Multi-Model Subseasonal Prediction Skill Assessment of Water Vapor Transport Associated With Atmospheric Rivers Over the Western U.S.

Zhenhai Zhang[1] , **Michael J. DeFlorio[1]** , **Luca Delle Monache[1]** , **Aneesh C. Subramanian[2]**, **F. Martin Ralph[1]** , **Duane E. Waliser[3,4]**, **Minghua Zheng[1]** , **Bin Guan[3,4]**, **Alexander Goodman[3]**, **Andrea M. Molod[5]**, **Frederic Vitart[6]** , **Arun Kumar[7]** , and **Hai Lin[8]**

[1]Center for Western Weather and Water Extremes, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA, [2]University of Colorado Boulder, Boulder, CO, USA, [3]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, [4]Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, CA, USA, [5]Global Modeling and Assimilation Office, NASA/GSFC, Greenbelt, MD, USA, [6]European Centre for Medium-Range Weather Forecasts, Reading, UK, [7]Climate Prediction Center, NCEP/NWS/NOAA, College Park, MD, USA, [8]Environment and Climate Change Canada, Montreal, QC, Canada

Abstract Subseasonal-to-seasonal (S2S) forecasts of atmospheric rivers (ARs) are in high demand in the water supply management and flood control communities. This study focuses on a new metric, the accumulated water vapor transport associated with ARs, which is closely related to the winter precipitation over the western U.S., and provides a multi-model S2S prediction skill assessment. The prediction skill is evaluated at lead time 1–4 weeks in four dynamical model hindcast data sets from National Centers for Environmental Prediction (NCEP), European Center for Medium-Range Weather Forecasts (ECMWF), Environment and Climate Change Canada (ECCC), and Global Modeling and Assimilation Office (GMAO) at National Aeronautics and Space Administration. Three reanalysis data sets are used to evaluate the uncertainty of prediction skills related to the choice of references. The AR-related water vapor transport is underestimated in ECMWF and ECCC over most of the investigated region, while its maximum has a southeastward shift in NCEP and GMAO at lead time 3–4 weeks. The root mean square error, anomaly correlation coefficient, and Brier skill score are calculated to quantify the prediction skill in both deterministic and probabilistic sense. At week-3 lead time, the models have significant skill near the lower latitudes (<40°N) of the eastern North Pacific, extending northeastward to the California coastal area. Models have higher skills in forecasting no and strong AR cases than weak cases. The Madden–Julian Oscillation can modulate the prediction skill at week-3 lead over central and Southern California, but with large uncertainties across models.

Plain Language Summary An atmospheric river (AR) is a long and narrow corridor of strong horizontal water vapor transport, which plays a key role in transporting water vapor from the tropics and subtropics to feed precipitation over the U.S. West Coast. Therefore, subseasonal-to-seasonal (S2S) forecasts of ARs are in high demand in water supply management and flood control communities. This study provides an S2S prediction skill assessment for the weekly accumulated water vapor transport from ARs at 1–4 weeks lead times in four dynamical model hindcast data sets. Three reanalysis data sets are used as a reference to evaluate the model hindcasts and test the impacts of different reference data sets on the evaluation results. Several metrics are used to quantify the prediction skills. The results show that the models have some useful skills in predicting the weekly water vapor transport related to ARs at week-3 lead over the lower latitudes of the eastern North Pacific, extending northeastward to the California coastal area. Models have higher skills in forecasting no and strong AR weeks than weak AR weeks. The Madden–Julian Oscillation has some impacts on the prediction skill at week-3 lead over central and Southern California, but the impacts vary across models.

1. Introduction

Predictions at subseasonal-to-seasonal (S2S, more than 2 weeks but less than a season lead time) times-cales are common challenges in both weather and climate models (Brunet et al., 2010; Mariotti et al., 2020; Robertson et al., 2015; Vitart et al., 2017, 2012). On the one hand, S2S lead times are too long for initialized weather prediction to retain the initial condition memory due to the chaotic nature of atmospheric variability (Lorenz, 1963, 1965; Palmer & Hagedorn, 2006; Vitart & Robertson, 2018). On the other hand, S2S lead times

are too short for prediction to respond to the changes of major components in the earth system (e.g., sea surface temperature, sea ice, soil moisture) that modulate predictability (Palmer & Hagedorn, 2006; Saha et al., 2014; Stockdale et al., 2011; W. Wang et al., 2010). Despite being a challenging time scale for predicting impactful hydrometeorological phenomena, information at S2S lead times is invaluable in many applications such as water supply management, flood control, and agricultural risk assessment (Merryfield et al., 2020; Pendergrass et al., 2020; White et al., 2017). Skillful S2S forecasts have the potential to be extremely helpful for policy-setting decision-making (DeFlorio et al., 2021; Mariotti et al., 2020; Sengupta et al., 2022; White et al., 2022).

To meet the operational and research needs for S2S products, a joint effort between the World Weather Research Program and the World Climate Research Program, referred to as the S2S Prediction Project, was developed in 2013 (Mariotti et al., 2018; Vitart et al., 2017; Vitart & Robertson, 2018). The S2S Prediction Project allows for comparisons of the prediction skill of each participating modeling system and investigations of the usefulness of ensemble or multi-model forecasts at the S2S time scale. This project recommends specific priorities for S2S research, such as improving understanding of the mechanisms of S2S predictability, systematically assessing S2S prediction skills across different models, and identifying "forecasts of opportunity", that is, the periods in the hindcast record with higher prediction skill than normal. Accordingly, many S2S studies have focused on the role of Madden–Julian Oscillation (MJO; Madden & Julian, 1971) in modulating S2S predictability (Kim et al., 2016; Mani et al., 2014; Vitart et al., 2017; Waliser et al., 2003).

Recent studies have investigated the regional S2S prediction skill of surface temperature anomalies, ridging events, and jet patterns over the North Pacific and the western U.S. (Albers et al., 2021; Gibson et al., 2020; Z. Zhang, Pierce, & Cayan, 2019). In addition, given the increased societal demands (e.g., water management, flood-risk reduction) in this region, there has been a growing need for improved longer lead time forecasts for atmospheric rivers (ARs, Ralph et al., 2020; Ralph et al., 2017a; Ralph et al., 2004; Zhu & Newell, 1998). An AR is a long, narrow, and transient corridor of strong water vapor transport, and is often identified with vertically integrated water vapor transport (IVT). Over the western U.S., landfalling ARs can contribute up to half of the annual precipitation (Dettinger et al., 2011; Rutz et al., 2014), which can provide beneficial water resources and cause damaging hazards such as floods and coastal surges. The accumulated water vapor transport measured as time-integrated IVT (T-IVT) during the passage of an AR over a location is highly correlated with nearby orographic precipitation over the western U.S. region. Variations in T-IVT can explain >70% of the variance in storm-total precipitation in AR storms (Ralph et al., 2013). Thus, assessing the forecast skill of T-IVT is an indicator of the forecast skill of precipitation with the advantage of understanding the type of storm producing it. More plainly, accurate dynamical forecasts of extreme precipitation in the region require accurate prediction of the ARs that produce it. Lavers et al. (2016) found that predictions of IVT have greater skill than the predictions of precipitation in forecast week 2 since it is not as strongly influenced by errors in the microphysics. Additionally, the representation of topography in the coarse-resolution climate models could be another issue for the simulation of precipitation (Wehner et al., 2010). However, S2S AR predictions could still be challenging due to (a) the unique transient, filamentary structure of ARs (Guan & Waliser, 2019; Zhu & Newell, 1998), (b) the interactions between tropical and extratropical circulation patterns which can modulate midlatitude ARs (Z. Zhang & Ralph, 2021; Z. Zhang, Ralph, & Zheng, 2019), (c) the existence of observational gaps for water vapor transport (M. Zheng et al., 2021), and (d) the uncertainties from AR detection methods (Collow et al., 2022; O'Brien et al., 2020; Shields et al., 2018).

In recent years, Baggett et al. (2017) demonstrated that the forecast skill of ARs might be extended to 3–5 weeks presuming the presence of certain phases of MJO and the quasi-biennial oscillation (QBO, Baldwin et al., 2001; Reed et al., 1961). Nardi et al. (2018) identified little skill for daily AR occurrence beyond a lead of 14 days using nine operational models from the S2S Prediction Project. In contrast, DeFlorio, Waliser, Guan, Ralph, and Vitart (2019) showed that the European Center for Medium-Range Weather Forecasts (ECMWF) S2S prediction skill is noticeably higher than a reference forecast in several regions up to 3-week lead time by using weekly aggregates (the number of AR days that accumulated over a week-long period) instead of daily AR occurrence. In a follow-up study, DeFlorio et al. (2019) further validated the S2S skill of ARs occurrence with three operational hindcast data sets. Cao et al. (2021) investigated the subseasonal prediction skill of AR-related flooding in a hydrological model driven by SubX (Pegion et al., 2019) hindcast system and demonstrated skillful probabilistic predictions out to week 2. Some other studies explored the sources of S2S predictability for AR activity over the U.S. West Coast, including the MJO and QBO (e.g., Mundhenk et al., 2018), the El Niño-Southern Oscillation

Table 1
Characteristics of the Four Dynamical Model Subseasonal-To-Seasonal (S2S) Hindcast Data Sets Used in This Study

Model	Ensemble size	Hindcast frequency	Period covered	Total number of hindcasts	Horizontal resolution
NCEP	4	Daily	1999–2010	1,740	$1.0° \times 1.0°$
ECCC	4	Weekly	1995–2014	420	$0.45° \times 0.45°$
ECMWF	11	Twice per week	1996–2015	840	$0.5° \times 0.5°$
GMAO	4	Every 5 days	1999–2016	522	$0.5° \times 0.5°$

(ENSO; e.g., Huang et al., 2021; Zhou & Kim, 2018), and the sea surface temperature of the North Pacific (e.g., Liu et al., 2021; Sun et al., 2021).

Past studies of S2S validation for ARs were mainly focused on occurrence-based quantities (e.g., DeFlorio et al., 2019) and disagree on the limit of skillful S2S prediction of AR occurrence. Aggregated AR occurrence metrics could be complemented with physical quantities as validation metrics, such as the T-IVT, which is one of the variables that best characterizes ARs (Cordeira & Ralph, 2021; Ralph et al., 2017b). Thus, this study aims to investigate the S2S prediction skill of the T-IVT associated with ARs over the western U.S. and the eastern North Pacific. Different from the occurrence-based quantity in DeFlorio et al. (2019), T-IVT associated with ARs is a physical quantity that takes both AR occurrence and AR intensity into account, which is closely related to the precipitation amount and plays a dominant role in winter precipitation over the western U.S. This evaluation is completed using the hindcast output from four dynamical S2S models (Table 1). In addition to three operational hindcast data sets from the S2S Prediction Project, the hindcast output of the Global Earth Observing System (GEOS, Molod et al., 2020) from the Global Modeling and Assimilation Office (GMAO) at the National Aeronautics and Space Administration (NASA) is also evaluated in this study. GEOS modeling and data assimilation system from NASA's GMAO was developed with the primary goal of improving the use of satellite assimilation for seasonal predictions. Its hindcast data set has not been evaluated in previous studies focused on S2S AR prediction skills. Three different reanalysis data sets are used to evaluate the prediction skill in the context of reanalysis uncertainty. The impact of MJO on the prediction skill of T-IVT associated with ARs is also explored. Thus, this study provides a baseline assessment to document systematic errors and biases in representing ARs at S2S scales, explores "forecasts of opportunity" where S2S prediction skill is higher than normal conditions during certain MJO phases, and informs decision makers about the reliability and confidence of each S2S system.

2. Data and Methods

2.1. Data

In this study, four S2S hindcast data sets were evaluated to investigate the subseasonal prediction skill of the water vapor transport associated with ARs over the western U.S. and the eastern North Pacific. Three of them are from the operational centers participating in the S2S Prediction Project (Vitart et al., 2017), including National Centers for Environmental Prediction (NCEP), ECMWF, Environment and Climate Change Canada (ECCC). The fourth hindcast data set is from NASA GMAO (Molod et al., 2020). The characteristics of the four hindcast data sets are summarized in Table 1, including the ensemble size, hindcast frequency, hindcast period, and the total number of hindcasts evaluated in this study. There are several major differences among the parameters of the four hindcast data sets. ECMWF has 11 ensemble members while the other three systems have four ensemble members. Hindcasts are initialized daily for NCEP, weekly for ECCC, twice a week (every Monday and Thursday) for ECMWF, and every 5 days for GMAO. The hindcast period varies from 12 to 20 yr in the different hindcast data sets. Due to those differences, it is not proper to directly compare the prediction skill between different models. This study aims to provide an assessment of prediction skill for individual models based on their hindcasts. It is noteworthy that the ECCC and ECMWF systems considered here have been superseded by upgraded operational versions. In this study, the hindcast versions of NCEP, ECCC, and ECMWF are consistent with the hindcasts used in DeFlorio et al. (2019) so that the results can be fairly compared.

We focus on ARs during the cool season (November–March), and therefore only hindcasts in the cool season were used. As a result, the number of hindcasts used in this study is 1,740 for NCEP, 420 for ECCC, 840 for ECMWF, and 522 for GMAO. Hindcast output data from all four models consists of daily instantaneous values at 00 UTC interpolated to a common 1° longitude × 1° latitude horizontal grid using bilinear interpolation.

The subseasonal prediction skill of the water vapor transport associated with ARs is verified using three reanalysis data sets: the fifth generation ECMWF atmospheric reanalysis (ERA5, Hersbach et al., 2020), the Modern-Era Retrospective Analysis for Research and Applications Version 2 (MERRA2, Gelaro et al., 2017) from NASA, and the Climate Forecast System Reanalysis (CFSR, Saha et al., 2010) from NCEP. CFSR spans from January 1979 to December 2010, therefore the NCEP Climate Forecast System Version 2 analysis (Saha et al., 2014), which spans from January 2011 to present, was used as the extension of CFSR after January 2011. All three reanalysis data sets were obtained on a 1° longitude × 1° latitude horizontal grid to match the resolution of the interpolated model hindcast data.

Madden–Julian Oscillation (MJO) is a dominant mode of the subseasonal variability in the tropical atmosphere (C. Zhang, 2005). The MJO also has a significant influence on the extratropical weather and climate, which can provide a useful signal to the subseasonal prediction (Waliser et al., 2003). Recent study (J. Wang et al., 2023; Zhou et al., 2021) found that MJO has large impacts on the AR activities over the North Pacific region. Thus, in this study we explored the impacts of MJO on the prediction skill of ARs. MJO phases were defined based on the daily Real-time Multivariate MJO (RMM) index (Wheeler & Hendon, 2004), obtained from the Center for Australian Weather and Climate Research. The RMM is a function of daily outgoing longwave radiation and zonal winds at 850 and 200 hPa in the tropics, which provides the information of the MJO phases (8-phase cycle) and amplitude. In this study, only active MJO phases (RMM > 1.0) were used to estimate the impacts of MJO on the subseasonal prediction skill of the water vapor transport associated with ARs.

2.2. AR Detection

Vertically IVT was calculated in both reanalysis and model data sets following Rutz et al. (2014) using specific humidity, zonal, and meridional wind components at 300, 500, 700, 850, and 1,000 hPa pressure levels because only limited pressure levels are available from the S2S hindcast data set. If a data set has values at the pressure levels below the ground, those values are masked out so that the IVT is integrated from the lowest pressure level above the surface to 300 hPa. ARs were next identified using the detection algorithm developed by Guan and Waliser (2015) and refined in Guan et al. (2018). In this detection scheme, multiple and sequentially higher IVT thresholds (85th–95th percentiles) with a fixed lower limit of 100 kg m^{-1} s^{-1} specific to each location and season were applied to the IVT fields to find the potential AR objects. The IVT fields were scanned as needed with the 85th, 87.5th, 90th, 92.5th, and 95th percentile thresholds such that too large or less well-defined AR objects could be identified (Guan et al., 2018). The potential AR objects were further filtered based on several requirements, such as on IVT direction of the object (within 45° of the direction of the AR object), length (>2,000 km), and length/width ratio (>2). DeFlorio et al. (2019) used a common set of 85th–95th percentile IVT thresholds derived from one reanalysis product (ECMWF Reanalysis Interim) to detect ARs in all models. However, in this study the 85th–95th percentile IVT thresholds for AR detection were computed using the three reanalyses and four model hindcast data sets respectively. This change is implemented because different reanalyses and models have unique climatologies. Using data set-based IVT percentile thresholds, the above algorithm is shown to be among a few Atmospheric River Tracking Method Intercomparison Project (ARTMIP; O'Brien et al., 2020; Shields et al., 2018) algorithms that are particularly robust to the choice of the input data product (Collow et al., 2022; their Figures 3 and 4), which is a desirable attribute for the current study. Meanwhile, different requirements from different detection methods (e.g., 85th–95th percentile relative threshold or 250 kg m^{-1} s^{-1} absolute threshold) may cause some uncertainties in the identified ARs and thus the prediction skill. Further study using different AR detection methods, such as those participating in ARTMIP, is needed to quantify those uncertainties.

2.3. Subseasonal Prediction Skill Metrics

Since this study focuses on the subseasonal time scale, the water vapor transport is quantified using time-integrated IVT (T-IVT; accumulated IVT) over a week-long period (7 days). At each grid cell, the IVT within the detected AR objects (Guan & Waliser, 2015) is calculated as T-IVT associated with ARs (AR T-IVT). In other words, at a particular location the weekly AR T-IVT is the water vapor transport that only occurred under AR conditions during that week. For the hindcast data sets, the AR T-IVT at forecast week-1 (days 1–7) to week-4 (days 22–28) lead time is evaluated. As summarized in Table 1, the four hindcast data sets have different hindcast frequencies. Thus, the reanalysis data is resampled and reorganized to match the forecast initialization time and forecast time period for each of the four hindcast data sets.

The weekly AR T-IVT in the four hindcast data sets is evaluated using four prediction skill metrics: model bias, root-mean-square error (RMSE), anomaly correlation coefficient (ACC), and Brier skill score (BSS). In this study, the skill calculations are not cross-validated, so the true out-of-sample skills might be lower than the skills shown in the results section. Model bias is defined as the difference in the climatological mean weekly AR T-IVT between the model hindcast and the reanalysis. RMSE is defined as the square root of the averaged squared differences between hindcasts and reanalysis, representing the accuracy of the hindcasts. As RMSE approaches zero, the hindcasts are more accurate.

ACC is defined as the correlation between the temporal anomalies of the model hindcasts and the reanalysis, which is a concise metric to quantify the skill in forecasting the patterns of departures from the mean climatological conditions. If the variation of the forecasted AR T-IVT anomalies is perfectly coincident with that of reanalysis, ACC = 1 (a perfect forecast); if the variation of forecasted anomalies is opposite of that of reanalysis, ACC = −1.

BSS is utilized to verify hindcast data sets in a probabilistic framework and it indicates the degree of forecast improvements in reference to the climatological forecast. Weigel et al. (2007) found that BSS is negatively biased for the forecasts with small ensemble size, so we followed their formula to avoid that bias since the four hindcast data sets have relatively small ensemble size. For a given location, the weekly AR T-IVT is grouped into three probabilistic categories: (a) no AR T-IVT, the weeks with no ARs; (b) weak AR T-IVT, the lower 50% of the weeks with (weaker) AR T-IVT; and (c) strong AR T-IVT, the upper 50% of the weeks with (stronger) AR T-IVT. Following Weigel et al. (2007), the BSS is calculated for each category, and it represents the probabilistic skill of each model in predicting these specific categories of weekly AR T-IVT in reference to the climatological forecast. If the model forecast has a lower skill than the climatological forecast, BSS < 0; if the forecast has the same skill with the climatological forecast, BSS = 0; if the forecast has a higher skill than the climatological forecast, BSS >0; and if the forecast is perfect, BSS = 1.

3. Results

3.1. Water Vapor Transport in Three Reanalysis Data Sets

Before evaluating the model performance, the water vapor transport in the ERA5, CFSR, and MERRA2 reanalysis data sets were compared to examine the uncertainties in the reanalysis estimate. Figure 1 shows the climatology of weekly total T-IVT in the cool season (November–March) from 1995 to 2016, which covers the hindcast periods of all the four models. The weekly total T-IVT is calculated with the original IVT field without considering AR conditions. The three reanalyses have a similar spatial pattern of the T-IVT over the eastern North Pacific, including a maximum (T-IVT >14 × 10^7 kg m^{-1} per week) near 40°N over the ocean and a decrease from the ocean to the coastal region (Figures 1a–1c). Along the U.S. West Coast, the T-IVT is strongest (∼10 × 10^7 kg m^{-1} per week) over the coast of Oregon, Washington, and Northern California, and decreases to ∼6 × 10^7 kg m^{-1} per week at the coast of Southern California. Meanwhile, the T-IVT decreases substantially from the coast to the inland area.

Although the three reanalyses have a similar spatial pattern of the total T-IVT, some differences exist in a few regions (Figures 1 d–1f). Over most of the eastern North Pacific, the total T-IVT is slightly higher in MERRA2 (<5%) and CFSR (5%–10%) than in ERA5 (Figures 1e and 1f). Along the coast, MERRA2 is ∼5% lower in total T-IVT over the coastal region of Oregon and around the San Francisco Bay, and 5%–15% lower over Baja California than ERA5 and CFSR (Figures 1d and 1f). The difference between MERRA2 and ERA5 over the eastern North Pacific (Figure 1f) is roughly consistent with the results in Collow et al. (2022), who found that the global mean IVT in MERRA2 is 5%–10% higher than in ERA5. However, our results also show that there are large regional variabilities in the difference, especially over the continent. Over the inland continent, MERRA2 is 5%–15% lower in the northern Idaho and eastern Montana than ERA5 and CFSR. The differences of total T-IVT over the other inland regions between the three reanalyses are relatively small, around or below 5%. These differences in T-IVT between the reanalysis data sets may be due to many factors, such as the data assimilation methods and assimilated observations, model topography, land data assimilation and the surface scheme, and model resolution.

In addition to the total T-IVT, the climatology of weekly AR T-IVT in 1995–2016 cool seasons is also examined in these three reanalyses (Figure 2). Over the eastern North Pacific, the AR T-IVT has a maximum
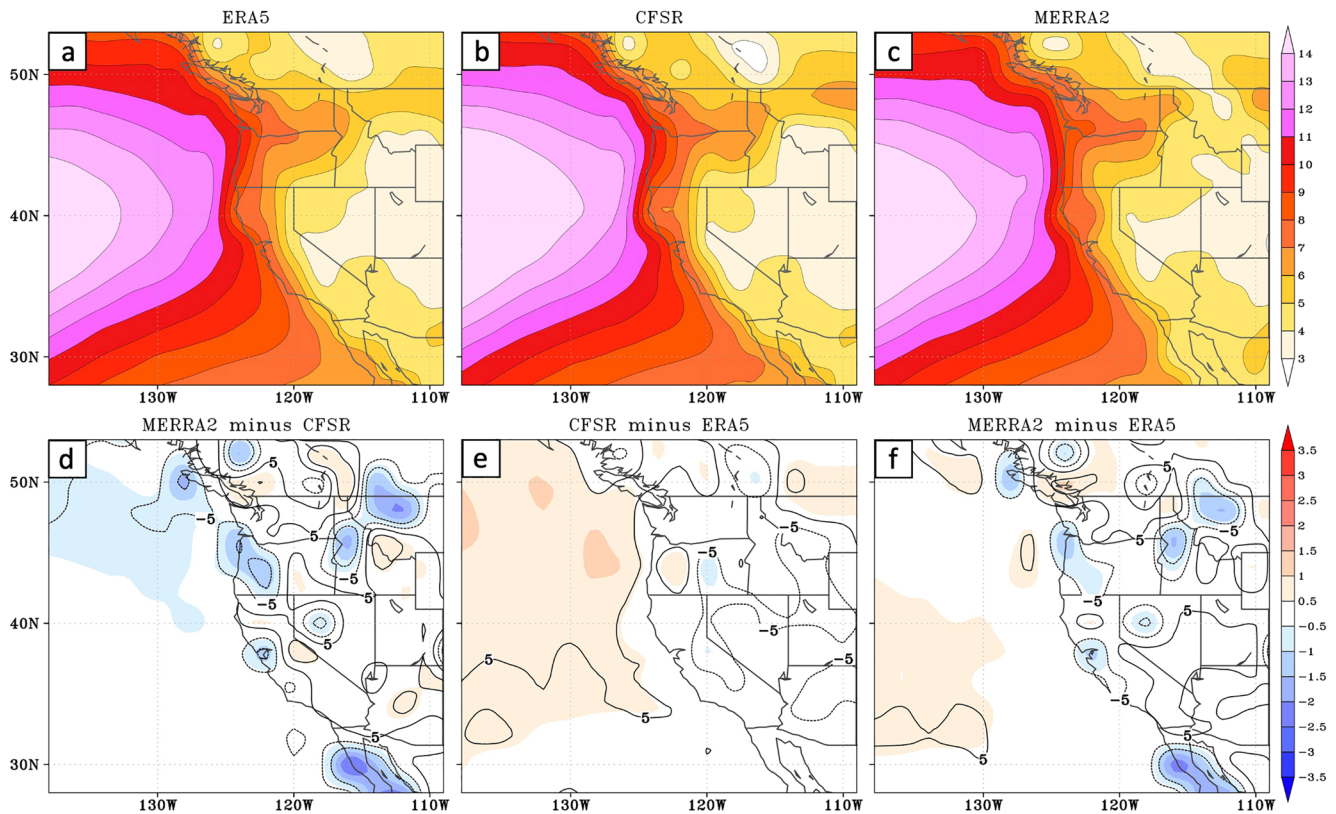
Figure 1. (a–c) The climatology of weekly T-IVT (colors, 10^7 kg m^{-1} per week) in 1995–2016 cool seasons (November–March) from (a) ERA5, (b) CFSR, and (c) MERRA2. (d–f) The differences (colors, 10^7 kg m^{-1} per week) and percentage differences (contours every 10%) of the climatological mean T-IVT for (d) MERRA2 minus CFSR, (e) CFSR minus ERA5, and (f) MERRA2 minus ERA5.

($>4.2 \times 10^7$ kg m^{-1} per week) located at the same place as the total T-IVT maximum, around 40°N over the ocean in all the three reanalyses (Figures 2a–2c). This is the eastern portion of the high-frequency area of AR activity over the North Pacific (Guan & Waliser, 2015; Guan et al., 2018), which is located over the south side of the North Pacific storm track (Hoskins & Hodges, 2002). Along the U.S. West Coast, the strongest AR T-IVT ($\sim3.0 \times 10^7$ kg m^{-1} per week) occurs over the coastal region of Oregon, Washington, and Northern California, which contributes to $\sim30\%$ of the total T-IVT. In southern Washington and northern Oregon, there is a clear inland penetration of AR T-IVT, which is similar with the feature there in Rutz et al. (2014, 2015). This inland penetration might be related to the topography over that region (e.g., Columbia River basin). The contributions (i.e., percentages) of the AR T-IVT to the total T-IVT along the coast in the three reanalyses similar in magnitude and spatial structure. This part of water vapor transport (AR T-IVT) contributes 20%–35% of the total T-IVT along the coast; however, it always directly contributes to the precipitation over the western U.S. Therefore, in this study we focus on the prediction skill of AR T-IVT rather than total T-IVT. The overall differences for AR T-IVT (Figures 2d–2f) among the three reanalyses are generally consistent with that for total T-IVT (Figures 1d–1f). Meanwhile, the percentage difference for AR T-IVT along the coast between CFSR/MERRA2 and ERA5 is slightly larger than that for total T-IVT, indicating large uncertainties in AR activity near the coast across the reanalyses.

For the T-IVT not associated with ARs (Figure S1 in Supporting Information S1), the spatial patterns are very close to the total T-IVT in all three reanalyses. Meanwhile, the distribution of differences in T-IVT not associated with ARs between different reanalyses are also similar with the differences in total T-IVT. It indicates that the three reanalyses have a good agreement in large-scale background of T-IVT.

Overall, the difference of total and AR T-IVT between different reanalyses over the eastern North Pacific and the western U.S. is relatively small (around or below 5% over most areas), although the difference may reach 10% over some specific regions. In this study, ERA5 is used as the primary verification data since previous studies
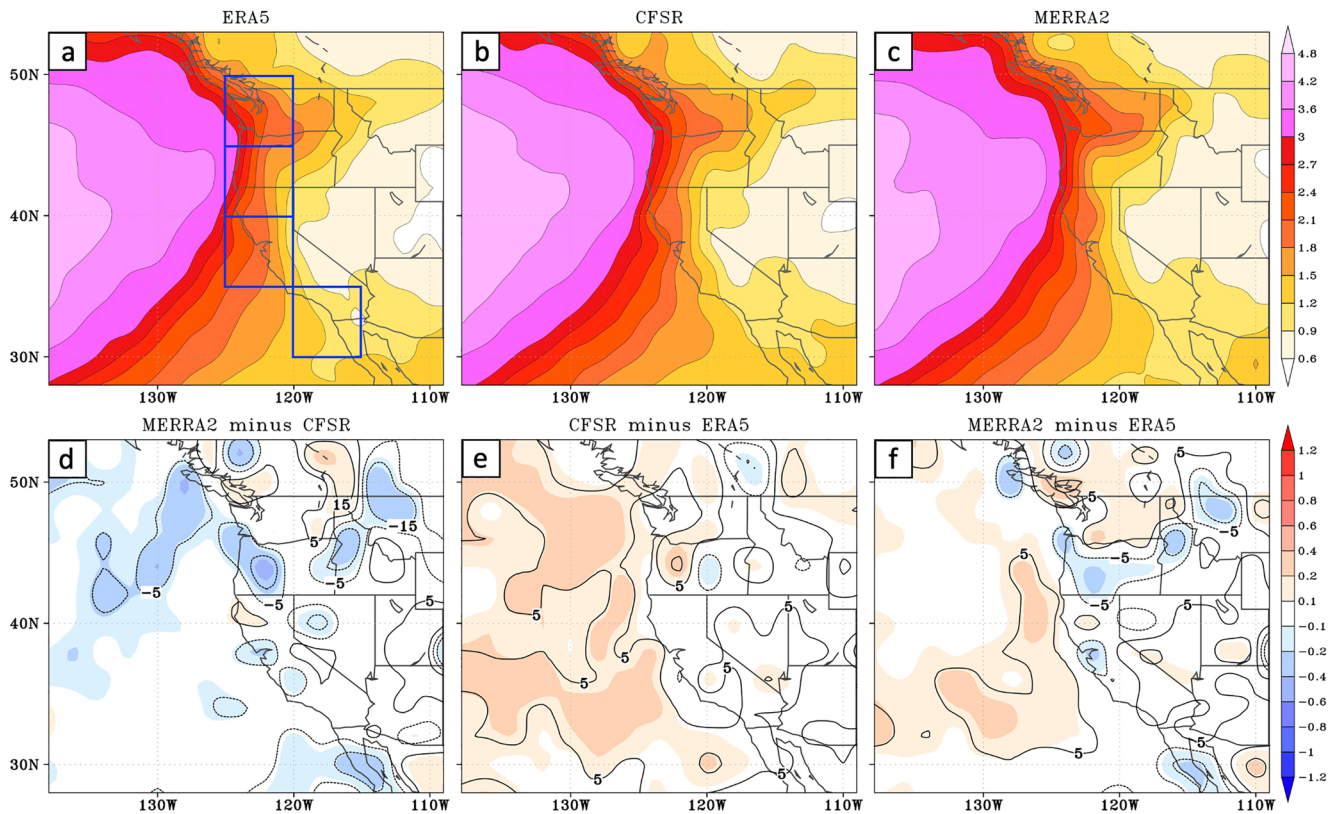
Figure 2. As in Figure 1, but for the T-IVT associated with ARs. The four blue boxes along the U.S. West Coast in (a) are used to calculate the area-mean AR T-IVT for Figures 10 and 11.

suggest that ERA5 performs relatively well when verified with various observations (e.g., Cobb et al., 2021). However, if any of the skill metrics in this study are sensitive to the choice of reanalysis data set, we report these sensitivities in the corresponding sections.

3.2. Model Biases

The climatological mean AR T-IVT in all four models has a similar spatial pattern as the reanalyses for weeks 1–4 lead time (contours in Figure 3), which implies that all the models can capture the overall distribution of water vapor transport associated with ARs over the eastern North Pacific. However, the models have different biases in AR T-IVT over different regions and at different lead times. The colors in Figure 3 show the model bias for weeks 1–4 lead time with respect to ERA5. In NCEP, the negative bias of AR T-IVT over the inland region is relatively small and persistent from week-1 to week-4 lead (Figures 3a, 3e, 3i, and 3m). Meanwhile, NCEP has a negative bias over the north side and a positive bias over the southeast side of the AR T-IVT maximum, indicating a shift of the maximum to the southeast. This shift increases slightly with the lead time. GMAO has a similar southeastward shift of the AR T-IVT maximum, and the shift increases with lead time more substantially than in NCEP (Figures 3d, 3h, 3l, and 3p). The positive bias over the southeast side of AR T-IVT maximum extends from the low-latitude ocean to the west coast of U.S., and further to the inland region (Nevada and Utah) at week-3 and week-4 lead in GMAO. This southeastward shift of the AR T-IVT maximum might be related to bias of the North Pacific storm track since extratropical cyclones have a great impact on the position of ARs (Z. Zhang, Ralph, & Zheng, 2019). Different from NCEP and GMAO, ECMWF has negative biases in AR T-IVT over most areas of the domain at week-1 lead (Figure 3b). The negative biases become larger at week-2 lead and reach a peak at week-3 lead (Figures 3f, 3j, and 3n). This is consistent with the results in DeFlorio et al. (2019), which shows that ECMWF underpredicted the percentages of weeks with AR days over most areas of the western U.S. through lead time weeks 1–4. ECCC has a negative bias over the continent and this negative bias does not change much with lead time (Figures 3c, 3g, 3k, and 3o). However, the weak negative bias over the southeast
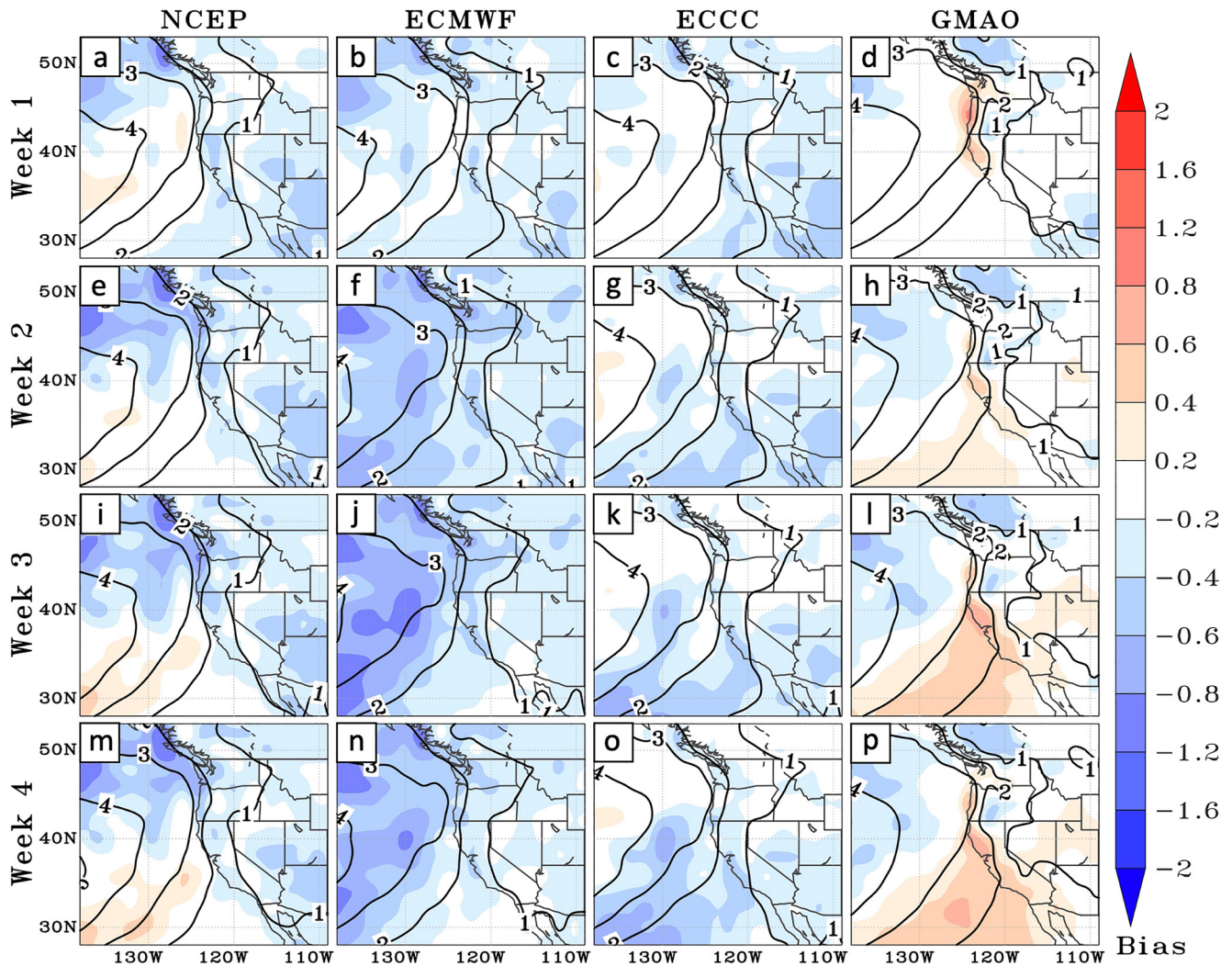
Figure 3. (a–d) Climatological mean AR T-IVT (contours, 10^7 kg m^{-1} per week) and the bias (colors, 10^7 kg m^{-1} per week) of the four models (NCEP, ECMWF, ECCC, and GMAO) with respect to ERA5 at week-1 lead time during the cool seasons of the hindcast periods. (e–h), (i–l), and (m–p) are the same as (a–d) but for week-2, week-3, and week-4 lead times, respectively.

side of the AR T-IVT maximum becomes larger with the increase of lead time in ECCC. DeFlorio et al. (2019) found that ECCC underpredicted the percentage of weeks with AR days (the frequency of ARs) across the entire domain through week–1 to week–4 lead. But in this study, we found that the bias of AR T-IVT is neutral in the AR T-IVT maximum area over the ocean, which indicates that the ECCC may overpredicted the intensity of the ARs there. Overall, AR T-IVT is determined by the duration and intensity (IVT, as a function of horizontal wind and specific humidity) of ARs. The biases of AR T-IVT might be related to some inherent model biases, such as the bias in subtropical jets, blocking, etc., in each model. Quinting and Vitart (2019) found that most of the S2S Project models underestimated the atmospheric blocking frequency over the North Atlantic but overestimated the blocking frequency over the eastern North Pacific at S2S time scale due to the models' bias in Rossby Wave decay. The overestimated blocking in the eastern North Pacific can be an important reason for the overall negative bias in the AR T-IVT (less or weaker ARs) in weeks 2–4 lead time over this region in NCEP, ECCC, and ECMWF.

As discussed in Section 3.1, there are differences in the magnitude of the AR T-IVT in the three reanalyses, especially over some specific regions. As a result, the model biases of the four models have some uncertainties using different reanalysis data sets as a reference. Figure 4 demonstrates the difference in the model bias of AR T-IVT when different reanalyses are used at week-3 lead. In Figures 4a–4d, the colors show the model bias with respect with ERA5, which are the same with the bias (colors) in Figures 3i–3l at week-3 lead. The bias in
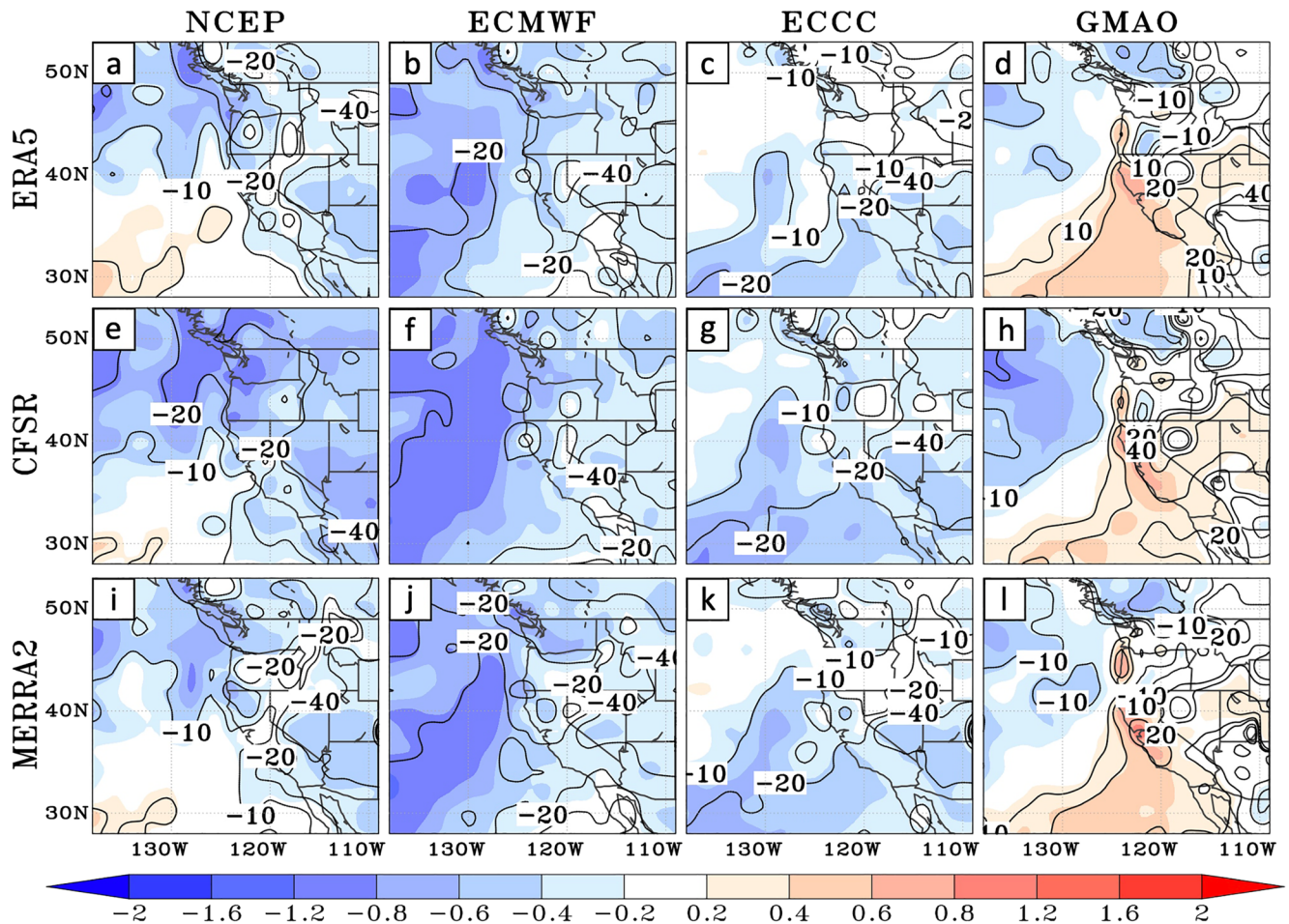
Figure 4. (a–d) The model bias of climatological mean AR T-IVT (colors, 10^7 kg m^{-1} per week) and the percentage bias (contours, %) of the four models with respect to ERA5 at week-3 lead time during the cool seasons of the hindcast periods. (e–h) and (i–l) are the same as (a–d) but for the model bias with respect to CFSR and MERRA2, respectively.

Figures 4e–4h and 4i–4l are the same model bias but using CFSR and MERRA2 as a reference respectively. If CFSR is used as the reference data instead of ERA5, the negative bias in all models becomes stronger, especially over the eastern North Pacific (Figures 4a–4h), because CFSR has 5%–10% stronger AR T-IVT over the ocean than ERA5 (Figure 2e). As a result, ECMWF and ECCC have negative biases over most areas of the domain. The negative bias in NCEP and GMAO is also enhanced if CFSR is used as a reference. However, GMAO still has a clear southeastward shift of the AR T-IVT maximum. The magnitude of the model bias with respect to MERRA2 (Figures 4i–4l) is generally between the magnitude of the biases with respect to ERA5 and CFSR. Overall, while using different reanalyses as references does exhibit some uncertainties in model biases, especially in the magnitude of model bias. The results have good agreements in the sign of model bias in many regions over the domain, including the negative biases in the western U.S. and high latitudes of the eastern North Pacific in NCEP, negative biases over most areas of the domain in ECWMF, negative biases at the lower latitudes in ECCC, and the southeastward shift of the AR-TIVT maximum in GMAO.

The differences of model biases due to different reference data at weeks 1, 2, and 4 lead (not shown) are similar to the results at week-3 lead.

3.3. Root-Mean-Square Error

Figure 5 shows the RMSE of the weekly AR T-IVT using ERA5 as the reference data set at weeks 1–4 lead time for the four models. The spatial patterns of the RMSE in the four models are in general consistent. The RMSE
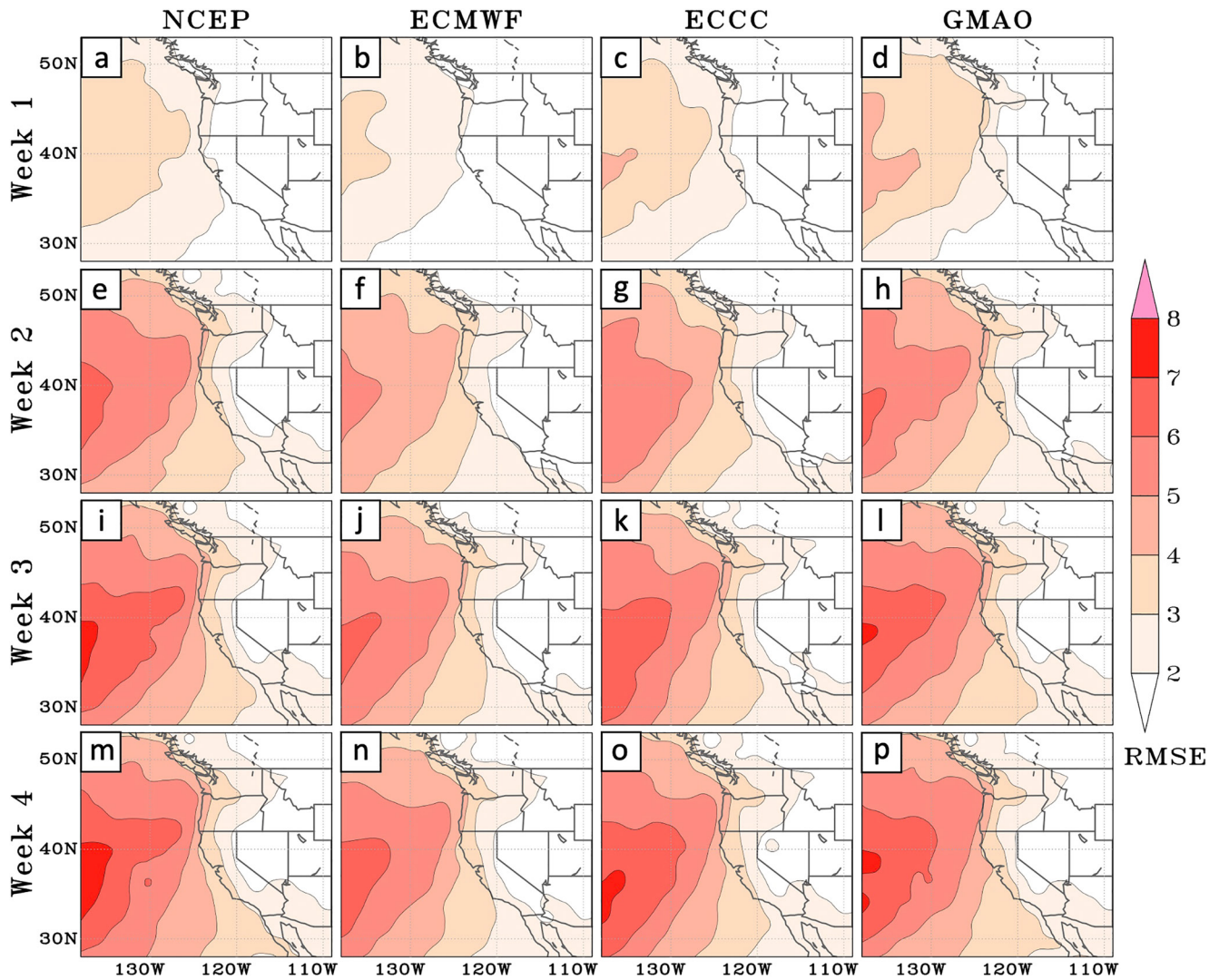
Figure 5. (a–d) RMSE of the AR T-IVT (colors, 10^7 kg m^{-1} per week) of the four models with respect to ERA5 at week-1 lead time during the cool seasons of the hindcast periods. (e–h), (i–l), and (m–p) are the same as (a–d) but for week-2, week-3, and week-4 lead times respectively.

is at a maximum around 40°N over the eastern North Pacific Ocean and decreases from the ocean to the coastal and inland regions, which is also consistent to the spatial pattern of the climatological mean AR T-IVT (contours in Figure 3). That is because RMSE is highly influenced by the climatological mean AR T-IVT. The RMSE is relatively small at week-1 lead in all models. At week-2 lead, the RMSE increases substantially, and then from week-2 to week-4 lead the increase of RMSE is relatively small. For example, along the coast of Washington, Oregon, and North California, the RMSE increases from $\sim 2.5 \times 10^7$ kg m^{-1} at week-1 lead to $\sim 4.0 \times 10^7$ kg m^{-1} at week-2 lead, and then to $\sim 4.5 \times 10^7$ kg m^{-1} at week-4 lead in NCEP. Overall, ECMWF has relatively smaller RMSE over the eastern North Pacific compared to the other three models at all lead times.

To exclude the impacts of climatological mean AR T-IVT, we normalized RMSE with the standard deviation of AR T-IVT at each grid cell (Figure S2 in Supporting Information S1). The normalized RMSE is relatively small (below or around 1 standard deviation) over the eastern North Pacific and the West Coast in the models at week-1 lead, and then increases with the lead time. Different from the original RMSE, the normalized RMSE has a maximum located over the southeast side of the maximum climatological mean AR T-IVT in all models. This spatial pattern is due to the low AR T-IVT values and thus low variance/standard deviation there; meanwhile, it indicates that the models have relatively large errors at the southeast edge of the maximum AR T-IVT region (southeast edge of AR activities).
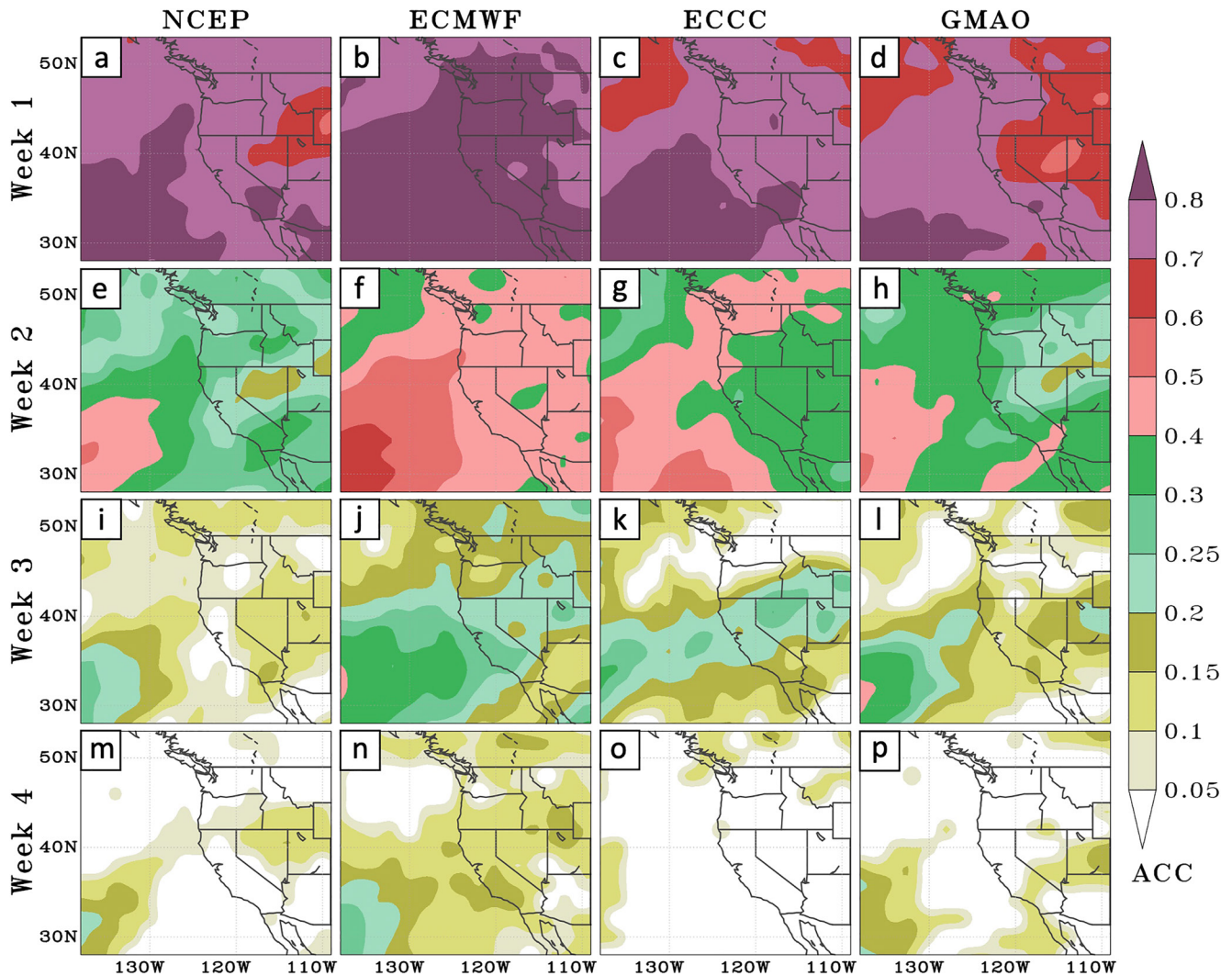
Figure 6. (a–d) Anomaly correlation coefficients (ACCs) of the AR T-IVT for the four models with respect to ERA5 at week-1 lead time during the cool seasons of the hindcast periods. (e–h), (i–l), and (m–p) are the same as (a–d) but for week-2, week-3, and week-4 lead times, respectively. Only ACC values at 95% confidence level based on a 1,000-resampling bootstrap statistical significance test are plotted.

Compared to the RMSE based on ERA5, the RMSEs calculated using CFSR and MERRA2 have similar spatial structure but are slightly higher in magnitude (not shown).

3.4. Anomaly Correlation Coefficient

We calculated the ACC of weekly AR T-IVT in the four models from week-1 to week-4 lead time using ERA5 as a reference (Figure 6). As expected, the ACC of AR T-IVT at week-1 lead is very high, ranging from 0.6 to nearly 0.9 over the eastern North Pacific and the western U.S. in all four models (Figures 6a–6d). At week-2 lead, the ACC in all models decreases substantially and the differences in ACC values across different models become larger (Figures 6e–6h). In ECMWF, the ACC is still above 0.4 over the western U.S. and is even higher (up to nearly 0.7) over most of the eastern North Pacific. The ACC in the other three models is lower than in ECMWF.

The ACC values decrease further at week-3 lead but are still statistically significant over many regions in the models (Figures 6i–6l). In all four models, the highest values of ACC are concentrated over the southwestern part of the domain (i.e., 28°N–40°N and 125°W–138°W) and extend northeastward into the continent. The high ACC values at the lower latitudes over the ocean might be related to the predictability source at the tropical and subtropical regions. In ECMWF and ECCC, the highest ACC values over the continent are ~0.30 and ~0.25

respectively, which are located around the California region and extend further inland. In NCEP and GMAO, the highest ACC values (~0.15 and ~0.20) over the continent are mainly located over Southern California and Utah, respectively. Overall, ECMWF has statistically significant ACC values at week-3 lead over the whole domain, and these values are higher than the ACC values in the other models. At week-4 lead, only ECMWF still has some statistically significant ACC values over of the western U.S., but those ACC values are relatively small (0.1–0.2) and explain about 2%–4% of the variance. NCEP and GMAO have low ACC values over limited areas; and ECCC has nearly no skill over the entire western U.S. region.

ACC values were also calculated using CFSR and MERRA2 as reference data. Figure S3 in Supporting Information S1 shows the comparison between ACCs using different reanalyses as reference data at week-3 lead. The difference in ACC values is generally below 0.05 and located within some small areas. For example, in GMAO over the northeastern Utah, the ACC with CFSR as a reference is slightly lower (<0.05) than the ACC with ERA5 or MERRA2 as a reference (Figures S3d, S3h, and S3l in Supporting Information S1). Overall, there were few significant differences in ACCs with different reanalyses as reference data. That is because the anomalies of AR T-IVT in the three reanalyses are highly correlated, although they have some differences in the magnitude of the climatological mean AR T-IVT (Figure 2).

3.5. Brier Skill Score

As described in Section 2.3, BSS values are calculated for three AR T-IVT categories (no AR T-IVT, weak AR T-IVT, and strong AR T-IVT) to assess the models' probabilistic skill in predicting AR T-IVT with respect to the climatological forecast (Figures 7–9). Positive BSS indicate that the model skill exceeds that of the climatological forecast. Following DeFlorio et al. (2019), a 100-sample bootstrapping significance test was used. Note that 1 represents a perfect forecast. Only BSS values greater than 0 at the 95% confidence level are plotted in Figures 7–9. Overall, the BSS values of the no AR T-IVT (Figure 7) and strong AR T-IVT (Figure 9) categories are higher than the BSS values of the weak AR T-IVT category (Figure 8) in all models and across all lead times (weeks 1–4).

For the no AR T-IVT category (Figure 7), BSS values are above 0.3 over most of the region at week-1 lead for all four models, with a maximum located over the southwestern part of the eastern North Pacific. The same location of maximum BSS and maximum ACC indicates relatively high prediction skill over that region in all models. At week-2 and week-3 lead, the differences of BSS values in different models become larger. In NCEP, BSS values decrease substantially at week-2 lead, and the maximum values are still located over the ocean, and at week-3 lead, there is almost no skill over the whole domain. In GMAO, BSS values are relatively high over most of the open ocean, the eastern Washington and Oregon, and the southern British Columbia at week-2 lead; then at week-3 lead, the skill is mainly concentrated over the southwestern part of the eastern North Pacific and the Southern California. ECMWF and ECCC have BSS values >0.1 over the whole domain at week-2 lead. At week-3 lead, ECMWF has BSS values >0.1 over the lower-latitude (<40°N) open ocean and California, while ECCC has BSS values >0.1 over the ocean and some continental regions. At week-4 lead, the models do not have any meaningful skill over the continent, although ECCC has weak positive BSS values at a very small area along the West Coast. It indicates that the models do not have any skill beyond the climatological forecast for no AR T-IVT cases at week-4 lead time.

For the weak AR T-IVT category (Figure 8), BSS values are generally much lower than the BSS values for the no AR T-IVT category in all four models. Even at week-1 lead, BSS values for the weak AR T-IVT category are below 0.3 over most of the domain in all models. At week-2 lead, BSS values decrease quickly. At week-3 lead, although ECCC, ECMWF, and GMAO still have some skill, the spatial distribution is noisy. The skill disappears in all models at week-4 lead. The low BSS for the weak AR T-IVT category suggest that the models are struggling in forecasting weak ARs. This is likely because it is more difficult to detect weak ARs, which could cause some uncertainties for the low skill for the weak AR T-IVT cases.

BSS values for the strong AR T-IVT category (Figure 9) are similar to the BSS values for the no AR T-IVT category. BSS values for the strong AR T-IVT category are above 0.3 over most of the domain in all the four models at week-1 lead, and then the values decrease rapidly with increasing lead time. In NCEP, the high BSS values (>0.15) are mainly located over the ocean, and relatively low BSS values are present over limited continental areas at week-2 lead, and at week-3 and week-4 lead, low BSS values (around or below 0.15) exist over
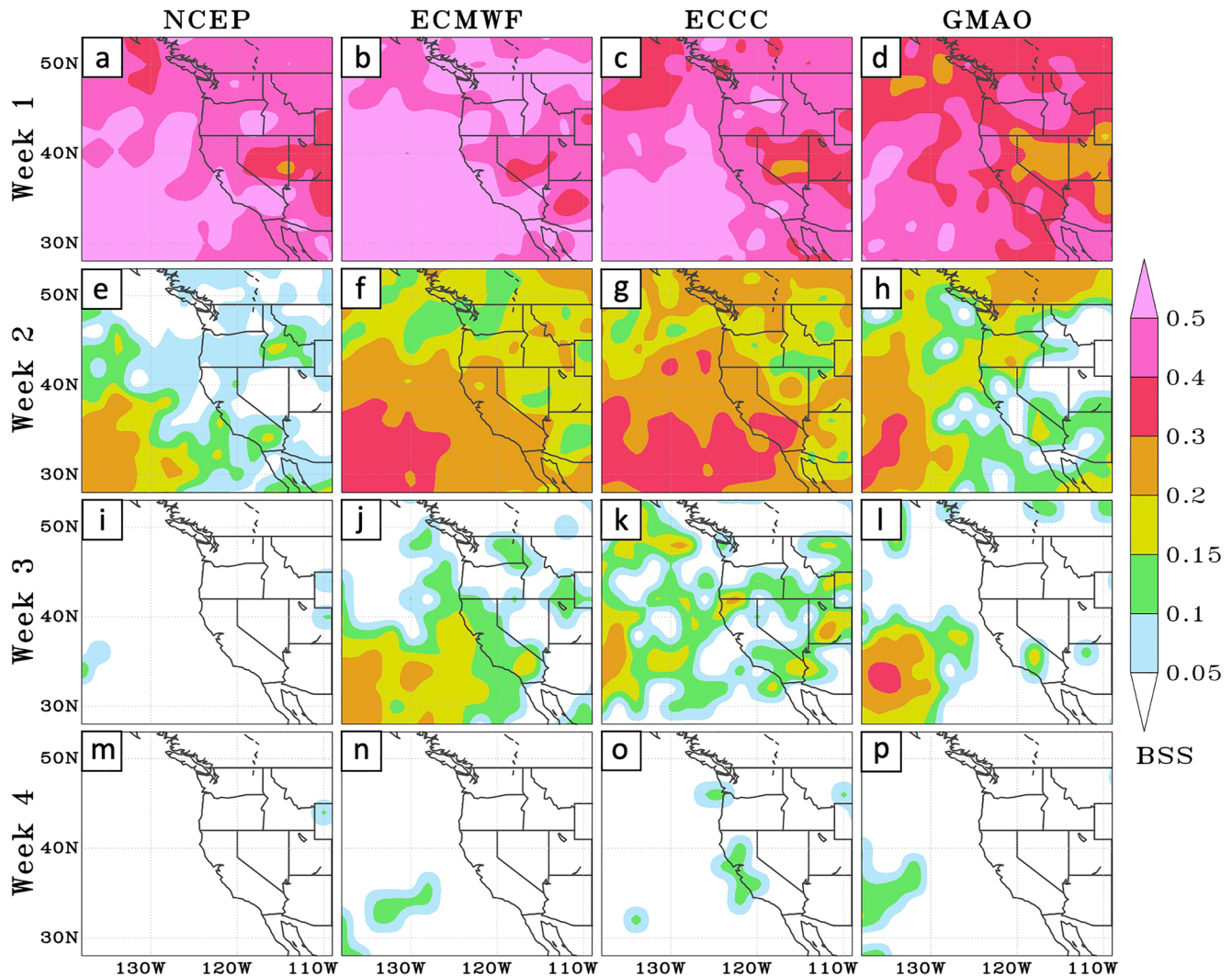
Figure 7. (a–d) Brier Skill Score of the no AR T-IVT category for the four models with respect to ERA5 at week-1 lead time during the cool seasons of the hindcast periods. (e–h), (i–l), and (m–p) are the same as (a–d) but for week-2, week-3, and week-4 lead times, respectively. Only BSS values that are statistically significant at the 95% confidence level are plotted.

the mountains in the inland western U.S. Due to the complex topography over that region and thus the uncertainties in AR detection there, the prediction skill might be sensitive to the AR detection methods. Further study is needed to understand the distribution of skill over that inland mountain region and test the impacts of different AR detection methods. In GMAO, BSS values >0.15 are mainly located over the eastern North Pacific, the U.S. West Coast, and the southern British Columbia at week-2 lead, and at week-3 lead, the skill is concentrated over the southwestern part of the eastern North Pacific, as well as the coastal regions of southern Oregon and northern California. In ECMWF, BSS values are above 0.15 over most area of the domain at week-2 lead, and at week-3 lead, positive BSS values are mainly located over the southwestern part of the eastern North Pacific and extend northeastward to California and further inland regions. In ECCC, BSS values are also above 0.15 over most area of the domain at week-2 lead, and at week-3 lead, positive BSS values are present over the ocean around 40°N, as well as across Southern California and the adjacent inland area. At week-4 lead, there are very few regions with positive BSS values, indicating that the models do not have much skill compared to the climatological forecast at this lead time. Overall, the relatively high prediction skill for AR T-IVT at the low latitudes in the North Pacific is consistent with the results in C. Zheng et al. (2019), which showed that the prediction skill of extratropical cyclones is mainly concentrated at the low latitudes over the ocean, given the substantial impacts of cyclones on ARs.
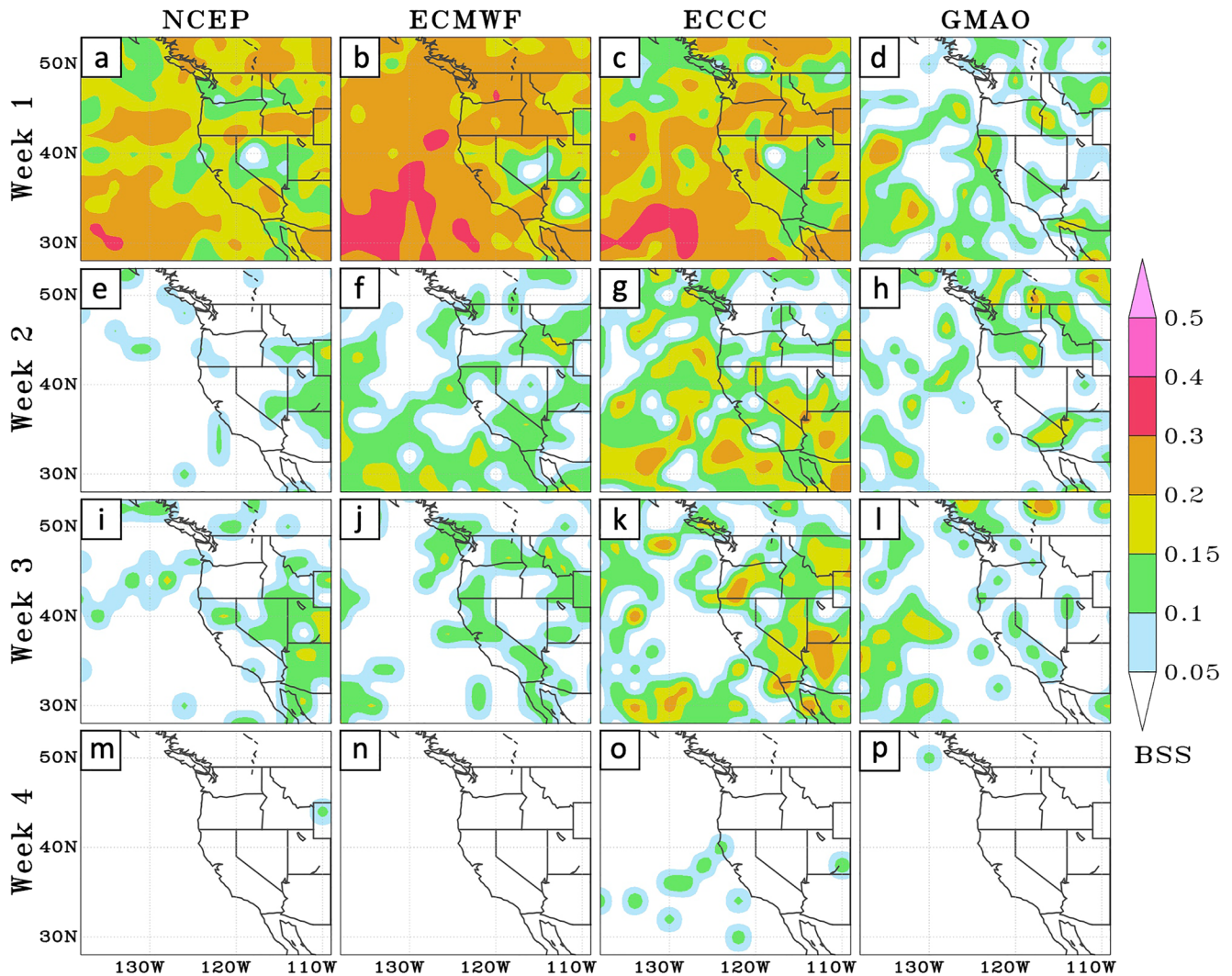
Figure 8. Same as Figure 7 but for the weak AR T-IVT category.

Generally, all models have higher BSS in predicting the no and strong AR T-IVT categories than in predicting the weak AR T-IVT category. The overall BSS skill in this study is consistent with the results in DeFlorio et al. (2019), who showed that NCEP, ECMWF, and ECCC hindcasts have relatively higher BSS skill in predicting the weeks with 0 or 3–7 AR days than the weeks with 1–2 AR days. It implies that the models might be struggling in predicting the weak AR events (e.g., AR landfalling time, AR duration) and do not have much higher skill in prediction those weak events than climatology forecasts. Similar to ACC, using different reanalyses data sets as reference data does not change the results of BSS significantly.

3.6. Modulation of Prediction Skill by MJO

The impact of the MJO on the subseasonal prediction skill of AR T-IVT is examined over four 5° × 5° domains along the U.S. West Coast (blue boxes in Figure 2a). These four domains were chosen following DeFlorio et al. (2019) to represent useful AR landfall domains that span the entire western U.S. coastline, including Washington (125°W–120°W, 45°N–50°N), Oregon-North California (125°W–120°W, 40°N–45°N), Central California (125°W–120°W, 35°N–40°N), and South California (120°W–115°W, 30°N–35°N). The metric of ACC is used in exploring the impacts of MJO since ACC quantifies the model skill in forecasting the anomalies of AR T-IVT, which we are most interested in at S2S lead times.

Before examining the impacts of the MJO, the ACC of the area-mean weekly AR T-IVT in the four domains along the West Coast is calculated using ERA5 as a reference for the four models at week-1 to week-4 lead time
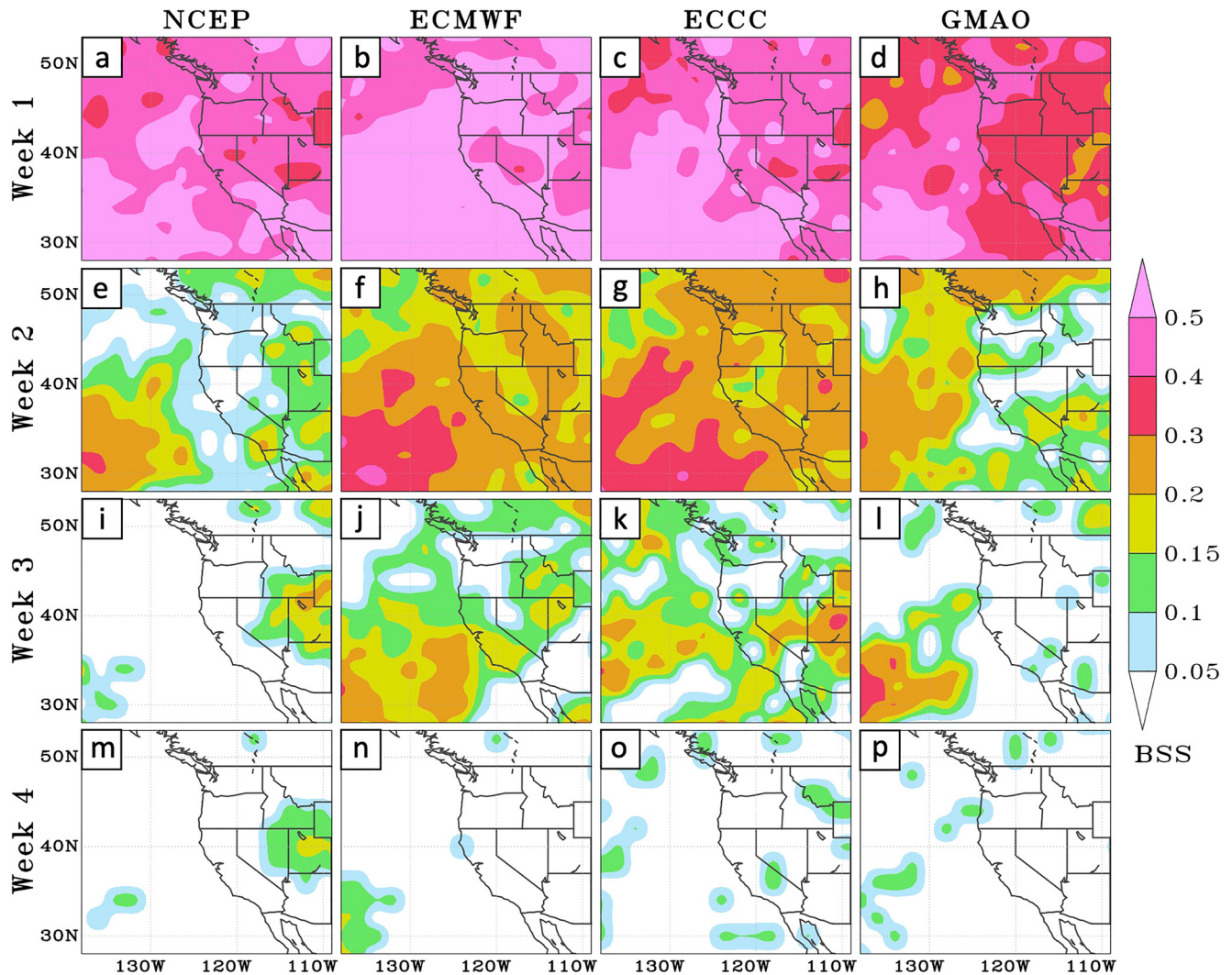
Figure 9. Same as Figure 7 but for the strong AR T-IVT category.

(Figure 10). At week-1 lead, all models have high ACCs in all four domains: 0.89–0.92 for ECMWF, 0.84–0.86 for ECCC, 0.83–0.85 for NCEP, and 0.79–0.83 for GMAO. At week-2 lead, the difference of the ACC values in different models becomes larger. In the Washington domain, ECCC (0.52) has a slightly (not significantly) higher ACC than ECMWF (0.48), while NCEP (0.28) is significantly lower than both ECCC and ECMWF. In the other three domains, ECMWF has higher skill than the other three models, and NCEP has a comparable skill with GMAO. Although the skill decreases substantially with lead time beyond the weather timescale (i.e., beyond week-2 lead time), all models still have positive ACC values that are significantly positive at week-3 lead, indicating that the models have some skill in forecasting weekly AR T-IVT anomalies along the U.S. West Coast beyond the weather timescale. ECMWF and ECCC have relatively higher ACC values in the domains of Central California, which is an extension of the high ACC values from the southwestern part of the eastern North Pacific Ocean to the continent (Figures 6j and 6k). At week-4 lead, only ECMWF still has positive ACC values that are significantly positive in all the four domains.

ACC values of the forecasts initialized in MJO phases 2 and 3, 4 and 5, 6 and 7, and 8 and 1 are calculated and compared with the ACC of the forecasts initialized across all days for each model and at weeks 1–4 lead time. MJO phases were combined as 2 and 3, 4 and 5, 6 and 7, and 8 and 1 so that each combined phase has more available forecast samples to conduct more robust statistical significance testing. A 1,000-resampling bootstrap statistical significance test was conducted to estimate 90% confidence intervals. The 90% confidence interval was utilized instead of 95% due to the relatively small sample size of the MJO cases. Table 2 summarizes the
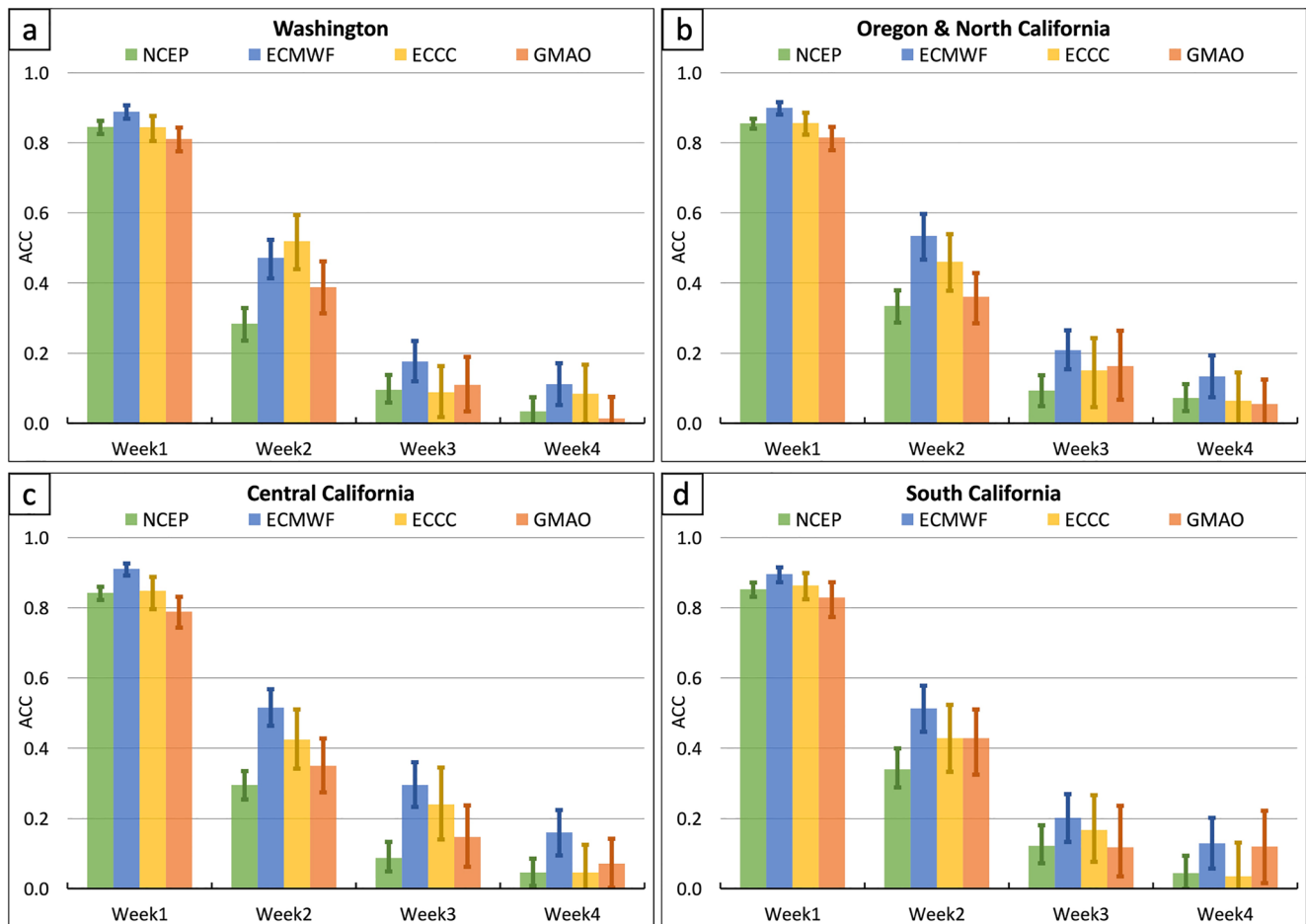
Figure 10. (a) Anomaly correlation coefficients (ACCs) of the area-mean AR T-IVT within the domain of Washington for the four models (different colors) and at different forecast lead times (weeks 1–4). The vertical bars show the 95% confidence intervals. (b–d) are the same as (a) but for (b) Oregon and North California, (c) Central California, and (d) South California.

forecast sample size of each model in each combined MJO phase. Only active MJO conditions (RMM > 1) are considered in this study.

Overall, at week-1 and week-2 lead the impact of MJO on the prediction skill of AR T-IVT is relatively small (Figures S4 and S5 in Supporting Information S1). At week-1 lead, although different MJO phases have some impacts on the prediction skill, nearly all of them are not statistically significant (Figure S4 in Supporting Information S1). At week-2 lead, ACC values of the NCEP forecasts initialized in MJO phases 4 and 5 are significantly higher than the ACC values of all days in the Central and Southern California domains. For the other MJO phases, the models do not show any significantly higher ACC values compared to the skill for all days.

The impact of the MJO on prediction skill (ACC) becomes larger at week-3 lead, especially for the southern part of the West Coast, including Central and South California domains (Figure 11). This is consistent with many previous studies that highlight the substantial impacts of MJO on the subseasonal variability in both tropical atmosphere and extratropical weather and climate (e.g., Waliser et al., 2003; C. Zhang, 2005; Zhou et al., 2021). However, the influence of MJO has uncertainties across different models. Specifically, the skill is significantly increased over Central California in the hindcasts initialized in MJO phases 6 and 7 for GMAO and ECCC, and in MJO phases 4 and 5 for NCEP and ECMWF. On the other hand, the skill is decreased over

Table 2
Number of Forecasts Initialized in Particular Madden–Julian Oscillation (MJO) Phases for the Four Models

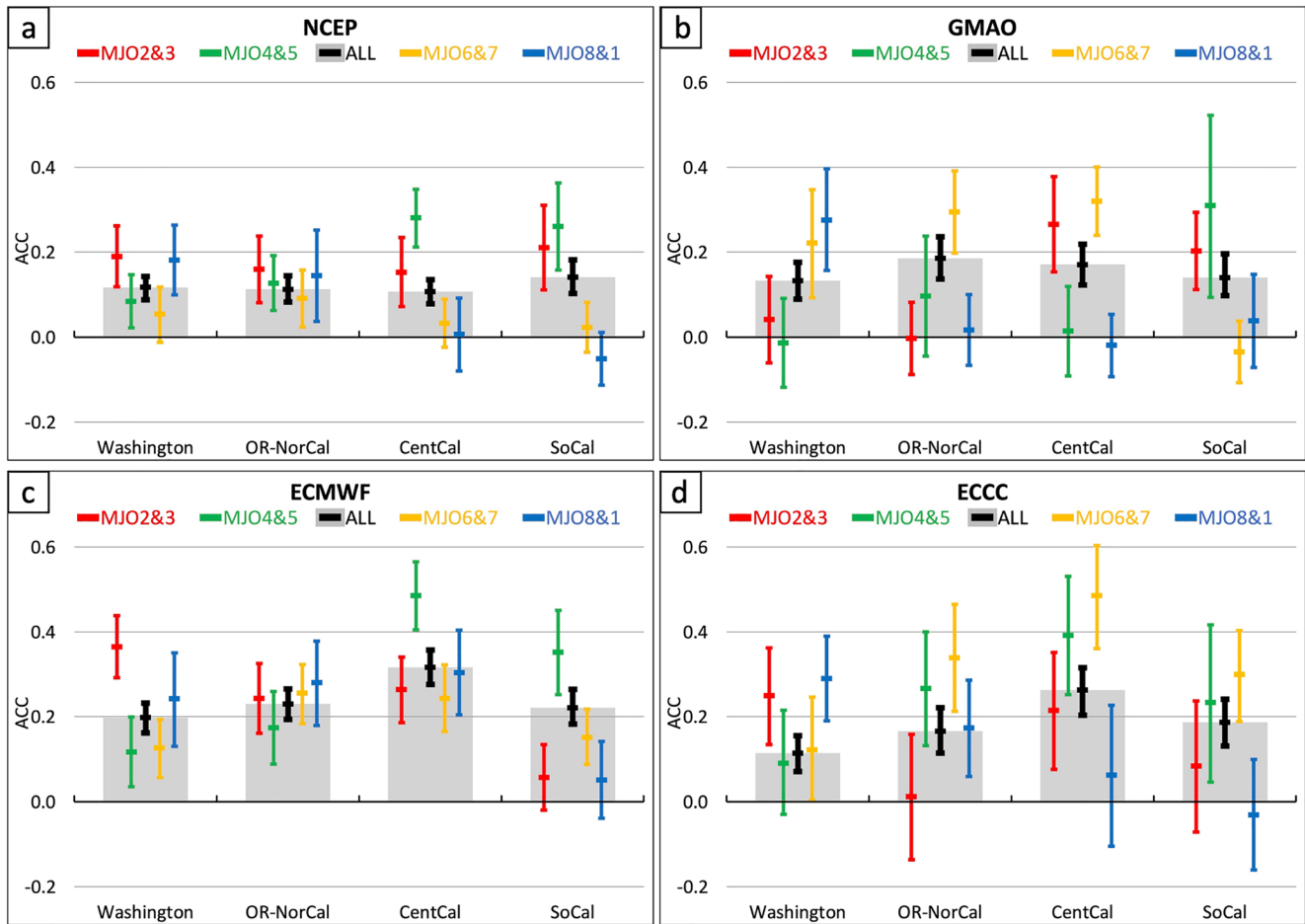| MJO phase	Number of forecasts			
	NCEP	ECMWF	ECCC	GMAO
Phases 2–3	271	124	65	80
Phases 4–5	310	141	69	91
Phases 6–7	327	160	76	104
Phases 8–1	217	103	52	63

Figure 11. (a) Anomaly correlation coefficients (ACCs) of the area-mean AR T-IVT from the NCEP hindcast system at week-3 lead time (gray histogram). The vertical bars show the 90% confidence intervals for the hindcasts initialized in all days (black), and for initialization days in MJO phases 2–3 (red), phases 4–5 (green), phases 6–7 (yellow), and phases 8–1 (blue). (b–d) are the same as (a) but for the GMAO, ECMWF, and ECCC hindcast systems respectively.

Southern California in the hindcasts initialized in MJO phases 8 and 1 for NCEP, ECMWF and ECCC, in MJO phases 6 and 7 for NCEP and GMAO, and in MJO phases 2 and 3 for ECMWF. In addition, the prediction skill is significantly increased in MJO phases 2 and 3 for ECMWF and phases 8 and 1 for ECCC over Washington, while the skill is significantly decreased in MJO phases 2 and 3 and 8 and 1 for GMAO over Oregon-Northern California. Small sample size (Table 2) could be one reason that cause the large uncertainties. For example, ECCC has much larger uncertainties (longer confidence intervals in Figure 11d) for the MJO impacts than NCEP (Figure 11a), which has a relatively larger sample size. In addition, difference in the models' ability to predict the MJO and reproduce the associated teleconnections could be a main factor responsible for the large uncertainties in the impacts of MJO on the prediction skill of AR T-IVT. Recent studies (e.g., Stan et al., 2022; Vitart et al., 2017) found that the models from the S2S Project have some advantages in predicting MJO, but biases in the predictions and the MJO teleconnections still exists, especially beyond week-2 lead time.

The impact of some specific MJO phases varies across locations (latitudes). For example, the prediction skill in MJO phases 4 and 5 increases from the Washington to the Central California domain in NCEP, ECMWF, and ECCC (green bars in Figures 11a, 11c, and 11d); and the prediction skill in MJO phases 8 and 1 decreases from the north to the south along the West Coast in NCEP and ECCC (blue bars in Figures 11a and 11d).

Different MJO phases also have some impacts on the prediction skill at week-4 lead. However, ACC values of the domain-averaged AR T-IVT are relatively low (<0.2) in most MJO phases and domains at week-4 lead. Thus, this is not shown and discussed here. The prediction skill (ACC) of domain-averaged AR T-IVT was also examined using CFSR and MERRA2 as reference data. The choice of reanalysis data set does not have much impact on the results shown in Figure 11.

The impacts of different MJO phases on prediction skills in this study are different from the results in DeFlorio et al. (2019), who found that the skill along the western U.S. is increased in hindcasts initialized during MJO Phases 1 and 8, and the skill is decreased over California in hindcasts initialized during MJO Phase 4. In addition to the different AR quantifies (AR T-IVT and AR occurrence), it is noteworthy that in this study the skill is quantifications using ACC of the area-mean AR T-IVT, while the skill in DeFlorio et al. (2019) is based on a Relative Operating Characteristic-like diagrams (comparing AR hit rate and false alarm rate). These differences might be important factors responsible for the different results.

4. Summary

In this study, the subseasonal (weeks 1–4 lead time) prediction skill of water vapor transport associated with atmospheric rivers (AR T-IVT) over the western U.S. and the eastern North Pacific was evaluated in four dynamical model hindcast data sets from NCEP, ECMWF, ECCC, and NASA GMAO (Table 1). Three reanalysis data sets (ERA5, CFSR, and MERRA2) were used as reference data to examine the impacts of different reference data on the evaluation. The model prediction skill of weekly AR T-IVT was verified using a number of deterministic and probabilistic skill metrics (bias, RMSE, ACC, and BSS). Additionally, the influence of the MJO on the prediction skill of AR T-IVT was explored over four domains along the U.S. West Coast (Washington, Oregon-North California, Central California, and Southern California). The main results are summarized below:

1. Using ERA5 as a reference, ECMWF and ECCC have weak negative biases of AR T-IVT over most of the eastern North Pacific and the western U.S. at week-1 lead. These biases increase with lead time (Figure 3). In GMAO and NCEP the AR T-IVT maximum has a southeastward shift, and the shift was enhanced with the increase of lead time in GMAO.
2. The hindcasts of the four S2S systems have similar spatial patterns of RMSE, with a maximum around 40°N over the eastern North Pacific and a lower RMSE over the inland area of the western U.S. (Figure 5), which is generally proportional to the climatological mean AR T-IVT.
3. ACC values of AR T-IVT are above 0.6 over almost the whole domain in all models at week-1 lead, then decrease with lead time substantially (Figure 6). However, at week-3 lead time statistically significant ACC values are still present with a maximum at the lower latitudes (<40°N) of the eastern North Pacific and extend northeastward to the continent, resulting a relatively high ACC/skill over the California region in ECMWF and ECCC.
4. BSS values for the no and strong AR T-IVT categories are higher than those for the weak AR T-IVT category in all hindcast data sets at weeks 1–3 lead (Figures 7–9). This implies that the models have higher skill in forecasting the no and strong AR T-IVT cases than the weak cases. At week-3 lead, ECMWF, ECCC, and GMAO still have significantly higher skill for the no and strong AR T-IVT categories than the climatological forecasts (BSS > 0) over limited areas.
5. The impact of the MJO on the prediction skill at week-3 lead time is larger than for week-1 and week-2 lead. At week-3 lead, the modulation of prediction skill by MJO mainly occurs over the central and Southern California, but with large uncertainties across models.
6. Differences of total T-IVT and AR T-IVT between the three reanalysis data sets (ERA5, CFSR, and MERRA2) are relatively small (around or below 5%) over most regions of the investigated domain, although they may reach nearly 10% over some specific regions. As a result, the model biases of AR T-IVT (especially for the magnitude) have some uncertainties when using different reanalysis as a reference (Figure 4).

5. Conclusion and Discussions

The prediction skill metrics of ACC, RMSE, and BSS are not significantly affected by the choice of reference reanalysis data, suggesting that the three reanalysis data sets have a generally good agreement in representing an AR and its vicinity. However, the uncertainties in reference reanalysis products do impact the model bias assessment of AR T-IVT, albeit overall not substantially. This may have some impacts on the regional research and applications (e.g., AR T-IVT bias correction) over some specific area.

This study demonstrates that the dynamical models have significant prediction skill on weekly AR T-IVT at week-3 lead time, which is beyond the weather forecast time scale (0–14-day lead). The prediction skill at week-3 lead mainly extends from the lower latitudes (<40°N) of the eastern North Pacific to the California coastal area

and the adjacent inland region. Although there are positive skill values over some other small areas, they are relatively weak and not spatially extensive or coherent. At week-4 lead, only ECMWF has some ACC skill over the western U.S. and the eastern North Pacific. It is noteworthy that the latest operational model versions of ECCC and ECMWF have been upgraded and may have a better prediction skill.

We found that the models have higher skill (BSS) in forecasting no and strong AR T-IVT events than weak AR T-IVT events. This is consistent with the results from DeFlorio et al. (2019), which shows that the NCEP, ECMWF, and ECCC models have higher skill in forecasting the weeks with 0 or 3–7 AR days than the weeks with 1–2 AR days, although their AR activity level depends only on the number of AR days and does not include the AR intensity. Taken together, results from this study and from DeFlorio et al. (2019) indicate that the models are struggling with the prediction of ARs with a short duration and/or weak intensity.

The MJO has a greater influence on the prediction skill of AR T-IVT at week-3 lead than at weeks 1–2 lead, which is consistent with previous studies that suggest the importance of the MJO in modulating subseasonal prediction skill of ARs and precipitation (DeFlorio et al., 2019; Pan et al., 2019). The impact of MJO is mainly concentrated over central and Southern California. However, there are large uncertainties in the MJO's influence across models, which might be caused by the different model performance in predicting the MJO and the relevant teleconnections. Huang et al. (2021) showed that ENSO can modulate the MJO-AR relationship. Therefore, the different hindcast periods (thus different ENSO phases) in different models (Table 2) may cause some uncertainty in the MJO's impacts on the prediction skill of AR T-IVT shown in Figure 11. Some studies (e.g., Arcodia et al., 2020; Guan et al., 2012; Zhou et al., 2021) emphasized the large impacts of MJO on the AR activities and the precipitation. Given the dominant role in the subseasonal variability, MJO could be a key source of predictability for AR T-IVT at subseasonal time scale. However, further studies are needed to understand the uncertainties across models and the underlying physical mechanism of the MJO's impacts on the prediction skill of AR T-IVT.

Previous studies (Collow et al., 2022; O'Brien et al., 2022) showed that the uncertainties from different AR detection methods could be larger than the uncertainties due to different reanalysis data sets or models. In this study, we used one AR detection method, which might cause some uncertainties in the prediction skills shown in this study. Further studies with different AR detection methods could be useful to quantify those uncertainties.

The Center for Western Weather and Water Extremes (CW3E) and the NASA Jet Propulsion Laboratory (JPL) have closely collaborated to develop experimental S2S AR forecast products for the western U.S. since November 2017. In addition to the AR activity forecast introduced in DeFlorio et al. (2019, their Figure 12), the near real-time forecasts of weekly AR T-IVT and AR Scale (Ralph et al., 2019) are generated for 85 locations over the western U.S. at week-3 lead time. These forecasts are created based on the real-time S2S forecast data from the four model centers (NCEP, ECMWF, ECCC, and NASA GMAO). The multi-model prediction skill assessment of AR-related water vapor transport in this study provides a baseline for the CW3E/JPL experimental S2S AR forecast products and their future development.

Data Availability Statement

Hindcast data sets of NCEP, ECCC, and ECMWF can be accessed on the ECMWF S2S reforecasts website portal (https://apps.ecmwf.int/datasets/data/s2s/levtype=sfc/type=cf/) and hindcast data set of GMAO is obtained from the NASA GMAO group (https://gmao.gsfc.nasa.gov/gmaoftp/gmaofcst/subx/GEOS_S2S_V2.1_fcst/AtmRiv/hindcast/). The CFSR reanalysis data set can be accessed on the UCAR Research Data Archive website (https://rda.ucar.edu); ERA5 reanalysis data set can be accessed on the ECMWF Climate Data Store website (https://cds.climate.copernicus.eu/#!/search?text=ERA5&type=dataset); and the MERRA2 reanalysis data set can be accessed on the NASA GES DISC website (https://disc.gsfc.nasa.gov). The AR detection code (Guan, 2021) is available via the Global Atmospheric Rivers Dataverse (https://dataverse.ucla.edu/dataverse/ar).

References

Albers, J. R., Butler, A. H., Breeden, M. L., Langford, A. O., & Kiladis, G. N. (2021). Subseasonal prediction of springtime Pacific–North American transport using upper-level wind forecasts. *Weather and Climate Dynamics*, *2*(2), 433–452. https://doi.org/10.5194/wcd-2-433-2021

Arcodia, M. C., Kirtman, B. P., & Siqueira, L. S. (2020). How MJO teleconnections and ENSO interference impacts U.S. precipitation. *Journal of Climate*, *33*(11), 4621–4640. https://doi.org/10.1175/jcli-d-19-0448.1

Baggett, C. F., Barnes, E. A., Maloney, E. D., & Mundhenk, B. D. (2017). Advancing atmospheric river forecasts into subseasonal-to-seasonal time scales. *Geophysical Research Letters*, *44*(14), 7528–7536. https://doi.org/10.1002/2017gl074434

Baldwin, M. P., Gray, L. J., Dunkerton, T. J., Hamilton, K., Haynes, P. H., Randel, W. J., et al. (2001). The quasi-biennial oscillation. *Reviews of Geophysics*, *39*(2), 179–229. https://doi.org/10.1029/1999rg000073

Brunet, G., Shapiro, M., Hoskins, B., Moncrieff, M., Dole, R., Kiladis, G. N., et al. (2010). Collaboration of the weather and climate communities to advance subseasonal-to-seasonal prediction. *Bulletin of the American Meteorological Society*, *91*(10), 1397–1406. https://doi.org/10.1175/2010bams3013.1

Cao, Q., Shukla, S., DeFlorio, M. J., Ralph, F. M., & Lettenmaier, D. P. (2021). Evaluation of the subseasonal forecast skill of floods associated with atmospheric rivers in coastal western U.S. watersheds. *Journal of Hydrometeorology*, *22*(6), 1535–1552. https://doi.org/10.1175/jhm-d-20-0219.1

Cobb, A., Delle Monache, L., Cannon, F., & Ralph, F. M. (2021). Representation of dropsonde-observed atmospheric river conditions in reanalyses. *Geophysical Research Letters*, *48*(15), e2021GL093357. https://doi.org/10.1029/2021gl093357

Collow, A. B. M., Shields, C. A., Guan, B., Kim, S., Lora, J. M., McClenny, E. E., et al. (2022). An overview of ARTMIP's Tier 2 Reanalysis Intercomparison: Uncertainty in the detection of atmospheric rivers and their associated precipitation. *Journal of Geophysical Research: Atmospheres*, *127*(8), e2021JD036155. https://doi.org/10.1029/2021JD036155

Cordeira, J. M., & Ralph, F. M. (2021). A summary of GFS ensemble integrated water vapor transport forecasts and skill along the U.S. West Coast during water years 2017–2020. *Weather and Forecasting*, *36*(2), 361–377. https://doi.org/10.1175/waf-d-20-0121.1

DeFlorio, M. J., Ralph, F. M., Waliser, D. E., Jones, J., & Anderson, M. L. (2021). Better subseasonal-to-seasonal forecasts for water management. *EOS*, *102*. https://doi.org/10.1029/2021EO159749

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), 3039–3060. https://doi.org/10.1007/s00382-018-4309-x

DeFlorio, M. J., Waliser, D. E., Ralph, F. M., Guan, B., Goodman, A., Gibson, P. B., et al. (2019). Experimental subseasonal-to-seasonal (S2S) forecasting of atmospheric rivers over the western United States. *Journal of Geophysical Research: Atmospheres*, *124*(21), 11242–11265. https://doi.org/10.1029/2019jd031200

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

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The Modern-Era Retrospective Analysis for Research and Applications Version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/jcli-d-16-0758.1

Gibson, P. B., Waliser, D. E., Goodman, A., DeFlorio, M. J., Delle Monache, L., & Molod, A. (2020). Subseasonal-to-seasonal hindcast skill assessment of ridging events related to drought over the western United States. *Journal of Geophysical Research: Atmospheres*, *125*(22), e2020JD033655. https://doi.org/10.1029/2020jd033655

Guan, B. (2021). [Code] tracking atmospheric rivers globally as Elongated Targets (tARget), version 2. UCLA Dataverse. https://doi.org/10.25346/S6/PSP5N0

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*(24), 12514–12535. https://doi.org/10.1002/2015jd024257

Guan, B., & Waliser, D. E. (2019). Tracking atmospheric rivers globally: Spatial distributions and temporal evolution of life cycle characteristics. *Journal of Geophysical Research: Atmospheres*, *124*(23), 12523–12552. https://doi.org/10.1029/2019jd031205

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. https://doi.org/10.1175/jhm-d-17-0114.1

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hoskins, B. J., & Hodges, K. I. (2002). New perspectives on the Northern Hemisphere winter storm tracks. *Journal of the Atmospheric Sciences*, *59*(6), 1041–1061. https://doi.org/10.1175/1520-0469(2002)059<1041:npotnh>2.0.co;2

Huang, H., Patricola, C. M., Bercos-Hickey, E., Zhou, Y., Rhoades, A., Risser, M. D., & Collins, W. D. (2021). Sources of subseasonal-to-seasonal predictability of atmospheric rivers and precipitation in the western United States. *Journal of Geophysical Research: Atmospheres*, *126*(6), e2020JD034053. https://doi.org/10.1029/2020jd034053

Kim, H. M., Kim, D., Vitart, F., Toma, V. E., Kug, J. S., & Webster, P. J. (2016). MJO propagation across the Maritime Continent in the ECMWF ensemble prediction system. *Journal of Climate*, *29*(11), 3973–3988. https://doi.org/10.1175/jcli-d-15-0862.1

Lavers, D. A., Pappenberger, F., Richardson, D. S., & Zsoter, E. (2016). ECMWF extreme forecast index for water vapor transport: A forecast tool for atmospheric rivers and extreme precipitation. *Geophysical Research Letters*, *43*(22), 11–852. https://doi.org/10.1002/2016gl071320

Liu, X., Ma, X., Chang, P., Jia, Y., Fu, D., Xu, G., et al. (2021). Ocean fronts and eddies force atmospheric rivers and heavy precipitation in western North America. *Nature Communications*, *12*(1), 1–10. https://doi.org/10.1038/s41467-021-21504-w

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, *20*(2), 130–141. https://doi.org/10.1175/1520-0469(1963)020<0130:dnf>2.0.co;2

Lorenz, E. N. (1965). A study of the predictability of a 28-variable atmospheric model. *Tellus*, *17*(3), 321–333. https://doi.org/10.1111/j.2153-3490.1965.tb01424.x

Madden, R. A., & Julian, P. R. (1971). Detection of a 40–50 days oscillation in the zonal wind in the tropical Pacific. *Journal of the Atmospheric Sciences*, *28*(5), 702–708. https://doi.org/10.1175/1520-0469(1971)028<0702:doadoi>2.0.co;2

Mani, J. M., Lee, J. Y., Waliser, D., Wang, B., & Jiang, X. (2014). Predictability of the Madden–Julian Oscillation in the Intraseasonal Variability Hindcast Experiment (ISVHE). *Journal of Climate*, *27*(12), 4531–4543. https://doi.org/10.1175/JCLI-D-13-00624.1

Mariotti, A., Baggett, C., Barnes, E. A., Becker, E., Butler, A., Collins, D. C., et al. (2020). Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond. *Bulletin of the American Meteorological Society*, *101*(5), E608–E625. https://doi.org/10.1175/bams-d-18-0326.1

Mariotti, A., Ruti, P. M., & Rixen, M. (2018). Progress in subseasonal to seasonal prediction through a joint weather and climate community effort. *NPJ Climate and Atmospheric Science*, *1*(1), 1–4. https://doi.org/10.1038/s41612-018-0014-z

Merryfield, W. J., Baehr, J., Batté, L., Becker, E. J., Butler, A. H., Coelho, C. A., et al. (2020). Current and emerging developments in subseasonal to decadal prediction. *Bulletin of the American Meteorological Society*, *101*(6), E869–E896. https://doi.org/10.1175/bams-d-19-0037.1

Molod, A., Hackert, E., Vikhliaev, Y., Zhao, B., Barahona, D., Vernieres, G., et al. (2020). GEOS-S2S version 2: The GMAO high-resolution coupled model and assimilation system for seasonal prediction. *Journal of Geophysical Research: Atmospheres*, *125*(5). https://doi.org/10.1029/2019jd031767

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), 1–7. 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

O'Brien, T. A., Payne, A. E., Shields, C. A., Rutz, J., Brands, S., Castellano, C., et al. (2020). Detection uncertainty matters for understanding atmospheric rivers. *Bulletin of the American Meteorological Society*, *101*(6), E790–E796. https://doi.org/10.1175/bams-d-19-0348.1

O'Brien, T. A., Wehner, M. F., Payne, A. E., Shields, C. A., Rutz, J. J., Leung, L. R., et al. (2022). Increases in future AR count and size: Overview of the ARTMIP Tier 2 CMIP5/6 experiment. *Journal of Geophysical Research: Atmospheres*, *127*(6), e2021JD036013. https://doi.org/10.1029/2021jd036013

Palmer, T., & Hagedorn, R. (Eds.). (2006). *Predictability of weather and climate*. Cambridge University Press.

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

Pegion, K., Kirtman, B. P., Becker, E., Collins, D. C., LaJoie, E., Burgman, R., et al. (2019). The Subseasonal Experiment (SubX): A multi-model subseasonal prediction experiment. *Bulletin of the American Meteorological Society*, *100*(10), 2043–2060. https://doi.org/10.1175/bams-d-18-0270.1

Pendergrass, A. G., Meehl, G. A., Pulwarty, R., Hobbins, M., Hoell, A., AghaKouchak, A., et al. (2020). Flash droughts present a new challenge for subseasonal-to-seasonal prediction. *Nature Climate Change*, *10*(3), 191–199. https://doi.org/10.1038/s41558-020-0709-0

Quinting, J. F., & Vitart, F. (2019). Representation of synoptic-scale Rossby wave packets and blocking in the S2S prediction project database. *Geophysical Research Letters*, *46*(2), 1070–1078. https://doi.org/10.1029/2018gl081381

Ralph, F. M., Cannon, F., Tallapragada, V., Davis, C. A., Doyle, J. D., Pappenberger, F., et al. (2020). West Coast forecast challenges and development of atmospheric river reconnaissance. *Bulletin of the American Meteorological Society*, *101*(8), E1357–E1377. https://doi.org/10.1175/bams-d-19-0183.1

Ralph, F. M., Coleman, T., Neiman, P. J., Zamora, R. J., & Dettinger, M. D. (2013). Observed impacts of duration and seasonality of atmospheric-river landfalls on soil moisture and runoff in coastal Northern California. *Journal of Hydrometeorology*, *14*(2), 443–459. https://doi.org/10.1175/jhm-d-12-076.1

Ralph, F. M., Dettinger, M., Lavers, D., Gorodetskaya, I. V., Martin, A., Viale, M., et al. (2017a). 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., Iacobellis, S. F., Neiman, P. J., Cordeira, J. M., Spackman, J. R., Waliser, D. E., et al. (2017b). Dropsonde observations of total integrated water vapor transport within North Pacific atmospheric rivers. *Journal of Hydrometeorology*, *18*(9), 2577–2596. https://doi.org/10.1175/jhm-d-17-0036.1

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 1997/1998. *Monthly Weather Review*, *132*(7), 1721–1745. https://doi.org/10.1175/1520-0493(2004)132<1721:sacaoo>2.0.co;2

Ralph, F. M., Rutz, J. J., Cordeira, J. M., Dettinger, M., Anderson, M., Reynolds, D., et al. (2019). A scale to characterize the strength and impacts of atmospheric rivers. *Bulletin of the American Meteorological Society*, *100*(2), 269–289. https://doi.org/10.1175/bams-d-18-0023.1

Reed, R. J., Campbell, W. J., Rasmussen, L. A., & Rodgers, D. G. (1961). Evidence of downward-propagating annual wind reversal in the equatorial stratosphere. *Journal of Geophysical Research*, *66*(3), 813–818. https://doi.org/10.1029/jz066i003p00813

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

Rutz, J. J., Steenburgh, W. J., & Ralph, F. M. (2014). Climatological characteristics of atmospheric rivers and their inland penetration over the western United States. *Monthly Weather Review*, *142*(2), 905–921. https://doi.org/10.1175/mwr-d-13-00168.1

Rutz, J. J., Steenburgh, W. J., & Ralph, F. M. (2015). The inland penetration of atmospheric rivers over western North America: A Lagrangian analysis. *Monthly Weather Review*, *143*(5), 1924–1944. https://doi.org/10.1175/mwr-d-14-00288.1

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

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., et al. (2014). The NCEP climate forecast system version 2. *Journal of Climate*, *27*(6), 2185–2208. https://doi.org/10.1175/jcli-d-12-00823.1

Sengupta, A., Singh, B., DeFlorio, M. J., Raymond, C., Robertson, A. W., Zeng, X., et al. (2022). Advances in subseasonal to seasonal prediction relevant to water management in the western United States. *Bulletin of the American Meteorological Society*, *103*(10), E2168–E2175. https://doi.org/10.1175/bams-d-22-0146.1

Shields, C. A., Rutz, J. J., Leung, L. Y., Ralph, F. M., Wehner, M., Kawzenuk, B., et al. (2018). Atmospheric River Tracking Method Intercomparison Project (ARTMIP): Project goals and experimental design. *Geoscientific Model Development*, *11*(6), 2455–2474. https://doi.org/10.5194/gmd-11-2455-2018

Stan, C., Zheng, C., Chang, E. K. M., Domeisen, D. I., Garfinkel, C. I., Jenney, A. M., et al. (2022). Advances in the prediction of MJO teleconnections in the S2S forecast systems. *Bulletin of the American Meteorological Society*, *103*(6), E1426–E1447. https://doi.org/10.1175/bams-d-21-0130.1

Stockdale, T. N., Anderson, D. L. T., Balmaseda, M. A., Doblas-Reyes, F., Ferranti, L., Mogensen, K., et al. (2011). ECMWF seasonal forecast system 3 and its prediction of sea surface temperature. *Climate Dynamics*, *37*(3–4), 455–471. https://doi.org/10.1007/s00382-010-0947-3

Sun, R., Subramanian, A. C., Cornuelle, B. D., Mazloff, M. R., Miller, A. J., Ralph, F. M., et al. (2021). The role of air-sea interactions in atmospheric rivers: Case studies using the SKRIPS regional coupled model. *Journal of Geophysical Research: Atmospheres*, *126*(6), e2020JD032885. https://doi.org/10.1029/2020jd032885

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., et al. (2017). The subseasonal to seasonal (S2S) prediction project database. *Bulletin of the American Meteorological Society*, *98*(1), 163–173. https://doi.org/10.1175/bams-d-16-0017.1

Vitart, F., & Robertson, A. W. (2018). The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. *NPJ Climate and Atmospheric Science*, *1*(1), 1–7. https://doi.org/10.1038/s41612-018-0013-0

Vitart, F., Robertson, A. W., & Anderson, D. L. (2012). Subseasonal to seasonal prediction project: Bridging the gap between weather and climate. *Bulletin of the World Meteorological Organization*, *61*(2), 23.

Waliser, D. E., Lau, K. M., Stern, W., & Jones, C. (2003). Potential predictability of the Madden–Julian Oscillation. *Bulletin of the American Meteorological Society*, *84*(1), 33–50. https://doi.org/10.1175/BAMS-84-1-33

Wang, J., DeFlorio, M. J., Guan, B., & Castellano, C. M. (2023). Seasonality of MJO impacts on precipitation extremes over the western United States. *Journal of Hydrometeorology*, *24*(1), 151–166. https://doi.org/10.1175/jhm-d-22-0089.1

Wang, W., Chen, M., & Kumar, A. (2010). An assessment of the CFS real-time seasonal forecasts. *Weather and Forecasting*, *25*(3), 950–969. https://doi.org/10.1175/2010waf2222345.1

Wehner, M. F., Smith, R. L., Bala, G., & Duffy, P. (2010). The effect of horizontal resolution on simulation of very extreme U.S. precipitation events in a global atmosphere model. *Climate Dynamics*, *34*(2), 241–247. https://doi.org/10.1007/s00382-009-0656-y

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)132<1917:aarmmi>2.0.co;2

White, C. J., Carlsen, H., Robertson, A. W., Klein, R. J., Lazo, J. K., Kumar, A., et al. (2017). Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteorological Applications*, *24*(3), 315–325. https://doi.org/10.1002/met.1654

White, C. J., Domeisen, D. I., Acharya, N., Adefisan, E. A., Anderson, M. L., Aura, S., et al. (2022). Advances in the application and utility of subseasonal-to-seasonal predictions. *Bulletin of the American Meteorological Society*, *103*(6), E1448–E1472. https://doi.org/10.1175/bams-d-20-0224.1

Zhang, C. (2005). Madden–Julian Oscillation. *Reviews of Geophysics*, *43*(2). https://doi.org/10.1029/2004RG000158

Zhang, Z., Pierce, D. W., & Cayan, D. R. (2019). A deficit of seasonal temperature forecast skill over west coast regions in NMME. *Weather and Forecasting*, *34*(4), 833–848. https://doi.org/10.1175/waf-d-18-0172.1

Zhang, Z., & Ralph, F. M. (2021). The influence of antecedent atmospheric river conditions on extratropical cyclogenesis. *Monthly Weather Review*, *149*(5), 1337–1357. https://doi.org/10.1175/mwr-d-20-0212.1

Zhang, Z., Ralph, F. M., & Zheng, M. (2019). The relationship between extratropical cyclone strength and atmospheric river intensity and position. *Geophysical Research Letters*, *46*(3), 1814–1823. https://doi.org/10.1029/2018gl079071

Zheng, C., Chang, E. K. M., Kim, H., Zhang, M., & Wang, W. (2019). Subseasonal to seasonal prediction of wintertime Northern Hemisphere extratropical cyclone activity by S2S and NMME models. *Journal of Geophysical Research: Atmospheres*, *124*(22), 12057–12077. https://doi.org/10.1029/2019jd031252

Zheng, M., Delle Monache, L., Wu, X., Ralph, F. M., Cornuelle, B., Tallapragada, V., et al. (2021). Data gaps within atmospheric rivers over the northeastern Pacific. *Bulletin of the American Meteorological Society*, *102*(3), E492–E524. https://doi.org/10.1175/bams-d-19-0287.1

Zhou, Y., & Kim, H. M. (2018). Prediction of atmospheric rivers over the North Pacific and its connection to ENSO in the North American multi-model ensemble (NMME). *Climate Dynamics*, *51*(5–6), 1623–1637. https://doi.org/10.1007/s00382-017-3973-6

Zhou, Y., Kim, H., & Waliser, D. E. (2021). Atmospheric river lifecycle responses to the Madden–Julian Oscillation. *Geophysical Research Letters*, *48*(3), e2020GL090983. https://doi.org/10.1029/2020gl090983

Zhu, Y., & Newell, R. E. (1998). A proposed algorithm for moisture fluxes from atmospheric rivers. *Monthly Weather Review*, *126*(3), 725–735. https://doi.org/10.1175/1520-0493(1998)126<0725:apafmf>2.0.co;2