

# Geophysical Research Letters<sup>®</sup>



## RESEARCH LETTER

10.1029/2024GL110925

### Special Collection:

Analyzing Big Data for Understanding Climate Variability, Natural Phenomena and Rapid Environmental Changes

### Key Points:

- Statistical tools trained with large ensemble simulations outperform those trained with observations in predicting observed precipitation
- Machine learning tools exhibit slightly higher skill than multiple linear regression
- ENSO makes larger contributions to the prediction than the MJO, with the MJO's influence diminishing as forecast lead time increases

### Supporting Information:

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

### Correspondence to:

C. Zheng,  
cheng.zheng.1@stonybrook.edu

### Citation:

Zheng, C., Kim, H., LaJoie, E., He, S., & Chang, E. K. M. (2025). Improving statistical prediction of subseasonal CONUS precipitation based on ENSO and the MJO by training with large ensemble climate simulations. *Geophysical Research Letters*, 52, e2024GL110925. <https://doi.org/10.1029/2024GL110925>

Received 22 JUN 2024

Accepted 8 NOV 2024

## Improving Statistical Prediction of Subseasonal CONUS Precipitation Based on ENSO and the MJO by Training With Large Ensemble Climate Simulations

C. Zheng<sup>1</sup> , H. Kim<sup>1,2</sup> , E. LaJoie<sup>3</sup>, S. He<sup>1</sup> , and E. K. M. Chang<sup>1</sup> 

<sup>1</sup>School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY, USA, <sup>2</sup>Department of Science Education, Ewha Womans University, Seoul, Republic of Korea, <sup>3</sup>Climate Prediction Center, NOAA/NWS/NCEP, College Park, MD, USA

**Abstract** Previous studies have highlighted the significant impacts of El Niño–Southern Oscillation (ENSO) and the Madden–Julian Oscillation (MJO) on wintertime precipitation over the contiguous United States (CONUS). Here, we demonstrate skillful statistical prediction of subseasonal precipitation over the CONUS using the information of ENSO and the MJO. Simple statistical tools, such as multiple linear regression, exhibit significant improvement in prediction when trained with large ensemble climate simulations, surpassing those trained solely on observational data. Despite the biases in ENSO and MJO teleconnections in the climate simulations, the abundance of data, exceeding observational records by 100 times, allows more robust statistical relationships to be established, leading to such improvement. The utilization of machine learning tools yields additional gains in prediction skill beyond multiple linear regression. ENSO emerges as a dominant contributor to prediction skill, surpassing the influence of the MJO, whose impact diminishes with increasing forecast lead time.

**Plain Language Summary** Forecasting weather patterns on subseasonal timescales, ranging from 2 weeks to 2 months, has long been challenging due to the inherent unpredictability of the atmosphere. However, accurate predictions during these timescales are crucial for managing water resources and preparing for disasters. It has been well recognized that variability in the tropics can influence precipitation in the extratropics on subseasonal timescales, meaning we can predict extratropical precipitation by considering tropical conditions. Here, we demonstrate that by using extensive climate model simulations for training, far exceeding the amount of available observational data, statistical tools can better predict observed precipitation, compared to those only trained with observational data. The abundance of these simulations enables us to establish more reliable relationships between tropical conditions and extratropical precipitation, which is unlikely to be reached with observational records alone. Machine learning techniques can further improve the prediction accuracy compared to simple regression tools.

## 1. Introduction

Skillful subseasonal-to-seasonal (S2S) prediction of precipitation is valuable for disaster preparedness, water management, agriculture, and renewable energy (White et al., 2017, 2022). However, the subseasonal timescale has long been considered as a “predictability desert” (Vitart et al., 2012). Numerical weather prediction models perform reasonably well when the forecast lead time is within 10 days, as they rely on initial atmospheric conditions. Forecast skill beyond 1 month mostly comes from slowly varying boundary conditions, such as El Niño–Southern Oscillation (ENSO). However, subseasonal prediction covering 2–8 weeks, poses significant challenges due to the rapid growth of initial errors and the insufficient time for the effects of boundary conditions to reach large signal-to-noise ratio, leading to difficulties in producing skillful forecasts.

Recent studies have identified several sources of predictability on subseasonal timescales. These include conditions in the tropics, such as ENSO and the Madden-Julian Oscillation (MJO). The MJO (Madden & Julian, 1971, 1972) is the dominant mode of tropical subseasonal variability, characterized by eastward propagating convection. It can be skillfully predicted out to a few weeks (e.g., Jiang et al., 2020; Kim et al., 2018; Vitart, 2017). Both ENSO and the MJO modulate tropical convective activity, inducing the Pacific–North American (PNA) like patterns in the extratropics (e.g., Jin & Hoskins, 1995; Mori & Watanabe, 2008; Riddle et al., 2013) which significantly influence precipitation over the contiguous United States (CONUS) on

© 2025 The Author(s). This article has been contributed to by U.S. Government employees and their work is in the public domain in the USA.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

subseasonal time scales (e.g., Ropelewski & Halpert, 1986; Zheng et al., 2018). Other sources of predictability include stratosphere-troposphere interactions (Domeisen et al., 2020), the quasi-biennial oscillation (QBO; Wang et al., 2018), and land surface conditions (e.g., Koster et al., 2011).

The Climate Prediction Center (CPC) produces official Weeks 3 and 4 temperature and precipitation outlooks derived from various sources, including dynamical models, forecaster experience, and statistical tools. Statistical tools usually employ predictors discussed above, such as ENSO and the MJO, and make predictions based on the observed relationships between these predictors and the predictands, temperature and precipitation. One example of statistical tools predicting subseasonal temperature over North America is developed by Johnson et al. (2014), who employed the multi-decadal trend and observed composites under different ENSO and MJO phases. Alternatively, statistical forecasts can be generated using linear regression, based on the observed connections between tropical conditions (ENSO and the MJO) and extratropical temperature/precipitation. In this study, we will first construct a multiple linear regression (MLR) forecast tool based on the observed linear relationship between tropical conditions and precipitation over the CONUS, serving as a baseline tool. Then, we will develop new statistical tools and evaluate them against this baseline tool to verify the improvement in skill.

Machine learning tools have gained increasing popularity for S2S prediction (Cohen et al., 2019). These tools are frequently trained using extensive climate model simulations (e.g., Arcodia et al., 2024; Mayer & Barnes, 2022). In the context of real-world S2S prediction, previous studies have shown that statistical tools trained with a large amount of climate model data can provide skillful forecasts of a large scale temperature pattern for weeks 3 and 4 (Buchmann & DelSole, 2021) over the CONUS, large scale circulation patterns over the North Pacific (Mayer et al., 2024), as well as large scale precipitation patterns in the seasonal timescale (Gibson et al., 2021) and the subseasonal time scale (Arcodia et al., 2023) over western North America. However, given model biases, it is not clear that statistical tools trained solely using climate model data can provide skillful S2S predictions of grid-point scale precipitation, and be useful for real-time forecasting, since the spatial distribution of precipitation is noisy. In this study, we will test the hypothesis that even biased climate model data can be used to train statistical models to provide skillful S2S forecasts of grid-point precipitation.

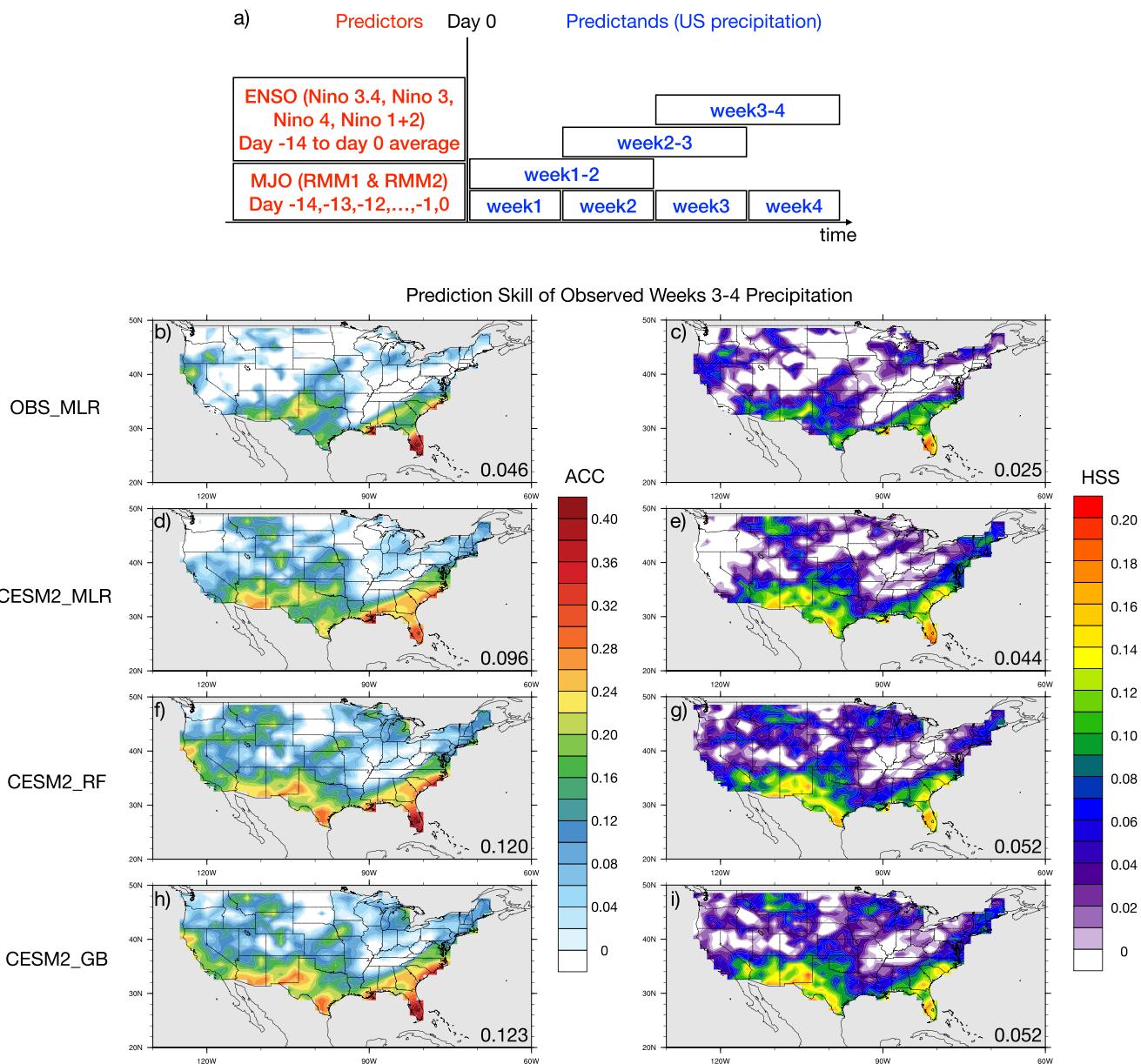
## 2. Data and Methods

### 2.1. Predictors and Predictands

Figure 1a provides a summary of the predictors representing tropical conditions and the predictands targeting CONUS precipitation. Tropical conditions encompass four ENSO indices (Niño 1 + 2, Niño 3, Niño 4, Niño 3.4) averaged from 14 days before to the reference day (day 0), along with daily real-time multivariate MJO (RMM) index (Wheeler & Hendon, 2004) from 14 days before to the reference day. Since past evolution of the MJO before the reference day could influence the MJO teleconnection after the reference day (Zheng & Chang, 2019), the RMM index before the reference day is also included as predictors. See Text S1 in Supporting Information S1 for more details about the selection of predictors. Winter (December–February; DJF) weekly and bi-weekly precipitation at each grid point out to week 4 will be predicted by using these tropical conditions.

To represent the observed precipitation, we utilize the CPC Unified Gauge-Based Analysis of Daily Precipitation over CONUS (M. Chen, Shi, et al., 2008; M. Chen, Xie, & CPC Precipitation Working Group, 2008; 0.25° resolution). Normalized ENSO indices are derived from the NOAA Daily Optimum Interpolation Sea Surface Temperature Version 2.1 (Huang et al., 2021; 0.25° resolution). The observed RMM index is obtained by applying multivariate empirical orthogonal functions (EOF) analysis on outgoing longwave radiation (OLR) and zonal wind at 200 and 850-hPa in the tropics, following the procedures in Gottschalck et al. (2010). The data sets used for calculating the RMM index are NOAA Interpolated OLR (Liebmann & Smith, 1996) and zonal wind from the European Center for Medium-Range Weather Forecasts reanalysis 5 (ERA5; Hersbach et al., 2020). The prediction focuses on winters from 1982/83 to 2021/22.

To explore how much additional skill can be obtained by training statistical tools with a large amount of climate model simulations, we utilize simulations from the Community Earth System Model version 2 large ensemble (CESM2-LE; Rodgers et al., 2021). CESM2-LE comprises 100 ensemble members representing different realizations of the climate system with the same model physics and nearly identical external forcings. We utilize CESM2-LE simulations spanning 1950–2022, which include both historical simulations (ending in 2014) and future projections with Shared Socioeconomic Pathways (SSP) 3–7.0 forcings (beginning in 2015). Winter



**Figure 1.** (a) Schematic figure summarizing the predictors and predictands used in statistical prediction. (b and c) Prediction skill (ACC and HSS) of OBS\_MLR for observed weeks 3 and 4 precipitation. (d and e), (f and g), and (h and i) Similar to (b and c), but for CESM2\_MLR, CESM2\_RF, and CESM2\_GB, respectively. The numbers at the bottom right corners of panels (b–i) represent the area-weighted domain average score.

precipitation from the model output serves as the predictands in the statistical tools. Normalized ENSO indices are derived from the model SST. The RMM index of each CESM2-LE ensemble member is constructed by projecting model anomalies onto the EOF eigenvectors from observation, following the procedures of Henderson et al. (2017). These procedures are similar to those employed for deriving RMM indices from S2S prediction models (Gottschalck et al., 2010; see Text S2 in Supporting Information S1).

Observed precipitation data is regressed to the CESM2 horizontal resolution (approximately 1°) with an area-conserved method. Predictands used in the statistical tools (both observation and CESM2) are weekly or biweekly anomalies derived from detrended daily anomalies. Daily anomalies are calculated by subtracting the first four harmonics of the seasonal cycle, and then detrended at each calendar day by removing the trend of a 91-day running mean anomaly centered on the calendar day.

## 2.2. Statistical Tools

One broadly applied and straightforward statistical approach is multiple-linear regression, which predicts precipitation based on the linear relationship between predictors and precipitation. MLR models can be constructed using observed tropical conditions and CONUS precipitation, employing a leave-one-season-out cross-validation (LOOCV) procedure to assess performance. We refer to this tool as OBS\_MLR. In addition to MLR, we apply two decision tree-based machine learning methods, namely random forests (RF; Breiman, 2001; Text S3 in Supporting Information S1) and gradient boosting (GB; T. Chen & Guestrin, 2016; Text S4 in Supporting Information S1). When utilizing these statistical tools on CESM2-LE data, we randomly select 80 ensemble members from the total 100 members for training, reserving the remaining 20 members for evaluation (testing). These tools trained with CESM2-LE data will be referred to as CESM2\_MLR, CESM2\_RF and CESM2\_GB, respectively. All statistical tools are trained and optimized at individual grid points for each lead time.

Following training with CESM2-LE data, the statistical tools can be used to predict observed precipitation. We “freeze” the statistical tools after training with CESM2-LE data and utilize observed tropical conditions as predictors to make predictions of observed precipitation. Such predictions are based solely on the relationship between tropical conditions and CONUS precipitation “learned” from CESM2-LE.

## 2.3. Skill Evaluation

We use two metrics to evaluate the prediction skill of the statistical tools: the anomaly correlation coefficient (ACC) and the Heidke skill score (HSS; Wilks, 2011), at each grid point. The ACC quantifies the association of anomalies between the forecast and the observation. The HSS, a common performance metric used by the CPC to evaluate extended-range probabilistic forecasts, assesses the proportion of categories forecasted correctly relative to a climatology forecast, or reference climatology. Each forecast and observation are assigned to one of the three forecast categories (top, middle, or bottom tercile) based on the corresponding climatological distribution. The number of categories forecasted correctly is denoted as  $H$ . For three-category forecasts, the expected number of categories forecasted correctly by a random forecast, denoted as  $E$ , is one-third of the total number of forecasts, denoted as  $T$ . Then the HSS can be written as,

$$\text{HSS} = \frac{H - E}{T - E}$$

For three-category HSS, values range from  $-0.5$  (a completely wrong set of forecasts) to  $1$  (perfect forecasts). An HSS value of zero implies that the skill is equivalent to a random draw from three equiprobable forecast categories, or a climatology forecast. Hence, forecasts are deemed skillful when HSS values exceed zero.

## 2.4. MJO Teleconnection

To quantify the MJO teleconnection, we employ an index developed by Zheng et al. (2018). MJO composites are constructed for the 8 RMM phases and lag days 0–27 (4 weeks) at each grid point, specifically when the RMM index amplitude exceeds  $1.0$  at lag 0. The standard deviation of these composites across all phases and lag days at each grid point is considered as the variability associated with the MJO, representing the MJO teleconnection. To better capture the MJO influence on precipitation which varies in magnitude across different regions, we use standard deviation instead of variance, which was employed by Zheng et al. (2018). This method can be considered as a simplified version of the approach of Jenney et al. (2019).

## 3. Results

### 3.1. Prediction Skill and the Dependence on the Amount of Training and Testing Data

To illustrate the performance of the different statistical tools on subseasonal times scales, the spatial patterns of prediction skill for observed weeks 3 and 4 precipitation are shown in Figures 1b–1i. For both skill metrics (ACC and HSS), CESM2\_MLR performs better than OBS\_MLR in the spatial patterns and domain averaged score (bottom right corners of each panel). The two machine learning tools, CESM2\_RF and CESM2\_GB, perform similarly, and are better than CESM2\_MLR. OBS\_MLR shows skill in predicting precipitation in the southern and southeastern parts of the US, while the statistical tools trained with CESM2-LE data improve the skill over the

southern US and make skillful predictions over regions in the northern and central US, as well as regions along the east coast.

Using the same MLR algorithm, why does CESM2\_MLR outperform OBS\_MLR? CESM2\_MLR benefits from a significantly larger training data set, comprising 5,760 winters, approximately 150 times more than OBS\_MLR, which relies on only 39 winters for training. To investigate to what extent the increased training data enhances prediction skill, we vary the amount of training data used in CESM2\_MLR. Different subsets of training data, ranging from 1 to 30 ensemble members, are randomly selected from the pool of the original 80 training members, repeated 50 times for each size of training data. Figure 2a illustrates how forecast skill varies with the amount of training data, by displaying the average skill (domain averaged ACC for weeks 3 and 4) across these random samples. The spread of the random samples (shadings) depicts the uncertainty range. As shown, the skill in predicting CESM2-LE testing data (20 members; in blue) and observed precipitation (in red) notably improves with increasing training data. The skill nearly matches that achieved with 80 training members (dashed lines) when the number of training members approaches 30. Thus, the low prediction skill of OBS\_MLR (black cross) is expected as its training data (39 winters) is only about half of an “ensemble member” (72 winters). The gray box plot shows the skill distribution when applying the LOOCV approach within each ensemble member of CESM2-LE (use CESM2-LE to predict CESM2-LE) during the same period (1982/83–2021/22) as in observation. Such an LOOCV approach for CESM2-LE can be considered as an analogue of OBS\_MLR. The skill of OBS\_MLR falls well within the distribution of the LOOCV approach for CESM2-LE, further confirming that the prediction skill is low when the amount of “observed” training data is limited.

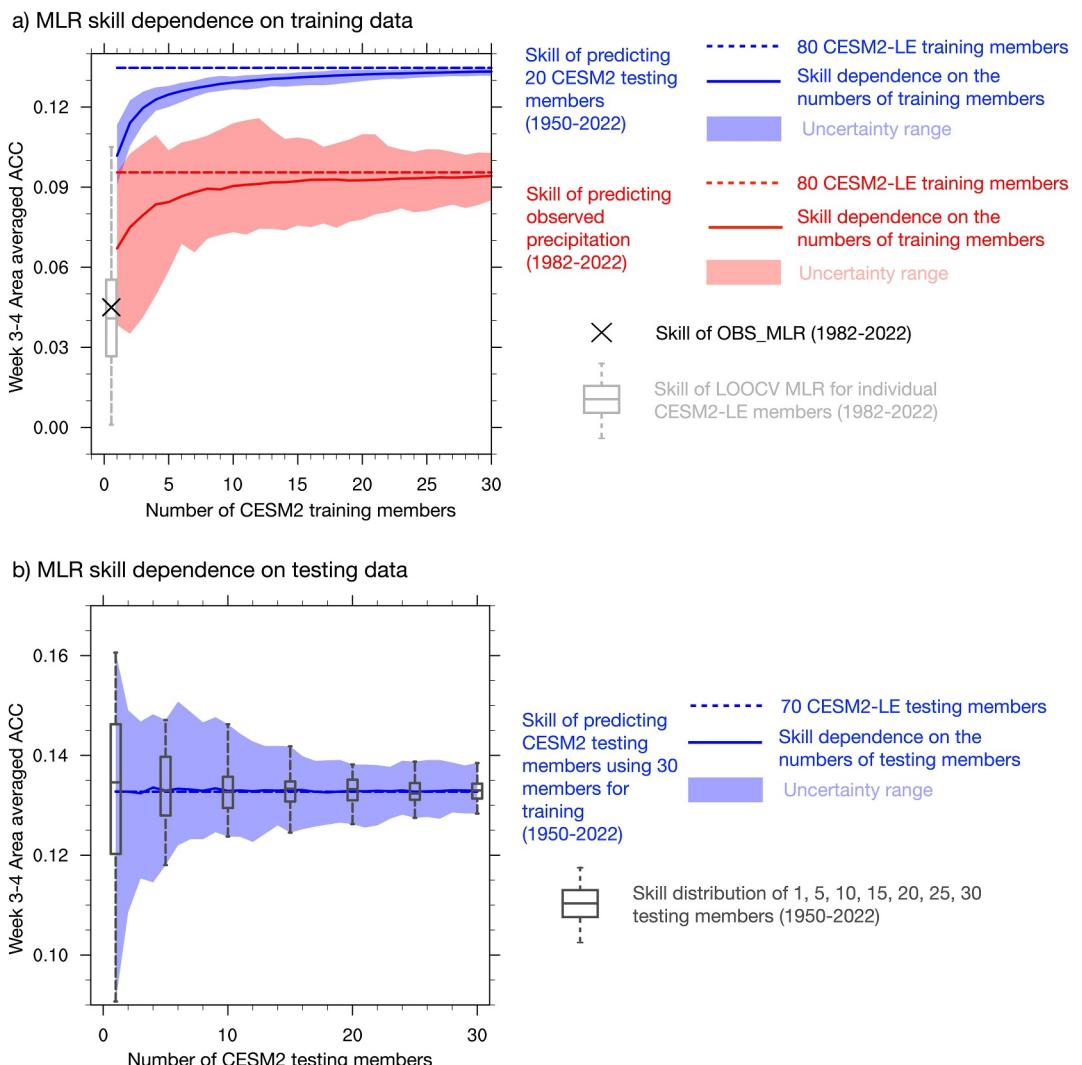
The results in Figure 2a demonstrate that a large amount of training data (about 30 members or 2000 seasons; Text S5 in Supporting Information S1) is required to establish a robust linear statistical relationship between tropical conditions and CONUS precipitation for subseasonal prediction. Machine learning tools capturing nonlinear relationships probably require even more data. Despite the potential biases in CESM2-LE (see Section 3.2), the statistical relationship derived from this large data set still significantly outperforms the relationship derived from limited observational data in predicting observed precipitation.

The uncertainty range in Figure 2a is noticeably larger in predicting observed precipitation (red shadings) compared to predicting CESM2-LE testing members, which can be attributed to the size of the validation data set. We illustrate how the uncertainty range of skill depends on the amount of testing data in Figure 2b. To expand the pool of testing members from which random subsamples can be drawn, we utilize 30 randomly selected training members, as demonstrated to be sufficient (Figure 2a). Subsequently, similar to the approach described above, subsets of testing data, ranging from 1 to 30 ensemble members, are randomly selected from the remaining 70 members designated for testing, repeated 50 times for each size of training data. The uncertainty range, shown by the shadings and box plots, largely decreases from 1 member to 20 members, without much reduction beyond 20 members.

The results indicate that a substantial amount of validation data (approximately 20 members or 1500 seasons) is necessary to achieve a robust estimation of the prediction skill. The current amount of observational data falls well below this requirement, probably resulting in large uncertainties when evaluating prediction skill for observed precipitation. This poses an additional challenge for skill evaluation, as the uncertainty arises not only from statistical algorithms or training data, but from the limited availability of observational data. Nevertheless, in Figure 2a, the differences in skill between predicting 20 CESM2-LE testing members (dashed blue line) and observed precipitation (dashed red line), could not be solely attributed to uncertainty arising from limited observational data, as CESM2\_MLR shows better skill in predicting any of the 20 testing members within a 40-year period than in predicting observations (Figure S4 in Supporting Information S1). This also suggests biases in CESM2-LE, which are discussed further in Section 3.2. Note that the findings presented above do not depend on the skill evaluation metric. Results based on HSS (Figure S5 in Supporting Information S1) exhibit similar behaviors.

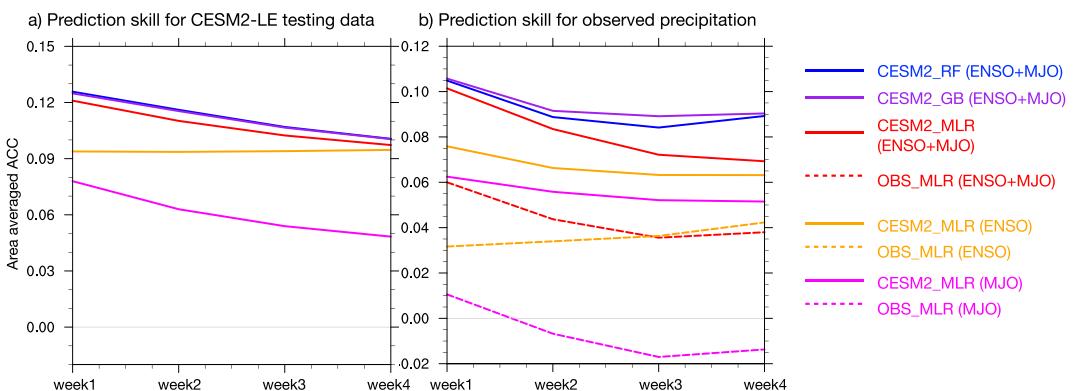
### 3.2. Contribution by ENSO and the MJO

Figure 3a illustrates the domain-averaged prediction skill of weekly precipitation in the 20 CESM2-LE testing members. Among the three statistical tools, CESM2\_RF (blue) and CESM2\_GB (purple) exhibit slightly higher performance compared to CESM2\_MLR (red). The decline in skill from week 1 to week 4 prompts an exploration into the contribution of ENSO and the MJO to the prediction skill. We retrain the prediction tools by only using



**Figure 2.** (a) Skill dependence on the amount of training data for the MLR algorithm. The y-axis shows the domain-averaged weeks 3 and 4 ACC score, while the x-axis shows the amount of training data in terms of CESM2-LE members (72 winters per member). The blue lines and shading depict the skill for predicting 20 CESM2-LE testing members, while the red lines and shadings represent the skill for predicting observed precipitation. The dashed lines illustrate the skill when using 80 training members, as in Figure 1. The solid lines depict the dependence on the amount of training data, averaged across 50 randomly selected samples for each different size of training data. The shadings represent the uncertainty range from the 50 random samples. The black cross denotes the skill of OBS\_MLR. The gray box plot shows the distribution of prediction skill when applying the LOOCV approach within each CESM2-LE member during the time period of 1982/83–2021/22, as in observation. (b) Skill dependence on the amount of testing data for MLR algorithms. The y-axis represents the domain-averaged weeks 3 and 4 ACC score, while the x-axis indicates the amount of testing data in terms of CESM2-LE members. The dashed blue line illustrates the skill when using 70 testing members, while the solid blue line illustrates the dependence on the amount of testing data, averaged across 50 randomly selected samples for each different size of testing data. The shadings represent the uncertainty range from the 50 random samples. The box plots demonstrate the distribution of the 50 random samples for 1, 5, 10, 15, 20, 25, and 30 testing members.

ENSO or the MJO as predictors within the MLR framework, represented by the orange and magenta lines, respectively. Figure 3a reveals that ENSO makes a significant contribution to the prediction skill, consistently greater than that of the MJO. Similarly, ENSO shows larger contribution than the MJO in RF and GB tools (Text S6). Such findings are similar to that in Johnson et al. (2014) and Mayer et al. (2024). The skill is relatively stable when ENSO indices are the sole predictors (orange) across week 1–4, which is not surprising as ENSO generally does not have large variations within a few weeks. In contrast, the skill decreases over time when MJO indices are the only predictors. Therefore, the combined use of both ENSO and MJO as predictors results in higher skill in the



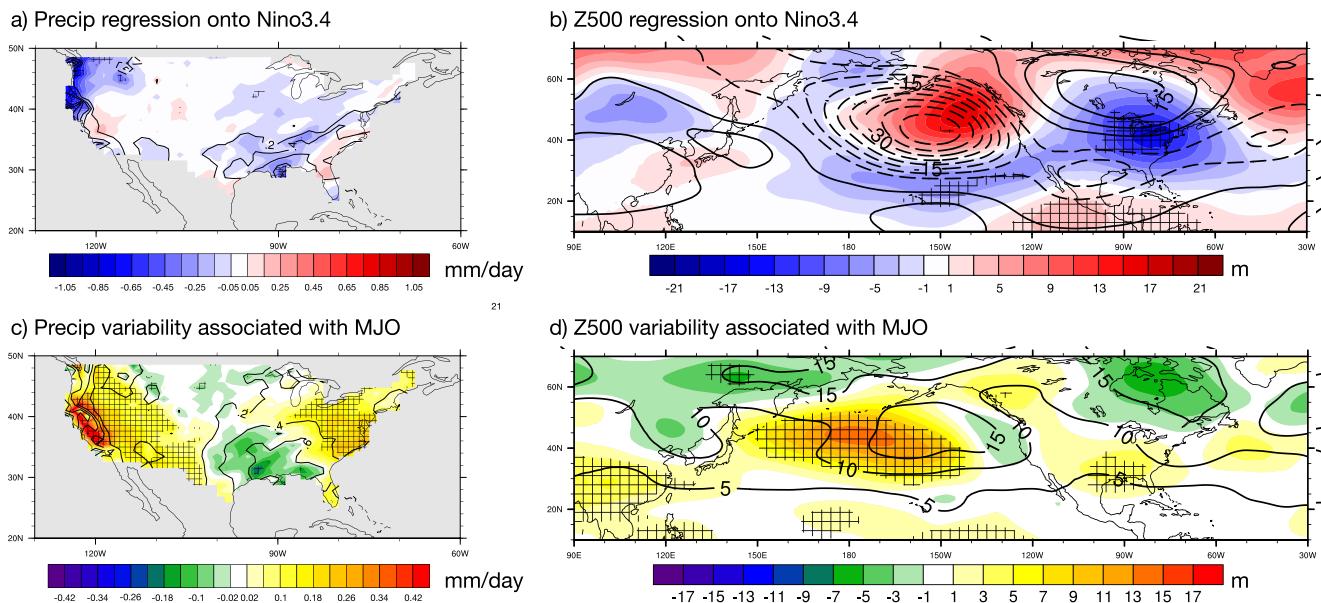
**Figure 3.** (a) Domain averaged prediction skill (ACC) for 20 CESM2-LE testing members during week 1–4. The solid blue and purple lines represent the skill of CESM2\_RF and CESM2\_GB, respectively, using both ENSO and MJO as predictors. The solid red, orange, and magenta lines depict the skill of CESM2\_MLR, using both ENSO and the MJO, only ENSO, and only the MJO as predictors, respectively. (b) Similar to (a), but for prediction skill for observed precipitation. The dashed red, orange, and magenta lines depict the skill of OBS\_MLR, using both ENSO and the MJO, only ENSO, and only the MJO as predictors, respectively.

first 2 weeks, however by week 4, the skill is only slightly better than that achieved by using only ENSO indices as predictors.

Figure 3b shows the domain-averaged prediction skill of weekly precipitation for observed precipitation. When utilizing CESM2-LE data to train the statistical tools, the findings from Figure 3b mirror those from Figure 3a with a few exceptions: (a) the skill is less stable across week 1–4 when using ENSO indices as the only predictors; (b) the skill increases from week 3 to week 4 in the two machine learning tools, CESM2\_RF and CESM2\_GB. The large uncertainty when evaluating prediction skill for observed precipitation due to a limited amount of data, as discussed in Section 3.1, could be responsible for such exceptions. Thus, Figure 3b should be interpreted qualitatively rather than quantitatively.

The skill of OBS\_MLR with different input predictors is represented by the dashed lines in Figure 3b. By comparing CESM2\_MLR and OBS\_MLR with the same input predictors (the solid lines vs. the dashed lines in the same color), it is evident that CESM2\_MLR consistently outperforms OBS\_MLR. This indicates that statistical tools trained by large ensembles of CESM2-LE data perform better than those trained by a limited amount of observational data, regardless of whether the predictors include ENSO or the MJO. This is confirmed by repeating the analysis in Figure 2a using only ENSO or the MJO as predictors (Figure S7 and Text S7 in Supporting Information S1). The two machine learning tools further improve the prediction skill over CESM2\_MLR. The contour intervals are 0.2 mm/day. Positive contours are represented by solid lines, and negative contours are represented by dashed lines. Zero contour omitted. The conclusions reached above from the ACC metric are similar to the conclusions reached from the results based on the HSS metric (Figure S8 in Supporting Information S1).

As shown above, statistical tools trained with a large amount of CESM2-LE data outperform those trained with limited observational data. But a question remains, is CESM2 correctly capturing the relationships between tropical conditions and CONUS precipitation? To address this, we examine the observed ENSO-driven precipitation anomalies, represented by the seasonal regression of CONUS precipitation onto the Niño 3.4 index (illustrated by contours in Figure 4a). To assess CESM2's representation of this relationship, regressions are computed within each ensemble member over the same period as the observations (1982/83–2021/22). Subsequently, the differences between the regression coefficients of the 100-member mean and the observations are depicted by the shadings in Figure 4a. It is important to note that these differences do not necessarily indicate model biases. Previous studies (e.g., Coats et al., 2013; Deser et al., 2017, 2018) revealed substantial variations in model-simulated ENSO teleconnections across different time periods due to internal variability. To account for internal variability in teleconnections for the 40-year period, model biases that cannot be solely attributed to internal variability are identified when observed values fall outside of the distribution that is derived from 100 CESM2-LE members. These instances are denoted by the cross-hatched regions in Figure 4a. Despite CESM2-LE



**Figure 4.** (a) The contours show the regression coefficients of seasonal (DJF) precipitation onto Niño 3.4 in observational data during winters from 1982/83 to 2021/22. Contour interval is 0.2 mm/day. Positive contours are represented by solid lines, and negative contours are represented by dashed lines. The zero contour is omitted. The shadings indicate the differences between CESM2-LE ensemble mean regression coefficients and those in observations. Ensemble mean regression coefficients are the mean regression coefficients of each individual ensemble member. The crossed regions indicate where the regression coefficients in observations fall outside of the distribution of 100 CESM2-LE ensemble members. (b) Similar to (a), but for Z500. Contour interval is 5 m. (c and d) Similar to (a and b), but for the MJO teleconnection. Instead of regression coefficients, an index representing the variability associated with the MJO teleconnection is applied. See Section 2.4 for more details of the index.

showing weaker ENSO teleconnection in precipitation in general (negative shadings over positive contours), model biases can only be confidently identified over a few grid points over the northwest US and near the Gulf of Mexico.

A similar analysis is conducted using the 500-hPa geopotential height (Z500) field to investigate how CESM2 captures the ENSO teleconnection in the large-scale circulation (Figure 4b). CESM2-LE exhibits weaker teleconnections in Z500, as the large values in the shadings generally have opposite signs compared to the contours, along with a slight shift in the teleconnection patterns. In the extratropics, model biases can only be confidently identified over eastern North America. In short, much of the inconsistency in the ENSO teleconnection between CESM2-LE and observations could not be easily distinguished from internal variability. However, in some regions model biases which cannot be explained by internal variability are identified. This suggests that CESM2-LE is at least slightly biased in simulating the ENSO teleconnection.

The approach above is also applied to the MJO teleconnections by using the index introduced in Section 2.4. CESM2-LE overestimates the precipitation variability associated with the MJO over the eastern and western parts of the US (Figure 4c), which cannot be explained solely by internal variability, and thus indicates true model biases. In Figure 4d, CESM2-LE also exhibits biases in the MJO teleconnection in Z500 over the North Pacific and parts of the southern US, noting that the overestimation in CESM2-LE is out of the range of internal variability. In summary, CESM2-LE displays significant biases in simulating the MJO teleconnections. Note the analysis above could be influenced by biases in model simulated precipitation variability (see Text S8 in Supporting Information S1). In addition, the MJO teleconnection index used here only accounts for the total variability associated with different RMM phases and lag times. Model-simulated teleconnection could also be biased when the MJO teleconnection is not captured at the correct phase or lag time (e.g., Zheng et al., 2019), possibly not reflected by this index.

Overall, CESM2-LE exhibits minor biases in simulating the ENSO teleconnections, and significant biases in capturing the MJO teleconnections. Nevertheless, statistical tools trained with a large amount of biased model data still outperforms those trained with limited observational data. This highlights the importance of providing a large training data set for statistical tools, even when training data have biases.

#### 4. Conclusions

We use MLR and two machine learning methods to make statistical predictions of CONUS precipitation based on tropical conditions (ENSO and the MJO). The major findings are as follows:

1. Increasing the amount of training data improves the prediction skill. When predicting observed precipitation, statistical tools (e.g., MLR) trained with a large amount of climate model simulation data (e.g., CESM2-LE) outperform those trained solely with observational data. Despite biases in climate model simulations, especially in the MJO-teleconnections, the abundance of climate model data allows establishment of more robust statistical relationships which improves prediction skill. Our results indicate that about 2000 seasons of training data are required to saturate the prediction skill.
2. The scarcity of observational data may lead to significant uncertainties when evaluating prediction skill. Our analysis based on CESM2-LE indicates that approximately 1500 seasons of validation data are necessary to achieve a robust estimation of the prediction skill.
3. Prediction skill is highest during week 1 and decreases with longer lead times by using both ENSO and the MJO as predictors. ENSO consistently contributes more to the prediction skill than the MJO. The contribution from ENSO remains stable from week 1 to 4, while the contribution from the MJO decreases with longer lead times. These findings are consistent with Johnson et al. (2014) and Mayer et al. (2024).
4. Machine learning tools can improve the skill compared to MLR.

Note the estimated numbers of years required are specific to CESM2-LE using the MLR tool. Points (1) and (2) likely stem from the large internal variability in tropical-extratropical teleconnections, especially ENSO teleconnections as highlighted by many previous studies (e.g., Coats et al., 2013; Deser et al., 2017, 2018). Large internal variability implies a small sample of training data does not adequately capture the teleconnection, meaning robust statistical relationship cannot be reached, which results in low prediction skill. Similarly, a small sample of evaluation data may also lead to large uncertainty. Point (2) poses additional challenges when evaluating the prediction skill for observed precipitation of statistical tools, as validation using only the current observational record (about 40 years since the satellite era) may introduce large uncertainties. Our results suggest that future improvements of climate models which reduce the biases in ENSO- and MJO-teleconnections could also be useful in improving the skill of statistical tools trained using climate model data.

Previous studies (Buchmann & DelSole, 2021; Gibson et al., 2021) have shown that training statistical tools with a large amount of climate model data can provide skillful forecasts of large-scale patterns of temperature and precipitation on S2S timescales. Here, we further demonstrate that grid-point scale precipitation, which could be noisy and sensitive to model biases, can also be skillfully predicted by statistical tools trained with large amounts of climate model data. These tools can significantly outperform those trained only with observational data.

#### Data Availability Statement

The website of NOAA PSL provides the CPC Unified Gauge-Based Analysis of Daily Precipitation over CONUS (M. Chen, Shi, et al., 2008; M. Chen, Xie, & CPC Precipitation Working Group, 2008; <https://psl.noaa.gov/data/gridded/data.unified.daily.conus.html>) and NOAA interpolated OLR (Liebmann & Smith, 1996; <https://psl.noaa.gov/data/gridded/data.olrcdr.interp.html>). ERA5 reanalysis data (Hersbach et al., 2023) is available from the Copernicus Climate Change Service Climate Data Store (<https://doi.org/10.24381/cds.bd0915c6>). CESM2-LE (Rodgers et al., 2021) data can be accessed from the NCAR website (<https://www.cesm.ucar.edu/community-projects/lens2/data-sets>). All statistical tools used for predicting precipitation, including multiple linear regression, random forests, and gradient boosting, are performed by using the scikit-learn package (Pedregosa et al., 2011) in python (<https://scikit-learn.org/stable/>).

#### Acknowledgments

We acknowledge the support from the National Oceanic and Atmospheric Administration (NOAA) Joint Technology Transfer Initiative (JTTI) Award NA22OAR4590168. H. Kim is supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (NRF-RS-2023-00278113). Computing resources, including the SeaWulf cluster, are provided by Stony Brook University.

#### References

- Arcodia, M., Barnes, E. A., & Durack, P. J. (2024). Sea surface salinity provides subseasonal predictability for forecasts of opportunity of U.S. Summertime precipitation. *ESS Open Archive*. <https://doi.org/10.22541/au.170602087.73596352/v1>
- Arcodia, M., Barnes, E. A., Mayer, K., Lee, J., Ordonez, A., & Ahn, M.-S. (2023). Assessing decadal variability of subseasonal forecasts of opportunity using explainable AI. *Environmental Research*, 2(4), 045002. <https://doi.org/10.1088/2752-5295/aced60>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/a:1010933404324>

- Buchmann, P., & DelSole, T. (2021). Week 3–4 prediction of wintertime CONUS temperature using machine learning techniques. *Frontiers in Climate*, 3, 81. <https://doi.org/10.3389/fclim.2021.697423>
- Chen, M., Xie, P., & CPC Precipitation Working Group. (2008). CPC unified gauge-based analysis of global daily precipitation. In *Western Pacific Geophysics Meeting* (p. 14). American Geophysical Union. Retrieved from [http://ftp.cpc.ncep.noaa.gov/precip/CPC\\_UNI\\_PRCP/GAUGE\\_GLB/DOCU/Chen\\_et\\_al\\_2008\\_Daily\\_Gauge\\_Anal.pdf](http://ftp.cpc.ncep.noaa.gov/precip/CPC_UNI_PRCP/GAUGE_GLB/DOCU/Chen_et_al_2008_Daily_Gauge_Anal.pdf)
- Chen, M., Shi, W., Xie, P., Silva, V. B. S., Kousky, V. E., Wayne Higgins, R., & Janowiak, J. E. (2008). Assessing objective techniques for gauge-based analyses of global daily precipitation. *Journal of Geophysical Research*, 113(D4), D04110. <https://doi.org/10.1029/2007JD009132>
- Chen, T., & Guestrin, C. (2016). *XGboost: A scalable tree boosting system*. Association for Computing Machinery. <https://doi.org/10.1145/2939672.2939785>
- Coats, S., Smerdon, J. E., Cook, B. I., & Seager, R. (2013). Stationarity of the tropical pacific teleconnection to North America in CMIP5/PMIP3 model simulations. *Geophysical Research Letters*, 40(18), 4927–4932. <https://doi.org/10.1002/grl.50938>
- Cohen, J., Coumou, D., Hwang, J., Mackey, L., Orenstein, P., Totz, S., & Tziperman, E. (2019). S2S reboot: An argument for greater inclusion of machine learning in subseasonal to seasonal forecasts. *Wiley Interdisciplinary Reviews: Climate Change*, 10(2), 1–15. <https://doi.org/10.1002/wcc.567>
- Deser, C., Simpson, I. R., McKinnon, K. A., & Phillips, A. S. (2017). The Northern Hemisphere extra-tropical atmospheric circulation response to ENSO: How well do we know it and how do we evaluate models accordingly? *Journal of Climate*, 30(13), 5059–5082. <https://doi.org/10.1175/JCLI-D-16-0844.1>
- Deser, C., Simpson, I. R., Phillips, A. S., & McKinnon, K. A. (2018). How well do we know ENSO's climate impacts over North America and how do we evaluate models accordingly? *Journal of Climate*, 31(13), 4991–5014. <https://doi.org/10.1175/JCLI-D-17-0783.1>
- Domeisen, D. I., Butler, A. H., Charlton-Perez, A. J., Ayarzagüena, B., Baldwin, M. P., Dunn-Sigouin, E., et al. (2020). The role of the stratosphere in subseasonal to seasonal prediction: 2. Predictability arising from stratosphere-troposphere coupling. *Journal of Geophysical Research: Atmospheres*, 125(2), e2019JD030923. <https://doi.org/10.1029/2019jd030923>
- Gibson, P. B., Chapman, W. E., Altinok, A., Delle Monache, L., DeFlorio, M. J., & Waliser, D. E. (2021). Training machine learning models on climate model output yields skillful interpretable seasonal precipitation forecasts. *Communications Earth & Environment*, 2(1), 1–13. <https://doi.org/10.1038/s43247-021-00225-4>
- Gottschalck, J., Wheeler, M., Weickmann, K., Vitart, F., Savage, N., Lin, H., et al. (2010). A framework for assessing operational Madden-Julian Oscillation forecasts: A CLIVAR MJO Working Group Project. *Bulletin of the American Meteorological Society*, 91(9), 1247–1258. <https://doi.org/10.1175/2010BAMS2816.1>
- Henderson, S. A., Maloney, E. D., & Son, S.-W. (2017). Madden-Julian Oscillation Pacific teleconnections: The impact of the basic state and MJO representation in general circulation models. *Journal of Climate*, 30(12), 4567–4587. <https://doi.org/10.1175/JCLI-D-16-0789.1>
- Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., et al. (2023). ERA5 hourly data on pressure levels from 1940 to present [Dataset]. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. <https://doi.org/10.24381/cds.bd0915c6>
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/qj.3803>
- Huang, B., Liu, C., Banzon, V., Freeman, E., Graham, G., Hankins, B., et al. (2021). Improvements of the daily optimum interpolation sea surface temperature (DOISST) version 2.1. *Journal of Climate*, 34(8), 2923–2939. <https://doi.org/10.1175/JCLI-D-20-0166.1>
- Jenney, A. M., Randall, D. A., & Barnes, E. A. (2019). Quantifying regional sensitivities to periodic events: Application to the MJO. *Journal of Geophysical Research: Atmospheres*, 124(7), 3671–3683. <https://doi.org/10.1029/2018JD029457>
- Jiang, X., Adames, Á. F., Kim, D., Maloney, E. D., Lin, H., Kim, H., et al. (2020). Fifty years of research on the Madden-Julian Oscillation: Recent progress, challenges, and perspectives. *Journal of Geophysical Research: Atmospheres*, 125(17), e2019JD030911. <https://doi.org/10.1029/2019JD030911>
- Jin, F.-F., & Hoskins, B. J. (1995). The direct response to tropical heating in a baroclinic atmosphere. *Journal of the Atmospheric Sciences*, 52(3), 307–319. [https://doi.org/10.1175/1520-0469\(1995\)052%3C307:TDRTH%3E2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052%3C307:TDRTH%3E2.0.CO;2)
- Johnson, N. C., Collins, D. C., Feldstein, S. B., L'Heureux, M. L., & Riddle, E. E. (2014). Skillful wintertime North American temperature forecasts out to 4 weeks based on the state of ENSO and the MJO. *Weather and Forecasting*, 29(1), 23–38. <https://doi.org/10.1175/WAF-D-13-00102.1>
- Kim, H., Vitart, F., & Waliser, D. E. (2018). Prediction of the Madden-Julian Oscillation: A review. *Journal of Climate*, 31(23), 9425–9443. <https://doi.org/10.1175/JCLI-D-18-0210.1>
- Koster, R. D., Mahanama, S. P. P., Yamada, T. J., Balsamo, G., Berg, A. A., Boisserie, M., et al. (2011). The second phase of the global land-atmosphere coupling experiment: Soil moisture contributions to subseasonal forecast skill. *Journal of Hydrometeorology*, 12(5), 805–822. <https://doi.org/10.1175/2011JHM1365.1>
- Liebmann, B., & Smith, C. A. (1996). Description of a complete (interpolated) outgoing longwave radiation dataset. *Bulletin of the American Meteorological Society*.
- Madden, R. A., & Julian, P. R. (1971). Detection of a 40–50 day oscillation in zonal wind in tropical Pacific. *Journal of the Atmospheric Sciences*, 28(5), 702–708. [https://doi.org/10.1175/1520-0469\(1971\)028<702:DOADOI>2.0.CO;2](https://doi.org/10.1175/1520-0469(1971)028<702:DOADOI>2.0.CO;2)
- Madden, R. A., & Julian, P. R. (1972). Description of global-scale circulation cells in tropics with a 40–50 day period. *Journal of the Atmospheric Sciences*, 29(6), 1109–1123. [https://doi.org/10.1175/1520-0469\(1972\)029<1109:DGSCCC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1972)029<1109:DGSCCC>2.0.CO;2)
- Mayer, K. J., & Barnes, E. A. (2022). Quantifying the effect of climate change on midlatitude subseasonal prediction skill provided by the tropics. *Geophysical Research Letters*, 49(14), e2022GL098663. <https://doi.org/10.1029/2022GL098663>
- Mayer, K. J., Chapman, W. E., & Manriquez, W. A. (2024). Exploring the relative importance of the MJO and ENSO to North Pacific subseasonal predictability. *Geophysical Research Letters*, 51(10), e2024GL108479. <https://doi.org/10.1029/2024GL108479>
- Mori, M., & Watanabe, M. (2008). The growth and triggering mechanisms of the PNA: A MJO-PNA coherence. *Journal of the Meteorological Society of Japan Series II*, 86(1), 213–236. <https://doi.org/10.2151/jmsj.86.213>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. Retrieved from <https://arxiv.org/abs/1201.0490v4>
- Riddle, E. E., Stoner, M. B., Johnson, N. C., L'Heureux, M. L., Collins, D. C., & Feldstein, S. B. (2013). The impact of the MJO on clusters of wintertime circulation anomalies over the North American region. *Climate Dynamics*, 40(7), 1749–1766. <https://doi.org/10.1007/s00382-012-1493-y>
- Rodgers, K. B., Lee, S.-S., Rosenbloom, N., Timmermann, A., Danabasoglu, G., Deser, C., et al. (2021). Ubiquity of human-induced changes in climate variability. *Earth System Dynamics Discussions*, 12(4), 1–22. <https://doi.org/10.5194/esd-12-1393-2021>
- Ropelewski, C. F., & Halpert, M. S. (1986). North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). *Monthly Weather Review*, 114(12), 2352–2362. [https://doi.org/10.1175/1520-0493\(1986\)114<2352:NAPATP>2.0.CO;2](https://doi.org/10.1175/1520-0493(1986)114<2352:NAPATP>2.0.CO;2)

- Vitart, F. (2017). Madden-Julian oscillation prediction and teleconnections in the S2S database. *Quarterly Journal of the Royal Meteorological Society*, 143(706), 2210–2220. <https://doi.org/10.1002/qj.3079>
- Vitart, F., Robertson, A. W., & Anderson, D. L. T. (2012). Subseasonal to seasonal prediction project: Bridging the gap between weather and climate. *WMO Bulletin*, 61(61).
- Wang, J., Kim, H.-M., & Chang, E. K. M. (2018). Interannual modulation of Northern Hemisphere winter storm tracks by the QBO. *Geophysical Research Letters*, 45(6), 2786–2794. <https://doi.org/10.1002/2017GL076929>
- Wheeler, M. C., & Hendon, H. H. (2004). An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Monthly Weather Review*, 132(8), 1917–1932. [https://doi.org/10.1175/1520-0493\(2004\)132<1917:AARMMI>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2)
- 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(3), 315–325. <https://doi.org/10.1002/met.1654>
- White, C. J., Domeisen, D. I. V., Acharya, N., Adefisan, E. A., Anderson, M. L., Aura, S., et al. (2022). Advances in the application and utility of subseasonal-to-seasonal predictions. *Bulletin of the American Meteorological Society*, 103(1448–1472), 1448–1472. <https://doi.org/10.1175/BAMS-D-20-0224.1>
- Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (p. 676). Elsevier. New York, NY: Academic Press.
- Zheng, C., & Chang, E. K. M. (2019). The role of MJO propagation, lifetime, and intensity on modulating the temporal evolution of the MJO extratropical response. *Journal of Geophysical Research: Atmospheres*, 124(10), 5352–5378. <https://doi.org/10.1029/2019JD030258>
- Zheng, C., Chang, E. K.-M., Kim, H., Zhang, M., & Wang, W. (2019). Subseasonal to seasonal prediction of wintertime northern hemisphere extratropical cyclone activity by S2S and NMME models. *Journal of Geophysical Research: Atmospheres*, 124(22), 12057–12077. <https://doi.org/10.1029/2019JD031252>
- Zheng, C., Chang, E. K.-M., Kim, H., Zhang, M., & Wang, W. (2018). Impacts of the Madden-Julian Oscillation on storm-track activity, surface air temperature, and precipitation over North America. *Journal of Climate*, 31(15), 6113–6134. <https://doi.org/10.1175/jcli-d-17-0534.1>

## References From the Supporting Information

- Lin, H., Brunet, G., & Derome, J. (2008). Forecast skill of the Madden–Julian Oscillation in two Canadian atmospheric models. *Monthly Weather Review*, 136(11), 4130–4149. <https://doi.org/10.1175/2008MWR2459.1>