Utility of medium range atmospheric river forecasting for water management in northern California

Zachary P. Brodeur1, Scott Steinschneider2

Department of Biological and Environmental Engineering, Cornell University, Ithaca, NY

1. Graduate Research Assistant, 111 Wing Drive, Riley-Robb Hall, Department of Biological and Environmental Engineering, Cornell University, Ithaca, NY, 14853. Email: zpb4@cornell.edu, Phone: 607-255-2155 (Corresponding Author).

2. Assistant Professor, 111 Wing Drive, Riley-Robb Hall, Department of Biological and Environmental Engineering, Cornell University, Ithaca, NY, 14853. Email: ss3378@cornell.edu, Phone: 607-255-2155.

**1. Introduction**

Water resource management in northern California has been and will continue to be a challenge for those engaged in the analysis, formulation, and implementation of effective policy. The extreme historical intra and inter-annual variation of precipitation coupled with the effects of anthropogenic climate change and inevitable socio-economic fluxes in demand promise increased strain on an already critically allocated system. Further, the massive transfers of water from the northern reaches of the state to its southern population centers remain a fixture in the water management architecture of the region not likely to be changed soon. Recent research has made significant strides in characterizing the effect of ‘atmospheric river’ events on the precipitation regime of this region. In just the past two years, researchers have finally decided on a formal definition of ARs as well as a categorization scheme similar to that used for hurricanes (Ralph et al. 2018a; Ralph et al. 2018b). For this region in particular, ARs provide up to 50% of annual precipitation critical for water management considerations and most of it comes in the form of ‘extreme precipitation’ event as defined by (Hecht and Cordeira 2017). Necessarily, a variety of negative effects result from these events as well, including flooding, high winds, and mudslides. Though the categorization of ARs is a significant leap in regards to formulating policies to address such effects, much remains to be discovered regarding accurate characterization of precipitation from ARs and its resultant implications for water resource management. Further, advances on the forecasting front will only be useful if leveraged by concomitant improvements in the control schema used to store and allocate water effectively. To this end, a strategy that has recently garnered heightened interest is the use of medium range (1-14 day) and sub-seasonal to seasonal (S2S, 15-90 day) forecasts of extreme precipitation events to inform reservoir control policy decisions.

In general, the use of extended forecasts to inform operations has been limited by the relatively slow increase in forecast skill at long leads (approximately one day in forecast skill per decade) (Bauer and Brunet 2015). However, in specific regions where extreme events are governed by a limited number of synoptic-scale atmospheric mechanisms, measures of those mechanisms can yield gains in forecast skill beyond the ‘average’ value that exists globally. Along the West Coast, and particularly in northern California, the coincidence of extreme precipitation events and ARs suggests these type of gains may be possible. Because ARs are defined by prominent synoptic scale structures of integrated vapor transport (IVT), there is a highly observable signature associated with their formation and promulgation. Since these atmospheric features are also explicitly resolved in forecast models, albeit at a fairly coarse level, it is logical to surmise that they could be more accurately predicted at longer lead times than unresolved features like local extreme precipitation that have to be parameterized at the grid scale (Bauer and Brunet 2015). Further, numerous studies have shown significant correlations between measurable synoptic features of an AR, such as IVT magnitude, direction, and duration and their resultant precipitation effects(Mundhenk et al. 2016; Dacre et al. 2015; Nardi et al. 2018; Brands et al. 2017; DeFlorio et al. 2018). These effects manifest in different ways depending on the local topography and positioning in the AR occurrence regions. Northern California has been especially well studied in relation to these teleconnections with numerous effects such as the Sierra Barrier Jet (SBJ) identified as crucial amplifiers of the bulk effects of a landfalling AR (Nardi et al. 2018; Hecht and Cordeira 2017; Ralph et al. 2016). Thus, improvements in AR predictions, including their occurrence, landfall location, and characteristics at landfall, hold great promise for the management of hydrologic extremes in the region. While the use of large-scale atmospheric predictors like IVT fields are not likely to produce accurate forecasts of extreme event occurrences at specific locations at long lead times (1-2 weeks), such information could be used to inform probabilistic estimates of the likelihood of those extremes across broad regions. These probabilistic forecasts could then be embedded into risk-based operating strategies for water infrastructure (Lavers et al. 2017).

Two prominent features emerge from the discussion above regarding what constitutes ‘useful’ information for water resource management in relation to an AR. First is the accuracy with which the location of the AR’s landfall can be predicted. Multiple analyses have shown that the effects of seasonality and geographical location on AR prediction skill are significant. Overall, AR frequency tends to be correlated with predictability and in the western US, specifically, the only reliable AR prediction skill is obtained in the winter months (NDJFM) (Mundhenk et al. 2016; Nardi et al. 2018; Deflorio et al. 2018). A number of mesoscale and synoptic indicators are important in predicting AR landfall location. The positioning and strength of both surface and mid-level pressure centers relates strongly to the location of landfall of ARs (Hu et al. 2017; Ralph et al. 2011; Guirguis et al. 2018; Hecht and Cordeira 2017). When these realizations are categorized via Rotated Empirical Orthogonal Functions (REOF) of 500 mb height anomalies, consistent patterns emerge between REOF phases and AR likelihood. For instance, more than 70% of ARs in northern California are associated with a single dipole of 2 of the 15 REOF modes (Guiruis et al. 2018). Similarly, characterization of AR correlated Rossby Wave Breaking (RWB) as either anti-cyclonic (AWB) or cyclonic (CWB) has shown a marked correlation with landfall latitude (Hu et al. 2017). CWB correlated ARs tend to landfall substantially farther south than their counterparts, reflecting an important teleconnection that is also highly correlated to the ENSO index. Further delimiting of AR occurrences by phases of atmospheric oscillatory modes may provide additional refinement of the spatial accuracy of IVT forecasts (Mundhenk et al. 2016; Mundhenk et al. 2017; Baggett et al. 2017). Depending on allowable spatial error, studies demonstrate that useful locational forecasts can be obtained for ARs out to approximately 7 days with significant decreases in skill thereafter (Deflorio et al. 2018). While results and methods vary substantially between these studies, the categorization of ‘allowable error’ consistently emerges as a key criterion in determining AR forecast utility. At longer lead times (7-14 days), accurately locating the landfall of an AR event is likely to be more realistic given current capabilities and more useful to a water resource manager than a full characterization of its structure.

The second feature of concern to water management is the accuracy with which physical characteristics of an AR can be resolved. Within spatially and geographically constrained subsets, a variety of atmospheric teleconnections have been identified that influence the intensity and orientation of ARs and their associated IVT fields. As above, positioning of geopotential height anomalies and RWB characteristics also prove to be highly important for strength and orientation of IVT at landfall (Guirguis et al. 2018; Hu et al. 2017). In a sense, geopotential height anomalies ‘steer’ IVT, influencing both its direction and intensity. Similarly, occurrences of AWB or CWB with an AR are correlated with average westerly or southwesterly oriented IVT respectively (Hu et al. 2017). Studies in northern California watersheds have documented the importance of this IVT orientation in determining precipitation distribution (Hecht and Cordeira 2017, Ralph et al. 2016). For the Sacramento River watershed, the amplification of a southwesterly oriented IVT field by a low-level southerly barrier jet is strongly correlated to the most extreme precipitation events in the historical record (Ralph et al. 2016). Unsurprisingly, IVT magnitude is also directly related to AR precipitation effects as the categorization scheme in (Ralph et al. 2018a) attests. Additional characteristics of the AR including temperature are also of seminal importance (Nardi et al. 2018). As ARs tend to be warmer than non-AR precipitation events, they often can produce Rain on Snow (ROS) occurrences that can increase streamflow by 50% over a non-ROS event. The most extreme cases of ROS events are all associated with warm ARs of high IVT magnitude owing to the exponential relationship between temperature and saturation vapor pressure (Guan et al. 2016). Warmer and stronger ARs tend to be those with the greatest connectivity to tropical moisture, though analyses differ on exactly how much tropical moisture is actually entrained in the land-falling portion of ARs and to what extent local or tropical/sub-tropical sources influence the magnitude of IVT (Dacre et al. 2015; Nusbaumer and Noone 2018; Ralph et al. 2011). Nevertheless, at shorter lead times (<7 days), it is much more likely that physical characteristics of a land-falling AR can be forecasted with precision and used to update decision processes related to water management. Since 3 day forecast informed reservoir operations have already shown significant improvement over baseline rule-curve operations, it is likely that an adaptive control structure could leverage refined forecast information continuously until the landfall of an AR event (Herman and Giuliani 2018; Nayak et al. 2018) .

In this study, I assess the utility of multiple aspects of AR forecasting in relation to the Sacramento River watershed, an area of intense water management. Previous studies have made substantial inroads into describing the predictability, characteristics and effects of ARs in various contexts ranging from global to sub-grid scales. Many have detailed specific processes and effects in this particular watershed. However, my goal is to conduct a functional evaluation of spatial and structural error in medium to extended range forecasts (1-14 days) of AR events and directly link this to the quality of information it can provide for water resource management in the region. First, I define the allowable spatial error in IVT forecasting that differentiates between an AR that produces significant effects in the watershed and one that does not. Then, I analyze clustering of error structures based on a set of extreme precipitation events in a 12 year period correlated with high resolution reforecasts. These clusters are used to highlight correlated synoptic-scale and larger weather phenomena that can provide useful forecast information regarding the location and extent of AR structures. Furthermore, I extend this logic to highlight these correlations with respect to some of the measurable dynamical components of ARs. Finally, I relate the results to the long-term goal of assimilating systematic error structures into more robust forecast-informed operating policies for water resources infrastructure.

**2. Data and Methods**

2.1. Study Region.

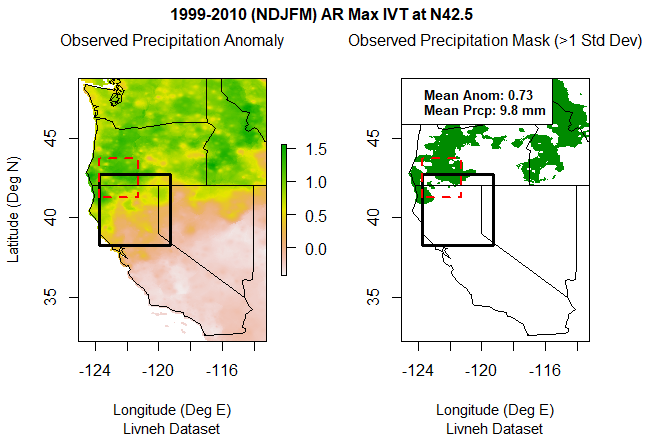
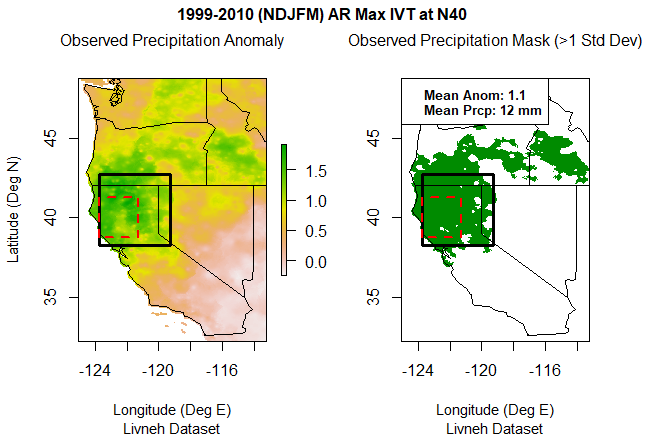
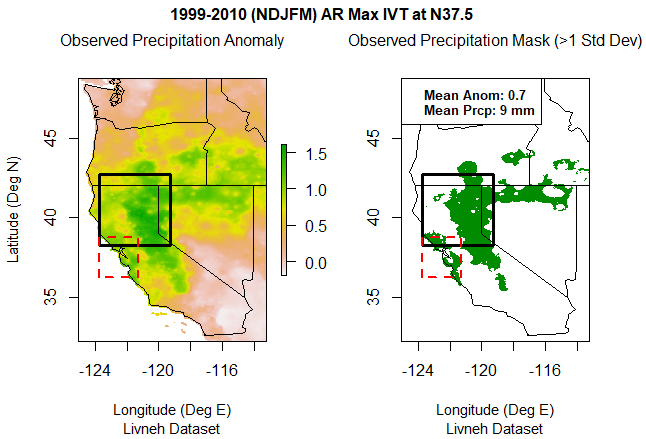


*Figure 1: Sacramento River Watershed*

The Sacramento River watershed, which terminates in San Francisco bay, is bounded by the northern California coastal range to the west, the Sierra Nevada range to the east, and the Trinity Alps/Cascade range to the north. To approximate this area, I used a geographical region bounded by N38.25 - 42.75 and W 119.25 – 123.75. This region corresponds with a three by three box of 1.5 degree grid cells (9 total grid cells) to match with the European Center for Medium Range Weather Forecasting (ECMWF) National Center for Environmental Prediction (NCEP) reforecast dataset and is about 400 km wide by 500 km long. Additionally, it is slightly larger than the actual confines of the watershed as shown by the graphical representation in Figure 1. The Folsom Reservoir in this catchment serves as an example indicative of common reservoir operating policies in the region due to the variable nature of precipitation effects. Current control rules requires that 40% (37.5% by special allowance) of the storage capacity behind the dam be left empty to capture flood flows in the winter rainy season. Thus, control schema that could make use of more of this flood pool safely offer a tremendous potential in water storage considering the oscillations between periods of abundant precipitation and multi-year droughts typical of the region. Quantification of this potential has already proven fruitful, with 3-day forecast informed operations showing a marked boost in water management efficiency when allowed to use the entire active storage of the reservoir (Nayak et al. 2018). More accurate predictions at longer lead times of extreme events would capitalize on these improvements, especially when considering that an AR propagated storm can produce inflows capable of filling the entire reservoir capacity once over during a multi-day event. Further, advanced knowledge of such events would allow manageable allocations of water to other uses, including storage in groundwater reservoirs (Whately et al. 2015).

2.2. Regional AR effects

Before commencing an event based analysis of forecast error, I characterized the extent of AR associated precipitation effects in the region to provide a reference frame. To accomplish this, I used gridded data from the Livneh dataset (Livneh et al. 2015) to provide accurate precipitation estimates in an area dominated by significant mountainous regions. Orographic effects can produce up to 50% errors between estimates of precipitation from models and gaged data (Livneh et al. 2015), a fact which will be significant in later analyses as well. To categorize AR events in the 1999 – 2010 time period, I used the SIO-R1 tabulated AR archive data to identify AR events land-falling at specific latitudes. This archive uses a 250 kg m/s threshold for IVT and a 15 mm threshold for IWV to classify individual grid boxes as AR objects with a length criterion of 1500 km and a persistence criterion of 18 hours. Landfall location is assigned to the 2.5 degree grid box along the coastline where maximum IVT is observed at any given 6 hour time step (Gershunov et al. 2015). In the time period indicated, I composited precipitation anomalies associated with AR landfalls centered at the latitudes of N35, 37.5, 40, and 42 as indicated in Figure 2 below. Precipitation anomalies were based on the gridded precipitation scaled and centered during the winter months of NDJFM. Based on these plots, it is evident that significant precipitation effects occur within the watershed with AR landfalls centered at the four indicated latitudes. Mean precipitation across the region varies between 7.6 and 12 mm (0.3 – 0.47 in) but would be much higher locally in mountainous regions and extend over many days during persistent events. Landfalls centered at N40 produce the most substantial precipitation effects across the region while those further south produce the least effects. Interestingly, precipitation effects tend to be concentrated along the latitude of landfall and northwards with lesser effects to the south. This fact suggests that there is more room for IVT prediction error to the south than to the north of the region. In distance terms, this equates to an allowable northward error of approximately 300 km (2.5 degree) and a southward error up to 600 km (5 degree) based on an initial forecast of IVT centered on the region.



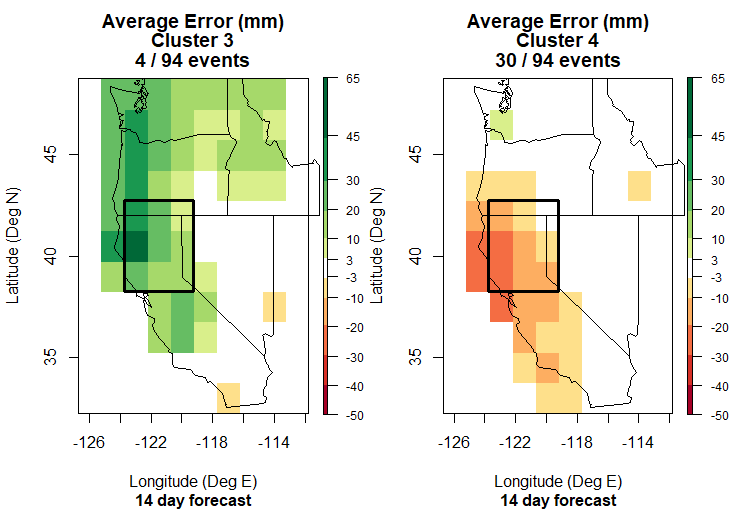
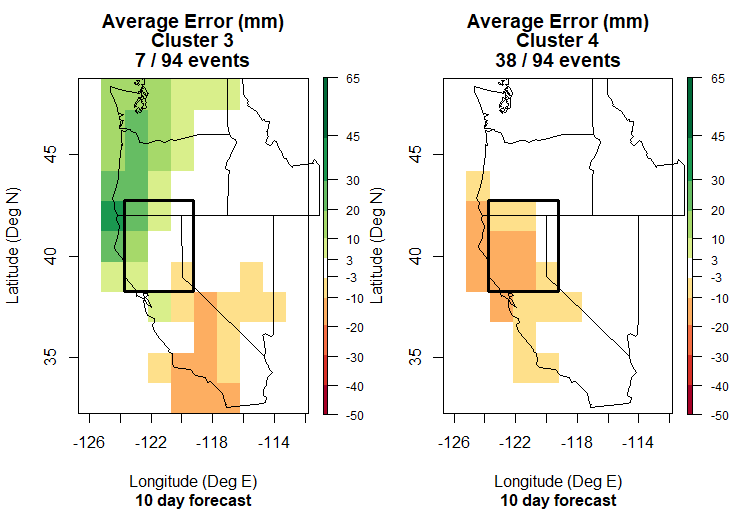
*Figure 2. Latitude of AR landfall maximum IVT (red dashed box) and associated precipitation anomalies. Right panel of each plot shows anomalies greater than one standard deviation above the mean.*

2.3. Characterization of extreme precipitation events

The remainder of the analyses in the study are based on a set of extreme precipitation events in the watershed region that occurred between 1999 and 2010. To correlate the observed Livneh data with the NCEP forecasts, I up-scaled the data to 1.5 degree grid cells. Then, the top 98th percentile precipitation events from each of the nine grid cells were selected, filtered to remove duplicate events, and de-clustered to remove multi-day occurrences. Furthermore, all events were indexed to the first day of any cluster occurrence and any events outside of the NDJFM period were removed. Therefore, the final set of 94 extreme precipitation days represented both events that impacted the entire region as well as some events that may have only impacted one or two grid cells around the edges. All of the 50 ‘extreme daily precipitation’ events occurring in NDJFM found in (Ralph et al. 2016) were represented in my selected events and correlation with ARs was high. Only 10 of the 94 events were not associated (+/- 1 day) with an AR event as classified in the SIO-R1 dataset.

2.4. Cluster Analysis

To identify error clusters in the extreme precipitation events, I performed k-means clustering on arrays of forecast error. Error arrays were created by calculating differences between NCEP forecasts at 1 -14 day lead times and the resultant observed precipitation for each grid cell. Importantly, I synchronized the Livneh dataset (based on daily observations recorded at midnight local time) to the NCEP forecast (based on GMT) at a 06:00Z reference. On PST in the winter, this equates to a two hour error in precipitation recording times between the two datasets, which was deemed insignificant for the purposes of this study. Each array was then centered using each grid cell’s mean error over the entire 12 year period. In executing the k-means clustering algorithm, I analyzed ratios of the sum of squared errors (SSE) between clusters to that of the total SSE at each forecast lead time. On average, clusters of 4 aligned with the ‘elbow’ in the scree plots where further increases in cluster size resulted in only marginal gains in the between to total SSE ratio. For consistency, this cluster size of 4 was used for all forecast leads and seemed to provide a qualitatively good delineation of error structures in a spatial sense. Thus, for each forecast lead time a subset of the 94 extreme precipitation dates was assigned to clusters 1 – 4 based upon grouping of similar error structures by the k-means algorithm. The error plots were then composited over all events assigned to a given cluster for a given lead time, resulting in a total of 56 cluster plots. Since the k-means algorithm arbitrarily assigns cluster numbers, I manually corrected cluster numbers based on qualitative similarity between cluster structures. These manual corrections were confined to forecast leads of 1, 5, 10, and 14 days for comparative purposes. Figure 3 shows a comparison between clusters 3 and 4 of the 10 and 14 day forecast respectively for illustration. Notably, cluster 3 for both forecast periods has a relatively small number of events while cluster 4 has more than 30% of the events. Also, for both forecast lead times cluster 3 displays a significant over-prediction structure at the latitude of the watershed and northwards while cluster 4 shows a prominent under-prediction structure at the watershed and southwards. These error structures and those of the other two clusters appear to persist across forecast lead times with some variation in the size of the clusters.



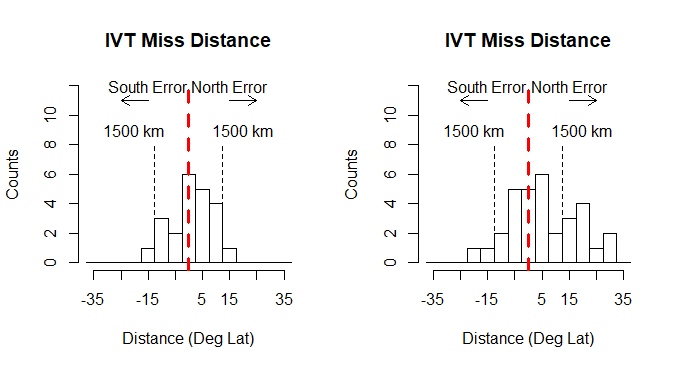
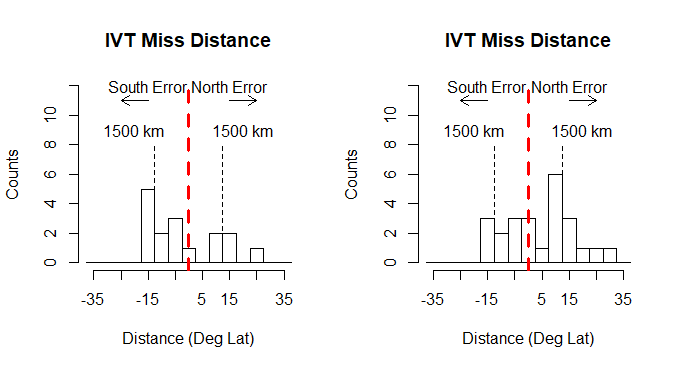
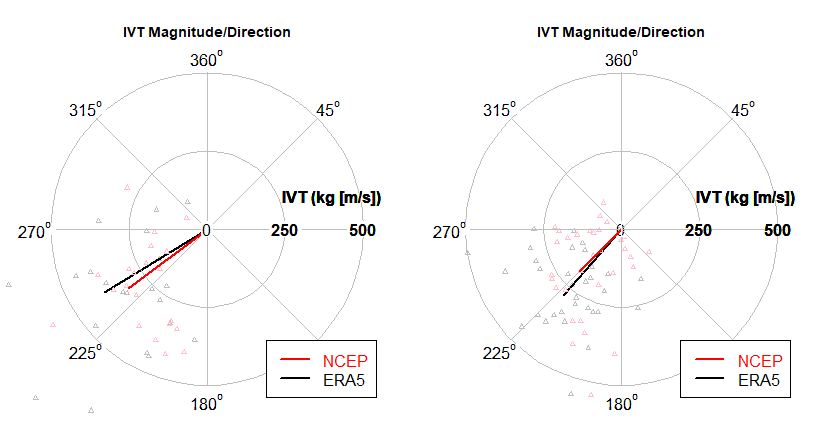
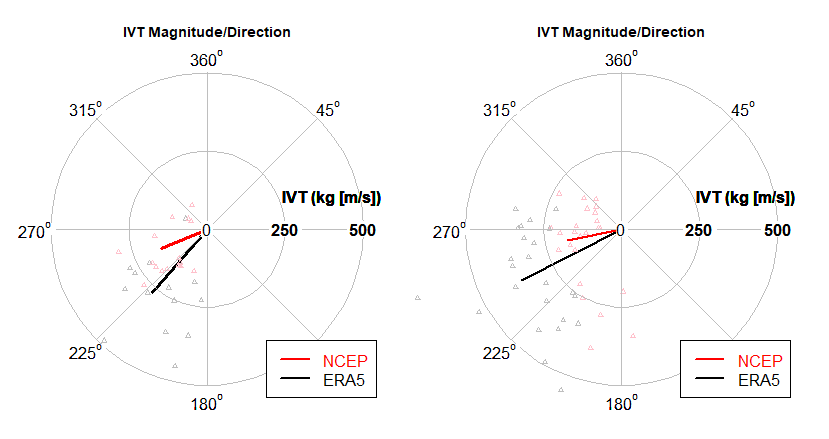
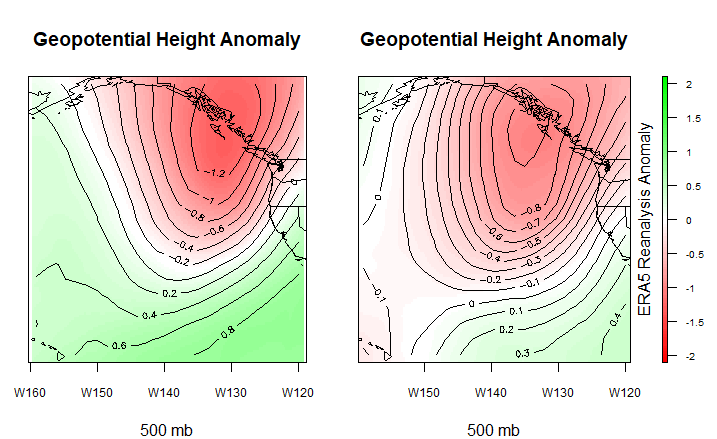
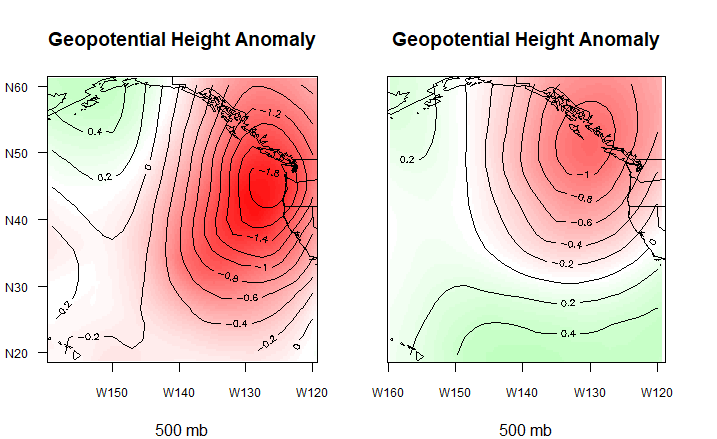
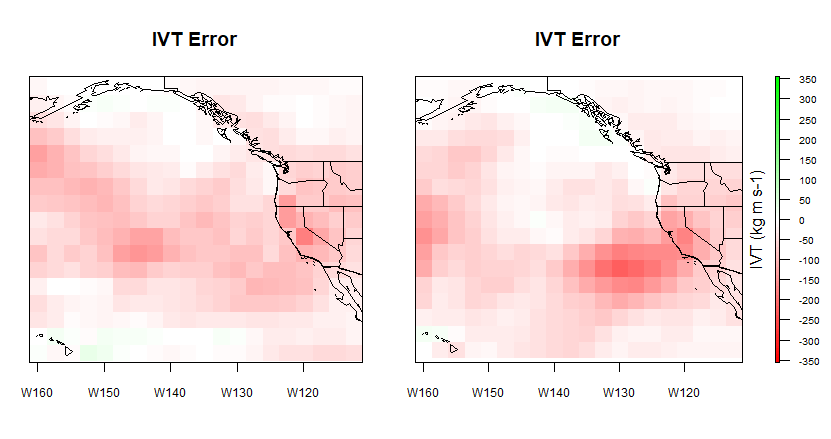
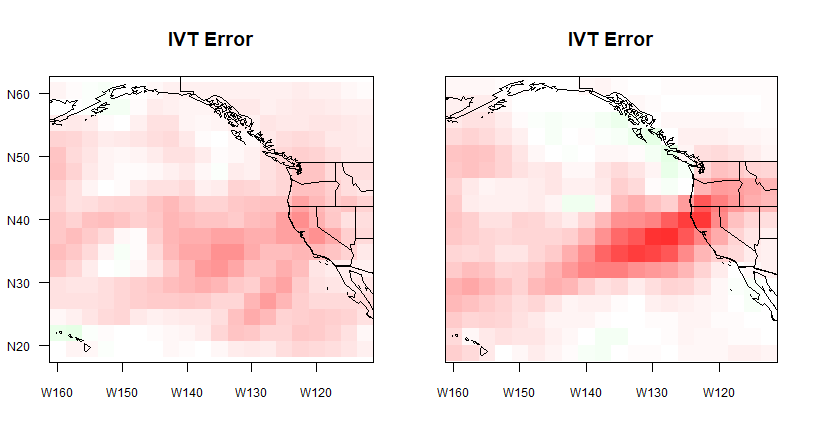
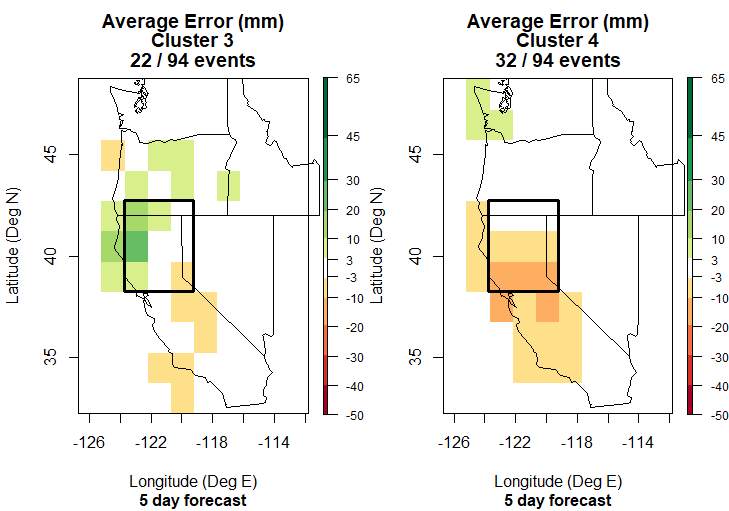
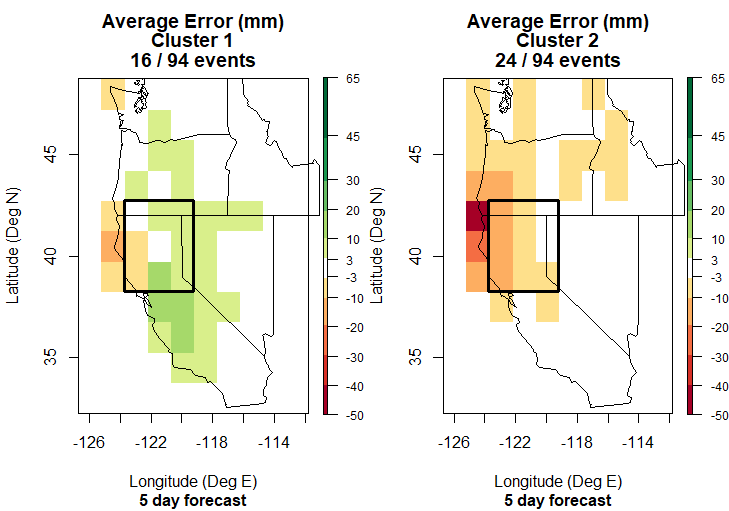
*Figure 3: Error clusters 3 and 4 for forecast lead times of 10 and 14 days. Positive (over-prediction) errors are indicated in green and negative (under-prediction) errors are indicated in orange.*

*Figure 3: Error clusters 3 and 4 from 10 and 14 day forecast lead times. Positive (over-predict) error are indicated in green and negative (under-predict) errors are indicated in orange*

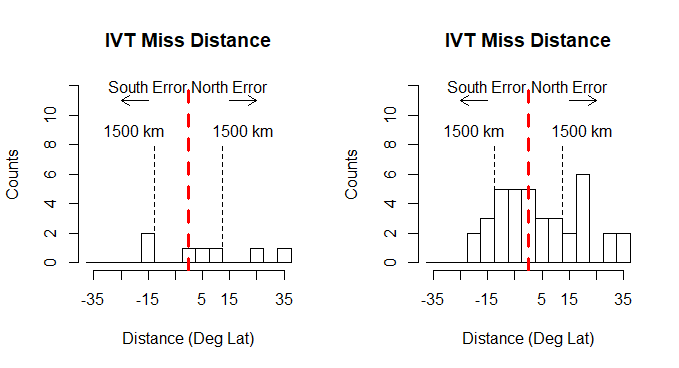
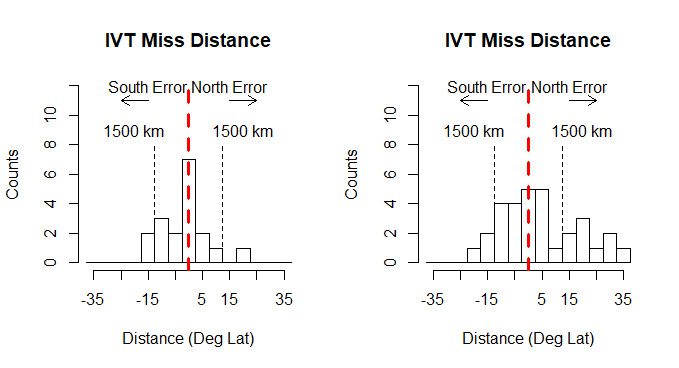
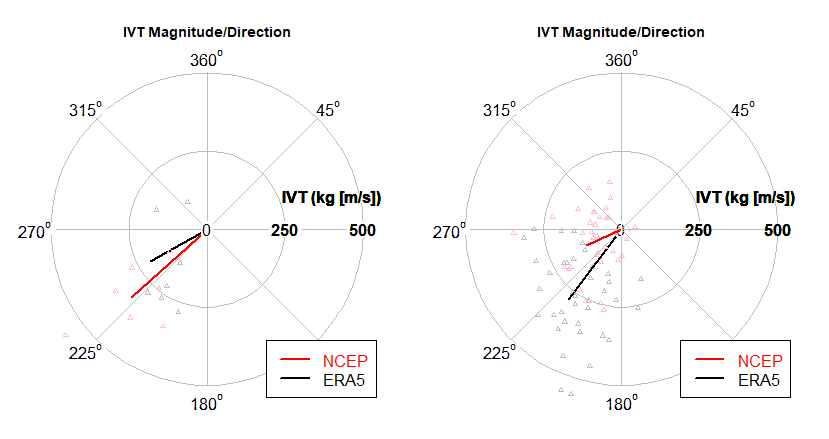
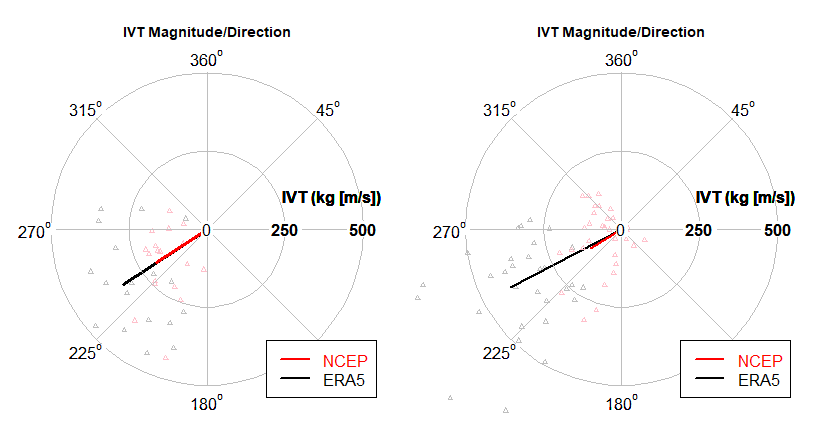
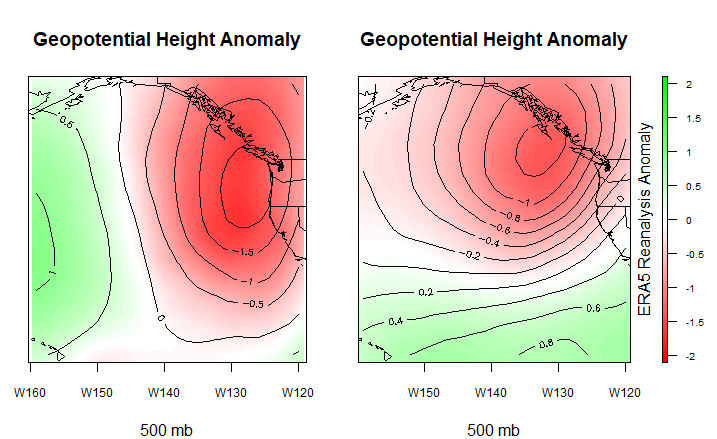
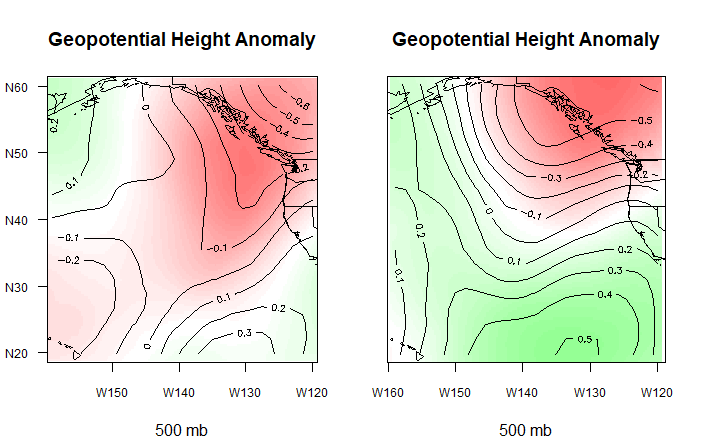
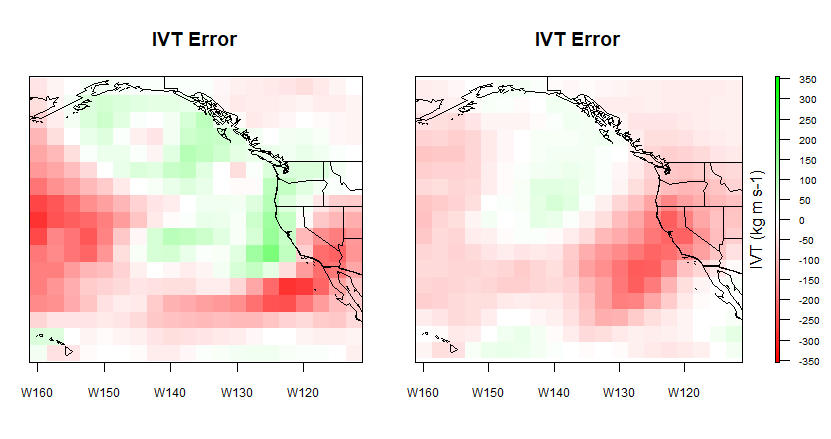
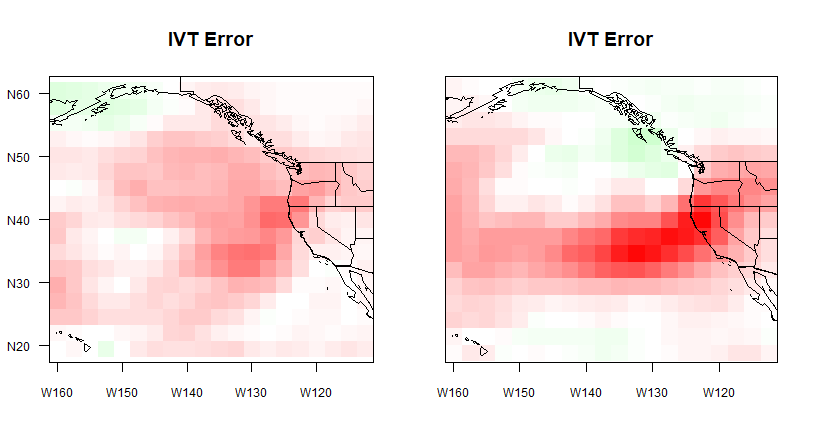
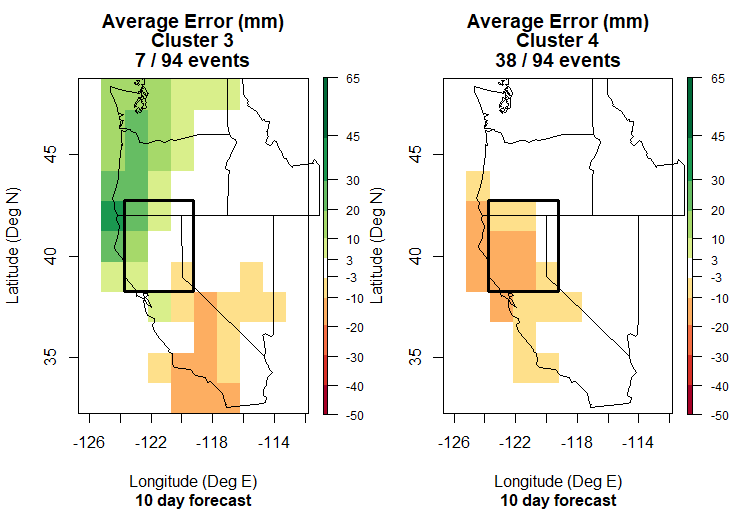
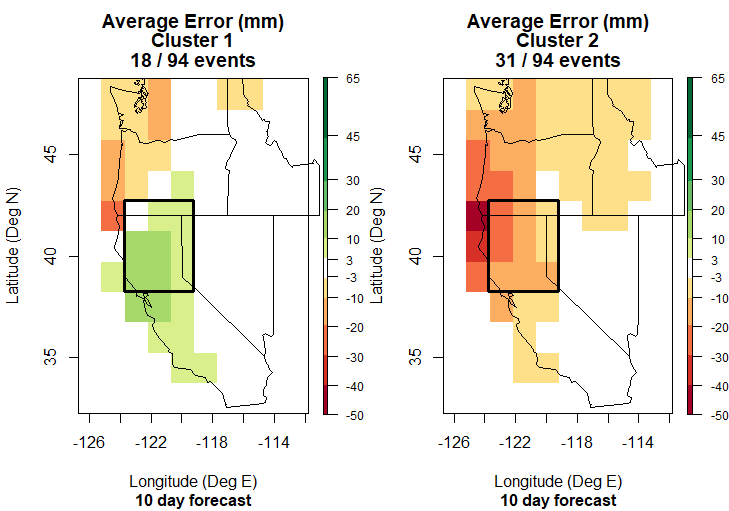
2.5. Mesoscale and synoptic correlations

Utilizing the cluster date indices, I assessed correlated structures of IVT error and geopotential height anomalies. Three components of IVT error were assessed with the first of these being the spatial error between forecast IVT magnitude and the observed values. Forecast values of IVT were obtained again from the ECMWF NCEP reforecast database and calculated via the integration procedures detailed in (Brands et al. 2017) and others. Observed values were extracted from the SIO-R1 archive which is derived from the NOAA-20C reanalysis products (Gershunov et al. 2017). Since reforecasts are only recorded at 24 hour time steps, it was impossible to perform a more precise synchronization with the observed Livneh data as was highlighted in section 2.4. Therefore, IVT features and those of geopotential height were synchronized to the 00:00 GMT time step that is 8 hours prior to the locally recorded precipitation observations. Because these signatures generally serve as precursors to precipitation events, this seemed to be a reasonable tradeoff. Additionally, the fact that only the first day of each extreme precipitation event was analyzed ensured that effects associated with persistent ARs or serial ARs were filtered out. The second row in Figures 4a-c shows the plots of composited IVT error across the eastern Pacific basin for the 5, 10, and 14 day lead times respectively. Plots associated with a 1 day lead time are catalogued in Appendix A since they did not show features of significant interest for my analysis. Correlated geopotential height anomalies are depicted in the third row with the colored shading indicating the ERA5 observed anomaly and the contours showing the NCEP forecast anomaly based on a monthly mean. In the fourth row, the radial plots display the difference between the mean NCEP forecast values of IVT magnitude and direction and the mean ERA-5 observed values in a geographical box defined by N36 – 42 and W121.5 – 126. This area projects over the western portion of the watershed and approximately 300 km into the Pacific Ocean while spanning a region longitudinally that is slightly larger than the watershed. Lightly shaded points on the radial plot indicate individual IVT magnitude and direction values for each date in the cluster whereas the darkly shaded lines indicate the average value for the forecast (red) and the observed (black) data. Finally, the histograms in the fifth row show the latitude error between the forecasted landfall location (latitude of maximum IVT) and the observed landfall location. These errors are shown in a relative sense since the actual location of observed and predicted landfall is not depicted in the plot.

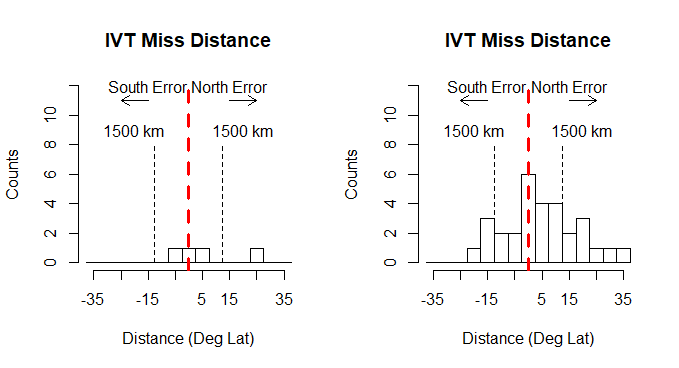
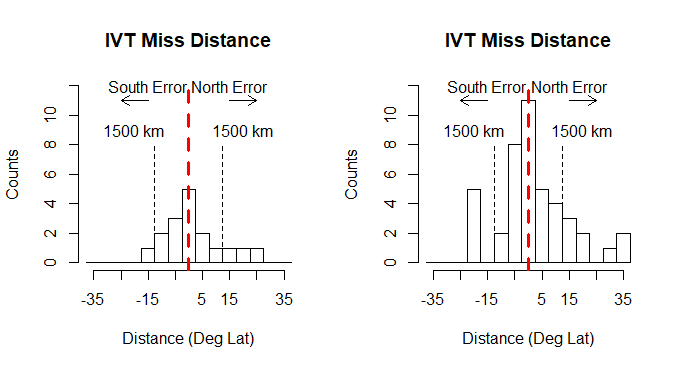
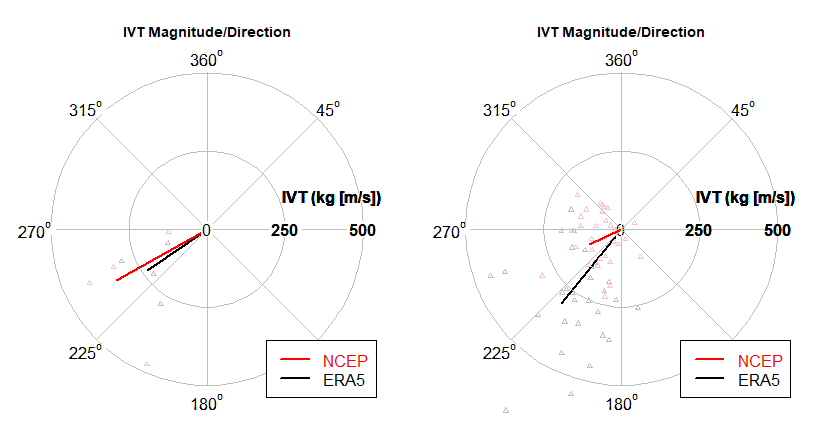
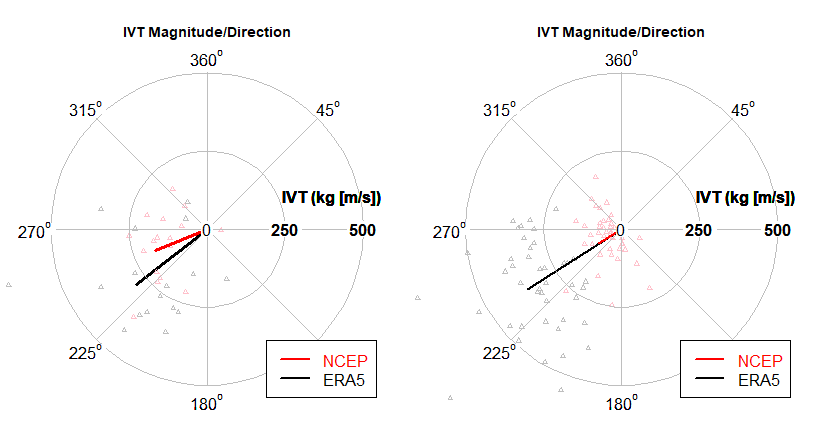
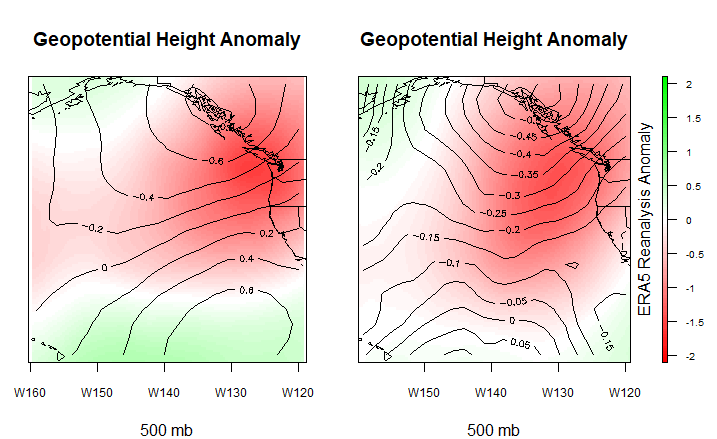
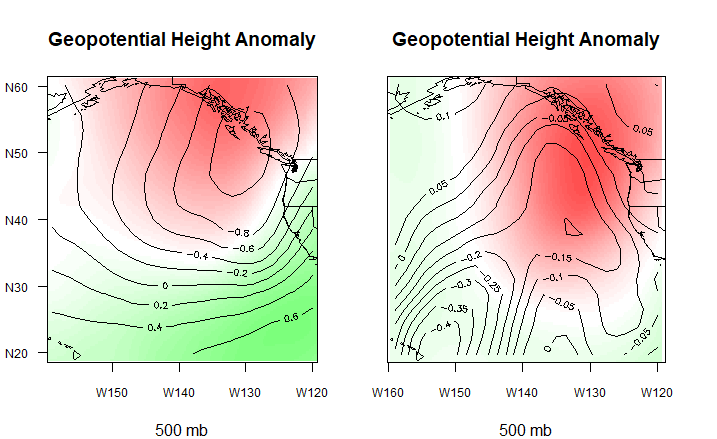
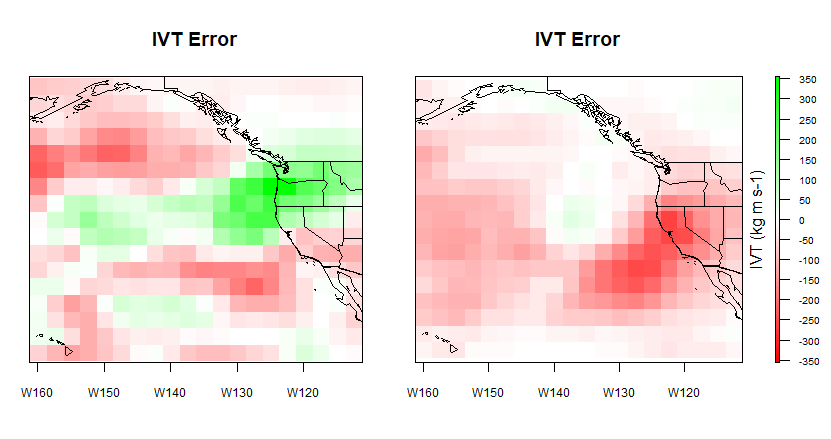
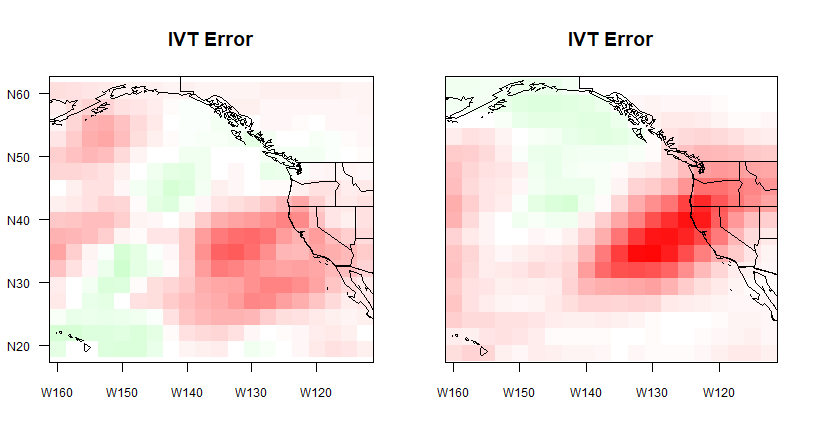
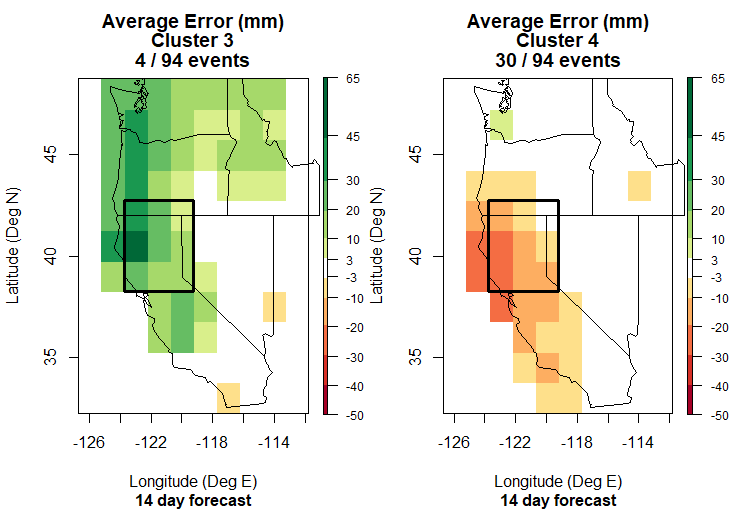
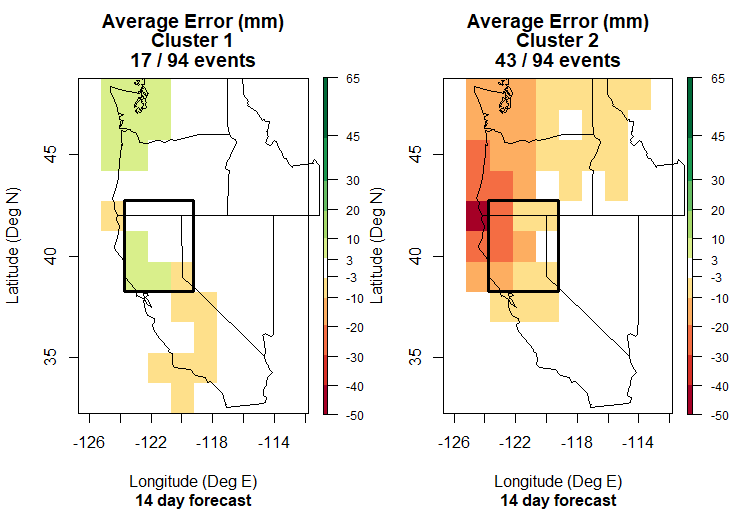
*Figure 4a: 5 day forecast lead plots of composite error, IVT error, and geopotential height anomaly in top 3 rows; Comparison of forecast to observed IVT magnitude and direction in row 4 and histogram of IVT landfall miss distance in degrees of latitude in row 5.*



*Figure 4b: 10 day forecast lead plots of composite error, IVT error, and geopotential height anomaly in top 3 rows; Comparison of forecast to observed IVT magnitude and direction in row 4 and histogram of IVT landfall miss distance in degrees of latitude in row 5.*



*Figure 4c: 14 day forecast lead plots of composite error, IVT error, and geopotential height anomaly in top 3 rows; Comparison of forecast to observed IVT magnitude and direction in row 4 and histogram of IVT landfall miss distance in degrees of latitude in row 5.*



**3. Results**

One of the interesting results already hinted at is the significant correlation across forecast lead times of the spatial error structure. Clusters 2 and 4 in particular show a consistent spatial signature in precipitation error. Essentially, both clusters show a subset of precipitation under-predictions, but cluster 2’s under-prediction region is centered at the northwestern corner of the watershed area whereas cluster 4’s under-prediction region is at the southwest corner. In line with these precipitation errors, both clusters also show a discernible IVT error structure in both a spatial sense as well as in the landfall metrics from the radial plots. In regards to the spatial IVT error structure, Cluster 2 shows a long narrow region of zonally oriented under-prediction of IVT centered on northern California. Cluster 4 also shows a long narrow structure of IVT error but it is more southwesterly oriented and impinges the coastline slightly further south. This observation is bolstered by the radial plots which show observed land-falling IVT between 225 and 270 degrees for cluster 2 versus IVT between 180 and 225 degrees for cluster 4. Moreover, in the longer lead forecasts (10 and 14 days), the NCEP forecast error in cluster 2 is almost entirely an under-prediction in land-falling IVT magnitude while in cluster 4 there is an error in both magnitude and direction. More precisely, the NCEP forecast in cluster 4 predicts a much weaker and more westerly IVT structure than that which actually materializes. Weak IVT error over-predict structures to the north of the strong under-predict structures suggest that either the model forecast the IVT signature too far to the north or did not forecast an AR type event at all. This notion is bolstered in clusters 2 and 4 by the histograms at both the 10 and 14 day forecast lead times showing significant northward skew in the miss distance histograms. In other words, the model predicted the IVT centroid well north of its actual landfall location with errors as large as 35 degrees of latitude. This northward skew appears to be present in cluster 3 as well at longer lead times, but is less significant. Since clusters 2 and 4 comprise well more than 50% of the cases, this strong northward error in IVT prediction is particularly relevant to an overall analysis and in underscored by analyses in (Nardi et al. 2018).

Out to 10 days, forecasts of geopotential height show a qualitatively high level of skill as described by the close agreement between the contours (NCEP forecast) and color shading (ERA-5 reanalysis). Furthermore, at a 10 day forecast lead time the geopotential height contour show marked similarities to the IVT error structure. While all clusters display the strong high-low dipole characteristic of land-falling AR events on the west coast (Guirguis et al. 2018), it is evident that each error cluster is associated with a geopotential height anomaly structure that varies considerably in zonal and meridional location as well as the curvature and strength of gradients. Interestingly, cluster 1 has the least defined geopotential structure at a 10 day lead time whereas it has the most defined and southerly structure at a 1 and 5 day lead. It is likely that this fact results from changes in cluster membership through lead times and the less defined error structures with shorter lead times. As can be seen in the 1 day forecast error structures in Appendix A, the cluster patterns in precipitation error are almost indiscernible at this short of a lead time. This fact may point to use of other clustered error data such as geopotential height anomaly or simply commonality of membership as a better delineator of a given cluster index across forecast lead times. Nonetheless, the geopotential height anomalies for clusters 2 – 4 at a 10 day lead do show patterns that can be visually correlated to errors in IVT prediction. In cluster 2, the geopotential height low anomaly is positioned further north than the other clusters and the gradient between the high and low anomalies is more zonally oriented and of relatively moderate strength. A superposition of the IVT under-prediction error structure over this gradient shows a marked correlation in the orientation and positioning of the IVT error at landfall. Similar super-positions are easily envisaged for clusters 3 and 4, with cluster 3 having the most southward and curved profile while cluster 4 sits somewhere between clusters 2 and 3. At a 10 day lead, this result suggests that large scale information exists in climate models showing an accurate relationship between upper level structures associated with an AR and its landfall dynamics. Also, it is notable that this strong agreement exists even though the NCEP forecasts of IVT are compared to NOAA-20C reanalysis data whereas the forecasts of geopotential height anomaly are compared to ERA-5 reanalysis data. Moreover, the agreement in these structures breaks down considerably going from the 10 day forecast lead to the 14 day lead time. The cluster 1 agreement remains reasonable but clusters 2 – 4 show significant disagreement with a trend of projecting the low pressure anomaly too far to the north, particularly in clusters 3 and 4. In light of this, the pronounced precipitation error structures in clusters 2 – 4 and the relatively small ones in cluster 1 further accentuate the connection between accurate resolution of large scale features and accurate prediction of local precipitation effects.

**4. Discussion**

Returning to the water resource management premise of this study, the results highlight important areas where forecast skill is evident in predicting certain aspects of the AR related atmosphere and decidedly inaccurate in others. At a 5 day forecast lead, performance in a majority of the AR forecast elements analyzed was generally reasonable with only cluster 2 showing any substantial inaccuracies of concern in IVT and precipitation error. Since this cluster is an ‘under-predict’ for the watershed, this type of error would be of concern for flood management and deserves further analysis. Additionally, it comprises 24 of 94 events and so is a significant portion of the total event pool. Cluster 1 also shows significant error in prediction of IVT magnitude and direction but its errors in forecasting precipitation and spatial aspects of IVT are less pronounced. Though the resultant effects were not as significant as those in cluster 2, the nature of the error is worthy of discussion. The actual IVT magnitude was much greater than forecasted and the actual IVT impinged from a much more southwesterly direction than forecasted. As noted, these landfall characteristics tend to produce the greatest precipitation effects for the Sacramento watershed. Errors of this type further amplified, as shown by cluster 4 at longer lead times can result in substantial under-predictions of precipitation and again the possibility for reservoir inflows much higher than anticipated. Otherwise, forecasts of these extreme events at a 5 day lead suggest sufficient skill to be useful for forecast informed reservoir operations, especially when 3 day forecasts have already shown considerable improvement potential.

At longer lead times, certain aspects of forecast skill break down more notably than others. For a 10 day lead, precipitation forecast shows much more distinct clusters of spatial error which is apparent in the IVT forecasts too. This is tempered by strong agreement between the forecast and observed geopotential height anomaly that may offer opportunities for probabilistic updating of the potential for an extreme precipitation event when an AR event is forecasted outside the spatial extent of the watershed. The northward bias extant in many of the clusters at longer lead times indicates that a persistent spatial bias materializes in the forecast model, at least for the extreme events analyzed herein. Therefore, it would be reasonable to infer that a forecast of a strong IVT event or AR north of a watershed incurs some substantial probability of shifting further south. Considering the analysis in section 2.2 that showed precipitation effects were strongest at the latitude of impinging IVT and northwards, this northward bias is less of a concern if an AR is initially forecast to impact the watershed. If it shifts further south from the forecast, it is still likely to have effects reaching to the north of its landfall location and bring precipitation to the region. From 10 to 14 days, the breakdown in forecast skill across the analyzed components is substantial. One consistent element in the 14 day forecast is a northward biased spatial over-prediction of IVT which is matched in clusters 3 and 4 by a northward bias in prediction of the geopotential low. Additionally, cluster 1 maintains a low prediction error for precipitation across 17 of 94 events and reasonable agreement in the geopotential height anomaly. Further analysis is warranted to see what atmospheric conditions existed to allow this high level of forecast skill to exist out to 14 days. On the other hand, because the majority of errors at 14 days are significant under-predictions, a determination of whether AR like structures were predicted at all during these events might yield important information regarding model characteristics. If ARs or AR like events are being predicted, then the probabilistic framework discussed for the 10 day lead might make sense at a 14 day lead with some additional uncertainty. If these type of events are not being predicted at all, then there is likely little useful information regarding them at this long of a lead time. Case by case analyses of these events suggest that the former is more likely to be true than the latter.

**5. Conclusion**

In line with other research, this study has highlighted the correlation between extreme precipitation events in the Sacramento River watershed and AR weather events. By clustering these events in relation to their spatial precipitation error structure, I found that correlated synoptic scale weather features show interpretable signatures of error that remain structured out to 10 days of forecast lead time. Even at 14 days of forecast lead, some error signatures are discernible, but would require additional analyses to quantify. The existence of these error patterns implies that there may be underlying biases in the forecast model that if understood more thoroughly, would allow a probabilistic determination of the risk of an AR associated extreme precipitation event for a given watershed region. Overall, the results show a tendency for a northward bias in forecasts that appears to increase with forecast lead time. This fact may allow a useful likelihood estimate for the potential impact of an AR event that is projected to be outside of the watershed. Furthermore, the reliability of certain climactic indices, especially geopotential height anomaly, was much greater at longer lead times than that of localized precipitation. Perhaps such sources of quality information could be used to improve probabilistic estimates of a high impact event traditionally based on local forecasts of precipitation.

At the very least, these results underscore the idea that synoptic scale features offer a more resolvable forecast signature than model outputs at the grid scale. Certainly this is a feature that could be leveraged in water resources management architectures utilizing forecast informed operating policies. While it is true that water resources systems can be managed effectively with no forecast information, various types of more nuanced policies like groundwater injection would be significantly benefitted by long lead forecast knowledge. For future work, this study opens a number of pathways. Within the study framework itself, comparisons of different forecast and reanalysis models would improve the scope of the results. In addition, more quantitative comparisons of error patterns are needed to determine relative significance of outcomes while more qualitative comparisons of individual cases could illuminate other aspects of the results. Subsequent studies could flesh out the probabilistic framework discussed and apply it to actual reservoir operation policies. Finally, the documented connections between sub-seasonal, seasonal, and inter-annual atmospheric modes and their relationship to ARs could similarly be applied to such forecast informed operations. Again, it is likely that subtler aspects of water management could be significantly improved by these efforts while realizing improvements in day to day operating efficiency.

**References**

Baggett, C. F., Barnes, E. A., Maloney, E. D., & Mundhenk, B. D. (2017). Advancing atmospheric river forecasts into subseasonal-to-seasonal time scales. *Geophysical Research Letters*, *44*(14), 7528–7536. <https://doi.org/10.1002/2017GL074434>

Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, *525*(7567), 47–55. <https://doi.org/10.1038/nature14956>

Brands, S., Gutiérrez, J. M., & San-Martín, D. (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 Dynamics*, *48*(9–10), 2771–2795. https://doi.org/10.1007/s00382-016-3095-6

Dacre, H. F., Clark, P. A., Martinez-Alvarado, O., Stringer, M. A., & Lavers, D. A. (2015). How do atmospheric rivers form? *Bulletin of the American Meteorological Society*, *96*(8), 1243–1255. <https://doi.org/10.1175/BAMS-D-14-00031.1>

DeFlorio, M. J., Waliser, D. E., Guan, B., Lavers, D. A., Ralph, F. M., & Vitart, F. (2018). Global assessment of atmospheric river prediction skill. *Journal of Hydrometeorology*, JHM-D-17-0135.1. <https://doi.org/10.1175/JHM-D-17-0135.1>

Gershunov, A., Shulgina, T., Ralph, F. M., Lavers, D. A., & Rutz, J. J. (2017). Assessing the climate-scale variability of atmospheric rivers affecting western North America. *Geophysical Research Letters*, *44*(15), 7900–7908. https://doi.org/10.1002/2017GL074175

Guan, B., Waliser, D. E., Ralph, F. M., Fetzer, E. J., & Neiman, P. J. (2016). Hydrometeorological characteristics of rain-on-snow events associated with atmospheric rivers. *Geophysical Research Letters*, *43*(6), 2964–2973. https://doi.org/10.1002/2016GL067978

Guirguis, K., Gershunov, A., Clemesha, R. E. S., Shulgina, T., Subramanian, A. C., & Ralph, F. M. (2018). Circulation Drivers of Atmospheric Rivers at the North American West Coast. *Geophysical Research Letters*. <https://doi.org/10.1029/2018GL079249>

Hecht, C. W., & Cordeira, J. M. (2017). Characterizing the influence of atmospheric river orientation and intensity on precipitation distributions over North Coastal California. *Geophysical Research Letters*, *44*, 9048–9058. <https://doi.org/10.1002/2017GL074179>

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

Hu, H., Dominguez, F., Wang, Z., Laversa, D. A., Zhang, G., & Ralph, F. M. (2017). Linking atmospheric river hydrological impacts on the U.S. West Coast to Rossby wave breaking. *Journal of Climate*, *30*(9). <https://doi.org/10.1175/JCLI-D-16-0386.1>

Lavers, D. A., Zsoter, E., Richardson, D. S., & Pappenberger, F. (2017). An Assessment of the ECMWF Extreme Forecast Index for Water Vapor Transport during Boreal Winter. *Weather and Forecasting*, *32*(4), 1667–1674. <https://doi.org/10.1175/WAF-D-17-0073.1>

Livneh, B., et al. (2015). A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950-2013. *Scientific Data*, *2*, 1–12. https://doi.org/10.1038/sdata.2015.42

Mundhenk, B. D., Barnes, E. A., Maloney, E. D., & Baggett, C. F. (2017). Skillful empirical subseasonal prediction of landfalling atmospheric river activity using the Madden–Julian oscillation and quasi-biennial oscillation. *Npj Climate and Atmospheric Science*, *1*(1), 7. <https://doi.org/10.1038/s41612-017-0008-2>

Mundhenk, B. D., Barnes, E. A., & Maloney, E. D. (2016). All-season climatology and variability of atmospheric river frequencies over the North Pacific. *Journal of Climate*, *29*(13), 4885–4903. <https://doi.org/10.1175/JCLI-D-15-0655.1>

Nardi, K. M., Barnes, E. A., & Ralph, F. M. (2018). Assessment of Numerical Weather Prediction Model Reforecasts of the Occurrence, Intensity, and Location of Atmospheric Rivers along the West Coast of North America, (2011), 3343–3363. <https://doi.org/10.1175/MWR-D-18-0060.1>

Nayak, M. A., Herman, J. D., & Steinschneider, S. (2018). Balancing flood risk and water supply in California: Policy search integrating short-term forecast ensembles with conjunctive use. *Water Resources Research*, 1–20. <https://doi.org/10.1029/2018WR023177>

Nusbaumer, J., & Noone, D. (2018). Numerical Evaluation of the Modern and Future Origins of Atmospheric River Moisture Over the West Coast of the United States. *Journal of Geophysical Research: Atmospheres*, *123*(12). <https://doi.org/10.1029/2017JD028081>

Ralph, F. M., Cordeira, J. M., Neiman, P. J., & Hughes, M. (2016). Landfalling Atmospheric Rivers , the Sierra Barrier Jet , and Extreme Daily Precipitation in Northern California ’ s Upper Sacramento River Watershed. *Journal of Hydrometeorology*, *17*, 1905–1914. https://doi.org/10.1175/JHM-D-15-0167.1

Ralph, F. M., Neiman, P. J., Kiladis, G. N., Weickmann, K., & Reynolds, D. W. (2011). A Multiscale Observational Case Study of a Pacific Atmospheric River Exhibiting Tropical–Extratropical Connections and a Mesoscale Frontal Wave. *Monthly Weather Review*, *139*(4), 1169–1189. <https://doi.org/10.1175/2010MWR3596.1>

Ralph, F. M. et al. (2018a). A Scale to Characterize the Strength and Impacts of Atmospheric Rivers. *Bulletin of the American Meteorological Society*, *100*(2), 269–289. <https://doi.org/10.1175/bams-d-18-0023.1>

Ralph, F. M., Dettinger, M. D., Cairns, M. M., Galarneau, T. J., & Eylander, J. (2018b). Defining “Atmospheric River”: How the Glossary of Meteorology Helped Resolve a Debate. *Bulletin of the American Meteorological Society*, *99*(4), 837–839. https://doi.org/10.1175/bams-d-17-0157.1

Whateley, S., Palmer, R. N., & Brown, C. (2015). Seasonal Hydroclimatic Forecasts as Innovations and the Challenges of Adoption by Water Managers. *Journal of Water Resources Planning and Management*, *141*(5), 04014071. <https://doi.org/10.1061/(ASCE)WR.1943-5452.0000466>

Appendix A. 1 day forecast plots

