Spatial Bias in Medium-Range Forecasts of Heavy Precipitation in the Sacramento River Basin: Implications for Water Management⊘

ZACHARY P. BRODEUR AND SCOTT STEINSCHNEIDER

Department of Biological and Environmental Engineering, Cornell University, Ithaca, New York

ABSTRACT

Forecasts of heavy precipitation delivered by atmospheric rivers (ARs) are becoming increasingly important for both flood control and water supply management in reservoirs across California. This study examines the hypothesis that medium-range forecasts of heavy precipitation at the basin scale exhibit recurrent spatial biases that are driven by mesoscale and synoptic-scale features of associated AR events. This hypothesis is tested for heavy precipitation events in the Sacramento River basin using 36 years of NCEP medium-range reforecasts from 1984 to 2019. For each event we cluster precipitation forecast error across western North America for lead times ranging from 1 to 15 days. Integrated vapor transport (IVT), 500-hPa geopotential heights, and landfall characteristics of ARs are composited across clusters and lead times to diagnose the causes of precipitation forecast biases. We investigate the temporal evolution of forecast error to characterize its persistence across lead times, and explore the accuracy of forecasted IVT anomalies across different domains of the North American west coast during heavy precipitation events in the Sacramento basin. Our results identify recurrent spatial patterns of precipitation forecast error consistent with errors of forecasted synoptic-scale features, especially at long (5–15 days) leads. Moreover, we find evidence that forecasts of AR landfalls well outside of the latitudinal bounds of the Sacramento basin precede heavy precipitation events within the basin. These results suggest the potential for using medium-range forecasts of large-scale climate features across the Pacific–North American sector, rather than just local forecasts of basin-scale precipitation, when designing forecast-informed reservoir operations.

1. Introduction

Water resources management in northern California is uniquely challenging because regional precipitation varies dramatically on both intra and interannual time scales, inducing stark tradeoffs between flood risk reduction and water supply services provided by regional infrastructure. The variability in precipitation is intimately tied to the space–time dynamics of atmospheric rivers (ARs), defined as long, narrow, and transient corridors of strong horizontal integrated water vapor transport (IVT) that are typically associated with a low-level jet ahead of the cold front of an extratropical cyclone (American Meteorological Society 2020). Approximately half of the precipitation in the region is delivered by a relatively small number of AR events during the cold season (Dettinger et al. 2011), and so the absence of a handful of ARs can produce annual precipitation deficits and reductions in snowpack and subsequent streamflow that require drawdowns of surface reservoir and groundwater storage to meet demand (Ralph et al. 2014). In addition, ARs often deliver precipitation in the form of heavy or extreme events (Steinschneider et al. 2016; Hecht and Cordeira 2017), and so pose a direct threat to communities by raising the risk of flooding, high winds, and mudslides (Ralph et al. 2019; Waliser and Guan 2017; Konrad and Dettinger 2017).

The dual role of ARs as the primary driver of both water supplies and flood hazards has prompted significant interest in their characterization (Dacre et al. 2015; Guirguis et al. 2018; Hecht and Cordeira 2017), classification (Dettinger et al. 2018; Ralph et al. 2019), and predictability (Baggett et al. 2017; DeFlorio et al. 2018a,b; Lavers et al. 2016, 2017). Predictive skill at medium range (1–14 days) and subseasonal to seasonal (S2S; 15–90 days) time scales has become a topic of particular interest because of its relevance to water infrastructure management

decisions, such as forecast-informed reservoir operations (Nayak et al. 2018; Forecast Informed Reservoir Operations Steering Committee 2017). In this study we focus on characterizing the spatial bias of medium-range forecasts of heavy precipitation in the Sacramento River basin, with the goal of developing insights for how forecast-informed reservoir operations can better leverage forecasts of ARs and associated precipitation at longer lead times.

The use of extended forecasts to inform reservoir operations has been limited by the relatively slow increase in forecast skill at long leads (approximately one day in forecast skill per decade; Bauer et al. 2015; Alley et al. 2019). However, in specific regions like California where extreme events are governed by a limited number of synoptic-scale atmospheric mechanisms (i.e., ARs), measures of those mechanisms can yield gains in forecast skill beyond the global average. Because ARs are defined by synoptic-scale filaments of IVT, there is a highly observable signature associated with their formation and promulgation. These atmospheric features are also explicitly resolved in forecast models, and so can be more accurately predicted at longer lead times than unresolved processes (e.g., localized convection) (Bauer et al. 2015; Holton and Hakim 2013). Still, at long leads (1–2 weeks), the use of large-scale atmospheric predictors like IVT fields are not likely to produce accurate forecasts of heavy precipitation at specific locations, in part because localized features that can significantly alter precipitation distributions (e.g., the Sierra barrier jet; IVT impingement angle on the local topography) are unlikely to be accurately predicted at those lead times (Nardi et al. 2018; Hecht and Cordeira 2017; Ralph et al. 2016). Rather, long-lead AR forecasts are likely to be most useful in estimating a shift in the odds of heavy precipitation events, which could then be embedded into risk-based operating strategies for water infrastructure (Forecast Informed Reservoir Operations Steering Committee 2017; Nayak et al. 2018).

For example, the ability to simply know in advance whether or not a significant AR event is likely to impact a given basin may be sufficient to inform certain decisions. Nayak et al. (2018) recently showed that optimal, forecast-informed reservoir operations favored reservoir levels that were far above the status quo conservation pool most of the year to bolster water supplies, except when 3-day ahead forecast information suggested an impending high flow event. Reservoir levels were then drawn down to the conservation pool ahead of the event to make space behind the dam to capture flood waters. This policy was effective even with a significant degree of forecast uncertainty, because the only information that was needed to initiate the drawdown was the detection of an incoming event, rather than a precise

forecast of its magnitude. If physical and environmental ramping constraints prevent rapid changes in downstream releases, such a policy would likely require longer lead information (5–15 days) that would suffer from a higher degree of uncertainty. However, given the conservative nature of water managers (Ralph et al. 2014; Loucks and Van-Beek 2017), there is likely a preference to err on the side of flood risk safety, i.e., a willingness to accept a higher number of incorrect forecasts of AR hits that lead to unnecessary reservoir drawdowns (false positives) in order to reduce or eliminate the number of forecast misses that result in an inability to manage flood waters (false negatives).

Under this asymmetric set of preferences, there are two prominent features of AR forecasts that may constitute "useful" information for water resource management. First is the accuracy with which the location of an AR landfall can be predicted. The positioning and strength of large-scale pressure anomalies relate strongly to the location of landfall, raising the potential for longer lead predictability driven by teleconnections (Hu et al. 2017; Ralph et al. 2011; Guirguis et al. 2018; Hecht and Cordeira 2017). For instance, more than 70% of ARs in northern California are associated with 2 of the 15 rotated empirical orthogonal functions (REOF) of 500-hPa height anomalies identified by Guirguis et al. (2018). Similarly, the characterization of Rossby wave breaking (RWB) as either anticyclonic (AWB) or cyclonic (CWB) has shown a marked correlation with landfall latitude. CWB-correlated ARs tend to landfall farther south than their AWB counterparts, reflecting an important teleconnection that is also highly correlated to ENSO (Hu et al. 2017). Other teleconnections between AR landfall location and various atmospheric modes show promise for "forecasts of opportunity," including the coupled impacts of the MJO and the QBO (Mundhenk et al. 2016, 2018; Baggett et al. 2017), as well as less predictable modes such as the PDO, PNA, EPO, and WPO (Guirguis et al. 2019; DeFlorio et al. 2018a; Brands et al. 2017).

The second feature of concern for water management is the accuracy with which physical characteristics of an AR can be resolved, such as its intensity and orientation, which can significantly impact precipitation distributions (Ralph et al. 2016, 2019). As above, the positioning of pressure anomalies and RWB characteristics influence the strength and orientation of IVT at landfall (Guirguis et al. 2018; Hu et al. 2017), but these influences are weaker than for landfall location. Thus, these features are more likely be forecasted with precision at shorter (<7 days) lead times.

Regardless of lead time, water resources management decisions need to be based on AR forecasts relevant at the basin level. The degree of forecast resolution required will

be highly dependent on the basin being considered and the inherent biases in the forecasting model for that basin. If a model tends to project precipitation or AR features with systematic biases relative to that basin, these biases should be incorporated into forecast-informed control rules. For instance, release decisions tailored for conservative flood risk management could be triggered by a longer-lead forecast of AR landfall outside the basin of interest if forecasts consistently show spatial biases in landfall at that lead time. Recent studies have categorized forecast biases in AR landfalls, but they have generally concentrated on all AR events across large spatial scales such as the entire North American west coast (Nardi et al. 2018; DeFlorio et al. 2018a). There remains a gap in our understanding of systematic errors in AR forecasts at the river basin scale, which is of critical importance to water resource applications.

The goal of this work is to evaluate structural biases in medium-range forecasts (1–15 days) of heavy precipitation events and associated AR characteristics, with an emphasis on their implications for forecast-informed water management. In particular, we address the following research questions: 1) What spatial patterns of error exist in precipitation forecasts across western North America during heavy precipitation events in the Sacramento River basin? 2) How do these patterns correspond to systematic forecast biases in synoptic and mesoscale features of ARs? 3) How persistent are these error structures and how do they evolve through forecast lead times? and 4) How does the selection of landfall domain influence the utility of forecast information? To answer these questions, we identify spatial clusters of precipitation forecast error from a 36-yr hindcast using the NCEP Global Ensemble Forecast System Reforecast (GEFS/R). These clusters are used to highlight synoptic-scale features that can help diagnose the causes of major patterns of precipitation forecast error. We investigate the temporal evolution of error clusters to characterize their persistence across forecast lead times. Finally, we assess the accuracy of forecasted IVT anomalies across different domains of the North American west coast during heavy precipitation events in the Sacramento basin, in order to explore whether forecast-informed water management could utilize a broader set of forecast information that extends beyond forecasts of climate directly over the basin of interest. We conclude with a discussion of future research needed to assimilate systematic forecast errors into more robust forecast-informed operating policies for water resources infrastructure.

2. Data and methods

a. Data

Precipitation in the Sacramento basin was gathered from the 1° × 1° GPCC Full Data Daily Product V.2018

and First Guess Daily datasets (Schamm et al. 2013, 2015) within the geographical region bounded by 38°–42°N and 120°–123°W. This region is approximately 135 000 km^2 and corresponds to 12 different 1° grid cells (Fig. 1) (USGS 2013). The Full Data Daily Product is based on up to 35 000 stations per month and undergoes a full quality control before distribution (Schneider et al. 2016). GPCC's First Guess Daily Product uses approximately 7000 stations per month that have automated quality control features and a simplified interpolation scheme. These data are available near real time at the expense of some additional gauge-measurement and sampling error (Schamm et al. 2014; Schneider et al. 2018). For our study, the Full Data product was used from 1 December 1984 to 31 December 2016 with the First Guess product filling the remaining period through 31 March 2019. In the period of overlap between the Full Data and First Guess daily products (2009–16), we found the errors between them to be minimal, with a mean (standard deviation) of −0.098 mm (3.85 mm) across western North America.

We also collected GPCC data for a broader region of western North America that spans 30°–62°N and is bounded in the west by the coastline and in the east by a transect that runs between 114° and 128°W (see yellow region in Fig. 1). These data were used to examine precipitation forecast errors across a broader geographic region during heavy precipitation events in the Sacramento basin, and extend far enough inland to cover regions most often impacted by landfalling ARs (Rutz et al. 2015).

We opted to use the coarse-resolution GPCC products because their spatial and temporal extent allows us to examine forecasted precipitation errors across most of western North America and during extreme precipitation years like the cold season of 2016/17. Moreover, the coarse resolution of the available reforecast data negates many of the benefits of higher-resolution observations for our particular application. However, we did compare the GPCC estimates to the higher-resolution (0.0625°) data of Livneh et al. (2015b) in the region (30°–53°N, 115°–125°W) and time periods (1984–2013) where the two data products overlap. We found the errors to be relatively small, with a mean (standard deviation) of −0.33 mm (3.49 mm) across the region.

Reforecast data were retrieved from the NCEP GEFS/R version 2 reforecast dataset (NOAA/NCEP 2013). These 1° × 1° reforecasts are based on an 11-member ensemble and are catalogued daily from 1 December 1984 to present. We collected the following variables: accumulated precipitation, u wind, v wind, specific humidity, and geopotential height. The NCEP reforecasts are based on a single version
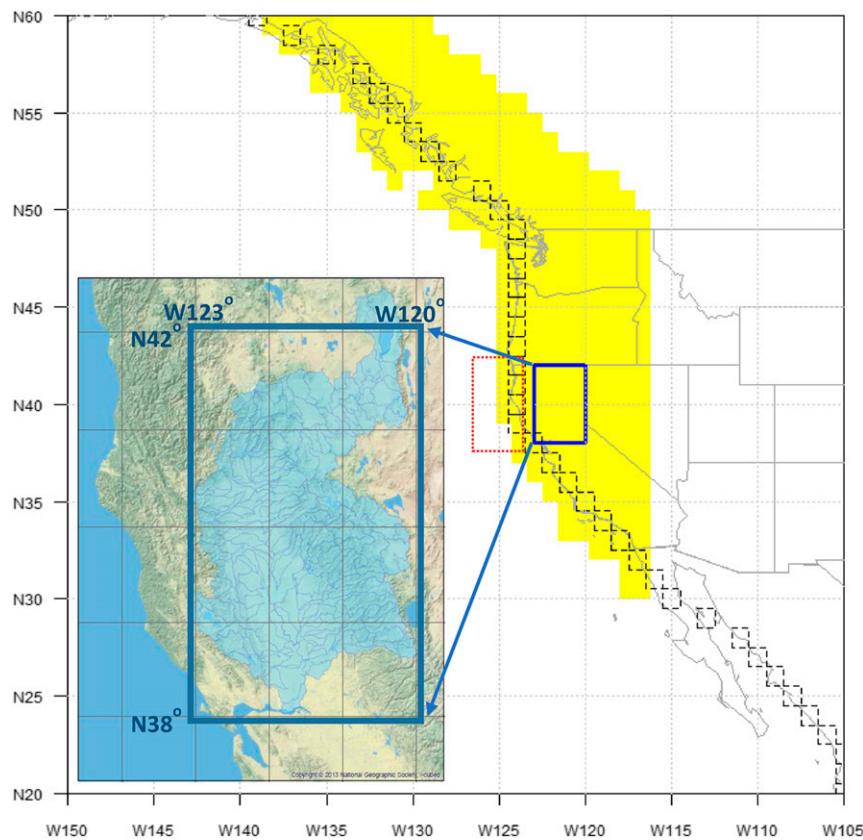
Fig. 1. Overview of study area. The yellow overlay indicates spatial extent of analyzed precipitation data; the dark blue box shows grid cells used to approximate the Sacramento River basin with light gray lines in inset showing individual 1° grid cells. Black dashed boxes along coast are ''landfall'' grid cells, and the red dashed box shows landfall region used to calculated IVT metrics for the Sacramento basin.

of the model initialized at 0000 UTC on each day in the reforecast period, which lends itself to categorization of bias and error since incremental changes to the model are not a confounding factor (Hamill et al. 2013). We utilized both the control and ensemble mean reforecast data in this study. The control was used in our primary analysis to better understand patterns of forecast bias and error structures in a single forecast trace and to avoid excessive smoothing in synoptic features across ensemble members. However, patterns of forecast error from the control were also confirmed using data from the ensemble mean across all analyses.

Reanalysis data were retrieved from the NCEP Reanalysis-2 (NOAA/NCEP 2002; Kanamitsu et al. 2002) archive at a 2.5° × 2.5° resolution for all variables (u wind, v wind, relative humidity, temperature, geopotential height) except accumulated precipitation, which was retrieved at its native T62 Gaussian resolution. These data were chosen over other reanalysis products because they utilize the same basic underlying model structure as the reforecast data, although with

an earlier version of the NCEP model. Reforecast and reanalysis values for IVT were calculated via the integration of u wind, v wind, and specific humidity across 1000–300 hPa as detailed in Zhu and Newell (1998). All datasets were subset for the period between 1 December 1984 and 31 March 2019 and bilinearly interpolated to match the 1° GPCC dataset for precipitation analyses or the 1° NCEP reforecast dataset for IVT and geopotential height analyses.

Finally, we used the Scripps Institution of Oceanography–Reanalysis-1 (SIO-R1) AR catalog (Gershunov et al. 2017) as our reference for the location and timing of landfalling AR events. To identify ARs, this catalog first identifies grid cells with absolute IVT > 250 kg m^{-1} s^{-1} and integrated water vapor (IWV) > 15 mm, and then identifies landfalling AR objects based on contiguous grid cells that cross the western coastline of North America and exhibit a length of at least 1500 km and a persistence of at least 18 h. The SIO-R1 AR catalog was particularly useful for this study because it includes tabular maximum IVT at landfall locations that can

be compared against similar information from the reforecasts.

b. Characterization of heavy precipitation events

We began the analysis by identifying historic occurrences of heavy precipitation in the Sacramento basin, defined as any date where precipitation rose above the 95th percentile of cold season [October–April (ONDJFMA)] events with accumulations greater than 1 mm for any grid cell in the basin (Hoerling et al. 2016). The final event tally was then filtered to remove duplicate events between grid cells. For multiday events, we identified all clusters of two or more consecutive heavy precipitation days, retained the first day of that cluster, and disregarded the remaining days. This focused our analysis on forecast skill at the start of heavy precipitation events. Of the resultant 295 events, two were manually removed due to missing data (3 December 2014, 26 February 2018) and seven events were removed for erroneous data (6 October 2010, 15 December 2011, 24 January 2013, 20 February 2013, 7 December 2013, 28 January 2015, 3 January 2019). The erroneous events were removed based on visual analysis that revealed either unrealistically high precipitation or distinct disagreement between the GPCC and Livneh gridded data. The final set of 286 heavy precipitation days contain all of the 49 "extreme daily precipitation" events found in Ralph et al. (2016), and only 24 of the 286 events were not associated (±1 day) with an AR event somewhere along western North America, as classified in the SIO-R1 dataset produced by Gershunov et al. (2017). For consistency, we compared heavy precipitation events assigned by GPCC to those assigned by the Livneh dataset in the same time period and found >85% commonality between the two catalogues.

c. Cluster analysis

For each heavy precipitation day, we calculated forecast error (i.e., differences between GPCC and control reforecast precipitation) for all 250 GPCC grid cells across western North America (see yellow region in Fig. 1). These forecast errors were produced separately for lead times between 1 and 15 days, although we focus primarily on lead times of 1, 5, 10, and 15 days. When comparing the observed and reforecast precipitation, we synchronized the GPCC dataset (based on daily observations recorded at midnight local time) to the NCEP reforecasts (based on UTC) at a 0600 UTC reference. This equates to a 2-h error in precipitation recording times between the two datasets, which was deemed acceptable for the purposes of this study. Additionally, since the NCEP reforecasts are only initialized at 0000 UTC, the forecast lead time includes an additional 6-h lag

(i.e., a 1-day forecast is actually a 30-h forecast; a 2-day forecast is a 54-h forecast). However, we continue to refer to forecast lead times by day (i.e., 1-day lead, 2-day lead, etc.).

For each lead time, the 286 heavy precipitation dates were clustered using a K-means algorithm based on the first 50 principal components (PCs) of the 286×250 forecast error matrix, which accounted for 79%–89% of the variance depending on forecast lead. PCs were used instead of raw error across all 250 grid cells to reduce data dimensionality and mitigate the influence of highly variable grid cells (Hannachi et al. 2007; Wilks 2011). The number of clusters to use was selected using scree plots of the ratio of between and total sum of squared errors. For most lead times there were only marginal gains in explained variance with more than four clusters. Therefore, four clusters were used for all forecast lead times for consistency.

Precipitation forecast errors were then composited over all events in each cluster for a given lead time. We expanded the spatial domain for composite maps to include the Alaskan coastline to illustrate precipitation error patterns across the entire northeastern Pacific coast. However, grid cells along the Alaskan coast were not used in the clustering process because of the high uncertainty for observed precipitation in that region (Sun et al. 2018). Errors in each cluster that differed significantly from 0 mm (95% confidence level) were highlighted based on the Wilcoxon signed-rank test (Wilks 2011) corrected for test multiplicity using a false discovery rate (FDR) procedure (Wilks 2016). Precipitation error composites were also recreated for two additional cases to help interpret the results: 1) composites were made based on the clusters identified using the GPCC and control reforecast data, but using NCEP Reanalysis-2 precipitation in place of GPCC precipitation to create the composite maps, and 2) composites were made based on new clusters identified using the GPCC data and the ensemble mean (rather than control) reforecast data. These analyses were used to determine 1) whether precipitation error patterns were driven by local errors in model precipitation or larger-scale forecast errors of synoptic-scale climate drivers and 2) whether patterns of precipitation forecast error differed significant between a single forecast trace and the ensemble mean.

d. Synoptic-scale error patterns

We examined associated features of forecasted and reanalysis IVT and geopotential height to analyze synoptic conditions related to the precipitation error patterns in each cluster and lead time. Again, reforecast and reanalysis data were synchronized at the 0600 UTC time reference to match the GPCC precipitation. Three components of

IVT error were assessed, including errors in magnitude, direction, and landfall location.

IVT magnitude errors were composited by cluster for grid cells across the Pacific–North American sector and were calculated as the difference in standardized anomalies of IVT magnitude between the reforecast and reanalysis data. Statistically significant forecast errors were highlighted using the same procedure as for precipitation. Standardized anomalies were based on localized monthly means and standard deviations for each product across the study period, and were used to help highlight IVT errors in northern latitudes that have lower absolute values due to colder temperatures. Further, anomalies in the reforecast data were calculated separately for each forecast lead time due to biases that are prominent in the ensemble mean IVT data (Fig. S1 in the online supplemental material). For consistency, this method was applied to all reforecast anomaly calculations.

Errors in IVT magnitude and direction near the Sacramento basin were calculated based on differences in reforecast and reanalysis IVT averaged over a geographic area to the west of the basin over the Pacific (red box in Fig. 1). We examined the distribution of both types of error within each cluster. The statistical significance of differences in the mean of these characteristics was tested using a paired t test. Errors in landfall location across all of western North America were calculated based on the difference in latitude between the reforecast and reanalysis projection of maximum IVT magnitude in a set of 41 grid cells along the coast (see black grids in Fig. 1).

Finally, 500-hPa geopotential height anomalies were calculated in the same manner as IVT anomalies. Height anomalies were composited for both the reforecast and reanalysis data for each cluster.

e. Evolution of forecast error

To understand the persistence of forecast error patterns, we analyzed how the spatial pattern of precipitation errors evolved across lead times. We focused on event clusters at a 15-day lead and tracked error composites for these same event clusters at 10-, 5-, and 1-day lead times. For example, suppose N of the 286 events were grouped into Cluster 1 at a 15-day lead. For these same N events, we composited the precipitation error at each of the other lead times. These composites will differ from those for Cluster 1 at other lead times found using the method of section 2c, since they were defined by errors at those leads. Using this procedure, we assessed the degree to which events associated with a particular spatial error pattern at a 15-day lead continued to exhibit that pattern at shorter lead times.

f. Anomalous IVT landfall fraction

In the final analysis, we assessed the fraction of days having a landfalling IVT anomaly over a high threshold (1 standard deviation above the mean) across four spatial domains that span different lengths of the western North American coastline. These domains included Northern California (38°–42°N), California–Oregon–Washington (CA–OR–WA; 33°–49°N), CA–OR–WA and approximately half of the British Columbian coast (33°–55°N), and the entire study area including coastal Alaska (20°–60°N). If any grid cell in the domain had an IVT anomaly over the specified threshold for a given day, that day was considered to have a positive occurrence of anomalously high landfalling IVT. The fraction of landfall occurrences was calculated for each of the four domains using both reforecast data at 1–15-day lead times and reanalysis data. Moreover, we assessed this fraction only for heavy precipitation days in the Sacramento basin, as well as across all days in ONDJFMA in the 1984–2019 timespan excluding the heavy precipitation events. The goal of this analysis was to investigate how different spatial domains could be used to extract useful forecast information about elevated IVT with relevance for heavy precipitation events in the Sacramento basin, while accounting for known spatial biases in AR landfalls and controlling for the climatology of landfalls in the cold season months.

3. Results

a. Regional AR effects

Prior to assessing patterns of forecast error, we illustrate the spatial extent of AR driven precipitation across the region associated with different AR landfall locations. Figure 2 shows composites of standardized anomalies of GPCC precipitation for all landfalling AR events over the cold season (ONDJFMA) between 1984 and 2019 centered at latitudes 32.5°, 35°, 37.5°, and 40°N. AR landfalls are based on the SIO-R1 AR archive developed in Gershunov et al. (2017).

Landfalls outside of these latitudes produce insignificant precipitation in the Sacramento basin, besides minor precipitation in the northwest of the basin for AR landfalls at 42.5°N (not shown). There is significant precipitation within the basin when ARs make landfall between 32.5° and 40°N. Mean precipitation across the basin varies between 9.2 and 13 mm, although these values can be much higher locally in the mountainous regions and can extend over multiple days during persistent events. Landfalls centered at 35°N produce the most substantial precipitation over the Sacramento basin, while those centered at 32.5° and 40°N produce the least.
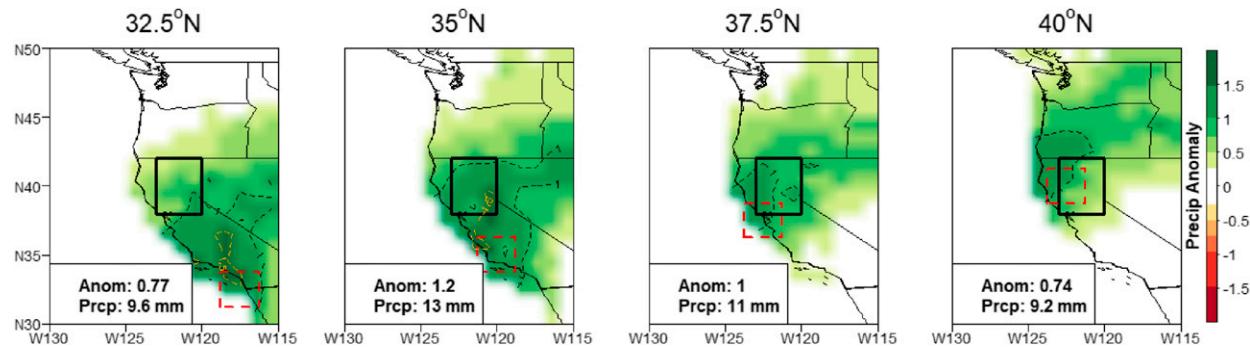
FIG. 2. Precipitation anomalies associated with landfalling ARs between 1984 and 2019 at latitudes of 32.5°, 35°, 37.5°, and 40°N. The solid black box indicates the study region, and the dashed red box is the landfall grid as per the SIO-R1 AR database. The black contour line shows extent of anomaly > 1 standard deviation and orange contour shows anomalies > 1.5 standard deviation. Mean anomaly and total precipitation in the study region are indicated in the bottom left of each plot.

Precipitation tends to be concentrated at or to the north of the latitude of landfall, with smaller precipitation accumulations to the south. The spatial pattern and magnitude of precipitation in Fig. 2 show the relatively strong effect of southwesterly AR events that typically strike south and central California versus the somewhat weaker effect of more zonally oriented ARs that are characteristic of northern California landfalls. This finding aligns well with those in previous studies (Nardi et al. 2018; Hu et al. 2017; Kim et al. 2017).

b. Cluster analysis

Patterns of precipitation forecast error for lead times of 1, 5, 10, and 15 days are shown in Fig. 3. Detailed results for each lead time are discussed sequentially in Figs. 4–7 . Where possible, the clusters are arranged so that spatial patterns of precipitation forecast error are qualitatively similar across lead times (i.e., Cluster 1 at a 1-day lead has a similar pattern of forecast error as Cluster 1 at 5-, 10-, and 15-day leads). However, cluster patterns across lead times do not always align, as each cluster is developed independently for each lead time. These discrepancies are noted below. Still, there are many similarities in error patterns across lead times, suggesting that similar processes govern major patterns of forecast error at medium-range time scales.

Cluster 1 generally has the highest number of events and shows a relatively low degree of precipitation error at a 1-day lead in both magnitude (small underpredictions) and spatial bias (localized error around the Sacramento basin) (Fig. 3). The magnitude and extent of underprediction error is larger at longer lead times.

At short lead times (1 and 5 days), Cluster 2 exhibits overpredictions across much of the Pacific Northwest and southern British Columbia, with underpredictions along the coast. Cluster 3 in the same timeframe exhibits a regional underprediction that extends along

most of the U.S. coastline and is particularly prominent along the Coastal and Sierra Nevada ranges of California, with overpredictions to the north in British Columbia. At long lead times (10 and 15 days), Clusters 2 and 3 both exhibit a north/south dipole in precipitation error, with underpredictions centered around the Sacramento basin and overpredictions to the north. These two clusters differ primarily based on the latitude and extent of the dipole centers.

For lead times of 1, 5, and 10 days, Cluster 4 exhibits a precipitation overprediction centered near the southeastern corner of the basin that extends to the northeast, with an opposing underprediction that extends north along the coast from Northern California to Washington and southern British Columbia. However, at a 15-day lead, Cluster 4 exhibits a spatial pattern unlike the other lead times, with underpredictions across most of California and little systematic error elsewhere.

We also assessed the consistency of the clusters in Fig. 3 after removing systematic forecast bias separately for each lead time. Forecast bias for each grid cell and lead time was calculated in three different ways, including basing the bias on forecasts over all cold season days, all cold season days with greater than 1 mm of precipitation, and only the 286 heavy precipitation days identified in this study. In all cases, the spatial patterns of forecast error based on the debiased data were very similar to those patterns seen in Fig. 3 (see Figs. S2–S4), suggesting that the spatial error patterns are not determined by localized forecast biases.

1) 1-DAY FORECAST LEAD

Unsurprisingly, forecasts of synoptic-scale fields (IVT magnitude, 500-hPa heights) at a 1-day lead time are very accurate (Fig. 4). In the region directly west of the Sacramento basin (red box in Fig. 1), there are no significant errors in IVT magnitude and errors in the
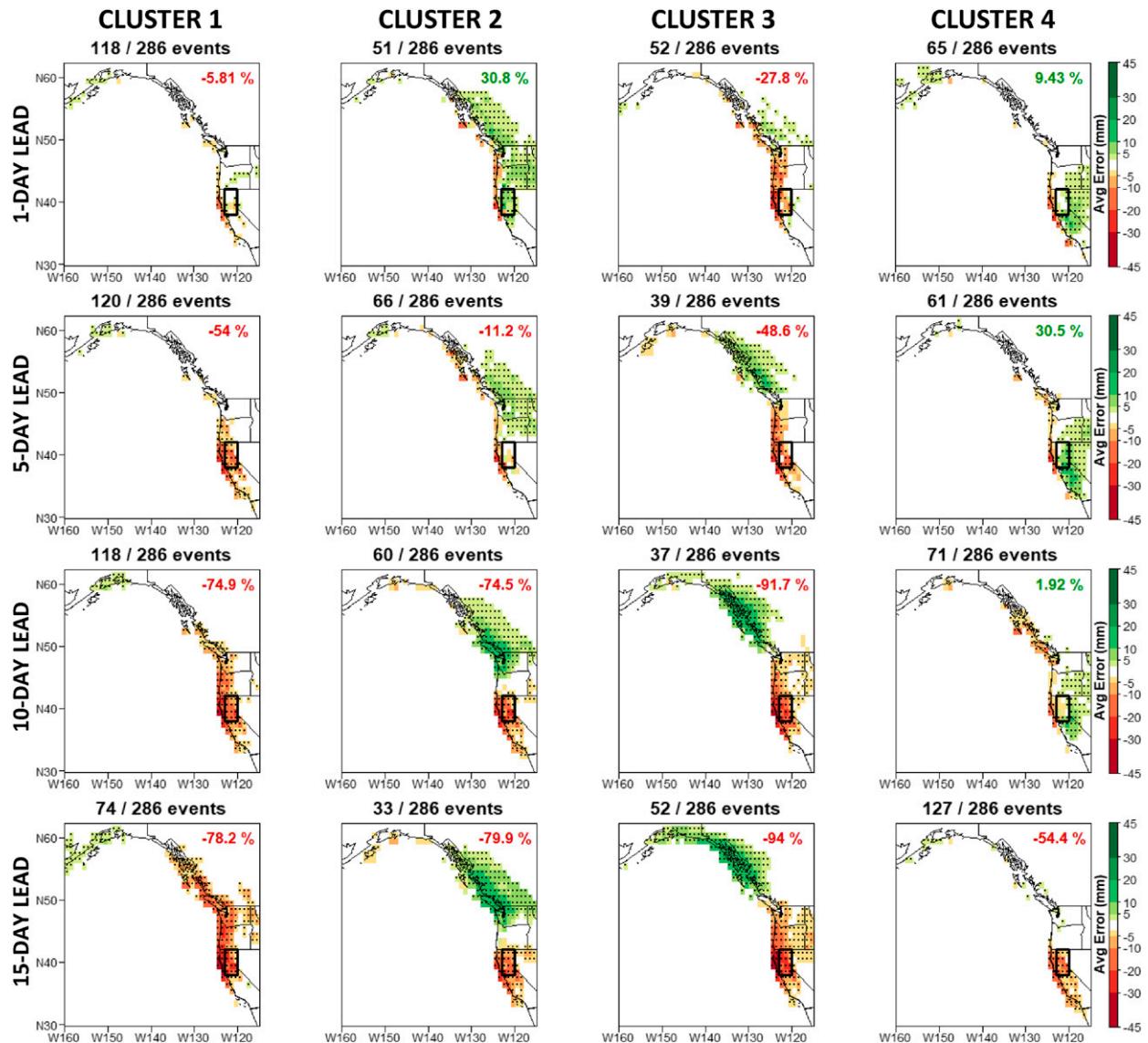
Fig. 3. NCEP GEFS v2 reforecast vs GPCC precipitation error cluster composites at 1-, 5-, 10-, and 15-day lead times with red (green) colors indicating model underpredictions (overpredictions). The number in the upper-right corner of each plot shows the average error [(average forecast − average observed precipitation)/average observed precipitation] across the Sacramento basin, with negative (positive) errors in red (green).

average angle of IVT near the study basin are minimal with the exception of Cluster 4. With respect to precipitation, Cluster 1 contains over 40% of the events and represents the most accurate forecasts across lead times, with relatively small precipitation errors over the basin. However, average precipitation errors in the other clusters are substantial, with as great as 31% average overprediction in Cluster 2 and −28% underprediction in Cluster 3. Average errors over the Sacramento basin in Cluster 4 are smaller (9%), but this cluster also displays a spatially expansive error pattern that extends beyond the basin and penetrates into eastern California and Nevada.

Precipitation errors across all clusters do not appear to be driven by prediction error in large-scale circulation or the overall magnitude of IVT. Rather, precipitation errors are more likely to be attributable to either the inability of the forecast model to 1) accurately resolve subgrid microphysics or orographic enhancement of precipitation (Livneh et al. 2015a; Holton and Hakim 2013), or 2) predict more nuanced attributes of IVT or other localized dynamics (e.g., the Sierra barrier jet). To assess whether subgrid-scale parameterizations are responsible for the large precipitation errors shown in Fig. 4, we reexamine the forecast errors for the same
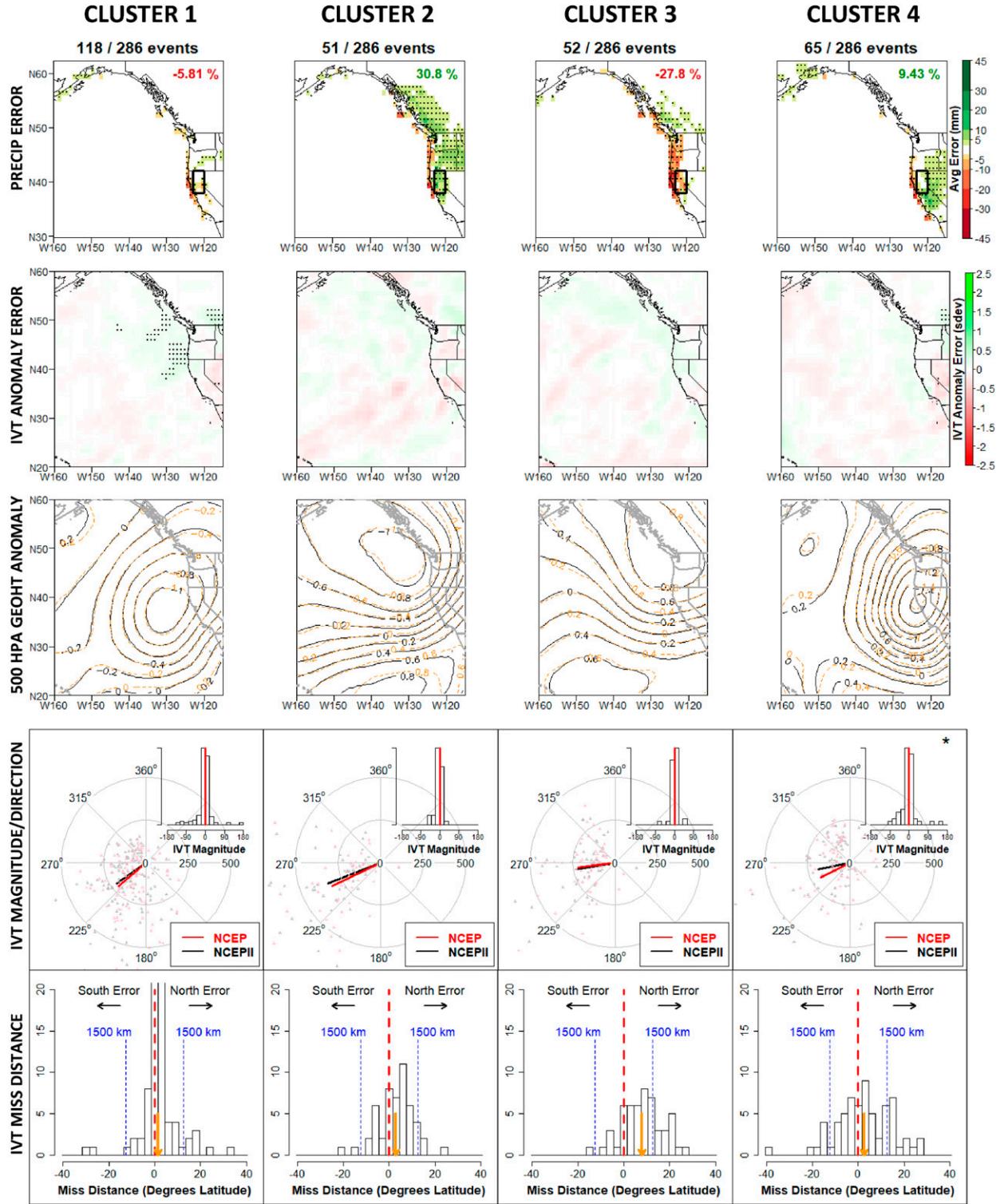
FIG. 4. 1-day forecast lead. (row 1) As in Fig. 3. (row 2) Composite error between forecast and observed IVT anomalies. (row 3) Contour of observed (solid black) vs forecast (dashed orange) 500-hPa geopotential height anomaly composite (units of standard anomaly). (row 4) Landfalling IVT magnitude/direction for individual cluster events forecasted (pink) and observed (gray) with cluster average indicated by solid red (black) for forecasted (observed); histogram in the upper-right corner show errors between forecasted and observed direction, and the red line is zero error. Significant differences ($p < 0.05$) in magnitude, direction, or both are indicated by +, *, and ** in upper-right corner, respectively. (row 5) Histogram of latitude error between forecasted and observed landfalling IVT maxima. The dashed red line indicates zero error, and the orange arrow shows mean miss distance.
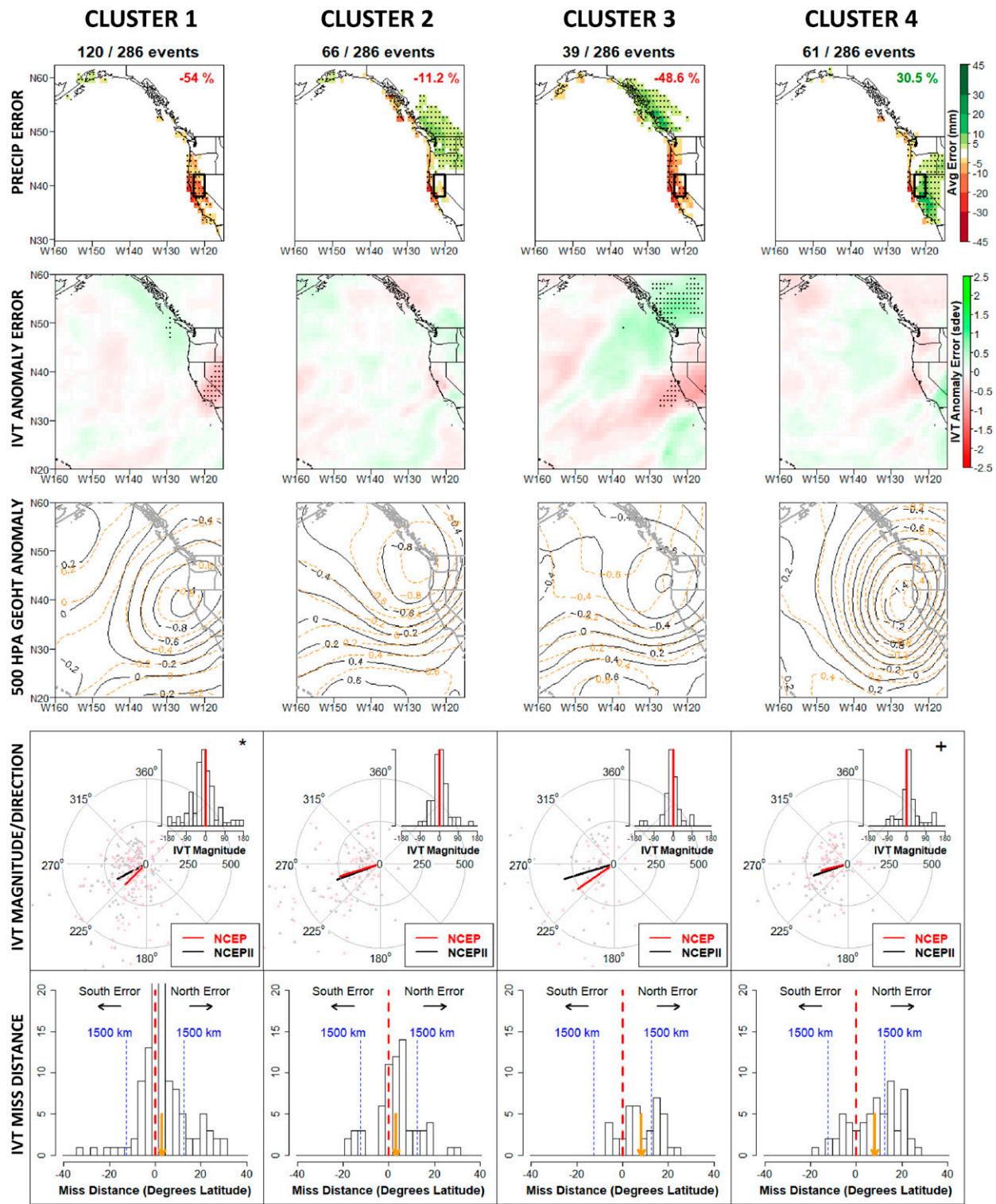
FIG. 5. As in Fig. 4, but for 5-day forecast lead.
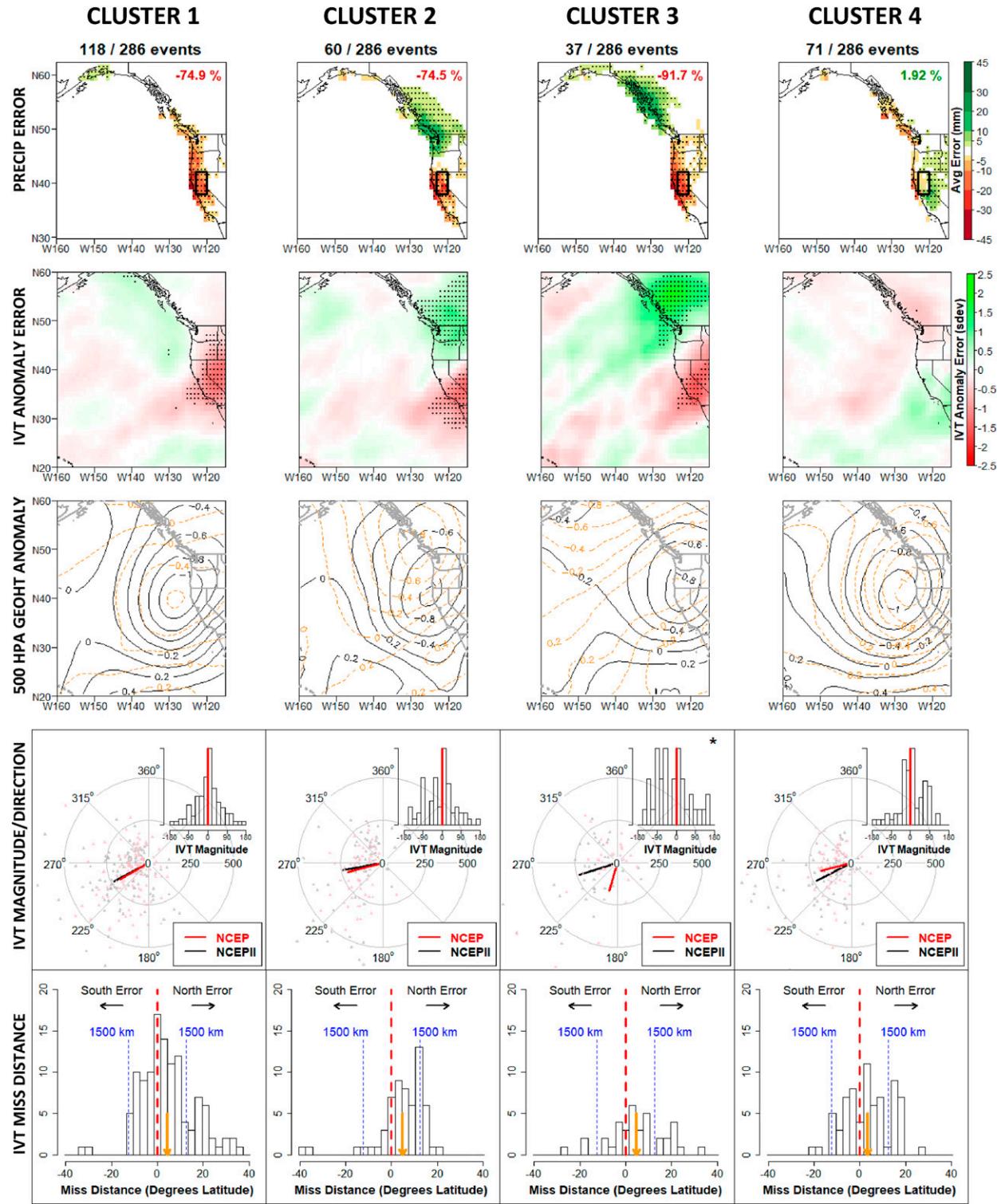
FIG. 6. As in Fig. 4, but for 10-day forecast lead.
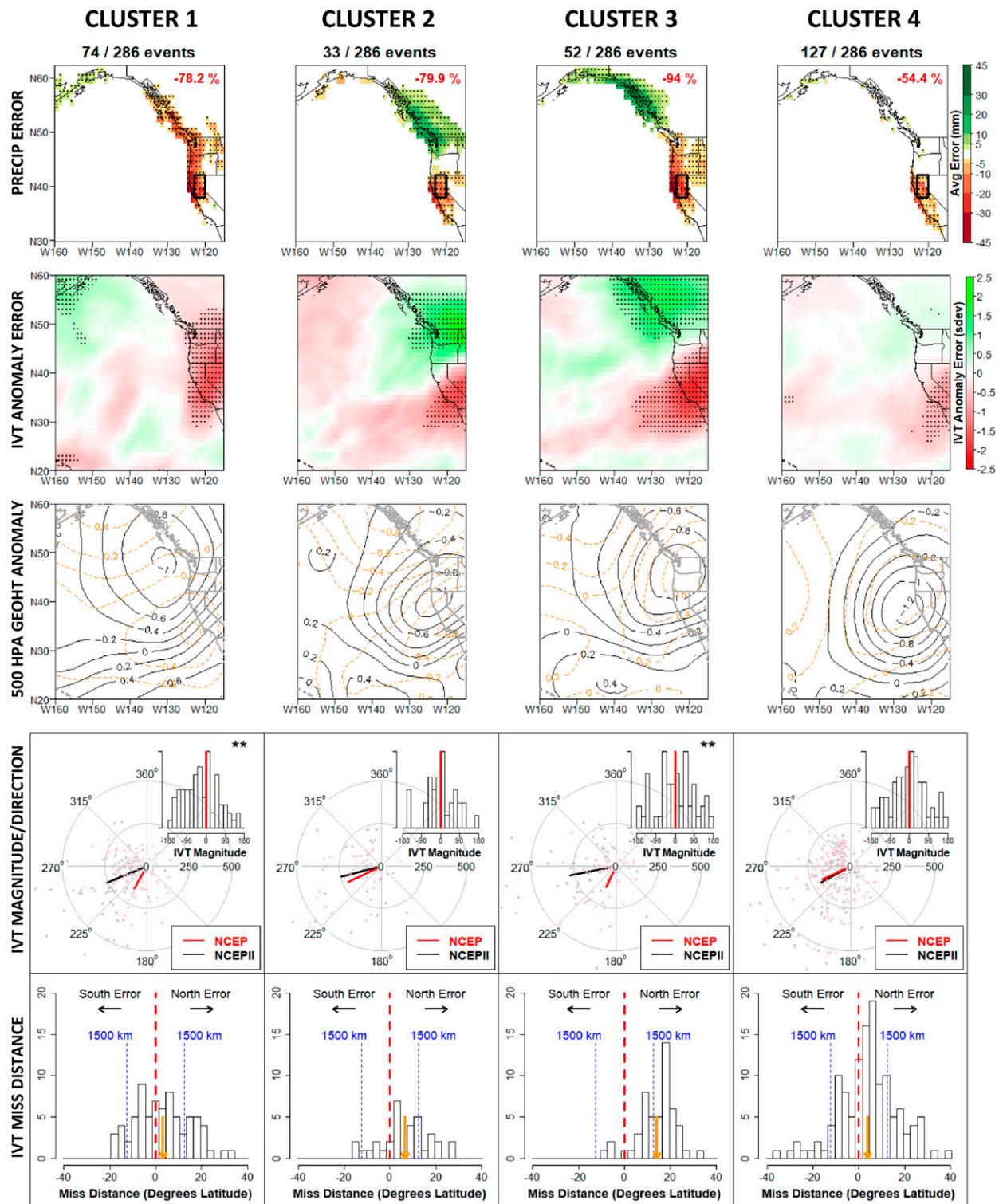
FIG. 7. As in Fig. 4, but for 15-day forecast lead.

events but against NCEP Reanalysis-2 precipitation rather than GPCC precipitation (Fig. S5). By replacing GPCC observations with Reanalysis-2 precipitation, we can better control for errors from subgrid processes, which are similar in the reforecast and reanalysis datasets (Kanamitsu et al. 2002; Hamill et al. 2011, 2013). Interestingly, the patterns of precipitation error in Fig. 4 are very similar to those between the reforecast and reanalysis precipitation, although there are fewer underpredictions along the Pacific Coastal Range when using the reanalysis precipitation. This suggests that subgrid processes are not the driving factor behind some of the larger spatial patterns of precipitation error in Fig. 4, besides perhaps in some coastal areas linked to poorly represented orographic effects.

Instead, most of the precipitation errors appear to be linked to relatively minor differences in the orientation of IVT or other localized dynamics poorly resolved by the model. For instance, Cluster 4 has the deepest 500-hPa low centered on the Northern California coastline, supporting a southwest to northeast flow along the axis of the precipitation overpredictions. However, the more zonal orientation of the observed IVT as compared to the reforecast (Fig. 4, row 4, $p < 0.05$) may explain the overpredicted precipitation along the Sierra Nevada. Clusters 2 and 3 show distinctly different precipitation error while having comparable upper level structure and IVT patterns. The only prominent difference is a stronger northward bias in maximum IVT in Cluster 3 (Fig. 4, row 5). Together, these seemingly small variations produce very different outcomes in basin precipitation. These results suggest that even at a 1-day lead, precipitation error patterns can be very sensitive to local dynamics, even when large-scale attributes of these features are forecasted accurately.

2) 5-DAY FORECAST LEAD

At a 5-day lead, forecasts of height anomalies and IVT remain relatively accurate. The precipitation error clusters and number of events at a 5-day lead are similar to the 1-day lead, although all of the error patterns are deeper, especially Cluster 1. Also, Cluster 3 exhibits a clear north/south dipole error pattern that was not as evident at a 1-day lead.

In Cluster 1, a significant IVT underprediction is apparent over southern California and Nevada with a commensurate underprediction in basin precipitation. The depth of the low is also less than observed, which may explain the IVT underprediction. Though the IVT near the Sacramento basin is more meridional than the observed ($p < 0.05$), it is also underpredicted in magnitude ($p < 0.1$), which may be contributing to the precipitation underprediction.

Cluster 2 exhibits positive precipitation anomalies in the Pacific Northwest. A small northward bias in the midlevel trough could be the cause of these precipitation errors, but forecasted IVT shows no significant differences from observations. In addition, some of the precipitation forecast error may be related to errors in subgrid-scale processes, since the error pattern changes sign near the Sacramento basin and along the Pacific Coastal Range when errors are composited using reanalysis precipitation (Fig. S5).

In contrast, Cluster 3 shows a distinct dipole of precipitation underpredictions centered over the study region and overpredictions centered near the middle of the British Columbian coast, and this pattern does not change with reanalysis precipitation (Fig. S5). This spatial error pattern is supported by the spatial pattern in IVT error and the $>5°$ bias in landfalling IVT location. The underprediction of IVT magnitude west of the Sacramento basin is statistically significant at the $p < 0.1$ level, although not at the $p < 0.05$ level. Forecasts of 500-hPa heights are biased to the northwest and are consistent with the errors in IVT.

Cluster 4 precipitation errors are very similar to those at a 1-day lead. Slightly more organized patterns of IVT error are apparent, particularly along the California coast, and there is a slight eastward bias in the forecasted heights. Overall though, precipitation errors still appear somewhat disconnected from errors in the larger-scale circulation.

3) 10-DAY FORECAST LEAD

At a 10-day lead, precipitation forecast errors are broader and more intense, although the spatial patterns of error are very similar to those for 5-day lead times. In Cluster 1, underpredictions of precipitation intensity around the Sacramento basin coincide with a trough that is forecasted in a similar location to the observations but is weaker in magnitude. Forecasted IVT landfall location also exhibits a high degree of variability, and the IVT magnitude south of the basin is weaker than the observations.

In Cluster 2, a more distinct dipole in precipitation and IVT error is observed than that at a 5-day lead. Similarly, a stronger IVT error dipole is seen for Cluster 3, albeit the dipole is located farther north compared to Cluster 2. A northwest displacement in the forecasted location of the midlevel trough appears to be the primary driver for both dipole patterns, and the degree of this displacement explains the spatial differences in IVT and precipitation error between the two clusters.

Finally, Cluster 4 displays a pattern similar to that observed at a 5-day lead, but with underpredictions expanding north along the coast and farther into the basin.

Also similar to the 5-day pattern, the IVT errors weakly align with the precipitation errors and now appear driven in part by an eastward bias in the forecast low. However, IVT errors still remain insignificant and do not emerge as a clear driver of the spatial pattern in precipitation forecast error.

4) 15-DAY FORECAST LEAD

The error structures in Clusters 1–3 at a 15-day lead are spatially similar but deeper and broader than those at a 10-day lead. For Cluster 1, underpredictions of precipitation extend from California to British Columbia, but there is a clearer overprediction along the coast of Alaska that was only weakly observed at a 5- and 10-day lead. In Clusters 2 and 3, the dipole error patterns are driven by northwest errors in the forecasted location of the trough, which are again similar to the 10-day lead but with even greater displacement. These errors align well with significant biases in IVT magnitude, direction, and the latitude of maximum IVT landfall.

Unlike the other clusters, Cluster 4 at a 15-day lead does not have a consistent error pattern with the events grouped into Cluster 4 at shorter lead times. Rather, this cluster, which accounts for approximately 40% of events (127/286), has a spatial pattern of precipitation error that is rather concentrated around the Sacramento basin. Notably, the forecasted low under this cluster and the IVT south of the basin are relatively weak, while the variance of latitudinal error in IVT landfall is the largest of any cluster at any lead time. These results suggest that Cluster 4 at a 15-day lead may be averaging over a wide range of events that are not distinctly similar in their error patterns, with the effect of smoothing and dampening signals in the IVT and height fields. Since all events were defined such that the Sacramento basin was experiencing heavy precipitation, the only signal that is not canceled out after averaging is an underprediction in IVT and precipitation around the basin.

5) ENSEMBLE MEAN VERSUS CONTROL FORECAST

A similar analysis to that in Figs. 3–7 and Fig. S5 was conducted for the ensemble mean forecast and is shown in Figs. S6–S11. The results between the ensemble mean and the control are very similar, although features of the IVT and height field composites are more finely resolved in the control forecast results because of smoothing effects under the ensemble mean. In addition, the ensemble mean exhibits a somewhat different pattern for Cluster 4, particularly at the 10-day lead time, where the ensemble mean shows an underprediction for the basin while the control shows a split underprediction/overprediction pattern. Still, these results suggest that many of the basic forecast error patterns found for one member of an ensemble can reflect the patterns of error found in the ensemble mean.

c. Evolution of forecast error

Figure 8 shows the persistence of spatial error patterns across lead times when events remain in the same cluster to which they were assigned at a 15-day lead. From a 15- to 10-day lead, the strong underprediction along the U.S. West Coast through British Columbia remains coherent in Cluster 1. A similar result is seen for Cluster 4. For Clusters 2 and 3, the dipole structures in precipitation error are still apparent at a 10-day lead but are weaker. At a 5-day lead, the spatial patterns in Clusters 1, 2, and 4 defined at the 15-day lead are still visible but are extremely weak, and the spatial error pattern in Cluster 3 is no longer distinguishable. At a 1-day lead, there is no relation between the error composites defined at the 15-day lead, and basin-average precipitation error falls dramatically (−1.5%, −0.9%, 3.3%, 0.3%). These results suggest that forecast errors at a 15-day lead time may persist to a 10-day lead to some degree, but not to shorter lead times.

d. Anomalous IVT landfall occurrence frequency

Figure 9 depicts the fraction of days exhibiting anomalously high landfalling IVT across four spatial domains, where any occurrence of IVT over a standard deviation of 1.0 across the landfalling grid cells in that domain yields a positive signal. These fractions are shown for the reanalysis and reforecasts (both control and ensemble mean) at 1–15-day lead times, and are calculated separately for heavy precipitation days and nonheavy precipitation days in the Sacramento basin. For a particular lead time and forecast product (control or ensemble mean), a high fraction of events for the heavy precipitation days and a low fraction of events for nonheavy precipitation days (i.e., climatology) suggests that the forecasts are able to detect high IVT activity somewhere in the domain with relevance to heavy precipitation events in the Sacramento basin, while excluding high IVT events that are not associated with heavy precipitation in the basin. That is, large differences in the fractions for heavy and nonheavy precipitation days suggest a larger signal to noise ratio in the forecasted IVT. The results for the reanalysis provide the observed fractions, and so can be used as target values against which to compare the reforecast results. We note that the results in Fig. 9 are similar if we change the threshold for anomalously high IVT (see Fig. S12).

Several insights emerge from Fig. 9. First, for all domains and for both the control (Fig. 9, row 2) and the ensemble mean (Fig. 9, row 3), the forecast signal degrades with lead time. That is, a lower fraction of high IVT events over the
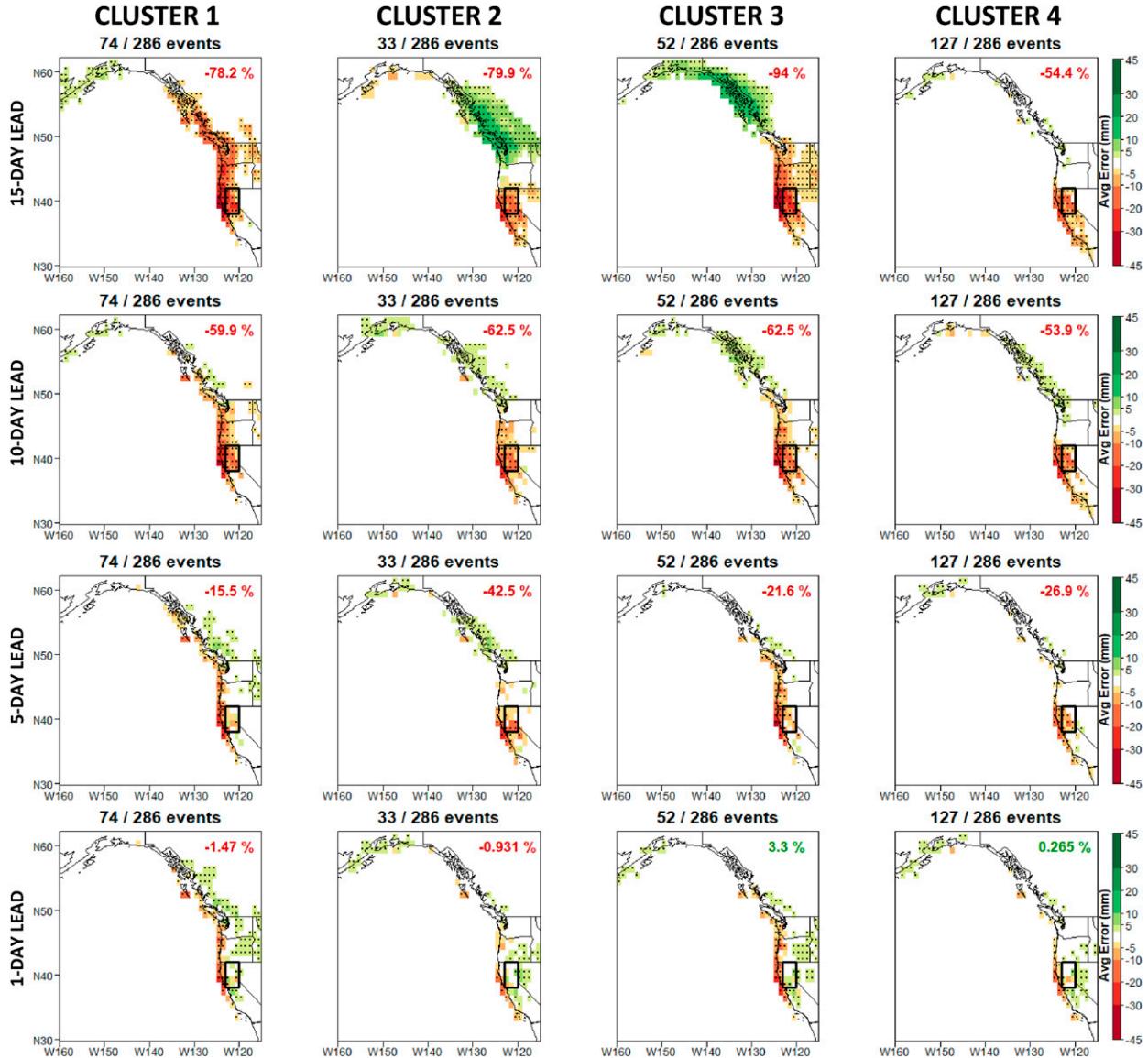
FIG. 8. NCEP GEFS v2 reforecast vs GPCC precipitation error cluster composites at 15-, 10-, 5-, and 1-day lead times based on clusters at a 15-day lead.

climatological rate is detected at longer lead times. Second, the differences between the control and the ensemble mean reforecast are substantial. The control forecast produces a much weaker signal during the heavy precipitation events versus nonheavy precipitation events across all spatial domains. Even in the case of the strongest signal to noise ratio associated with the CA–OR–WA domain and a 1-day lead, there is only a 20% gain in signal above the climatological rate of high IVT events (around 32%) for the control forecast. The ensemble mean detects a much higher fraction of high IVT events (78%) for the same domain and lead time, which is close to the reanalysis rate. In general,

the ensemble mean forecast maintains a higher signal to noise ratio for longer leads, regardless of domain.

Finally, the signal to noise ratio depends strongly on the domain over which high IVT events are tracked, particularly for the ensemble mean. If the domain is too small (e.g., Northern California only), forecasted events of high IVT that impact the basin are often missed. If the domain is too large (e.g., the entire North American west coast), most of the high IVT events associated with heavy precipitation in the Sacramento basin are captured at all lead times, but so are a large majority of high IVT events that are not associated with heavy precipitation in the basin. That is, there are a very large number
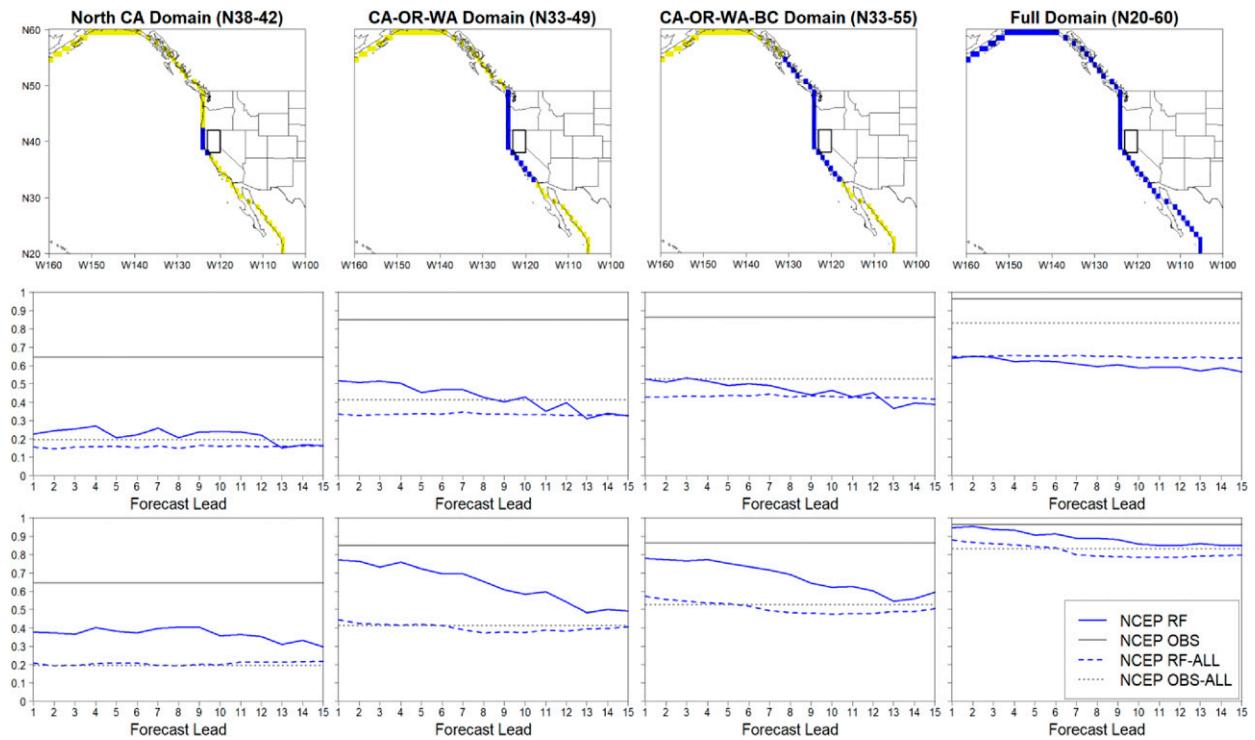
FIG. 9. Anomalous IVT landfall fraction plots based on NCEP GEFS v2 reforecast and NCEP Reanalysis-2. (row 1) Visual depiction of landfall domain; yellow grid cells depict all landfall grids with blue highlighted grids and dashed red box showing the area of consideration. (rows 2–3) Blue (gray) lines show fraction of days that IVT was over 1 standard deviation in the indicated domain for reforecast (reanalysis) data. The solid lines depict the fraction occurring during the heavy precipitation days while the dashed lines depict the fraction occurring across all days excluding the heavy precipitation days in the ONDJFMA periods from 1984 to 2019. Row 2 is based on the *control* reforecast whereas row 3 is based on the *ensemble mean* reforecast.

of false positives if the domain is too large. Signal to noise appears largest when the domain spans the U.S. West Coast (Fig. 9, column 2). Here, for leads between 1 and 10 days, there is a significant increase in the rate of high IVT events for heavy precipitation days compared to the smaller Northern California domain, but a more modest rise in the rate of false positives associated with nonheavy precipitation days. This result is very consistent with the patterns of forecast bias for lead times of 5 and 10 days (Figs. 5 and 6, Cluster 2), which showed considerable latitudinal error in AR landfall location that could be captured if high IVT was tracked in the Pacific Northwest.

4. Discussion and conclusions

In this study we identified recurrent spatial biases in medium-range forecasts of precipitation across western North America during heavy precipitation events in the Sacramento River basin. At 1-day leads and to a lesser extent at 5-day leads, many of the forecasts are quite accurate, and those that exhibit substantial error are more closely related to the inability to resolve localized

dynamics and subgrid processes. However, at longer leads, a set of clear and spatially distinct error patterns are evident that are stronger at longer lead times and closely mirror spatial patterns of IVT forecast error, which are in turn consistent with spatial biases in forecasted features of large-scale circulation. In particular, there is a tendency for the models to forecast AR landfall too far to the north, leading to overpredicted precipitation and IVT in the Pacific Northwest, British Columbia, and even Alaska. The implication is that forecasts of significant precipitation or IVT north of the Sacramento basin at long leads may be associated with elevated risk of heavy precipitation within the basin. Where large-scale error patterns fail to explain patterns of forecasted precipitation error, mesoscale errors in landfalling IVT (e.g., IVT orientation) can help to reconcile these differences.

The patterns of forecast error suggest that large-scale forecasts of circulation or features of IVT could be used to improve forecast-informed water resources management. While patterns of forecast error do not tend to persist from longer to shorter lead times (Fig. 8), similar spatial patterns of error are observed independently at

lead times between 5 and 15 days. This suggests the possibility of creating probabilistic estimates of risk for heavy precipitation events in the Sacramento basin conditional on a forecast at a particular lead. As shown in Fig. 9, the spatial domain over which forecast information is analyzed could be used tailor these probabilistic estimates. If the domain is too small then many events are missed, whereas if the domain is too large, the signal is saturated by climatological noise. For the Sacramento basin, an expanded domain that extends across the U.S. coastline appears to provide a good balance of hits and false positives, although this choice will strongly depend on the risk tolerance and preferences of water managers. Moreover, the enhanced signal apparent in the ensemble mean versus the control further supports the use of multimember ensembles to provide better forecast information. While this study was limited to a single model, future work will explore how multimodel ensembles can be used to further improve the selection of forecast information for forecast-informed reservoir operations, for instance by using emerging input variable selection procedures (Giuliani et al. 2015; Denaro et al. 2017; Herman and Giuliani 2018).

Importantly, this study did not directly address forecast error in other fields relevant to streamflow, which is the primary hydrologic driver of surface water systems in California. For instance, streamflow can be strongly influenced by storm duration and anomalously high temperatures and snowmelt, both factors which are strongly impacted by ARs but were not considered here (Guan et al. 2016; Lamjiri et al. 2017; Gonzales et al. 2019). Future work will consider how forecast errors in these fields also relate to patterns of forecast error in synoptic-scale features of associated ARs.

Finally, it is worthwhile to note that although some of the spatial signatures of forecast error seen at 5- and 10-day lead times are still present at a 15-day lead, the most common type of forecast error at a 15-day exhibits significant variability in large-scale circulation and the location of AR landfalls (see Fig. 7, Cluster 4). Therefore, while 15-day lead forecasts may provide some information for management, the utility of this information is likely limited. However, our results suggest that forecast information at large spatial scales and lead times of less than 10 days has real potential to inform water management decisions, and this may prove to be a sufficient lead time for key actions like reservoir release adjustments needed for flood risk reduction.

Data availability statement: The data used in this manuscript are in the following repository: HydroShare, http://www.hydroshare.org/resource/8780d5b2648c4c39acabc37bd6774cf8, or are available on request from the corresponding author.

REFERENCES

Alley, B. R. B., K. A. Emanuel, and F. Zhang, 2019: Advances in weather prediction. *Science*, **363**, 342–344, https://doi.org/10.1126/science.aav7274.

American Meteorological Society, 2020: Atmospheric river. Glossary of Meteorology, http://glossary.ametsoc.org/wiki/Atmospheric_river.

Baggett, C. F., E. A. Barnes, E. D. Maloney, and B. D. Mundhenk, 2017: Advancing atmospheric river forecasts into subseasonal-to-seasonal time scales. *Geophys. Res. Lett.*, **44**, 7528–7536, https://doi.org/10.1002/2017GL074434.

Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525**, 47–55, https://doi.org/10.1038/nature14956.

Brands, S., J. M. Gutiérrez, and D. San-Martín, 2017: Twentieth-century atmospheric river activity along the west coasts of Europe and North America: Algorithm formulation, reanalysis uncertainty and links to atmospheric circulation patterns. *Climate Dyn.*, **48**, 2771–2795, https://doi.org/10.1007/s00382-016-3095-6.

Dacre, H. F., P. A. Clark, O. Martinez-Alvarado, M. A. Stringer, and D. A. Lavers, 2015: How do atmospheric rivers form? *Bull. Amer. Meteor. Soc.*, **96**, 1243–1255, https://doi.org/10.1175/BAMS-D-14-00031.1.

DeFlorio, M. J., D. E. Waliser, B. Guan, D. A. Lavers, F. M. Ralph, and F. Vitart, 2018a: Global assessment of atmospheric river prediction skill. *J. Hydrometeor.*, **19**, 409–426, https://doi.org/10.1175/JHM-D-17-0135.1.

——, ——, ——, F. M. Ralph, and F. Vitart, 2018b: Global evaluation of atmospheric river subseasonal prediction skill. *Climate Dyn.*, **52**, 3039–3060, https://doi.org/10.1007/s00382-018-4309-x.

Denaro, S., D. Anghileri, M. Giuliani, and A. Castelletti, 2017: Informing the operations of water reservoirs over multiple temporal scales by direct use of hydro-meteorological data. *Adv. Water Resour.*, **103**, 51–63, https://doi.org/10.1016/j.advwatres.2017.02.012.

Dettinger, M. D., F. M. Ralph, T. Das, P. J. Neiman, and D. R. Cayan, 2011: Atmospheric rivers, floods and the water resources of California. *Water*, **3**, 445–478, https://doi.org/10.3390/w3020445.

——, ——, and J. J. Rutz, 2018: Empirical return periods of the most intense vapor transports during historical atmospheric river landfalls on the U.S. West Coast. *J. Hydrometeor.*, **19**, 1363–1377, https://doi.org/10.1175/JHM-D-17-0247.1.

Forecast Informed Reservoir Operations Steering Committee, 2017: Preliminary viability assessment of Lake Mendocino forecast informed reservoir operations. Scripps Institute Center for Western Weather and Water Extremes, 75 pp., https://cw3e.ucsd.edu/FIRO_docs/FIRO_PVA.pdf.

Gershunov, A., T. Shulgina, F. M. Ralph, D. A. Lavers, and J. J. Rutz, 2017: Assessing the climate-scale variability of atmospheric rivers affecting western North America. *Geophys. Res. Lett.*, **44**, 7900–7908, https://doi.org/10.1002/2017GL074175.

Giuliani, M., F. Pianosi, and A. Castelletti, 2015: Making the most of data: An information selection and assessment framework to improve water systems operations. *Water Resour. Res.*, **51**, 9073–9093, https://doi.org/10.1002/2015WR017044.

Gonzales, K. R., D. L. Swain, K. M. Nardi, E. A. Barnes, and N. S. Diffenbaugh, 2019: Recent warming of landfalling atmospheric rivers along the west coast of the United States. *J. Geophys. Res. Atmos.*, **124**, 6810–6826, https://doi.org/10.1029/2018JD029860.

Guan, B., D. E. Waliser, F. M. Ralph, E. J. Fetzer, and P. J. Neiman, 2016: Hydrometeorological characteristics of rain-on-snow events associated with atmospheric rivers. *Geophys. Res. Lett.*, **43**, 2964–2973, https://doi.org/10.1002/2016GL067978.

Guirguis, K., A. Gershunov, R. E. S. Clemesha, T. Shulgina, A. C. Subramanian, and F. M. Ralph, 2018: Circulation drivers of atmospheric rivers at the North American west coast. *Geophys. Res. Lett.*, **45**, 576–584, https://doi.org/10.1029/2018GL079249.

——, ——, T. Shulgina, R. E. S. Clemesha, and F. M. Ralph, 2019: Atmospheric rivers impacting Northern California and their modulation by a variable climate. *Climate Dyn.*, **52**, 6569–6583, https://doi.org/10.1007/s00382-018-4532-5.

Hamill, T. M., J. S. Whitaker, M. Fiorino, and S. G. Benjamin, 2011: Global ensemble predictions of 2009's tropical cyclones initialized with an ensemble Kalman filter. *Mon. Wea. Rev.*, **139**, 668–688, https://doi.org/10.1175/2010MWR3456.1.

——, G. T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau, Y. Zhu, and W. Lapenta, 2013: NOAA's second-generation global medium-range ensemble reforecast dataset. *Bull. Amer. Meteor. Soc.*, **94**, 1553–1565, https://doi.org/10.1175/BAMS-D-12-00014.1.

Hannachi, A., I. T. Jolliffe, and D. B. Stephenson, 2007: Empirical orthogonal functions and related techniques in atmospheric science: A review. *Int. J. Climatol.*, **27**, 1119–1152, https://doi.org/10.1002/joc.1499.

Hecht, C. W., and J. M. Cordeira, 2017: Characterizing the influence of atmospheric river orientation and intensity on precipitation distributions over North Coastal California. *Geophys. Res. Lett.*, **44**, 9048–9058, https://doi.org/10.1002/2017GL074179.

Herman, J. D., and M. Giuliani, 2018: Policy tree optimization for threshold-based water resources management over multiple timescales. *Environ. Modell. Software*, **99**, 39–51, https://doi.org/10.1016/j.envsoft.2017.09.016.

Hoerling, M., J. Eischeid, J. Perlwitz, X. W. Quan, K. Wolter, and L. Cheng, 2016: Characterizing recent trends in U.S. heavy precipitation. *J. Climate*, **29**, 2313–2332, https://doi.org/10.1175/JCLI-D-15-0441.1.

Holton, J. R., and G. J. Hakim, 2013: *An Introduction to Dynamic Meteorology*. 5th ed. Elsevier, 532 pp.

Hu, H., F. Dominguez, Z. Wang, D. A. Laversa, G. Zhang, and F. M. Ralph, 2017: Linking atmospheric river hydrological impacts on the U.S. West Coast to Rossby wave breaking. *J. Climate*, **30**, 3381–3399, https://doi.org/10.1175/JCLI-D-16-0386.1.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP–DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, **83**, 1631–1643, https://doi.org/10.1175/BAMS-83-11-1631.

Kim, H. M., Y. Zhou, and M. A. Alexander, 2017: Changes in atmospheric rivers and moisture transport over the Northeast Pacific and western North America in response to ENSO diversity. *Climate Dyn.*, **52**, 7375–7388, https://doi.org/10.1007/s00382-017-3598-9.

Konrad, C. P., and M. D. Dettinger, 2017: Flood runoff in relation to water vapor transport by atmospheric rivers over the western United States, 1949–2015. *Geophys. Res. Lett.*, **44**, 11 456–11 462, https://doi.org/10.1002/2017GL075399.

Lamjiri, M. A., M. D. Dettinger, F. M. Ralph, and B. Guan, 2017: Hourly storm characteristics along the U.S. West Coast: Role of atmospheric rivers in extreme precipitation. *Geophys. Res. Lett.*, **44**, 7020–7028, https://doi.org/10.1002/2017GL074193.

Lavers, D. A., D. E. Waliser, F. M. Ralph, and M. D. Dettinger, 2016: Predictability of horizontal water vapor transport relative to precipitation: Enhancing situational awareness for forecasting western U.S. extreme precipitation and flooding. *Geophys. Res. Lett.*, **43**, 2275–2282, https://doi.org/10.1002/2016GL067765.

——, E. Zsoter, D. S. Richardson, and F. Pappenberger, 2017: An assessment of the ECMWF Extreme Forecast Index for water vapor transport during boreal winter. *Wea. Forecasting*, **32**, 1667–1674, https://doi.org/10.1175/WAF-D-17-0073.1.

Livneh, B., T. J. Bohn, D. W. Pierce, F. Muñoz-Arriola, B. Nijssen, R. Vose, D. R. Cayan, and L. Brekke, 2015a: A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950-2013. *Sci. Data*, **2**, 150042, https://doi.org/10.1038/sdata.2015.42.

——, ——, ——, ——, ——, ——, ——, and ——, 2015b: A spatially comprehensive, meteorological data set for Mexico, the U.S., and southern Canada (NCEI Accession 0129374). NOAA/NCEI, accessed 20 August 2019, https://doi.org/10.7289/v5x34vf6.

Loucks, D. P., and E. Van-Beek, 2017: *Water Resources Systems Planning and Management: An Introduction to Methods, Models, and Applications*. Springer, 624 pp.

Mundhenk, B. D., E. A. Barnes, and E. D. Maloney, 2016: All-season climatology and variability of atmospheric river frequencies over the North Pacific. *J. Climate*, **29**, 4885–4903, https://doi.org/10.1175/JCLI-D-15-0655.1.

——, ——, ——, and C. F. Baggett, 2018: Skillful empirical subseasonal prediction of landfalling atmospheric river activity using the Madden–Julian oscillation and quasi-biennial oscillation. *NPJ Climate Atmos. Sci.*, **1**, 20177, https://doi.org/10.1038/s41612-017-0008-2.

Nardi, K. M., E. A. Barnes, and F. M. Ralph, 2018: Assessment of numerical weather prediction model reforecasts of the occurrence, intensity, and location of atmospheric rivers along the west coast of North America. *Mon. Wea. Rev.*, **146**, 3343–3362, https://doi.org/10.1175/MWR-D-18-0060.1.

Nayak, M. A., J. D. Herman, and S. Steinschneider, 2018: Balancing flood risk and water supply in California: Policy search integrating short-term forecast ensembles with conjunctive use. *Water Resour. Res.*, **54**, 7557–7576, https://doi.org/10.1029/2018WR023177.

NOAA/NCEP, 2002: NCEP-DOE AMIP-II Reanalysis 2 (updated daily). NOAA/OAR/ESRL/PSL, accessed 20 August 2019, https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.html.

——, 2013: NCEP Global Ensemble Forecasting System (GEFS, version 10, updated daily). NOAA/OAR/ESRL/PSL, accessed 22 August 2019, https://psl.noaa.gov/forecasts/reforecast2/download.html.

Ralph, F. M., P. J. Neiman, G. N. Kiladis, K. Weickmann, and D. W. Reynolds, 2011: A multiscale observational case study of a Pacific atmospheric river exhibiting tropical–extratropical connections and a mesoscale frontal wave. *Mon. Wea. Rev.*, **139**, 1169–1189, https://doi.org/10.1175/2010MWR3596.1.

——, and Coauthors, 2014: A vision for future observations for western U.S. extreme precipitation and flooding. *J. Contemp.*

Water Res. Educ., **153**, 16–32, https://doi.org/10.1111/j.1936-704X.2014.03176.x.

——, J. M. Cordeira, P. J. Neiman, and M. Hughes, 2016: Landfalling atmospheric rivers, the Sierra Barrier Jet, and extreme daily precipitation in northern California's upper Sacramento River watershed. *J. Hydrometeor.*, **17**, 1905–1914, https://doi.org/10.1175/JHM-D-15-0167.1.

——, J. J. Rutz, J. M. Cordeira, M. Dettinger, M. Anderson, D. Reynolds, L. J. Schick, and C. Smallcomb, 2019: A scale to characterize the strength and impacts of atmospheric rivers. *Bull. Amer. Meteor. Soc.*, **100**, 269–289, https://doi.org/10.1175/BAMS-D-18-0023.1.

Rutz, J. J., W. J. Steenburgh, and F. M. Ralph, 2015: The inland penetration of atmospheric rivers over Western North America: A Lagrangian analysis. *Mon. Wea. Rev.*, **143**, 1924–1944, https://doi.org/10.1175/MWR-D-14-00288.1.

Schamm, K., Ziese, M., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., and U. Schneider, 2013: GPCC First Guess Daily Product at 1.0°: Near real-time first guess daily land-surface precipitation from rain-gauges based on SYNOP data. Global Precipitation Climatology Centre, accessed 20 August 2019, https://doi.org/10.5676/DWD_GPCC/FG_D_100.

——, M. Ziese, A. Becker, P. Finger, A. Meyer-Christoffer, U. Schneider, M. Schroder, and P. Stender, 2014: Global gridded precipitation over land: A description of the new GPCC First Guess Daily product. *Earth Syst. Sci. Data*, **6**, 49–60, https://doi.org/10.5194/essd-6-49-2014.

——, ——, K. Raykova, A. Becker, P. Finger, A. Meyer-Christoffer, and U. Schneider, 2015: GPCC Full Data Daily Version 1.0 at 1.0°: Daily land-surface precipitation from rain-gauges built on GTS-based and historic data. Global Precipitation Climatology Centre, accessed 20 August 2019, https://doi.org/10.5676/DWD_GPCC/FD_D_V1_100.

Schneider, U., M. Ziese, A. Meyer-Christoffer, P. Finger, E. Rustemeier, and A. Becker, 2016: The new portfolio of global precipitation data products of the Global Precipitation Climatology Centre suitable to assess and quantify the global water cycle and resources. *Proc. Int. Assoc. Hydrol. Sci.*, **374**, 29–34, https://doi.org/10.5194/piahs-374-29-2016.

——, A. Becker, M. Ziese, and B. Rudolf, 2018: Global precipitation analysis products of the GPCC. Global Precipitation Climatology Centre, 15 pp., https://opendata.dwd.de/climate_environment/GPCC/PDF/GPCC_intro_products_v2018.pdf.

Steinschneider, S., M. Ho, E. R. Cook, and U. Lall, 2016: Can PDSI inform extreme precipitation?: An exploration with a 500 year long paleoclimate reconstruction over the U.S. *Water Resour. Res.*, **52**, 3866–3880, https://doi.org/10.1002/2016WR018712.

Sun, Q., C. Miao, Q. Duan, H. Ashouri, S. Sorooshian, and K. L. Hsu, 2018: A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Rev. Geophys.*, **56**, 79–107, https://doi.org/10.1002/2017RG000574.

USGS, 2013: Watershed Boundaries, HUC 10, for California. USGS, accessed 28 August 2019, https://www.sciencebase.gov/catalog/item/5696a727e4b039675d00a4ef.

Waliser, D., and B. Guan, 2017: Extreme winds and precipitation during landfall of atmospheric rivers. *Nat. Geosci.*, **10**, 179–183, https://doi.org/10.1038/ngeo2894.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International Geophysics Series, Vol. 100, Elsevier, 676 pp.

——, 2016: ''The stippling shows statistically significant grid points'': How research results are routinely overstated and overinterpreted, and what to do about it. *Bull. Amer. Meteor. Soc.*, **97**, 2263–2273, https://doi.org/10.1175/BAMS-D-15-00267.1.

Zhu, Y., and R. E. Newell, 1998: A proposed algorithm for moisture fluxes from atmospheric rivers. *Mon. Wea. Rev.*, **126**, 725–735, https://doi.org/10.1175/1520-0493(1998)126<0725:APAFMF>2.0.CO;2.