

Geophysical Research Letters

RESEARCH LETTER

10.1029/2021GL094000

Special Section:

Atmospheric Rivers: Intersection of Weather and Climate

Key Points:

- Atmospheric river frequency can be skillfully forecast at least 9 months in advance
- The skills are significant over certain regions of western North America such as Alaska and California
- El Niño-Southern Oscillation and Interdecadal Pacific Oscillation are important predictability sources

Supporting Information:

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

Correspondence to:

K.-C. Tseng,
kaichiht@princeton.edu

















Citation:

Tseng, K.-C., Johnson, N. C., Kapnick, S. B., Delworth, T. L., Lu, F., Cooke, W., et al. (2021). Are multiseasonal forecasts of atmospheric rivers possible? *Geophysical Research Letters*, 48, e2021GL094000. <https://doi.org/10.1029/2021GL094000>

Received 26 APR 2021

Accepted 26 JUL 2021

Are Multiseasonal Forecasts of Atmospheric Rivers Possible?

Kai-Chih Tseng^{1,2} , Nathaniel C. Johnson¹ , Sarah B. Kapnick¹ , Thomas L. Delworth¹ , Feiyu Lu^{1,2} , William Cooke¹ , Andrew T. Wittenberg¹ , Anthony J. Rosati^{1,3} , Liping Zhang^{1,3} , Colleen McHugh^{1,4} , Xiaosong Yang¹ , Matthew Harrison¹ , Fanrong Zeng¹ , Gan Zhang^{1,2,5} , Hiroyuki Murakami^{1,3,6}, Mitchell Bushuk^{1,3} , and Liwei Jia^{1,3} 

¹Geophysical Fluid Dynamics Laboratory, National Oceanic and Atmospheric Administration, Princeton, NJ, USA,

²Department of Geosciences, Program in Atmospheric and Oceanic Science, Princeton University, Princeton, NJ, USA,

³University Corporation for Atmospheric Research, Boulder, CO, USA, ⁴SAIC, Science Applications International Corporation, Reston, VA, USA, ⁵Citadel Americas, LLC, Chicago, IL, USA, ⁶Meteorological Research Institute, Tsukuba, Japan

Abstract Atmospheric rivers (ARs) exert significant socioeconomic impacts in western North America, where 30% of the annual precipitation is determined by ARs that occur in less than 15% of wintertime. ARs are thus beneficial to water supply but can produce extreme precipitation hazards when making landfall. While most prevailing research has focused on the subseasonal (≤ 5 weeks) prediction of ARs, only limited efforts have been made for AR forecasts on multiseasonal timescales (≥ 3 months) that are crucial for water resource management and disaster preparedness. Through the analysis of reanalysis data and retrospective predictions from a new seasonal-to-decadal forecast system, this research shows the existing potential of multiseasonal AR frequency forecasts with predictive skills 9 months in advance. Additional analysis explores the dominant predictability sources and challenges for multiseasonal AR prediction.

Plain Language Summary Atmospheric rivers (ARs), narrow corridors of intense moisture transport and heavy precipitation, are an important water resource but also a cause of flooding-related disasters for western North America. Consequently, predictions of AR frequency several seasons in advance potentially would be of great value, but such operational forecasts are currently lacking due to the challenges in simulating such intense, small-scale weather phenomena and their predictability sources on seasonal timescales. In this study, we examine the forecast skill of AR frequency on seasonal-to-multiseasonal timescales (≥ 3 months) in a new generation seasonal-to-decadal prediction system developed at the Geophysical Fluid Dynamics Laboratory. We find that AR frequency can be skillfully forecast at least 9 months in advance over certain regions of the west coast of North America, such as California and Alaska, while the forecasts are only reliable for the first season in other regions. This regional variability can be further explained by the large-scale climate variability pattern that is responsible for much of the skill, which is strongly modulated by slowly varying sea surface temperature (SST) variations. A prototype probabilistic seasonal AR forecast product is proposed.

1. Introduction

Atmospheric rivers (ARs)—corridors of intense moisture flux with transport intensity of over 180,000 tons of moisture per second—are uniquely characterized by plume-like features with a length over 1,000 km and a width about one-third of its length (Newell et al., 1992; Ralph et al., 2018; Zhu & Newell, 1998). AR intensity and duration span a wide range and sometimes can lead to different socio-economic impacts. Previous work has categorized ARs into five categories according to certain intensity and duration criteria (Ralph et al., 2019). The flood damage caused by ARs grows exponentially from one category to the next (Corringham et al., 2019). Due to the shallow scale-height of specific humidity (~ 2.2 km), the strongest transport is confined in the bottom 3 km of the troposphere (Ralph et al., 2004). The bottom-heavy moisture transport causes AR activity to be strongly influenced by the topography over the land. Although the term AR derives from the nature of its geometry, its transport is not continuous (Dacre et al., 2015). In fact, the duration and

trajectory vary from one event to another, indicating the complexity of the dynamics that determines AR evolution (Ralph et al., 2013). The complexity of AR dynamics and the great socioeconomic influence of ARs indicate both the challenge and societal need to improve AR prediction over a range of time horizons.

In the past decade, great efforts have been made in AR prediction to move beyond the weather timescale (≤ 2 weeks) toward subseasonal prediction problems (2–5 weeks) (Baggett et al., 2017; DeFlorio et al., 2019; Lang et al., 2020; Mundhenk et al., 2018; White et al., 2017). The societal benefits of skillful subseasonal and longer timescales prediction, particularly for hydrological extremes, extend to many sectors, including resource management, agriculture, and industry (Corringham et al., 2019; Dettinger, 2013). For example, water managers could benefit from advanced lead time to prepare reservoirs for extreme precipitation conditions or aquifer recharge. Subseasonal hydrological extremes prediction has progressed rapidly, particularly with respect to theory (Lavers et al., 2016; Pasquier et al., 2019; Payne & Magnusdottir, 2014; Zheng et al., 2018), empirical prediction (Baggett et al., 2017; Mundhenk et al., 2018), and the production of archived reforecast data from operational forecast systems (Vitart & Robertson, 2018; Vitart et al., 2017). However, the seasonal-to-multiseasonal forecast of ARs has received comparatively little attention. A previous study demonstrated the potential to skillfully predict wintertime seasonal AR activity for lead times of 1–3 months in dynamical forecast models participating in the North American Multi-Model Ensemble (NMME; Zhou & Kim, 2018). However, biases in sea surface temperature (SST) and large-scale circulations lead to marginal skills at this forecast lead. In addition, the performance of extreme precipitation forecasts is resolution-dependent (O'Brien et al., 2016). The question remains whether the skill of predictions of AR statistics on seasonal and longer timescales (≥ 3 months) is hindered by intrinsic predictability limits or whether improving global climate models will lead to enhanced prediction skill to inform multiseasonal outlooks of AR-related hydrological extremes.

Here, we address this latter possibility by focusing on seasonal-to-multiseasonal AR predictions that span forecast lead times of 0–9 months. ARs are closely related to extratropical storm activity, with 85% of AR events paired with one or multiple extratropical cyclones (Z. Zhang et al., 2019). Skillful predictions of extratropical storm activity within this time frame generally rely on boundary conditions from the ocean and show less dependence on the atmospheric initial state (Bach et al., 2019; Yang et al., 2015). Indeed, predictive power on seasonal and longer timescales is mostly associated with the far-reaching teleconnections generated by interannual variability of tropical SST such as the El Niño–Southern Oscillation (ENSO; Hoerling & Ting, 1994; Hoskins & Karoly, 1981). In addition, the re-emergence of extratropical oceanic subsurface temperature from one year to the next is also a potential source of predictability on seasonal timescales (Alexander & Deser, 1995; Byju et al., 2018; Newman et al., 2016).

Seasonal-to-multiseasonal predictions have been developed for hurricanes over the last 20 years since (Gray, 1984a, 1984b; also see discussion in Camargo & Barnston, 2009). On the other hand, the prediction science of ARs targeting seasonal and longer timescales (≥ 3 months), has not been developed despite a high societal need for early warnings. This paper directly addresses this need, quantifying seasonal-to-multiseasonal prediction skill in a new prediction system and developing a framework for new operational seasonal outlooks. The hope for skillful seasonal-to-multiseasonal AR predictions has similar roots as recent progress made in seasonal tropical cyclone forecasts (Murakami et al., 2015), as the increasing realism and atmospheric resolution of global climate models in the past decade has made explicit simulation of tropical cyclones and the skillful prediction of their seasonal statistics possible. These advances indicate that global climate models may now also be able to simulate complex AR structures and their interactions with slowly varying boundary conditions and complex topography. In addition, increasing ocean subsurface observations from Argo (2020) enable climate models to leverage long-term memory from the ocean to generate seasonal forecasts and permit more thorough evaluations of their performance.

In this study, we analyze AR seasonal-to-multiseasonal prediction skill in retrospective forecasts produced by a coupled general circulation model (CGCM) developed at the Geophysical Fluid Dynamics Laboratory (GFDL), the Seamless System for Prediction and Earth System Research (SPEAR, Methods). We demonstrate that winter AR frequency over certain regions of western North America can be skillfully predicted 9 months in advance by SPEAR. We also explore the potential predictability sources and challenges for multiseasonal AR forecast. A prototype probabilistic forecast product of multiseasonal AR prediction is

proposed at the end of this study, which shows the potential to be incorporated with the existing storm-based outlook systems.

2. Methods and Model

2.1. AR Detection Algorithm

ARs are identified with a detection algorithm (Mundhenk et al., 2016), which tests whether gridded integrated vapor transport (IVT) anomaly data meet particular intensity and geometry criteria (i.e., area $\geq 300,000 \text{ km}^2$, length $\geq 1,400 \text{ km}$, and aspect ratio ≥ 1.4) that resemble the plume-like nature of ARs. The IVT is defined as

$$\text{IVT} = \frac{1}{g} \sqrt{\left(\int_{1000}^{250} q u dp \right)^2 + \left(\int_{1000}^{250} q v dp \right)^2} \quad (1)$$

where u is the zonal wind, v is the meridional wind, q is the specific humidity, and g is the gravitational acceleration. We remove the first three harmonics of the IVT seasonal cycle to calculate the anomaly data. Following Mundhenk et al. (2018), the 94th percentile of the anomalous IVT intensity over the Pacific basin is used as the minimum threshold for detecting ARs. In this study, the European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis (ERA5; Hersbach et al., 2020; Supporting Information S1 for details), covering the period from 1995 to 2018, is used as the observational reference for validation and is processed following the steps above. For the SPEAR ensemble hindcast, the IVT is processed in the same way as ERA5, where the first three harmonics of seasonal cycle is derived from the ensemble mean as a function of forecast lead. The drift of daily climatology in SPEAR is small (figure not shown), which may benefit from the new bias correction technique used in the SPEAR ocean model (see next section). Both ERA5 and SPEAR data are gridded at a spatial resolution of 50 or 100-km and at a daily temporal resolution before applying the detection algorithm. Each time step (i.e., daily IVT map) is scrutinized independently. An AR event is recorded when the criteria are satisfied or exceeded. The AR frequency is then defined as the number of AR days that are labeled for a grid cell in a given period (e.g., 3 months, unit = day/day).

2.2. Climate Model Retrospective Predictions

The retrospective forecast data used in this study is produced by GFDL's latest seasonal-to-decadal prediction system, SPEAR, which incorporates component models recently developed at GFDL, including the AM4 atmosphere model, MOM6 ocean model, LM4 land model, and SIS2 sea ice model (Delworth et al., 2020). AM4 simulates reasonable AR statistics (Zhao, 2020). SPEAR is optimized to study seasonal-to-decadal variability and prediction. The ocean model is run at $1^\circ \times 1^\circ$ horizontal resolution while the atmospheric horizontal resolution is either 50 km (SPEAR_MED) or 100 km (SPEAR_LO). The analysis of the 50-km SPEAR_MED simulations are shown in this study; since February 2021, this version of SPEAR has contributed to the NMME, a seasonal prediction system composed of coupled models from North American centers that provide real-time seasonal forecasts each month (Kirtman et al., 2014). For the inter-comparison of different model resolutions, we regrid all variables to the same common grid (100 km) before calculating AR occurrence. The results remain qualitatively unchanged between the native grid (50 km) and coarsened (100 km) SPEAR_MED. Further results based on the 100 km SPEAR_LO simulations are documented in Supporting Information S1. Simulations with 15 ensemble members are initialized each January 1, April 1, July 1, and October 1 from 1995 to 2018 and then run for 12 months. For initialization, an updated ocean data assimilation using ocean tendency adjustment (OTA; Lu et al., 2020) was developed for the MOM6 ocean model, which ingests data from NOAA's Optimum Interpolation SST (OISST), tropical buoys, and three-dimensional temperature and salinity profiles acquired from Argo (2020).

2.3. Anomaly Correlation

In this study, we use the anomaly correlation coefficient (ACC) to evaluate the model prediction skill, which is defined as follows:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2)$$

In 2, i is the time index, and n is the total number of initializations (i.e., 23, one for each year). x_i and y_i are the seasonal AR anomaly at a given grid point from SPEAR and ERA5, respectively. \bar{x} and \bar{y} are defined as the mean over the 23 years of data. The linear trend is removed when evaluating the model skills, but the results change little if the linear trend is retained. The ACC measures the ability of the forecasts to capture the observed time evolution of AR frequency for a given location or region. In addition, using ACC as a metric avoids the systematic biases in forecast amplitude and enables us to focus on the variability since the variables have been standardized. In general, an ACC from 0.4 to 0.6 is usually considered as skillful and actionable for seasonal prediction (rule of thumb).

2.4. Climatological Reference Forecast and Persistence Forecast

From 2, one can find that a random draw from the climatological distribution as predictors is expected to have an ACC of zero. Thus, if the ACC is significantly higher than 0, it also suggests the forecast given by SPEAR is significantly better than a climatological forecast. In addition to the climatological reference forecast, we also compare the skill to a persistence forecast. Specifically, the observed AR frequency in the initialized season is used as the forecast for the following 12 months. A persistence forecast can be considered as the forecast reference provided by the memory in the initial states.

3. Result

3.1. Atmospheric River Climatology and Bias

ARs impact western North America over the whole year (Mundhenk et al., 2016). However, these impacts exhibit pronounced seasonality over different regions. Figure 1 illustrates the seasonal climatology and biases of AR frequency simulated by SPEAR in forecasts initialized on January 1. In general, the model reproduces the seasonal evolution of AR frequency shown in reanalysis data. Coastal California (British Columbia) has maximum values around December–January (September–October) and minimum values around August–September (April–June, details can be found in Figure S2). Contrary to the winter maximum AR frequency along the West Coast of the United States, the Great Plains show maximum frequency around July–September, which is associated with the Great Plains low-level jet (Figure 1c). The low-level jet is most active during boreal summer and fall due to prevailing southerly winds and the diurnal oscillation of wind direction by elevation-dependent east-west pressure gradient force (Jiang et al., 2007). All of these features are reasonably captured by SPEAR (i.e., contours and shading in the left panels of Figure 1 are spatially collocated), indicating the capability of the model in reproducing the observed climatology.

However, despite the reasonable simulation of AR climatology, a few systematic biases are still evident. The AR frequency is underestimated over the regions with the most active occurrence (e.g., to the northeast of Hawaii in Figure 1e), which may result, in part, from the difference in spatial resolution between SPEAR (50 km) and ERA5 (31 km). In addition, the frequency is slightly overestimated to the north of Hawaii. This north-south dipole pattern of frequency bias suggests a potential bias in model storm track position. The biases over western North America and the Great Plains are also resolution-dependent, which are more evident in the 100-km simulations (see Text S2 in Supporting Information S1 for details). The resolution-dependent biases indicate that medium to high spatial resolutions (25–50 km) are necessary for simulating hydrological extremes, consistent with previous studies (O’Brien et al., 2016; Zhou & Kim, 2018).

In spite of the systematic biases, SPEAR shows a promising ability to simulate the AR climatology, especially over the regions of interest. This motivates us to further explore the potential for seasonal to multiseasonal AR prediction.

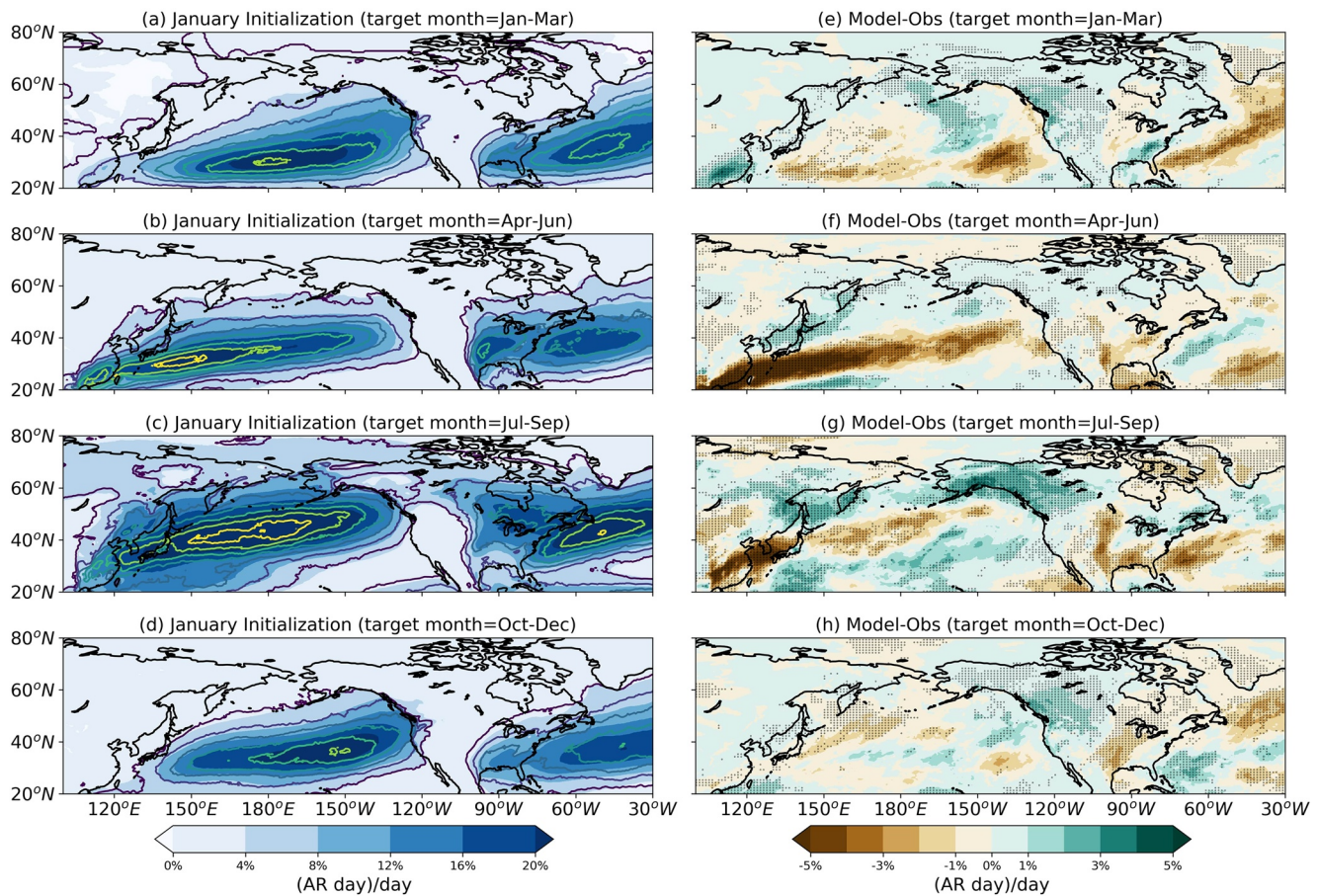


Figure 1. (a)–(d) The monthly climatology of Atmospheric river (AR) frequency for (a) January–March, (b) April–June, (c) July–September, and (d) October–December in Seamless System for Prediction and Earth System Research (SPEAR) forecasts initialized on 1 January. The shading shows the SPEAR output and contours show the European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis (ERA5) reanalysis climatology with an interval of 4% starting from 0. (e)–(h) The model biases (i.e., SPEAR minus ERA5 climatology) corresponding to each panel on the left side. Stippling indicates that all ensemble members agree on the sign of the bias and the values are significantly different from 0 at the 5% level based on a two-tailed t -test.

3.2. AR Prediction Skill in Western North America

Given that Figure 1 reveals the model capability in simulating the large-scale patterns of AR climatology, we next focus on the ability to predict local AR activity, particularly over western North America from January to March where ARs are most active and have significant socio-economic impacts. Figures 2a–2d show maps of the ACC for forecasts targeting the January–March season and initialized the previous January, October, July, and April. Figure 2 reveals that high prediction skill is related to a two-stripe feature (marked with arrows), which extends from the subtropical Pacific to the west coast of North America. California is located near the eastern ends of the two stripes, while Alaska is slightly to the north of high prediction skill regions. The British Columbia and Washington/Oregon regions lie between these two stripes. Figures 2e–2h show the ACC as a function of initialization and forecast lead time over four locations along the West Coast of North America ordered from north to south. The selection of locations is based on Mundhenk et al. (2018). The four panels reveal significant regionally averaged AR prediction skill (≥ 0.5 , triangle markers), even out to lead times of several seasons in some cases. Alaska and California in particular have the highest skill after the first two seasons, with significant skill for winter forecasts extending to leads of at least 9 months. Specifically, the January–March forecast made in April is significantly higher than the climatological prediction skills (Figures 2e–2h, white dots). The high prediction skills of January–March forecasts in California and Alaska are robust over all initializations (Figures 2e and 2h) and show strong dependence on target season. However, the prediction skills are only significant in the first two seasons for

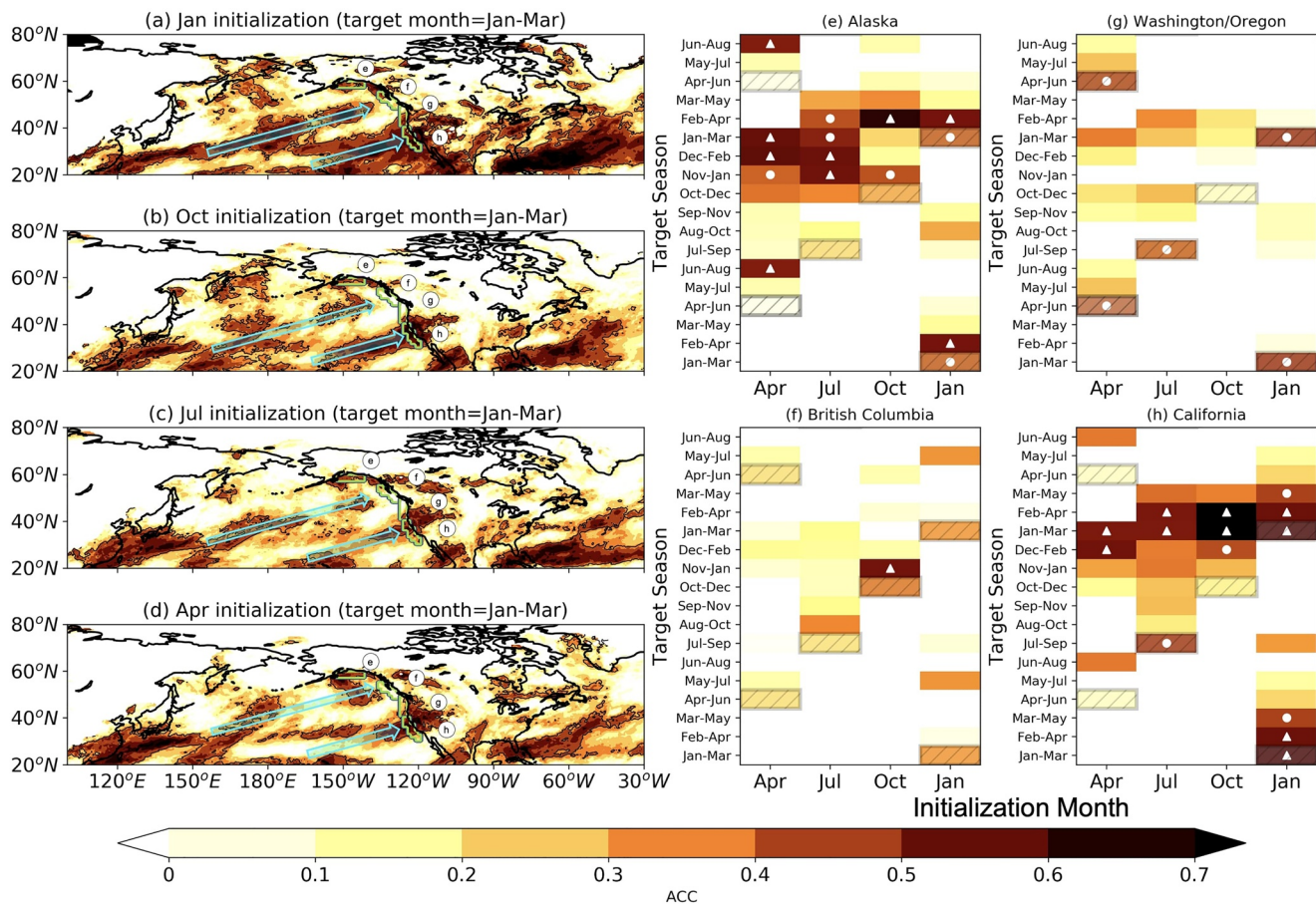


Figure 2. Prediction skills of seasonal Atmospheric river (AR) frequency measured by the anomaly correlation coefficient (ACC). (a)–(d) The ACC for January–March forecasts made in (a) January, (b) October, (c) July and (d) April. (e)–(h) are averaged over four offshore regions near to (e) Alaska, (f) British Columbia, (g) Washington/Oregon, and (h) California (the four regions are outlined in green in the left panels). Only values above zero are plotted in color. Hatching in (e)–(h) is the initialization month. Black contours in (a)–(d) and white dots in (e)–(f) represent values that are significantly higher than 0 at the 5% level based on a two-tailed t -test with a Fisher z -transformation. White triangles are ACC higher than 0.5. In addition, both white dots and white triangles indicate Seamless System for Prediction and Earth System Research (SPEAR) outperforms the persistence forecast. The effective degrees of freedom are seasonally stratified and defined as $N \frac{1 - \rho(1)}{1 + \rho(1)}$, where $N = 23$ years and $\rho(1)$ is the lag-1 autocorrelation of the AR time series. The two arrows in (a)–(d) represent the axes of elevated ACC discussed in the main text.

British Columbia and Washington/Oregon. We also investigate the skill score based on persistence forecasts (not shown). Persistence outperforms both SPEAR and climatological reference forecast over at the first season but quickly decays to 0 at longer lead times, suggesting the short memory of AR activity. Referring to the spatial ACC structure (Figures 2a–2d), one can find this regional variation reflects the broad pattern of ACC. This regional and seasonal dependence of ACC is closely related to the leading variability of AR frequency, which will be discussed in the next section.

3.3. Sources of AR Prediction Skill in Western North America

To identify the dominant predictability sources of AR activity, we apply an empirical orthogonal function (EOF) analysis to seasonal AR anomalies from 1995 to 2018 over the North Pacific and western North America. Figures 3a and 3b show the EOF1 patterns from the ERA5 reanalysis and SPEAR model output respectively. The two EOF1 patterns are nearly identical, showing the capability of the model in simulating the dominant pattern of AR variability. The local extrema lie on the northern and southern flanks of the climatological AR frequency maximum (contours in Figures 3a and 3b). The EOF1 pattern therefore represents a meridional shift of the subtropical jet, which is consistent with the pattern identified in previous

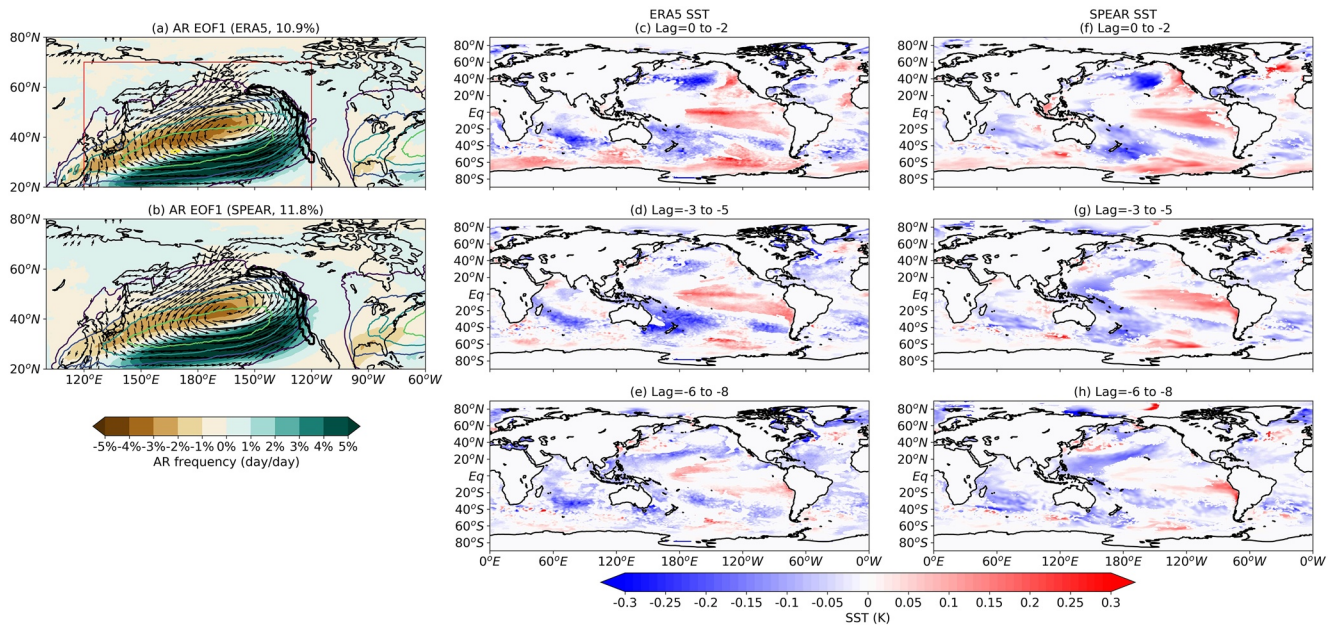


Figure 3. The regression pattern of Atmospheric river (AR) frequency on the first AR frequency principal component (PC1) from the (a) European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis (ERA5) reanalysis and (b) Seamless System for Prediction and Earth System Research (SPEAR) reforecasts. The domain for the empirical orthogonal function (EOF) analysis is 120°E – 120°W , 20°N – 70°N (red box in Figure 4a). The numbers inside the parentheses provide the variance explained by the EOF1 patterns. All seasons from 1995 to 2018 are used in the EOF analysis. Vectors show the 950-hPa wind regressed onto PC1. Vector sizes are proportional to the wind intensity with amplitude smaller than 0.8 ms^{-1} omitted. The contours are climatological AR frequency with an interval of 4% . The right two columns are the lagged SST anomaly (K) composites from (c)–(e) ERA5 and (f)–(h) SPEAR. Lag 0 is defined when the January–March AR PC1 is greater than 1 standard deviation. Only regions where both the forecast and observational data agree on the sign of sea surface temperature (SST) anomaly are shaded.

studies (W. Zhang & Villarini, 2018). The variance explained by EOF1 is significantly higher than any of the other EOFs in both ERA5 and SPEAR (Figure S6). The extrema of EOF1 are spatially collocated with the regions of high ACC in Figures 2a–2d, indicating a strong link between the dominant mode of AR variability and the AR seasonal prediction skill.

Because the initial state of the atmosphere primarily benefits predictions of the first season, the persistence of skill at longer leads is likely due to SST forcing from the ocean. We therefore investigate the lagged relationships between the leading EOF of AR frequency and global SST anomalies. Figures 3c–3h illustrate the lagged composites of seasonal SST anomalies from ERA5 and SPEAR, respectively. Lag 0 is defined when the amplitude of the first principal component (PC1) exceeds 1 standard deviation in the target season of January to March. Negative lags indicate the number of months before lag 0. For example, if the forecast PC1 in January–March 1998 is greater than 1 standard deviation, then the SST pattern from November 1997 to January 1998 is considered as lag –2 to lag 0 in our composite. Only regions for which both SPEAR and ERA5 agree on the sign of the SST anomalies are shaded.

In the preceding November to January, we see prominent SST anomalies in both the tropical and extratropical Pacific that potentially could serve as sources of seasonal AR predictability. The anomalous warming in the tropical eastern Pacific indicates an association between the positive phase of EOF1 and El Niño, particularly of the Central Pacific (CP) type with maximum warming near the date line (Capotondi et al., 2015). It also indicates a projection onto the warm phase of the Interdecadal Pacific Oscillation (IPO; Text S4 in Supporting Information S1; Henley et al. (2015)), with a warm anomaly along the equator flanked by extratropical cold anomalies in both hemispheres. During the positive (warm) phase of the IPO (CP El Niño), the extratropical Pacific is typically characterized by an anomalously deepened Aleutian low (vectors in Figures 3a and 3b). This circulation pattern, which strongly modulates the AR activity, is also identified by previous research based on the NMME (Zhou & Kim, 2018) and observational reanalysis (Gershunov et al., 2017; W. Zhang & Villarini, 2018).

SST signals over the extratropical Pacific can be traced back to 8 months before EOF1 reaches its maximum amplitude (lag -6 to -8 , Figures 3e and 3h) while the signal around the equatorial Pacific shows less agreement between SPEAR and ERA5. This difference might be associated with the spring predictability barrier, where the tropical ENSO signal is weak during its transition stage around late spring, or that 8-month ENSO predictability is not strongly reflected in tropical SST composites at these lags. Indeed, a previous study (Lu et al., 2020) has shown that SPEAR can skillfully predict ENSO 9 months in advance (see also Figures S7 and S8). We demonstrate in Figure S7 that SPEAR also skillfully predicts the IPO at these lead times, which supports that both ENSO and the IPO may be potential seasonal predictability sources for AR activity. To further demonstrate the connection between the AR EOF1 and ENSO (IPO) we calculate the predictability of three time series, where predictability is defined as the ratio of the predictable signal (variance of ensemble mean) to the climatological noise (variance of each member, see Text S5 in Supporting Information S1). The result is shown in Figure S8. The temporal evolution of predictability is similar for all three time series, with maximum around the forecast targeting January–March, which is consistent with Figures 2e and 2h. The result in Figure S8 suggests the high ACC in California and Alaska might be attributed to AR EOF1 and its connections with ENSO and the IPO.

3.4. Challenges of Seasonal AR Prediction Over British Columbia and Washington/Oregon

The EOF1 pattern provides some clues as to why the ARs are more predictable over California than over British Columbia and Washington/Oregon. California is located near the east end of EOF1, which represents a relative maximum in the amount of AR variance explained by the leading EOF. On the other hand, British Columbia and Washington/Oregon are located near the nodal region of EOF1, indicating a relative minimum in the amount of AR variance explained by the leading mode.

Thus, to further elucidate that EOF1 reflects the major predictability source of ARs over the extratropical Pacific and adjacent regions, we calculate the AR variance explained by PC1. Figure S10 demonstrates the spatial pattern of seasonal AR variance explained by EOF1 from ERA5. We focus on January–March, since that season is characterized by the highest AR prediction skill. Figure S10 shows a similar two-stripe pattern as in Figure 2, but with a few subtle differences. EOF1 contributes limited AR variance over British Columbia and Washington/Oregon, which is consistent with the ACC patterns in Figure 2. However, EOF1 also explains little Alaska AR variability, which implies other EOFs must be necessary to explain AR variability in this region. In spite of the discrepancy in Alaska, the correspondence between the Pacific pattern of AR prediction skill (Figures 2a–2d) and the AR activity variance explained by EOF1 (Figure S10) is quite strong. Therefore, the reduced skill at long leads in British Columbia (Figure 2f) and Washington/Oregon (Figure 2g) can largely be explained by their position within a nodal region of EOF1. This challenge is also reflected in seasonal snow pack predictions (Kapnick et al., 2018).

3.5. Probabilistic Predictions of Seasonal AR Activity

The evaluations of AR seasonal prediction skill in Figure 2 are based on deterministic predictions. However, practical implementation of seasonal AR forecasts requires a probabilistic outlook for the likelihood of different scenarios (e.g., above-normal or below-normal AR activity). A challenge in probabilistic climate forecasting is calibrating a reliable outlook, by comparing the relative frequency of an event to the predicted probability. A reliable forecast has a close correspondence between forecast probability (e.g., 50% probability of above-normal AR activity) and observed occurrence (e.g., occurrence 50% of the time when such a forecast is made). An attributes diagram, which shows the observed relative frequency plotted against forecast probability (see Text S7 in Supporting Information S1; van den Dool et al., 2017), can be used to evaluate the reliability of the ensemble forecast. Figure 4a shows the attributes diagram for January–March forecasts, aggregated over the Pacific and western North America, for probabilities calculated as the relative frequency of the ensemble members within below-normal, near-normal, and above-normal terciles of AR activity. This is a standard format used in the NOAA Climate Prediction Center's seasonal outlooks. The attributes diagram in Figure 4a shows that the forecast probabilities tend to exceed the observed relative frequencies for all three forecast categories, indicating that the uncalibrated probabilistic forecasts are overconfident.

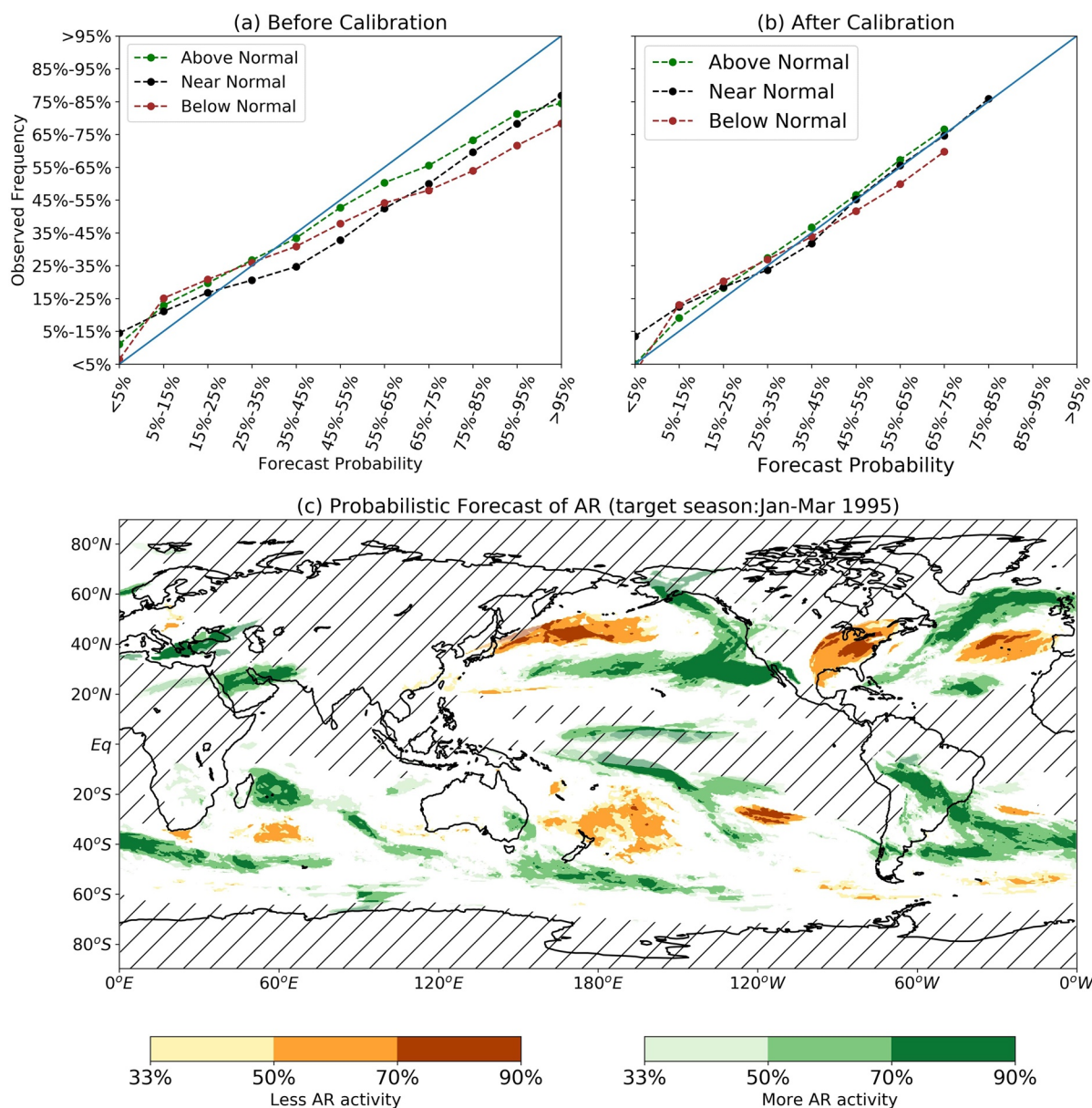


Figure 4. Attributes diagrams for (a) uncalibrated and (b) linearly calibrated Seamless System for Prediction and Earth System Research (SPEAR) probabilistic forecasts, for all initializations and with a target season of January to March. The uncalibrated probabilities are based on ensemble member relative frequency, whereas the method of linear calibration is described in Text S7 (Supporting Information S1). The data are aggregated over all grid points within 120°E – 120°W , 20°N – 70°N . The x-axis is the forecast probability, and the y-axis is the observed relative frequency of the event. The three lines represent the three forecast categories: above-normal (green), near-normal (black) and below-normal (brown). (c) An example of probabilistic forecast product initialized in January 1995 with a target season of January–March, 1995. Orange shading indicate enhanced probabilities of below-normal AR activity, whereas green shading indicates enhanced probability of above-normal AR activity. White regions are enhanced probability of near-normal activity. Hatching represents regions where the frequency of AR climatology is less than 2% (AR day/day) in both SPEAR and European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis (ERA5).

To remove the conditional bias in the probabilistic forecast, we apply a linear calibration based on the probability anomaly correlation (van den Dool et al., 2017; see Text S7 in Supporting Information S1). This improves the forecast reliability (Figure 4b) so that the forecast probability nearly equals the observed relative frequency for all forecast probability bins and for all three categories. The forecasts for below-normal (lower tercile) AR activity remain slightly overconfident after calibration, but still noticeably better than prior to calibration. Figure 4c illustrates an example of calibrated probabilistic forecast for the target season

of January–March, 1995. Overall, these probabilistic forecasts exhibit substantial positive skill, as measured by the Heidke Skill Score, throughout much of western North America for all initializations (Figure S11). This prototype product with simple linear calibration illustrates the potential of integrating the probabilistic forecast of seasonal AR activity with existing storm-based outlook systems.

4. Conclusions and Discussion

Through the analysis of retrospective forecasts from a new dynamical seasonal forecast system, GFDL SPEAR, we find strong potential to provide skillful seasonal-to-multiseasonal forecasts of AR activity over western North America. We first demonstrate that SPEAR skillfully forecasts broad patterns of wintertime AR activity at least 9 months in advance, and certain regions (such as California and Alaska) are characterized by high prediction skill. We further show that this predictability is largely rooted in the strength of the Aleutian low, as captured in the leading EOF pattern of seasonal AR activity, and is strongly tied to both tropical (ENSO) and extratropical Pacific (IPO) SST patterns. This leading EOF also appears to set the pattern of seasonal forecast skill, as the fluctuations in Aleutian low strength are connected to meridional shifts of the Pacific storm track. Regions like California and Alaska that are strongly affected by these meridional shifts have the highest skill at long leads, whereas other regions like Washington/Oregon and British Columbia have a more rapid decay of skill beyond the first season.

The demonstration of significant AR forecast skill out to at least 9 months provides a rationale for developing probabilistic seasonal hazard outlooks for ARs or integrating them into existing storm-based outlook systems. We show that a simple linear calibration on the 15-member ensemble is sufficient for calibrating the forecast distribution and generating reliable probabilistic AR seasonal forecasts. We also note, however, that some regions are characterized by skills not high enough to be useful for decision makers (e.g., white regions with negative HSS in Figure S11). This feature is also reflected in the predictability and prediction skill of AR EOF1, which are generally lower than those of ENSO and IPO (Figures S7 and S8). One possible explanation is that the nature of low signal-to-noise ratio in AR activity limits the utility of basin-wide predictions. To foster development of the most effective probabilistic AR forecasts, interdisciplinary collaborations between model developers, operational forecasters as well as stakeholders are required.

As ARs can trigger a wide range of impacts, the gains in fundamental understanding of seasonal AR predictability and the demonstration of significant forecast skill potentially may benefit diverse sectors ranging from water resource management to disaster preparedness. Future work will attempt to expand from forecasts of AR activity to forecasts of AR impacts (like flooding) that are more directly relevant to users. Such an expansion raises additional scientific questions worthy of pursuit. For example, are there differences in the dynamics and seasonal predictability of AR-related and AR-unrelated precipitation over western North America? Such differences could support partitioning hydroclimate forecasts into AR-related and AR-unrelated components. Improvements are also needed in understanding the impacts of land processes, model resolution, and model physics on seasonal AR predictions. Increasing resolution beyond what is presented here may help illuminate regional AR variability and improve understanding of the intrinsic limits of AR predictability at longer leads, in particular for apparently less-predictable regions such as Washington/Oregon and British Columbia.

Data Availability Statement

The AR detection algorithm is available at <https://mountainscholar.org/handle/10217/170619>. The post-processed SPEAR hindcast data used in the main text is stored at <https://zenodo.org/record/5217636#.YR5rZ45KiUk>, and a subset of the data will be publicly available via the NMME project (<https://www.cpc.ncep.noaa.gov/products/NMME/>) when SPEAR becomes operational. ERA5 data are available at <https://cds.climate.copernicus.eu/#!/search?text=ERA5&type=dataset>.

Acknowledgments

This research study is supported by award NA18OAR4320123 from NOAA, U.S. Department of Commerce and by the NOAA Office of Water and Air Quality FACETs program. This research study is also supported by the NOAA OAR Weather Portfolio grant.

References

- Alexander, M. A., & Deser, C. (1995). A mechanism for the recurrence of wintertime midlatitude SST anomalies. *Journal of Physical Oceanography*, 25(1), 122–137. [https://doi.org/10.1175/1520-0485\(1995\)025<0122:amftro>2.0.co;2](https://doi.org/10.1175/1520-0485(1995)025<0122:amftro>2.0.co;2)
- Argo. (2020). Argo float data and metadata from global data assembly centre (argo GDAC). Retrieved from <https://www.seanoe.org/data/00311/42182/>
- Bach, E., Motesharrei, S., Kalnay, E., & Ruiz-Barradas, A. (2019). Local atmosphere-ocean predictability: Dynamical origins, lead times, and seasonality. *Journal of Climate*, 32(21), 7507–7519. <https://doi.org/10.1175/JCLI-D-18-0817.1>
- Baggett, C. F., Barnes, E. A., Maloney, E. D., & Mundhenk, B. D. (2017). Advancing atmospheric river forecasts into subseasonal-to-seasonal time scales. *Geophysical Research Letters*, 44(14), 7528–7536. <https://doi.org/10.1002/2017GL074434>
- Byju, P., Dommenget, D., & Alexander, M. A. (2018). Widespread reemergence of sea surface temperature anomalies in the global oceans, including tropical regions forced by reemerging winds. *Geophysical Research Letters*, 45(15), 7683–7691. <https://doi.org/10.1029/2018GL079137>
- Camargo, S. J., & Barnston, A. G. (2009). Experimental dynamical seasonal forecasts of tropical cyclone activity at Iri. *Weather and Forecasting*, 24(2), 472–491. <https://doi.org/10.1175/2008WAF2007099.1>
- Capotondi, A., Wittenberg, A. T., Newman, M., Di Lorenzo, E., Yu, J.-Y., Braconnot, P., et al. (2015). Understanding ENSO diversity. *Bulletin of the American Meteorological Society*, 96(6), 921–938. <https://doi.org/10.1175/BAMS-D-13-00117.1>
- Corringham, T. W., Ralph, F. M., Gershunov, A., Cayan, D. R., & Talbot, C. A. (2019). Atmospheric rivers drive flood damages in the western United States. *Science Advances*, 5(12), eaax4631. <https://doi.org/10.1126/sciadv.aax4631>
- Dacre, H. F., Clark, P. A., Martinez-Alvarado, O., Stringer, M. A., & Lavers, D. A. (2015). How do atmospheric rivers form? *Bulletin of the American Meteorological Society*, 96(8), 1243–1255. <https://doi.org/10.1175/bams-d-14-00031.1>
- DeFlorio, M. J., Waliser, D. E., Ralph, F. M., Guan, B., Goodman, A., Gibson, P. B., et al. (2019). Experimental subseasonal-to-seasonal (s2s) forecasting of atmospheric rivers over the western United States. *Journal of Geophysical Research: Atmospheres*, 124(21), 11242–11265. <https://doi.org/10.1029/2019JD031200>
- Delworth, T. L., Cooke, W. F., Adcroft, A., Bushuk, M., Chen, J.-H., Dunne, K. A., et al. (2020). Spear: The next generation GFDL modeling system for seasonal to multidecadal prediction and projection. *Journal of Advances in Modeling Earth Systems*, 12(3), e2019MS001895. <https://doi.org/10.1029/2019MS001895>
- Dettinger, M. D. (2013). Atmospheric rivers as drought busters on the U.S. West Coast. *Journal of Hydrometeorology*, 14(6), 1721–1732. <https://doi.org/10.1175/JHM-D-13-02.1>
- Gershunov, A., Shulgina, T., Ralph, F. M., Lavers, D. A., & Rutz, J. J. (2017). Assessing the climate-scale variability of atmospheric rivers affecting western North America. *Geophysical Research Letters*, 44(15), 7900–7908. <https://doi.org/10.1002/2017GL074175>
- Gray, W. M. (1984a). Atlantic seasonal hurricane frequency. Part I: El Niño and 30 mb quasi-biennial oscillation influences. *Monthly Weather Review*, 112(9), 1649–1668. [https://doi.org/10.1175/1520-0493\(1984\)112<1649:ashfpi>2.0.co;2](https://doi.org/10.1175/1520-0493(1984)112<1649:ashfpi>2.0.co;2)
- Gray, W. M. (1984b). Atlantic seasonal hurricane frequency. Part II: Forecasting its variability. *Monthly Weather Review*, 112(9), 1669–1683. [https://doi.org/10.1175/1520-0493\(1984\)112<1669:ashfpi>2.0.co;2](https://doi.org/10.1175/1520-0493(1984)112<1669:ashfpi>2.0.co;2)
- Henley, B. J., Gergis, J., Karoly, D. J., Power, S., Kennedy, J., & Folland, C. K. (2015). A tripole index for the Interdecadal Pacific Oscillation. *Climate Dynamics*, 45, 3077–3090. <https://doi.org/10.1007/s00382-015-2525-1>
- 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>
- Hoerling, M. P., & Ting, M. (1994). Organization of extratropical transients during El Niño. *Journal of Climate*, 7(5), 745–766. [https://doi.org/10.1175/1520-0442\(1994\)007<0745:ooetde>2.0.co;2](https://doi.org/10.1175/1520-0442(1994)007<0745:ooetde>2.0.co;2)
- Hoskins, B. J., & Karoly, D. J. (1981). The steady linear response of a spherical atmosphere to thermal and orographic forcing. *Journal of the Atmospheric Sciences*, 38(6), 1179–1196. [https://doi.org/10.1175/1520-0469\(1981\)038<1179:tslroa>2.0.co;2](https://doi.org/10.1175/1520-0469(1981)038<1179:tslroa>2.0.co;2)
- Jiang, X., Lau, N.-C., Held, I. M., & Ploshay, J. J. (2007). Mechanisms of the Great Plains low-level jet as simulated in an AGCM. *Journal of the Atmospheric Sciences*, 64(2), 532–547. <https://doi.org/10.1175/JAS3847.1>
- Kapnick, S. B., Yang, X., Vecchi, G. A., Delworth, T. L., Gudgel, R., Malyshev, S., et al. (2018). Potential for western US seasonal snowpack prediction. *Proceedings of the National Academy of Sciences*, 115(6), 1180–1185. <https://doi.org/10.1073/pnas.1716760115>
- Kirtman, B. P., Min, D., Infanti, J. M., Kinter, I., James, L., Paolino, D. A., et al. (2014). The North American multimodel ensemble: Phase-1 seasonal-to-interannual prediction; Phase-2 toward Developing Intraseasonal Prediction. *Bulletin of the American Meteorological Society*, 95(4), 585–601. <https://doi.org/10.1175/BAMS-D-12-00050.1>
- Lang, A. L., Pegion, K., & Barnes, E. A. (2020). Introduction to special collection: “bridging weather and climate: Subseasonal-to-seasonal (s2s) prediction”. *Journal of Geophysical Research: Atmospheres*, 125(4), e2019JD031833. <https://doi.org/10.1029/2019JD031833>
- Lavers, D. A., Waliser, D. E., Ralph, F. M., & Dettinger, M. D. (2016). Predictability of horizontal water vapor transport relative to precipitation: Enhancing situational awareness for forecasting western U.S. extreme precipitation and flooding. *Geophysical Research Letters*, 43(5), 2275–2282. <https://doi.org/10.1002/2016GL067765>
- Lu, F., Harrison, M. J., Rosati, A., Delworth, T. L., Yang, X., Cooke, W. F., et al. (2020). Gfdl’s spear seasonal prediction system: Initialization and Ocean Tendency Adjustment (OTA) for coupled model predictions. *Journal of Advances in Modeling Earth Systems*, 12(12), e2020MS002149. <https://doi.org/10.1029/2020MS002149>
- Mundhenk, B. D., Barnes, E. A., & Maloney, E. D. (2016). All-season climatology and variability of atmospheric river frequencies over the north Pacific. *Journal of Climate*, 29(13), 4885–4903. <https://doi.org/10.1175/JCLI-D-15-0655.1>
- Mundhenk, B. D., Barnes, E. A., Maloney, E. D., & Baggett, C. F. (2018). Skillful empirical subseasonal prediction of landfalling atmospheric river activity using the Madden-Julian oscillation and quasi-biennial oscillation. *Atmospheric and Climate Sciences*, 1(20177). <https://doi.org/10.1038/s41612-017-0008-2>
- Murakami, H., Vecchi, G. A., Underwood, S., Delworth, T. L., Wittenberg, A. T., Anderson, W. G., et al. (2015). Simulation and prediction of category 4 and 5 hurricanes in the high-resolution GFDL HiFLOR coupled climate model*. *Journal of Climate*, 28(23), 9058–9079. <https://doi.org/10.1175/JCLI-D-15-0216.1>
- Newell, R. E., Newell, N. E., Zhu, Y., & Scott, C. (1992). Tropospheric rivers?—A pilot study. *Geophysical Research Letters*, 19(24), 2401–2404. <https://doi.org/10.1029/92GL02916>
- Newman, M., Alexander, M. A., Ault, T. R., Cobb, K. M., Deser, C., Di Lorenzo, E., et al. (2016). The Pacific Decadal Oscillation, revisited. *Journal of Climate*, 29(12), 4399–4427. <https://doi.org/10.1175/JCLI-D-15-0508.1>
- O’Brien, T. A., Collins, W. D., Kashinath, K., Rübel, O., Byna, S., Gu, J., et al. (2016). Resolution dependence of precipitation statistical fidelity in hindcast simulations. *Journal of Advances in Modeling Earth Systems*, 8(2), 976–990. <https://doi.org/10.1002/2016MS000671>

- Pasquier, J. T., Pfahl, S., & Grams, C. M. (2019). Modulation of atmospheric river occurrence and associated precipitation extremes in the North Atlantic region by European weather regimes. *Geophysical Research Letters*, 46(2), 1014–1023. <https://doi.org/10.1029/2018GL081194>
- Payne, A. E., & Magnusdottir, G. (2014). Dynamics of landfalling atmospheric rivers over the North Pacific in 30 years of MERRA reanalysis. *Journal of Climate*, 27(18), 7133–7150. <https://doi.org/10.1175/JCLI-D-14-00034.1>
- Ralph, F. M., Coleman, T., Neiman, P. J., Zamora, R. J., & Dettinger, M. D. (2013). Observed impacts of duration and seasonality of atmospheric-river landfalls on soil moisture and runoff in coastal northern California. *Journal of Hydrometeorology*, 14(2), 443–459. <https://doi.org/10.1175/JHM-D-12-076.1>
- Ralph, F. M., Dettinger, M. D., Cairns, M. M., Galarneau, T. J., & Eylander, J. (2018). Defining “atmospheric river”: How the glossary of meteorology helped resolve a debate. *Bulletin of the American Meteorological Society*, 99(4), 837–839. <https://doi.org/10.1175/bams-d-17-0157.1>
- Ralph, F. M., Neiman, P. J., & Wick, G. A. (2004). Satellite and caljet aircraft observations of atmospheric rivers over the eastern north pacific ocean during the winter of 1997/98. *Monthly Weather Review*, 132(7), 1721–1745. [https://doi.org/10.1175/1520-0493\(2004\)132<1721:sacao>2.0.co;2](https://doi.org/10.1175/1520-0493(2004)132<1721:sacao>2.0.co;2)
- Ralph, F. M., Rutz, J. J., Cordeira, J. M., Dettinger, M., Anderson, M., Reynolds, D., et al. (2019). A scale to characterize the strength and impacts of atmospheric rivers. *Bulletin of the American Meteorological Society*, 100(2), 269–289. <https://doi.org/10.1175/BAMS-D-18-0023.1>
- van den Dool, H., Becker, E., Chen, L.-C., & Zhang, Q. (2017). The probability anomaly correlation and calibration of probabilistic forecasts. *Weather and Forecasting*, 32(1), 199–206. <https://doi.org/10.1175/WAF-D-16-0115.1>
- Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., et al. (2017). The subseasonal to seasonal (s2s) prediction project database. *Bulletin of the American Meteorological Society*, 98(1), 163–173. <https://doi.org/10.1175/BAMS-D-16-0017.1>
- Vitart, F., & Robertson, A. (2018). The sub-seasonal to seasonal prediction project (s2s) and the prediction of extreme events. *Atmospheric and Climate Sciences*, 1(3). <https://doi.org/10.1038/s41612-018-0013-0>
- White, C. J., Carlsen, H., Robertson, A. W., Klein, R. J., 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>
- Yang, X., Vecchi, G. A., Gudgel, R. G., Delworth, T. L., Zhang, S., Rosati, A., et al. (2015). Seasonal Predictability of extratropical storm tracks in GFDL's High-resolution climate prediction model. *Journal of Climate*, 28(9), 3592–3611. <https://doi.org/10.1175/JCLI-D-14-00517.1>
- Zhang, W., & Villarini, G. (2018). Uncovering the role of the East Asian jet stream and heterogeneities in atmospheric rivers affecting the western United States. *Proceedings of the National Academy of Sciences*, 115(5), 891–896. <https://doi.org/10.1073/pnas.1717883115>
- Zhang, Z., Ralph, F. M., & Zheng, M. (2019). The relationship between extratropical cyclone strength and atmospheric river intensity and position. *Geophysical Research Letters*, 46(3), 1814–1823. <https://doi.org/10.1029/2018GL079071>
- Zhao, M. (2020). Simulations of atmospheric rivers, their variability and response to globalwarming using GFDL's new high resolution general circulation model. *Journal of Climate*, 33, 1–10303. <https://doi.org/10.1175/JCLI-D-20-0241.1>
- Zheng, C., Kar-Man Chang, E., Kim, H.-M., 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>
- Zhou, Y., & Kim, H. (2018). Prediction of atmospheric rivers over the North Pacific and its connection to ENSO in the North American multi-model ensemble (NMME). *Climate Dynamics*, 51, 1623–1637. <https://doi.org/10.1007/s00382-017-3973-6>
- Zhu, Y., & Newell, R. E. (1998). A proposed algorithm for moisture fluxes from atmospheric rivers. *Monthly Weather Review*, 126(3), 725–735. [https://doi.org/10.1175/1520-0493\(1998\)126<0725:apafmf>2.0.co;2](https://doi.org/10.1175/1520-0493(1998)126<0725:apafmf>2.0.co;2)

References From the Supporting Information

- Guan, B., & Waliser, D. E. (2015). Detection of atmospheric rivers: Evaluation and application of an algorithm for global studies. *Journal of Geophysical Research: Atmospheres*, 120, 12514–12535. <https://doi.org/10.1002/2015jd024257>
- Guan, B., Waliser, D. E., & Ralph, F. M. (2018). An intercomparison between reanalysis and drop-sonde observations of the total water vapor transport in individual atmospheric rivers. *Journal of Hydrometeorology*, 19, 321–337. <https://doi.org/10.1175/jhm-d-17-0114.1>
- Jia, L., Yang, X., Vecchi, G. A., Gudgel, R. G., Delworth, T. L., Rosati, A., et al. (2015). Improved Seasonal prediction of temperature and precipitation over land in a high-resolution GFDL climate model. *Journal of Climate*, 28, 2044–2062. <https://doi.org/10.1175/jcli-d-14-00112.1>
- North, G. R., Bell, T. L., Cahalan, R. F., & Moeng, F. J. (1982). Sampling errors in the estimation of empirical orthogonal functions. *Monthly Weather Review*, 110, 699–706. [https://doi.org/10.1175/1520-0493\(1982\)110<0699:seiteo>2.0.co;2](https://doi.org/10.1175/1520-0493(1982)110<0699:seiteo>2.0.co;2)