Evaluation of Forecasts of the Water Vapor Signature of Atmospheric Rivers in Operational Numerical Weather Prediction Models

GARY A. WICK, PAUL J. NEIMAN, F. MARTIN RALPH, AND THOMAS M. HAMILL

NOAA/Earth System Research Laboratory/Physical Sciences Division, Boulder, Colorado

ABSTRACT

The ability of five operational ensemble forecast systems to accurately represent and predict atmospheric rivers (ARs) is evaluated as a function of lead time out to 10 days over the northeastern Pacific Ocean and west coast of North America. The study employs the recently developed Atmospheric River Detection Tool to compare the distinctive signature of ARs in integrated water vapor (IWV) fields from model forecasts and corresponding satellite-derived observations. The model forecast characteristics evaluated include the prediction of occurrence of ARs, the width of the IWV signature of ARs, their core strength as represented by the IWV content along the AR axis, and the occurrence and location of AR landfall. Analysis of three cool seasons shows that while the overall occurrence of ARs is well forecast out to a 10-day lead, forecasts of landfall occurrence are poorer, and skill degrades with increasing lead time. Average errors in the position of landfall are significant, increasing to over 800 km at 10-day lead time. Also, there is a 1°–2° southward position bias at 7-day lead time. The forecast IWV content along the AR axis possesses a slight moist bias averaged over the entire AR but little bias near landfall. The IWV biases are nearly independent of forecast lead time. Model spatial resolution is a factor in forecast skill and model differences are greatest for forecasts of AR width. This width error is greatest for coarser-resolution models that have positive width biases that increase with forecast lead time.

1. Introduction

Accurately forecasting extreme precipitation and flooding is important for the protection of human lives and property, and is a critical mission of weather services throughout the globe. Depending on the geographic region, different types of weather systems are responsible for the most significant precipitation events. Along the coastal regions of the western United States, the most extreme precipitation typically occurs during the wintertime cool season (Ralph and Dettinger 2012). Recent research has demonstrated that major winter flooding events both in California and the Pacific Northwest were accompanied by the presence of features termed atmospheric rivers (Ralph et al. 2006; Neiman et al. 2008a; Neiman et al. 2011).

Atmospheric rivers (ARs) are long, narrow regions of intense water vapor transport within the lower atmosphere (e.g., Zhu and Newell 1998; Ralph et al. 2004) that represent a subset corridor within a broader region of generally poleward heat transport in the warm sector of extratropical cyclones (Neiman et al. 2008b). Combining moist neutrality, strong horizontal winds, and large water vapor content, ARs can foster heavy orographic precipitation when they make landfall (e.g., Ralph et al. 2005; Neiman et al. 2008a). While the extreme precipitation can result in flooding as cited above, the results can also be beneficial, as AR events contribute significantly to the seasonal water supply in the western United States (Dettinger et al. 2011; Guan et al. 2010).

Given their critical role in the global water cycle and extreme precipitation, it is important to understand how well ARs are represented and predicted in current numerical weather prediction (NWP) models. Analysis of quantitative precipitation forecasts (QPF) by Ralph et al. (2010) showed that some of the largest QPF forecast errors were associated with landfalling atmospheric rivers. Forecasters rely heavily on the model guidance of AR activity for issuance of flood warnings along the U.S. west coast. While the accuracy of forecasts of the intensity and landfall location of tropical cyclones are closely monitored, quantitative evaluations of forecasts of ARs

Corresponding author address: Gary A. Wick, NOAA/ESRL/ Physical Sciences Division, 325 Broadway, Boulder, CO 80305.
E-mail: gary.a.wick@noaa.gov

and Pacific landfalling winter storms are found to be lacking. Real-time web pages at our laboratory monitoring forecast fields for the presence of ARs are frequently viewed, but more guidance on the forecast reliability and categorization with respect to climatology has been requested. This paper fills an important gap in understanding the cause of large QPF errors in extreme events on the U.S. west coast by documenting the uncertainties in predicting the phenomena primarily responsible for the extreme precipitation: atmospheric rivers.

Forecast verification can be approached in many ways, both with respect to variables and the validation source. Previous model forecast evaluations for storms in the northeastern Pacific have largely focused on sea level pressure and cyclone center characteristics associated with extratropical cyclones (e.g., McMurdie and Mass 2004; McMurdie and Casola 2009; Froude 2010). To specifically consider ARs, object- or feature-based techniques can be employed to identify distinct events and evaluate important characteristics such as position, extent, and frequency of occurrence. Ideally, ARs are best characterized in terms of the water vapor transport, which defines the features and is closely related to their potential for orographic precipitation. In one recent object-based effort, Clark et al. (2011) evaluate the errors in forecasts of water vapor transport associated with ARs from the Global Forecast System (GFS) model through comparisons against the model analysis corresponding to the valid time of the forecast. While comparisons against analyses or a reanalysis product enable very direct comparisons, to achieve complete model independence and facilitate comparisons with multiple models, validation against direct observations is desirable.

Aircraft flight-level and dropsonde observations allow a detailed evaluation of the representation of ARs (including their water vapor transport) in NWP forecasts and reanalyses for a few select recent cases. The National Oceanic and Atmospheric Administration (NOAA) led Winter Storms and Pacific Atmospheric Rivers (WISPAR) experiment in February–March 2011 resulted in the sampling of three ARs with dropsondes deployed from the National Aeronautics and Space Administration (NASA) unmanned Global Hawk aircraft along with additional observations of one of the ARs with dropsondes deployed from the NOAA G-IV aircraft flying a Winter Storms Reconnaissance mission. Two earlier ARs were also sampled with dropsondes from the NOAA P-3 aircraft. The limited number of these cases, however, precludes a comprehensive evaluation of model forecast accuracy.

Because aircraft data like these are only available episodically and there are no other direct observations of water vapor transport over the oceans, validation of

significant numbers of forecasts of ARs against direct observations requires the use of fields different than the water vapor transport. Fields common to both the models and observations are required. Current satellite sensors are unable to quantify the water vapor transport due to a lack of information on the vertical wind profile. Based on comparisons with aircraft observations, Ralph et al. (2004) demonstrated that satellite-based retrievals of the vertically integrated water vapor (IWV) content could be successfully used as a proxy for the identification of ARs over the ocean and developed objective criteria to identify the features. Employing this approach, Neiman et al. (2008b) demonstrated that, in a composite sense, the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis accurately depicts the position and orientation of AR plumes, but they did not compare individual events in detail. This paper builds upon this work and utilizes a feature-based method to evaluate the representation of ARs based on their IWV signature on a case-by-case basis.

Visually evaluating the performance of multiple models over an extended period across a large geographic area is very time consuming, particularly if considering several different lead times, because of the large number of fields that must be examined. To facilitate the verification of AR forecasts, a new automated, objective technique for the identification and characterization of the IWV signature of ARs in both model and observational fields has been developed and validated (Wick et al. 2013). The technique, the AR Detection Tool for IWV (ARDT-IWV), utilizes basic image-processing techniques such as thresholding and skeletonization to implement and extend the objective criteria for the length (>2000 km), width (<1000 km), and IWV content (>2 cm) for ARs that was first defined by Ralph et al. (2004) and used in multiple later studies. In initial evaluations, the ARDT-IWV proved highly successful (92.4% critical success index and a 98.5% probability of detection) in identifying landfalling AR events in satellite IWV imagery in comparison with a manually derived climatology (Wick et al. 2013). Moreover, the ARDT-IWV is particularly well suited for the verification of model fields since identical objective criteria are applied to both the model and observation fields.

This paper employs the ARDT-IWV to evaluate the ability of several operational ensemble prediction systems to accurately forecast the IWV signature of ARs in the northeast Pacific Ocean roughly in the region from Hawaii to the west coast of North America. The existence and characteristics of ARs in the model fields are compared against corresponding satellite-derived observations of IWV. Control forecasts from five prominent

ensemble prediction systems over the months of October–March for the three cool seasons from 2008–09 to 2010–11 are evaluated as a function of lead time out to 10 days. The key characteristics of the forecasts evaluated include 1) the predictions of AR occurrence, 2) the width of the IWV signature of ARs, 3) the core strength of the ARs as represented by the IWV content along the AR axis, and 4) the occurrence and location of landfall of the ARs. Particular attention is given to the impact of the spatial resolution of the different forecast models. Section 2 introduces the specific NWP models evaluated and the satellite-derived IWV products used in their verification. Section 3 describes the approach and methods employed in the model assessment, and the results are presented in section 4. Implications and conclusions from this work are summarized in section 5.

2. Data

a. NWP models

The ability of five different operational ensemble prediction systems to accurately forecast and represent ARs is evaluated in this study. These models include those from NCEP, the European Centre for Medium-Range Weather Forecasts (ECMWF), the Met Office (UKMO), the Canadian Meteorological Center (CMC), and the Japan Meteorological Agency (JMA). The evaluation was performed for the models' analysis fields and the control forecast fields at lead times of 1, 3, 5, 7, and 10 days, where available. Multiple ensemble members and the ensemble mean were not tested. To achieve the best temporal coincidence with the available satellite observations, the 1200 UTC model runs were used in each case.

The forecast model data were obtained through The Observing System Research and Predictability Experiment (THORPEX) Interactive Grand Global Ensemble (TIGGE; e.g., Bougeault et al. 2010). Data were downloaded from the portal maintained at ECMWF. The extracted model fields were labeled as total column water. Evaluations were performed for the months of October–March for the three cool seasons from 2008–09 to 2010–11.

The following text describes the basic relevant characteristics of each model and a summary is included in Table 1. Since the models are operational, changes impacting the model characteristics may have occurred over the 3-yr verification period.

The ECMWF Ensemble Prediction System (EPS) uses the ECMWF Integrated Forecast System model, employing four-dimensional variational data assimilation (4DVAR). The horizontal resolution of the model

TABLE 1. Summary of key characteristics of the operational ensemble prediction systems evaluated in this study.

Model center	Model horizontal resolution	Grid resolution for comparison	Forecast fields evaluated (h)
ECMWF	TL399	0.5° × 0.5°	0, 24, 72, 120, 168, 240
NCEP	T126	1° × 1°	0, 24, 72, 120, 168, 240
UKMO	1.25° × 0.833°	1° × 1°	0, 24, 72, 120, 168, 240
CMC	0.9°	1° × 1°	0, 24, 72, 120, 168, 240
JMA	T319	1° × 1°	0, 24, 72, 120, 168

was TL399 out through 10 days, corresponding to approximately 0.45°. The data were regridded onto a 0.5° × 0.5° grid for application of the ARDT and comparison with the satellite-derived observations. The model contained 62 vertical levels. An overview of the ensemble system is included in Buizza et al. (2007). Forecasts extended out to 15 days (at reduced horizontal resolution), but only the first 10 days were considered.

The NCEP Global Ensemble Forecast System (GEFS) employs the Global Forecast System (GFS) model incorporating the gridded statistical interpolation (GSI) assimilation method. The model resolution was T126 with output on a 1° grid. Verification of the forecasts was conducted on this 1° grid. The model contained 28 vertical levels. Forecasts were available out to 16 days.

The UKMO EPS uses the Met Office Global and Regional Ensemble Prediction System (MOGREPS; e.g., Bowler et al. 2008) employing 4DVAR. The horizontal resolution of the model was 0.833° latitude × 1.25° longitude on a regular grid. For comparison with the satellite-derived fields, the data were regridded onto a 1° grid. Vertically, 38 levels were used. Forecasts were produced out to 15 days.

The CMC Global Ensemble Prediction System (GEPS) uses the Global Environmental Multiscale Model (GEM) and an ensemble Kalman filter assimilation method. The model resolution was 0.9° with 58 vertical levels. Output was remapped onto a 1° grid for verification. Forecasts were available out to 16 days. Additional details on the Canadian EPS are included in Charron et al. (2010).

The JMA 1-week ensemble prediction system (WEPS) uses the Global Spectral Model (GSM) and a 4DVAR scheme. The model resolution was T319 with 60 vertical levels. Output was obtained at 1.25° resolution and remapped onto a 1° grid. Forecasts were only produced out to 9 days, making this the only of the models evaluated without a 10-day forecast.

b. Satellite data

The reference product for validation of the various models' representation of ARs was taken to be

satellite-derived fields of IWV. The IWV over the oceans was retrieved from passive microwave observations from the Special Sensor Microwave Imager (SSM/I) and the Special Sensor Microwave Imager/Sounder (SSM/IS) flying on the Defense Meteorological Satellite Program (DMSP) satellites using the statistical algorithm of Wentz (1995). Retrievals were generated from the orbital data from the SSM/I on the *F-13* satellite (through November 2009) and SSM/IS on the *F-16* and *F-17* satellites and then mapped onto grids at 0.5° and 1.0° resolution for comparison with the model fields. The native resolution of the IWV retrievals is approximately 40 km. Due to calibration issues with the SSM/IS (e.g., Kunkee et al. 2008), the brightness temperatures were linearly mapped to those of the SSM/I *F-13* using a technique closely modeled after Yan and Weng (2008) and employed in real-time diagnostics at NOAA's Earth System Research Laboratory (ESRL).

All available data from the different sensors for the period from 1200 to 2359 UTC each day were combined into single grids. For the region of study in the northeastern Pacific, the data typically corresponded to between 1300 and 1600 UTC, providing good coincidence with the forecast fields initialized at 1200 UTC and valid at 24-h intervals. An average effective time difference between the satellite data and forecast fields was approximately 2 h. While this time difference could be eliminated entirely through comparison against reanalysis data, the comparison against observations was determined to be a priority for this study so that numerical models do not influence the "observed" AR structure in any way. Nevertheless, the time difference does introduce a source of uncertainty that will be explicitly addressed in the analysis. For a typical propagation speed of ~50 km h^{-1}, 2 h would correspond to a difference of approximately 100 km in position. The high degree of consistency in time of the contributing satellite observations resulting from the similarity of the orbits and overpass times of the different sensors limits the amount of blurring of propagating IWV plumes over time, making it possible to reasonably assess the extent and location of features in the data. Combining the data from the different satellites eliminates gaps between the individual swaths that would hamper application of the ARDT-IWV.

Satellite-derived retrievals of the total precipitable water vapor or IWV are sufficiently advanced to serve as a reliable source for verification of the model-based fields. The IWV can potentially be retrieved from passive microwave observations with a high degree of accuracy (Wentz 1997) as the relationship is more direct than for many remotely derived quantities. Though there is limited comprehensive validation against in situ data over the open oceans, the physically based IWV product generated by Remote Sensing Systems (Wentz 1997; Wentz et al. 2007) has been broadly applied in climatological studies (Trenberth et al. 2005; Wentz et al. 2007; Mears et al. 2007). We have also successfully employed satellite-derived retrievals of IWV in several observationally based AR studies, as cited in the introduction. While we generated our own retrievals from the orbital data using the Wentz (1995) optimal statistical algorithm to facilitate blending of the data from the different sensors over specific time intervals, our comparisons (not shown) have demonstrated very good agreement between the RSS gridded physical product and our product. Additionally, comparison of our IWV product with independent estimates of the IWV from global positioning system radio occultation (GPSRO) soundings from the Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC) mission showed strong agreement with zero mean bias and a 3-mm RMS difference (Wick et al. 2008). Such differences are small relative to the variations across ARs and, in any event, any biases should not hinder accurate representation of the AR spatial structure.

3. Methodology

Our approach to evaluating the ability of the operational forecast models to predict and represent the IWV signature of ARs employs a feature-based verification methodology to directly compare the characteristics of the IWV plumes in the model fields against those in the satellite-derived observations. The key questions we seek to answer include the following:

- Are the occurrence and timing of ARs accurately predicted in current models?
- How accurately is the width of the features reproduced given the available resolution?
- Are there any biases in the modeled strength of the ARs as represented by the IWV content along its axis?
- How well is the occurrence and location of landfall of ARs forecast along the west coast of North America?

We particularly examine how these results are affected by the model resolution and forecast lead time.

To address these goals, we apply the ARDT-IWV developed by Wick et al. (2013) to identify and characterize ARs in individual fields of IWV from both the forecast models and corresponding satellite observations. The suitability of the procedure for this task was demonstrated by its success in replicating a visually derived climatology of landfalling AR events as shown in Wick et al. (2013). The comparison requires consistent processing of both the model and satellite data. The

ARDT-IWV (hereafter simply ARDT) is run independently for each available model-based and satellite-derived IWV field. The primary outputs include the number of ARs present in the scene, and, for each AR, the coordinates of the identified AR axis, corresponding IWV values, and multiple estimates of the AR width at each location. Width values are derived for where the IWV first falls below thresholds at 2.0, 2.33, and 2.67 cm, and drops to 0.37 times (the e-folding scale) the difference between the peak and mean values along the AR cross section. Additionally, a flag indicates if the detected AR makes contact with a continental landmass somewhere along its extent. The corresponding model- and satellite-based results are grouped together and evaluated for each model and distinct lead time. Any forecast–observation field pair where the satellite-derived IWV field has significant gaps due to data outages or overlap of orbits from the different satellites is discarded because of the potential for the gaps to compromise the automated AR detection.

Comparison of the model- and satellite-based results considered both the accurate prediction of occurrence and the specific characteristics of the ARs. With respect to AR occurrence, the key parameter considered is the number of days with at least one AR present over the domain (in both the observational and forecast fields). This quantity is easier to assess and subject to less uncertainty than the number of distinct ARs but gives much the same information. Specific characteristics such as the width, position, and IWV content along the AR axis were compared for days with exactly one AR detected in both the forecast and corresponding satellite-derived field. This restriction was imposed to minimize the potential comparison of different features present in different portions of the spatial domain. There is still the slight possibility that single different, unrelated features could be detected in the corresponding model and satellite-based fields, but the fields were generally similar. Width and core IWV content were characterized both as averages over the entire length of the AR and for distinct latitudinal bands. The width estimates used in the comparisons were the minimum width computed relative to the 2.0-, 2.33-, and 2.67-cm IWV thresholds. This value enables a good comparison of the AR width for the different models as it allows for slight variations in the background IWV values. Further comparisons of results were then performed for the specific case of landfalling events. For all quantities, the results were stratified by forecast lead time and the performance of the different models compared.

The ability of the models to reproduce the number of days with ARs present was evaluated in terms of traditional forecast diagnostics (e.g., Wilks 2006) including threat score (TS), probability of detection (POD), and false alarm ratio (FAR). The threat score (or critical success index) is defined as the number of cases where an AR was identified both in the forecast field and in the satellite observations divided by the number of occasions on which an AR was either forecast and/or observed. The POD represents the fraction of those cases of an observed AR where it was accurately forecast to occur. The FAR is defined as the fraction of forecast AR cases that did not occur in the observations. The representation of the width, IWV content, and location were evaluated simply in terms of bias, standard deviation, and RMS error.

The geographic domain for the study is 15°–55°N, 110°–160°W. This domain corresponds to that of the climatological analysis of landfalling ARs conducted by Neiman et al. (2008a) and encompasses those events contributing to precipitation along the west coast of North America. The domain is also that for which the accurate performance of the ARDT was validated in Wick et al. (2013).

Slightly different preprocessing is applied to the satellite and model data because of their inherent difference in smoothness. A median filter is applied to the input IWV field as part of the ARDT to facilitate identification of continuous features. Because of the higher noise level in actual satellite observations, a 7×7 window is used for satellite data while a 3×3 window is applied to the model fields.

To minimize apparent errors in the forecast AR widths and other quantities due to the coarser resolution of the model fields compared to the observations, primary analyses of the satellite data were conducted on the data mapped onto a 0.5° grid corresponding to the finest-resolution model field. Because this still placed the 1°-resolution models at a disadvantage, the analyses were also repeated for the observations gridded at 1° resolution. Internal to the function of the ARDT, all input fields at a resolution coarser than 0.5° are regridded at 0.5° resolution. Bilinear interpolation is used in this process to smoothly fill the grid cells between the original coarser grid-cell centers.

The overall nature of the comparison is illustrated in Fig. 1 for a landfalling AR on 7 January 2009. This AR produced record rains and flooding across parts of the Pacific Northwest and was responsible for compromising the integrity of Washington's Howard Hanson Dam on the Green River (e.g., Mastin et al. 2010; White et al. 2012). The observed IWV field from the satellite-derived data is shown on top along with the 1-, 3-, and 7-day control forecast fields from each of the five ensemble prediction systems considered. Within each panel, the objectively determined axis of the AR is shown with the overplotted gray circles. Each of the forecast models
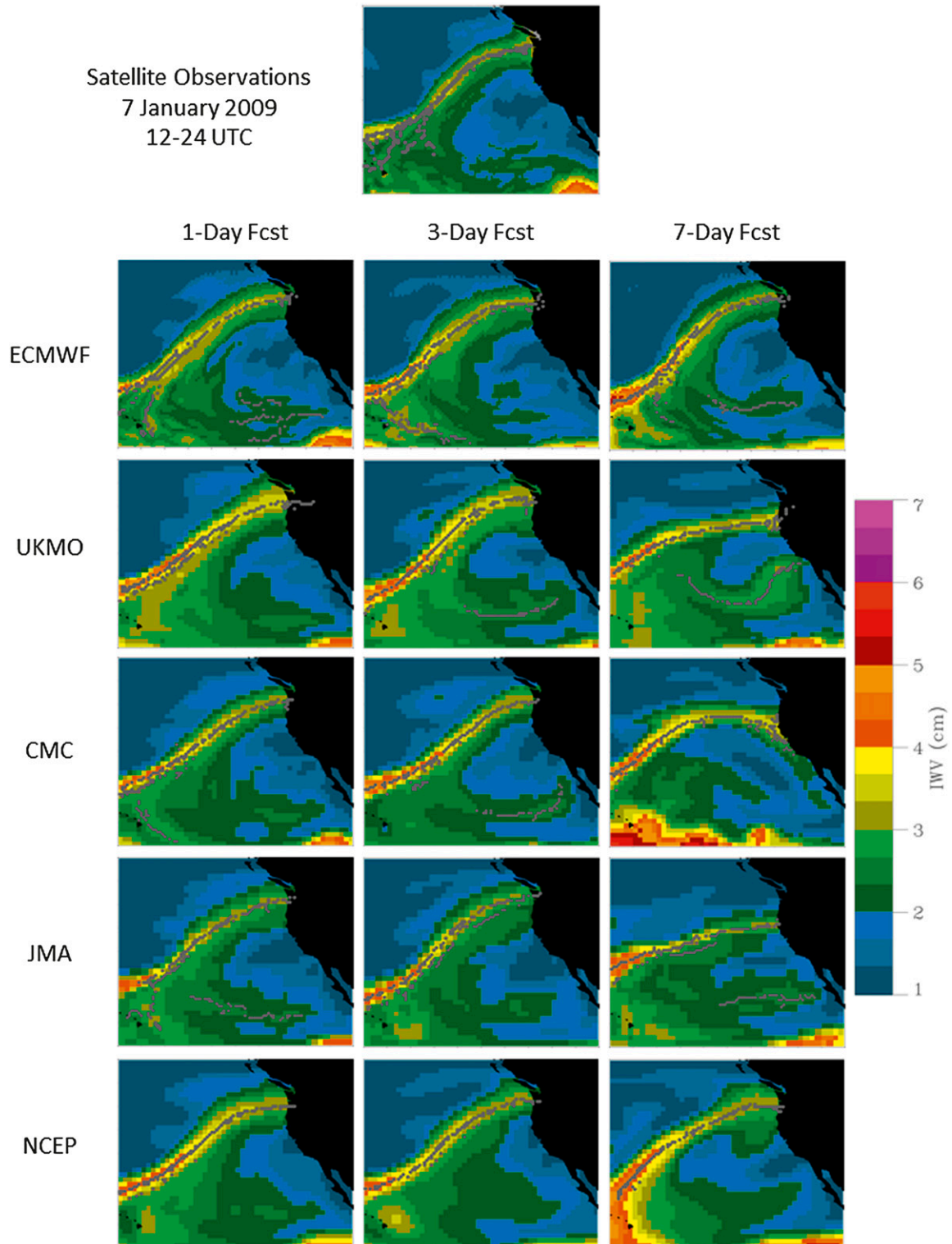
FIG. 1. Visual illustration comparing the modeled representation of an AR observed on 7 Jan 2009. (top) The satellite-observed IWV with the detected AR axis indicated with the gray circles. (bottom) The (left) 1-day, (middle) 3-day, and (right) 7-day forecasts all valid at the time of the observations. Separate rows correspond to the different ensemble forecast systems as shown. While the AR is clearly represented by all models at all lead times, increasing variability in the AR position and strength is observed at the longer lead times.

clearly predicts the occurrence of an AR even out to 7 days, but the predicted position, width, orientation, and IWV content of the AR vary notably with model and lead time. The location of landfall, in particular, is highly variable across models and forecast lead times. The results in the following section attempt to quantify these differences in forecast accuracy over the three cool seasons considered.

4. Results

Since the results of the ARDT are highly dependent on the IWV threshold values used in detecting the ARs, the model-based total column water outputs were first evaluated for potential overall biases relative to the satellite-derived IWV product. Grid-cell by grid-cell differences in IWV content were computed and averaged over the three cool seasons' worth of data for each model and forecast lead time. This comparison was performed over a broader domain (55°N–55°S, 110°E–70°W) encompassing the entire Pacific Ocean to better capture the total variability in the IWV fields. Because of the approach to gridding the satellite data, time differences between the observations and forecast fields were slightly greater in the western portion of the domain. The resulting derived IWV biases are shown in Fig. 2. Results limited to the localized study domain were similar and are not shown. Overall, the biases are generally small with values typically less than 1 mm. The biases are smallest and most consistent for the ECMWF fields. The largest biases, approaching 2 mm, are observed for the UKMO fields which, as for NCEP, show a tendency for decreasing bias with increasing lead. The CMC fields, in contrast, show an increase in bias with increasing lead. Because the biases were generally small and variable with forecast lead, no bias adjustments were applied to the model fields prior to application of the ARDT. The use of multiple IWV thresholds within the ARDT (Wick et al. 2013) should lessen the impact of the remaining biases.

a. Prediction of occurrence

The first question addressed was how well the occurrence of ARs is forecast. The TS, POD, and FAR for forecasts of days with at least one AR present within the analysis domain averaged over the three cool seasons analyzed are plotted in Fig. 3 as a function of forecast lead time. Overall, the results are very positive, demonstrating that the occurrence of ARs is generally well forecast by each model even out to 10-day lead time. The probability of detecting ARs is greater than 84% and false alarms are less than 12% at all lead times. While model performance is generally similar, the TS and POD values are typically highest for ECMWF and
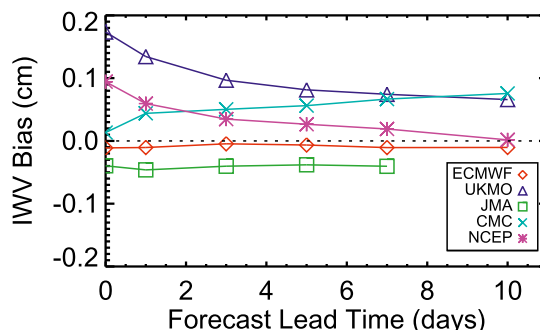

FIG. 2. Bias in the modeled IWV fields as a function of forecast lead time computed relative to the satellite-derived observations averaged over the three cool seasons of data for a domain encompassing the entire Pacific Ocean. Different models are reflected by the different colors and symbols.

lowest for NCEP. The high POD values for ECMWF are accompanied by a slightly increased FAR, though the difference relative to the other models may not be significant. Not surprisingly, the model performance does tend to degrade with increasing lead time but the degradation is not that great. One exception to this tendency is observed for the UKMO results moving from the analysis out to 3-day lead time. The poorer performance for the analysis and 1-day lead results is likely related to the increased IWV bias at these times observed in Fig. 2, which leads to IWV features in the model data that are too wide to be classified as ARs.

We next looked specifically at the accuracy of predictions of the occurrence of those ARs making landfall along the west coast of North America within the analysis domain. This approximately corresponds to latitudes from 25° to 55°N. The TS, POD, and FAR for these comparisons are shown in Fig. 4. The accuracy of predictions of days with a landfalling AR are significantly poorer than those for overall AR occurrence, and the decrease in performance of the models with increasing lead time is larger. Threat scores are decreased (relative to that for overall AR occurrence) by values ranging from 0.25 at 1-day lead to over 0.4 at 10-day lead. While the probability of detection is still near 80% at 1-day lead, values decrease to near just 60% for 10-day forecasts. Similarly, false alarms increase from ~27% at 1 day to near 45% for 10-day forecasts. The relative performance of the models remains generally similar. While small differences in time should not significantly impact the overall occurrence statistics, they could potentially have an effect on the values for landfall. If anything, however, there is a slight tendency for the models to overpredict landfall, and the nature of the time offset is such that the observations are slightly later than the forecasts, giving the observed ARs more time to make landfall.
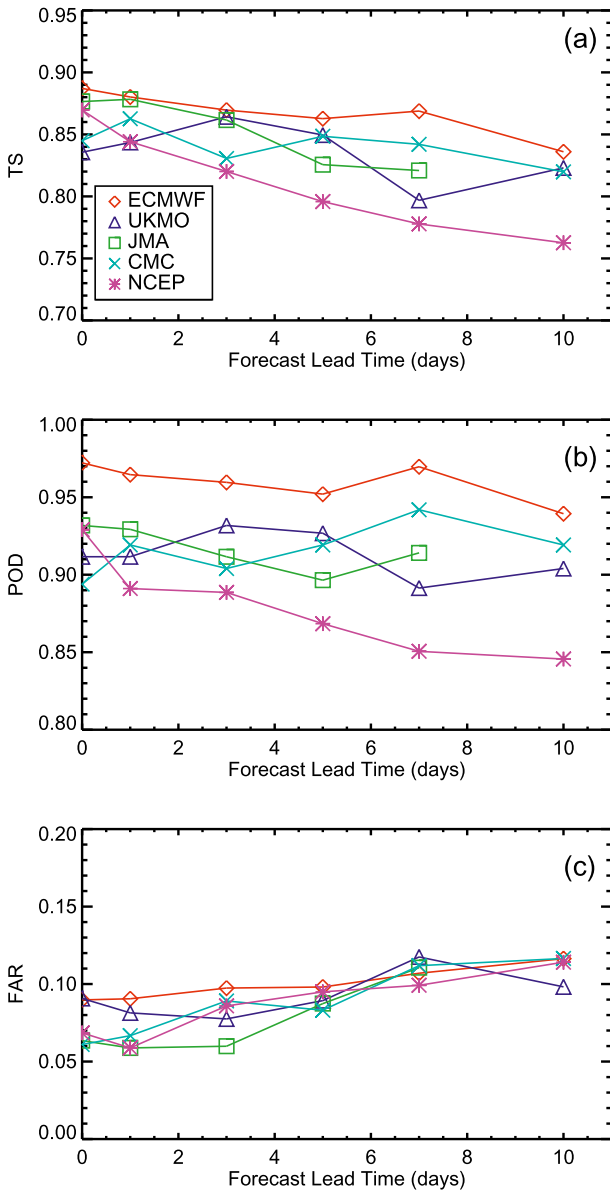
FIG. 3. Forecast verification statistics reflecting the ability of the models to predict the overall occurrence of at least one AR somewhere within the analysis domain on a given day. Quantities include (a) TS, (b) POD, and (c) FAR. The scores were computed using all three cool seasons of observations and are plotted as a function of forecast lead time.

FIG. 4. As in Fig. 3, but for the prediction of at least one landfalling AR within the domain on a given day.

If the criterion for accurate prediction of AR landfall occurrence is relaxed to occurrence within a 2-day window rather than on the same day, the behavior of the TS, POD, and FAR shown in Fig. 5 is obtained. At longer lead times, knowing that an AR would make landfall within a couple of days can be almost as valuable as knowing the exact day. The scores are all improved markedly with a threat score and a probability of detection at 10-day lead of ~0.58 and 0.73–0.78, respectively.
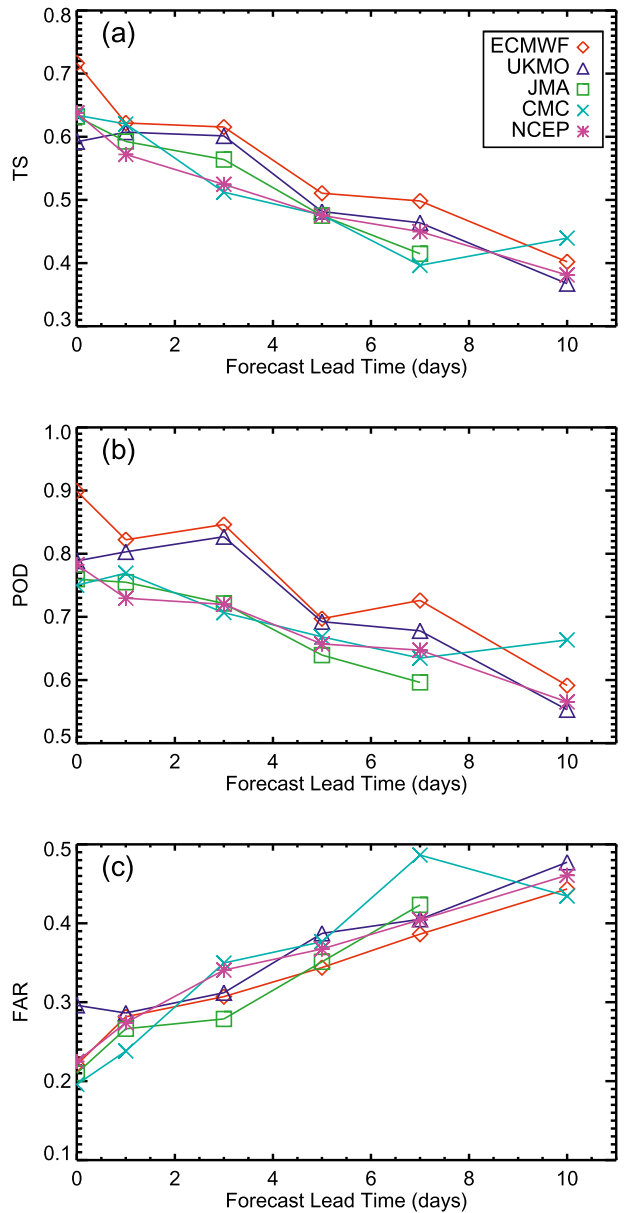
While the scores are still degraded relative to overall occurrence, the results suggest that the landfall forecasts can still be very valuable for planning purposes.

The ability of the models to reproduce the seasonal and latitudinal distributions of the occurrence of landfalling ARs is presented in Fig. 6. The frequency of occurrence of landfalling ARs was computed for individual 10° bins of latitude and 2-month seasonal blocks. The model results are shown for 7-day forecasts. The largest fraction of days with observed landfalling ARs occur in the early season (October–November) along the most
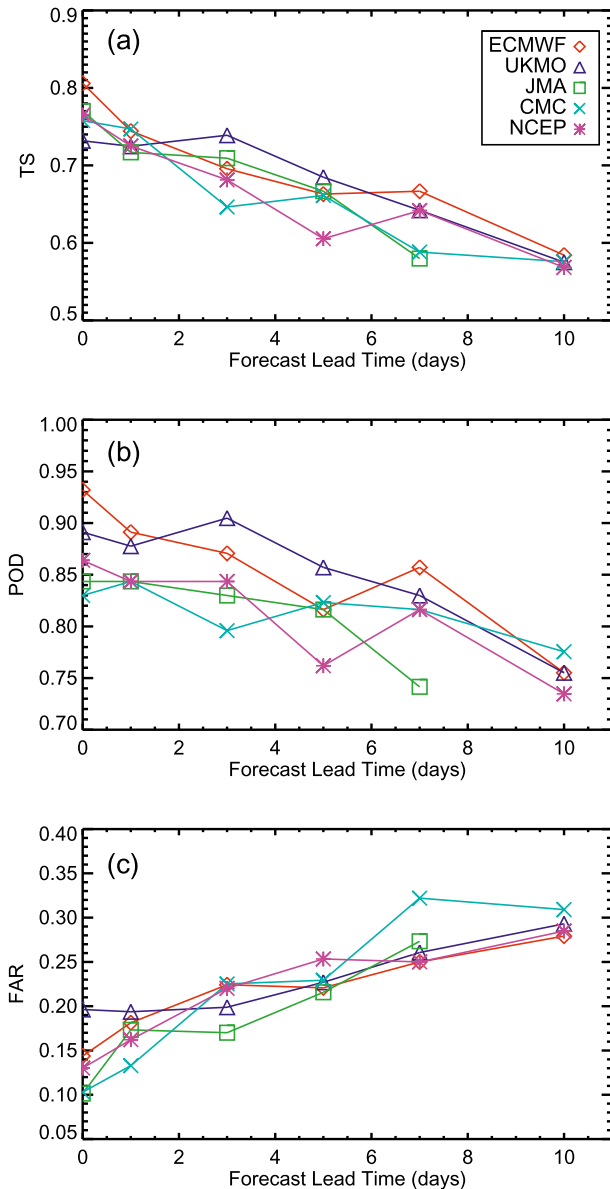
FIG. 5. As in Fig. 4, but for the prediction of landfall within a 2-day window rather than for 1 day.

northern part of the coast. In this early part of the season the number of days with landfalling ARs decreases moving south along the coast with very few events along the southernmost part of the coast. These tendencies are captured by all the models except CMC, which predicts the largest fraction of landfalling ARs in the early season to occur in the middle coast region. All models reproduce the observed tendency for the largest percentage of days with landfalling ARs in the southernmost coast to occur in the late part of the cool season. The increase in the number of ARs in the south later in the cool season is consistent with the climatological pattern of an expanding

circumpolar vortex during the winter with a resulting more southern storm track with time. Considering land-falling ARs over all seasons, the models predict a peak in the number of landfalls in the central part of the coast while observations suggest that the north and central bins have more similar numbers of landfalling ARs. The models generally underpredict landfall in the north part of the coast while overpredicting landfall in the center and south coast. Combining all latitude bands, the number of days with observed landfalling ARs is greatest in the early season and decreases as the season progresses. This is reproduced by all models except NCEP, which has nearly equal numbers in December–January and February–March.

A total number of 455 days with suitable gap-free satellite observations were available over the three cool seasons. A landfalling AR was observed somewhere along the coast on 208 (or 46%) of those days. In contrast, the presence of at least one AR anywhere in the analysis domain was detected in the satellite observations on a remarkable 87% of those days. The occurrence of ARs is clearly very common over the open ocean.

b. IWV content and width

The remaining analyses examine the ability of the different ensemble prediction systems to reproduce specific characteristics of the ARs and target the questions related to the accuracy of forecast strength, width, and position of the ARs. An initial broad measure of the models' ability to capture the position of ARs can be obtained by computing a simple pattern correlation between the satellite-derived IWV fields and each individual model forecast. Comparing the gridpoint by gridpoint IWV values over the entire northeastern Pacific domain for all available scenes over the three cool seasons analyzed yields the correlation coefficients shown in Fig. 7. The results illustrate excellent short-term performance but a steady decrease in the skill of the forecast models as the lead time increases. While over 80% of the observed variance in IWV is reproduced in the models' analysis and 1-day forecast fields, the percentage decreases to less than 35% at 10-day lead. In this comparison the models all perform very similarly with the exception of CMC, which possesses lower correlation coefficients, particularly as the lead time increases. The lack of absolute coincidence in time between the satellite observations and model fields could be responsible for a slight degradation in the absolute correlation values. Limiting the analysis to only those days where an AR was found to be present reduced the correlations only by about 0.01 since ARs were detected on such a large fraction of the days.

To address whether the models possess any bias in the forecast strength of ARs as represented by the IWV
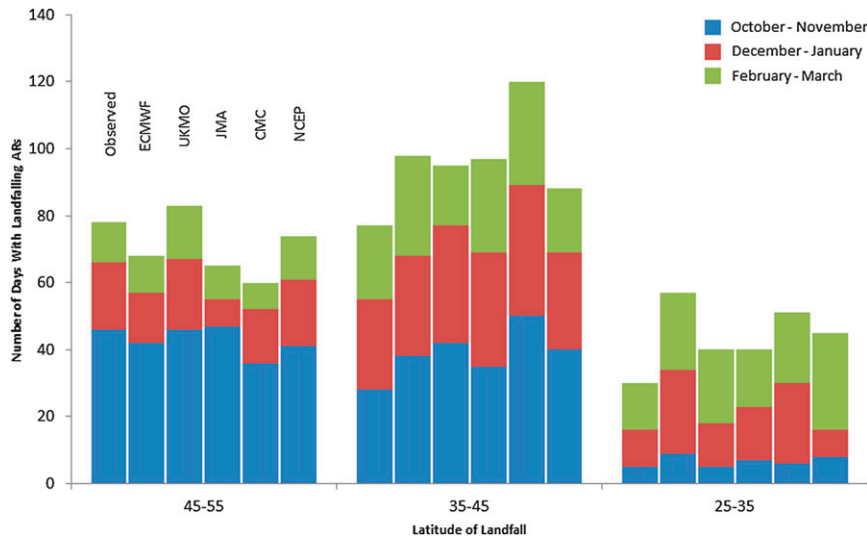
FIG. 6. Occurrence of observed and modeled landfalling ARs stratified by latitude of landfall and season. Model results are shown for 7-day forecasts. The heights of the bars represent the numbers of days with landfalling ARs detected in the individual bins. A total of 455 days were included in the analysis. Individual columns reflect the satellite observations and different models as shown. Time periods are October–November (blue), December–January (red), and February–March (green).

content along the AR axis, we compared both average values along different portions of the AR and peak values near landfall. Forecast minus observed differences were computed for each AR and then averaged. The apparent bias in the IWV content along the axis of the AR averaged over the entire length of the detected AR is presented as a function of forecast lead time in Fig. 8a. The corresponding average observed value of the core IWV content is 3.1 cm. The results suggest a slight overall moist bias (up to ~6%) along the AR axis for each of the models. The biases here are all typically greater than the average overall bias in the IWV fields presented in Fig. 2, suggesting a slight overestimate of the strength of the ARs. With the exception of the CMC and JMA results, which suggest an increase in the bias with increasing forecast lead, the bias is largely independent of the forecast lead time. The average results shown here are largely consistent with those obtained for core IWV values within individual latitude bands (not shown).

The corresponding comparison of the peak IWV values in the vicinity of landfall is shown in Fig. 8b. The peak values were taken as the maximum IWV content along the detected AR axis within a distance of 100–200 km offshore from the coastline. Observed IWV values within 100 km of the coast are subject to potential sidelobe contamination in the microwave satellite data resulting from the proximity of the bright land surface. Comparisons were only performed for those cases where a landfalling AR was detected in both the observed and forecast

fields. The average observed peak IWV value corresponding to the computed biases is 2.8 cm. Visually subtracting the overall bias in the IWV fields from Fig. 2 results in relatively consistent biases near or just below 0 cm for all the models. This suggests little systematic bias in the peak IWV values near landfall in contrast to the overall moist bias noted above. There again is no significant consistent trend for this bias to change with forecast lead. If observations within 100 km of the coast are included, there is an increased suggestion of a dry bias in the models. While this could be a result of the
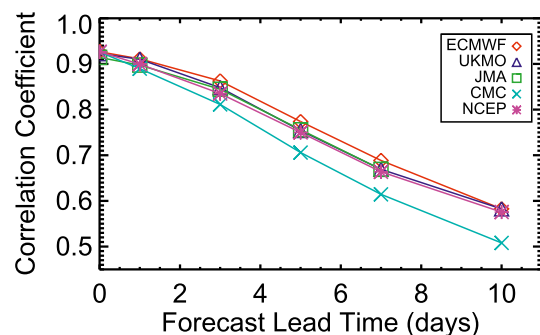

FIG. 7. Correlation between modeled and observed IWV fields over the entire analysis domain computed on a grid-cell by grid-cell basis for data mapped onto a common grid. The correlation coefficients were computed using all collocated data over the three cool seasons analyzed. Results were virtually identical if the analysis was restricted to days with detected ARs.

Average Value Over Entire AR

Peak Value Near Landfall

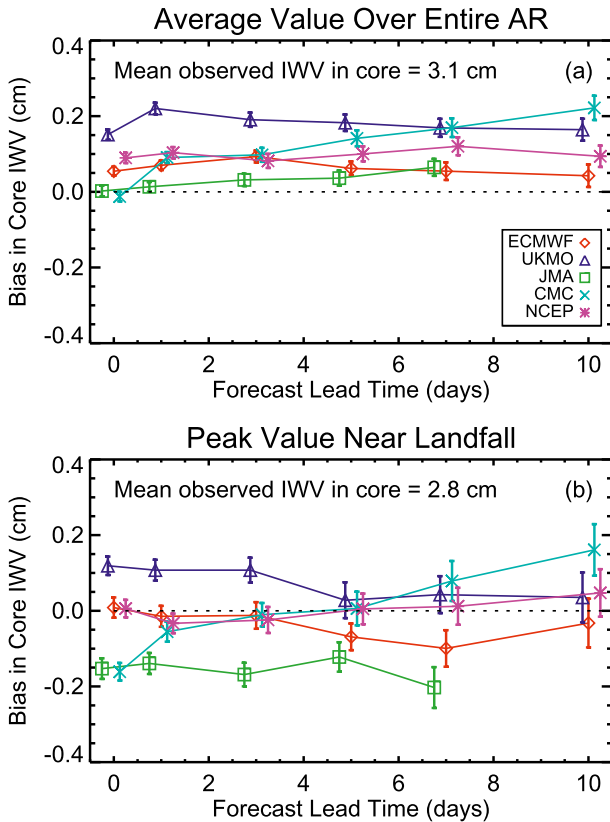## Average Value Over Entire AR

Average Value Near Landfall

FIG. 8. Bias in the modeled AR strength as represented by the IWV content along the axis of the AR relative to satellite-derived observations computed both (a) for an average over the entire length of the AR and (b) for the maximum value along the detected AR axis between 100 and 200 km of the coast. The corresponding average values for the observed IWV contents are annotated. The error bars represent ±1 standard deviation of the mean to reflect uncertainty in the mean value. The horizontal position of the points has been offset slightly to avoid overlap of the error bars.
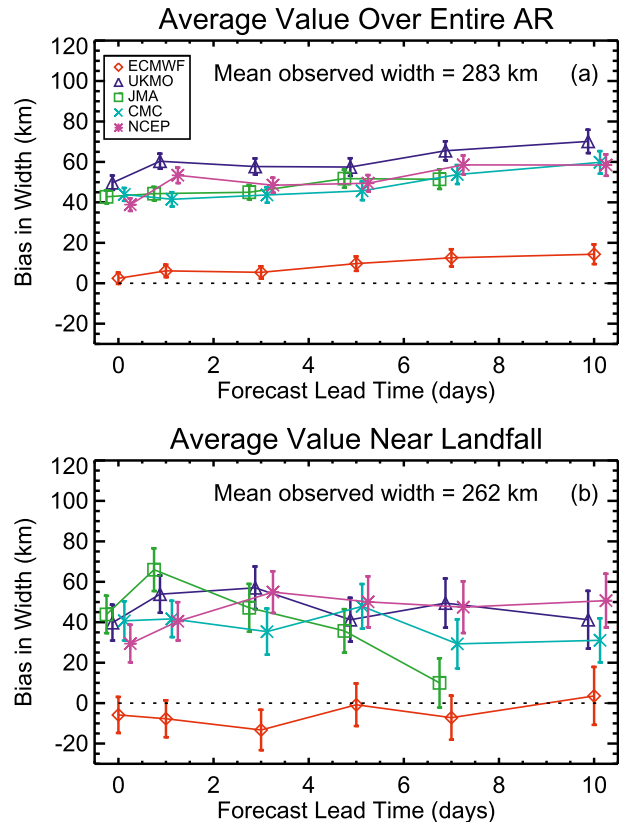
FIG. 9. Bias in the forecast AR width relative to satellite-derived observations computed both (a) for an average over the entire length of the AR and (b) for an average within 100–200 km offshore of the coast. The width estimate used was the minimum width computed relative to the 2.0-, 2.33-, and 2.67-cm IWV thresholds. The corresponding average values for the observed width are annotated. The error bars represent ±1 standard deviation of the mean to reflect uncertainty in the mean value. The horizontal position of the points has been offset slightly to avoid overlap of the error bars.

models not capturing convergence of the IWV near landfall, this influence cannot be adequately distinguished from the possibility of a moist bias in the observations close to land.

The reduced average peak IWV value near landfall relative to the overall average IWV value reflects the fact that the largest core IWV values are typically observed in the more southern "upstream" region of an AR in areas of greater background water vapor content. Enhancement of the core IWV value can occur near landfall due to convergence and orographic influences but this enhancement is typically smaller than the variation over the entire length of the AR.

We next considered the question of how well the models reproduced the AR width, particularly given their differences in resolution. As for the IWV content along the AR axis, the bias in the estimated AR width averaged over the entire length of the AR is shown in Fig. 9a. The

average observed width based on the selected minimum threshold approach is 283 km and the biases can be interpreted relative to this value. The results demonstrate the superior performance of the ECMWF predictions and a significant overestimate of the width by all other models. The biases clearly reflect the impact of model resolution as the models with near 1° resolution spread the signature of the AR over a larger region than the ECMWF model at 0.5° resolution. It is important to note that the average AR widths involved in this comparison were on the order of a few degrees so that the coarser-resolution models are still able to resolve the features. An additional comparison (not shown) of the ECMWF results against width estimates derived from satellite observations gridded at 0.25° resolution showed a positive bias on the order of 30 km, suggesting there is still spreading of the features in 0.5°-resolution models. Regridding the

ECMWF results at 1° resolution resulted in biases of similar order to the other 1°-resolution models. The models all (including the ECMWF results relative to the observations gridded at 0.25° resolution) exhibit a general tendency for the width of the features to be over-estimated by a larger amount at longer forecast lead times, but it is unclear if the increase in bias is statistically significant. The variability in the differences of the width estimates relative to the observations is also observed to increase with increasing lead as reflected by the size of the error bars.

The results constrained to differences in width between 100 and 200 km offshore of the coastline for landfalling ARs are shown in Fig. 9b. The corresponding average observed width is 262 km. The biases again reflect the difference in model resolution with greater spreading of the AR features in the models with near 1° resolution. Near the coastline, however, the apparent positive biases are slightly reduced and the ECMWF results now demonstrate a small negative bias. Visual analysis of the satellite observations suggests some widening of ARs as they approach large-scale terrain and the results here might indicate that the models do not capture this change. There is no significant tendency for changes in width bias at landfall with increasing lead time but the variability in the differences again increases. The variability in the width estimates is notably larger near landfall than for the overall average.

The peak IWV content and average width within 100–200 km offshore of the coastline for landfalling ARs (with no model–observed coincidence requirement) were then stratified and compared as a function of landfall latitude and season as for the frequency of occurrence. The observations suggested no clear dependence of the width at landfall with latitude, so the results are presented only as a function of month of occurrence in Fig. 10. The model-based results are shown for a forecast lead time of 7 days. This lead was selected since model differences are greater at longer leads and this is the longest lead for which data from all five models are available. The observed IWV content along the AR axis (Fig. 10a) shows a clear decrease as the cool season progresses and this tendency is captured by all the models. The relative performance of the different models with respect to the core IWV content shows no apparent change with month. The observed and modeled AR widths (Fig. 10b) show no clear seasonal trend but the model overestimates of width appear greater in the middle and later portions of the season. For the 1°-resolution models the average AR width is generally close to the observations in October–November but then are overestimated in the later months, especially December–January. It is unclear at present if there is any physical reason to expect this behavior.
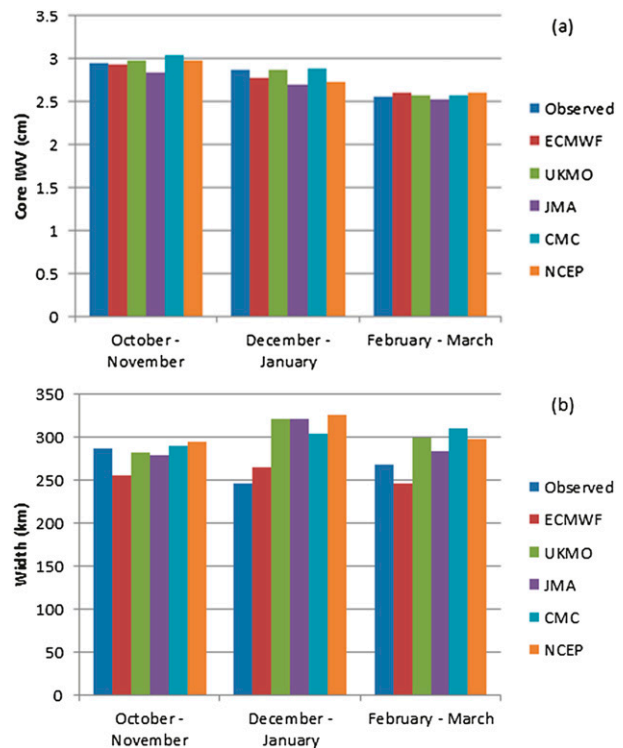

FIG. 10. Mean observed and modeled (a) peak IWV content along the AR axis and (b) average width, within 100–200 km offshore of the coastline for landfalling ARs stratified by season of occurrence. Values were computed independently over all ARs detected within each separate data type. No coincidence in date was required between the observed and forecast results. Modeled results are shown for forecasts with 7-day lead.

While selected to facilitate comparison of observed and modeled results, the AR width computed as the minimum width relative to the different IWV thresholds as displayed in Fig. 10 is not necessarily the most appropriate value for physical interpretation. AR widths of approximately 250–300 km obtained via this method are generally narrower than those discussed in previous publications (e.g., Ralph et al. 2004) relative to a single threshold of 2 cm. For averages within 100–200 km offshore of the coastline for observed landfalling ARs, the average width computed relative to the 2-cm threshold is 510 km. This contrasts with a corresponding 262 km for the minimum width relative to the multiple IWV thresholds. Different width estimates should be considered for AR climatological studies.

c. Position

Finally, we considered the question of how well the positions of ARs are forecast, looking specifically at the predicted location of landfall. Unique data points were considered for every day an AR contacted the coastline, meaning that a single AR could contribute multiple
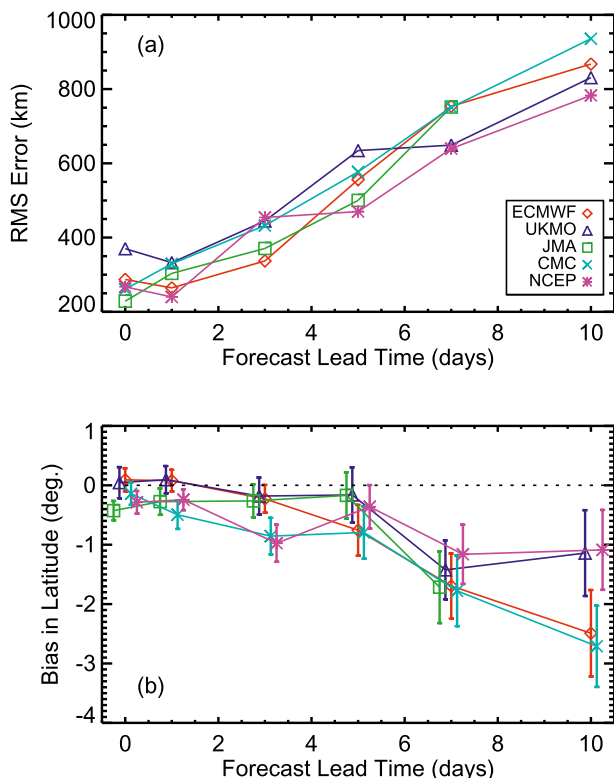
FIG. 11. Estimates of error in forecast AR landfall location as a function of lead time. (top) The total RMS error (km) in the detected landfall location along the west coast. (bottom) The bias in the latitude of the detected landfall relative to the satellite observations. The error bars represent ±1 standard deviation of the mean to reflect uncertainty in the mean value.

times over the course of its lifetime. The errors in forecast landfall location are presented in terms of an overall RMS error in landfall position and a latitudinal bias in Fig. 11. The detected landfall location is taken as the point where the detected AR axis makes contact with the coastline. If multiple contiguous axis points were found to touch the coast, the position was taken as the median of these values. The derived RMS error in landfall position is seen to increase significantly with forecast lead, growing from values near 200 km to more than 800 km at 10-day lead. No single model clearly performs best, though the NCEP model performs reasonably well in this comparison. Interestingly, these estimated position errors are generally larger than the National Hurricane Center official annual average track errors for tropical storms and hurricanes, which, in 2011, varied from ~200 km at 3-day lead to ~400 km at 5-day lead (see http://www.nhc.noaa. gov/verification/verify5.shtml), while largely consistent (after day 1) with extratropical cyclone position errors determined by Froude (2010).

While difficult to directly quantify, it is important to consider the uncertainties contributing to the derived

landfall position errors. First, errors associated with the time of landfall cannot be resolved in this analysis and could be a component of the apparent spatial biases. Assuming a potential time difference of ~2 h between the observations and forecast valid times would correspond to a spatial offset of ~100 km, as noted previously. Additionally, the dilation of the AR axis (see Wick et al. 2013) to ensure proper continuity detection and landfall identification leads to potential spreading of the axis position by one or two grid cells, adding additional variability on the order of ~50 km. Coupled with the fact that the landfall could be naturally spread over a range of latitudes, it is challenging to determine a precise landfall location for comparison. Considering these factors, the landfall position errors are not inconsistent with hurricane track forecast errors. Even without allowing for all the possible uncertainties, the errors at 1-day lead agree with forecaster experience of errors in predictions of frontal passage along the west coast (D. Reynolds 2013, personal communication).

The errors in position generally correspond to a southerly bias in landfall location, except for ECMWF and UKMO, which exhibit little mean bias out to about 3-day forecasts. The results suggest an increase in this southerly bias with increasing lead time for each of the models. Any potential negative or dry bias in core IWV content near landfall as suggested previously does not appear to be related to a bias in landfall location. A southerly bias in landfall location would be expected to correspond to slightly moister IWV values due to mean IWV distributions. A delay in the time of the observations relative to the forecasts could contribute to an apparent southerly bias.

Considering positional uncertainty in the open ocean, we also examined any longitudinal bias in the position within specific latitudinal bands. The results were rather noisy, revealing little systematic behavior, but the agreement was typically within 2°. Differences in orientation angle were also quite variable, yielding little additional insight into the relative model performance. For both quantities (not shown), the variability (as reflected by the standard deviations of the mean) did increase notably with increasing forecast lead time.

5. Discussion and conclusions

The ability of ensemble forecast systems from five leading forecasting centers to accurately predict and reproduce the water vapor signature of ARs was evaluated through feature-based comparisons between their IWV fields and satellite-derived observations from SSM/I and SSM/IS. The assessment focused on the models' representation of the occurrence of AR events, the AR

strength as represented by the IWV content along the AR axis, the AR width, and the position at landfall. The study was performed for a region of the northeast Pacific Ocean bordering the coast of North America over three cool seasons from 2008–09 to 2010–11. The overall presence of ARs was well forecast, even out to 10-day lead times, but the forecasts for landfall occurrence were less accurate. Significant errors were observed in the forecast position of landfall, particularly at longer lead times.

Comparisons were based on ''AR days'' rather than distinct ARs. The prediction of occurrence of ARs was evaluated in terms of the number of days where at least one AR was found to be present. The direct comparison of AR characteristics was performed on a daily basis for those cases where exactly one AR was detected in both the observed and forecast IWV fields. Landfall occurrence was also evaluated using a relaxed 2-day window. Further enhancements to application of the feature detection technique are required to perform analyses on distinct ARs. While comparison of the landfall statistics between the 1- and 2-day arrival windows provides some insight into time errors in the forecasts, the lack of isolation of distinct ARs prevents us from formally assessing model uncertainties related to leads or lags in time of landfall or duration of occurrence.

Model resolution was found to be important for accurate representation of detailed AR characteristics, but realistic ARs were still predicted by the coarser-resolution models. Indeed, Dettinger (2011) has also shown the presence of ARs in climate models. While the performance for each of the different models was largely similar, the ECMWF forecasts benefited in several tests from the model's finer 0.5° resolution. The difference between models was greatest for the apparent width of the ARs with all models except ECMWF significantly overestimating the AR width, both overall and near landfall, due to their coarser resolution. While comparisons with observations gridded at the same resolution as ECMWF suggested small biases in forecast AR width, comparisons with observations at their native resolution suggested that there may still be some spreading in the width of features in the ECMWF forecasts. The overall degradation in forecast skill with increasing lead time was similar for all models for all quantities evaluated, with each model generally maintaining its same performance relative to the other models. Model resolution had less of an impact on the IWV content along the AR axis, with all models demonstrating a moist bias for averages over the entire length of the AR.

This work represents the first detailed diagnostic of AR predictability on the U.S. west coast and the results have significant implications for forecasting of extreme precipitation events in that region. Some results highlight

an encouraging level of predictability, such as whether an AR will hit the West Coast within the coming week and whether it will contain relatively large water vapor content. Conversely, a major forecast challenge is pinpointing the location of landfall and the duration of AR conditions in key watersheds. Forecasts of the location, timing, and duration of landfall are critical to precipitation forecasts and the issuance of warnings. The results at landfall also have major implications for forecast-based dam operations, which typically require many days lead time. Although position errors decrease as landfall approaches, errors on the order of hundreds of kilometers pose a serious problem for water managers and flood forecasters whose watersheds are smaller than that error. Improvements in the model predictions of the location (and timing) of AR landfall are desirable for improved warnings. The identified forecast biases offer potential for future work to diagnose their causes and yet also hold promise for forecasters to consider the results directly as they prepare their human-in-the-loop forecasts. NOAA's Hydrometeorology Testbed (HMT; hmt.noaa.gov) and its major partners are pursuing several avenues of research and related actions aimed at improving both our understanding of ARs as well as their monitoring and prediction.

Given that ARs are intimately connected with the dynamics of midlatitude cyclones, one would expect a relationship between the accuracy of forecasts of ARs and extratropical cyclones. In fact, the landfalling AR position errors derived here broadly agree with position errors of extratropical cyclones in ensemble prediction system forecasts found by Froude (2010). While slightly larger, the landfalling position errors are also not inconsistent with current hurricane track forecast errors. Increased errors in landfalling AR position relative to tropical cyclones could be associated with contortions and bending occurring during frontal passage over the coast and limitations in the models' ability to treat lower boundary phenomena.

The results are clearly predicated on the performance of the ARDT. While the performance of this automated tool has been found to be very good and the direct comparisons performed here are the precise application for which it was designed, there are clearly uncertainties related to its function. We have attempted to document and limit these as much as possible, but it is important to emphasize that any manual approach is also subjective in nature.

The implications for assessing AR prediction are also affected by utilizing IWV patterns rather than water vapor transport. A visual comparison of IWV and water vapor transport patterns in real-time forecasts, however, supports past work suggesting that IWV generally

reflects the existence of ARs quite well. The strength is less well characterized and a relatively small percentage of the ARs identified at the lowest IWV thresholds fail to have a corresponding well-defined IVT signature. Where there is a water vapor transport signature, the ARs at low IWV thresholds also often correspond to more isolated regions of transport rather than major transport corridors extending deeper into the tropics.

Though additional comparisons with other methods would be valuable, the comparisons with cyclone position errors are encouraging, and this work provides a baseline for the evaluation of future AR predictions. Conducting a similar study of the skill of existing climate models to reproduce ARs would also be very valuable to help determine to what degree projections of changing AR activity in a changing climate might be reliable.

REFERENCES

Bougeault, P., and Coauthors, 2010: The THORPEX Interactive Grand Global Ensemble. *Bull. Amer. Meteor. Soc.,* **91,** 1059–1072.

Bowler, N. E., A. Arribas, K. R. Mylne, K. B. Robertson, and S. E. Beare, 2008: The MOGREPS short-range ensemble prediction system. *Quart. J. Roy. Meteor. Soc.,* **134,** 703–722.

Buizza, R., J.-R. Bidlot, N. Wedi, M. Fuentes, M. Hamrud, G. Holt, and F. Vitart, 2007: The new ECMWF VAREPS (Variable Resolution Ensemble Prediction System). *Quart. J. Roy. Meteor. Soc.,* **133,** 681–695.

Charron, M., G. Pellerin, L. Spacek, P. L. Houtekamer, N. Gagnon, H. L. Mitchell, and L. Michelin, 2010: Toward random sampling of model error in the Canadian Ensemble Prediction System. *Mon. Wea. Rev.,* **138,** 1877–1901.

Clark, W., H. Yuan, T. L. Jensen, G. Wick, E. I. Tollerud, R. G. Bullock, andE. Sukovich, 2011: Evaluation of GFS water vapor forecasting errors during the 2009-2010 West Coast cool season using the MET/MODE object analysis package. *Preprints, 25th Conf. on Hydrology,* Seattle, WA, Amer. Meteor. Soc., 378. [Available online at https://ams.confex.com/ams/91Annual/webprogram/Paper183894.html.]

Dettinger, M. D., 2011: Climate change, atmospheric rivers and floods in California—A multimodel analysis of storm frequency and magnitude changes. *J. Amer. Water Resour. Assoc.,* **47,** 514–523.

——, F. M. Ralph, T. Das, P. J. Neiman, and D. Cayan, 2011: Atmospheric rivers, floods, and the water resources of California. *Water,* **3,** 455–478.

Froude, L. S. R., 2010: TIGGE: Comparison of the prediction of Northern Hemisphere extratropical cyclones by different ensemble prediction systems. *Wea. Forecasting,* **25,** 819–836.

Guan, B., N. P. Molotch, D. E. Waliser, E. J. Fetzer, and P. J. Neiman, 2010: Extreme snowfall events linked to atmospheric rivers and surface air temperature via satellite measurements. *Geophys. Res. Lett.,* **37,** L20401, doi:10.1029/2010GL044696.

Kunkee, D. B., S. D. Swadley, G. A. Poe, Y. Hong, and M. F. Werner, 2008: Special Sensor Microwave Imager Sounder (SSMIS) radiometric calibration anomalies—Part I: Identification and characterization. *IEEE Trans. Geosci. Remote Sens.,* **46,** 1017–1033.

Mastin, M. C., A. S. Gendaszek, and C. R. Barnas, 2010: Magnitude and extent of flooding at selected river reaches in western Washington, January 2009. U.S. Geological Survey Scientific Investigations Rep. 2010–5177, 34 pp.

McMurdie, L. A., and C. Mass, 2004: Major numerical forecast failures over the northeast Pacific. *Wea. Forecasting,* **19,** 338–356.

——, and J. H. Casola, 2009: Weather regimes and forecast errors in the Pacific Northwest. *Wea. Forecasting,* **24,** 829–842.

Mears, C. A., B. D. Santer, F. J. Wentz, K. E. Taylor, and M. F. Wehner, 2007: Relationship between temperature and precipitable water changes over tropical oceans. *Geophys. Res. Lett.,* **34,** L24709, doi:10.1029/2007GL031936.

Neiman, P. J., F. M. Ralph, G. A. Wick, Y.-H. Kuo, T.-K. Wee, Z. Ma, G. H. Taylor, and M. D. Dettinger, 2008a: Diagnosis of an intense atmospheric river impacting the Pacific Northwest: Storm summary and offshore vertical structure observed with COSMIC satellite retrievals. *Mon. Wea. Rev.,* **136,** 4398–4420.

——, ——, ——, J. D. Lundquist, and M. D. Dettinger, 2008b: 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. *J. Hydrometeor.,* **9,** 22–47.

——, L. J. Schick, F. M. Ralph, M. Hughes, and G. A. Wick, 2011: Flooding in western Washington: The connection to atmospheric rivers. *J. Hydrometeor.,* **12,** 1337–1358.

Ralph, F. M., and M. D. Dettinger, 2012: Historical and national perspectives on extreme West Coast precipitation associated with atmospheric rivers during December 2010. *Bull. Amer. Meteor. Soc.,* **93,** 783–790.

——, P. J. Neiman, and G. A. Wick, 2004: Satellite and CALJET aircraft observations of atmospheric rivers over the eastern North Pacific Ocean during the winter of 1997/98. *Mon. Wea. Rev.,* **132,** 1721–1745.

——, ——, and R. Rotunno, 2005: Dropsonde observations in low-level jets over the northeastern Pacific Ocean from CALJET-1998 and PACJET-2001: Mean vertical-profile and atmospheric-river characteristics. *Mon. Wea. Rev.,* **133,** 889–910.

——, G. A. Wick, S. I. Gutman, M. D. Dettinger, D. R. Cayan, and A. B. White, 2006: Flooding on California's Russian River: The role of atmospheric rivers. *Geophys. Res. Lett.,* **33,** L13801, doi:10.1029/2006GL026689.

——, E. Sukovich, D. Reynolds, M. Dettinger, S. Weagle, W. Clark, and P. J. Neiman, 2010: Assessment of extreme quantitative precipitation forecasts and development of regional extreme event thresholds using data from HMT-2006 and COOP observers. *J. Hydrometeor.,* **11,** 1288–1306.

Trenberth, K. E., J. Fasullo, and L. Smith, 2005: Trends and variability in column integrated atmospheric water vapor. *Climate Dyn.,* **24,** 741–758, doi:10.1007/s00382-005-0017-4.

Wentz, F. J., 1995: The intercomparison of 53 SSM/I water vapor algorithms. Remote Sensing Systems Tech. Rep. on the WetNet Water Vapor Intercomparison Project (VIP), Remote Sensing Systems, Santa Rosa, CA, 19 pp.

——, 1997: A well-calibrated ocean algorithm for Special Sensor Microwave/Imager. *J. Geophys. Res.,* **102** (C4), 8703–8718.

——, L. Ricciardulli, K. Hilburn, and C. A. Mears, 2007: How much more rain will global warming bring? *Science,* **317,** 233–235.

White, A. B., and Coauthors, 2012: NOAA's rapid response to the Howard A. Hanson Dam flood risk management crisis. *Bull. Amer. Meteor. Soc.,* **93,** 189–207.

Wick, G. A., Y.-H. Kuo, F. M. Ralph, T.-K. Wee, and P. J. Neiman, 2008: Intercomparison of integrated water vapor retrievals from SSM/I and COSMIC. *Geophys. Res. Lett.,* **35,** L21805, doi:10.1029/2008GL035126.

——, P. J. Neiman, and F. M. Ralph, 2013: Description and validation of an automated objective technique for identification and characterization of the integrated water vapor signature of atmospheric rivers. *IEEE Trans. Geosci. Remote Sens.,* **51,** 2166–2176, doi:10.1109/TGRS.2012.2211024.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences.* 2nd ed. Academic Press, 627 pp.

Yan, B., and F. Weng, 2008: Intercalibration between Special Sensor Microwave Imager/Sounder and Special Sensor Microwave Imager. *IEEE Trans. Geosci. Remote Sens.,* **46,** 984–995.

Zhu, Y., and R. E. Newell, 1998: A proposed algorithm for moisture fluxes from atmospheric rivers. *Mon. Wea. Rev.,* **126,** 725–735.