability time (APT) analysis—a statistical optimization method that determines components that maximize predictability. This method was developed by DelSole and Tippett (2009a,b), and has been widely used in predictability studies (Jia et al. 2015; Jia and DelSole 2011, 2012; Yang et al. 2015; Wu et al. 2016; Xiang et al. 2019). Here, we summarize the APT analysis as follows. In the context of initialized ensemble forecasts, the predictability is measured by

$$P(\tau) = 1 - \frac{\sigma_\tau^2}{\sigma_\infty^2}, \tag{1}$$

where σ_τ^2 is the forecast variance at lead time τ and σ_∞^2 is the climatological variance. It can be shown that $P(\tau)$ is related to the signal-to-noise ratio (Jia 2011), which is often used to measure predictability. APT is then defined as twice the sum of $P(\tau)$ over all lead times:

$$\text{APT} = 2\sum_{\tau=0}^{\infty}\left(1 - \frac{\sigma_\tau^2}{\sigma_\infty^2}\right)\Delta\tau. \tag{2}$$

It yields a time scale that agrees with the usual e-folding time for an exponentially decaying forecast signal. Note that the APT is computed from model forecasts that are independent of the observations. In this study, for the target season of June–August (JJA), the sum is from lead 0 to 9 months. The

lead 0-month forecasts correspond to forecasts initialized on 1 June, lead 1-month forecasts correspond to forecasts initialized on 1 May, and so on.

To maximize APT, we seek the projection vector $\mathbf{q}$, such that the time series $\mathbf{q}^T\mathbf{x}(\tau, e)$ maximizes APT, where $\mathbf{x}(\tau, e)$ is the state vector specifying the amplitudes of ensemble forecasts at fixed lead time τ and ensemble member e. The superscript T denotes the transpose operation. The component $\mathbf{q}^T\mathbf{x}(\tau, e)$ has respective forecast and climatological variances as follows:

$$\sigma_\tau^2 = \mathbf{q}^T\boldsymbol{\Sigma}_\tau\mathbf{q} \quad \text{and} \quad \sigma_\infty^2 = \mathbf{q}^T\boldsymbol{\Sigma}_\infty\mathbf{q}, \tag{3}$$

where $\boldsymbol{\Sigma}_\tau$ is the estimated forecast covariance matrix at τ. Substituting (3) into (2) gives

$$\text{APT} = 2\sum_{\tau=0}^{\infty}\left(\frac{\mathbf{q}^T(\boldsymbol{\Sigma}_\infty - \boldsymbol{\Sigma}_\tau)\mathbf{q}}{\mathbf{q}^T\boldsymbol{\Sigma}_\infty\mathbf{q}}\right)\Delta\tau. \tag{4}$$

DelSole and Tippett (2009b) showed that maximizing APT leads to a generalized eigenvalue problem:

$$2\sum_{\tau=0}^{\infty}(\boldsymbol{\Sigma}_\infty - \boldsymbol{\Sigma}_\tau)\Delta\tau\mathbf{q} = \lambda\boldsymbol{\Sigma}_\infty\mathbf{q}. \tag{5}$$

The eigenvalue λ gives the APT value, and each eigenvector $\mathbf{q}$ corresponds to a component. Equation (5) above has more than a single solution; thus, it yields multiple components. The time series of each component is uncorrelated with one another. We order the eigenvalues and their associated eigenvectors by decreasing order such that the first component maximizes APT, the second maximizes APT subject to being uncorrelated with the first, and so on. That is, the first component is most predictable (such as a trend or low-frequency variability), the second component is the second most predictable component, and the last component is the least predictable component. The spatial pattern associated with each component is obtained by projecting the component time series $\mathbf{q}^T\mathbf{x}(\tau, e)$ on the original data, which gives

$$\mathbf{p} = \boldsymbol{\Sigma}_\infty\mathbf{q}. \tag{6}$$

When solving the above eigenvalue problem, we project the forecast data onto 15 leading principal components (PCs) and then maximize APT only in the subspace spanned by the 15 PCs. This is done because the number of grid points in the forecast data exceeds the number of samples, which results in singular covariance matrices and an eigenvalue problem that cannot be solved. We have determined that the APT values are not very sensitive to the number of PCs used when using more than 15 PCs.

To test the significance of the predictability of each component, we adopt the Monte Carlo methods under the null hypothesis that forecasts are drawn independently from a white noise process (DelSole et al. 2011). More specifically, for M spatial dimensions, N time steps, E ensemble members, and L lead times, an $M \times N \times E \times L$ data array was created by drawing independent random numbers from a normal
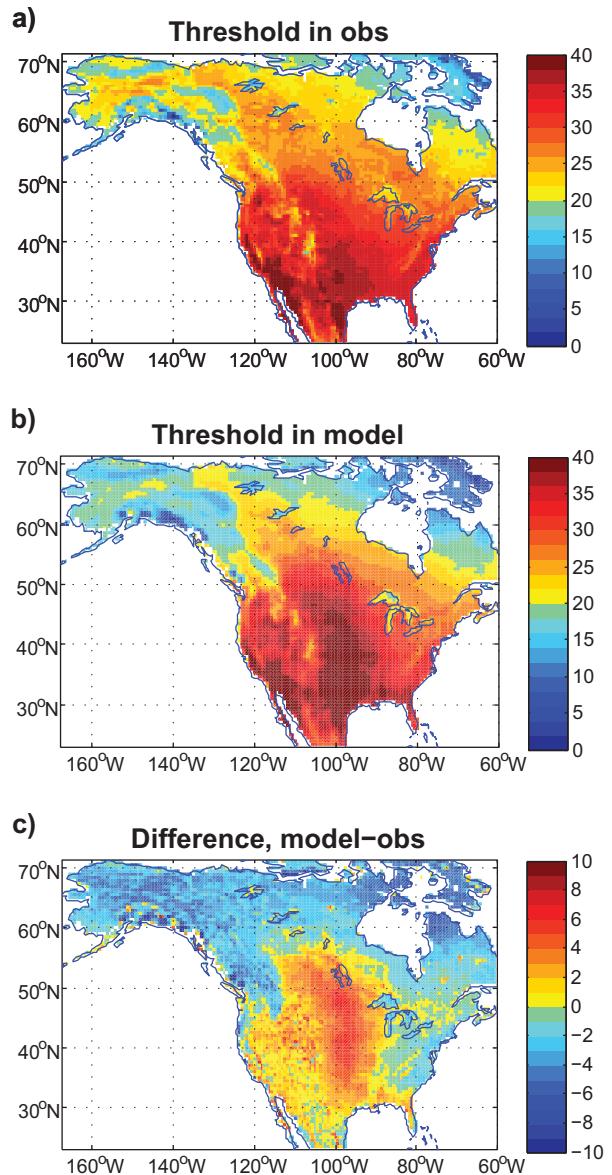

FIG. 1. The 90th-percentile daily Tmax (°C) threshold for defining summer hot days in (a) the observations (ERA5), (b) the SPEAR model hindcast initialized on 1 Jun, and (c) the difference between the model and the observations.

distribution with zero mean and unit variance. APT analysis was applied to this array to produce an ordered sequence of APT values. The procedure was repeated 1000 times. The 5% significance level for each APT associated with each component was determined by selecting the 95th percentile of the APT values derived from the random data. A component is then considered to be potentially predictable if its APT value is high enough to reject the null hypothesis at a 5% significance level. The application of Monte Carlo methods in predictability study can be found in a number of studies (DelSole et al. 2011; Jia and DelSole 2011; Yang et al. 2015; Jia et al. 2015).
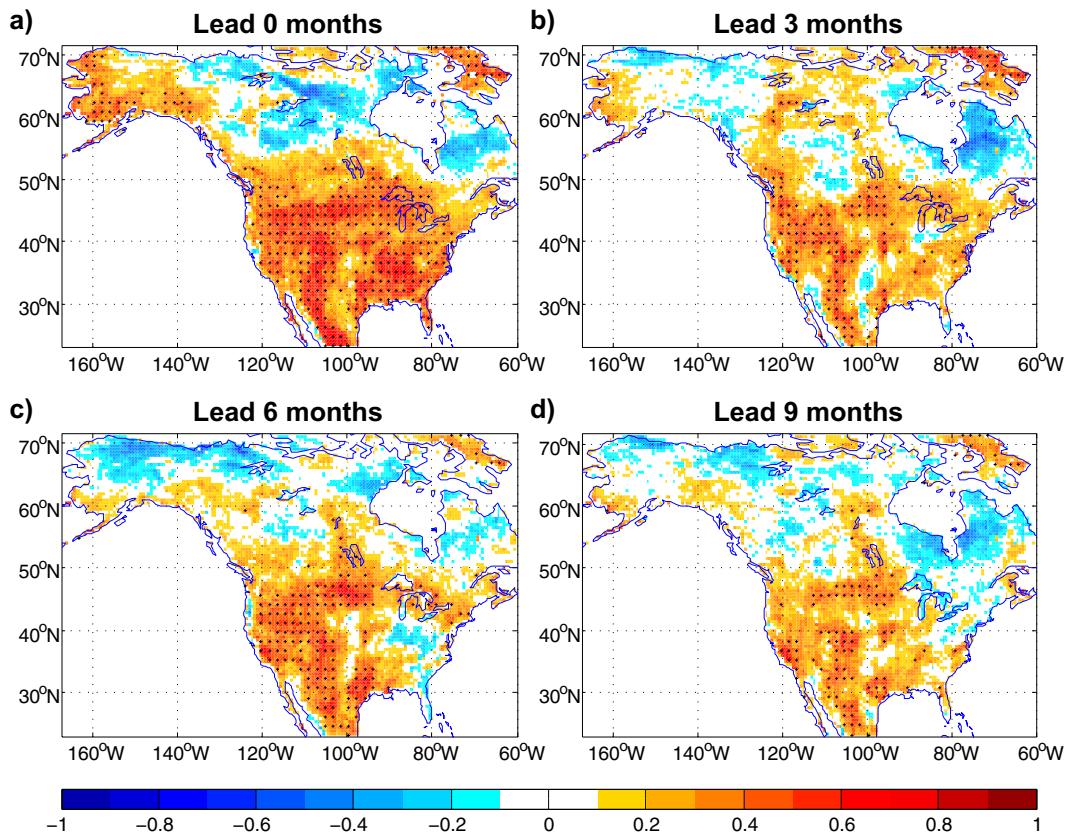
FIG. 2. Pointwise rank correlation of TX90p (percentage of days when daily Tmax exceeds the 90th percentile) over North America between the observations and the ensemble-mean hindcasts at (a) lead 0 (initialized on 1 Jun), (b) lead 3 (initialized on 1 Mar), (c) lead 6 (initialized on 1 Dec of the previous year), and (d) lead 9 (initialized on 1 Sep of the previous year) months. Stippling indicates correlations are locally significant at 5% significance level based on a one-tailed Student's *t* test. The correlation patterns are field significant with a false discovery rate of 10% for all leads (Benjamini and Hochberg 1995).

3. Results

a. Definition of summertime heat index

Following Hamilton et al. (2012) and Pepler et al. (2015), we define a "summer hot day" as the occurrence of the daily maximum temperature (Tmax) anytime during JJA exceeding the 90th-percentile threshold of the climatological daily Tmax distribution from all days in JJA for all years during 1992–2019. In this study, we choose a moderate 90th-percentile threshold to allow a sufficient sample for verification. Based on the threshold, the percentage of days in JJA when daily Tmax exceeds the threshold, denoted as TX90p, is calculated at each grid point and for each year. This study assesses the predictive skill of the TX90p. According to the definition, the TX90p averaged over all years (1992–2019) is 10 at each grid point, but varies with years. To give an idea of the magnitude of TX90p, in the 2012 North American summer heatwave, TX90p is over 40 (i.e., 40% of the days in JJA are hot days) in many areas of the midwestern United States.

In the hindcasts, the threshold is calculated by considering daily Tmax distribution from all days in JJA during 1992–2019 and from all ensemble members, and is computed for each

lead time. The TX90p is then computed for each ensemble member and lead time independently. Since the thresholds in the observations and hindcasts are calculated separately, there is no need to remove model biases when calculating the TX90p.

As used in previous studies (Pepler et al. 2015; Zhang et al. 2018, 2019; Hamilton et al. 2012), a static threshold is used in this study. Using a static threshold might lead to the TX90p varying throughout the season for places that show seasonality of Tmax. To know if a day is an extreme relative to the same day in the historical period, a moving threshold that changes day by day can be used to define a hot day if it exceeds the 90th percentile of its own daily Tmax distribution. Accounting for the seasonal cycle would allow each day in the summer season to have an equal chance to be an extreme, but it would introduce a possible disadvantage of including days with less extreme heat during climatologically cooler times of the season. Another disadvantage of the moving threshold is that it is generally noisy because of limited sample size on a specific day, such that a temporal/spatial smoothing is required. The choice of static or moving threshold would depend on the application. In many applications, a threshold
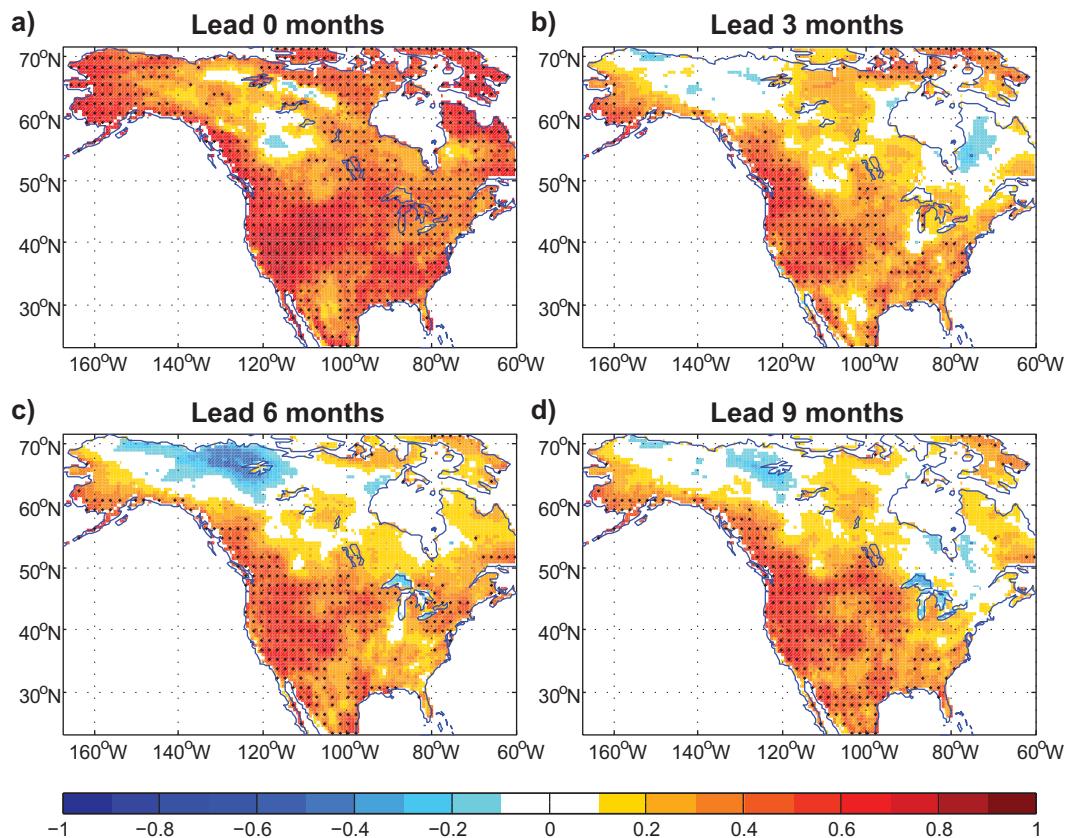
FIG. 3. Pointwise anomaly correlation of JJA mean 2-m air temperature over North America between the observations and the ensemble-mean hindcasts at (a) lead 0 (initialized on 1 Jun), (b) lead 3 (initialized on 1 Mar), (c) lead 6 (initialized on 1 Dec of the previous year), and (d) lead 9 (initialized on 1 Sep of the previous year) months. Stippling indicates correlations are locally significant at 5% significance level based on a one-tailed Student's t test. The correlation patterns are field significant with a false discovery rate of 5% for all leads (Benjamini and Hochberg 1995).

that is constant throughout the season is of interest. As an example, power lines have lower capacities under extreme heat, so an absolute temperature threshold is most relevant.

The 90th-percentile thresholds of daily Tmax over North America in the observations, SPEAR hindcasts, and their difference are shown in Fig. 1. Here, the modeled threshold is calculated from SPEAR hindcasts initialized on 1 June (i.e., lead 0 months). There is little difference in the thresholds at different lead times (not shown). In the observations, a Tmax of approximately 35°C and above can be considered a hot day for most areas south of 40°N. The Tmax threshold generally decreases with latitude. The spatial structures of the threshold in the model bear strong similarities with the observations (Figs. 1a,b), indicating the model represents the observations well. The model demonstrates warm biases over the central United States and Canada and cold biases over high latitudes in simulating the 90th percentile of Tmax. The summer warm biases over the central United States are common in many climate models and the cause of the biases are not fully understood (Lin et al. 2017; Cheruy et al. 2014). Since the thresholds in the observations and model are calculated separately, these model biases do not affect the TX90p.

b. Pointwise correlation skill of North American TX90p and its relationship with the skill of JJA mean air temperature

The pointwise correlation skill of the JJA TX90p over North America at lead 0, 3, 6, and 9 months is shown in Fig. 2. We choose Spearman's rank correlation to measure the skill of TX90p because it is more appropriate for count data. The TX90p shows significant correlation skill over most areas of the United States at lead 0 months (i.e., initialized on 1 June). The correlation skill decreases at longer leads. Some areas over the United States show significant skill even at the 9-month lead (initialized on 1 September of the previous year). These results suggest that TX90p values in many areas of North America, particularly over the United States, are skillfully predictable in the SPEAR forecast system.

It is natural to ask the question if the skill in predicting the TX90p is related to the skill in predicting the summer mean air temperature. We show in Fig. 3 the pointwise correlation skill of JJA mean 2-m air temperature over North America. Overall, it shows higher skill in JJA mean temperature at all leads than that of TX90p (Fig. 2). Their skill structures are similar over the western and central United
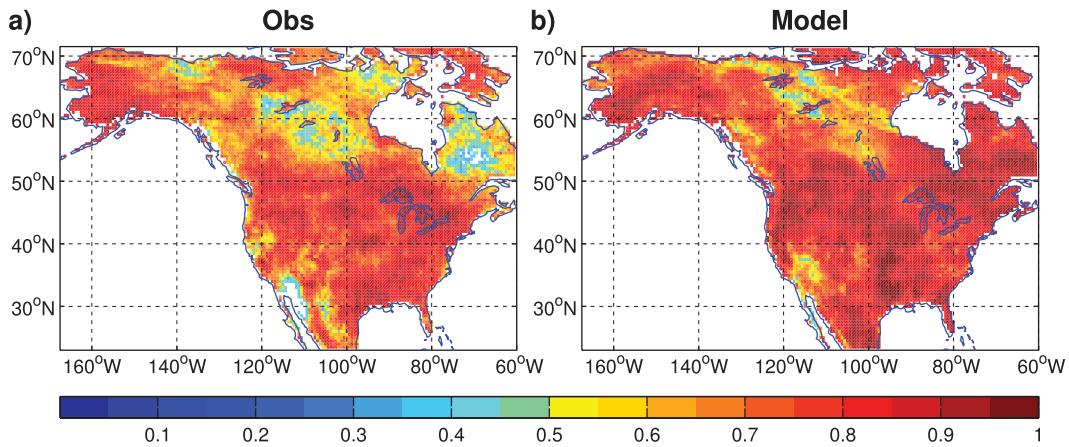
FIG. 4. Pointwise anomaly correlation between the TX90p and JJA mean 2-m air temperature over North America in (a) the observations and (b) model hindcasts (initialized on 1 Jun). Insignificant areas at 5% level based on the Student's t test are in white.

States. In fact, the TX90p in both the observations and model hindcasts is significantly correlated with the mean temperature nearly everywhere over North America (Fig. 4), which agrees with the findings in earlier studies (Johnson et al. 2018; Hamilton et al. 2012). The strong relationship between TX90p and mean temperature suggests that skillful prediction

of mean temperature contributes to the skill in predicting the TX90p.

c. Predictable components of North American TX90p

With the point-to-point skill of TX90p demonstrated above, we now identify the dominant modes of North American
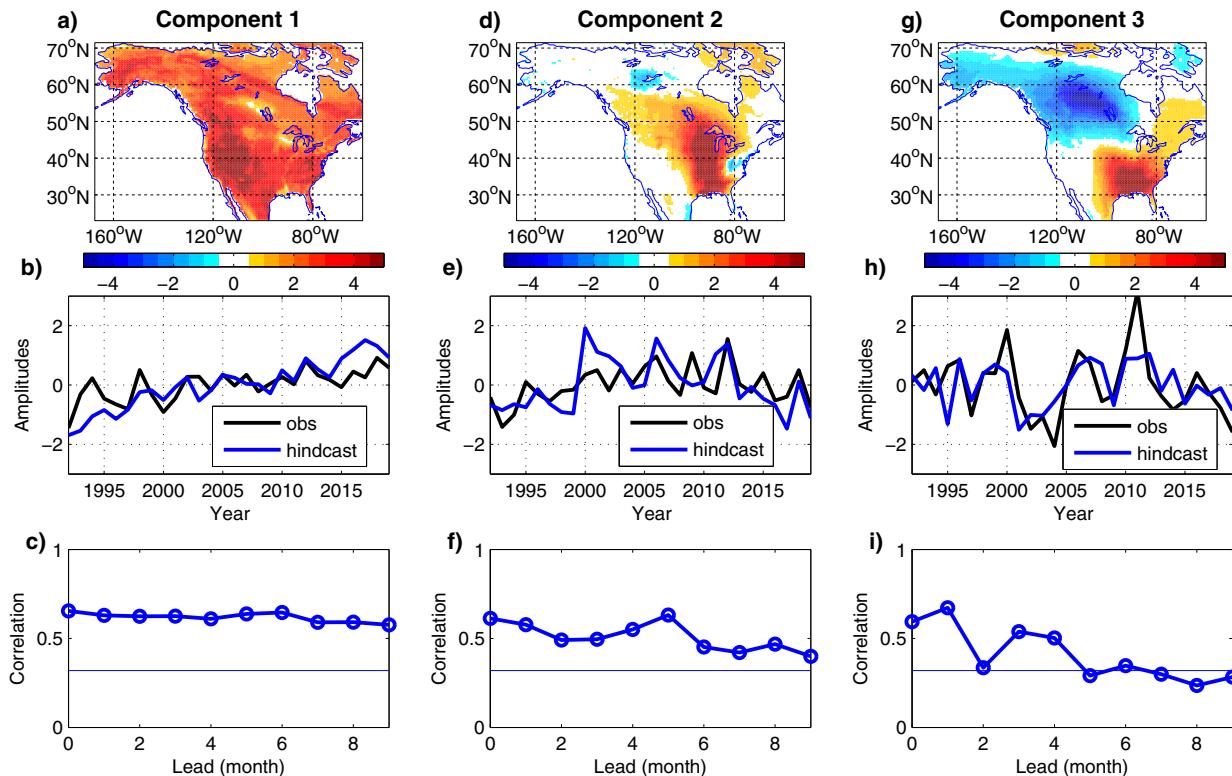

FIG. 5. The (a),(d),(g) spatial patterns (in percentage of days in JJA season), (b),(e),(h) time series, and (c),(f),(i) rank correlation skill of three predictable components of TX90p over North American land areas. The hindcasted time series shown in (b), (e), and (h) are the ensemble mean of lead 0-month forecasts. The horizontal lines in (c), (f) and (i) indicate the 5% significance level based on a one-tailed Student's t test.

summertime TX90p, which helps identify the sources of the prediction skill. We apply the APT analysis to decompose the model hindcasted TX90p into components based on their predictability. Figure 5 shows the spatial structures of three components, their time series, and correlation skill. These three components are statistically significant at 5% level according to the Monte Carlo test. They explain 14%, 5%, and 6% of the total variance, respectively. The associated time series in the observations for each component is computed by projecting the ERA5 onto the component. The prediction skill of the components at each lead time is measured by the Spearman's rank correlation between the observational and the ensemble-mean time series at each lead time in the hindcasts. We find significant prediction skill in the three components. The rest of the components do not show significant skill on seasonal time scales. Thus, we only discuss these three predictable components in this study.

The spatial pattern of the first predictable component shows positive amplitudes over all of North America (Fig. 5a). The associated time series of the first component shows a warming trend in both the hindcasts and the observations, indicating its relationship with the warming climate (Fig. 5b). This trend component is highly predictable with significant correlation skill at all leads from 0 to 9 months (Fig. 5c). Its high predictive skill is not surprising because trends tend to be highly predictable. The spatial pattern of the second component shows the largest positive magnitudes over the central United States. (Fig. 5d). Its time series shows low-frequency variability (Fig. 5e). Figure 5f reveals that this component is predictable with significant skill 9 months in advance. The third component displays a southeast–northwest dipole structure over North America (Fig. 5g). Its time series varies primarily on interannual time scales (Fig. 5h). The third component is skillfully predictable up to 4 months (Fig. 5i).

d. Predictability source of the North American summertime heat extremes: Radiative forcing

As showed above, the first predictable component demonstrates a warming signal. We further explore the predictability source of the first component. Using the signal-to-noise maximizing EOF technique (Ting et al. 2009; Chang et al. 2000; Venzke et al. 1999), we estimate the externally forced pattern of JJA TX90p over North American land area in SPEAR historical simulations. The signal-to-noise maximizing EOF method extracts the forced pattern by maximizing the ratio of signal variance to the noise variance in the ensemble model simulations, where the signal variance is defined as the variance of ensemble mean and the noise variance is estimated as the variance of the deviations of each ensemble member from the ensemble mean. This method applies a spatial prewhitening transformation that removes spatial coherence in the atmospheric noise contained in the ensemble mean. As a result, the noise contamination in the ensemble mean is reduced. This method isolates the forced pattern better than a simple ensemble average.

It can be seen from Fig. 6 that the externally forced pattern of TX90p shows similarities to the first predictable component
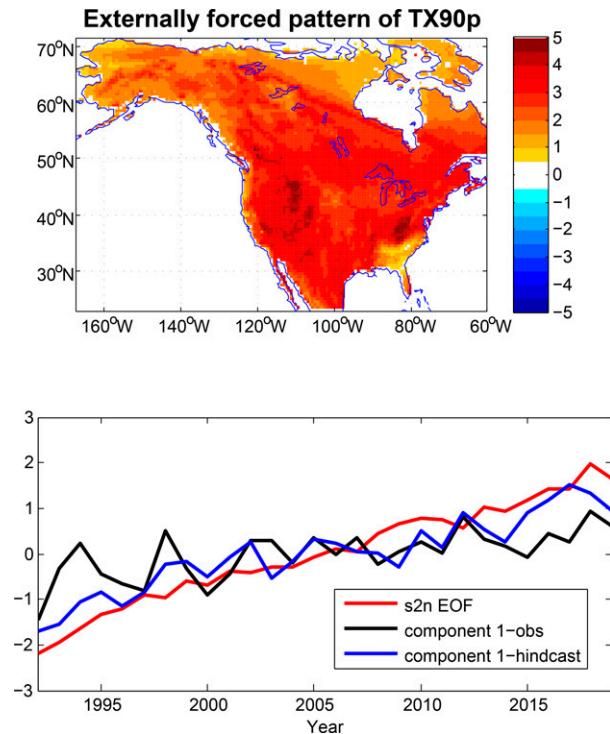

FIG. 6. (top) Externally forced pattern of JJA TX90p anomalies (in percentage of days in JJA season) estimated from the 15-member historical simulations of SPEAR during 1992–2019 using the signal-to-noise maximizing EOF analysis. (bottom) The time series of the externally forced pattern (red) and the time series of the first predictable component in the observations and model hindcasts. The time series of the first predictable component are identical to those in Fig. 5b, which are included here for ease of comparison.

(Fig. 5a) with a pattern correlation of 0.59. The pattern correlation is statistically significant at 5% level according to the bootstrapping metric under the null hypothesis that the first predictable component is uncorrelated with the externally forced pattern. To test the null hypothesis, we randomly resample the TX90p in time and across ensemble members in the historical simulations and then apply the signal-to-noise maximizing EOF to the resampled data. The resulting pattern is then correlated with the pattern of the first predictable component (Fig. 5a). This procedure is repeated 1000 times to produce an ordered sequence of pattern correlations. The 5% significance level is selected as the 95th percentile of the correlations. In addition to the significant pattern correlation, the time series of the forced pattern (bottom of Fig. 6) is also highly correlated with the time series of the first predictable component. These results suggest that this component is likely the response to the changes in external radiative forcing.

e. Predictability source of the North American summertime heat extremes: Sea surface temperature

To understand the source of the predictive skill of the second and the third predictable components, we first correlate the components with a potential large-scale driver—SST. We
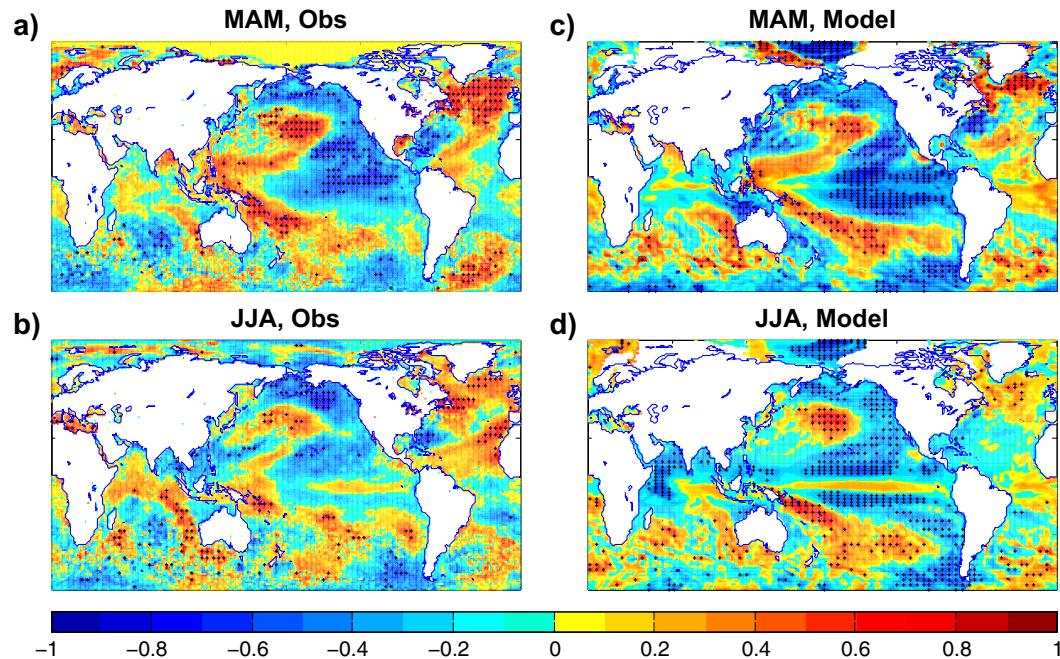
FIG. 7. The correlation coefficients between global SST anomalies in (top) MAM and (bottom) JJA seasons and the time series of the second predictable component of JJA TX90p in the (a),(b) observations and (c),(d) model hindcasts. The model-hindcasted correlation map for each season is initialized on the first day of the season. Stippling indicates correlations are significant at 10% level based on the Student's t test.

focus on SST because on seasonal and longer time scales, anomalous atmospheric conditions are often linked to SST anomalies (McKinnon et al. 2016; Johnson et al. 2018; Kamae et al. 2014; Trenberth and Fasullo 2012). For example, studies have found ENSO is associated with climate extremes (Goddard and Gershunov 2020; Jia et al. 2016; Hamilton et al. 2012). Due to the long memory of the ocean, skillful prediction of oceanic variables can provide skill in predicting atmospheric extremes.

Figure 7 shows the correlation maps of the global SSTs in the March–May (MAM) and JJA seasons with the time series of the second predictable component of JJA TX90p. The SSTs are linearly detrended to remove the influence of linear trend. The MAM correlation map of SST is calculated by correlating the time series of the second predictable component of JJA TX90p with the SST in MAM at each grid point. Similarly, the correlation map in the JJA season is calculated by correlating the time series of the second predictable component of JJA TX90p with the SST in JJA at each grid point. Each time step in the time series of the predictable component or the SST represents a season of a year, and the spacing to the next time step is 1 year. As the temporal autocorrelations of the time series of the second component and the detrended SST are small, the time steps are thus assumed to be independent. The degrees of freedom are $N - 2$, where N indicates the number of years.

The SSTs over the Pacific are reminiscent of a negative phase of the PDO (Mantua et al. 1997; Mantua and Hare 2002) in both seasons in the observations as well as in model

hindcasts. The model also well represents the associated teleconnection pattern in 500-hPa geopotential height (figure not shown). The observed time series of the second predictable component is correlated with the observed PDO index in MAM ($R = -0.42$) and JJA ($R = -0.38$) seasons, which means the PDO explains about 15% of the variance of the second component. The PDO has been described by some as a long-lived El Niño–like pattern of Pacific climate variability, varying on decadal time scales (Zhang et al. 1997; Mantua and Hare 2002) and interactions between the tropical Pacific and extratropical Pacific (Newman et al. 2016). The PDO may also be related to dynamical air–sea interactions in the extratropics (Zhang and Delworth 2015). It has widespread impacts on global climate. Unlike ENSO, which has primary impacts in the tropics and secondary impacts in the extratropics, the climatic fingerprints of the PDO are most visible in the extratropics, especially the North Pacific–North American sector, but secondary signatures exist in the tropics (Mantua and Hare 2002). Studies have shown that the PDO affects Pacific marine ecosystems (Mantua and Hare 2002), western U.S. extreme precipitation (DeFlorio et al. 2013), and droughts in the U.S. Great Plains (Hu and Huang 2009).

To further demonstrate the connection between the PDO and North American heat extremes, we display the observed correlations between the JJA PDO index and North American TX90p in Fig. 8a. The PDO is significantly correlated with the TX90p over the central United States, where the second component shows largest loadings (cf. Figs. 8a and 5d). As can be seen from the observed regression map of 500-hPa
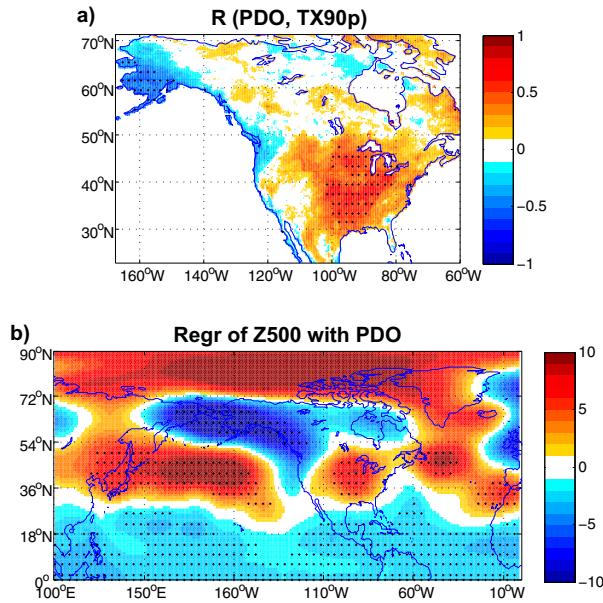
FIG. 8. (a) The observed correlation between the negative PDO index in JJA and TX90p and (b) the observed regression map of 500-hPa geopotential height anomalies (in m) with the negative PDO index in JJA. Stippling indicates correlations/regressions are locally significant at 10% level. The correlation pattern in (a) [regression pattern in (b)] is field significant with a false discovery rate of 10% (20%); see Benjamini and Hochberg 1995.

geopotential height anomalies with the negative PDO index in JJA season (Fig. 8b), the negative phase of the PDO is associated with a wave train–like pattern over the North Pacific and North America. Anomalous low pressures are shown over the subpolar Pacific (around 60°N), Alaska, and the northeast of North America. Anomalous high pressures are located over the North Pacific (around 40°N) as well as the central United States. The high pressure system over the central United States provides clear and dry conditions that increase the radiative heating of the surface and reduce precipitation, hence favoring the development of heat extremes. Note that the PDO-related wave train–like pattern in summer is weaker than that in winter (not shown), and does not extend as far poleward as in winter.

In fact, the SPEAR model is able to skillfully predict the PDO index months in advance. Figure 9 is the correlation skill of the PDO index in SPEAR as a function of initial month and target season. It shows the PDO index is predictable with significant skill in the SPEAR seasonal forecast system for all initial months and leads from 0 to 9 months, except for the lead 8- and 9-month forecasts initialized in July. The skillful prediction of the PDO serves as the source of the prediction skill of central U.S. summertime heat extremes.

In addition to the Pacific basin, the North Atlantic also demonstrates correlations with the second component of TX90p in both the observations and model hindcasts (Fig. 7). The SST structures over the North Atlantic are reminiscent of the AMO pattern. The observed AMO index is significantly correlated with the time series of the second component with
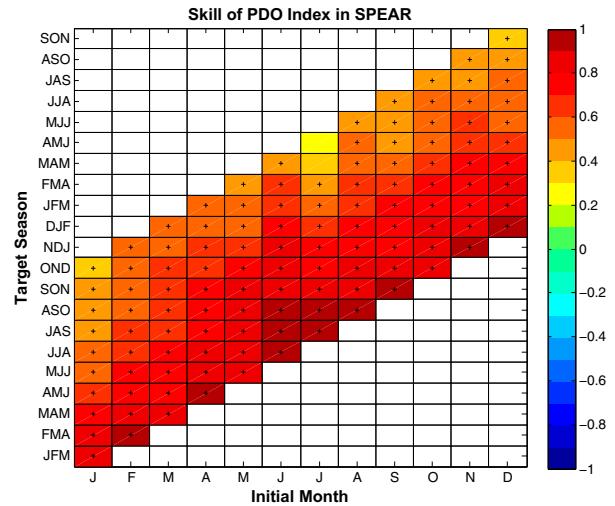

FIG. 9. Anomaly correlation skill of the PDO index in the SPEAR seasonal forecast system as a function of initial month and target season. Stippling indicates the correlations are statistically significant at 5% level based on the Student's t test.

a simultaneous correlation of 0.43, meaning the AMO-like SSTs also contribute to the predictability of central U.S. summertime heat extremes. The impacts of AMO on North American summer climate and extreme events have been documented in numerous earlier studies (Johnson et al. 2018; Curtis 2008; Zhang et al. 2018; Ruprich-Robert et al. 2018).

We now show in Fig. 10 the correlation maps of global SSTs in MAM and JJA seasons with the time series of the third predictable component of JJA TX90p. The SSTs in the tropical Pacific show a central Pacific La Niña pattern in all cases, except for the MAM season in the model. This suggests that the third component of TX90p with a southeast–northwest dipole structure is ENSO-related. The observed Niño-4 index is significantly correlated with the observed time series of the third component with $R = -0.62$ in JJA and $R = -0.37$ in MAM. These results are consistent with earlier findings that ENSO can influence the frequency of temperature extremes (Goddard and Gershunov 2020; Pepler et al. 2015; Jia et al. 2016; Hamilton et al. 2012).

f. Predictability source of the North American summertime heat extremes: Soil moisture

As found in earlier studies, both the large-scale drivers (such as SSTs) and local-to-regional feedback contribute to the development of extremes (Sillmann et al. 2017). Another possible predictability source of heat extremes arises from the local land–atmosphere feedback. Many studies have revealed that soil moisture is an important predictor of summer temperature extremes (Zhang et al. 2018, 2019; Fischer et al. 2007). The low soil moisture levels in spring and early summer lead to reduced evaporation, preventing cloud formation, which allows more insolation to further warm and dry out the land surface (Hanlon et al. 2013; Fischer et al. 2007; Seneviratne et al. 2006). Here, we correlated the second and the third
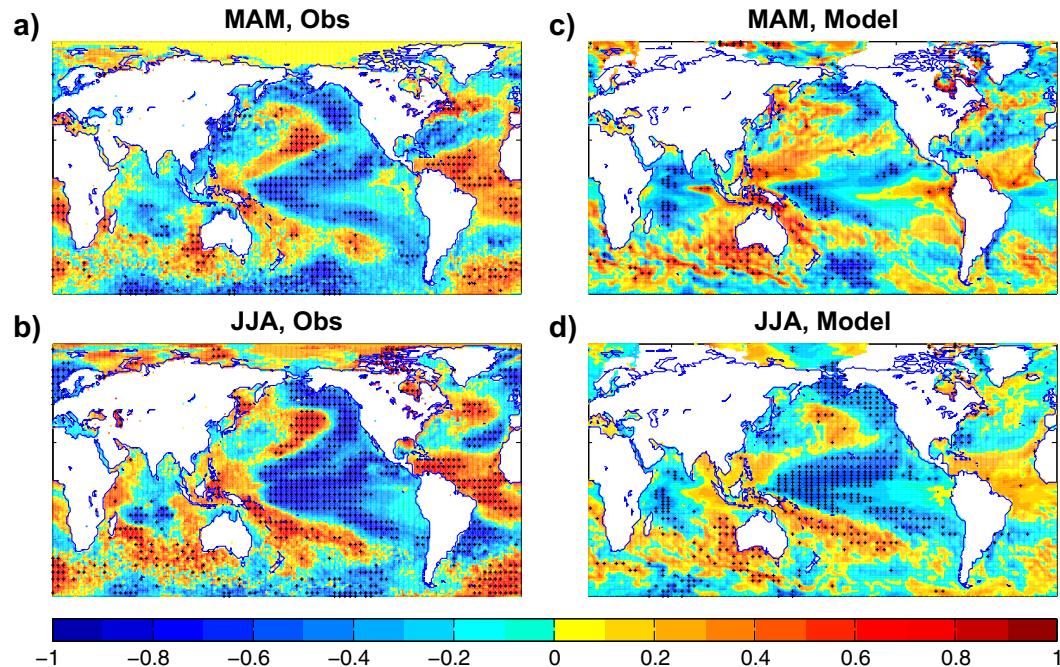
FIG. 10. As in Fig. 7, but for the third predictable component.

predictable components of North American summer TX90p with North American soil moisture.

Figure 11 shows the correlation maps of North American soil moisture in MAM and JJA seasons with the time series of the second (C2) and the third (C3) predictable components of TX90p. The correlation patterns in the hindcasts bear strong similarities with those in the observations, meaning the relationship between North American summer heat extremes and the soil moisture is well captured in the model. For the correlations between the soil moisture and the second component (Figs. 11a–d), the largest negative correlations are shown over the central United States in both seasons. The highest negative correlations over the central United States correspond well to the largest loadings over the central United States as seen in the spatial pattern of the second component of TX90p (Fig. 5d), implying the local atmosphere–land feedback may also serve as a predictability source of the central U.S. summer heat extremes. In other words, there is a high frequency of central U.S. heat extremes when the local land conditions are dry. The dry land conditions in spring and summer seasons favor the occurrence of summer heat extremes.

The local soil moisture is also related to the third predictable component of TX90p. As shown from the correlation maps of North American soil moisture with the time series of the third predictable component of TX90p (Figs. 11e–h), both the observations and model hindcasts reveal negative correlations in the southeast and positive correlations in the U.S. Northwest, corresponding well with the spatial structure of the third component of TX90p (Fig. 5g). The land conditions persisting from the spring season contribute to the prediction skill of summer heat extremes.

g. Reconstructing TX90p predictions based on three predictable components

Having identified three predictable components of North American summertime TX90p and demonstrated prediction skill of these components, it is compelling to reconstruct predictions based upon the three skillful components. The hypothesis is that the reconstructed predictions have higher skill than the raw predictions directly from the model because the unskillful components are filtered out. Similarly, Scaife et al. (2014) has demonstrated the prediction skill of North American and European winter surface climate using only the forecast North Atlantic Oscillation (NAO) is higher than the skill directly from model forecast.

Figure 12 shows the rank correlations averaged over lead times from 0 to 9 months in model raw predictions, reconstructed predictions, and their difference. The reconstructed predictions show higher skill than the raw predictions over Alaska and parts of the central and southeastern United States (Fig. 12c), although both show positive correlations over most of the western United States. To examine the skill improvements from another perspective, we plot the percentage of North American land area with significant correlation skill (at the 5% level) as a function of lead time in raw model predictions and reconstructed predictions (Fig. 12d). The reconstructed predictions using only the three predictable components have higher skill than the raw predictions at all leads except for the lead 0 months. The improvements in skill are more prominent at long leads than short leads. This is because at short leads, the skill primarily comes from model initialization. With the increase in lead time, model initialization has less impact on the predictions and the unpredictable
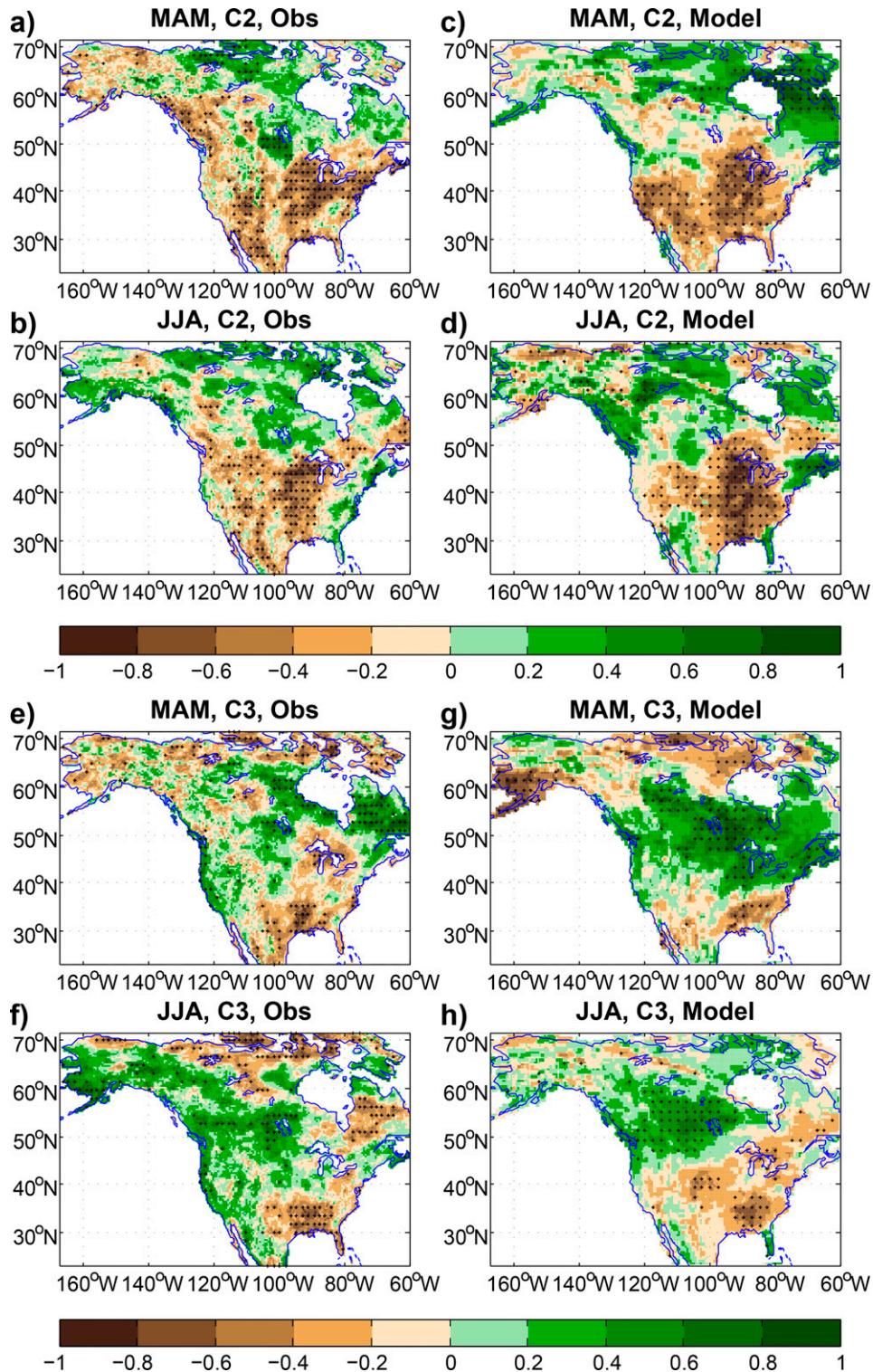
FIG. 11. The correlation maps of North American soil moisture in MAM and JJA seasons with the time series of the (a)–(d) second and (e)–(h) third predictable component of JJA TX90p in the (left) observations and (right) model hindcasts. The model-hindcasted soil moisture at each season is initialized on the first day of the season. Stippling indicates correlations are locally significant at 10% level based on the Student's *t* test. The correlation patterns are field significant with a false discovery rate of 10% (20%) in JJA (MAM) in both the observations and model hindcasts; see Benjamini and Hochberg (1995).
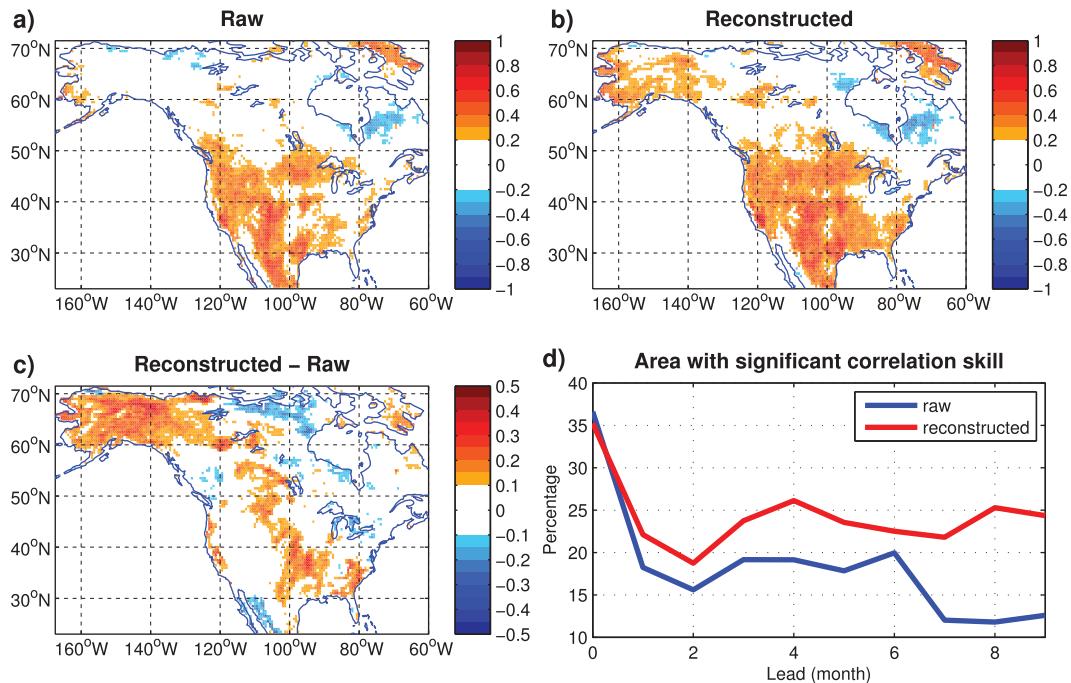
FIG. 12. The rank correlation skill of TX90p from (a) raw predictions directly from SPEAR and (b) reconstructed predictions from the three predictable components shown in Fig. 5, (c) the correlation skill from reconstructed predictions minus the correlation skill from raw predictions, and (d) the percentage of North American land area with significant rank correlation skill at 5% significance level based on a one-tailed Student's *t* test in model raw predictions (blue) and reconstructed predictions (red) as a function of lead time. The correlations in (a)–(c) are averaged over lead times from 0 to 9 months.

noise increases, and so filtering unpredictable components considerably improves skill. The above results suggest that making predictions with the three components advances prediction skill of North American summertime heat extremes.

4. Summary and discussion

We show in this study that the frequency of summertime daily maximum 2-m air temperature exceeding the 90th percentile of the climatological distribution (TX90p) over North America is skillfully predictable on seasonal time scales in the newly developed GFDL SPEAR seasonal forecast system. On grid point scale, the North American summer TX90p shows significant correlation skill over many areas of the United States at leads of 0–9 months. The TX90p demonstrates a good relationship with summer mean 2-m air temperature, meaning skillful prediction of mean temperature contributes to the skill in predicting the TX90p. To capture the large-scale structure of TX90p and explore the sources of the predictability of North American summer heat extremes, we further identify the large-scale predictable components of North American TX90p using a statistical optimization technique (APT) and explore their sources of the predictability. Three components of North American summer TX90p are found to be skillfully predictable on seasonal time scales. The first one is a trend component, the second component is a PDO-/AMO-like component, and another component is

related to the central Pacific El Niño. The first component shows a continent-wide increase in the frequency of summer heat extremes and is likely a response to external radiative forcing. This trend component is skillfully predictable at least 9 months in advance. The second predictable component shows a central U.S. pattern that is predictable with significant correlation skill of 9 months. The central U.S. summer TX90p is correlated with the PDO and AMO indices as well as the central U.S. soil moisture. The third component with a southeast–northwest dipole structure is associated with the central Pacific El Niño. This ENSO-related component is skillfully predicted up to 4 months. This study suggests that the radiative forcing, PDO-/AMO-like SSTs, ENSO, and local atmosphere–land feedback all contribute to the skillful seasonal prediction of the frequency of North American summertime heat extremes. Conducting predictions using the three skillful components (i.e., filtering out unpredictable noise) advances seasonal prediction skill of North American heat extremes.

This study uses a moderate threshold to define hot days, allowing sufficient samples for verification. One can choose a more extreme threshold, and the associated TX90p skill may vary. Studies suggested slightly lower skill in predicting temperature extremes when choosing more extreme thresholds (Hamilton et al. 2012). The seasonal prediction skill of the North American summer TX90p is diagnosed in the GFDL SPEAR seasonal forecast model, which has demonstrated skill in predicting many aspects of the climate system. The

actual prediction skill of TX90p might be different in other forecast models. The SST patterns related to the second predictable component show features that are reminiscent of the PDO/AMO patterns that vary primarily on decadal scale. However, due to the limited data size (1992–2019) in this study, caution should be given when interpreting the results that are related to PDO/AMO. Further investigations on the robustness of the results could be done with additional experiments. This study focuses on the prediction of the North American summertime heat extremes. The detailed mechanisms linking SSTs and soil moisture to heat extremes need to be further explored in future works.

Data availability statement. Data used in this study are available upon reasonable request.

REFERENCES

Adcroft, A., and Coauthors, 2019: The GFDL global ocean and sea ice model OM4.0: Model description and simulation features. *J. Adv. Model. Earth Syst.*, **11**, 3167–3211, https://doi.org/10.1029/2019MS001726.

Benjamini, Y., and Y. Hochberg, 1995: Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Stat. Soc.*, **57B**, 289–300, https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Bushuk, M., and Coauthors, 2021: Seasonal prediction and predictability of regional Antarctic sea ice. *J. Climate*, **34**, 6207–6233, https://doi.org/10.1175/JCLI-D-20-0965.1.

Chang, P., R. Saravanan, L. Ji, and G. C. Hegerl, 2000: The effect of local sea surface temperatures on atmospheric circulation over the tropical Atlantic sector. *J. Climate*, **13**, 2195–2216, https://doi.org/10.1175/1520-0442(2000)013<2195:TEOLSS>2.0.CO;2.

Changnon, S., K. Kunkel, and B. Reinke, 1996: Impacts and responses to the 1995 heat wave: A call to action. *Bull. Amer. Meteor. Soc.*, **77**, 1497–1506, https://doi.org/10.1175/1520-0477(1996)077<1497:IARTTH>2.0.CO;2.

Cheruy, F., J. L. Dufresne, F. Hourdin, and A. Ducharne, 2014: Role of clouds and land–atmosphere coupling in midlatitude continental summer warm biases and climate change amplification in CMIP5 simulations. *Geophys. Res. Lett.*, **41**, 6493–6500, https://doi.org/10.1002/2014GL061145.

Curtis, S., 2008: The Atlantic multidecadal oscillation and extreme daily precipitation over the US and Mexico during the hurricane season. *Climate Dyn.*, **30**, 343–351, https://doi.org/10.1007/s00382-007-0295-0.

DeFlorio, M. J., D. W. Pierce, D. R. Cayan, and A. J. Miller, 2013: Western U.S. extreme precipitation events and their relation to ENSO and PDO in CCSM4. *J. Climate*, **26**, 4231–4243, https://doi.org/10.1175/JCLI-D-12-00257.1.

DelSole, T., and M. K. Tippett, 2009a: Average predictability time. Part I: Theory. *J. Atmos. Sci.*, **66**, 1172–1187, https://doi.org/10.1175/2008JAS2868.1.

——, and ——, 2009b: Average predictability time. Part II: Seamless diagnoses of predictability on multiple time scales. *J. Atmos. Sci.*, **66**, 1188–1204, https://doi.org/10.1175/2008JAS2869.1.

——, ——, and J. Shukla, 2011: A significant component of unforced multidecadal variability in the recent acceleration of global warming. *J. Climate*, **24**, 909–926, https://doi.org/10.1175/2010JCLI3659.1.

Delworth, T., and Coauthors, 2020: The next generation GFDL modeling system for seasonal to multidecadal prediction and projection. *J. Adv. Model. Earth Syst.*, e2019MS001895, https://doi.org/10.1029/2019MS001895.

Enfield, D., A. Mestas-Nuñez, and P. Trimble, 2001: The Atlantic Multidecadal Oscillation and its relationship to rainfall and river flows in the continental U.S. *Geophys. Res. Lett.*, **28**, 2077–2080, https://doi.org/10.1029/2000GL012745.

Fischer, E. M., S. I. Seneviratne, D. Lüthi, and C. Schär, 2007: Contribution of land–atmosphere coupling to recent European summer heat waves. *Geophys. Res. Lett.*, **34**, L06707, https://doi.org/10.1029/2006GL029068.

Goddard, L., and A. Gershunov, 2020: Impact of El Niño on weather and climate extremes. *El Niño Southern Oscillation in a Changing Climate*, M. J. McPhaden, A. Santoso, and W. Cai, Eds., Wiley, 361–375, https://doi.org/10.1002/9781119548164.ch16.

Hamilton, E., R. Eade, R. J. Graham, A. A. Scaife, D. M. Smith, A. Maidens, and C. MacLachlan, 2012: Forecasting the number of extreme daily events on seasonal timescales. *J. Geophys. Res.*, **117**, D03114, https://doi.org/10.1029/2011JD016541.

Hanlon, H. M., G. C. Hegerl, S. F. B. Tett, and D. M. Smith, 2013: Can a decadal forecasting system predict temperature extreme indices? *J. Climate*, **26**, 3728–3744, https://doi.org/10.1175/JCLI-D-12-00512.1.

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

Hu, Z.-Z., and B. Huang, 2009: Interferential impact of ENSO and PDO on dry and wet conditions in the U.S. Great Plains. *J. Climate*, **22**, 6047–6065, https://doi.org/10.1175/2009JCLI2798.1.

Hudson, D., A. G. Marshall, and O. Alves, 2011: Intraseasonal forecasting of the 2009 summer and winter Australian heat waves using POAMA. *Wea. Forecasting*, **26**, 257–279, https://doi.org/10.1175/WAF-D-10-05041.

Jia, L., 2011: Robust multi-year predictability on continental scales. Ph.D. dissertation, George Mason University, 101 pp., https://www.researchgate.net/publication/258541704_Robust_multi-year_predictability_on_continental_scales.

——, and T. DelSole, 2011: Diagnosis of multiyear predictability on continental scales. *J. Climate*, **24**, 5108–5124, https://doi.org/10.1175/2011JCLI4098.1.

——, and ——, 2012: Multi-year predictability of temperature and precipitation in multiple climate models. *Geophys. Res. Lett.*, **39**, L17705, https://doi.org/10.1029/2012GL052778.

——, and Coauthors, 2015: Improved seasonal prediction of temperature and precipitation over land in a high-resolution GFDL climate model. *J. Climate*, **28**, 2044–2062, https://doi.org/10.1175/JCLI-D-14-00112.1.

——, and Coauthors, 2016: The roles of radiative forcing, sea surface temperatures, and atmospheric and land initial conditions

in U.S. summer warming episodes. *J. Climate*, **29**, 4121–4135, https://doi.org/10.1175/JCLI-D-15-0471.1.

Johnson, N. C., S.-P. Xie, Y. Kosaka, and X. Li, 2018: Increasing occurrence of cold and warm extremes during the recent global warming slowdown. *Nat. Commun.*, **9**, 1724, https://doi.org/10.1038/s41467-018-04040-y.

Kamae, Y., H. Shiogama, M. Watanabe, and M. Kimoto, 2014: Attributing the increase in Northern Hemisphere hot summers since the late 20th century. *Geophys. Res. Lett.*, **41**, 5192–5199, https://doi.org/10.1002/2014GL061062.

Lin, Y., W. Dong, M. Zhang, Y. Xie, W. Xue, J. Huang, and Y. Luo, 2017: Causes of model dry and warm bias over central U.S. and impact on climate projections. *Nat. Commun.*, **8**, 881, https://doi.org/10.1038/s41467-017-01040-2.

Lu, F., and Coauthors, 2020: GFDL's SPEAR seasonal prediction system: Initialization and ocean tendency adjustment (OTA) for coupled model predictions. *J. Adv. Model. Earth Syst.*, **12**, e2020MS002149, https://doi.org/10.1029/2020MS002149.

Mandal, R., S. Joseph, A. K. Sahai, R. Phani, A. Dey, R. Chattopadhyay, and D. R. Pattanaik, 2019: Real time extended range prediction of heat waves over India. *Sci. Rep.*, **9**, 9008, https://doi.org/10.1038/s41598-019-45430-6.

Mantua, N. J., and S. R. Hare, 2002: The Pacific Decadal Oscillation. *J. Oceanogr.*, **58**, 35–44, https://doi.org/10.1023/A:1015820616384.

——, ——, Y. Zhang, J. M. Wallace, and R. C. Francis, 1997: A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull. Amer. Meteor. Soc.*, **78**, 1069–1079, https://doi.org/10.1175/1520-0477(1997)078<1069:APICOW>2.0.CO;2.

McKinnon, K., A. Rhines, M. Tingley, and P. Huybers, 2016: Long-lead predictions of eastern United States hot days from Pacific sea surface temperatures. *Nat. Geosci.*, **9**, 389–394, https://doi.org/10.1038/ngeo2687.

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.

Pepler, A. S., L. B. Díaz, C. Prodhomme, F. J. Doblas-Reyes, and A. Kumar, 2015: The ability of a multi-model seasonal forecasting ensemble to forecast the frequency of warm, cold, and wet extremes. *Wea. Climate Extremes*, **9**, 68–77, https://doi.org/10.1016/j.wace.2015.06.005.

Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate*, **15**, 1609–1625, https://doi.org/10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2.

Riahi, K., and Coauthors, 2017: The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environ. Change*, **42**, 153–168, https://doi.org/10.1016/j.gloenvcha.2016.05.009.

Ruprich-Robert, Y., T. Delworth, R. Msadek, F. Castruccio, S. Yeager, and G. Danabasoglu, 2018: Impacts of the Atlantic multidecadal variability on North American summer climate and heat waves. *J. Climate*, **31**, 3679–3700, https://doi.org/10.1175/JCLI-D-17-0270.1.

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1058, https://doi.org/10.1175/2010BAMS3001.1.

Scaife, A. A., and Coauthors, 2014: Skillful long-range prediction of European and North American winters. *Geophys. Res. Lett.*, **41**, 2514–2519, https://doi.org/10.1002/2014GL059637.

Seneviratne, S. I., D. Lüthi, M. Litschi, and C. Schär, 2006: Land–atmosphere coupling and climate change in Europe. *Nature*, **443**, 205–209, https://doi.org/10.1038/nature05095.

——, and Coauthors, 2012: Changes in climate extremes and their impacts on the natural physical environment. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, C. B. Field, Eds., Cambridge University Press, 109–230.

Sillmann, J., and Coauthors, 2017: Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities. *Weather Climate Extremes*, **18**, 65–74, https://doi.org/10.1016/j.wace.2017.10.003.

Teng, H., G. Branstator, H. Wang, G. A. Meehl, and W. M. Washington, 2013: Probability of US heat waves affected by a subseasonal planetary wave pattern. *Nat. Geosci.*, **6**, 1056–1061, https://doi.org/10.1038/ngeo1988.

Ting, M., Y. Kushnir, R. Seager, and C. Li, 2009: Forced and internal twentieth-century SST trends in the North Atlantic. *J. Climate*, **22**, 1469–1481, https://doi.org/10.1175/2008JCLI2561.1.

Trenberth, K. E., and J. T. Fasullo, 2012: Climate extremes and climate change: The Russian heat wave and other climate extremes of 2010. *J. Geophys. Res.*, **117**, D17103, https://doi.org/10.1029/2012JD018020.

Tseng, K.-C., and Coauthors, 2021: Are multiseasonal forecasts of atmospheric rivers possible? *Geophys. Res. Lett.*, **48**, e2021GL094000, https://doi.org/10.1029/2021GL094000.

Venzke, S., M. R. Allen, R. T. Sutton, and D. P. Rowell, 1999: The atmospheric response over the North Atlantic to decadal changes in sea surface temperature. *J. Climate*, **12**, 2562–2584, https://doi.org/10.1175/1520-0442(1999)012<2562:TAROTN>2.0.CO;2.

Vitart, F., 2005: Monthly forecast and the summer 2003 heat wave over Europe: A case study. *Atmos. Sci. Lett.*, **6**, 112–117, https://doi.org/10.1002/asl.99.

——, and A. Robertson, 2018: The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. *npj Climate Atmos. Sci.*, **1**, 3, https://doi.org/10.1038/s41612-018-0013-0.

White, C., D. Hudson, and O. Alves, 2014: ENSO, the IOD and the intraseasonal prediction of heat extremes across Australia using POAMA-2. *Climate Dyn.*, **43**, 1791–1810, https://doi.org/10.1007/s00382-013-2007-2.

Wu, Y., M. Latif, and W. Park, 2016: Multiyear predictability of Northern Hemisphere surface air temperature in the Kiel Climate Model. *Climate Dyn.*, **47**, 793–804, https://doi.org/10.1007/s00382-015-2871-z.

Xiang, B., S. J. Lin, M. Zhao, N. Johnson, X. Yang, and X. Jiang, 2019: Subseasonal week 3–5 surface air temperature prediction during boreal wintertime in a GFDL model. *Geophys. Res. Lett.*, **46**, 416–425, https://doi.org/10.1029/2018GL081314.

Yang, X., G. A. Vecchi, R. G. Gudgel, T. Delworth, and S. Zhang, 2015: Seasonal predictability of extratropical storm tracks in GFDL's high-resolution climate prediction model. *J. Climate*, **28**, 3592–3611, https://doi.org/10.1175/JCLI-D-14-00517.1.

Zhang, G., and Coauthors, 2021: Seasonal predictability of baroclinic wave activity. *npj Climate Atmos. Sci.*, **4**, 50, https://doi.org/10.1038/s41612-021-00209-3.

Zhang, J., Z. Yang, and L. Wu, 2018: Skillful prediction of hot temperature extremes over the source region of ancient Silk Road. *Sci. Rep.*, **8**, 6677, https://doi.org/10.1038/s41598-018-25063-x.

——, ——, ——, and K. Yang, 2019: Summer high temperature extremes over northeastern China predicted by spring soil moisture. *Sci. Rep.*, **9**, 12577, https://doi.org/10.1038/s41598-019-49053-9.

Zhang, L., and T. Delworth, 2015: Analysis of the characteristics and mechanisms of the Pacific decadal oscillation in a suite of coupled models from the Geophysical Fluid Dynamics Laboratory. *J. Climate*, **28**, 7678–7701, https://doi.org/10.1175/JCLI-D-14-00647.1.

Zhang, Y., J. M. Wallace, and D. S. Battisti, 1997: ENSO-like interdecadal variability: 1900–93. *J. Climate*, **10**, 1004–1020, https://doi.org/10.1175/1520-0442(1997)010<1004:ELIV>2.0.CO;2.

Zhao, M., and Coauthors, 2018: The GFDL global atmosphere and land model AM4.0/LM4.0: 1. Simulation characteristics with prescribed SSTs. *J. Adv. Model. Earth Syst.*, **10**, 691–734, https://doi.org/10.1002/2017MS001208.