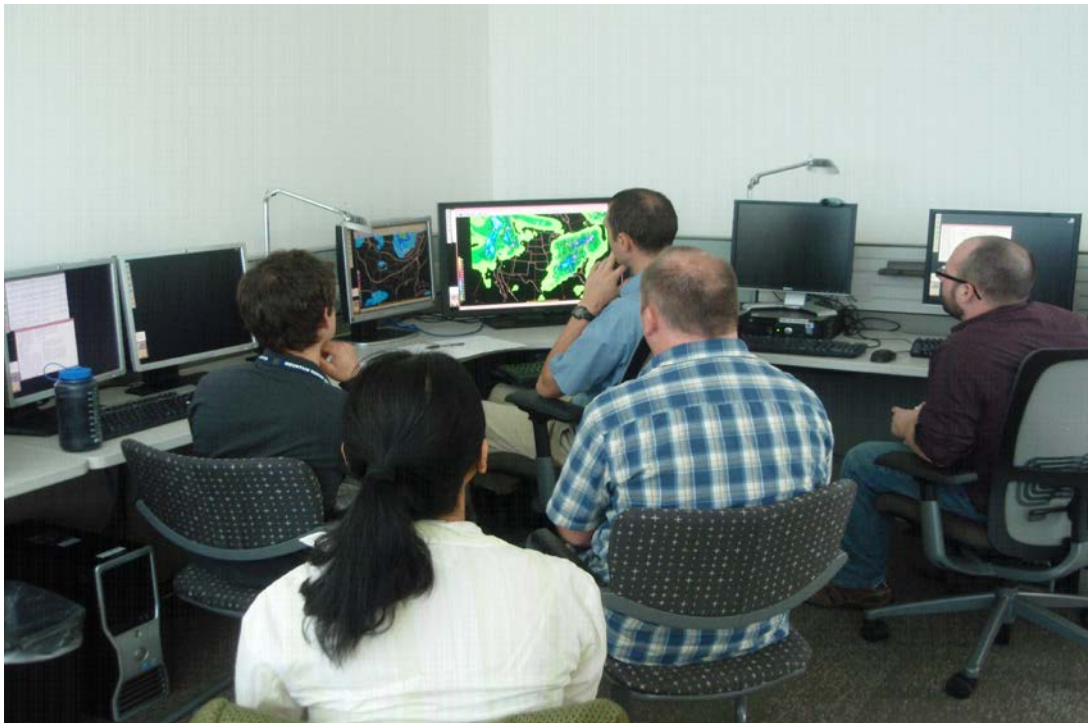


The 2012 Atmospheric River Retrospective Forecasting Experiment

Final Experiment Report



1. Introduction

In collaboration with the Earth Systems Research Laboratory (ESRL), the Hydrometeorological Testbed (HMT)-Hydrometeorological Prediction Center (HPC) hosted the Atmospheric River Retrospective Forecasting Experiment (ARRFEX) from September 17-28, 2012. The experiment brought together forecasters, modelers and researchers (Appendix A) from River Forecast Centers (RFC), Weather Forecast Offices (WFO), the Environmental Modeling Center (EMC), ESRL and the University of Utah to identify potential techniques to improve forecasts of atmospheric river (AR) induced extreme precipitation events along the U.S. West Coast.

The experiment portion of the project featured retrospective analysis of 8 pre-selected AR events that resulted in heavy precipitation along the U.S. West Coast during the 2009-2012 cool seasons (Appendix B). The participants were instructed to create a variety of forecast products, in pseudo-real time, for each event using archived operational and experimental numerical model guidance and datasets.

The goal of ARRFEX was to investigate two primary issues: 1) evaluate operational and experimental numerical modeling datasets in forecasting West Coast extreme precipitation events, and 2) discuss ways to provide better forecast information to customers. Specifically, the experiment focused on the following:

- Determining the predictability of quantitative precipitation forecasts (QPF) for heavy precipitation events using current operational deterministic and ensemble guidance at mid-range (3 and 5 day) lead times.
- Determining the utility of probabilistic QPF (PQPF) for heavy precipitation events.
- Examining the utility of the high resolution ensembles QPF/PQPFs at short to medium range lead times.
- Examining the utility of the ESRL 2nd generation reforecast dataset and techniques for QPF/PQPF at short to medium range lead times.
- Investigating the predictability of the timing and duration of AR-induced precipitation.

This report presents the activities and findings of the retrospective forecasting portion of the experiment.

2. Data and Methodology

A) Data

ARRFEX featured a variety of numerical guidance systems (Table 1), both operational and experimental. Archived numerical guidance data was provided (when available) from 12 UTC initializations (except for the experimental reforecast dataset) from dates 7, 5, 3 and 1 day prior to the occurrence of the event.

Several operational deterministic and ensemble guidance systems were featured in experiment:

- NCEP's Global Forecast System (GFS) deterministic model and its ensemble system (GEFS)
- The European Centre for Midrange Weather Forecasting (ECMWF) deterministic model and its ensemble system (ECENS)
- The Canadian Meteorological Centre Ensemble (CMCE)
- NCEP's North American Mesoscale model (NAM)

In addition to the operational guidance, several experimental guidance systems and forecast tools were used and evaluated. These are described in the section below.

Provider	Model	Resolution	Forecast Hours
NCEP	GFS	1.0 deg	180 h
NCEP	GEFS	70 km	180 h
ECMWF	ECMWF	1.0 deg	196 h
ECMWF	ECENS	70 km	196 h
NCEP	NAM	32 km	84 h
CMC	CMCE	1.0 deg	196 h
ESRL/GSD	HMT-Ensemble	9 km	84 h
ESRL/PSD	ESRL Reforecast Dataset	32 km	96 h
NCEP/CMCE/ECMWF	MMENS	70 km	180 h

Table 1) Numerical guidance used in ARRFEX. Experimental datasets denoted by gray shading.

Experimental Datasets and Forecast Tools

Hydrometeorology Testbed (HMT) Ensemble:

The HMT ensemble is a high-resolution (9 km), multiple-physics scheme ensemble produced by ESRL. Normally at 9 members, there were 7 members (Appendix C) in the ARRFEX version due to difficulty initializing the NMM core members retrospectively. The ensemble not only provides the advantage of higher resolution, but also provides multiple precipitation-physics schemes, potentially providing an advantage over single-physics, larger resolution ensembles in handling topography and quantifying forecast uncertainty and variability. The ensemble system was initialized by LAPS (<http://laps.noaa.gov/>) and was run out to 84 hours. However, the ensemble uses a relatively smaller domain (30°N, -134°W; 45°N, -112°W) (Figure 1), which limited its usage for events where precipitation was focused in northern Oregon and Washington. Data was not available for the October 2009 case (Appendix B, event #4).

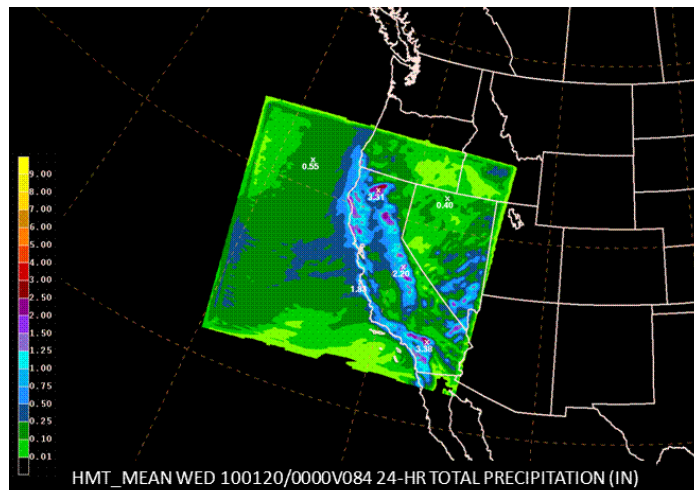


Figure 1) Showing the domain of the HMT ensemble with the mean 24 h QPF valid 00 UTC 20 January 2010.

Multi-Member Ensemble

The Multi-Member Ensemble (MMENS) is a multi-model ensemble which contained 90 members: the 20 perturbed members of the GFS Ensemble Forecast System (GEFS), the 20 perturbed members of the Canadian Meteorological Centre Ensemble (CMCE), and the 50 perturbed members of the European Center for Midrange Weather Forecasting (ECMWF) Ensemble (ECENS). The system was created only for the 8 archived cases used in the experiment in an effort to examine the potential benefit of multi-model ensembles in creating PQPF. Multi-member ensembles are candidates to improve PQPF forecasts, as they capture the true range of potential QPFs across several guidance systems. A recent study by Hamill (2012) concludes that multi-model ensembles do provide benefit to probability forecasts, and suggests further investigation into their creation and usage.

Reforecast guidance from the ESRL 2nd Generation Reforecast Dataset (GEFS)

PSD/ESRL recently released their 2nd Generation Reforecast Dataset, which uses version 9.0.1 of the GEFS (implemented 14 February, 2012). Further details on the dataset can be found at:

http://www.esrl.noaa.gov/psd/forecasts/reforecast2/README.GEFS_Reforecast2.pdf

The reforecast guidance featured in ARRFEX consisted of 24 h PQPF and 24 h mean QPF based off the ESRL 2nd Generation Dataset. For each of the specific forecasts and lead times used in the experiment, the output of the 00 UTC initialization of the GEFS mean was compared statistically, at each grid point, to a collection of relevant reforecasts of the same forecast lead-time. This collection contains all reforecasts within 3 months of the initialization date for the entire 25 year period contained in the dataset. Forecasts of the same lead time (e.g. 72 hours) for all relevant 2275 cases were compared to the current forecast using a ranked-analog technique. Once the dates of the top 50 closest analogs were determined, observed precipitation data (from the North American Regional Reanalysis dataset) from the 50 dates was used to calculate probability of precipitation and mean precipitation at that grid point.

Recent studies (Hamill and Whitaker, 2006; Hamill, 2012) have shown favorable results in regards to using reforecast products in PQPF.

Standardized Anomalies

Standardized anomalies are used operationally at HPC to evaluate NWP model forecasts. The anomaly data are computed from a 1948-2011 climatology derived from the NCAR/NCEP Reanalysis 1 dataset. The forecast model output is interpolated onto the 2.5 degree reanalysis grid, and a 15-day centered average for each time in the daily climatology is used in computing the anomalies. The anomalies are quantified as standard deviations (SDs) from normal. The climatological means and standard deviations are used to compute the standardized anomalies, as described in Hart and Grumm (2001), and Grumm and Hart (2001).

Standardized anomalies of moisture flux (at 850 mb and 700 mb levels) and precipitable water, as well as U and V component flux values, from the GFS and ECMWF deterministic models were created for use in ARRFEX.

B) Daily Activities

ARRFEX consisted of three separate forecasting activities, each accompanied with corresponding subjective verification. The daily schedule of activities is outlined in Appendix D. Forecasting and verification exercises took place on Monday-Thursday of both weeks, with one event analyzed per day.

For each case, a different HPC forecaster participated in the experiment (Appendix A). The events investigated each day were chosen at random from the list of the 8 events prior to the experiment; the participants did not know which particular events they would be assigned beforehand.

a) Experimental Forecasts

The focus of the forecast portion of the experiment was on three forecasting topics: (1) Day 5 and Day 3 24-h probabilistic QPFs, (2) a 72-h cumulative QPF covering Days 1-3, and (3) timing (i.e., start and end times) of precipitation associated with land falling ARs.

Forecast Task #1: Create 24 hour probability of QPFs (00 UTC to 00 UTC) at 5 day and 3 day lead times

The forecast team created two separate PQPF forecasts for a pre-determined 24 hour (00 UTC to 00 UTC) period; one at a 5 day lead time and another at a 3 day lead time. To complete this task, the team was provided operational and experimental numerical model guidance and datasets initialized at 12 UTC on 5 and 3 days prior to the forecast period. The PQPF forecasts were based on the probability of greater than 3 inches of precipitation falling during the 24 hour period of interest. Participants drew contours based on a 10% and 40% chance of exceedance (Figure 2). There was no set domain of interest for the PQPF forecast, although the overall forecast area focused on the West Coast and Intermountain West.

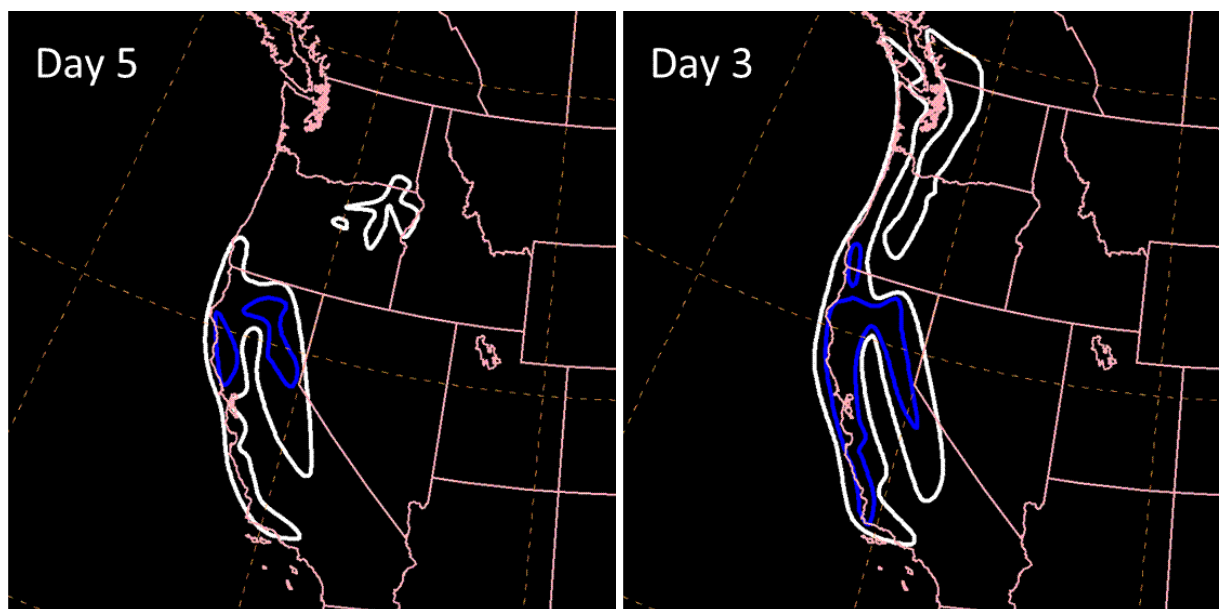


Figure 2) Experimental probability of QPF indicating the probability of >3" falling in the 24 hour period ending 00 UTC 26 October 2012 at (a) 5 and (b) 3 day lead times. The white line represents >10% probability, the blue line >40%.

Forecast Task #2: Create a Day 1-3 72 hour QPF

Participants created a 72 hour cumulative QPF for a pre-determined (00 UTC to 00 UTC) period. Forecasters drew isohyets based on expected precipitation amounts of 4", 8", >12" (Figure 3). This forecast was designed to mirror the Day 1-3 cumulative QPF generated by HPC. The forecast team was given operational and experimental guidance initialized at 12Z the day immediately before the

prescribed forecast period in order to make the available data and time requirements as realistic as possible.

Forecast Task #3: Predict precipitation duration at a specific point location

The participants forecasted the time of precipitation onset and end at a specified location (refer to Appendix B) using 6 hour windows (e.g. 00-06 UTC, 06-12 UTC, 12-18 UTC, and 18-00 UTC). When relevant, forecasts of the start/stop time of the 'heaviest precipitation' at the specific point were also created. The team was given guidance data using the same model initializations used in creating the 72 h QPF forecast.

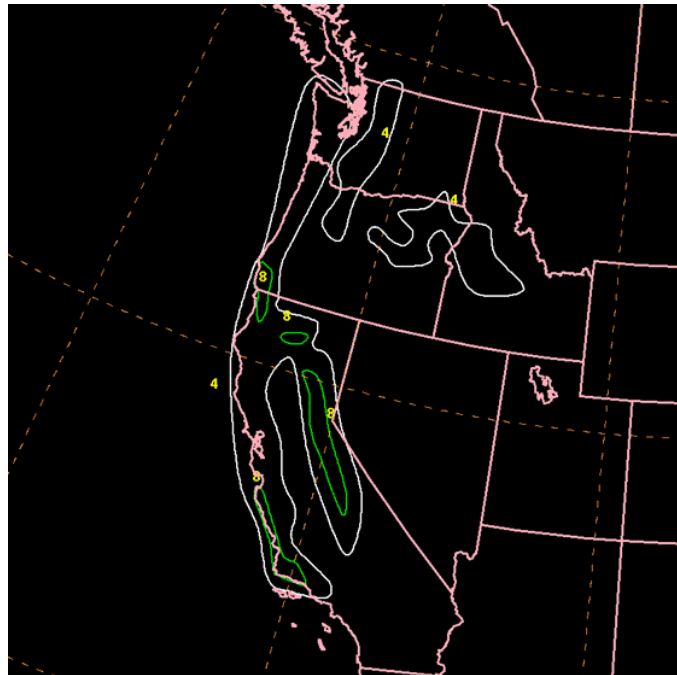


Figure 3) Experimental 72 h QPF indicating the predicted total precipitation for the 72 h period ending 00 UTC 26 October 2012. The white line represents >4" of total precipitation, the green line >8".

b) Subjective Forecast and Model Verification

Verification of Day 5 and Day 3 PQPF

Participants were asked to subjectively evaluate the performance of both their experimental forecasts and the available ensemble guidance for each case. This was done by comparing the experimental and guidance forecasts to the observed Stage IV precipitation data (displayed at 32 km) for the relevant 24 hour time period. The team was asked a series of survey questions requiring them to assign a grade to their forecast (good, fair, and poor), and compare and contrast the accuracy of the model ensemble guidance probability of exceedance forecasts.

Verification of Day 1-3 QPF

Subjective evaluation consisted comparing the experimental and model guidance forecasts against the observed Stage IV precipitation data (displayed at 4 km) for the 72 hour period of interest. Participants were asked series of questions focusing on how accurate the forecasts were in terms of locations of the precipitations maximums, as well as accuracy of the accumulated precipitation amounts. Furthermore,

the team’s experimental forecast was compared against the archived HPC Day 1-3 forecast, in order to evaluate if the addition of the experimental guidance led to an improved forecast.

Verification of AR duration forecast

Atmospheric River Observatory (ARO) data was used from the location of interest (Appendix B) as the primary form of verification to accurately identify when the precipitation began and ended at the specific location. Additionally, the participants were shown a series of 850 mb moisture flux forecasts (and corresponding standardized anomalies) from the GFS and ECMWF at 6 hour intervals overlaid with the precipitation (Stage IV) that was observed during the immediately following 6 hour period. They were then asked a series of questions as to how well the moisture flux and standardized anomaly forecasts correlated to the observed precipitation location.

3. Experiment Results

A) Probability of QPF

Participant Forecasts

Overall, participants felt that their PQPFs performed well when validated against the Stage IV observations (Figure 4). When evaluating their Day 5 PQPF, 7 (of 8) forecasts were classified as “good”, while just one was classified as “fair”, and none of the forecasts were classified as “poor.” The results were similar for the Day 3 forecasts, where 6 were classified as “good”, 2 “fair”, and none “poor.”

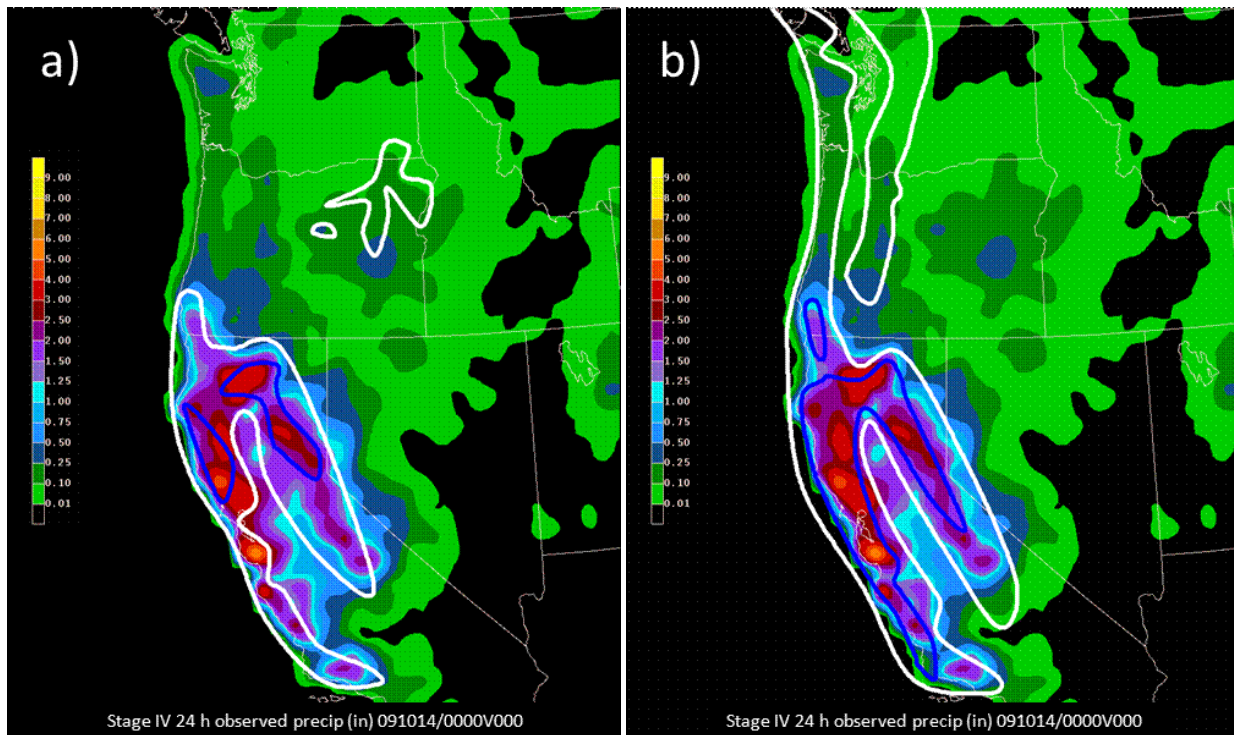


Figure 4) Experimental probability of QPF indicating the probability of >3” falling in the 24 hour period ending 00 UTC 26 October 2012 at (a) 5 day and (b) 3 day lead times overlaid on Stage IV observed precipitation.

Not all rating designations had a group consensus, as the subjective nature of how to ‘rate’ a forecast initiated discussion amongst participants. In most cases, the forecast was judged by whether or not the probability lines, particularly the 10% contour, captured all areas where greater than 3” of precipitation

was observed. If the 40% contour also captured all areas >3", this made the forecast even stronger. However, forecasts that missed >3", or missed areas of heavier precipitation that didn't quite total 3" amounts (e.g. 2-3"), with the 10% probability contour were considered inferior (especially at the 3 day lead time).

There was also diversity in the forecasts issued. Some forecasts were very spatially generous with the 10% probability contour, particularly at the 5 day lead time, while other forecasts focused on very small areal domains. This was dependent on the size of the expected precipitation field itself, but also varied by how the forecasters viewed the definition of a probability forecast. Some of the issues discussed regarding probability forecasts included:

- >Is a probability forecast based on a spatial probability, rainfall amount probability, or both?
- >Is a 10% contour worthwhile? What exactly does it tell the customer?
- >What information does the consumer get from a PQPF? Is this the same information as the forecaster is trying to convey?

The general consensus was that PQPFs provide a viable way for forecasters to address the potential for extreme events at the 5 and 3 day lead-time. Forecasters found it advantageous to use probabilities in order to convey forecast uncertainty, as well as communicate risk to consumers, without being held to specific spatial, timing or precipitation amount requirements associated with creating prototypical deterministic QPFs.

Operational and Experimental Guidance

Figure 5 displays the subjective performance for the guidance PQPF, and shows that the reforecast dataset and HMT ensemble provided consistently better guidance than the operational ensembles. The MMENS provided an upgrade over the GEFS and ECENS, which is to be expected since it is composed of all three operational ensembles (GEFS, ECENS, CMCE), but trailed behind the other experimental guidance in its ability to consistently identify areas of >3" of precipitation.

An example of the benefits provided by the HMT ensemble and reforecast datasets compared to the operational ensembles is shown in Figure 6. The higher resolution (9 km) of the HMT ensemble allowed it to focus the heavier precipitation, and therefore higher probabilities, around the areas of higher topographies. In the case shown in Figure 6, it identifies the high potential for >3" of precipitation throughout the Sierra Nevada Mountains in interior California, but also identifies the second maximum of >3" in the northern Sierra Nevada that was missed by the other operational ensembles. The higher probabilities displayed by the HMT ensemble, however, were noted to be a bit misleading; since the model only contains 7 members, it is much more feasible to achieve higher probability values than it is for the operational ensembles that contain 20 (GEFS, CMCE) and 50 (ECENS) members. Also, the higher probabilities of the HMT ensemble caused concern that the model itself may have a wet bias, which could increase the potential for false alarms in extreme precipitation events.

The reforecast dataset consistently outperformed all other guidance, being chosen as the 'most helpful' guidance in 6 of the 8 cases. While the probabilities were consistently low (mostly between 5-15%), it outperformed all model guidance in its ability to alert forecasters to areas where the heaviest precipitation could potentially fall. This can also be seen in Figure 6, as the reforecast data is the only guidance that suggests the potential for >3" of precipitation in the Sierra Madre and San Gabriel Mountains in southwestern California.

Overall, the operational guidance, particularly the GEFS and ECENS, struggled to provide helpful guidance at either lead time as their probability forecasts consistently missed areas that were later

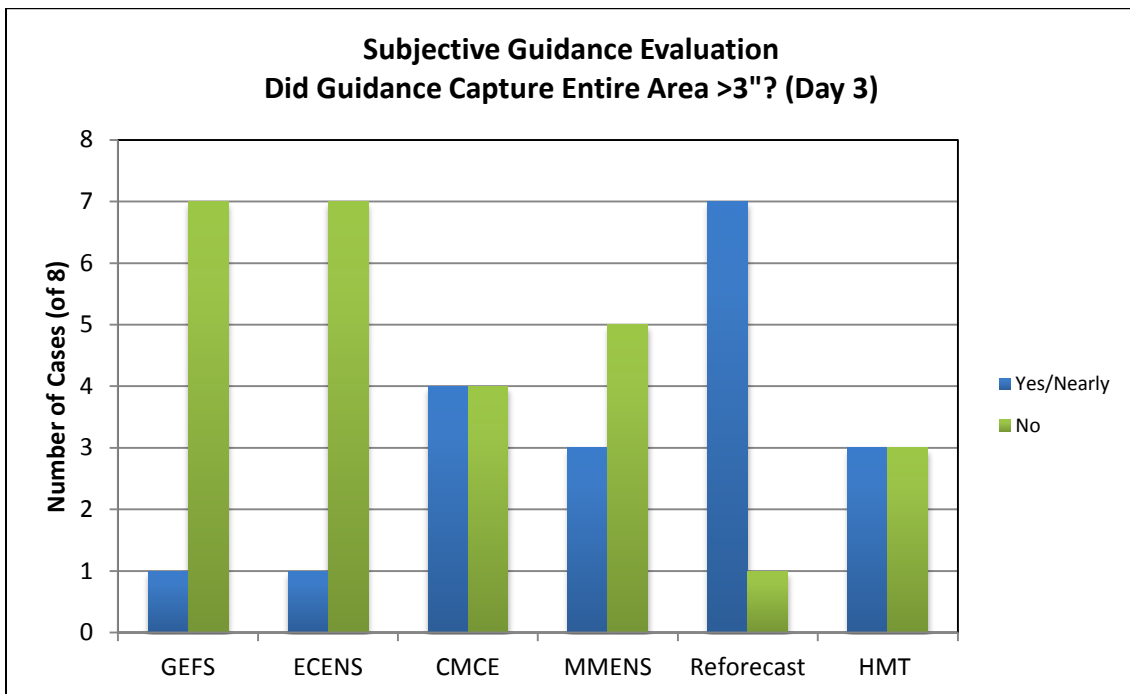
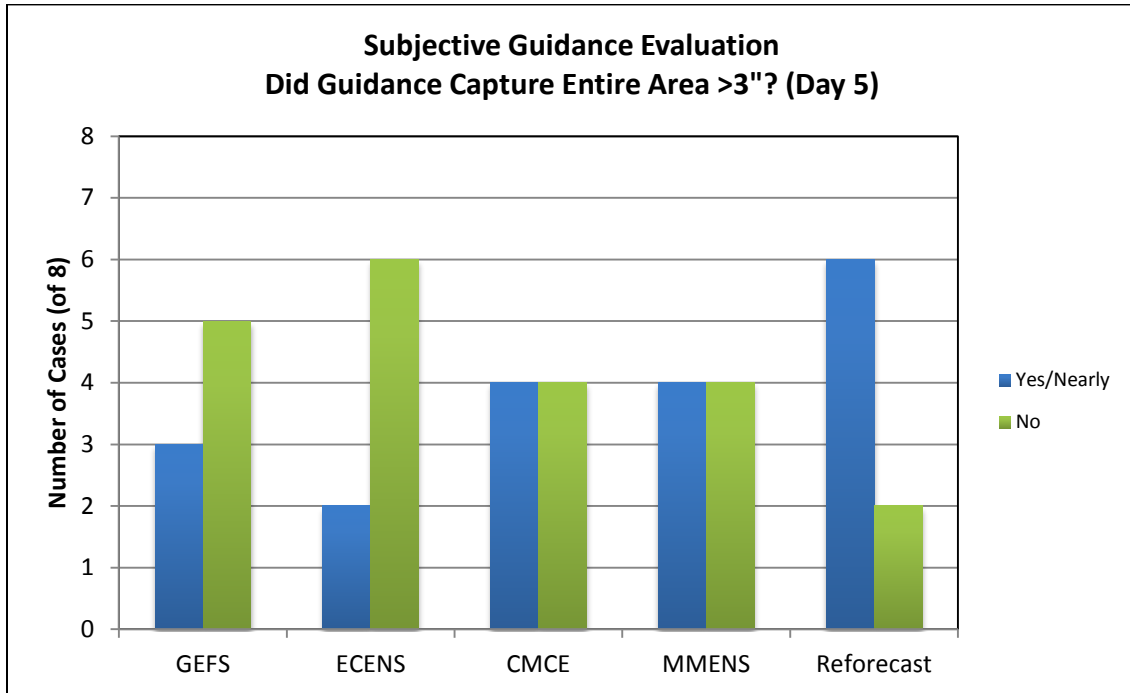


Figure 5) Ability of the ensemble guidance to forecast the area which received >3" of precipitation in a 24 hour period at a 5 day (top) and 3 day (bottom) lead time.

observed to receive >3" of precipitation (Figure 6). Also of note was the possibility of the probability forecasts degrading from a 5 day lead time to a 3 day lead time. Figure 5 shows that the GEFS and

ECENS were able to capture all/nearly all of the >3" area more often at 5 day lead time than at 3 days. Several participants noted that they had seen this forecast degradation between 5 and 3 days lead times, particularly in regards to geopotential height fields in the north Pacific and Gulf of Alaska, but were surprised at this signal in the precipitation forecast fields. The prevailing thought from the participants was that the lack of observations available to capture mid and upper level energy as it crosses the northern Pacific leads to a degradation of forecasts, which then improve as the guidance is able to ingest more observational data as systems approach the West Coast.

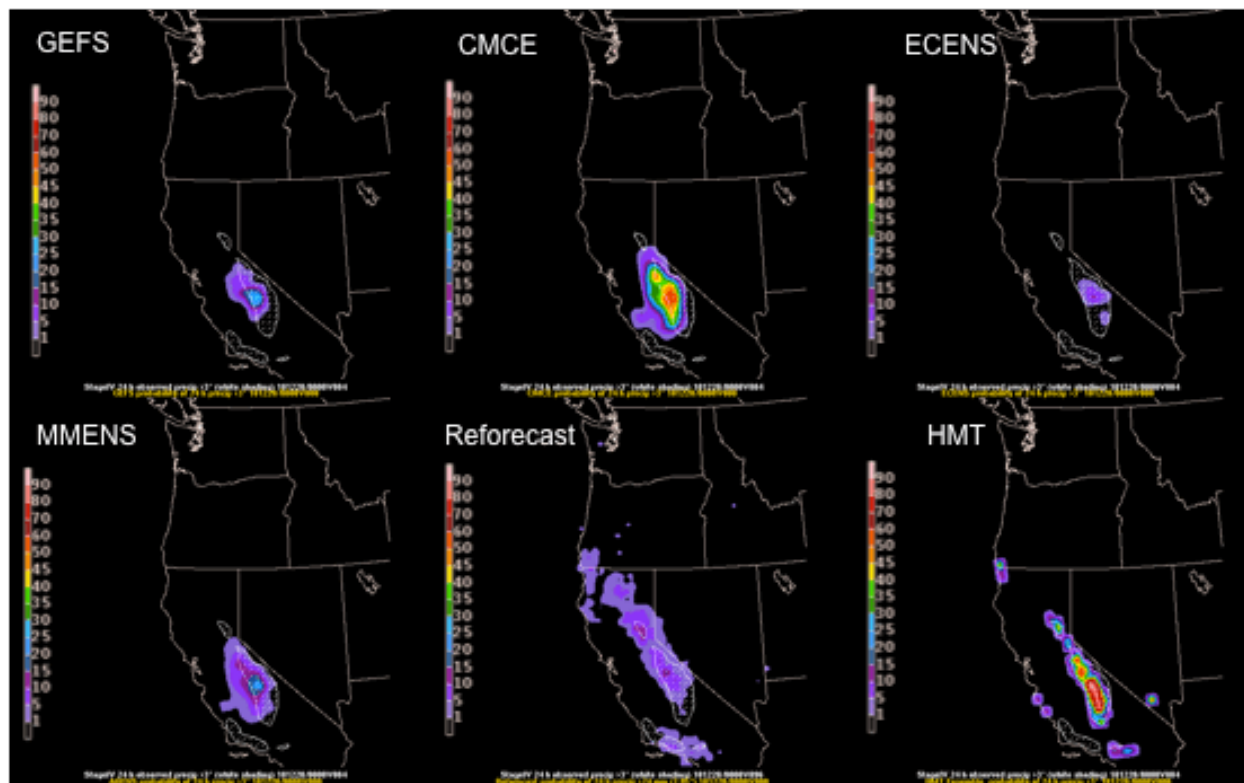


Figure 6) The probability of QPF (QPF) for >3" in 24 hours at a 3 day lead time for the GEFS, CMCE, ECENS, MMENS, Reforecast and HMT ensemble systems valid at 00 UTC 13 December 2010. The probability forecasts are overlaid with the observed area of >3" from the Stage IV data (white dashed area).

B) Day 1-3 (72 hour) QPF

Participant Forecasts

Of the 8 QPFs created in ARRFEX, 5 were subjectively rated as "good", with 2 being rated "fair" and one "poor." The result of this ranking exercise revealed that QPFs struggle with total precipitation amounts more than spatial distribution. The three forecasts that were rated as "fair" or "poor" all noted that the main axis of heavy rainfall was captured, but amounts, particularly in the areas of maximum precipitation, were under-represented. A reason for this is the difference between the high-resolution 4 km Stage IV verification data and the more general, lower resolution that the QPF contours were drawn with. However, participants acknowledged that this is a major drawback in trying to accurately forecast extreme events with prototypical QPFs.

Comparison with the archived HPC Day 1-3 QPFs (Figure 7) revealed that forecasts created with the help of the experimental guidance were generally an improvement (7 of 8 cases). The main improvement was in the QPF amounts; while the forecast maximum values were still consistently low, guidance from

the experimental datasets and tools gave forecasters enough confidence to increase forecast amounts closer to what was observed.

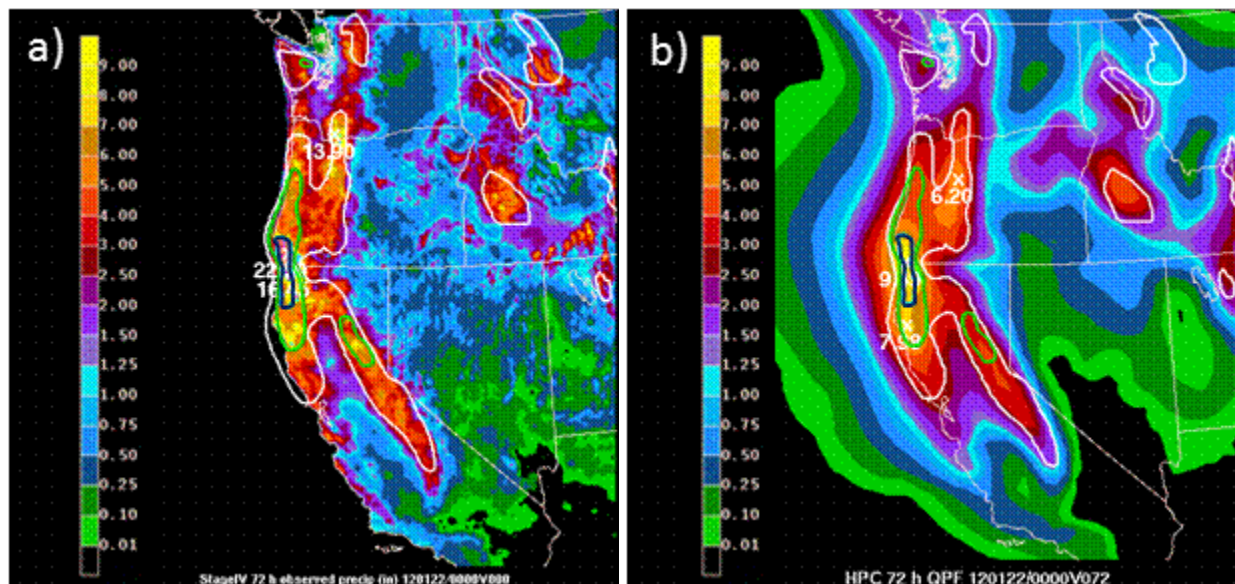


Figure 7) Experimental 72 h QPF forecast valid 00 UTC 22 January 2012 overlaid on the Stage IV 72 h observed precipitation (a) and the archived Day 1-3 HPC forecast (b).

In terms of experimental datasets and tools, the standardized anomaly fields were deemed helpful by most participants. Forecasters tended to use them to identify extreme values (relative to NARR climatology) of moisture flux and precipitable water quickly, which suggested the potential of heavy precipitation, regardless the model QPF guidance. The higher resolution of the HMT ensemble produced QPFs that were more aggressive with amounts and more spatially refined to the topography than the lower resolution operational models. This resulted in what the participants often deemed a more “realistic” looking model forecast. This can be seen in Figure 8, as the HMT ensemble (Figure 8b) identifies an area of extreme precipitation >15” associated with Klamath Mountains along the northern California and southern Oregon coast that is in good agreement with the Stage IV observations (Figure 8a). The deterministic GFS and ECMWF (Figures 8d and 8e) hint at a local precipitation maximum in that location, but under-predict the amounts. Despite the apparent benefits of the HMT ensemble, however, there was continued concern of a wet bias as it consistently produced considerably higher precipitation amounts, despite being a mean ensemble value. Preliminary examination into individual member QPF during the experiment revealed that, depending on the case, there could be noticeable differences in max QPF amounts and location between members; this is hypothesized to be due to the different physics schemes implemented, but further evaluation is needed.

Figure 9 shows that the operational guidance struggled to produced quality 72 hour QPFs. The deterministic QPFs from the operational GFS and ECWFM were consistently rated as only “fair” or “poor” forecasts. Their total precipitation amounts were significantly low, in some cases as much as 10-12” below what was observed. While their coarse resolution is an explanation for their struggles to produce extreme precipitation amounts and adapt the precipitation to the topography, participants were often disappointed at the quality of guidance from the ECWFM and GFS in both amounts and location. It should be noted that both deterministic versions of the models used in ARRFEX were displayed at 1° resolution, which is coarser than what is currently used operationally.

While participants were often disappointed at that quality of the ECWMF and GFS guidance, the NAM (32km) performed well, consistently providing “fair” and “good” quality forecasts (Figure 8c, Figure 9).

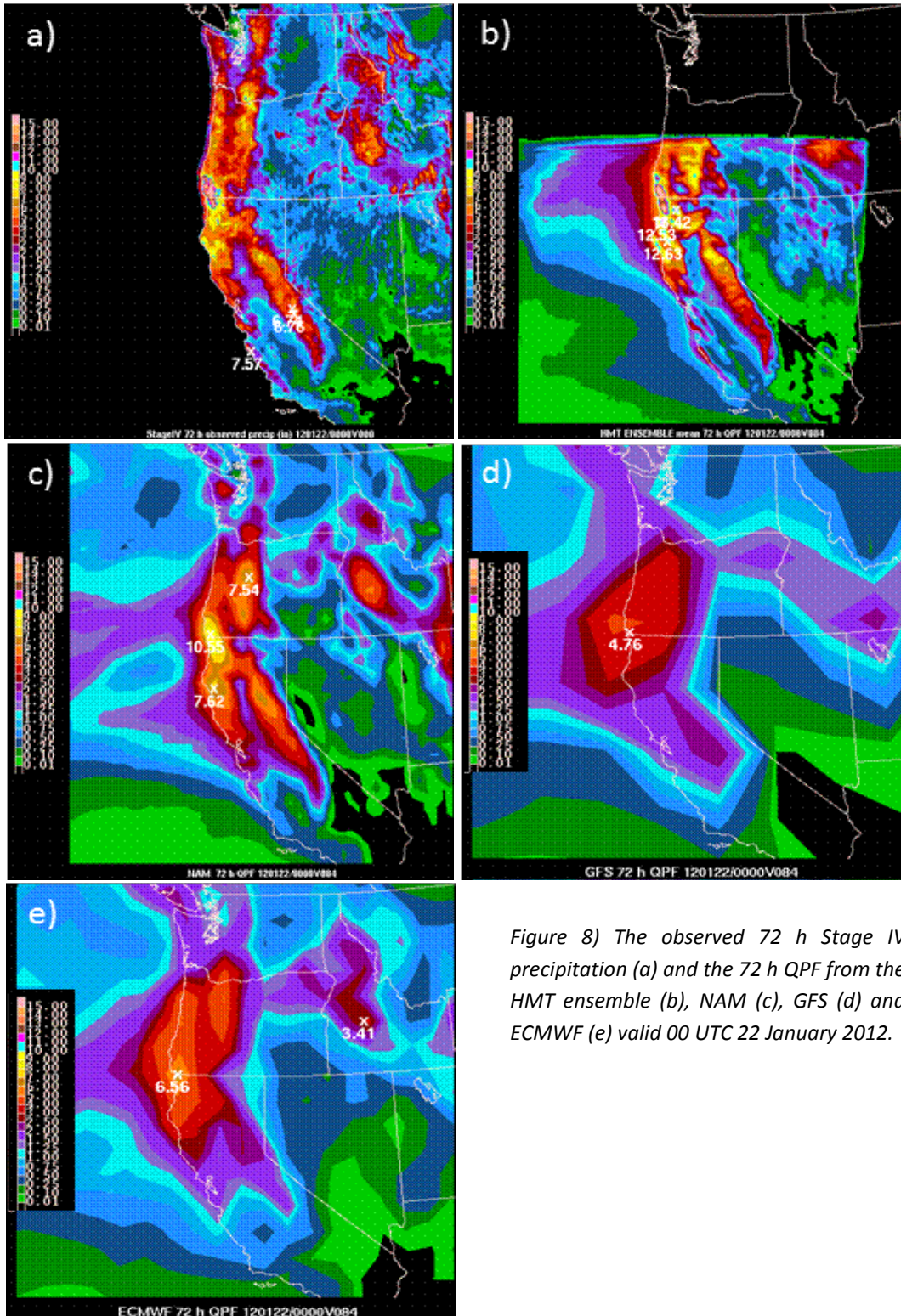


Figure 8) The observed 72 h Stage IV precipitation (a) and the 72 h QPF from the HMT ensemble (b), NAM (c), GFS (d) and ECMWF (e) valid 00 UTC 22 January 2012.

The increased resolution allowed the NAM to accurately represent precipitation maximums in favored topographical locations, as well as provide higher total QPFs. Additionally, the expertise of the local topography and climatology of some participants was also extremely valuable in adjusting QPF amounts and locations, highlighting how valuable small-scale and topographical details are in AR forecasting.

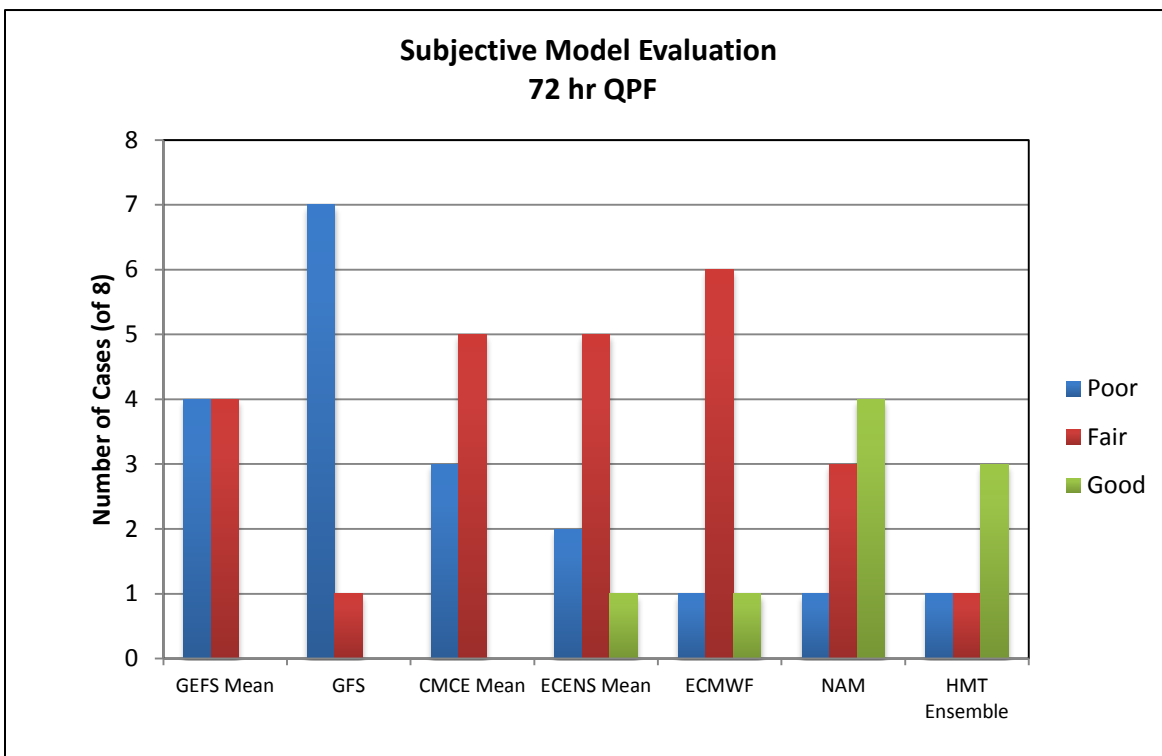


Figure 9) Subjective model performance of the 72 h QPF for each of the guidance systems used in ARFEX.

C) AR Duration Forecasts

Forecasts of AR-induced precipitation start and stop times revealed that model guidance struggles to accurately depict AR timing and duration. Of the 7 cases where forecasters were asked to predict the start and stop time of the precipitation in a 6-hour window (refer to section 2B), two forecasts were able to correctly depict the start time, and only one was able to correctly depict the stop time. These results were not correlated to the forecaster’s confidence in their forecast; 5 cases resulted in forecaster’s having “medium” confidence and two cases had “high” confidence in the start and stop forecasts, respectively. There were no forecasts of “low” confidence. However only one “high” confidence forecast (a forecast of precipitation start) was successful. Forecaster confidence was mostly correlated to model consensus, as forecasters had higher confidence if several models agreed on the start/stop timing.

Investigation into the applicability of using GFS and ECMWF forecasts of 850 mb moisture flux and associated standardized anomalies to identify locations of precipitation maximums throughout AR events showed promise. Participant’s felt that the standardized anomalies were a little easier to use operationally, and their magnitudes were more consistent in correlating with heavy precipitation. Further investigation into model forecasts of moisture parameters at differing atmospheric levels was encouraged.

4. Summary

A) Experiment Findings and Operational Impacts

The 2012 Atmospheric River Retrospective Forecasting Experiment (ARRFEX) was conducted September 17-28. The experiment focused on evaluating operational and experimental datasets to forecast atmospheric river (AR) induced extreme precipitation events along the West Coast, as well as diagnosing ways to better provide information to consumers at mid-range timeframes. The experimental datasets featured in ARRFEX were all found to provide value in AR forecast process, although to varying levels of consistency (Figure 10). The results of the project are summarized in the following:

- The current operational global guidance (both deterministic and ensemble) struggles in AR events, consistently showing a low bias in their QPF maxima. The coarser resolution of the global models limits their ability to resolve topography-driven precipitation with the desired detail, in particular causing them to smooth over small-scale shifts in wind direction and localized areas of higher topography. As a result, they consistently under-produce precipitation. However, the higher resolution operational models, such as the NAM, do provide benefits in AR cases. Their ability to resolve topography helps identify areas favorable for precipitation maximums, although high-resolution models carry a known wet bias, so forecasts must be used accordingly.
- Knowledge of local topography, climate and seasonal precipitation regimes along the West Coast is vital in AR forecasting. Interaction and discussion among participants was vital to creating successful experimental forecasts.
- Model forecasts of moisture parameters may be helpful in identifying the potential for extreme events, even when the model QPF does not forecast large precipitation amounts. However, the accuracy of model forecasts of moisture parameters in AR events (e.g. moisture flux, precipitable water) needs to be further investigated for potential model biases.
- PQPFs appear to be a valid way to aid in forecasting heavy precipitation events at mid-range lead times. Forecasters appreciated that they provide a platform to express forecast uncertainty, while still conveying the risk of heavy rainfall events. HPC currently utilizes probability forecasts for their Winter Weather and Excessive Rainfall products, but additional value may lie at the WFO and RFC levels in communicating risk with emergency management officials at longer forecast lead times.
- Forecasting AR duration is problematic. Models often err in forecasting the start/stop of precipitation, especially in situations where the precipitation is topographically forced or enhanced. Model consensus did not correlate well to accurate timing forecasts.
- Participants felt the HMT Ensemble provided helpful guidance, when applicable. The 9 km resolution allowed the ensemble system to capture the topography, and therefore the topographically driven/enhanced precipitation, in a way that forecasters trusted. However, there was a concern that the model may contain a wet bias, given its high member and mean precipitation forecasts.

- The Multi-Member Ensemble system showed promise, but is limited by the members which are used to construct it. While forecasters like the idea of showing a ‘true’ probability and the ease of seeing ensembles combined and displayed at once, its application in extreme precipitation forecasting is limited due to the fact the global ensembles struggle to produce accurate QPFs in these cases.
- ESRL’s Reforecast Dataset was widely considered the most helpful experimental guidance featured in ARRFEX (Figure 10). Participants liked the fact that the guidance was created from actual observations, therefore eliminating any potential model biases. Integration of the reforecast dataset into HPC operations is considered high priority, and HPC is currently working with ESRL on gaining real-time access to the guidance.
- Standardized anomalies were also considered to provide helpful guidance to forecasters in extreme events. Already operational at HPC, future efforts will involve investigating the use of the model climate, in partner with the reanalysis climate currently used to create the anomalies.

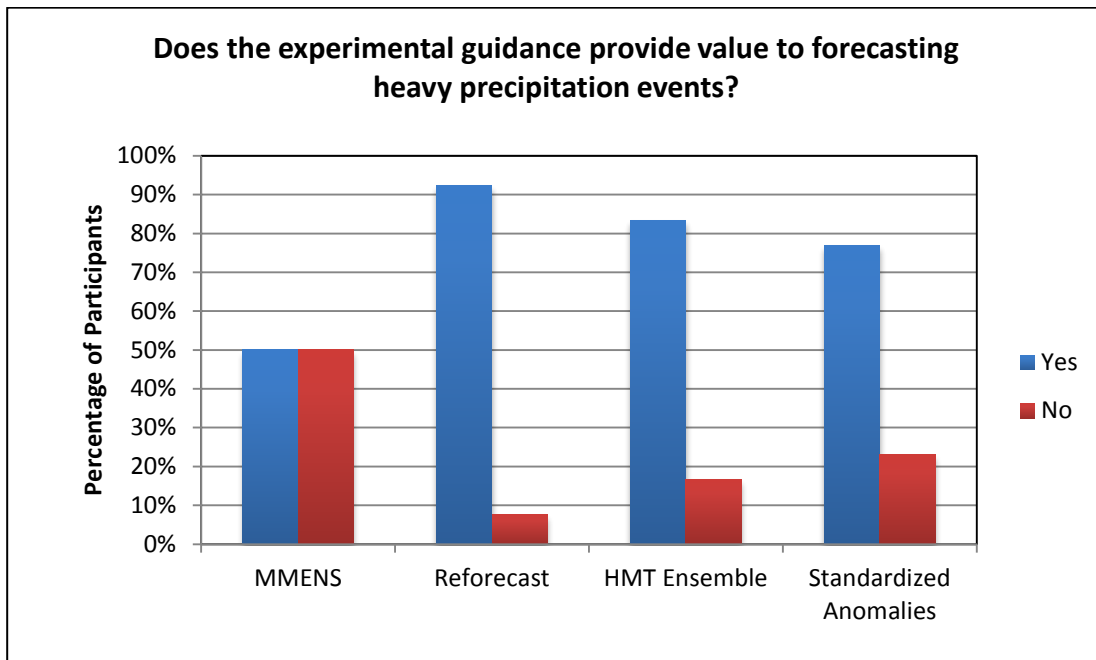


Figure 10) Participant feedback on the use and value of the experimental guidance systems featured in ARRFEX.

B) Ongoing and Future Work

HPC and ESRL are continuing to test and explore the forecast methods and guidance datasets. Some of the ongoing work and projects include:

- HPC and ESRL will continue to conduct subjective and objective analysis of the model forecasts for the ARRFEX cases, including a focus on moisture parameters (e.g. moisture fluxes).
- HPC will continue to apply and test high-resolution guidance, both experimentally and operationally, with an increased application in West Coast heavy precipitation events.

- Investigate potential correlation between model moisture flux forecasts and observed precipitation in order to evaluate the use of moisture flux as an indicator of precipitation amounts and location.
- ESRL and HPC will conduct objective analysis of the HMT ensemble precipitation fields (both the mean and individual members) to identify any potential bias from the ARRFEX cases; future work will expand to real-time application and analysis. ESRL is continuing to develop the system, with plans to expand to a North American domain and run in real-time.
- Additional future work will continue to objectively and subjectively analyze the model guidance from the ARRFEX cases, as well as examine the events from a climate perspective; in particular investigate potential relationships between the AR events and climate indices such as the El Nino Southern Oscillation (ENSO) and Madden-Julian Oscillation (MJO).

Acknowledgments

The 2012 Atmospheric River Retrospective Forecasting Experiment was the product of the ideas, collaboration and effort of a host of individuals from HPC and ESRL, including Tom Workoff (HPC), Ellen Sukovich (ESRL), Mike Bodner (HPC), Ben Moore (ESRL), Faye Barthold (HPC), David Novak (HPC) and Marty Ralph (ESRL). Mike Bodner was responsible for the creation of the standardized anomaly fields, while Tom Hamill (ESRL) provided the 2nd generation reforecast dataset and Ben Moore created the reforecast guidance used in the experiment. Brian Etherton and Linda Wharton (ESRL) provided the archived HMT ensemble guidance that was featured.

References

Grumm, R.H., and R. Hart, 2001: Standardized anomalies applied to significant cold season weather events: Preliminary findings. *Wea. Forecasting*, **16**, 736-754.

Hamill, Thomas M., Jeffrey S. Whitaker, 2006: Probabilistic Quantitative Precipitation Forecasts Based on Reforecast Analogs: Theory and Application. *Mon. Wea. Rev.*, **134**, 3209–3229.

Hamill, Thomas M., 2012: Verification of TIGGE Multimodel and ECMWF Reforecast-Calibrated Probabilistic Precipitation Forecasts over the Contiguous United States*. *Mon. Wea. Rev.*, **140**, 2232–2252.

Hart, R., and R.H. Grumm, 2001: Using climatological anomalies to rank synoptic scale events objectively. *Mon. Wea. Rev.*, **129**, 2426-2442.

Appendix A) ARRFEX Participants

	HPC	ESRL	WFO/RFC	EMC	Academia
Sept 17-20	Mike Musher (9/17) Brendon Rubin-Oster (9/18) Patrick Burke (9/19) David Roth (9/20)	Ellen Sukovich	Mel Nordquist (EKA) Tom Wright (MFR)	Kate Howard	----
Sept 24-27	Kenny James (9/24) Rich Otto (9/25) Bob Oravec (9/26) Dan Peterson (9/27)	Ellen Sukovich Ben Moore	Victor Stegemiller (NWRFC)	Yan Luo (9/24-25) Jun Du (9/26-27)	Jon Rutz (Univ. of Utah)

Appendix B) ARRFEX Cases and Point Forecast (ARO) Locations

ARRFEX Cases

Event	Dates of Event	Dates of 24 h PQPF	Dates of 72 h QPF	Initialization #1	Initialization #2	Initialization #3	Point Forecast Location
4	13-14 Oct 2009	13-14	13-16	10/08	10/10	10/12	CZC
7	17-23 Jan 2010	19-20	17-20	01/12	01/14	01/16	CZC
1	23-25 Oct 2010	25-26	23-26	10/18	10/20	10/22	CZC
5	10-14 Dec 2010	12-13	10-13	12/05	12/07	12/09	WPT
2	16-23 Dec 2010	19-20	17-20	12/12	12/14	12/16	PPB
8	15-19 Jan 2011	16-17	16-19	1/11	1/13	1/15	WPT
3	18-26 Mar 2011	19-20	19-22	3/14	3/16	3/18	TPK
6	14-20 Jan 2012	21-22	19-22	1/14	1/16	1/18	WTP

Atmospheric River Observatory Locations

Abbreviation	Location	Elevation
CZC	Cazadero, CA	475 m
WPT	Westport, WA	5 m
PPB	Pt. Piedras Blancas, CA	11 m
TPK	Three Peaks, CA	1021 m

Appendix C) HMT Ensemble Information

Member	Core	Moist Physics	Boundary Conditions
0	ARW	Thompson	GEFS member 1
1	ARW	Ferrier	GEFS member 1
2	ARW	Schultz	GEFS member 2
3	ARW	Thompson	GEFS member 3
4	ARW	Ferrier	GEFS member 5
5	ARW	Schultz	GEFS member 6
6	ARW	Thompson	GEFS member 7

Appendix D) Daily Schedule

8:00 am (Monday only) – Orientation

8:30-10:00 am – Create 24-h PQPFs (00 UTC to 00 UTC) for 5 day and 3 day lead times

10:15-11:30 am – Create Day 1-3 72-hour QPF

Optional: 11:00am – HPC Map Discussion

11:30-12:30 pm – Lunch

12:30-1:30 pm – Verify 24-h PQPFs (00 UTC to 00 UTC) for 5 day and 3 day lead times

1:45-2:30 pm – Verify 72-hour QPF

2:45-4:15 pm – Make and verify precipitation duration forecast at specific locations

4:15-4:30 pm – Group discussion and/or exit questions