

JGR Atmospheres

RESEARCH ARTICLE

10.1029/2019JD031200

Special Section:

Bridging Weather and Climate: Subseasonal-to-Seasonal (S2S) Prediction

Key Points:

- A multimodel subseasonal-to-seasonal (S2S) hindcast skill assessment of atmospheric river (AR) activity over the western United States is presented
- The modulation of hindcast skill conditioned on phases of active MJO and ENSO events is quantified
- An experimental S2S AR activity forecast product designed with California Department of Water Resources stakeholders is introduced

Supporting Information:

- Supporting Information S1

Correspondence to:

M. J. DeFlorio,
mdeflorio@ucsd.edu

Citation:

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, 11,242–11,265. <https://doi.org/10.1029/2019JD031200>

Received 19 JUN 2019











Accepted 11 OCT 2019

Accepted article online 22 OCT 2019

Published online 13 NOV 2019

©2019. American Geophysical Union.
All Rights Reserved.

Experimental Subseasonal-to-Seasonal (S2S) Forecasting of Atmospheric Rivers Over the Western United States

Michael J. DeFlorio¹ , Duane E. Waliser^{2,3} , F. Martin Ralph¹ , Bin Guan^{3,2} , Alexander Goodman², Peter B. Gibson², Shakeel Asharaf^{3,2} , Luca Delle Monache¹ , Zhenhai Zhang¹ , Aneesh C. Subramanian⁴ , Frederic Vitart⁵, Hai Lin⁶ , and Arun Kumar⁷ 

¹Center for Western Weather and Water Extremes, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, USA, ²Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, ³Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, CA, USA, ⁴University of Colorado Boulder, Boulder, CO, USA, ⁵European Centre for Medium-Range Weather Forecasts, Reading, UK, ⁶Environment and Climate Change Canada, Montreal, Québec, Canada, ⁷Climate Prediction Center, NCEP/NWS/NOAA, College Park, MD, USA

Abstract A multimodel evaluation of subseasonal-to-seasonal (S2S) hindcast skill of atmospheric rivers (ARs) out to 4-week lead over the western United States is presented for three operational hindcast systems: European Centre for Medium-Range Weather Forecasts (ECMWF; Europe), National Centers for Environmental Prediction (NCEP; U.S.), and Environment and Canada Climate Change (ECCC; Canada). Ensemble mean biases and Brier Skill Scores are examined for no, moderate, and high levels of AR activity (0, 1–2, and 3–7 AR days/week, respectively). All hindcast systems are more skillful in predicting no and high AR activity relative to moderate activity. There are isolated regions of skill at week-3 over 150–125°W, 25–35°N for the no and high AR activity levels, with larger magnitude and spatial extent of the skill in ECMWF and ECCC compared to NCEP. The spatial pattern of this skill suggests that for high AR activity, a southwest-to-northeast orientation is more predictable at subseasonal lead times than other orientations, and for no AR activity, more skill exists in the subtropical North Pacific, upstream of central and southern California. AR hindcast skill along the western U.S. is most strongly increased in hindcasts initialized during Madden-Julian Oscillation (MJO) Phases 1 and 8, and hindcast skill is substantially decreased over California in hindcasts initialized during MJO Phase 4. Skill modulations in the ECMWF hindcast system conditioned on El Niño-Southern Oscillation phase are weaker than those conditioned on particular MJO phases. This work provides hindcast skill benchmarks and uncertainty quantification for experimental real-time forecasts of AR activity during winters 2019–2021 as part of the S2S Prediction Project Real-time Pilot Initiative in collaboration with the California Department of Water Resources.

1. Introduction

Atmospheric rivers (ARs) are elongated regions of intense horizontal water vapor transport, which are often associated with a baroclinic midlatitude cyclone (Neiman et al., 2008; Ralph et al., 2004; Ralph et al., 2017; Zhang et al., 2019; Zhu & Newell, 1998). Many studies have revealed the important link between ARs and annual precipitation over different regions around the globe (Dettinger et al., 2011; Lavers et al., 2011; Ralph & Dettinger 2012; Neiman et al., 2013; Lavers & Villarini, 2015; Lamjiri et al., 2018; and others), their association with global floods and water availability (Ralph et al., 2006; Leung & Qian, 2009; Ralph & Dettinger 2011; Paltan et al., 2017; Corringham, 2018), their relationship to snowpack over the western U. S. (Goldenson et al., 2018; Guan et al., 2012; Huning et al., 2017, 2019), their importance to extratropical climate and hydrology (Gorodetskaya et al., 2014; Nash et al., 2018), and their projected changes in a warming climate (e.g., Espinoza et al., 2018; Guirguis et al., 2018). In addition, a new AR scale has recently been introduced which groups ARs into categories based on their intensity, duration, and beneficial or hazardous impacts on society (Ralph et al., 2019).

In recent years, the far-reaching impacts and the importance of predicting AR occurrence and intensity have expanded beyond the 7–10-day weather forecasting time horizon and into subseasonal-to-seasonal (S2S) regime. This hybrid weather-climate scale of interaction encompasses lead times of 2 weeks to 2 years and

Table 1
Model Characteristics for the Hindcast Systems Used in This Study

Model	Ensemble size	Frequency	Period covered	Total sample size
ECMWF	11	Twice/week	1996–2015	880
NCEP	4	Daily	1999–2010	1,812
ECCC	4	Weekly	1995–2014	440

Note. All data sets are regridded to a $1^\circ \times 1^\circ$ horizontal grid using a nearest-neighbor interpolation.

has been recognized by the academic community as a forecast period requiring additional research and process understanding (NAS S2S, 2016; Robertson et al., 2015; Robertson & Vitart, 2018; Vitart et al., 2017) for phenomenon such as ARs (Baggett et al., 2017; DeFlorio et al., 2018; DeFlorio et al., 2019; Mundhenk et al., 2018; Nardi et al., 2018), cold spells (Ferranti et al., 2018), heat waves (Ford et al., 2018), precipitation (Pan et al., 2019; Wang et al., 2017), atmospheric ridging (Gibson et al., 2019), tropical cyclones (Jiang et al., 2018; Lee et al., 2018), severe weather activity (Baggett et al., 2018), and others. Contemporaneously, forecasts at these lead times are in demand from various sectors including water resource management, agriculture and hazard insurance, and energy supply

(NAS S2S, 2016).

There has been considerable attention in recent work to quantify prediction skill of ARs and extreme precipitation for both short-term and S2S lead times over the western U.S. region. Pan et al. (2019) conducted a multimodel analysis of operational hindcast prediction skill (from daily to extended range lead times) of precipitation in western U.S. coastal regions using data from the S2S Project database (Vitart et al., 2017) and quantified the modulation of prediction skill on particular phases of the Madden-Julian Oscillation (MJO) and El Niño-Southern Oscillation (ENSO). Other studies over the past decade have focused on prediction skill of ARs, since it has been shown that the potential predictability of integrated vapor transport (IVT), a defining characteristic of ARs, is higher than that of the resultant precipitation at all lead times from 1 to 15 days (Lavers et al., 2016). These recent studies have investigated both shorter-range prediction skill of ARs on weather timescales out to 14-day lead time (e.g., DeFlorio et al., 2018; Nardi et al., 2018; Wick et al., 2013) and longer S2S timescales focusing on prediction skill of AR activity relative to that of a reference climatology at lead times ranging from 3 to 5 weeks over the North Pacific and western U.S. regions (Baggett et al., 2017; DeFlorio et al., 2019; Mundhenk et al., 2018).

The work presented here builds on these previous studies by providing hindcast skill benchmarks for AR activity at 1–4-week lead time over the western U.S. using uncalibrated ensemble hindcasts from three operational hindcast systems in the S2S Project database. Three levels of AR activity are defined: *0 AR days/week*, no AR activity; *1–2 AR days/week*, moderate AR activity; and *3–7 AR days/week*, high AR activity.

Stakeholders in various communities (water resource management, agriculture, insurance, and many others) are interested in receiving information from the research community regarding *forecasts of opportunity* and *forecasts of avoidance*, which are periods of time when predictions of quantities of interest are more or less skillful, respectively, than normal. This work identifies forecasts of opportunity for particular phases of the MJO and ENSO during which prediction skill of AR activity is improved. Information regarding AR activity at S2S lead times is highly valuable to water resource managers across the state of California and the western U.S. As a direct response to this demand from stakeholders, experimental near-real-time forecasts of S2S AR activity have been produced by the Center for Western Weather and Water Extremes (CW3E) during winter 2017–2018 and 2018–2019. The hindcast skill assessment provided in this study is essential in providing baseline skill estimates, which can be used as a reference for future forecasts of AR activity during winter 2019–2020 and 2020–2021 as part of a near-real-time pilot under the S2S Prediction Project.

2. Materials and Methods

2.1. Hindcast Systems and Reanalysis Data

To assess S2S prediction skill of AR activity, global ensemble hindcast data from three operational centers participating in the World Climate Research Program/World Weather Research Program's S2S Project database are used (Vitart et al., 2017). The three operational centers that produced the data used in this study are the European Centre for Medium-Range Weather Forecasts (ECMWF), the National Centers for Environmental Prediction (NCEP), and the Environment and Canada Climate Change (ECCC). Near-real-time data from these centers have been obtained for the production of experimental S2S AR forecasts. The main characteristics of each hindcast system (ensemble size, hindcast frequency, and period covered) are described in Table 1; more detail on the ECMWF hindcast system is described in section 2.1 of DeFlorio

et al. (2018, 2019). The ECMWF Reanalysis-Interim (ERA-I) data set (Dee et al., 2011), spanning the period 1996–2015, is used as a reference to verify the hindcast skill of AR activity. To evaluate each hindcast system, the ERA-I data are subsetted to match the period of years covered and the hindcast frequency of each hindcast system. The sensitivity of the ensemble mean bias and Brier Skill Score (BSS) calculations to the choice of the reference data set is examined in Figures S1–S3 in the supporting information using the NCEP Climate Forecast System Reanalysis (Saha et al., 2010) as an alternative reanalysis data set, which spans the period 1979–2009.

The following steps are common for producing AR activity fields for all data sets:

1. Retrieve specific humidity (q) and zonal (U) and meridional (V) wind components at 00Z at the 300-, 500-, 700-, 850-, and 1,000-hPa vertical levels.
2. Compute IVT, which is defined as the vertically integrated product of specific humidity and horizontal wind (Zhu & Newell, 1998).
3. Apply the Guan and Waliser (2015) AR detection algorithm, with the refinements in Guan et al. (2018), to the IVT fields for each hindcast system, and for the contemporaneous reanalysis data set used to verify the hindcasts. AR objects are detected based on the seasonal 85th percentile threshold of IVT magnitude from the ERA-I data set at each grid cell and based on geometric constraints described in Guan and Waliser (2015).
4. Aggregate the number of AR days per week for each individual ensemble member from a given hindcast system as a function of lead time for each forecast.

We restrict our analysis to cover the November–March (NDJFM) period, when AR activity and predictive information for water resources management is of most interest over the western U.S. region.

2.2. Regridding Procedure

To enable a comparison of multimodel hindcast skill assessment, all AR activity fields (following Step 4 in section 2.1) are regridded to a common $1^\circ \times 1^\circ$ horizontal grid using the *nearestod* (i.e., nearest source to destination method; see, e.g., Valcke et al., 2012) nearest-neighbor algorithm in NCAR Command Language. This method ensures that AR activity values are retained as integer values, as opposed to, for example, bilinear interpolation, which produces some noninteger values. The retention of integer values is desirable because AR activity is defined based on integer values (see section 2.3). A common $1^\circ \times 1^\circ$ grid is chosen to facilitate a comparison between these hindcast skill metrics and near-real-time forecast data from the three modeling centers in future work.

2.3. AR activity Levels

Three levels of AR activity are defined and used in this study based on the number of AR days per week (AR1wk):

AR1wk AR activity levels

“0 AR days/week”= no AR activity

“1–2 AR days/week”= moderate AR activity

“3–7 AR days/week”= high AR activity

It is useful to group AR activity into levels of no, moderate, and high activity from the perspective of a water resource manager/stakeholder in the western U.S. region. In addition, the binning of AR activity in this fashion allows for the calculation of several skill metrics used in this study and described in the next section.

2.4. Forecast Assessment and Skill Metrics

To assess S2S AR hindcast prediction skill over the western United States region, the following assessment and skill metric approaches are implemented:

1. Ensemble mean model bias, which evaluates the systematic error of each hindcast system in predicting AR activity.
2. Debaised BSS (Weigel et al., 2007), which measures the accuracy of each hindcast system in predicting AR activity.

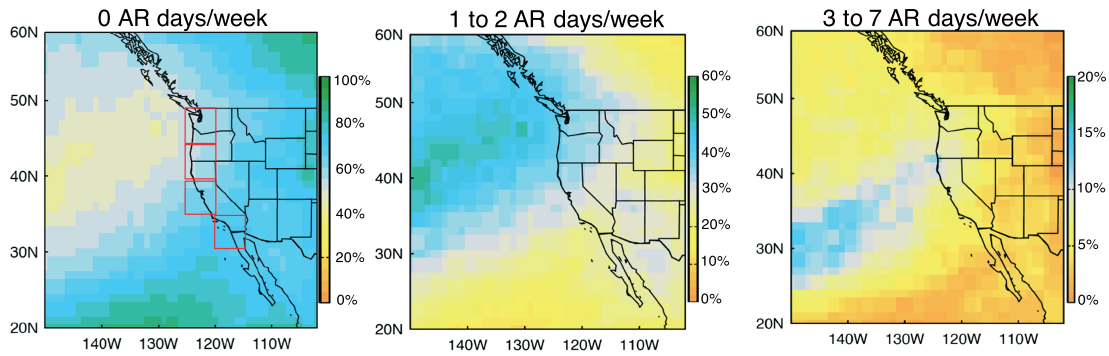


Figure 1. ERA-I November–March (NDJFM) 1996–2015 average number of AR days per week (AR1wk) for each of the AR activity levels used to assess S2S hindcast skill in this study. Color bar ranges are unique for each activity level. Red boxes indicate coastal regions where area-weighted average biases are computed (see Table 2), and where hits, misses, false alarms, and correct rejections of AR activity are computed in Figures 9–11. The sum of the three activity levels at each grid cell equals 100%.

3. Relative Operating Characteristic (ROC)-like diagrams, that is, ensemble mean hit rate and false alarm rate for the lead-dependent occurrence of an AR anywhere inside four western U.S. landfall regions defined in Figure 1a, depicted in more detail in Figure 2 and described in sections 2.5 and 2.7, which measures the ability of each hindcast system to discriminate between the occurrence or nonoccurrence of AR activity.

The lead-time-dependent model bias is defined as the difference between the forecast and observed percentages of the weeks that fall into a particular level of AR activity. The values reported are absolute percent differences between the forecast and observed data. Forecast AR events are calculated for each individual ensemble member; then the ensemble mean of AR events is computed.

The BSS is a relative measure of probabilistic skill, which is based on the Brier Score (BS; Wilks, 2011) that measures the mean squared error of probability forecasts for a binary event. However, it has been found that the BSS is sensitive (negatively biased) for small ensemble sizes (Müller et al., 2005). To overcome this issue, Weigel et al. (2007) added a correction term in the denominator of the BSS, which is reciprocal to the

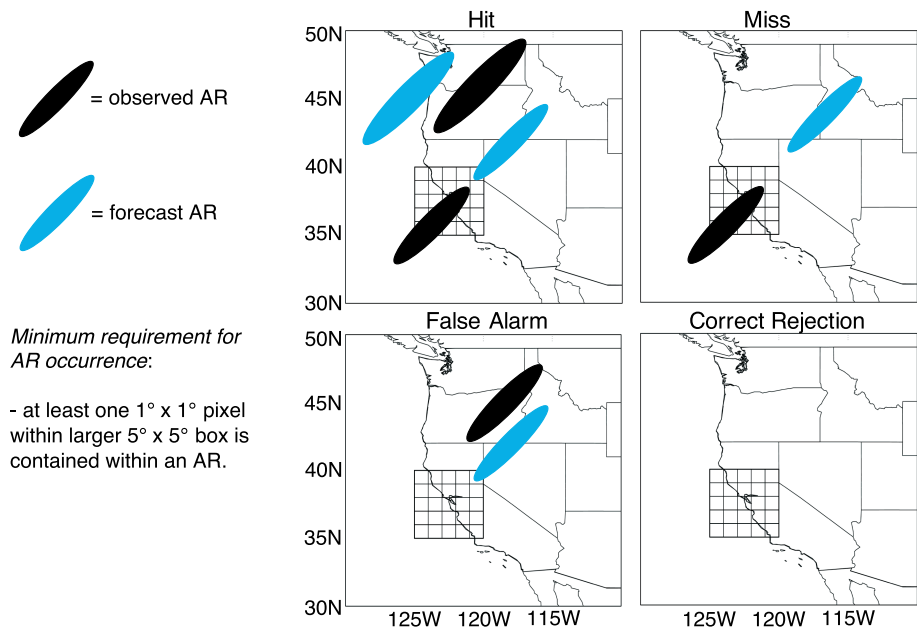


Figure 2. Schematic of Relative Operating Characteristic (ROC) statistics of AR activity (hit, miss, false alarm, and correct rejection) calculated in Figures 9–11 and described in sections 2.5, 2.7, and 4.1 of this study.

ensemble size and is thus weighted more strongly for hindcast systems with small ensemble size. More specifically, the debiased BSS is defined for each AR activity level as

$$\text{BSS} = 1 - \frac{\text{BS}}{\text{BS}_{\text{ref}} + D} \quad (1)$$

BS and BS_{ref} are the BSs for the hindcast and based on the reference climatology, respectively, and are defined as

$$\text{BS} = \frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2 \quad (2)$$

$$\text{BS}_{\text{ref}} = \frac{1}{N} \sum_{i=1}^N (P_{\text{clim}} - O_i)^2 \quad (3)$$

Following equation 16 of Weigel et al. (2007), the correction term D is defined as

$$D = \frac{1}{M} P_{\text{clim}}(1 - P_{\text{clim}})$$

In the above equations, N is the number of hindcast samples during all winter months (880 for ECMWF, 1,812 for NCEP, and 440 for ECCO, as reported in Table 1); P_i is the forecast probability of a particular level of AR activity for a single hindcast event, for example, the fraction of ensemble members, which predict AR activity falling into a particular category during a week-long lead period for a single hindcast; O_i is a binary representation of whether the reanalysis' AR activity fell into that specific level (1 if yes, 0 if no); M is the ensemble size; and P_{clim} is the probability of the reference climatology, defined as the long-term fraction of weeks during boreal winter (considering only NDJFM days) that fell into that specific AR activity level over the ERA-I period that is contemporaneous with each hindcast system. P_{clim} is similar to the result shown in Figure 1, but the period and sampling frequency of ERA-I are chosen to match each individual hindcast system.

The BSS values range from $-\infty$ to 1, and the interpretation of these values is as follows:

BSS < 0 hindcast skill is lower than skill from a forecast made using reference climatology.

BSS = 0 no skill in hindcast system; that is, hindcast skill is equal to skill from a forecast made using reference climatology.

0 < BSS < 1 hindcast skill is higher than skill from a forecast made using reference climatology.

BSS = 1 hindcast system is perfect.

In this analysis, the average BSS over all hindcast samples in the NDJFM hindcast record (and the contemporaneous reanalysis data) is calculated for each of the AR activity levels defined in section 2.3. As described in equations (1)–(4), the BSS value for each activity level is sensitive to (a) the squared difference between the AR activity hindcast probability and AR activity in the reanalysis data and (b) the squared difference between the reference climatology (the percentage of NDJFM weeks that fell into that specific AR activity level over the entire hindcast record) and reanalysis data.

2.5. ROC-Like Diagrams

ROC-like diagrams are created and described in sections 2.7 and 4.1 to compare ensemble mean hit and false alarm rates for four $5^\circ \times 5^\circ$ regions along the western U.S. coastline, outlined in red in Figure 1. The term *ROC-like* is used in this study to distinguish these diagrams from traditional ROC curves (Hanley & McNeil, 1982), which implement probabilistic trigger thresholds above which decision makers might issue warnings or take mitigating action for a potentially harmful event. A schematic depicting examples of a

hit, miss, false alarm, and correct rejection is shown in Figure 2 over the Central California $5^\circ \times 5^\circ$ box region.

The $5^\circ \times 5^\circ$ regions along the western U.S. coastline used to create these ROC-like diagrams are similar to those used in Pan et al. (2019), which examined precipitation prediction skill at both short and extended range lead times in the S2S Project database. For a given hindcast system, the occurrence of an AR at any grid cell within the $5^\circ \times 5^\circ$ region at any time during a given week-long lead window is considered for each ensemble member. This expanded $5^\circ \times 5^\circ$ domain is used so that penalties on the precise landfall location of ARs are reduced (similar to Ebert & Gallus, 2009). Considering forecasts over broader spatial domains in this fashion is advisable for longer lead prediction, as previous work has shown that the hit rate of predicting wintertime AR occurrence at 7-day lead time over the North Pacific/Western U.S. region drops by 30–35% when reducing the allowable error in predicted AR centroid location from 1,000 to 250 km (DeFlorio et al., 2018, Figure 4). Ensemble mean hit rates (defined as the total number of hits, normalized by the total number of hits + misses) and false alarm rates (defined as the total number of false alarms, normalized by the total number of false alarms + correct rejections) are then calculated.

The terms comprising these rates are defined as follows:

Hit = AR predicted in at least one grid cell within $5^\circ \times 5^\circ$ box, and AR occurred in at least one grid cell within this box.

Miss = AR not predicted in any grid cell within $5^\circ \times 5^\circ$ box, but AR occurred in at least one grid cell within this box.

False alarm = AR predicted in at least one grid cell within $5^\circ \times 5^\circ$ box, but AR did not occur in any grid cell within this box.

Correct rejection = AR not predicted in any grid cell within $5^\circ \times 5^\circ$ box, and AR did not occur in any grid cell within this box.

2.6. Climate Mode Composite Analysis

The sensitivity of the hit and false alarm rates for each region/lead time pair to the phase and magnitude of both the MJO and ENSO are evaluated in sections 4.1–4.2. Daily values of the Real-time Multivariate MJO (RMM) index are obtained from the Centre for Australian Weather and Climate Research. The RMM index is derived from a combined empirical orthogonal function analysis of equatorial 850-hPa zonal wind, 200-hPa zonal wind, and outgoing longwave radiation data (Wheeler & Hendon, 2004). Its values represent the phase and amplitude of the MJO and define its evolution as an eight-phase cycle based on large-scale anomalies of convection, which drive anomalies in vertical wind shear. Active MJO cases are defined as days when the magnitude of RMM is >1 . Monthly values of the Niño3.4 (5°S – 5°N , 170°W – 120°W) index are obtained from the National Oceanic and Atmospheric Administration Extended Reconstructed Sea Surface Temperature version 5 product (Huang et al., 2017). Positive and negative phases of ENSO (El Niño and La Niña, respectively) are defined as periods during which the magnitude of the 3-month running mean of the Niño3.4 index exceeds 0.5°C for a minimum of five consecutive overlapping 3-month running mean periods. The initial date of each hindcast is matched to the center month of the 3-month running mean period to perform the composite analysis described in section 4.1.2. It is noted that although the MJO RMM index is real time and can be used in a real-time forecasting environment, the Niño3.4 index is not, since ENSO conditions may exist for several months before El Niño or La Niña conditions are declared.

2.7. $5^\circ \times 5^\circ$ Western U.S. Coastline Regions Over Which Skill Is Assessed in This Study

In this work, S2S hindcast prediction skill of AR activity is made at both the $1^\circ \times 1^\circ$ grid cell native resolution of the regridded hindcast systems and over larger spatial domains in four $5^\circ \times 5^\circ$ boxes spanning the western U.S. coastline, outlined in red in Figure 1 (left panel). The four boxes are referred to as Washington (125°W – 120°W , 45°N – 50°N), Oregon-Northern California (125°W – 120°W , 40°N – 45°N), Central California (125°W – 120°W , 35°N – 40°N), and Southern California (120°W – 115°W , 30°N – 35°N). The choice of longitude and latitude boundaries for these boxes was agreed upon in coordination with stakeholders from the California Department of Water Resources to represent useful AR landfall domains that span the entire western U.S. coastline. They are also similar to coastal boxes that have been employed in previous similar studies (e.g., Nardi et al., 2018; Pan et al., 2019).

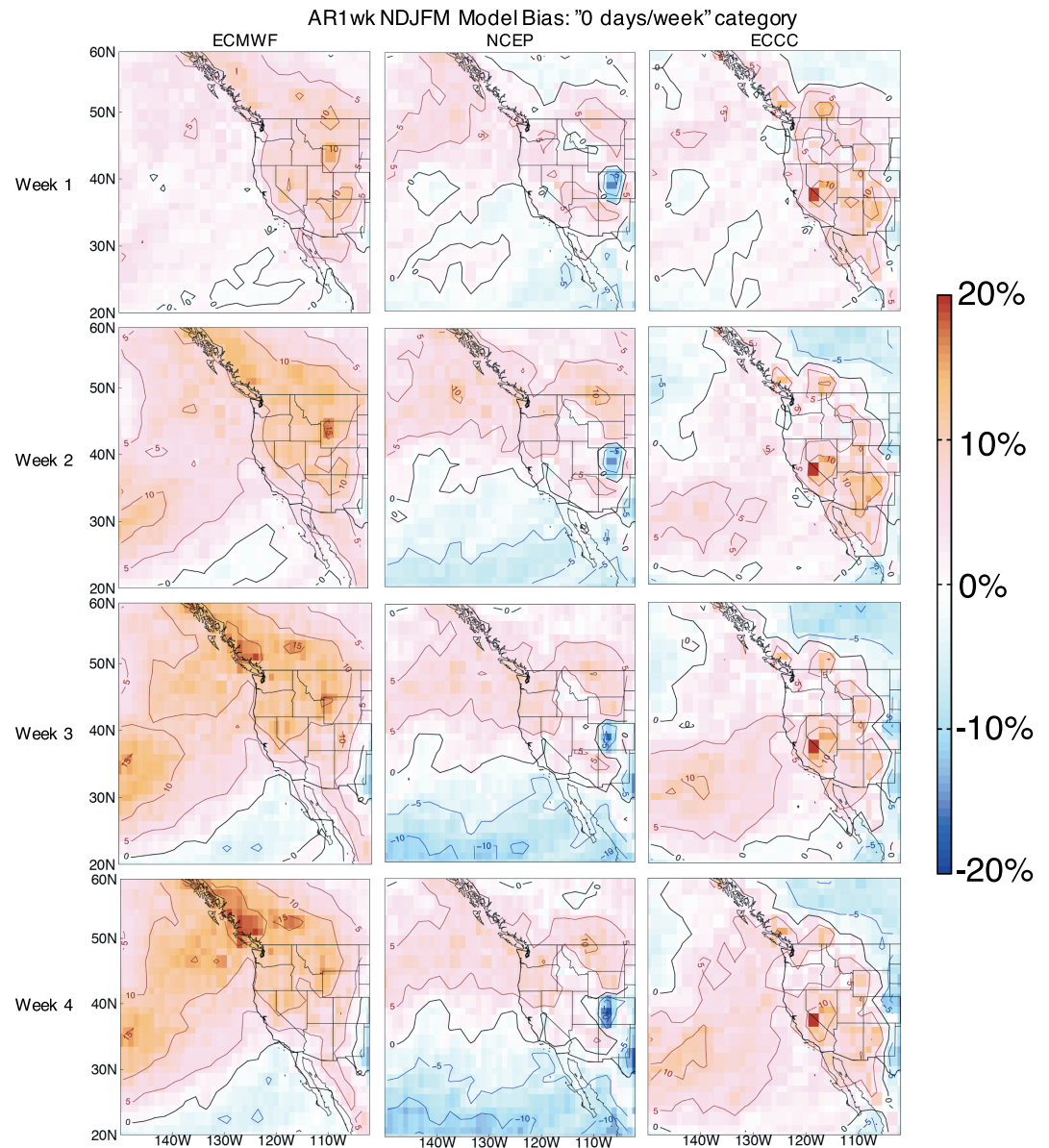


Figure 3. AR1wk NDJFM model bias (hindcast minus reanalysis; absolute percentage difference) in each hindcast system (columns) at each lead time (rows) for the 0 AR days/week activity level.

Each $5^\circ \times 5^\circ$ box contains 25 $1^\circ \times 1^\circ$ grid cells. For the analysis presented in Figures 3–8, area-weighted average biases and BSSs are computed over each box and are reported in Tables 2 and 3. For the analysis presented in Figures 9–11, the skill of each hindcast system in predicting AR occurrence at any grid cell within the larger $5^\circ \times 5^\circ$ box at any time during week-1, week-2, week-3, and week-4 is assessed. ROC-like diagrams are created for each box and are calculated according to the method outlined in section 2.5. An AR is counted as *occurring* inside a given $5^\circ \times 5^\circ$ box region if the criteria from the Guan and Waliser (2015) objective algorithm for AR occurrence is met within any of the $1^\circ \times 1^\circ$ pixels within the larger region (see Figure 2).

3. S2S Forecast Assessment at Grid Cell Level

S2S prediction of AR activity is evaluated in this section in a multimodel hindcast framework over each $1^\circ \times 1^\circ$ grid cell. It is emphasized that although differences between each hindcast system are described, readers

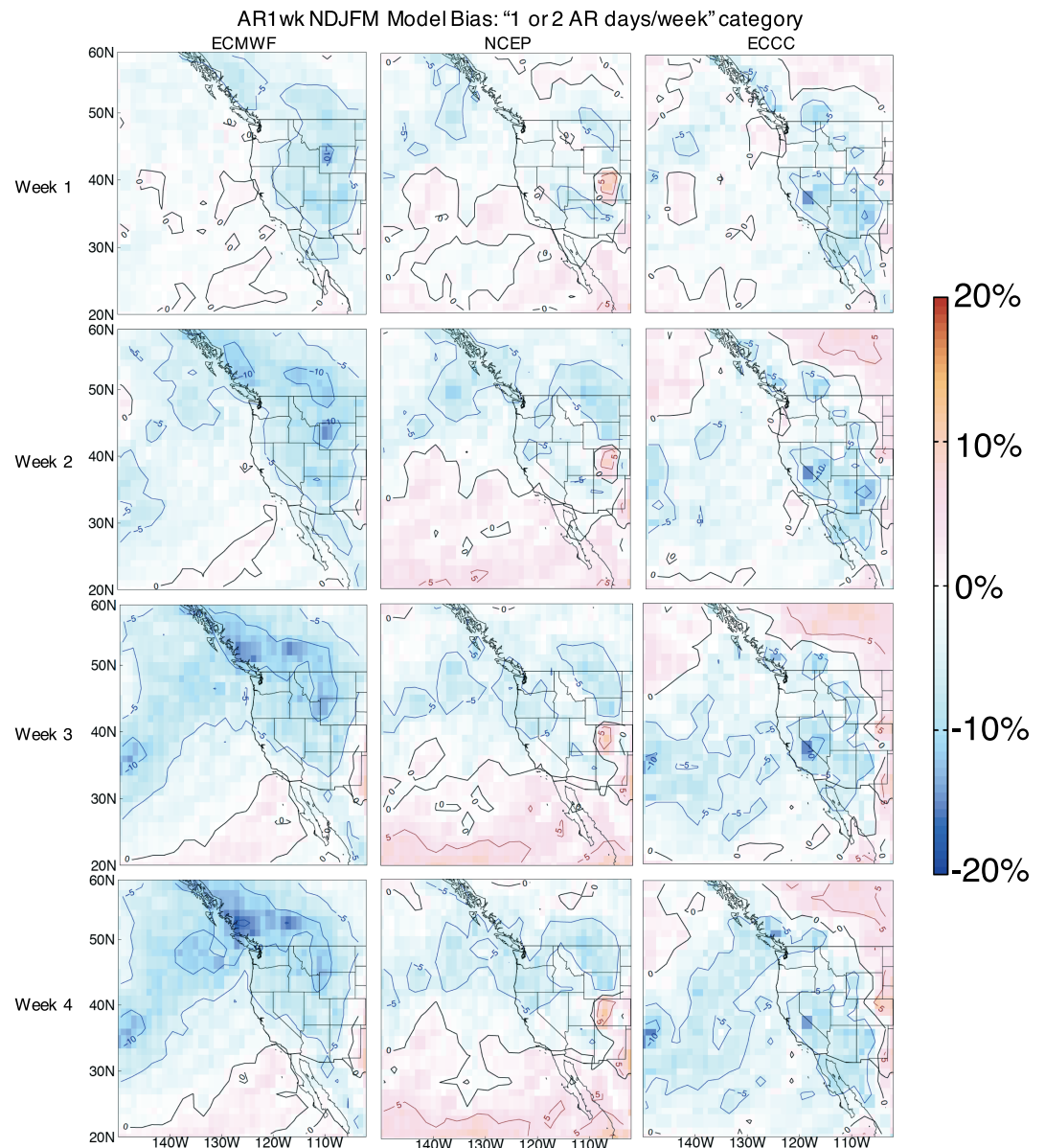


Figure 4. AR1wk NDJFM model bias (hindcast minus reanalysis; absolute percentage difference) in each hindcast system (columns) at each lead time (rows) for the 1-2 AR days/week activity level.

should bear in mind that the hindcast systems have varying ensemble sizes, hindcast frequencies, periods covered, and total number of hindcast samples, as described in Table 1. All available hindcast data from the S2S Project database are retained, but because of the differences between the design of these hindcast systems, these results should be interpreted not only as a forecast system intercomparison but also as a way to define baseline hindcast skill benchmarks that can be used as a reference for comparison against experimental real-time forecasts of S2S AR activity in future work.

3.1. AR Activity Climatology

The multimodel hindcast skill analysis of AR activity presented in this study is based on levels of AR activity over a week-long period. The average number of AR days per week (*AR1wk*) for each of the AR activity levels used to assess S2S hindcast skill in this study is the ERA-I NDJFM 1996–2015 data and is shown in Figure 1. The sum of the three activity levels at each grid cell equals 100%, by definition.

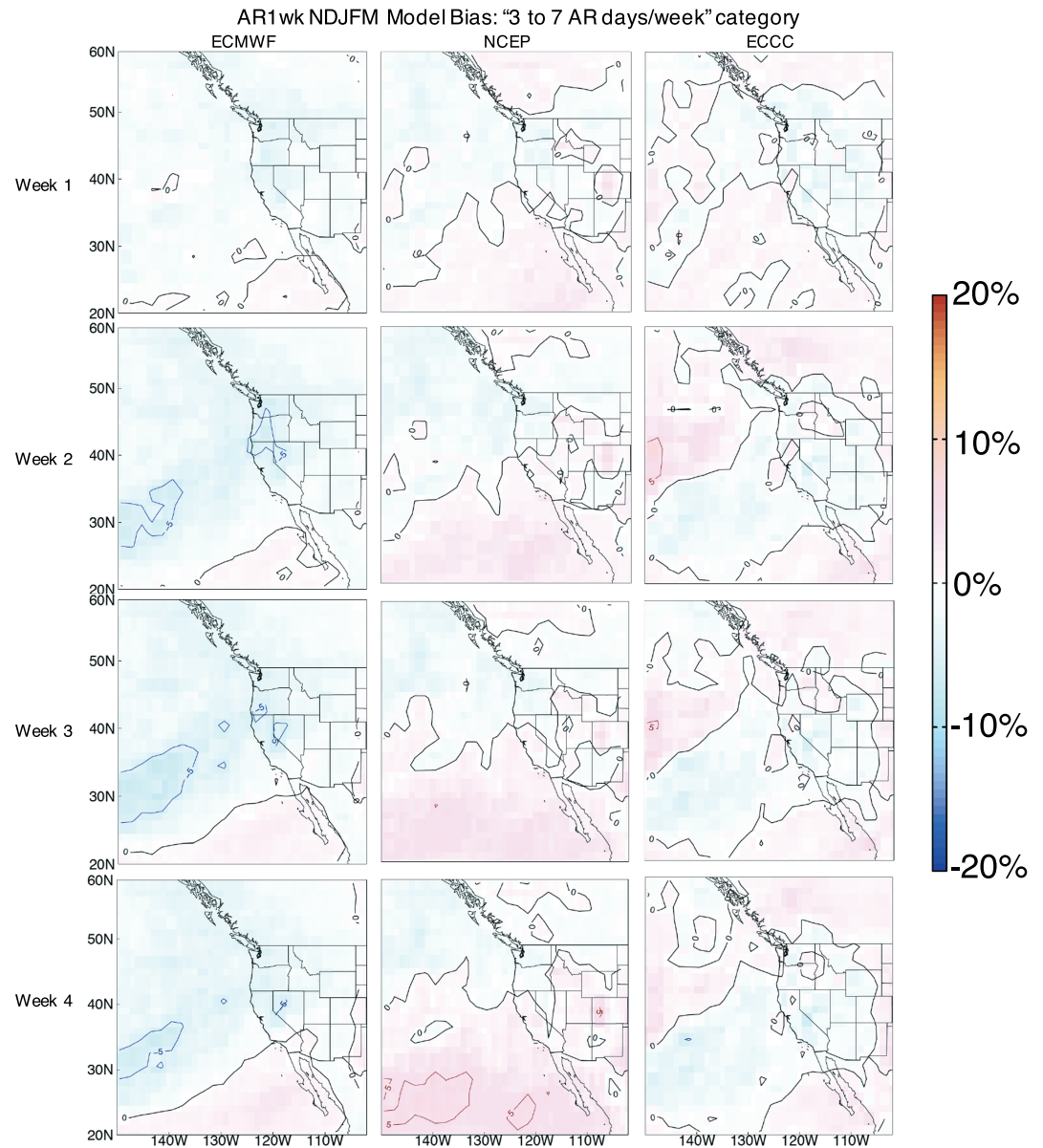


Figure 5. AR1wk NDJFM model bias (hindcast minus reanalysis; absolute percentage difference) in each hindcast system (columns) at each lead time (rows) for the 3-7 AR days/week activity level.

Minimum values for the 0 AR days/week activity level are located upstream of the Pacific Northwest coastline, centered around 145°W, 40°N. This region is collocated with a maximum in the percentage of weeks with 1–2 AR days/week. The percentage of weeks falling into the 0 AR days/week AR activity level generally increases with decreasing latitude along the western U.S. coastline, while the percentage of weeks falling into the 1–2 AR days/week activity level generally decreases with decreasing latitude. Interestingly, maximum values for the 3–7 AR days/week activity level are shifted southward to around 145°W (and extending to 125°W), 30°N, which is located approximately 10° south of the maximum in the 1–2 AR days/week activity level. The spatial structure of this high AR activity level suggests that the most active AR weeks in the climatological record are associated with southwest-to-northeast oriented AR regimes, with ARs that may propagate northeastward.

The sensitivity of AR activity climatology to the choice of reanalysis data set is explored in Figure S1, which shows the absolute percent difference between the ERA-I activity levels in Figure 1 and those computed

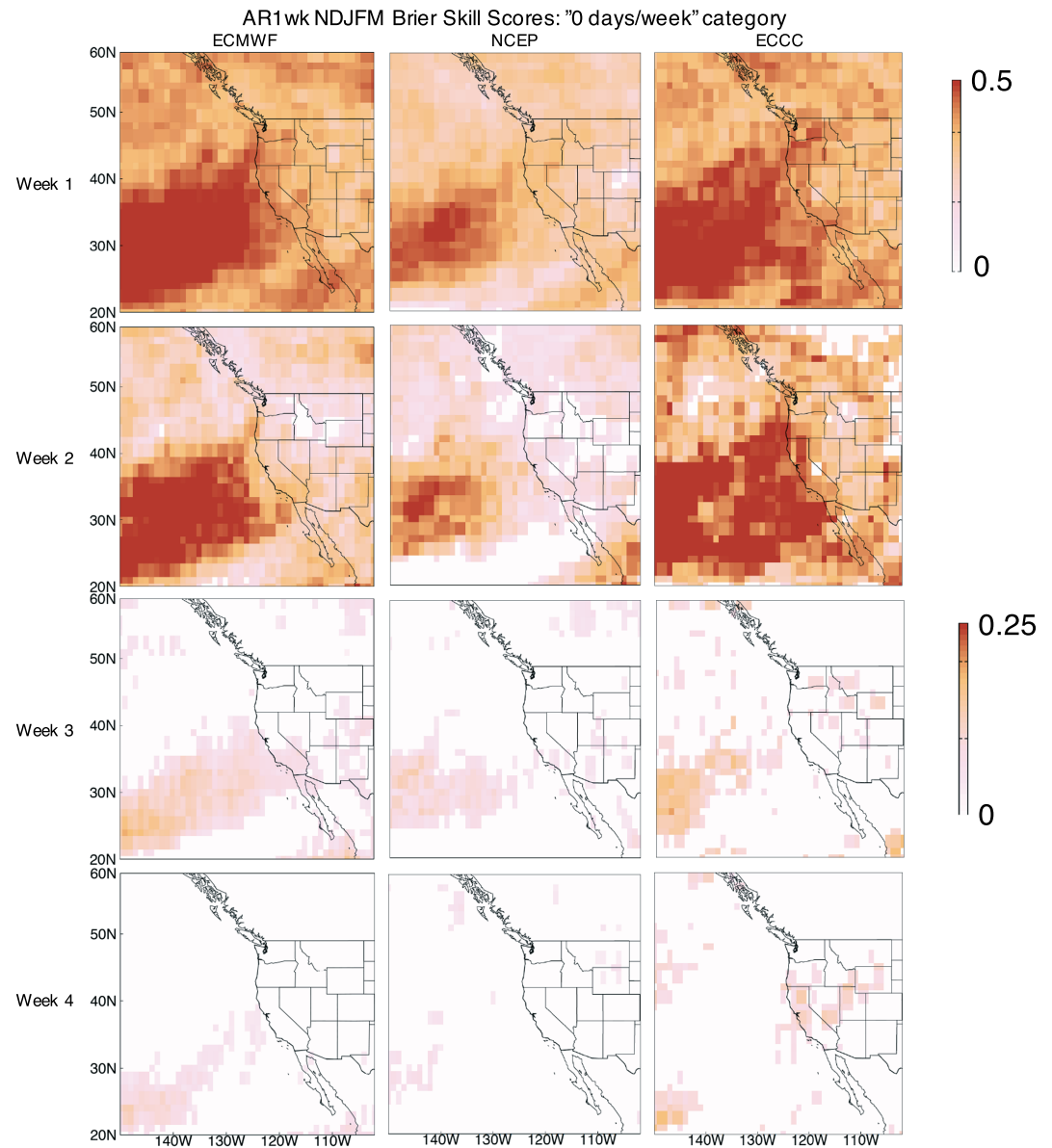


Figure 6. AR1wk NDJFM Brier Skill Score (BSS) values in each hindcast system (columns) at each lead time (rows) for the 0 AR days/week activity level. Values are only plotted if they are >0 (i.e., skillful with respect to ERA-I reference climatology) and if they are significant at the 95% level using a 100-sample bootstrapping procedure with replacement. The color bar range for week 1 (top row) is 0 to 0.5, while the color bar range for weeks 2–4 is 0 to 0.25.

using CFSR NDJFM data from 1979 to 2010. Absolute percent differences are generally less than 5% across all AR activity levels, except for a 10–15% increase (decrease) in the percentage of 0 (1–2) AR activity weeks over the Upper Colorado River Basin and Colorado Rocky Mountains regions. These differences do not project strongly onto differences in forecast skill based on the choice of reanalysis for computing a reference climatology (examples shown in Figures S2–S3 for BSS values of 0 AR days/week activity level at week-1 and week-3 leads, respectively).

3.2. AR Activity Hindcast System Bias

Hindcast system biases (hindcast system minus reanalysis) for the 0, 1–2, and 3–7 AR days/week activity levels are shown in Figures 3, 4, and 5, respectively, for each hindcast system (columns) and over the week-1 (1- to 7-day), week-2 (8- to 14-day), week-3 (15- to 21-day), and week-4 (22- to 28-day) lead windows (rows). Positive (negative) values represent an overprediction (underprediction) of weeks falling into each

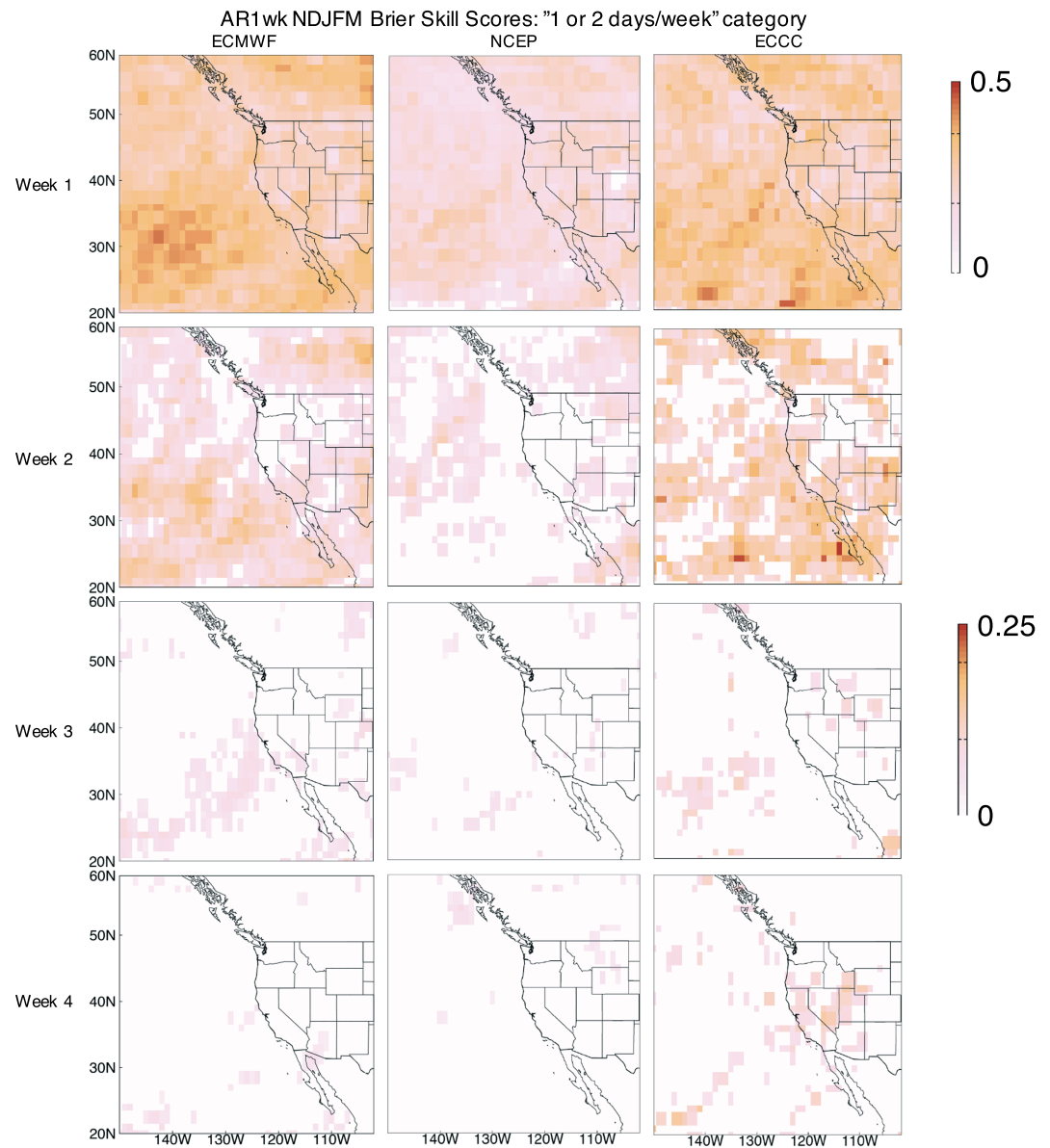


Figure 7. AR1wk NDJFM BSS values in each hindcast system (columns) at each lead time (rows) for the 1–2 AR days/week activity level. Values are only plotted if they are >0 (i.e., skillful with respect to ERA-I reference climatology) and if they are significant at the 95% level using a 100-sample bootstrapping procedure with replacement. The color bar range for week 1 (top row) is 0 to 0.5, while the color bar range for weeks 2–4 is 0 to 0.25.

particular activity level relative to the ERA-I reanalysis data. The area-weighted average biases over the $5^\circ \times 5^\circ$ regions shown in Figure 1 are reported in Table 2.

The ECMWF hindcast system generally overpredicts the percentage of weeks with no AR days by up to 15% (absolute percent difference) across the entire domain, with the highest values occurring over British Columbia, the Pacific Northwest, and the northern Rocky Mountains. These biases exhibit small increases with increasing lead time, except over the eastern subtropical North Pacific off the coast of Baja Mexico, where the bias changes from near neutral at 1–2-week lead time to weakly negative at 3–4-week lead time. Previous studies have found negative biases in AR occurrence in the ECMWF hindcast system over this general region at 1–4-week lead time (Nardi et al., 2018, Figure 3; DeFlorio et al., 2019, Figure 2), which is consistent with the overprediction of *no AR activity* weeks shown here. The NCEP hindcast system generally

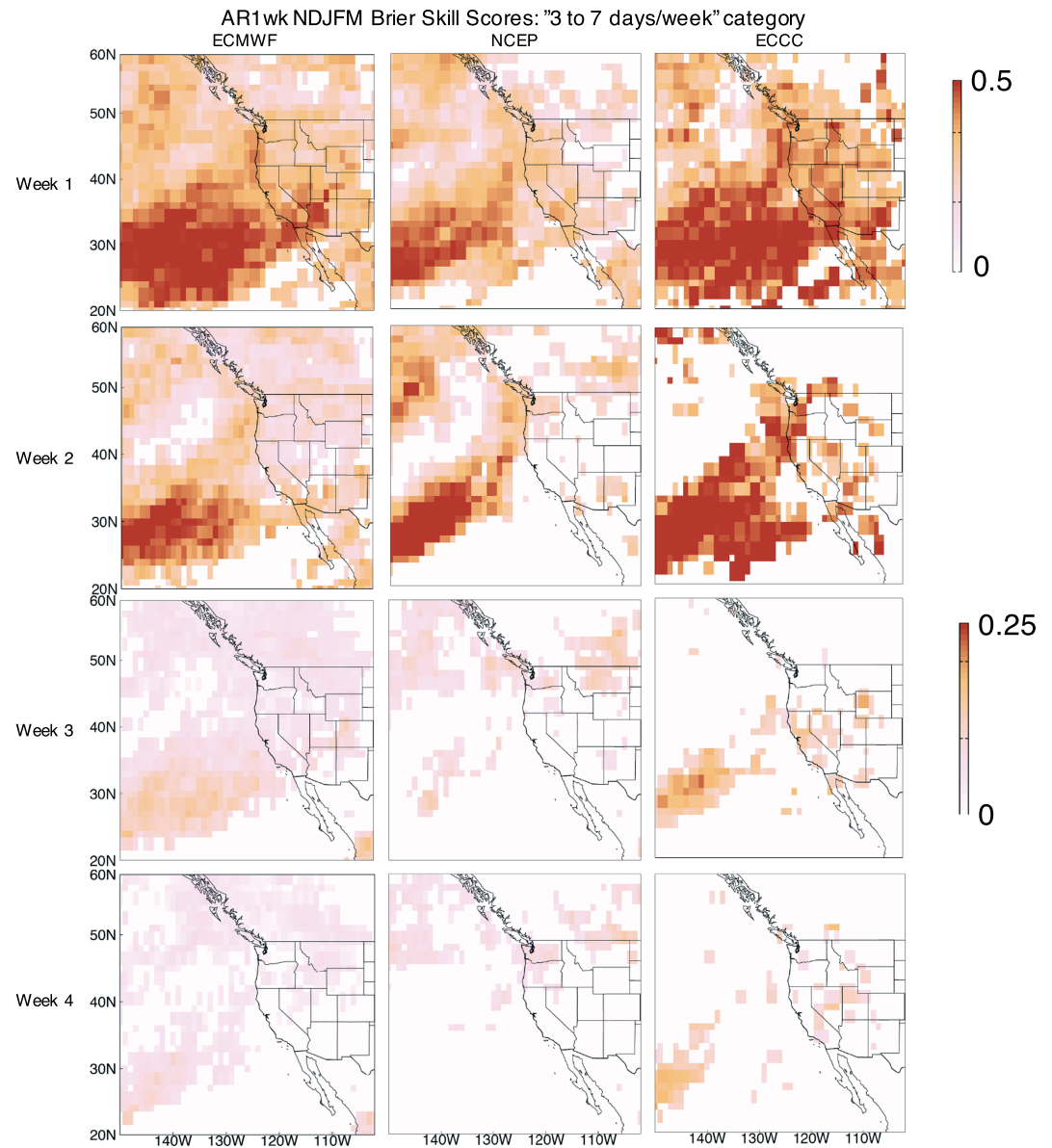


Figure 8. AR1wk NDJFM BSS values in each hindcast system (columns) at each lead time (rows) for the 3–7 AR days/week activity level. Values are only plotted if they are >0 (i.e., skillful with respect to ERA-I reference climatology) and if they are significant at the 95% level using a 100-sample bootstrapping procedure with replacement. The color bar range for week 1 (top row) is 0 to 0.5, while the color bar range for weeks 2–4 is 0 to 0.25.

overpredicts (underpredicts) the percentage of weeks with no AR days north (south) of approximately 35°N by 5–10%. The ECCM hindcast system generally overpredicts the percentage of weeks with no AR days across the entire domain, with the highest values occurring near the Sierra Nevada mountains. The spatial structure of the AR activity bias in each hindcast system for the 1–2 AR days/week activity level is consistent in magnitude and opposite in sign of the 0 AR days/week activity level over much of the domain. The bias of the percentage of weeks with 3–7 AR days is generally lower in all hindcast systems relative to the percentage of weeks with 0 or 1–2 AR days, which may be attributable to the fact that there are less events falling into this activity level relative to the 0 and 1–2 AR days/week activity levels. The causes of the model biases found in this study could be related to mean biases in the basic state of each hindcast system. For example, biases in the component variables of IVT (horizontal winds and specific humidity)

Table 2
Area Average Biases (Model Minus Reanalysis) of AR Activity Over Each 5° × 5° Region Shown in Figure 1 as a Function of Lead Time and AR Activity Level

Region	Lead	Activity level (AR days/week)	Area-weighted average bias (absolute %)	ECMWF	NCEP	ECCC
Washington	Week 0	0	3.2	2.4	1.8	
		1-2	-1.3	-1.4	-0.8	
		3-7	-1.9	-1.0	-1.1	
	Week 1	0	6.5	4.5	1.3	
		1-2	-3.5	-2.8	-1.0	
		3-7	-3.0	-1.8	-0.3	
	Week 2	0	8.0	4.8	1.8	
		1-2	-5.4	-3.3	-2.0	
		3-7	-2.6	-1.5	0.3	
	Week 3	0	8.2	4.8	3.0	
		1-2	-5.8	-3.7	-3.0	
		3-7	-2.4	-1.0	-0.1	
Oregon/ NorCal	Week 0	0	4.0	2.9	2.4	
		1-2	-1.8	-1.8	-1.8	
		3-7	-2.2	-1.1	-0.7	
	Week 1	0	6.8	4.5	1.6	
		1-2	-3.0	-3.4	-1.2	
		3-7	-3.8	-1.0	-0.4	
	Week 2	0	8.5	5.6	2.6	
		1-2	-5.0	-4.7	-2.4	
		3-7	-3.5	-0.9	-0.2	
	Week 3	0	7.6	5.2	2.9	
		1-2	-4.6	-4.9	-2.2	
		3-7	-3.0	-0.2	-0.7	
Cent. Cal	Week 0	0	2.4	0.5	1.9	
		1-2	-1.0	-0.3	-2.1	
		3-7	-1.4	-0.2	0.2	
	Week 1	0	4.3	1.3	2.5	
		1-2	-1.7	-1.7	-2.4	
		3-7	-2.6	0.3	-0.1	
	Week 2	0	5.3	1.7	4.0	
		1-2	-2.8	-2.5	-3.7	
		3-7	-2.6	0.8	-0.3	
	Week 3	0	4.5	1.3	4.0	
		1-2	-2.7	-2.6	-3.4	
		3-7	-1.8	1.3	-0.6	
So. Cal	Week 0	0	2.4	0.0	3.0	
		1-2	-1.9	-0.7	-2.8	
		3-7	-0.5	0.7	-0.2	
	Week 1	0	2.7	-1.9	3.4	
		1-2	-1.8	0.8	-2.4	
		3-7	-0.9	1.1	-1.0	
	Week 2	0	0.9	-1.9	4.4	
		1-2	-0.5	1.0	-3.6	
		3-7	-0.4	0.9	-0.8	
	Week 3	0	-0.2	-2.0	5.1	
		1-2	-0.2	0.6	-4.1	
		3-7	0.4	1.4	-1.0	

Note. Values shown in Figures 3–5 are first weighted by the cosine of latitude then averaged and reported in column 4 of this table. Values are rounded to the nearest tenth decimal place. The three values reported in column 4 represent the area-weighted average bias for the ECMWF, NCEP, and ECCC hindcast systems, respectively. Positive (negative) values represent an overprediction (underprediction) of weeks falling into each particular activity level relative to the ERA-I reanalysis data. The sum of the biases across all three activity levels for a given region, lead, and hindcast system $\approx 0\%$, by construction.

may be underestimated, which could contribute to the overprediction of no AR activity weeks over the western U.S. in all hindcast systems. Future studies might focus on understanding the underlying physics in each hindcast system that contribute to these errors in mean AR activity.

3.3. Brier Skill Scores

BSSs for the 0, 1–2, and 3–7 AR days/week activity levels are shown in Figures 6, 7, and 8, respectively, for each hindcast system (columns) and over the week-1 (1- to 7-day), week-2 (8- to 14-day), week-3 (15- to 21-day), and week-4 (22- to 28-day) lead windows (rows). Values are only plotted if they are >0 (i.e., skillful with respect to the long-term ERA-I 1996–2015 NDJFM reference climatology) and if all values of the 95% confidence interval, calculated using a 100-sample bootstrapping test with replacement, are >0 . Our bootstrapping test resamples the time elements of both the hindcast and reanalysis data with replacement; that is, the order of the resampled time elements is preserved and applied to both the hindcast and reanalysis data. The area-weighted average BSSs over the $5^\circ \times 5^\circ$ regions shown in Figure 1 are reported in Table 3, and only grid cells with positive BSSs values (i.e., with skill relative to the reference climatology) are included in this calculation. As described at the beginning of section 3, intercomparison of skill in each of the hindcast systems analyzed in this study is hindered by the differences in the ensemble hindcast system design. This is a trade-off with the desire to retain the maximum possible amount of data available from each operational center contributing to the S2S Project database.

In considering Figures 6–8 and Table 3, several important conclusions emerge:

- The NCEP hindcast system has lower skill relative to other hindcast system in predicting AR activity across all levels and all lead times, especially between $150\text{--}130^\circ\text{W}$, $25\text{--}40^\circ\text{N}$.
- All hindcast systems are generally more skillful in predicting no AR activity (0 AR days/week) or high AR activity (3–7 AR days/week) conditions relative to moderate AR activity (1–2 AR days/week) conditions.
- There are isolated regions of positive skill at week-3 over the $150\text{--}125^\circ\text{W}$, $25\text{--}35^\circ\text{N}$ regions in the all three hindcast systems in the 0 and 3–7 AR days/week activity levels, and the magnitude and spatial extent of the skill is larger in the ECMWF and ECCC hindcast systems compared to NCEP.
- The spatial structure of the offshore positive skill region at weeks 3–4 suggests a southwest-to-northeast orientation of skillful prediction of no or high AR activity upstream of the California coastline, with a lack of skill upstream of the Pacific Northwest coastline. The ECMWF hindcast system has additional positive skill over British Columbia and the Pacific Northwest regions for the 3–7 AR days/week activity level.
- It is likely that the enhanced skill for the no and high AR activity cases is driven by different processes being simulated well in the model; for example, the enhanced skill at weeks 3–4 in the no AR activity case may be related to the establishment of an offshore ridge (similar to the *West* ridge type described in Gibson et al., 2019), but the enhanced skill in the high AR activity case is likely not ridge related.
- There is considerable drop-off of positive skill beyond week-2 over land in all three hindcast systems relative to week-1 and week-2.

Table 3
Area Average Brier Skill Scores (BSS) of AR Activity Over Each 5° × 5° Region Shown in Figure 1 as a Function of Lead Time and AR Activity Level

Region	Lead	Activity level (AR days/week)	Area-weighted average BSS ECMWF NCEP ECCC
Washington	Week 1	0	0.29 (100%) 0.22 (100%) 0.31 (100%)
		1–2	0.18 (100%) 0.12 (100%) 0.20 (100%)
		3–7	0.25 (100%) 0.20 (100%) 0.29 (100%)
	Week 2	0	0.08 (100%) 0.05 (72%) 0.12 (100%)
		1–2	0.04 (39%) 0.03 (28%) 0.07 (36%)
		3–7	0.08 (100%) 0.09 (83%) 0.13 (69%)
	Week 3	0	0.05 (3%) X 0.07 (6%)
		1–2	0.04 (3%) X X
		3–7	0.03 (94%) 0.05 (64%) X
	Week 4	0	X X X
		1–2	X X X
		3–7	0.02 (53%) 0.05 (69%) 0.05 (17%)
Oregon/NorCal	Week 1	0	0.34 (100%) 0.28 (100%) 0.36 (100%)
		1–2	0.20 (100%) 0.15 (100%) 0.22 (100%)
		3–7	0.29 (100%) 0.26 (100%) 0.33 (100%)
	Week 2	0	0.10 (100%) 0.05 (64%) 0.16 (100%)
		1–2	0.04 (25%) X 0.09 (83%)
		3–7	0.08 (100%) 0.09 (64%) 0.16 (75%)
	Week 3	0	0.05 (3%) X 0.07 (6%)
		1–2	0.04 (8%) X 0.06 (3%)
		3–7	0.03 (97%) 0.04 (14%) 0.08 (36%)
	Week 4	0	X X 0.08 (50%)
		1–2	X X 0.07 (33%)
		3–7	0.03 (33%) 0.05 (42%) 0.06 (11%)
Cent. Cal	Week 1	0	0.38 (100%) 0.30 (100%) 0.37 (100%)
		1–2	0.23 (100%) 0.16 (100%) 0.24 (100%)
		3–7	0.33 (100%) 0.24 (100%) 0.35 (94%)
	Week 2	0	0.12 (100%) 0.08 (100%) 0.18 (100%)
		1–2	0.06 (56%) 0.05 (22%) 0.10 (92%)
		3–7	0.08 (92%) 0.09 (6%) 0.16 (39%)
	Week 3	0	0.06 (58%) 0.05 (8%) 0.05 (3%)
		1–2	0.05 (47%) 0.03 (6%), X
		3–7	0.05 (97%) X 0.10 (44%)
	Week 4	0	0.04 (28%) X 0.08 (67%)
		1–2	0.03 (8%) X 0.07 (56%)
		3–7	0.02 (8%) X 0.08 (3%)
So. Cal	Week 1	0	0.38 (100%) 0.28 (100%) 0.40 (100%)
		1–2	0.26 (100%) 0.17 (100%) 0.29 (100%)
		3–7	0.41 (100%) 0.30 (92%) 0.44 (100%)
	Week 2	0	0.16 (100%) 0.07 (92%) 0.19 (100%)
		1–2	0.09 (94%) 0.05 (17%) 0.12 (100%)
		3–7	0.13 (83%) 0.10 (11%) 0.16 (25%)
	Week 3	0	0.06 (92%) 0.04 (22%) 0.06 (8%)
		1–2	0.04 (25%) X X
		3–7	0.06 (53%) X 0.10 (31%)
	Week 4	0	0.03 (3%) X 0.07 (8%)
		1–2	0.03 (8%) X 0.07 (25%)
		3–7	X X 0.06 (17%)

Note. Values shown in Figures 6–8 are first weighted by the cosine of latitude then averaged and reported in column 4 of this table. Values are rounded to the nearest hundredth decimal place. The three values reported in column 4 represent the area-weighted average BSS for the ECMWF, NCEP, and ECCC hindcast systems, respectively. Positive values represent a skillful prediction relative to the reference climatology described in section 2.4. X symbols denote a region, lead, and activity level where no statistically significant positive BSS values are found. The percentage of grid cells within the 5° × 5° region with statistically significant positive BSS values is reported in parentheses after each value in column 4.

- There is generally very little skill at week-4 lead time, except over the 150–140°W, 20–30°N regions in the ECMWF and ECCC hindcast systems. There are also isolated grid cells over land with very weak positive skill during week-4 in the ECCC hindcast system across all three activity levels, and in the ECMWF and

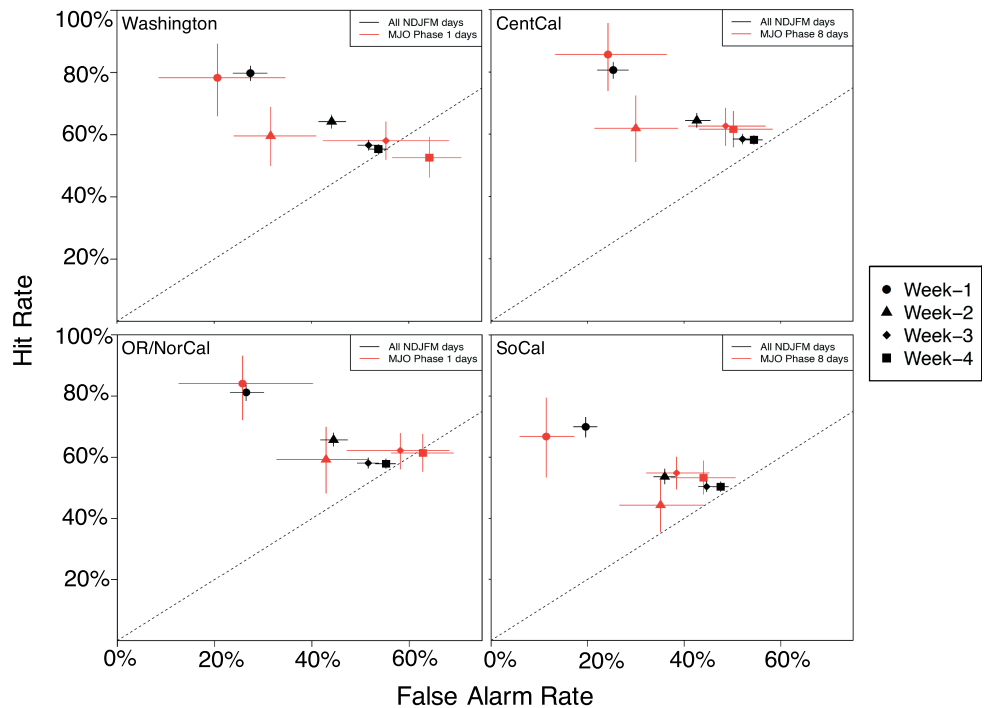


Figure 9. ROC-like diagrams (ensemble mean hit rate vs. false alarm rate) of AR activity over four $5^\circ \times 5^\circ$ regions along the western U.S. coastline displayed in Figure 1 for composites of ECMWF hindcasts where large differences between hindcasts initialized during particular phases of strong MJO activity (red) and all NDJFM days (black) exist. The 95% confidence intervals based on a 1000-sample bootstrapping procedure with replacement are shown as horizontal and vertical lines for false alarm rate and hit rate, respectively.

NCEP hindcast systems in the 3–7 AR days/week activity level, but there is no coherent pattern of positive skill at this lead time relative to other lead times and regions.

- The spatial structure and location of isolated positive skill over the subtropical North Pacific at weeks 3–4 are similar for the 0 and 3–7 AR days/week activity levels.
- The spatial structure of this positive skill region at weeks 3–4 lead is similar to that of the 3–7 AR days/week ERA-I climatology, but not the 0 AR days/week activity level climatology.

4. S2S Skill Assessment over $5^\circ \times 5^\circ$ Boxes Spanning the Western U.S. Coastline

4.1. ROC-Like Diagrams

ROC-like diagrams for the Washington, Oregon-Northern California, Central California, and Southern California regions described in section 2.7 are shown in Figures 9–11 for weeks 1–4 lead time (denoted by symbol type) during NDJFM in the ECMWF hindcast system. The criteria for hits, misses, false alarms, and correct rejections are summarized in section 2.5 and in Figure 2. The ROC values for the *all-days* case are denoted by black symbols in each figure; ROC values for strong MJO, active El Niño, and active La Niña conditions are shown as red, red, and blue symbols in Figures 9, 10, and 11, respectively. Ninety-five percent confidence intervals for the false alarm rate and hit rate are denoted by horizontal and vertical lines, respectively, for both the climate mode composite and all days cases based on a 1,000-sample bootstrapping procedure with replacement. For the MJO composite case, particular phases that exhibit large differences from the all-days case are chosen for display in Figure 9. The number of cases considered for each mode/phase combination is reported in Table 4, along with the total number of NDJFM all days cases in the ECMWF hindcast system. It is noted that while the nonoverlapping of climate mode composite and all days confidence intervals indicate statistically significant differences at the 95% level, the overlapping of the confidence intervals gives no indication on statistical significance or insignificance (Lanzante, 2005).

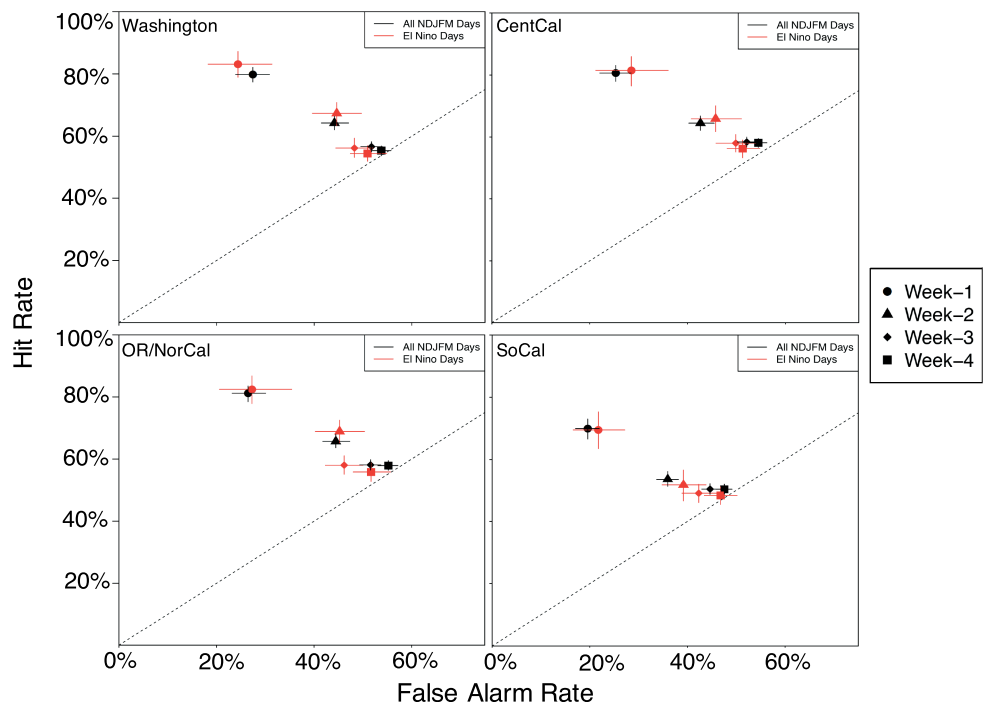


Figure 10. ROC-like diagrams of AR activity (ensemble mean hit rate versus false alarm rate) over four $5^\circ \times 5^\circ$ regions along the western U.S. coastline displayed in Figure 1 for composites of ECMWF hindcasts initialized during El Niño conditions (red) and all NDJFM days (black); 95% confidence intervals based on a 1000-sample bootstrapping procedure with replacement are shown as horizontal and vertical lines for false alarm rate and hit rate, respectively.

4.1.1. MJO Composites

The largest modulations in prediction skill over the Washington and Oregon/Northern California regions occur in the composite of hindcasts initialized during MJO Phase 1 (Figure 9, left column). Over the Washington region, false alarm rates are significantly lower (higher) at the 95% level at week-2 (week-4) lead time relative to the all days case (as evidenced by the nonoverlapping false alarm confidence intervals), while hit rates are not substantially different. This suggests that forecasts of AR activity in this region initialized during MJO Phase 1 are more (less) reliable at week-2 (week-4) lead time. Over the Oregon/Northern California region, false alarm rates are somewhat higher at week-4 lead relative to the all days case (though not at the 95% level), while hit rates are not substantially different. This is similar to the result found at week-4 lead over Washington, which suggests that week-4 forecasts of AR activity initialized during MJO Phase 1 are less reliable in this region.

The largest modulations in prediction skill over the Central California and Southern California regions occur in the composite of hindcasts initialized during MJO Phase 8 (Figure 9, right column). Over Central California, false alarm rates are significantly lower at the 95% level at week-2 lead relative to the all days case (as evidenced by the nonoverlapping false alarm confidence intervals), while hit rates are not substantially different, suggesting that forecasts of AR activity are more reliable under these conditions. There is also evidence that week-3 and week-4 hindcasts of AR activity in this region may be more reliable for the MJO Phase 8 initial condition composite case (as evidenced by the decreases in mean false alarm rate and increases in mean hit rate for the MJO-dependent cases), but the small sample size from the 20-year record here prevents any conclusive statements from being made regarding these lead times (as evidenced by the overlapping confidence intervals). Over Southern California, hindcasts of AR activity initialized during MJO Phase 8 are associated with substantially lower false alarm rates, but similar hit rates, at week-1 lead time (at nearly the 95% significant level, but the false alarm error bars slightly overlap). There is evidence that week-2 (week-3) hindcasts of AR activity are less (more) skillful due to a large decrease (increase) in hit (false alarm) alarm rate, but the relatively small sample size does not allow for a robust conclusion to be made regarding these lead times (as indicated again by the overlapping confidence intervals).

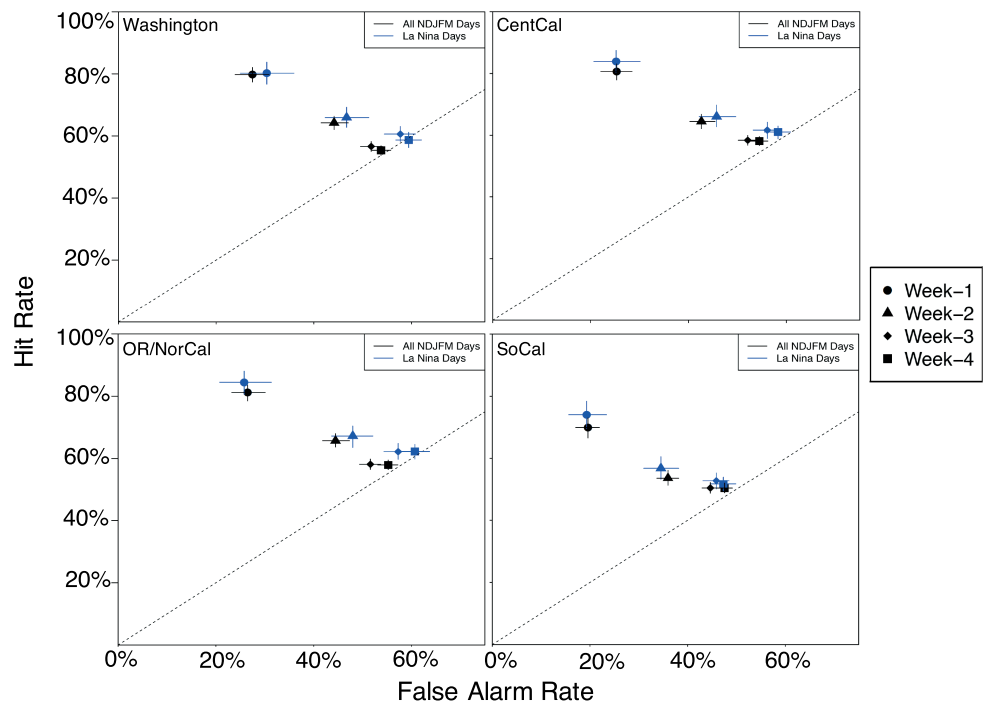


Figure 11. ROC-like diagrams of AR activity (ensemble mean hit rate versus false alarm rate) over four $5^\circ \times 5^\circ$ regions along the western U.S. coastline displayed in Figure 1 for composites of ECMWF hindcasts initialized during La Niña conditions (blue) and all NDJFM days (black); 95% confidence intervals based on a 1000-sample bootstrapping procedure with replacement are shown as horizontal and vertical lines for false alarm rate and hit rate, respectively.

4.1.2. ENSO Composites

Over the Washington region, modulations of prediction skill during active El Niño conditions (Figure 10) are not distinguishable from prediction skill during the all days case, as demonstrated by the substantial overlapping of the false alarm and hit rate confidence intervals at all lead times. In this same region during active La Niña conditions (Figure 11), false alarm rates are significantly higher at the 95% level at week-3 and week-4 lead times compared to the all days case, while hit rates are not significantly different, suggesting that forecasts of AR activity are less reliable under these conditions relative to the all days conditions. Over the Oregon/Northern California region, false alarm rates are substantially lower at week-3 lead time during active El Niño conditions relative to the all days case (at nearly the 95% level, but the false alarm error bars slightly overlap), while hit rates are not substantially different, suggesting that forecasts of AR activity are more reliable under these conditions relative to the all days case.

There are no substantial modulations of AR activity prediction skill during La Niña conditions in this region, as the hit rate and false alarm rate confidence intervals for the La Niña and all days cases substantially overlap at week-1 and week-2 lead times, and there are similar increases in hit rates and false alarm rates at week-3 and week-4 lead times for the La Niña composite relative to the all days case. Prediction skill of AR activity over the Central California and Southern California regions does not appear to be modulated strongly during either active El Niño or La Niña events at any lead time.

4.2. Summary of Forecasts of Opportunity of AR Activity Conditioned on Active MJO and ENSO Events and Comparison to Previous Related Studies

Several of our results in evaluating S2S hindcast prediction skill of ARs are consistent with results from previous studies. In describing these

Table 4

Sample Sizes of Active MJO, El Niño, and La Niña Composites Analyzed in This Study

Climate mode of variability	Number of samples
MJO Phase 1	48
MJO Phase 2	58
MJO Phase 3	85
MJO Phase 4	75
MJO Phase 5	70
MJO Phase 6	88
MJO Phase 7	81
MJO Phase 8	61
El Niño	249
La Niña	379

Note. Period covered is 1996–2015 during November–March. Total number of ECMWF hindcast dates during this period is 880, which is the sample size of the all days case considered in Figures 9–11.

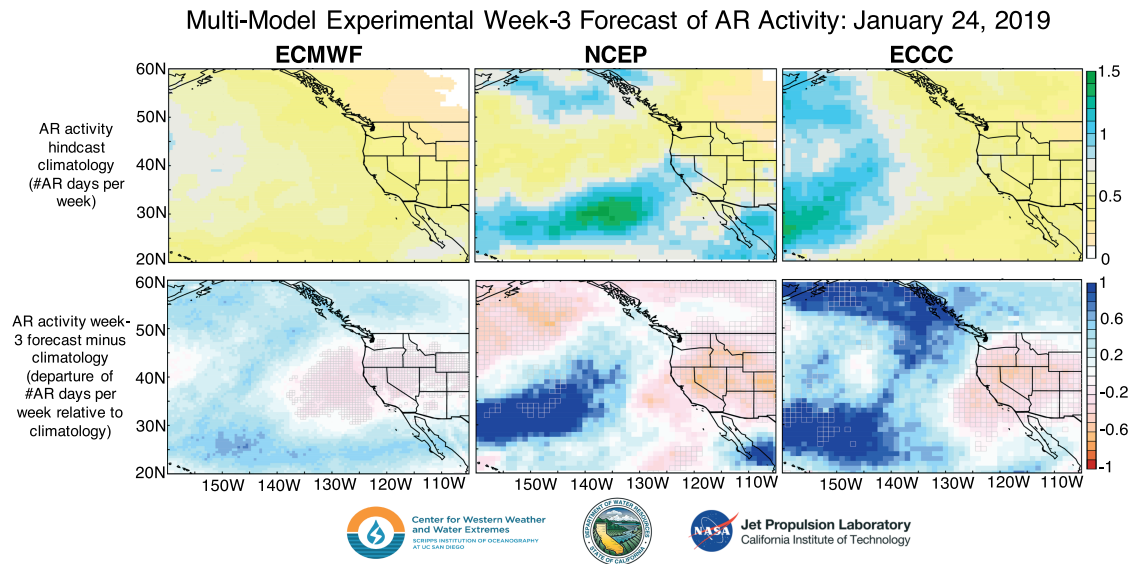


Figure 12. Example of multimodel experimental near real-time week-3 AR activity forecast from the CW3E Experimental Week-3 Outlook protected webpage. Forecast verification period is 8 to 14 February 2019.

consistencies (and some discrepancies) below, it is emphasized that the methodologies used in particular studies differ in important ways, and comparisons should be made cautiously in light of these differences. Specifically, the skill metrics used, the precise domains over which skill is evaluated, and the spatiotemporal averaging of skill employed in this work are not identical to those employed in previous studies.

For example, Pan et al. (2019) found a robust increase in prediction skill of precipitation in the ECMWF hindcast system at week-2 and combined weeks 3 and 4 lead time during MJO Phase 7–8 initial conditions over their *Northern California* region (roughly 125–120°W, 38–42°N; Figure 13 of their study). This region shares considerable overlap with the Central California region (125–120°W, 35–40°N) in the present study, where a robust increase in prediction skill of AR activity during MJO Phase 8 initial conditions is evident at week-2 lead time (substantially lower false alarm rate than the all days case, with similar hit rates). There is also evidence for increased prediction skill at weeks 3–4 lead time (higher hit rates and lower false alarm rates) in this study, but the error bars of the MJO Phase 8 composite case (red) overlap with the error bars of the all days case (black; see upper right panel of Figure 9). The current study also shows evidence for lower false alarm rates of AR activity in this region at week-4 lead time during MJO Phase 7 initial conditions (supporting information Figure S4). Figure 14b of DeFlorio et al. (2019) also showed that there is higher prediction skill of anomalous weekly AR activity at weeks 1–2 lead time in a composite of active MJO Phase 8 initial condition hindcasts in the ECMWF model over the North Pacific/western U.S. region (150–125°W, 35–45°N). This result is consistent with the ROC calculation here (Figure 9, Central California) over a similar domain to that used in DeFlorio et al. (2019) and suggests that a significant increase of AR prediction skill occurs along the western U.S. coastline, and especially near Northern California, 1–2 weeks after active MJO Phase 8 events manifest in the equatorial Pacific. It is also found that the skill modulations seen over Central and Southern California in this study in hindcasts initialized during MJO Phase 8 conditions are dampened if one combines MJO Phases 7–8 (not shown).

In addition, Pan et al. (2019) found a robust decrease in precipitation prediction skill in the ECMWF hindcast system at week-2 and combined weeks 3–4 lead time during MJO Phases 3–4 initial conditions over their *Washington* and *Oregon* regions. Over the Washington region used in this study, there is also evidence for decreased prediction skill of AR activity during MJO Phase 4 initial conditions, especially at weeks 1–2 and week-4 lead time (supporting information Figure S5, left panel). These modulations are less pronounced in hindcasts initialized during combined MJO Phases 3 + 4 conditions (not shown). Mundhenk et al. (2018) also found an absence of prediction skill of AR landfall over the entire California coastline at weeks 2–4 lead time in an empirical statistical model during MJO Phase 4 initial conditions (Figure 3b). Their work suggests

Verification of January 24, 2019 Week-3 AR Activity Outlook: IVT (color shading), SLP (contours)

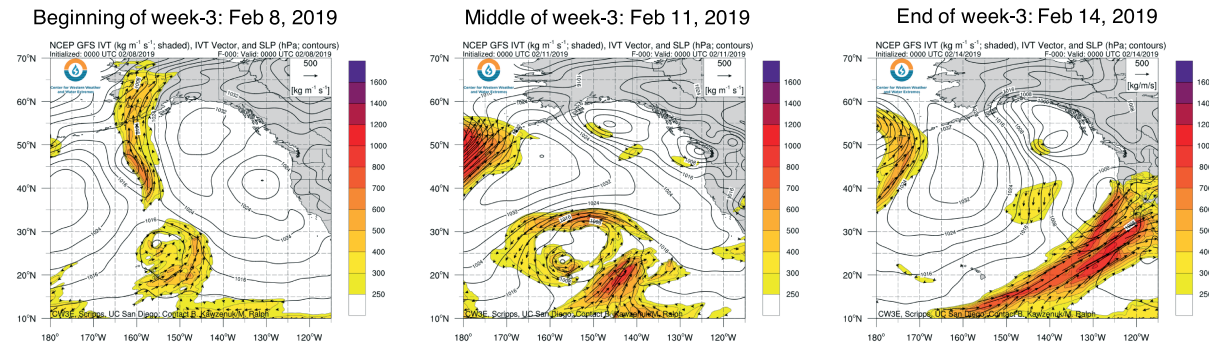


Figure 13. Observed integrated vapor transport (IVT; color shading) and sea level pressure (SLP; contours) fields from week-3 verification period for 24 January 2019 AR activity forecast shown in the bottom row of Figure 12.

that forecasts of S2S AR landfall are least predictable when they are made during MJO Phase 4 initial conditions relative to all other phases. Over the Oregon/Northern California region used in the present study, there is evidence for decreased prediction skill of AR activity at all lead times during MJO Phase 4 initial conditions (supporting information Figure S5, center panel), though not at the 95% confidence level. Over the Central California region, there is evidence for decreased prediction skill of AR activity at week-2 lead time during MJO Phase 4 initial conditions (supporting information Figure S5, right panel).

The weak modulation of AR prediction skill conditioned on ENSO phase found in this study is not in agreement with previous studies regarding the influence of ENSO phase on weeks 2–4 precipitation skill over California. This study does not find a strong modulation of AR prediction skill (using ensemble mean hit and false alarm rates as skill metrics) at weeks 1–4 lead time over Southern California in ECMWF hindcasts initialized during El Niño conditions relative to *all days* conditions. Pan et al. (2019) found that prediction skill of precipitation (using anomaly correlation coefficient skill metric) over Southern California at weeks 2–4 lead time is higher in ECMWF hindcasts initialized during El Niño conditions relative to both La Niña and all days initial conditions. This discrepancy could be related to the difference in skill metrics used or to the fact that AR prediction skill is less sensitive to ENSO phase than precipitation prediction skill in this region. In addition, the ECMWF hindcast system contains only 20 years of data (1996–2015), which is relatively short with respect to the periodicity of ENSO events themselves. Future work, especially using longer hindcast records with more cycles of ENSO included, should focus on exploring these differences and further crystallizing the role of ENSO in modulating both AR and precipitation prediction skill in this region.

To summarize, the results presented in Figure 9 in this study suggest that:

- Over Washington, false alarm rates are significantly lower (higher) at the 95% level at week-2 (week-4) lead time for hindcasts initialized during MJO Phase 1 relative to the all days case.
- Over Oregon/Northern California, false alarm rates are somewhat higher at week-4 lead for hindcasts initialized during MJO Phase 1 relative to the all days case (though not at the 95% level).
- Over Central California, false alarm rates are significantly lower at the 95% level at week-2 lead for hindcasts initialized during MJO Phase 8 relative to the all days case.
- Over Southern California, hindcasts of AR activity initialized during MJO Phase 8 are associated with substantially lower false alarm rates, but similar hit rates, at week-1 lead time (at nearly the 95% significant level, but the false alarm error bars slightly overlap).

The results presented in Figures 10 and 11 in this study suggest that:

- Over Washington, during active La Niña conditions, false alarm rates are significantly higher at the 95% level at week-3 and week-4 lead time compared to the all days case.
- Over Oregon/Northern California, during active El Niño conditions, false alarm rates are substantially lower at week-3 lead time relative to the all days case (at nearly the 95% level, but the false alarm error bars slightly overlap).
- Over Central and Southern California, no significant modulations in AR activity prediction skill are found in hindcasts initialized during active ENSO events

- In general, ENSO does not appear to be a strong modulator of weeks 1–4 lead AR prediction skill over much of the western U.S. coastline (especially over Southern California) during the 20-year period (1996–2015) considered in this study. However, the relatively short record (20 years) with respect to the typical periodicity of ENSO events (2–7 years) makes it difficult to state this conclusively. The constraint of the length of the ECMWF hindcast system used in this study precludes us from determining that the relative insensitivity of S2S AR prediction skill to ENSO phase is due simply to limited sample size. This highlights the critical need for an increased number of multidecadal (30–50+ years) hindcast records at S2S lead times in future databases.

This study builds upon the work of Mundhenk et al. (2018) and DeFlorio et al. (2019) by further resolving regions along the western U.S. coastline to assess prediction skill of AR activity, by assessing S2S prediction skill of AR activity in a multimodel dynamical hindcast system framework, and by examining the impact of ENSO (in addition to MJO) on prediction skill of AR activity. This study also builds on the work of Pan et al. (2019) by resolving both week-3 and week-4 lead times individually, and the individual MJO phases, and by assessing prediction skill of ARs rather than precipitation. Previous studies have shown that ARs are associated with significant annual variations in precipitation over the western U.S. (e.g., Dettinger et al., 2011) but also that IVT, which is a key ingredient for AR intensity, is more predictable than precipitation itself at all lead times spanning 1–15 days (Lavers et al., 2016). This has important implications for longer-range predictability, which is of very high interest to relevant stakeholders in the water resource management community over the western U.S. region. The fact that ARs are the drivers of annual precipitation variability over the western U.S. only enhances the value of considering S2S prediction skill of ARs themselves.

5. Experimental S2S AR Activity Forecast Products Over the Western U.S.

The California Department of Water Resources has supported the hindcast skill assessment performed in this study and has contributed to the design and curation of multimodel experimental week-3 AR activity forecasts. The analysis presented in this paper, along with the design of these experimental products, has been completed via a close partnership between CW3E and the NASA Jet Propulsion Laboratory.

CW3E/JPL experimental S2S AR activity, AR-related IVT, and AR scale category (Ralph et al., 2019) forecasts have been produced and archived since November 2017, and the hindcast skill assessment presented in this paper will serve as a benchmark against which winter season experimental AR activity forecast skill metrics can be compared. An example of a week-3 multimodel experimental forecast made on 24 January 2019 is shown in Figure 12. Each column represents data from a given operational modeling center (left to right: ECMWF, NCEP, and ECCO). The weekly resolved hindcast climatology of AR activity is shown in the top row, and the forecast anomaly of AR activity with respect to each center's hindcast climatology is shown in the bottom row. Across all three forecast systems, below average AR activity was predicted along the entire western U.S. coastline for the 8–14 February period. Verifying observations of IVT (shading) and sea level pressure (SLP; contours) for three stages of the week-3 forecast period (beginning: 8 February, left panel; middle: 11 February, center panel; and end: 14 February, right panel) over the Northeast Pacific/Western U.S. region are shown in Figure 13. An absence of AR conditions was observed over the beginning and middle of the week-3 forecast period, but toward the end of the week-3 period, an AR developed, strengthened, and eventually made landfall along the California coastline. The large-scale circulation prevalent during the beginning of the week-3 forecast period, including the high-pressure system centered just west of the Oregon/California border, was seemingly well predicted by the three hindcast systems at 15–18-day lead, as suggested by the negative AR activity anomalies that were predicted. However, the regional change in meteorological environment that was conducive to AR development at the end of the week-3 forecast period was not captured by any of the hindcast systems at the time of initialization of the forecast on 24 January. This case study emphasizes a key challenge in S2S prediction: that unexpected local and regional perturbations to seemingly well-predicted large-scale flow patterns may modulate or disrupt the skill of S2S-timescale predictions of relevant phenomena.

6. Discussion

The results presented in this study provide a multimodel hindcast skill assessment of AR activity over the western U.S. and the upstream region of the Northeast Pacific. An emphasis has been placed on

uncertainty quantification of probabilistic boreal winter prediction skill of AR activity (Figures 6–8) and of MJO- and ENSO-conditioned skill modulations of AR activity at weeks 1–4 lead time (Figures 9–11). This study also builds upon previous related work, which focused on shorter-term lead time of AR activity in this region using control runs from the S2S Project database (Nardi et al., 2018) and globally using only the ECMWF ensemble hindcast system (DeFlorio et al., 2018), and other studies, which provided insight into S2S AR prediction skill using empirical models based on observations (Mundhenk et al., 2018) or based on prediction of precipitation rather than ARs (Pan et al., 2019).

The ensemble mean biases and BSSs of no (0), moderate (1–2), and high (3–7) AR days/week activity levels at 1–4-week lead times are computed in this study for the ECMWF, NCEP, and ECCC hindcast systems. ROC-like diagrams are created for regions along the western U.S. coastline in the ECMWF hindcast system, and forecasts of opportunity are identified for region and lead time combinations where S2S prediction skill of AR activity is increased during particular phases of the MJO and the ENSO. A detailed summary of these results is provided in sections 3.2–3.3 and section 4.2, and several important conclusions are reiterated here:

ERA-I AR activity climatology

- Minimum values for the 0 AR days/week activity level are located upstream of the Pacific Northwest coastline, and this region is collocated with a maximum in the percentage of weeks with 1–2 AR days/week.
- Maximum values for the 3–7 AR days/week activity level are shifted approximately 10° south of the maximum in the 1–2 AR days/week activity level.
- the most active AR weeks in the climatological record are associated with southwest-to-northeast oriented AR regimes

AR activity bias

- The ECMWF hindcast system generally overpredicts the percentage of weeks with no AR days by up to 15% (absolute percent difference) across the entire domain.
- The NCEP hindcast system generally overpredicts (underpredicts) the percentage of weeks with no AR days north (south) of approximately 35N by 5–10%.
- The ECCC hindcast system generally overpredicts the percentage of weeks with no AR days across the entire domain, with the highest values occurring near the Sierra Nevada mountains.

AR activity BSS

- The NCEP hindcast system has lower skill relative to other hindcast systems in predicting AR activity across all levels and all lead times, especially between 150–130°W, 25–40°N.
- All hindcast systems are generally more skillful in predicting no AR activity (0 AR days/week) or high AR activity (3–7 AR days/week) conditions relative to moderate AR activity (1–2 AR days/week) conditions.
- The spatial structure of the offshore positive skill region at weeks 3–4 suggests a southwest-to-northeast orientation of skillful prediction of no or high AR activity upstream of the California coastline

AR activity ROC statistics in the ECMWF hindcast system

- Hindcasts initialized during MJO Phase 1 conditions show significant modulations in weeks 1–4 AR activity prediction skill over the Washington and Oregon/Northern California regions.
- Hindcasts initialized during MJO Phase 8 conditions show significant modulations in weeks 1–4 AR activity prediction skill over the Central and Southern California regions.
- Hindcasts initialized during MJO Phase 4 conditions show a significant reduction in weeks 1–4 AR activity prediction skill along the western U.S. coastline.
- Hindcasts initialized during El Niño or La Niña conditions generally do not show significant modulations in AR activity prediction skill over the western U.S. coastal regions, with few exceptions.
- The constraint of the length of the ECMWF hindcast system used in this study precludes us from determining that the relative insensitivity of S2S AR prediction skill to ENSO phase is due simply to limited sample size. This highlights the critical need for an increased number of multidecadal (30–50+ years) hindcast records at S2S lead times in future databases.

A multimodel experimental S2S AR activity outlook designed by the researchers in this study in conjunction with stakeholders from the California Department of Water Resources is introduced in section 5 and

Figure 12, and verifying observations of IVT and SLP for that week-3 period are included in Figure 13. These experimental outlooks will be expanded in the future to include a larger number of forecast systems from the S2S Database as part of the S2S Prediction Project 2019–2021 Real-time Pilot Initiative.

7. Future Directions

S2S prediction of ARs and their associated hydrological impacts is an emerging research topic that is attracting significant attention from both the research community (Baggett et al., 2017; DeFlorio et al., 2018; DeFlorio et al., 2019; Mundhenk et al., 2018; Nardi et al., 2018; Pan et al., 2019) and the applications and stakeholders communities (including commodities trading, agriculture, insurance and reinsurance, and water resource management; e.g., NAS, 2016; White et al., 2017). This study encourages continued investment by the research community into forecasts of opportunity of AR activity and AR intensity, conditioned upon certain states of climate variability. Future studies, which are already in preparation at the time of the writing of this manuscript, will focus on S2S hindcast prediction skill of AR-related IVT and AR scale category introduced in Ralph et al. (2019) rather than just AR occurrence. In addition, future work focusing on large ensemble operational hindcast generation would benefit the reliability and robustness of the statistical relationships suggested in this and related published work. Finally, exploiting postprocessing and machine learning techniques applied to S2S hindcast data (e.g., Chapman et al., 2019) has the potential to improve prediction skill of uncalibrated hindcast data from multiple operational centers.

Additionally, the methodology and hindcast skill assessment presented in this work can be repeated for other hindcast systems in the S2S Project Database. Our group intends to use several additional models from that database as part of the S2S Prediction Project 2019–2021 Real-time Pilot Initiative. This project grants our team access to real-time IVT data (instead of the current 3-week lag) from 9 of the 11 operational centers participating in the S2S Project. The IVT data from those additional centers can then be processed and analyzed similarly to the methodology outlined in this study to provide additional hindcast skill benchmarks against which the real-time experimental S2S AR outlooks can be compared.

Acknowledgments

We gratefully acknowledge support for this research from the California Department of Water Resources. We also thank the World Weather Research Programme (WWRP) and the World Climate Research Program (WCRP) for curating the S2S hindcast database, from which we downloaded and processed the data required to perform the analyses in this study. D. W., A. G., S. A., and P. G.'s contributions to this study were carried out on behalf of the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. We thank Emerson LaJoie, Dan Collins, and Jon Gottschalck (NCEP) for their productive discussions regarding S2S hindcast verification techniques. Hindcast data from this study can be accessed on the ECMWF S2S reforecasts website portal (<https://apps.ecmwf.int/datasets/data/s2s-forecasts-instantaneous-accum-ecmf>), and ERA-I data can be accessed via the ECMWF ERA-I portal (<https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim>).

References

- 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*, 7528–7536. <https://doi.org/10.1002/2017GL074434>
- Baggett, C. F., Nardi, K. M., Childs, S. J., Zito, S. N., Barnes, E. A., & Maloney, E. D. (2018). Skillful subseasonal forecasts of weekly tornado and hail activity using the Madden-Julian Oscillation. *Journal of Geophysical Research: Atmospheres*, *123*, 12,661–12,675. <https://doi.org/10.1029/2018JD029059>
- Chapman, W., Subramanian, A. C., Delle Monache, L., Xie, S.-P., & Ralph, F. M. (2019). Improving atmospheric river forecasts with machine learning. *Geophysical Research Letters*, *46*, 10627–10635. <https://doi.org/10.1029/2019GL083662>
- Corringham, T. W. (2018). Wildfires, floods, and climate variability. UC San Diego. ProQuest ID: Corringham_ucsd_0033D_17641. Merritt ID: ark:/13030/m5g78b7f. Retrieved from <https://escholarship.org/uc/item/49x991gx>
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. <https://doi.org/10.1002/qj.828>
- DeFlorio, M. J., Waliser, D. E., Guan, B., Lavers, D. A., Ralph, F. M., & Vitart, F. (2018). Global assessment of atmospheric river prediction skill. *Journal of Hydrometeorology*, *19*(2), 409–426. <https://doi.org/10.1175/JHM-D-17-0135.1>
- DeFlorio, M. J., Waliser, D. E., Guan, B., Ralph, F. M., & Vitart, F. (2019). Global evaluation of atmospheric river subseasonal prediction skill. *Climate Dynamics*, *52*(5–6), 3039–3060. <https://doi.org/10.1007/s00382-018-4309-x>
- Dettinger, M. D., Ralph, F. M., Das, T., Neiman, P. J., & Cayan, D. R. (2011). Atmospheric rivers, floods and the water resources of California. *Water*, *3*(2), 445–478. <https://doi.org/10.3390/w3020445>
- Ebert, E. E., & Gallus, W. A. Jr. (2009). Toward better understanding of the contiguous rain area (CRA) method for spatial forecast verification. *Weather and Forecasting*, *24*(5), 1401–1415. <https://doi.org/10.1175/2009WAF2222252.1>
- Espinoza, V., Waliser, D. E., Guan, B., Lavers, D. A., & Ralph, F. M. (2018). Global analysis of climate change projection effects on atmospheric rivers. *Geophysical Research Letters*, *45*, 4299–4308. <https://doi.org/10.1029/2017GL076968>
- Ferranti, L., Magnusson, L., Vitart, F., & Richardson, D. S. (2018). How far in advance can we predict changes in large-scale flow leading to severe cold conditions over Europe? *Quarterly Journal of the Royal Meteorological Society*, *144*(715), 1788–1802. <https://doi.org/10.1002/qj.3341>
- Ford, T. W., Dirmeyer, P. A., & Benson, D. O. (2018). Evaluation of heat wave forecasts seamlessly across subseasonal timescales. *npj Climate and Atmospheric Science*, *1*(1), 1–9. <https://doi.org/10.1038/s41612-018-0027-7>
- Gibson, P. G., Waliser, D. E., Guan, B., DeFlorio, M. J., & Swain, D. L. (2019). Ridging associated with drought in the Western U.S.: Characteristics, trends, and predictability. *Journal of Climate*. in revision
- Goldenson, N., Leung, L. R., Bitz, C. M., & Blanchard-Wrigglesworth, E. (2018). Influence of atmospheric rivers on mountain snowpack in the western United States. *Journal of Climate*, *31*(24), 9921–9940. <https://doi.org/10.1175/JCLI-D-18-0268.1>
- Gorodetskaya, I. V., Tsukernik, M., Claes, K., Ralph, M. F., Neff, W. D., & Van Lipzig, N. P. M. (2014). The role of atmospheric rivers in anomalous snow accumulation in East Antarctica. *Geophysical Research Letters*, *41*, 6199–6206. <https://doi.org/10.1002/2014GL060881>

- 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., Molotch, N. P., Fetzer, E. J., & Neiman, P. J. (2012). Does the Madden–Julian oscillation influence wintertime atmospheric rivers and snowpack in the Sierra Nevada? *Monthly Weather Review*, *140*(2), 325–342. <https://doi.org/10.1175/MWR-D-11-00087.1>
- Guan, B., Waliser, D. E., & Ralph, F. M. (2018). An intercomparison between reanalysis and dropsonde observations of the total water vapor transport in individual atmospheric rivers. *Journal of Hydrometeorology*, *19*(2), 321–337. Guirguis, K., Gershunov, A., Shulgina, T., Clemesha, R. E., & Ralph, F. M. (2018). Atmospheric rivers impacting Northern California and their modulation by a variable climate. *Climate Dynamics*, *52*(11), 6569–6583. <https://doi.org/10.1007/s00382-018-4532-5>
- Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29–36. <https://doi.org/10.1148/radiology.143.1.7063747>
- Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., et al. (2017). Extended reconstructed sea surface temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *Journal of Climate*, *30*(20), 8179–8205. <https://doi.org/10.1175/JCLI-D-16-0836.1>
- Huning, L. S., Guan, B., Waliser, D. E., & Lettenmaier, D. P. (2019). Sensitivity of seasonal snowfall attribution to atmospheric rivers and their reanalysis-based detection. *Geophysical Research Letters*, *46*, 794–803. <https://doi.org/10.1029/2018GL080783>
- Huning, L. S., Margulis, S. A., Guan, B., Waliser, D. E., & Neiman, P. J. (2017). Implications of detection methods on characterizing atmospheric river contribution to seasonal snowfall across Sierra Nevada, USA. *Geophysical Research Letters*, *44*, 10445–10453. <https://doi.org/10.1002/2017GL075201>
- Jiang, X., Xiang, B., Zhao, M., Li, T., Lin, S. J., Wang, Z., & Chen, J. H. (2018). Intraseasonal tropical cyclogenesis prediction in a global coupled model system. *Journal of Climate*, *31*(15), 6209–6227. <https://doi.org/10.1175/CLI-D-17-0454.1>
- Lamjiri, M. A., Dettinger, M. D., Ralph, F. M., Oakley, N. S., & Rutz, J. J. (2018). Hourly analyses of the large storms and atmospheric rivers that provide most of California's precipitation in only 10 to 100 hours per year. *San Francisco Estuary and Watershed Science*, *16*(4). <https://doi.org/10.15447/sfews.2018v16iss4art1>
- Lanzante, J. R. (2005). A cautionary note on the use of error bars. *Journal of Climate*, *18*, 3699–3703. <https://doi.org/10.1175/JCLI3499.1>
- Lavers, D. A., Allan, R. P., Wood, E. F., Villarini, G., Brayshaw, D. J., & Wade, A. J. (2011). Winter floods in Britain are connected to atmospheric rivers. *Geophysical Research Letters*, *38*, L23803. <https://doi.org/10.1029/2011GL049783>
- Lavers, D. A., & Villarini, G. (2015). The contribution of atmospheric rivers to precipitation in Europe and the United States. *Journal of Hydrology*, *522*, 382–390. <https://doi.org/10.1016/j.jhydrol.2014.12.010>
- 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*, 2275–2282. <https://doi.org/10.1002/2016GL067765>
- Lee, C. Y., Camargo, S. J., Vitart, F., Sobel, A. H., & Tippett, M. K. (2018). Subseasonal tropical cyclone genesis prediction and MJO in the S2S dataset. *Weather and Forecasting*, *33*(4), 967–988. <https://doi.org/10.1175/WAF-D-17-0165.1>
- Leung, L. R., & Qian, Y. (2009). Atmospheric rivers induced heavy precipitation and flooding in the western U.S. simulated by the WRF regional climate model. *Geophysical Research Letters*, *36*, L03820. <https://doi.org/10.1029/2008GL036445>
- Müller, W. A., Appenzeller, C., Doblas-Reyes, F. J., & Liniger, M. A. (2005). A debiased ranked probability skill score to evaluate probabilistic ensemble forecasts with small ensemble sizes. *Journal of Climate*, *18*(10), 1513–1523. <https://doi.org/10.1175/JCLI3361.1>
- 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. *npj Climate and Atmospheric Science*, *1*(1). <https://doi.org/10.1038/s41612-017-0008-2>
- Nardi, K. M., Barnes, E. A., & Ralph, F. M. (2018). Assessment of numerical weather prediction model reforecasts of the occurrence, intensity, and location of atmospheric rivers along the West Coast of North America. *Monthly Weather Review*, *146*(10), 3343–3362. <https://doi.org/10.1175/MWR-D-18-0060.1>
- Nash, D., Waliser, D., Guan, B., Ye, H., & Ralph, F. M. (2018). The role of atmospheric rivers in extratropical and polar hydroclimate. *Journal of Geophysical Research: Atmospheres*, *123*, 6804–6821. <https://doi.org/10.1029/2017JD028130>
- National Academies of Sciences, Engineering, and Medicine (Baltimore) (2016). *Next generation Earth system prediction: Strategies for subseasonal to seasonal forecasts*. Washington: National Academies Press.
- Neiman, P. J., Ralph, F. M., Moore, B. J., Hughes, M., Mahoney, K. M., Cordeira, J. M., & Dettinger, M. D. (2013). The landfall and inland penetration of a flood-producing atmospheric river in Arizona. Part I: Observed synoptic-scale, orographic, and hydrometeorological characteristics. *Journal of Hydrometeorology*, *14*(2), 460–484. <https://doi.org/10.1175/JHM-D-12-0101.1>
- Neiman, P. J., Ralph, F. M., Wick, G. A., Lundquist, J. D., & Dettinger, M. D. (2008). Meteorological characteristics and overland precipitation impacts of atmospheric rivers affecting the West Coast of North America based on eight years of SSM/I satellite observations. *Journal of Hydrometeorology*, *9*(1), 22–47. <https://doi.org/10.1175/2007JHM855.1>
- Paltan, H., Waliser, D., Lim, W. H., Guan, B., Yamazaki, D., Pant, R., & Dadson, S. (2017). Global floods and water availability driven by atmospheric rivers. *Geophysical Research Letters*, *44*, 10–387. <https://doi.org/10.1002/2017GL074882>
- Pan, B., Hsu, K., AghaKouchak, A., Sorooshian, S., & Higgins, W. (2019). Precipitation prediction skill for the West Coast United States: From short to extended range. *Journal of Climate*, *32*(1), 161–182. <https://doi.org/10.1175/JCLI-D-18-0355.1>
- Ralph, F. M., & Dettinger, M. D. (2011). Storms, floods, and the science of atmospheric rivers. *Eos, Transactions American Geophysical Union*, *92*(32), 265–266. Ralph, F. M., & Dettinger, M. D. (2012). Historical and national perspectives on extreme West Coast precipitation associated with atmospheric rivers during December 2010. *Bulletin of the American Meteorological Society*, *93*(6), 783–790. Ralph, F. M., Dettinger, M., Lavers, D., Gorodetskaya, I. V., Martin, A., Viale, M., et al. (2017). Atmospheric rivers emerge as a global science and applications focus. *Bulletin of the American Meteorological Society*, *98*(9), 1969–1973. <https://doi.org/10.1175/BAMS-D-16-0262.1>
- 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/10.1175/15200493\(2004\)1321721:SACAOO2.0.CO;2](https://doi.org/10.1175/15200493(2004)1321721:SACAOO2.0.CO;2)
- Ralph, F. M., Neiman, P. J., Wick, G. A., Gutman, S. I., Dettinger, M. D., Cayan, D. R., & White, A. B. (2006). Flooding on California's Russian River: Role of atmospheric rivers. *Geophysical Research Letters*, *33*, L13801. <https://doi.org/10.1029/2006GL026689>
- 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>
- Robertson, A., & Vitart, F. (2018). *Sub-seasonal to seasonal prediction: The gap between weather and climate forecasting*. Elsevier.

- Robertson, A. W., Kumar, A., Peña, M., & Vitart, F. (2015). Improving and promoting subseasonal to seasonal prediction. *Bulletin of the American Meteorological Society*, *96*(3), ES49–ES53. <https://doi.org/10.1175/BAMS-D-14-00139.1>
- Saha, S., Moorthi, S., Pan, H. L., Wu, X., Wang, J., Nadiga, S., et al. (2010). The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, *91*(8), 1015–1058. <https://doi.org/10.1175/2010BAMS3001.1>
- Valcke, S., Balaji, V., Craig, A., DeLuca, C., Dunlap, R., Ford, R. W., et al. (2012). Coupling technologies for Earth system modelling. *Geoscientific Model Development*, *5*(6), 1589–1596. <https://doi.org/10.5194/gmd-5-1589-2012>
- 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>
- Wang, S., Anichowski, A., Tippett, M. K., & Sobel, A. H. (2017). Seasonal noise versus subseasonal signal: Forecasts of California precipitation during the unusual winters of 2015–2016 and 2016–2017. *Geophysical Research Letters*, *44*, 9513–9520. <https://doi.org/10.1002/2017GL075052>
- Weigel, A. P., Liniger, M. A., & Appenzeller, C. (2007). The discrete Brier and ranked probability skill scores. *Monthly Weather Review*, *135*(1), 118–124. <https://doi.org/10.1175/MWR3280.1>
- 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\(2004\)1321917:AARMMI2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)1321917:AARMMI2.0.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>
- Wick, G. A., Neiman, P. J., Ralph, F. M., & Hamill, T. M. (2013). Evaluation of forecasts of the water vapor signature of atmospheric rivers in operational numerical weather prediction models. *Weather and Forecasting*, *28*(6), 1337–1352. <https://doi.org/10.1175/WAF-D-13-00025.1>
- Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences*. Academic Press.
- Zhang, Z., Ralph, F. M., & Zheng, M. (2019). The relationship between extratropical cyclone strength and atmospheric river intensity and position. *Geophysical Research Letters*, *46*, 1814–1823. <https://doi.org/10.1029/2018GL079071>
- 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/10.1175/15200493\(1998\)1260725:APAFMF2.0.CO;2](https://doi.org/10.1175/15200493(1998)1260725:APAFMF2.0.CO;2)