Geophysical Research Letter

Supporting Information for

Are multiseasonal forecasts of atmospheric rivers possible?

Kai-Chih Tseng[1,2], Nathaniel C. Johnson[1], Sarah B. Kapnick[1], Thomas L. Delworth[1], Feiyu Lu[1,2], William Cooke[1], Andrew T. Wittenberg[1], Anthony J. Rosati[1,3], Liping Zhang[1,3], Colleen McHugh[1,5], Xiaosong Yang[1], Matthew Harrison[1], Fanrong Zeng[1], Gan Zhang[6], Hiroyuki Murakami[1,3,4], Mitchell Bushuk[1,3] and Liwei Jia[1,3]

1. Geophysical Fluid Dynamics Laboratory, National Oceanic and Atmospheric Administration,Princeton, NJ 08540, USA
2. Program in Atmospheric and Oceanic Science, Department of Geosciences, Princeton University,Princeton, NJ 08540, USA
3. University Corporation for Atmospheric Research, Boulder, CO 80307, USA
4. Meteorological Research Institute, Tsukuba, Japan
5. SAIC, Science Applications International Corporation, Reston, VA 20190, USA
6. Citadel Americas, LLC

Contents of this file

Text S1 to S8
Figures S1 to S11

Introduction

This document is the supporting information for the main text including: (Text S1) Additional information about reanalysis data; (Text S2-S3, Figures S1-S5) Discussion about AR climatology, resolution-dependent bias and bias-correction methods; (Text S4) Deriving the IPO index in a 23-yr data; (Text S5, Figures S6-S10) The calculation and discussion of predictability; (Text S7-S9, Figure S10) Detailed information about multiseasonal AR probabilistic forecast.

Text S1. Reanalysis Data

ERA5 reanalysis data generated by the ECMWF reforecast system from 1995-2018 is used in this study. ERA5 is published within 3 months of real time and is available at: https://cds.climate.copernicus.eu/#!/search?text=ERA5&type=dataset. To calculate the IVT, we use the variables at six vertical levels (1000hPa, 925hPa, 850hPa, 700hPa, 500hPa, 250hPa) available from both ERA5 and SPEAR. Previous studies have demonstrated that the AR characteristics based on different reanalysis products (e.g., Modern-Era Retrospective analysis for Research and Applications, MERRA/MERRA-2, and ERA-Interim) are remarkably similar to each other (Guan and Guan 2015, Guan et al. 2018). Given that ERA5 is a new generation reanalysis product, benefiting from new data assimilation techniques, dynamical cores and model physics, data quality is expected to be better in ERA5 than in ERA-Interim.

Text S2. Resolution-dependent biases

The ERA5 data is generated by the ECMWF global reforecast system with spatial resolution of 31km; the higher atmospheric resolution than in SPEAR allows the simulation of strong IVT in more confined regions than in SPEAR. This suggests some events with narrow plumes but strong IVT might not be well-resolved in SPEAR, a circumstance that is expected over the regions with the most active AR occurrence (e.g., to the northeast of Hawaii in Fig.1e). Consistently, the biases in AR climatology over North America appear to depend on model resolution. At 50 km resolution, mountains are lower and smoother compared to those in ERA5. Thus, a landfalling AR is able to penetrate farther inland in SPEAR than in nature due to coarser topography. This likely explains why AR frequency is overestimated over western North America but underestimated over the adjacent ocean. On the other hand, the Great Plains low-level jet and associated ARs are less active due to a weaker east-west pressure gradient force and smoother topography than in nature. All of these biases are more evident in 100km simulations of SPEAR (Fig. S1), which supports that the biases depend on spatial resolution.

Text S3. ACC in 100 km simulations

The ACCs for seasonal AR activity, as in Fig. 2 but for the 100 km forecast simulations, are shown in Fig. S4. The differences in ACC between the 100 km and 50 km simulations are shown in Fig. S5. In general, the results of the 100 km simulations are qualitatively similar to those of the 50 km simulations, but with slightly higher skills, which may result from the higher signal-to-noise ratio resulting from the coarser model resolution. However, this difference is not statistically significant according to Fig. S5. The 50 km simulations are characterized by smaller climatological biases (Fig. S1), which is why the 50 km simulations are preferred for seasonal AR prediction.

Text S4. IPO index

The IPO index is traditionally calculated as the second principal component of low-pass filtered (6- or 13-year cutoff is usually used) global SST. Because the length of the SPEAR data record is limited to only 23 years, we use an alternative approach to derive the IPO index in both ERA5 and SPEAR. We first regressed monthly ERA5 SST anomaly onto the historical IPO index, which can be acquired from https://psl.noaa.gov/data/timeseries/IPOTPI/. Then, both the ERA5 and SPEAR monthly SST anomalies are projected onto the ERA5 regression pattern over the whole Pacific basin to derive the IPO index. The reconstructed index based on ERA5 is highly correlated with the original IPO time series (ρ>0.93).

Text S5. Predictability of AR PC1, IPO and ENSO

To define the predictability of AR PC1, IPO, and ENSO, we use the definition given by Jia et al. (2015)

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

where $P(\tau)$ is the predictability, σ_τ^2 is the forecast variance as a function of lead time τ, and σ_{clim}^2 is the climatological variance.

σ_τ^2 is further defined as $\frac{1}{TE}[x(\tau,t,e) - < x(\tau,t,e) >]^T[x(\tau,t,e) - < x(\tau,t,e) >]$. x is the variable of interest (i.e., AR PC1, IPO index or NINO3.4), which is a function of forecast lead time (τ), start time (t), and ensemble member (e). The angle brackets denote the average over ensemble members. $\frac{1}{T}$ and $\frac{1}{E}$ indicate the dot product is averaged over all start times and ensemble members, where T is the total number of start times and E is the ensemble size. σ_{clim}^2 defined as $\frac{1}{TE}[x(\tau,t,e)]^T[x(\tau,t,e)]$, which is the variance of x averaged over all t and e. As τ increases, $< x(\tau,t,e) > \rightarrow 0$, $\sigma_\tau^2 \rightarrow \sigma_{clim}^2$, and $P(\tau) \rightarrow 0$, indicating a loss of predictability. The results for AR PC1, IPO, and ENSO are shown in Fig. S8.

Text S6. Sample Size in Fig. 3

In Fig. 3, lag 0 is defined when the January - March AR PC1 is greater than 1 standard deviation. This definition can lead to an issue of small sample sizes while there is only four-year data in the ERA5 composite map. However, it is not case for SPEAR since each ensemble member is considered as an independent realization when we applied EOF analysis to the SPEAR data. This argument can justify by inspecting Fig. S8. In Fig. S8, one can find the predictability of IPO decays to 0.5 for April initialization targeting January to March suggesting the predictable signal and noise from interval variability have comparable amplitude at this lead time. This also indicates the ensemble spread is big enough to sample different phases of IPO SST and each ensemble member can be considered as an independent realization. Thus, the result shown in figure 3 is still robust even without ERA5.

Text S7. Attributes Diagram and Calibration of Probabilistic Forecasts

Fig. 4(a)-(b) in the main text shows the attributes diagram for the SPEAR probabilistic forecasts targeting January through March. Results are aggregated over all grid points within 120ºE-120ºW, 20ºN-70ºN. The probabilities are divided into 11 bins, starting from 0-5%with an interval of 5%. The three lines represent three forecast categories: above-normal (green), near-normal (black) or below-normal (brown) AR activity, which are defined as the top, middle or bottom 1/3 (i.e., terciles) of the seasonal AR anomaly distribution from SPEAR. The attributes diagram shows the observed AR frequency as a function of forecast probability. The uncalibrated probabilistic forecasts are calculated as the fraction of ensemble members falling into each of the three forecast categories. Before calibration (Fig. 4a), we find that the model is overconfident for all three categories. For example, when SPEAR forecasts a 95% chance of above-normal AR activity, the above-normal category verifies only 70%of the time. To remove the conditional bias in the probabilistic forecast, we use a simple linear calibration function based on the probability anomaly correlation (PAC) following van den Dool et al. (2017):

$$PAC = \frac{\frac{1}{N}\sum_{i=1}^{N} p_i o_i}{\frac{1}{N}\sqrt{\sum_{i=1}^{N} p_i^2 \sum_{i=1}^{N} o_i^2}} \quad (2)$$

$$p_i^* = PAC \frac{\sigma_0}{\sigma_p} p_i \quad (3)$$

where p_i is the uncalibrated forecast probability (e.g., the fraction of ensemble members forecasting above-normal AR activity), o_i is the observed event (0 for non-occurrence or 1 for occurrence), and N is the total number of forecasts. In Eq. S3, σ_0 is the standard deviation of o_i and σ_p is the standard deviation of p_i. Eq. S3 is a linear regression function which converts the p_i into the calibrated values, p_i^*. We use a leave-one-year-out cross-validation approach to construct and evaluate the calibration function. Specifically, the evaluation of the verification for a given year is based on the calibration from the statistics of all other years to ensure independence of the verification subset.

Fig. 4(b) in the main text illustrates the calibrated attributes diagram, and Fig. 4(c) is an example of a calibrated forecast map. In Fig. 4b, we find that the calibrated probabilistic forecasts fall near the diagonal line of the attributes diagram, indicating that the forecast probability is close to the observed relative frequency. This result indicates that the calibrated probabilistic forecasts are reliable.

Text S8. Heidke skill score

To evaluate the skill of the calibrated probabilistic forecasts, we use the Heidke skill score (HSS), defined as

$$HSS = \frac{H - E}{T - E} \times 100 \quad (4)$$

where H is the number of correct forecasts (i.e., how many seasons SPEAR successfully forecasts above-, near-, or below-normal AR activity), T is the total number of forecasts and E is the expected number of correct forecasts by random chance, which is $\frac{T}{3}$ in our three-category case. For the calculation of H, the probabilistic forecast is converted into a deterministic forecast based on the category with the highest probability. A set of perfect forecasts would receive a HSS of 100, and values above zero can be interpreted as skill relative to a climatological reference forecast. For example, a score of 50 indicates two times more correct forecasts than incorrect forecasts. The HSS maps for seasonal AR prediction are shown in Fig. S11. From Fig. S11(a)-(d), one can find a similar stripe pattern to that of Fig. 2 and of Fig. S4, while the subtropical branch is more evident. In addition, the seasonal dependence of high HSS values for January-March forecasts is also evident in Fig. S11(e)-S11(h), with California and Alaska showing the highest values. The results shown in Fig. S11 illustrate the potential of leveraging SPEAR for seasonal to multiseasonal probabilistic forecasts of AR activity.

References

1. L. Jia, et al., Improved Seasonal Prediction of Temperature and Precipitation over Land in a High-Resolution GFDL Climate Model. *J. Clim*. **28**, 2044–2062 (2015).

2. H. van den Dool, E. Becker, L. C. Chen, Q. Zhang, The probability anomaly correlation and calibration of probabilistic forecasts. *Weather. Forecast*. **32**, 199–206 (2017).

3. G. R. North, T. L. Bell, R. F. Cahalan, F. J. Moeng, Sampling errors in the estimation of empirical orthogonal functions. *Mon. Weather. Rev.***110**, 699–706 (1982)

4. Guan, B. & Waliser, D. E. Detection of atmospheric rivers: Evaluation and application ofan algorithm for global studies. *Journal of Geophysical Research: Atmospheres* **120**, 12514–12535 (2015).

5. Guan, B., Waliser, D. E. & Ralph, F. M. An Intercomparison between Reanalysis and Drop-sonde Observations of the Total Water Vapor Transport in Individual Atmospheric Rivers. *Journal of Hydrometeorology* **19**, 321–337 (2018).

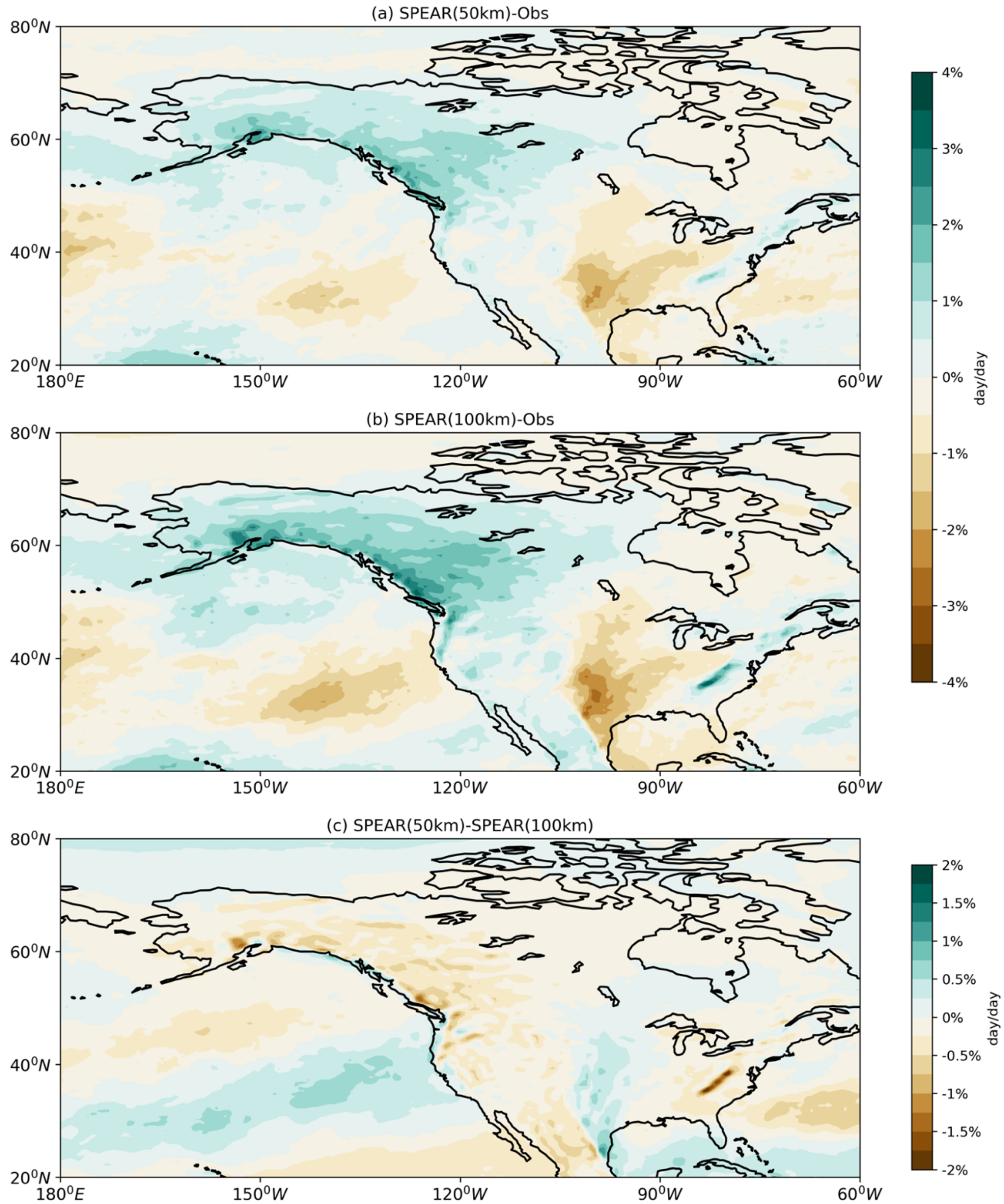

Figure S1. The bias in AR frequency, averaged over all forecast lead times, in the (a) 50-km, (b) 100-km SPEAR reforecasts and (c) the difference between 50-km and 100-km.

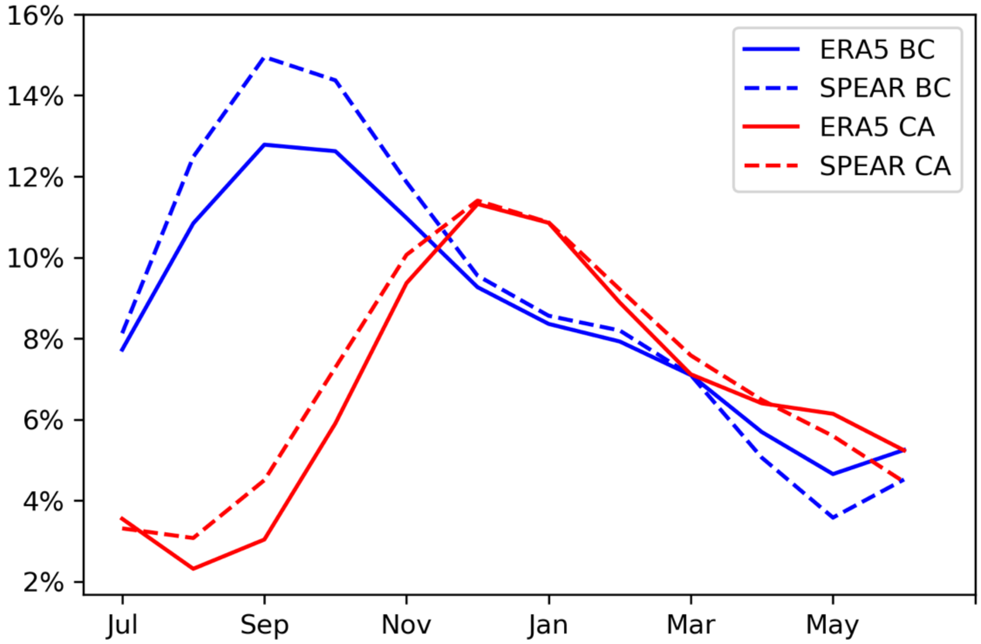

Figure S2. The monthly climatology of AR frequency from SPEAR (January initialization) and ERA5 in California (red) and British Columbia (blue).

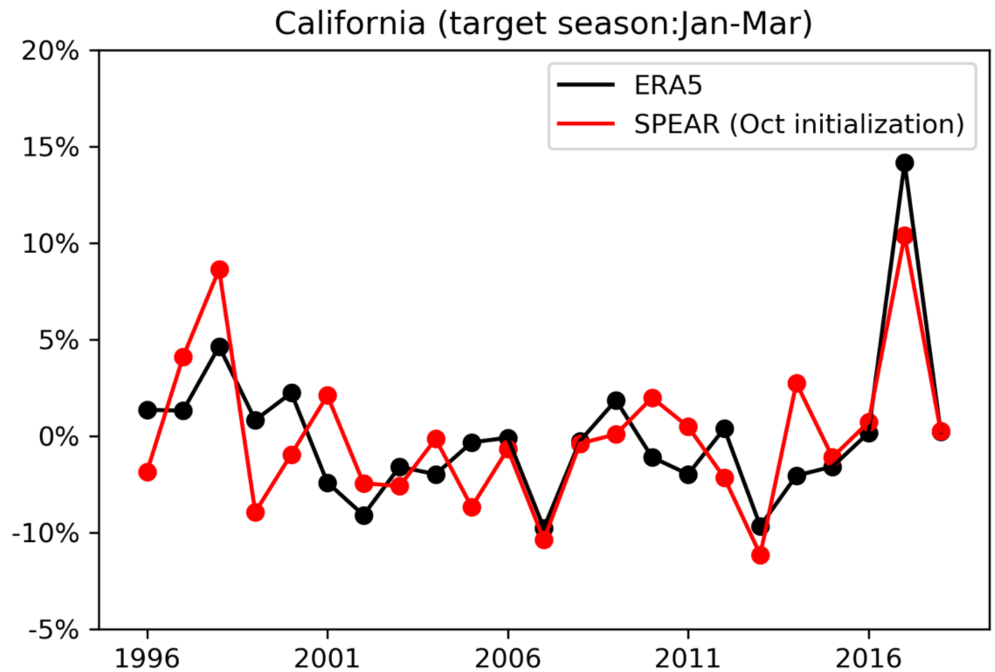

Figure S3. The January-March AR frequency anomaly over California from ERA5 and SPEAR. SPEAR is from October initializations (3 months lead time).

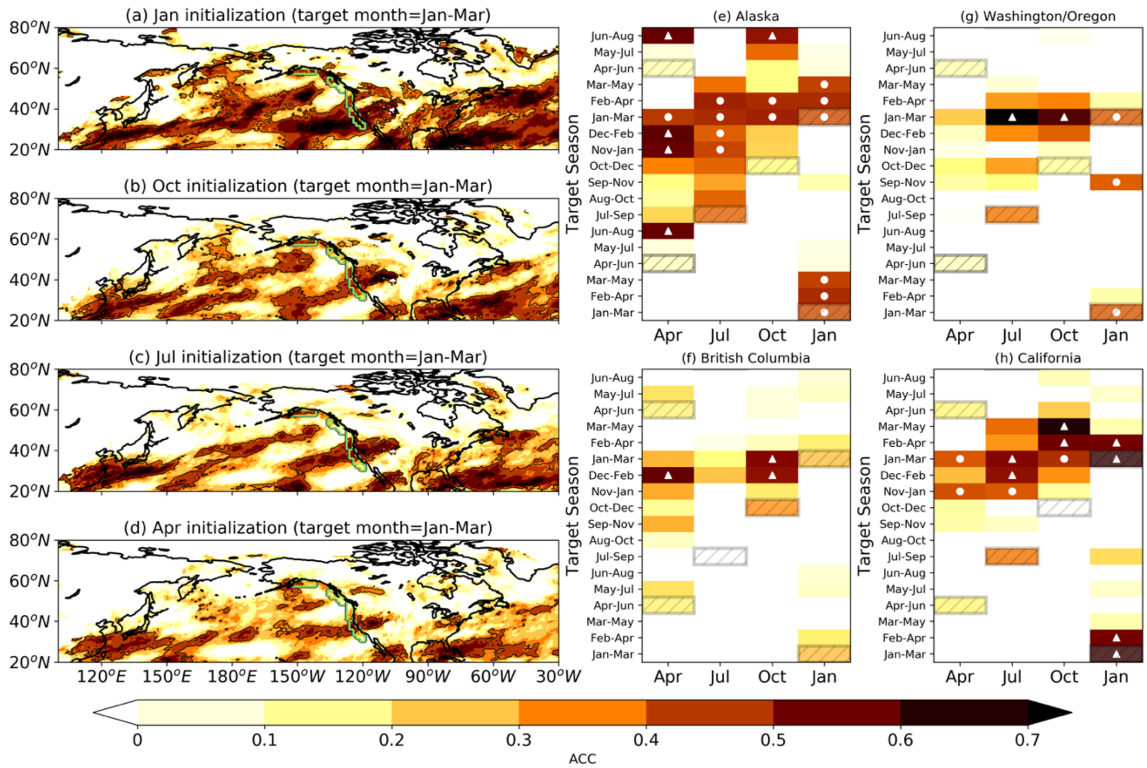

Figure S4. The same as Fig. 3, except for the 100-km SPEAR forecast simulations.

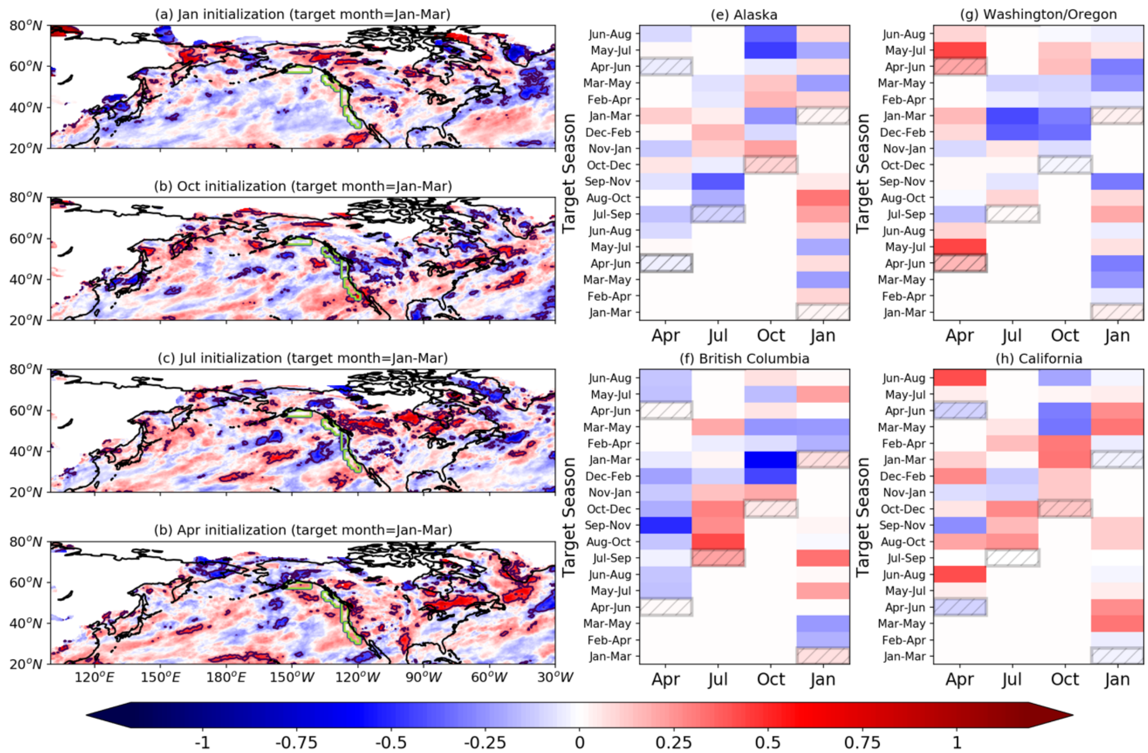

Figure S5. Difference in AR activity ACC between the 50-km SPEAR simulations (Fig. 2) and the 100-km SPEAR simulations (Fig. S4). Black contours represent the difference is statistically significant at the 5% level based on a two-tailed t-test.

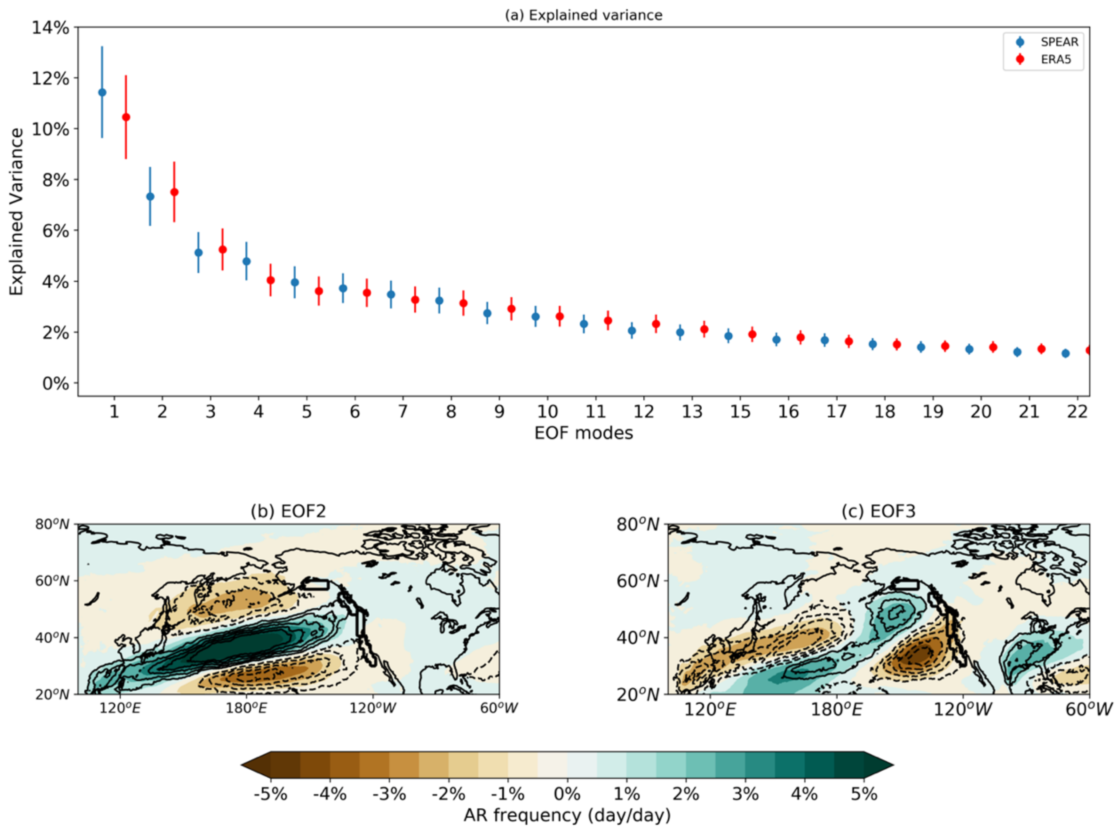

Figure S6. (a) The variance explained by the leading EOFs of seasonal AR activity in the 50-km SPEAR simulations (blue) and ERA5 (red). The error bars show the 95% confidence interval according to North et al. (1982). (b) and (c) show the AR EOF2 and EOF3 regression maps, respectively (shading for SPEAR and contour for ERA5). Shading and contours share the same intervals.

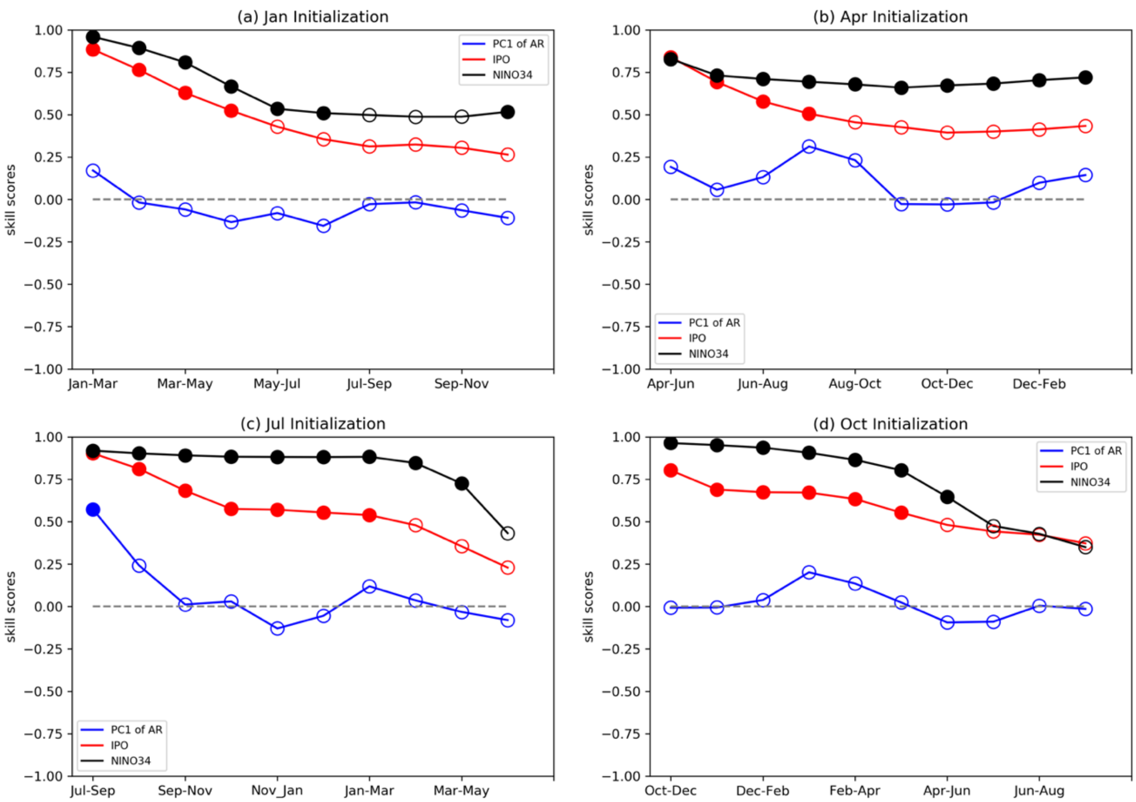

Figure S7. The seasonal prediction correlation skills of AR PC1 (blue solid line), the IPO index (red solid line) and NINO3.4 index (black solid line) for (a) January (b) April (c) July and (d) October initializations. The NINO3.4 index is defined as the SST anomaly averaged over 5°S-5°N and 170°W-120°W. The definition of the IPO index is provided in the SI text. The filled dots indicate the skill significantly higher than 0 at 1%significance level based on t-test with a Fisher-z transform for AR PC1 (blue), IPO (red) and NINO3.4 (black). The effective degrees of freedom is defined as $N\frac{1-\rho(1)}{1+\rho(1)}$, where N=23 years and ρ is the lag-1 autocorrelation of a given time series. The ρ is seasonally stratified.

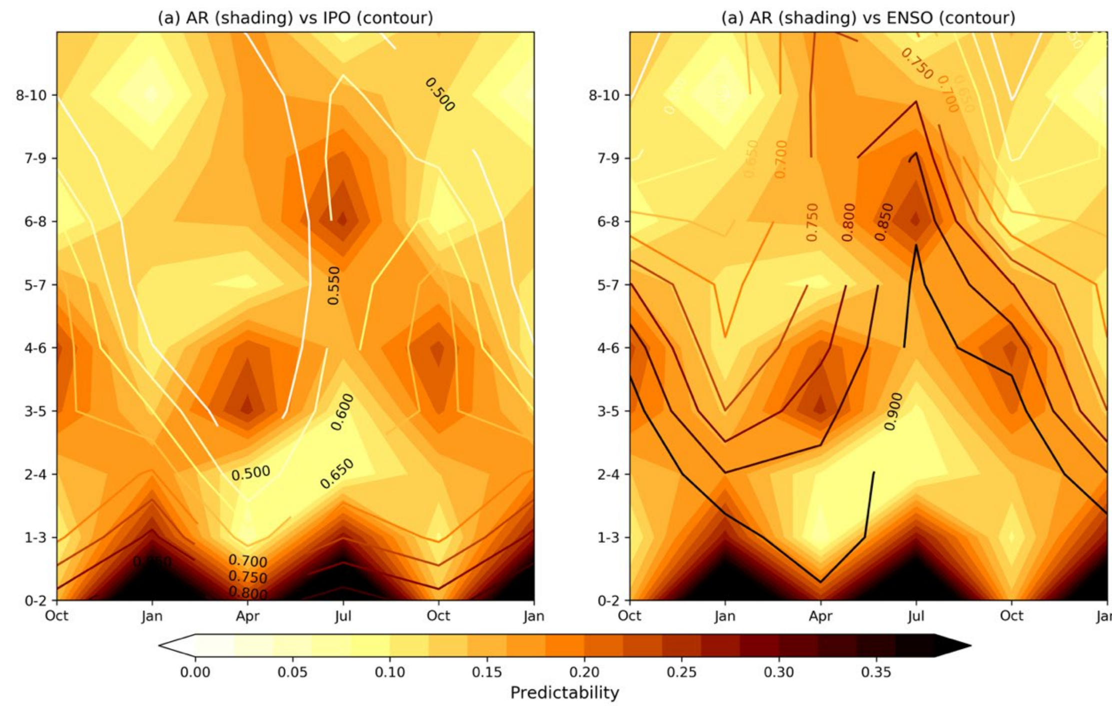

Figure S8. The predictability, as defined in equation (S2) of AR PC1 (shading), IPO (contour in a) and ENSO (contour in b). x-axis represents initialization season and y-axis represents forecast lead (months).

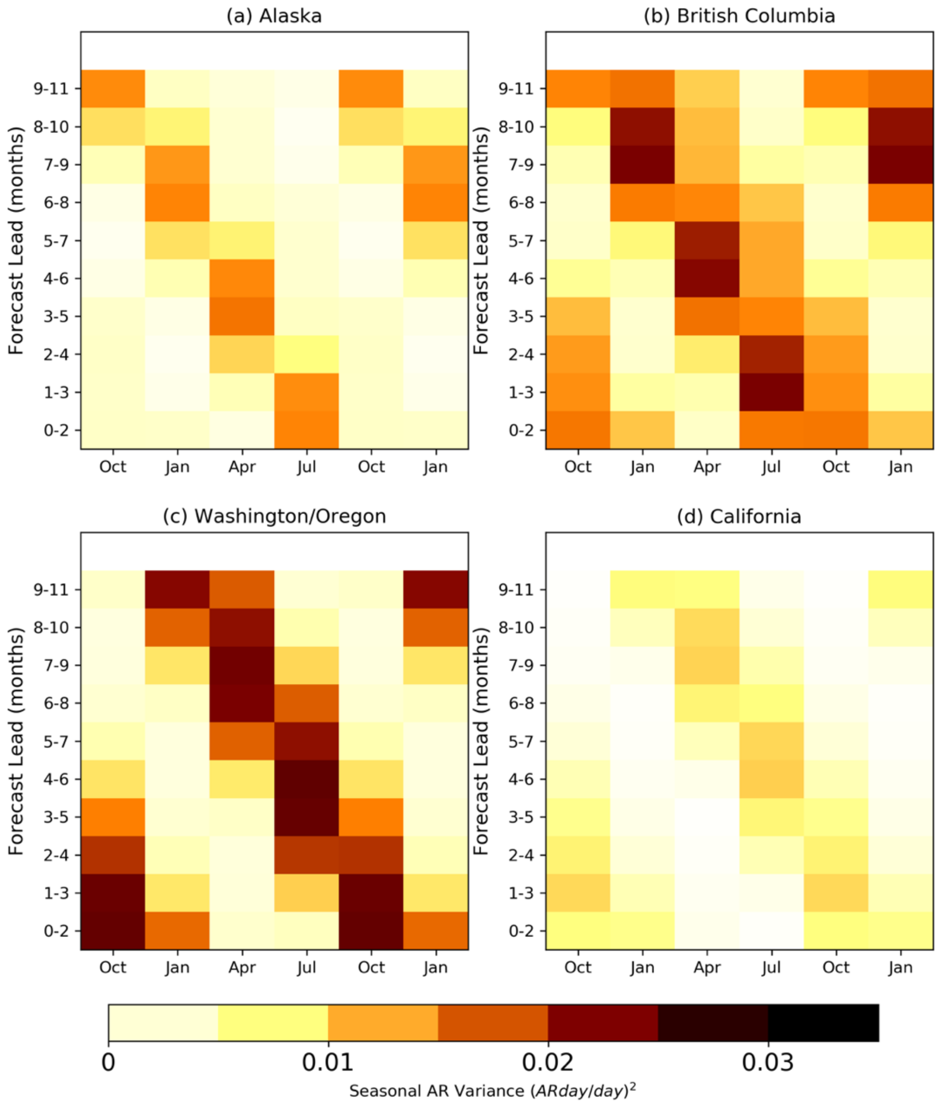

Figure S9. The year-to-year variance of seasonal AR anomalies from ERA5 ((AR day/day)^2) for the (a) Alaska, (b) British Columbia, (c) Washington/Oregon, and (d) California regions, shown as a function of initialization month (x-axis) and forecast lead (y-axis) for comparison with Fig.3. The boundaries of each region are outlined on the left side of Fig. 3.

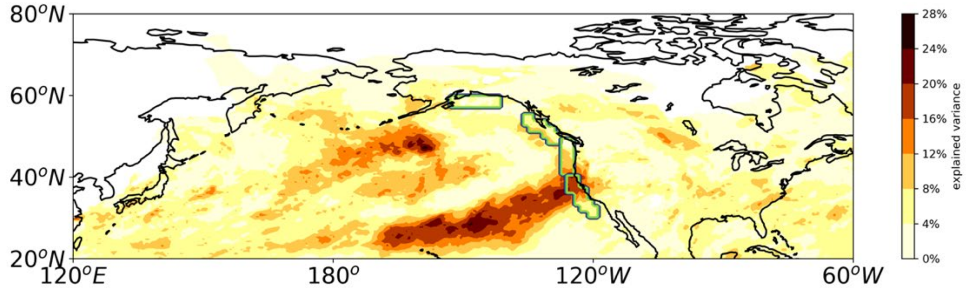

Figure S10. The fraction of local seasonal AR variance explained by EOF1 (%) during January through March in ERA5 reanalysis.

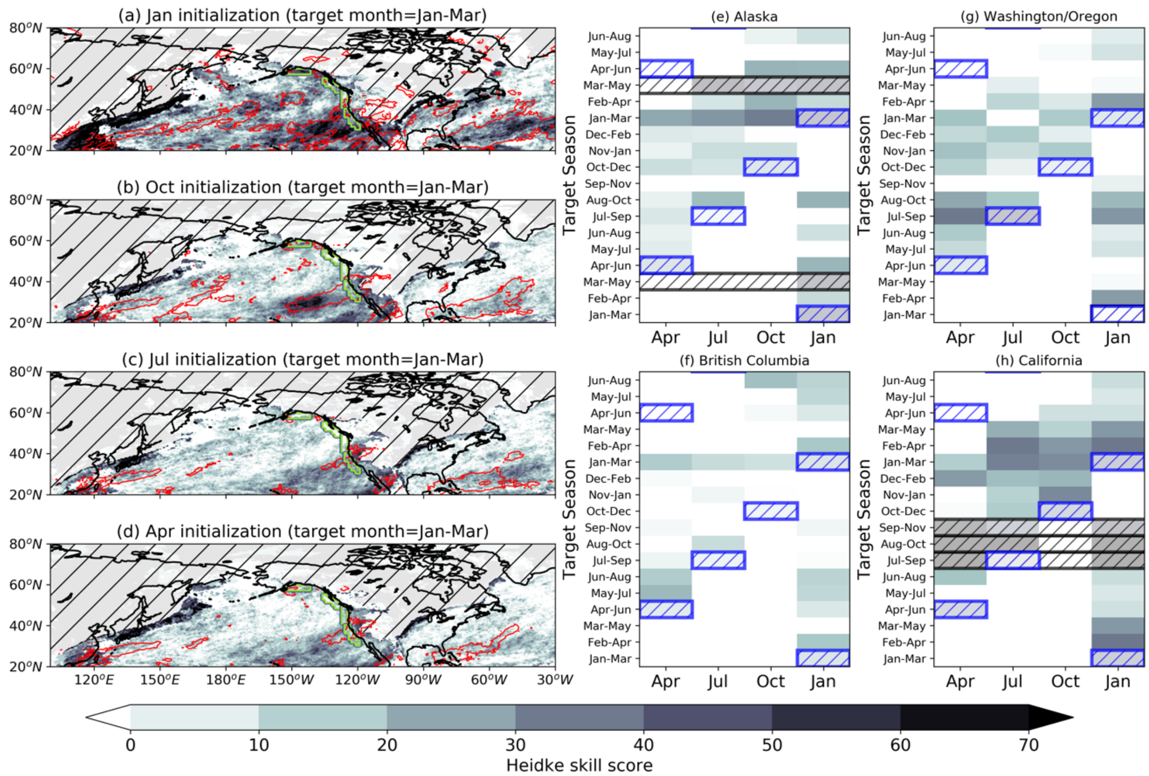

Figure S11. The Heidke Skill Scores of SPEAR calibrated probabilistic forecasts with the target season of January-March for all initializations. Because the regions with few (or no) AR events can lead to uninterpretable HSS values (see SI text), we mask those regions in Fig. S9. Black hatching represents regions where the frequency of AR climatology is less than 2% (AR day/day) in both SPEAR and ERA5. Blue hatching is the time of initialization. The red contours are the regions with ACC ≥ 0.5.