AGU ADVANCING EARTH AND SPACE SCIENCE

Geophysical Research Letters

RESEARCH LETTER

10.1029/2022GL098554

**Key Points:**

- Heavy rainfall explains the majority of the year-to-year variation of summer precipitation in the US Midwest
- SSS in the tropical Pacific and subtropical North Atlantic equally contribute to skillful predictions of Midwest heavy rain a season ahead
- Skillful prediction from SSS is realized through tropical-extratropical teleconnections and local soil moisture feedback

**Supporting Information:**

Supporting Information may be found in the online version of this article.

**Correspondence to:**

L. Li,
lf15240@psu.edu

**Citation:**

Li, L., Schmitt, R. W., & Ummenhofer, C. C. (2022). Skillful long-lead prediction of summertime heavy rainfall in the US Midwest from sea surface salinity. *Geophysical Research Letters*, *49*, e2022GL098554. https://doi.org/10.1029/2022GL098554

Received 4 MAR 2022
Accepted 30 MAY 2022

**Author Contributions:**

**Conceptualization:** Laifang Li, Raymond W. Schmitt
**Data curation:** Laifang Li
**Formal analysis:** Laifang Li
**Funding acquisition:** Laifang Li, Raymond W. Schmitt, Caroline C. Ummenhofer
**Investigation:** Laifang Li, Raymond W. Schmitt, Caroline C. Ummenhofer
**Methodology:** Laifang Li
**Project Administration:** Raymond W. Schmitt, Caroline C. Ummenhofer
**Resources:** Laifang Li
**Supervision:** Raymond W. Schmitt, Caroline C. Ummenhofer
**Validation:** Laifang Li, Raymond W. Schmitt, Caroline C. Ummenhofer
**Visualization:** Laifang Li
**Writing – original draft:** Laifang Li



# Skillful Long-Lead Prediction of Summertime Heavy Rainfall in the US Midwest From Sea Surface Salinity

**Laifang Li[1,2,3], Raymond W. Schmitt[4], and Caroline C. Ummenhofer[4]**

[1]Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, PA, USA, [2]Earth and Environmental Science Institute, The Pennsylvania State University, University Park, PA, USA, [3]Institute of Computational and Data Science, The Pennsylvania State University, University Park, PA, USA, [4]Department of Physical Oceanography, Woods Hole Oceanographic Institution, Woods Hole, MA, USA

**Abstract** Summertime heavy rainfall and its resultant floods are among the most harmful natural hazards in the US Midwest, one of the world's primary crop production areas. However, seasonal forecasts of heavy rain, currently based on preseason sea surface temperature anomalies (SSTAs), remain unsatisfactory. Here, we present evidence that sea surface salinity anomalies (SSSAs) over the tropical western Pacific and subtropical North Atlantic are skillful predictors of summer time heavy rainfall one season ahead. A one standard deviation change in tropical western Pacific SSSA is associated with a 1.8 mm day$^{-1}$ increase in local precipitation, which excites a teleconnection pattern to extratropical North Pacific. Via extratropical air-sea interaction and long memory of midlatitude SSTA, a wave train favorable for US Midwest heavy rain is induced. Combined with soil moisture feedbacks bridging the springtime North Atlantic salinity, the SSSA-based statistical prediction model improves Midwest heavy rainfall forecasts by 92%, complementing existing SSTA-based frameworks.

**Plain Language Summary** Predicting heavy rainfall one season ahead is challenging for the US Midwest, a primary crop production area in the world. In this study, we show a skillful prediction of Midwest heavy rain using sea surface salinity anomalies (SSSAs) in the western tropical Pacific and subtropical North Atlantic. Compared to sea surface temperature anomalies which traditionally form the basis for seasonal precipitation forecasts, our newly identified SSSA-based predictors improve the accuracy of heavy rainfall prediction by 92%. This superior skill of SSSA-based prediction appears to result from a close relationship between salinity variations and the oceanic water cycle, as well as related atmospheric circulation changes and soil moisture feedback.

## 1. Introduction

The hazardous impacts of heavy rainfall on agriculture and the economy in the US Midwest cannot be overstated (Andersen et al., 2001; Ray et al., 2015). The Great Flood of 1993 in the US Midwest was caused by inundation from extensive summertime rainfall and led to economic loss in excess of $12B in the region (Gumley & King, 1995; Kunkel et al., 1994). Thus, skillful long-lead prediction of heavy rainfall is particularly important (White et al., 2017).

These heavy rainfall events are generated in unique atmospheric circulation patterns associated with distinct thermodynamic conditions (Li & O'Gorman, 2020), making heavy rainfall challenging to predict (e.g., Doblas-Reyes et al., 2013). Studies found excessive US Midwest summer rain more likely following an El Niño sea surface temperature anomaly (SSTA) and during a negative Atlantic Multidecadal Oscillation (e.g., Barlow et al., 2001; Englehart & Douglas, 2003; Higgins et al., 2004; McCabe et al., 2004; Schubert et al., 2016). However, when implemented to empirical seasonal forecasts, these SSTA-based predictors provide moderate skill that hardly meets the socioeconomical needs (Carter et al., 2021; Li et al., 2018), which motivates this search for more skillful predictors. These predictors can be sought from the oceanic water cycle, as ocean is the ultimate moisture source for terrestrial precipitation (van der Ent et al., 2010; Gimeno et al., 2010, 2020), and the influence of oceanic moisture rises as rainfall intensity increases (Chan & Misra, 2010; Jana et al., 2018; Y. Liu et al., 2021). Thus, a variable indicative of ocean moisture export should provide predictive skills, complementing SSTA-based predictions.

**Writing – review & editing:** Laifang Li, Raymond W. Schmitt, Caroline C. Ummenhofer

Oceanic moisture export is also a forcing mechanism for sea surface salinity (SSS) (Durack, 2015; Vinogradova et al., 2019). Consequently, spatiotemporal SSS variations are highly sensitive to moisture flux from the ocean: high (low) SSS is located over the subtropical (tropical and midlatitude) oceans with a net moisture export (input) (Gordon, 2016; Schanze et al., 2010). Since the 1950s, this SSS spatial gradient has been enhanced by an intensifying oceanic water cycle (Durack et al., 2012; Yu et al., 2020).

The close connection between SSS and the oceanic water cycle, and the influence of the oceanic moisture on terrestrial precipitation suggest that the upstream SSS anomaly (SSSA) could be a skillful predictor of precipitation on land, with positive (negative) SSSA indicating greater (weaker) moisture export. In addition, the oceanic moisture export in the form of latent heat fuels atmospheric teleconnection, an important source of predictability for regional climate (Scaife et al., 2017). Encouraging skill of SSSA-based seasonal rainfall forecasts was shown for the African Sahel (Li et al., 2016a), the United States (Li et al., 2016b; T. Liu et al., 2018), China (Zeng et al., 2019), and Australia (Rathore et al., 2021). Here, we present an additional line of evidence that springtime SSSA in the tropical Pacific and subtropical North Atlantic are skillful predictors of summertime heavy rainfall in the US Midwest. The long-lead forecast based on SSSA surpasses that based on preseason SSTA by 92%, indicating a substantially increased prediction skill with important implications for water resource management in summer.

## 2. Data and Methods

### 2.1. Bayesian Inference on Summer-Season Heavy Rainfall Events Over the US Midwest

A Bayesian three-cluster Gaussian mixture model (GMM; L. Li & Li, 2013) is implemented in this study to empirically and objectively categorize Midwest daily precipitation (1948–2019) into light, moderate, and heavy rainfall. The likelihood function is formulated as:

$$y_i|\pi,\mu,\phi \sim \sum_{\mathrm{h}=1}^{3} \pi_{\mathrm{h}} \mathrm{Gauss}\left(y_{\mathrm{i}}|\mu_{\mathrm{h}},\phi_{\mathrm{h}}^{-1}\right) \tag{1}$$

where $y_i = \ln Pr_i$ is the natural log of precipitation in the $i$th day, $\pi$ is cluster weight (i.e., the frequency of each rainfall category), $\mu$ is cluster mean (i.e., the intensity of each rainfall category), and $\phi$ is the precision of a given Gaussian distribution. $h \in \{1,2,3\}$ is the cluster index, which respectively represents light, moderate, and heavy rainfall (Supporting Information). The distribution parameters ($\pi$,$\mu$,$\phi$) are derived using Bayesian inference and sampled using the Markov Chain Monte Carlo algorithm (Gelfand, 2000; Equations S1–S5 in Supporting Information S1).

The formulation of the GMM (Equation 1) implies that summertime cumulative precipitation equals the weighted sum of three rainfall clusters: $\sum_{i=1}^{N} Y_i|\pi,\mu,\phi = N\left(\sum_{h=1}^{3} \pi_h\mu_h\right)$, where $N$ is the total number of samples. Based on this relationship, the cluster contribution to US Midwest summertime cumulative precipitation can be evaluated.

### 2.2. Logistic Regression

To assess the skills of the preseason SSTA and SSSA in forecasting Midwest heavy rainfall frequency ($\pi_3 \in [0,1]$ from the GMM), a logistic regression model is constructed:

$$\pi_3 = \frac{1}{1 + e^{-\theta^T x + \varepsilon}} \tag{2}$$

In Equation 2, $x$ is the SSTA-based and SSSA-based predictors, and $\theta$ is the regression coefficient upon the predictors. $\varepsilon \sim N\,(0,\sigma^2)$ is the residual of the regression which is normally distributed with zero mean and variance of $\sigma^2$.

In constructing the regression model, we randomly draw half the data as training samples and the remaining half are independent samples for validation. The regression coefficients are first derived from the training sample and then applied to the validation samples to estimate the prediction errors. This training-validation process is repeated 10,000 times. The parameter set that minimizes the prediction errors of the validation samples is selected to create the final logistic regression model that is applied to the entire data set. With the validation

10.1029/2022GL098554

samples overlapping with half of the sample data set, the estimated prediction skill represents the upper bound of the logistic regression model.

The performance of the logistic regression is evaluated based on the portion of observed variance explained by the prediction model and calculated as $R^2 = 1 - (SS_{res}/SS_{tot})$. Here, $SS_{tot} = \sum_{j=1}^{N_Y} \left(\pi_{3j} - \overline{\pi}_{3j}\right)^2$ is the total variance of $\pi_3$, and $SS_{res} = \sum_{j=1}^{N_Y} (\hat{\pi}_{3j} - \pi_{3j})^2$ quantifies the variance of $\pi_3$ unexplained by the prediction model. The index $j$ refers to the $j$th year, the $\overline{\pi}_{3j}$ is the mean of the 1948–2019, and the $\hat{\pi}_{3j}$ the prediction by the logistic regression model.

### 2.3. Composite Analysis and Significance Test

Daily rainfall categories are assigned based on the cluster probability (Equation S4 in Supporting Information S1). Accordingly, atmospheric circulation and moisture fluxes (MFs) corresponding to light, moderate, and heavy rainfall are calculated using composite analysis. The vertically integrated MF is calculated as $\overrightarrow{Q} = \frac{1}{g}\int_{200\text{hPa}}^{P_s} q\overrightarrow{V}_h dp$, where $q$ is specific humidity (kg kg$^{-1}$), $\overrightarrow{V}_h$ is horizontal wind (m s$^{-1}$), $p$ is pressure, $P_s$ is surface pressure, and $g = 9.8$ m s$^{-2}$ is gravitational acceleration. The horizontal MF divergence (MFD) is expressed as $\frac{1}{g}\nabla_h \cdot \int_{200\text{hPa}}^{P_s} q\overrightarrow{V}_h dp$. The significance level of the composite analysis is determined using 1000-iteration Monte Carlo simulations (Livezey & Chen, 1983).

### 2.4. Data Sets

The precipitation data (0.25° × 0.25°) are from the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC) US Unified Precipitation for 1948–2006 and from the Real-Time US Daily Precipitation Analysis for 2007–2019 (Higgins et al., 2000). The US Midwest is defined as the area of 106°–90°W, 36°–49°N, and daily precipitation is area-averaged over this domain. The variables used to quantify moisture flux and atmospheric circulation are from ERA-Interim (1979–2019; Dee et al., 2011) and NCEP/NCAR reanalysis (1948–2019; Kalnay et al., 1996), which differ in temporal coverage and spatial resolution but produces similar results. The SST and SSS data are from the Met Office (UKMO) EN4.2.1 data set objectively analyzed with 1° × 1° spatial resolution and monthly temporal resolution for 1948–2019 (Good et al., 2013). EN4.2.1 is the only SSS product with sufficient temporal coverage for this study. In addition, we have verified our SSTA analysis using available SST data sets including ERSST v5, HadISST, and COBE. In all our analyses, the SSSA and SSTA are calculated by removing the seasonal cycle and linear trend.

## 3. Results

### 3.1. Heavy Precipitation Dominates US Midwest Summer Precipitation Variance

The Bayesian GMM (Equation 1 and Equations S1–S5 in Supporting Information S1) objectively classifies summertime (JJA) daily precipitation in the US Midwest into light, moderate, and heavy rainfall. During 1948–2019, 47% (±3%) of the events are classified as light rainfall, 39% (±3%) as moderate rainfall, and 14% (±2%) as heavy rainfall (Figure S1 in Supporting Information S1). Over the entire season, the US Midwest receives an accumulation of 261.7 mm of precipitation, with 38.5 (±7.2) mm from light rainfall, 121.9 (±23.7) mm from moderate rainfall, and 101.3 (±43.3) mm from the heavy rainfall cluster (Figure 1a and Figure S2 in Supporting Information S1). While the total amount of the seasonal precipitation is largely maintained by moderate rainfall, its interannual variation is almost exclusively explained by heavy rain. The $R^2$ between the heavy rainfall cluster and season-total precipitation is 0.81, that is, more than 80% of the summer rainfall variability can be explained by heavy rainfall (Figure 1a). In contrast, the variance explained by moderate (light) rainfall is lower at 0.29 (negative at −0.21) (Figure 1a).

Our analysis suggests that the impacts of heavy rainfall go beyond daily weather in the region and exert climatic impacts on water resources in the summer. Further, heavy rainfall frequency is overwhelmingly more important than intensity to interannual variation of season-mean precipitation (Figures 1b and 1c). Specifically, the correlation between seasonal-mean precipitation and heavy rainfall intensity is 0.18 ($p = 0.13$) but reaches 0.90

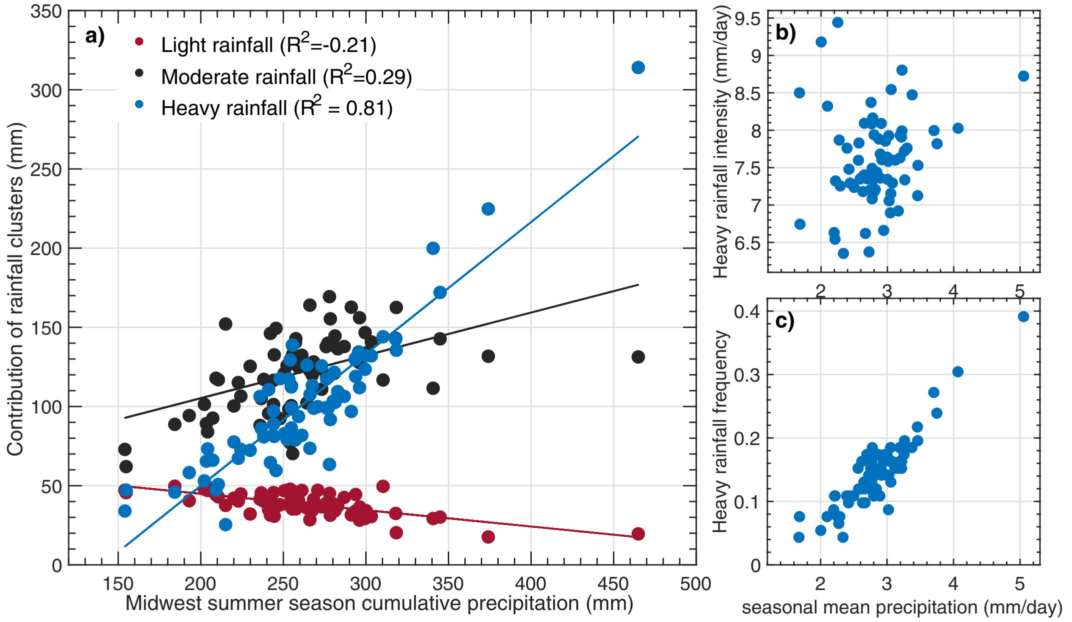


**Figure 1.** (a) Contribution of the light (red dots), moderate (black dots), and heavy (blue dots) rainfall to Midwest summertime cumulative precipitation. The lines are the linear regression of cluster total precipitation against summertime cumulative precipitation. (b) and (c) are, respectively, the scatter plots of summer-season precipitation versus heavy rainfall intensity and frequency.

($p \leq 0.00$) with heavy rainfall frequency (Figures 1b and 1c). The dominance of heavy rainfall frequency is exemplified by 1993, the wettest year during the 1948–2019 period, with more than 30 summer days classified as experiencing heavy rainfall—the most days in that interval.

### 3.2. Atmospheric Circulation Bridging the Oceanic Moisture to Precipitation in the US Midwest

Heavy rainfall in the US Midwest is fed by anomalously strong northward MF and increased MF convergence (MFC; Figure 2a). Accompanying the increased MFC in the US Midwest, the southwestern US and the adjacent oceans are supplying more moisture as evidenced by increased MFD in these regions (Figure 2a). The increased moisture export from these regions can be observed 3–6 days ahead of the heavy rainfall events, even though the precedent MFD is less apparent (Figures 2b and 2c).

The increased MF into the US Midwest is facilitated by a barotropic wave-train across North America (Figures 2d–2f and Figure S3 in Supporting Information S1). At 500 hPa, an anomalous high pressure (anticyclone) is observed off the Gulf of Alaska (GOA), with a low pressure (cyclone) over the western US and a high pressure (anticyclone) over the eastern US (Figure 2d). The pressure distribution over North America promotes strong ascending motion over the US Midwest, according to the quasi-geostrophic-Omega equation (Hoskins et al., 1985), which is a dynamic necessity to sustain the observed heavy rainfall (Li & O'Gorman, 2020). In the lower troposphere (925 hPa), the barotropic wave train enhances the Great Plains low-level jet (GPLLJ), a key feature responsible for MF and extreme precipitation in the US Midwest (Cook et al., 2011; Weaver & Nigam, 2008). At the same time, the anticyclone over the eastern US leads to anomalous mass divergence off Florida and in the Caribbean, favoring the oceanic region to export excessive moisture (Figures 2a–2c). Along with a westward extension of the Caribbean low-level jet into the Gulf of Mexico (GOM; Figure 2d), the excessive moisture from the subtropical ocean is conveyed to the Southern US and is connected by the GPLLJ to supply moisture to the Midwest (Dirmeyer & Kinter, 2010).

Overall, the wave train favors MF from adjacent land and ocean to the central US and promotes lift condensation for MFC over the Midwest. As a result, heavy precipitation preferentially occurs in this type of circulation pattern,

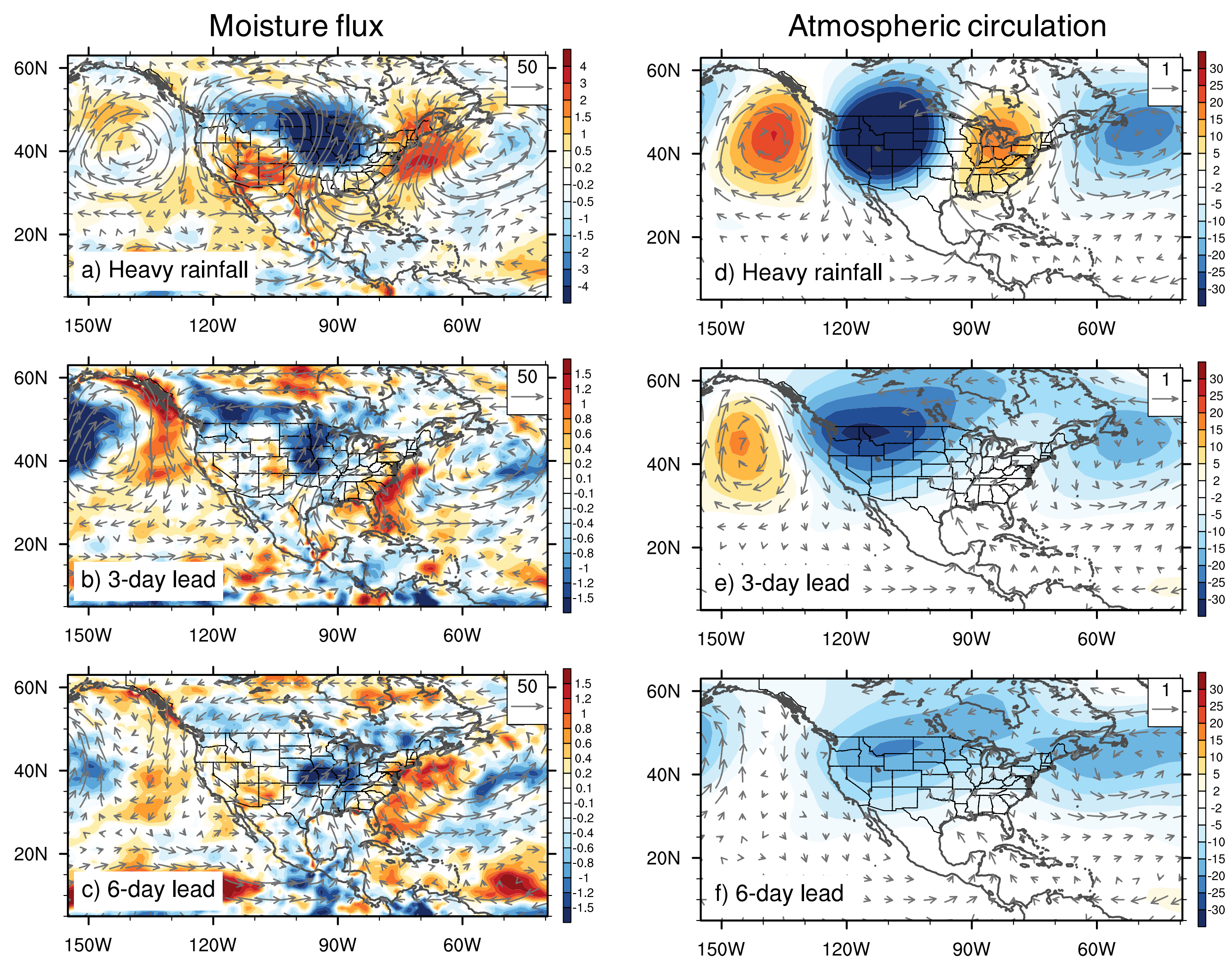


**Figure 2.** (a)–(c) are anomalous vertically integrated moisture flux (vectors; kg m$^{-1}$s$^{-1}$) and moisture flux divergence (shaded; mm day$^{-1}$) during and before summertime heavy rainfall events; (d)–(f) are 500 hPa geopotential height anomaly (shaded; gpm) and 925 hPa wind anomaly (vectors; m s$^{-1}$).

consistent with Zhang and Villarini (2019) recognizing this wave train as a weather regime favorable to floods in the US Midwest.

### 3.3. Springtime Salinity Provides Predictive Values to US Midwest Heavy Rainfall

We explore oceanic precursors of US Midwest heavy rainfall by regressing March-April-May (MAM) SSTA and SSSA upon summertime heavy rainfall frequency ($\pi_3$ in Equation 1 and Equation S3 in Supporting Information S1). More frequent heavy rainfall would be expected following an anomalously warm (cold) tropical central Pacific (GOM) (McCabe et al., 2004) (Figure 3a). Corresponding to the SSTA, the center of tropical convection is shifted near the dateline, resulting in a dipole SSSA with anomalously low salinity along the equator and high salinity to the south (Figure 3b; Delcroix et al., 2011; Guimbard et al., 2017). The low equatorial SSSA (150°E–180°W, 5°S–5°N) reaches −0.1 PSU (Figure 3b). With the typical SSS over this region being 33PSU, mixed layer depth of 15m, and assuming evaporation is relatively constant ($SSSA \sim -\frac{\overline{S}P'\Delta t}{\rho_0 h_m}$), the observed SSSA could correspond to a 1.8 mm day$^{-1}$ increase in precipitation ($P'$) in the spring, forcing atmospheric teleconnection patterns and thus the response in extratropical precipitation (Horel & Wallace, 1981; Liu & Alexander, 2007).

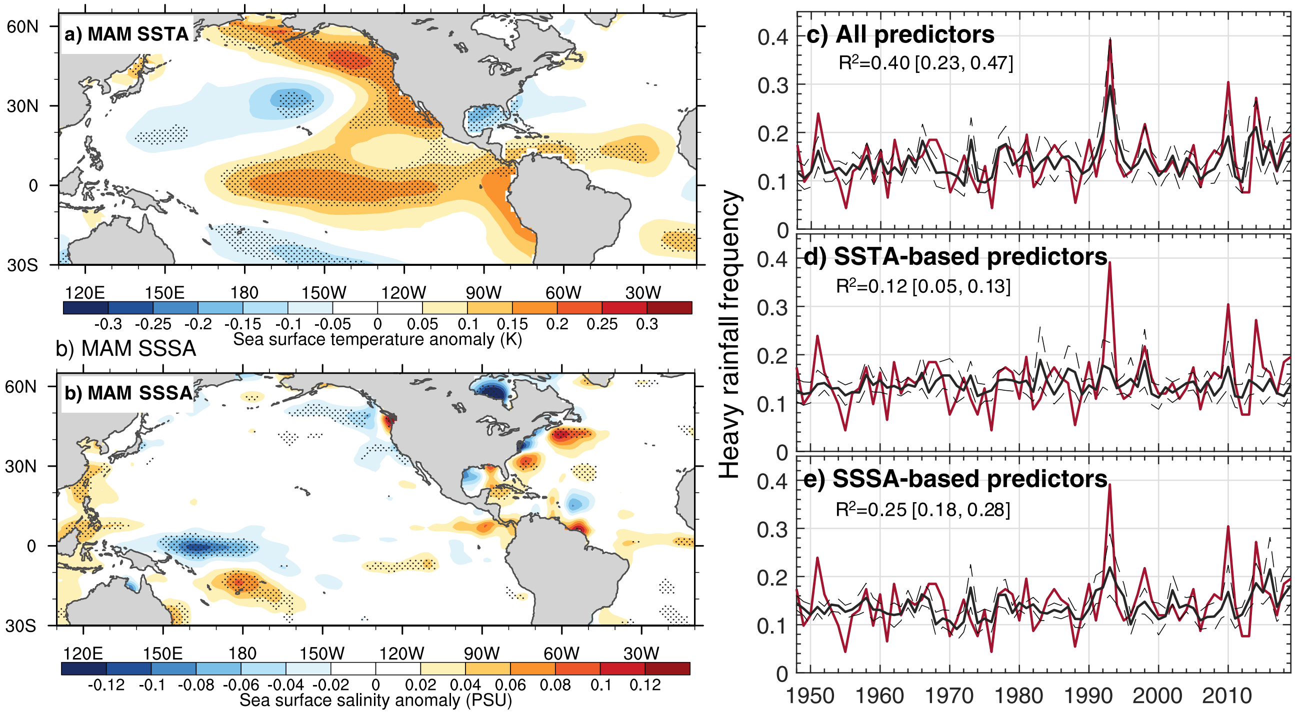


**Figure 3.** Pre-season SSTA (a) and SSSA (b) regressed upon Midwest heavy rainfall frequency. The shaded are the regression coefficients interpreted as SSTA or SSSA per standard deviation of heavy rainfall frequency. The regression coefficients significant at $\alpha = 0.05$ level per 1000-trial Monte Carlo simulations are stippled; (c)–(e) are logistic regression model's prediction of Midwest heavy rainfall frequency using SSTA-based and SSSA-based predictors: (c) all four predictors; (d) only SSTA-based predictors; (e) only SSSA-based predictors. The red curves are the observations, the solid black curves are the average of the 10,000-trial prediction by the logistic regression model, and the dashed black curves are the 95% uncertainty range. The $R^2$ reported is the median of the 10,000-trial prediction with the 95% uncertainty range in the bracket.

In the North Atlantic, the preseason SSSA is most significant off the Eastern coast of US, where a positive SSSA of 0.06–0.08 PSU is observed (Figure 3b), consistent with our previous studies linking a saltier subtropical North Atlantic to increased summer-season precipitation over the Midwest (Li et al., 2016b, 2018).

Based on preseason SSTA and SSSA (Figures 3a and 3b), we define four predictors: Niño 3.4 SSTA, GOM SSTA, tropical western Pacific SSSA dipole index, and North Atlantic SSSA as they show the most significant preseason signal (Figure S4 in Supporting Information S1). Using these predictors, we subsample the precipitation datasets and re-run the Bayesian GMM (Supporting Information) to evaluate the changes in heavy rainfall frequency corresponding to preseason SSTA and SSSA (Table S1 in Supporting Information S1). Overall, a combination of a warm tropical Pacific and cold GOM in spring tends to be followed by a 5.2% increase in summertime heavy rainfall frequency versus the opposite scenario (Figures S6a–S6c in Supporting Information S1). Compared to SSTA, the heavy rainfall frequency is more sensitive to SSSA-based predictors. The Pacific salinity dipole (North Atlantic SSSA) alone corresponds to a 5.1% (5.7%) change in heavy rainfall frequency (Figures S6d and S6e in Supporting Information S1). Collectively, a positive Pacific salinity dipole and a salty subtropical North Atlantic could increase the heavy rainfall frequency to 22.6%, almost doubling the occurrence in the opposite scenario (12.1%; Figure S6f in Supporting Information S1).

Further, we construct a logistic regression model (Equation 2) to predict summer-season heavy rainfall frequency (Figures 3c–3e). The average $R^2$ between the observations and the prediction reaches 0.40 [0.23, 0.47] (Figure 3c). In particular, the model predicts a doubling of heavy rainfall events in the wettest summer of 1993 and reduced heavy rainfall events in the dry summer of 2012 (Figure 3c), suggesting the capability of the logistic regression model to forecast extreme events one season ahead. It is noteworthy that SSSA-based predictors are twice as skillful ($R^2 = 0.25$ [0.18, 0.28]; Figure 3e) than SSTA-based predictors ($R^2 = 0.12$ [0.05, 0.13]; Figure 3d and Figure S6 in Supporting Information S1). Thus, incorporating the SSSA-based predictors into the statistical model could substantially improve the long-lead forecast of US Midwest heavy rainfall.

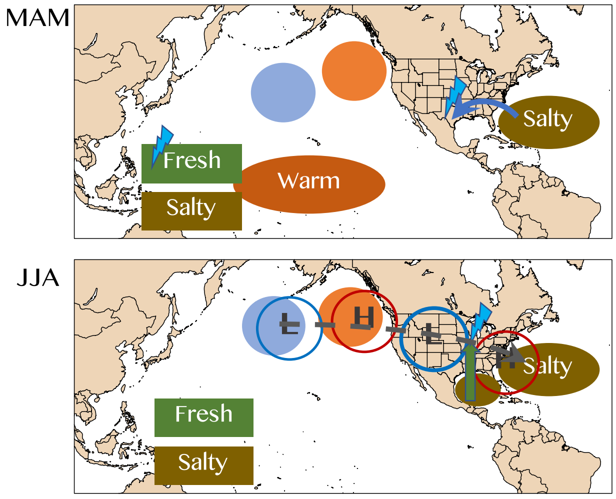


**Figure 4.** Schematic figure of salinity as a predictor of US Midwest heavy rainfall frequency. The red (blue) patches mark the regions with warm (cold) SSTA, and the brown (green) patches mark the high (low) SSSA region. The blue (green) arrow in the upper (lower) panel denotes the direction of moisture flux (intensified GPLLJ). The dashed arrow marks the direction of wave train across the Northeast Pacific and the US continent.

Among the four predictors, the Pacific SSSA dipole significantly correlates with Niño 3.4 (Figure S4 in Supporting Information S1) but is more skillful in predicting Midwest heavy rainfall. The superior skill of the North Pacific SSSA dipole is attributable to tropical convection as the dominant forcing on the SSSA (Sean-Martins & Stammer, 2015). Without a damping mechanism, the SSSA reflects cumulative tropical convection and latent heating released into the atmosphere, a driver of atmospheric teleconnection (Gill et al., 1980; Matsuno, 1966; Scaife et al., 2017). Via tropical-extratropical teleconnection patterns, the upstream SSSA dipole turns out to be a more skillful predictor of precipitation on land.

### 3.4. Mechanisms of SSSA-Heavy Rainfall Relationship

To better understand the sources of prediction skill from preseason SSSA, we explore the potential mechanisms. The one-season time lag between subtropical North Atlantic SSSA and summer-season precipitation in the US Midwest is bridged by the soil moisture feedback (Li et al., 2016b, 2018; Supporting Information). Here, we thus focus on the mechanism responsible for the Pacific SSSA dipole (Figures S7 and S8 in Supporting Information S1).

The seasonal evolution of SSTA shows that the positive SSSA dipole in the tropical Pacific is preceded by a decaying El Niño Modoki in the Central Pacific, as the warmest SSTA (1.5K) is located at 150°W along the equator in December-January-February (Ashok et al., 2007). Over the extratropical North Pacific, cold SSTA (−1.0K) is observed in the center of the subtropical gyre and surrounded by warm SSTA (0.4K) along the US West Coast and the GOA (Figure S7a in Supporting Information S1). In the subsequent seasons, tropical SSTA decays, while the SSTA in the extratropical North Pacific persists (Figures S7b–S7f in Supporting Information S1). In the summer, the most significant SSTAs are observed over the northern part of the subtropical gyre (−1.0K) and the GOA (0.6K) (Figure S7f in Supporting Information S1).

The initial El Niño Modoki and the different persistence time of tropical and extratropical SSTA can be responsible for the seasonal evolution of precipitation and atmospheric circulation (Figure S8 in Supporting Information S1). Before the formation of the Pacific SSSA dipole, increased precipitation is observed near the dateline (Figures S8a and S8b in Supporting Information S1), collocated with negative SSSA (Figure 4b). In response to the changes in tropical convection, the 500 hPa geopotential height shows a typical Pacific-North American pattern, with anomalous high pressure over the (sub) tropical North Pacific and subpolar regions, and low pressure over the mid-latitudes and southern US (Figures S8a–S8c). As the season progresses and tropical SSTA decays, the zonally symmetric structure in the 500 hPa geopotential height anomalies loses significance (Figures S8d–S8f in Supporting Information S1). Instead, a stationary wave train, resembling the typical circulation for US Midwest heavy rainfall (Figure 2d), is formed of the GOA and extends into the Conterminous US (Figures S8e and S8f in Supporting Information S1).

This wave train likely arises from a positive coupling between SSTA, precipitation, and pressure. The anomalous high pressure off the GOA depresses precipitation (Figures S8e and S8f in Supporting Information S1), and weakens the prevailing wind associated with the Aleutian Low, which increases local SSTA (Figures S7e and S7f in Supporting Information S1; Bond et al., 2015). The high SSTA, in turn, uplifts the 500 hPa isobaric surface and further intensifies the high pressure and suppresses precipitation (Figures S8e and S8f in Supporting Information S1). In addition, the decreased precipitation reduces the condensational heating released into the local atmospheric column, reinforcing the high pressure off the GOA (Figures S8e and S8f in Supporting Information S1). The SSTA, pressure patterns, and anomalous precipitation induce and maintain the 500 hPa wave train (Park & Lee, 2019), which modulates the vertical motion and GPLLJ over the central US, leading to increased heavy rainfall frequency over the US Midwest (Figure 2).

## 4. Conclusions

According to the Bayesian model (Equation 1 and Equations S1–S5 in Supporting Information S1) applied to summertime daily precipitation in the US Midwest, heavy rainfall events are primarily responsible for the year-to-year variation of cumulative seasonal precipitation, a meteorological variable that significantly impacts yields in this important crop production region. These heavy rainfall events preferentially occur in atmospheric circulation patterns that feature a barotropic wave train across the GOA and North America. The wave train promotes northward MF in the Central US and ascending motion over the US Midwest, and thus facilitates heavy rain.

Here, we explore preseason oceanic state variables for predicting summer-season heavy rainfall in the US Midwest. Based on the logistic regression, skillful prediction of US Midwest heavy rain is achieved one season ahead with the knowledge of SSTA over the Niño 3.4 region and GOM and SSSA over the tropical western Pacific and subtropical North Atlantic. The combination of the four predictors leads to prediction with $R^2 = 0.40$ [0.23, 0.47], in which the SSSA-based predictors ($R^2 = 0.25$ [0.18, 0.28]) provide superior skill compared to SSTA-based predictors ($R^2 = 0.12$ [0.05, 0.13]). Thus, including the newly identified SSSA indices would substantially improve seasonal forecasts of Midwest heavy rain.

The potential mechanism underlying the SSSA's predictive skill is summarized in Figure 4. Overall, the Pacific salinity dipole is indicative of tropical convection and the resultant tropical-extratropical teleconnection (Figure 4a and Figures S7 and S8 in Supporting Information S1). Through the long memory of extratropical SSTA, and the coupling between SSTA, pressure, and precipitation, a quasi-barotropic wave train that resembles the typical circulation for US Midwest flooding events is triggered and maintained (Figures S7 and S8 in Supporting Information S1). In addition, the subtropical North Atlantic SSSA is skillful in predicting summer-season heavy rain mainly due to a positive soil moisture feedback mechanism as identified previously (Figure 4; Supporting Information). With the cumulative effects of soil moisture feedback and tropical-extratropical teleconnection, the combination of Pacific SSSA dipole and North Atlantic SSSA provides indispensable value to predict US Midwest heavy rainfall, complementing the current statistical forecasting based solely on preseason SSTA.

Furthermore, the newly identified preseason SSSA could become more valuable as the climate warms: ENSO teleconnections are sensitive to a changing background climate in response to natural and anthropogenic forcing (McPhaden et al., 2011; Yeh et al., 2019). In addition, the oceanic water cycle is predicted to intensify in the future (Durack et al., 2012; Levang & Schmitt, 2015; Skliris et al., 2020), and the oceanic moisture export more significantly influences the moisture balance on land (Findell et al., 2019; Gimeno et al., 2020). SSS is a standard ocean state variable that has been routinely measured back to the late nineteenth century. The long observational records provide an enriched data source to train and validate prediction models, a critical step toward improved operational forecasts of high-impact hydroclimatic events. Thus, sustained measurement and monitoring of SSS will greatly benefit the prediction of summer-season heavy rainfall and early warning of flooding events for the US Midwest.

## Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

## Data Availability Statement

All of the data analyzed in this study are from publicly accessible data sets: The NOAA CPC US Unified Precipitation by Higgins et al. (2000); https://psl.noaa.gov/data/gridded/data.unified.daily.conus.html. The ERA-Interim reanalysis of atmospheric circulation by Dee et al. (2011); https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=ml/. The NCEP/NCAR reanalysis of atmospheric circulation by Kalnay et al. (1996); https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html. The UKMO EN4.2.1 ocean salinity and temperature by Good et al. (2013); https://www.metoffice.gov.uk/hadobs/en4/download-en4-2-1.html. *Code availability*: The code of the Bayesian GMM has been deposited and made available at: https://doi.org/10.5281/zenodo.5389218.

**Acknowledgments**
This study is supported by the NSF PREEVENTS program under ICER-1663138 (LL) and ICER-1663704 (RWS and CCU). Helpful comments from two anonymous reviewers on an earlier version of the manuscript are gratefully acknowledged.

## References

Andresen, J. A., Alagarswamy, G., Ritchie, J. T., Rotz, C. A., & LeBaron, A. W. (2001). Assessment of the impact of weather on maize, soybean, and alfalfa production in the Upper Great Lakes Region of the United States, 1895–1996. *Agronomy Journal*, *93*, 1059–1070. https://doi.org/10.2134/agronj2001.9351059x

Ashok, K., Behera, S. K., Rao, S. A., Weng, H., & Yamagata, T. (2007). El Niño Modoki and its possible teleconnection. *Journal of Geophysical Research*, *112*, C11007. https://doi.org/10.1029/2006jc003798

Barlow, M., Nigam, S., & Berbery, E. H. (2001). ENSO, Pacific decadal variability, and US summertime precipitation, drought, and stream flow. *Journal of Climate*, *14*, 2105–2128. https://doi.org/10.1175/1520-0442(2001)014<2105:epdvau>2.0.co;2

Bond, N. A., Cronin, M. F., Freeland, H., & Mantua, N. (2015). Causes and impacts of the 2014 warm anomaly in the NE Pacific. *Geophysical Research Letters*, *42*, 3414–3420. https://doi.org/10.1002/2015gl063306

Carter, E., Herrera, D. A., & Steinschneider, S. (2021). Feature engineering for subseasonal to seasonal warm-season precipitation forecasts in the midwestern US: Towards a unifying hypothesis of anomalous warm-season hydroclimatic circulation. *Journal of Climate*, *34*, 8291–8318. https://doi.org/10.1175/jcli-d-20-0264.1

Chan, S. C., & Misra, V. A. (2010). Diagnosis of the 1979-2005 extreme rainfall events in the Southeast US with isentropic moisture tracing. *Monthly Weather Review*, *138*, 1172–1185. https://doi.org/10.1175/2009mwr3083.1

Cook, B. I., Seager, R., & Miller, R. L. (2011). On the causes and dynamics of the early twentieth-century North American pluvial. *Journal of Climate*, *24*, 5043–5060. https://doi.org/10.1175/2011jcli4201.1

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*, 553–597. https://doi.org/10.1002/qj.828

Delcroix, T., Alory, G., Cravatte, S., Corrège, T., & McPhaden, M. J. (2011). A gridded sea surface salinity data set for the tropical Pacific with sample applications (1950–2008). *Deep Sea Research Part I: Oceanographic*, *58*, 38–48. https://doi.org/10.1016/j.dsr.2010.11.002

Dirmeyer, P. A., & Kinter, J. L. (2010). Floods over the U.S. Midwest: A regional water cycle perspective. *Journal of Hydrometeorology*, *11*, 1172–1181. https://doi.org/10.1175/2010jhm1196.1

Doblas-Reyes, F. J., Garcia-Serrano, J., Lienert, F., Biescas, A. P., & Rodrigues, L. R. L. (2013). Seasonal climate predictability and forecasting: Status and prospects. *WIREs Climate Change*, *4*, 245–268. https://doi.org/10.1038/ncomms2704

Durack, P. J. (2015). Ocean salinity and the global water cycle. *Oceanography*, *28*, 20–31. https://doi.org/10.5670/oceanog.2015.03

Durack, P. J., Wijffels, S. E., & Matear, R. J. (2012). Ocean salinities reveal strong global water cycle intensification during 1950 to 2000. *Science*, *336*, 455–458. https://doi.org/10.1126/science.1212222

Englehart, P. J., & Douglas, A. V. (2003). Assessing warm season drought episodes in the central United States. *Journal of Climate*, *16*, 1831–1842. https://doi.org/10.1175/1520-0442(2003)016<1831:awsdei>2.0.co;2

Findell, K. L., Keys, P. W., vander Ent, R. J., Lintner, B. R., Berg, A., & Krasting, J. P. (2019). Rising temperatures increase importance of oceanic evaporation as a source for continental precipitation. *Journal of Climate*, *32*, 7713–7726. https://doi.org/10.1175/jcli-d-19-0145.1

Gelfand, A. E. (2000). Gibbs sampling. *Journal of the American Statistical Association*, *95*, 1300–1304. https://doi.org/10.1080/01621459.2000.10474335

Gill, A. E. (1980). Some simple solutions for the heat induced tropical circulation. *Quarterly Journal of the Royal Meteorological Society*, *106*, 447–462. https://doi.org/10.1002/qj.49710644905

Gimeno, L., Drumond, A., Nieto, R., Trigo, R. M., & Stohl, A. (2010). On the origin of continental precipitation. *Geophysical Research Letters*, *37*, 1–7. https://doi.org/10.1029/2010gl043712

Gimeno, L., Nieto, R., & Sori, R. (2020). The growing importance of oceanic moisture sources for continental precipitation. *npj Climate and Atmospheric Science*, *3*, 27. https://doi.org/10.1038/s41612-020-00133-y

Good, S. A., Martin, M. J., & Rayner, N. A. (2013). EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research*, *118*, 6704–6716.

Gordon, A. L. (2016). The marine hydrological cycle: The ocean's floods and droughts. *Geophysical Research Letters*, *43*, 7649–7652.

Guimbard, S., Reul, N., Chapron, B., Umbert, M., & Maes, C. (2017). Seasonal and interannual variability of the Eastern Tropical Pacific Fresh Pool. *Journal of Geophysical Research*, *122*, 1749–1771.

Gumley, L. E., & King, M. D. (1995). Remote sensing of flooding in the U.S. Upper Midwest during the summer of 1993. *Bulletin of the American Meteorological Society*, *76*, 933–944.

Higgins, R. W., Kim, H., & Unger, D. (2004). Long-lead seasonal temperature and precipitation prediction using tropical Pacific SST consolidation forecasts. *Journal of Climate*, *17*, 3398–3414.

Higgins, R. W., Shi, W., Yarosh, E., & Joyce, R. (2000). Improved United States precipitation quality control system and analysis. *NCEP/Climate Prediction Center ATLAS Rep, 7*, 40.

Horel, J. D., & Wallace, J. M. (1981). Planetary-scale atmospheric phenomena associated with the Southern Oscill*ation. Monthly Weather Review*, *109*, 813–829.

Hoskins, B. J., McIntyre, M. E., & Robertson, A. W. (1985). On the use and significance of isentropic potential vorticity maps. *Quarterly Journal of the Royal Meteorological Society*, *111*, 877–946.

Jana, S., Rajagopalan, B., Alexander, M. A., & Ray, A. J. (2018). Understanding the dominant sources and tracks of moisture for summer rainfall in the southwest United States. *Journal of Geophysical Research: Atmospheres*, *123*, 4850–4870.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP-NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, *77*, 437–471.

Kunkel, K. E., Changnon, S. A., & Angel, J. R. (1994). Climate aspects of the 1993 upper Mississippi River basin flood. *Bulletin of the American Meteorological Society*, *75*, 811–822.

Levang, S. J., & Schmitt, R. W. (2015). Centennial changes of the global water cycle in CMIP5 models. *Journal of Climate*, *28*, 6489–6502.

Li, L., & Li, W. (2013). Southeastern United States summer rainfall framework and its implication for seasonal prediction. *Environmental Research Letters*, *8*, 044017.

Li, L., Schmitt, R. W., & Ummenhofer, C. C. (2018). The role of the subtropical North Atlantic water cycle in recent US extreme precipitation events. *Climate Dynamics*, *50*, 1291–1305.

Li, L., Schmitt, R. W., Ummenhofer, C. C., & Karnauskas, K. B. (2016a). North Atlantic salinity as a predictor of Sahel rainfall. *Science Advances*, *2*, e1501588.

Li, L., Schmitt, R. W., Ummenhofer, C. C., & Karnauskas, K. B. (2016b). Implications of North Atlantic sea surface salinity for summer precipitation over the US Midwest: Mechanisms and predictive values. *Journal of Climate*, *29*, 3143–3159.

Li, Z., & O'Gorman, P. A. (2020). Response of vertical velocities in extratropical precipitation extremes to climate change. *Journal of Climate*, *33*, 7125–7139.

Liu, T., Schmitt, R. W., & Li, L. (2018). Global search for autumn-lead sea surface salinity predictors of winter precipitation in southwestern United States. *Geophysical Research Letters*, *45*, 8445–8454.

Liu, Y., Zhang, C., Tang, Q., Hosseini-Moghari, S.-M., Haile, G. G., Li, L., et al. (2021). Moisture sources variation for summer rainfall in different intensity classes over Huaihe River Valley. *Climate Dynamics*, *57*, 1121–1133.

Liu, Z., & Alexander, M. (2007). Atmospheric bridge, oceanic tunnel, and global climatic teleconnections. *Reviews of Geophysics*, *45*, RG2005.

Livezey, R. E., & Chen, W. Y. (1983). Statistical field significance and its determination by Monte Carlo techniques. *Monthly Weather Review*, *111*, 46–59.

Matsuno, T. (1966). Quasi-geostrophic motions in the equatorial area. *Journal of the Meteorological Society of Japan*, *44*, 25–43.

McCabe, G. J., Palecki, M. A., & Betancourt, J. L. (2004). Pacific and Atlantic Ocean influence on multidecadal drought frequency in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 4136–4141.

McPhaden, M. J., Lee, T., & McClurg, D. (2011). El Niño and its relationship to changing background conditions in the tropical Pacific Ocean. *Geophysical Research Letters*, *38*, L15709.

Park, M., & Lee, S. (2019). Relationship between tropical and extratropical diabatic heating and their impact on stationary-transient wave interference. *Journal of the Atmospheric Sciences*, *76*, 2617–2633.

Rathore, S., Bindoff, N. L., Ummenhofer, C. C., Phillips, H. E., Feng, M., & Mishra, M. (2021). Improving Australian rainfall prediction using sea surface salinity. *Journal of Climate*, *34*, 2473–2490.

Ray, D. K., Gerber, J. S., MacDonald, G. K., & West, P. C. (2015). Climate variation explains a third of global crop yield variability. *Nature Communications*, *6*, 5989.

Scaife, A. A., Comer, R. E., Dunstone, N. J., Knight, J. R., Smith, D. M., MacLachlan, C., et al. (2017). Tropical rainfall, Rossby waves and regional winter climate predictions. *Quarterly Journal of the Royal Meteorological Society*, *143*, 1–11.

Schanze, J. J., Schmitt, R. W., & Yu, L. L. (2010). The global oceanic freshwater cycle: A state-of-the-art quantification. *Journal of Marine Research*, *68*, 569–595.

Schubert, S. D., Stewart, R. E., Wang, H., Barlow, M., Berbery, E. H., Cai, W., et al. (2016). Global meteorological drought: A synthesis of current understanding with a focus on SST drivers of precipitation deficits. *Journal of Climate*, *29*, 3989–4019.

Sena-Martins, M., & Stammer, D. (2015). Pacific Ocean surface freshwater variability underneath the double ITCZ as seen by satellite sea surface salinity retrievals. *Journal of Geophysical Research: Oceans*, *120*, 5870–5885.

Skliris, N., Marsh, R., Mecking, J. V., & Zika, J. D. (2020). Changing water cycle and freshwater transports in the Atlantic Ocean in observations and CMIP5 models. *Climate Dynamics*, *54*, 4971–4989.

van der Ent, R. J., Savenije, H. H. G., Schaefli, B., & Steele-Dunne, S. (2010). Origin and fate of atmospheric moisture over continents. *Water Resources Research*, *46*, W09525. https://doi.org/10.1029/2010wr009127

Vinogradova, N., Lee, T., Boutin, J., Kyla, D., Severine, F., Roberto, S., et al. (2019). Satellite salinity observing system: Recent discoveries and the way forward. *Frontiers in Marine Science*, *22*. https://doi.org/10.3389/fmars.2019.00243

Weaver, S. J., & Nigam, S. (2008). Variability of the great plains low-level jet: Large-scale circulation context and hydroclimate impacts. *Journal of Climate*, *21*, 1532–1551.

White, C. J., Carlsen, H., Robertson, A. W., Klein, R. J. T., Lazo, J. K., Kumar, A., et al. (2017). Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteorological Applications*, *24*, 315–325.

Yeh, S.-W., Cai, W., Min, S.-K., McPhaden, M. J., Dommenget, D., Dewitte, B., et al. (2019). ENSO atmospheric teleconnections and their response to greenhouse gas forcing. *Reviews of Geophysics*, *56*, 185–206.

Yu, L., Josey, S., Bingham, F. M., & Lee, T. (2020). Intensification of the global water cycle and evidence from ocean salinity: A synthesis review. *Annals of the New York Academy of Sciences*, *1472*, 76–94.

Zeng, L., Schmitt, R. W., Li, L., Wang, Q., & Wang, D. (2019). Forecast of summer precipitation in the Yangtze River Valley based on South China Sea springtime sea surface salinity. *Climate Dynamics*, *53*, 5495–5509.

Zhang, W., & Villarini, G. (2019). On the weather types that shape the precipitation patterns across the U.S. Midwest. *Climate Dynamics*, *53*, 4217–4232.

## References From the Supporting Information

Aligo, E. A., Gallus, W. A., & Segal, M. (2007). Summer rainfall forecast spread in an ensemble initialized with different soil moisture analyses. *Weather and Forecasting*, *22*, 299–314. https://doi.org/10.1175/waf995.1

Delworth, T., & Manabe, S. (1989). The influence of soil wetness on near-surface atmospheric variability. *Journal of Climate*, *2*, 1447–1462. https://doi.org/10.1175/1520-0442(1989)002<1447:tioswo>2.0.co;2

Ek, M. B., & Holtslag, A. A. M. (2004). Influence of soil moisture on boundary layer cloud development. *Journal of Hydrometeorology*, *5*, 86–99. https://doi.org/10.1175/1525-7541(2004)005<0086:iosmob>2.0.co;2

Fast, J. D., & McCorcle, M. D. (1990). A two-dimensional numerical sensitivity study of the Great Plains low-level jet. *Monthly Weather Review*, *118*, 151–164. https://doi.org/10.1175/1520-0493(1990)118<0151:atdnss>2.0.co;2

Fast, J. D., & McCorcle, M. D. (1991). The effect of heterogeneous soil moisture on a summer baroclinic circulation in the central United States. *Monthly Weather Review*, *119*, 2140–2167. https://doi.org/10.1175/1520-0493(1991)119<2140:teohsm>2.0.co;2

Lin, T. I., Lee, J. C., & Yen, S. Y. (2007). Finite mixture modeling using the skew normal distribution. *Statistica Sinica*, *17*, 909–927.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.

Meehl, G. A., & Washington, W. M. (1988). A comparison of soil moisture sensitivity in two global climate models. *Journal of the Atmospheric Sciences*, *45*, 1476–1492.

Melnykov, V., & Maitra, R. (2010). Finite mixture models and model-based clustering. *Statistics Surveys*, *4*, 80–116.

Richardson, S., & Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society*, *59*, 731–792.