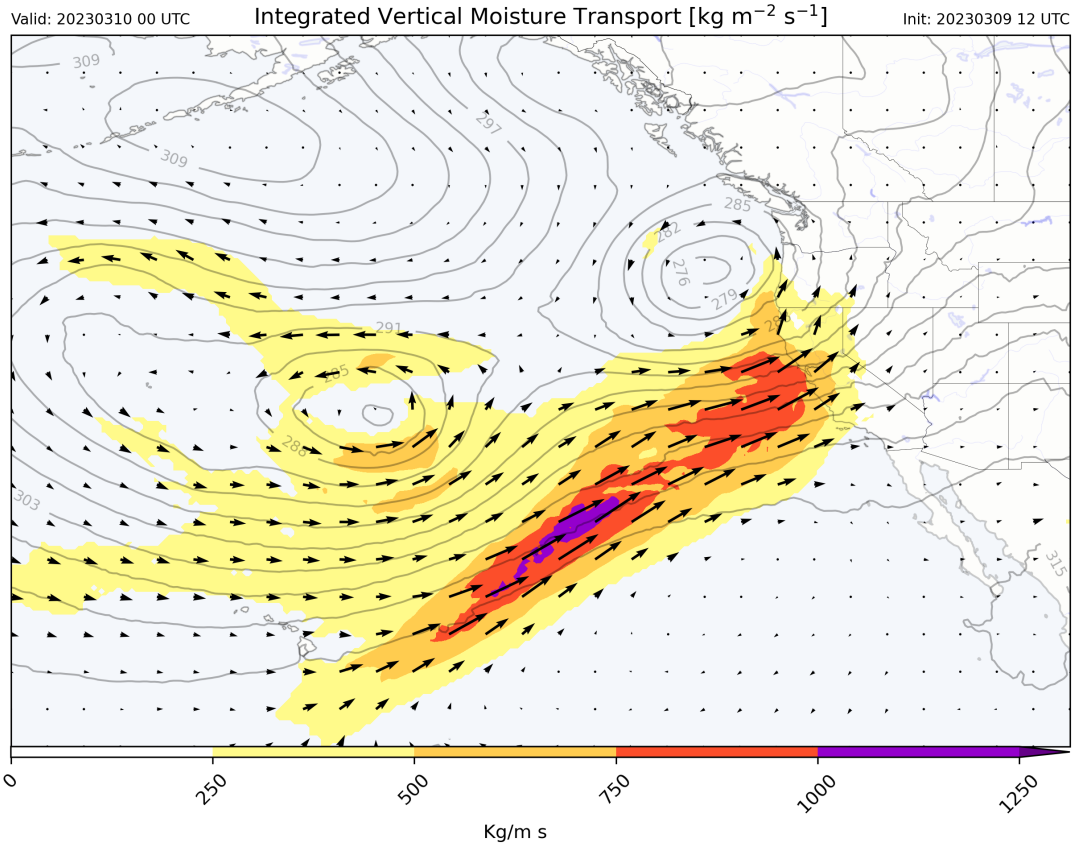


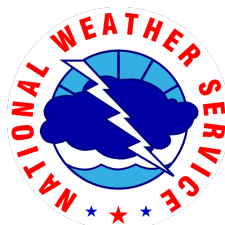


2024 Precipitation Experiment for Atmospheric Rivers (PEAR) Final Report: *Results and Findings*

October 29 - November 14, 2024
Weather Prediction Center (WPC)
HydroMeteorology Testbed (HMT)

Tomer Burg¹, James Correia Jr.¹, and Sarah Trojniak¹

¹CIRES-CIESRDS CU Boulder, NOAA/NWS/WPC/HMT



Contents

1	Introduction	2
2	Science and Operations	3
2.1	Daily Schedule	4
2.2	Case Selection	5
2.3	Data Overview	6
2.4	Science Questions and Goals	7
2.5	Focus Groups	8
2.6	Forecasting Activities	8
2.7	Verification Methods	11
3	Results	12
3.1	Focus Group	12
3.1.1	AR Forecasting by Lead Time	12
3.1.2	Model and Tool Utility	14
3.2	Forecast Activities	18
3.2.1	Case 1	18
3.2.2	Case 2	23
3.2.3	Additional discussion	28
4	Summary and Conclusions	30
4.1	Utility of AR-AFS	31
4.2	Activity Feedback and Future Objectives	35

1 Introduction

The Hydrometeorology Testbed (HMT) at the Weather Prediction Center (WPC) received support to develop and conduct a new forecast experiment for Atmospheric Rivers (ARs) beginning in 2024. While a one-off WPC forecast experiment for ARs occurred in 2012, this renewed support for an AR experiment comes as part of the Water in the West initiative, which includes multiple convection allowing models (CAMs) designed to predict ARs at medium range lead times.

As defined by the American Meteorological Society glossary, an AR is defined as “a long, narrow and transient corridor of strong horizontal water vapor transport that is typically associated with a low-level jet stream ahead of the cold front of an extratropical cyclone” (Ralph et al., 2019). Figure 1 shows a conceptual schematic of a typical AR derived from reconnaissance measurements and atmospheric reanalyses. ARs are sometimes but not always associated with a tropical moisture source, and frequently produce heavy precipitation in regions of orographic or warm conveyor belt-induced ascent (Ralph et al., 2019). On average, ARs generate 30–50% of the annual precipitation in the western United States (U.S.) (Lamjiri et al., 2017). Due to their major socioeconomic impacts, there is a vested interest to improve forecasts and communication of AR associated impacts (Ralph et al., 2020).

This new forecast experiment, the Precipitation Experiment for Atmospheric Rivers (PEAR), began with a two-week session in the fall of 2024. Building off of the template of existing forecast experiments at HMT, PEAR brought together forecasters, researchers, and model developers from across the country. The primary objectives of this first experiment were to better understand the primary forecast challenges for ARs at different lead times, learn more about the models and parameters forecasters use operationally, subjectively evaluate experimental models for two retrospective case studies from the 2022-2023 AR season in the West Coast, and gather feedback on the usefulness of the experimental models and the forecast activities to inform future PEAR activities and model development.

PEAR was in session for the following weeks:

Week 1: October 29 - 31 (virtual)

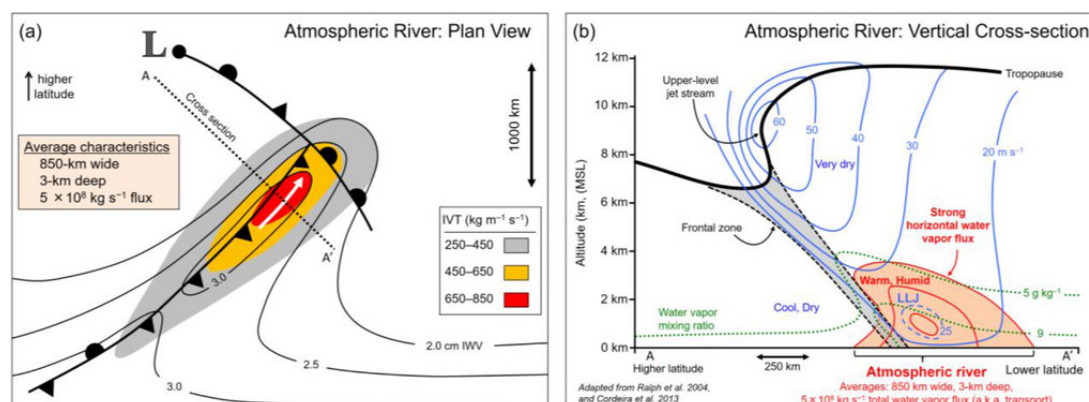


Figure 1: Schematic summary of the structure and strength of an AR based on dropsonde measurements analyzed in this study, and on corresponding reanalyses that provide the plan-view context. (a) Plan view including parent low-pressure system and associated cold, warm, stationary, and warm-occluded surface fronts. IVT is shown by color fill (magnitude; $\text{kg m}^{-1} \text{s}^{-1}$) and direction in the core (white arrow). IWV (cm) is contoured. A representative length scale is shown. The position of the cross section shown in (b) is denoted by the dashed line A-A'. (b) Vertical cross-section perspective, including the core of the water vapor transport in the AR (orange contours and color fill) and the pre-cold-frontal LLJ, in the context of the jet-front system and tropopause. Water vapor mixing ratio (green dotted lines; g kg^{-1}) and cross-section-normal isotachs (blue contours; m s^{-1}) are shown. Magnitudes of variables represent an average midlatitude AR with lateral boundaries defined using the IVT threshold of $250 \text{ kg m}^{-1} \text{s}^{-1}$. Depth corresponds to the altitude below which 75% of IVT occurs. Caption and figure adapted from Ralph et al. (2004).

2 Science and Operations

There were approximately 50 participants during the first PEAR. The participant makeup consisted of 23 Weather Forecast Office (WFO) participants, primarily based in the western U.S. and Alaska. Additional participation came from partners at the Environmental Modeling Center (EMC), the Physical Science Laboratory (PSL), and the Center for Western Weather and Water Extremes (CW3E). The full PEAR participant list can be found in Figure 1.

Week	WFO/RFC	WPC Forecasters	NOAA / Labs	DTB Helpers	Others
Week 1	Daniel Johnston – WFO AFC Sean Jones – WFO AJK Jeremy Buckles – WFO MRX Ed Townsend – WFO PDT Travis Wyatt – WFO PIH Brittany Whitlam – WFO REV Chris Smallcomb – WFO REV Anna Schneider – RFC RSA Brent Bower – WFO SEW Maddie Kristell – WFO SEW Tom Dang – WFO TWC Bill Leatham – RFC NE Neal Strauss – RFC NE	Allison Santorelli – WPC Scott Kleebauer – WPC	Keqin Wu – EMC Vijay Tallapragada – EMC Xingren Wu – EMC Benjamin Moore – PSL Kelly Mahoney – PSL Leif Swenson – PSL D. Alex Burrows – EPIC Chad Hecht – CW3E Matthew Simpson – CW3E Nora Mascioli – CW3E	Chris Smith – DTB Noah Brauer – DTB	Ben Pritchett – Avalanche Robert Hahn – Avalanche Jamekia Pritchard – NWC
Week 2	Virginia Rux – WFO AFC Aaron Jacobs – WFO AJK Nick Morgan – WFO AJK Matthew Kidwell – WFO EKA Zahaira Velez – WFO EKA Lamont Bain – WFO MTR Chris Burling – WFO PQR Colby Neuman – WFO PQR Brian Bong – WFO REV Jeremy Michael – WFO RLX Peyton Camden – WFO SGF Brandt Maxwell – WFO SGX	Jennifer Tate – WPC Josh Weiss – WPC	Matthew Simpson – CW3E (repeat) Rachel Weihs – CW3E Sam Bartlett – CW3E	Austin Coleman – DTB	Jason Konigsberg – Avalanche Jeff Davis – Avalanche Brian Hudec – Student (UMD)

Figure 2: List of the participants for both weeks of the fall 2024 PEAR experiment.

2.1 Daily Schedule

The schedule was largely consistent between weeks 1 and 2. Day 1 consisted of two focus group activities, one assessing AR forecasting by lead time and the second assessing models and model parameters used to forecast ARs. Day 2 consisted of text- and drawing-based forecast activities for a single case study evaluated at day 5, day 3, and day 1 forecast lead times. Before the forecast activity segment for each lead time, a pseudo-operational forecast briefing using the Global Forecast System (GFS) and experimental model data available at that lead time was provided by a WPC meteorologist. Additionally, satellite-derived layer precipitable water (PWAT) plots were made available for forecast briefings courtesy of the Cooperative Institute for Research in the Atmosphere (CIARA). Finally, day 3 began with verification of the previous day’s forecast activities, and a discussion on the utility of CAMs for ARs at multiple lead times for the case study and feedback for future PEAR experiments.

In both weeks, Marty Ralph from CW3E provided welcoming remarks with a motivation for the importance of AR forecasting challenges. On the second week, Julie Demuth and Andrea Schumacher from the National Center for Atmospheric Research (NCAR) provided a social science seminar on longitudinal surveys for how the public receives and responds to information about AR forecasts. As many of the participants were located in the western U.S. or Alaska, and the PEAR

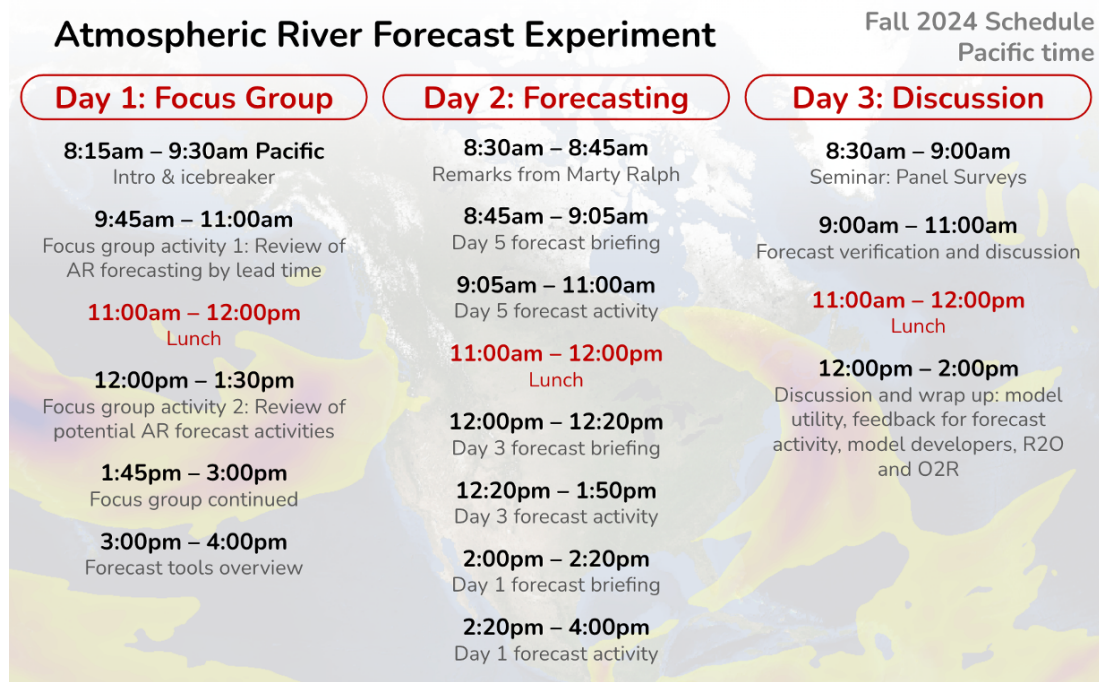


Figure 3: Daily PEAR schedule from week 2, in Pacific time.

facilitators and WPC forecasters were located in the eastern U.S., a compromise base time zone of Mountain time was chosen for the schedule.

2.2 Case Selection

Two retrospective case studies were selected for forecast activities for the first PEAR experiment:

Week 1: 1200 UTC 9 January – 1200 UTC 10 January 2023

Week 2: 1200 UTC 9 March – 1200 UTC 10 March 2023

These cases were selected from the available model simulations from the retrospective 2022–2023 AR season. Both cases predominantly affected California and the Pacific Northwest with the heaviest precipitation. The 9–10 January 2023 case was a transient but strong AR in the middle of a sequence of multiple AR events, associated with a strong zonal jet stream across the North Pacific Ocean. The 9–10 March 2023 case was a longer duration AR than the prior case, was preceded by dry conditions in the western U.S., and driven by the complex interaction of

two cutoff troughs in the Northeast Pacific equatorward of an upper-tropospheric block situated over Alaska. Both cases were associated with medium-range forecast uncertainty in the AR location, orientation, and magnitude, with uncertainty originating from the North Pacific.

2.3 Data Overview

Model guidance made available for PEAR participants included the (GFS) and three CAMs. The GFS is a global model run at 13 km horizontal resolution, with output provided at 0.25°grid spacing. Figure 4 has more details on the model configurations.

Model	Source	Horizontal Grid Spacing	Model Domain	Forecast Lead Time	Physics
GFS	NCEP/EMC (Operational)	13 km	Global	360 h	PBL: Scale-aware TKE-EDMF MP: GFDL Convection: Scale-Aware SAS
AR-AFS	NCEP/EMC	3 km	Limited Area	120 h	PBL: GFS EDMF MP: Thompson Convection: Scale-Aware Mass Flux (SAMF)
UFS-AR	GSL	3 km	Nested	60 h	PBL: MYNN MP: Thompson-Eidhammer Parameterization for Operations (TEMPO2) Cumulus: MYNN (shallow) C3 (deep)
West-WRF	CW3E	3 km inner nest 9 km outer nest	Nested	120 h inner nest 240 h outer nest	PBL: YSU MP: Thompson Convection: Grell-Freitas Ensemble

Figure 4: Information on models used for PEAR.

The primary model evaluated during PEAR was the Atmospheric River Analysis and Forecast System (AR-AFS), provided by the Environmental Modeling Center (EMC). AR-AFS is a limited-area model which inherits initial conditions (ICs) and lateral boundary conditions (LBCs) from GFS, uses the FV3 dynamical core, and is run out to 120 h forecast lead time at 3 km horizontal resolution.

The Atmospheric River Forecast System (UFS-AR) model was used for PEAR, courtesy of the Global Systems Laboratory (GSL). Unlike AR-AFS, UFS-AR is a nested model. UFS-AR is run out to 60 h forecast lead time at 3 km horizontal resolution. UFS-AR data was limited to only 6 initializations, so each PEAR case only has one UFS-AR cycle data available at day 1 forecast lead time.

Lastly, the West-WRF model from CW3E was provided to participants for reference as well. West-WRF is a nested model with an inner domain at 3 km horizontal resolution run out to 120 h forecast lead time, and an outer domain at 9 km horizontal resolution run out to 240 h forecast lead time.

For quantitative and subjective precipitation verification of participant forecasts, the Multi-Radar/Multi-Sensor Gauge Corrected (hereafter MRMS) Quantitative Precipitation Estimate (QPE) product was used, provided by the National Severe Storms Laboratory (NSSL).

2.4 Science Questions and Goals

The science objectives for PEAR relating to models evaluated were constrained to the 2 case studies used for the experiment.

The science objectives for PEAR are as follows:

- Understand the forecasters' primary forecast objectives, communication to partners, and challenges with ARs as a function of forecast lead time
- Understand the models and model variables forecasters use to forecast ARs
- Evaluate the utility of the AR-AFS in forecasting precipitation extremes, AR intensity, and location relative to the GFS from 1 through 5 day forecast lead times
- Compare the AR-AFS to other similar convection-allowing models, West-WRF and UFS-AR
- Understand the quantitative and subjective importance of IVT and defining an AR for forecasters

The first two items above were addressed through focus group activities. The subsequent objectives were addressed through forecast activities and group discussions.

Relating to model output and case study evaluation, an error in the calculation code for integrated vapor transport (IVT) rendered results from week 1 relating to

IVT magnitude and AR intensity unusable. This error was fixed prior to week 2 activities.

2.5 Focus Groups

Two focus group activities were conducted as part of the first day of PEAR, the first assessing AR forecasting by lead time, and the second assessing models and model parameters used to forecast ARs. Based on participant feedback, answers and discussions from week 1, in addition to slight differences in the daily schedule and time spent on each question between both weeks, the set of questions was slightly different for week 2. Fig. 5 shows the set of questions asked for both focus group activities for both weeks.

The questions for the first focus group were designed to understand how participants approach forecasting AR by lead time, specifically focusing on impacts and communication. For all poll questions, options of 1, 2, 3, 4, 5, 6, 8, and 10 day lead times were presented on Google Meet. After a few minutes for participants to answer the questions, the poll results were displayed anonymously to all participants, followed by an open discussion and some follow-up questions.

The questions for the second focus group were designed to better understand the parameters participants look at when forecasting AR impacts, and how they use global models and CAMs in their forecast process. For the first subset of poll questions, participants were provided with statements about using different variables and models for forecasting, such as “I care about IVT when forecasting ARs”, and presented with six options for each: strongly disagree, disagree, slightly disagree, slightly agree, agree, and strongly agree. The final poll question posed a hypothetical statement about using a deterministic CAM at day 6 lead time in the forecast process, as the AR-AFS does not extend beyond day 5 lead time.

2.6 Forecasting Activities

As this was the first PEAR experiment, all of the forecast activities used for PEAR were experimental in nature. One of the objectives of these activities was to keep them relatively open-ended to interpretation by the participants to

Format	Focus Group Activity 1	Format	Focus Group Activity 2
Poll	I get concerned about an AR event at day ____ lead time	Poll	I care about IVT when forecasting ARs
Poll	I communicate about an AR event at day ____ lead time	Poll	(WEEK 2 ONLY) I care about PWAT when forecasting ARs
Survey	To whom do you communicate an AR forecast? When do you communicate the forecast?	Poll	I care about the AR orientation/axis at landfall
Poll	I think about WFO-level impactful QPF at day ____ lead time	Poll	I care about the heaviest precipitation period during the AR event(s)
Poll	I think about specific location QPF at day ____ lead time	Poll	(WEEK 2 ONLY) I care about the snow level when making a forecast
Poll	I think about precipitation onset timing at day ____ lead time	Survey	What precipitation duration (heaviest, by 6, 12, 18, 24h+) and precipitation rate thresholds (inches per hour? Inches per 6h?) do you care about for impacts?
Survey	How important is precipitation onset timing to communicating impacts? Do you differentiate between precipitation onset and AR landfall timing?	Survey	What do you primarily use global models for when forecasting ARs? (e.g., trough variability - spatial or temporal, other large scale features, IVT plumes, etc.)
Poll	(WEEK 2 ONLY) I care about the AR landfall axis starting at day ____ lead time	Poll	(WEEK 1) Would you start to use CAMs for forecasting ARs at day 6? (WEEK 2) How likely are you to start to use CAMs for forecasting ARs at day 6?
Poll	I care about the AR inland extent starting at day ____ lead time	Survey	What do you primarily use CAMs for when forecasting AR-based events?
Poll	I care about the WFO-level impactful snow at day ____ lead time	Survey	What do you expect a convection allowing high-resolution model to struggle handling vs. a global model at day 5 lead time?
Poll	I care about highway or location specific snow levels starting at day ____ lead time		

Figure 5: List of questions asked for both focus group activities, and the question format.

understand how they identify ARs and what forecast metrics are important to them, and to help motivate participant feedback on their utility to inform future PEAR activities. These forecast activities were inspired by the Flash Flood and Intense Rainfall (FFaIR) forecast experiment activities, which are described in more detail in Trojnia et al. (2024), and the PEAR-related activities are described in more detail in the tutorial created for the participants¹.

Fig. 6 shows the text-based activities used for PEAR. The first two activities focused on probabilistic timing. Participants were provided 6-h time windows, six windows in the first week from 0000 UTC day 1 through 1200 UTC day 2, and four windows in the second week from 1200 UTC day 1 through 1200 UTC day 2. Participants were then tasked with assigning probabilities of the AR making landfall and the peak 6-h precipitation in the western United States occurring during each 6-h window, such that the sum of all probabilities adds up to 100%.

¹Tutorial location: https://www.wpc.ncep.noaa.gov/hmt/hmt_webpages/drawingtools/tutorial_pear.pdf

Figure 6: Text-based forecast activities used for week 2 of PEAR.

The next set of text-based activities were deterministic. Participants were tasked with forecasting the maximum 6-h and 24-h precipitation in the period of interest, between 1200 UTC day 1 and 1200 UTC day 2. Questions then became more open-ended, as participants were asked to predict the maximum IVT value at landfall, an IVT value they consider to be significant in making their forecast, and the AR event duration in hours. Participants were given an opportunity to explain their rationale for selecting the thresholds they did in the forecast activity survey.

Fig. 7 shows the map layout used for drawing activities, and a sample forecast from a participant. For this activity, participants were provided an interactive map to draw polygons for the AR contour at landfall and heavy precipitation risk. The latter contour was described to participants as drawing one or more contours for locations where they expect a risk of heavy precipitation, which is not necessarily the same as the excessive rainfall outlook (ERO). The objective of this activity was to have participants highlight a general corridor of heavy precipitation without too much of a focus on individual QPF contours, which given the complex terrain in the region risks becoming primarily contouring topographic lines. Participants were also tasked with drawing a single line corresponding to the orientation of the AR landfall axis and the inland extent of precipitation.

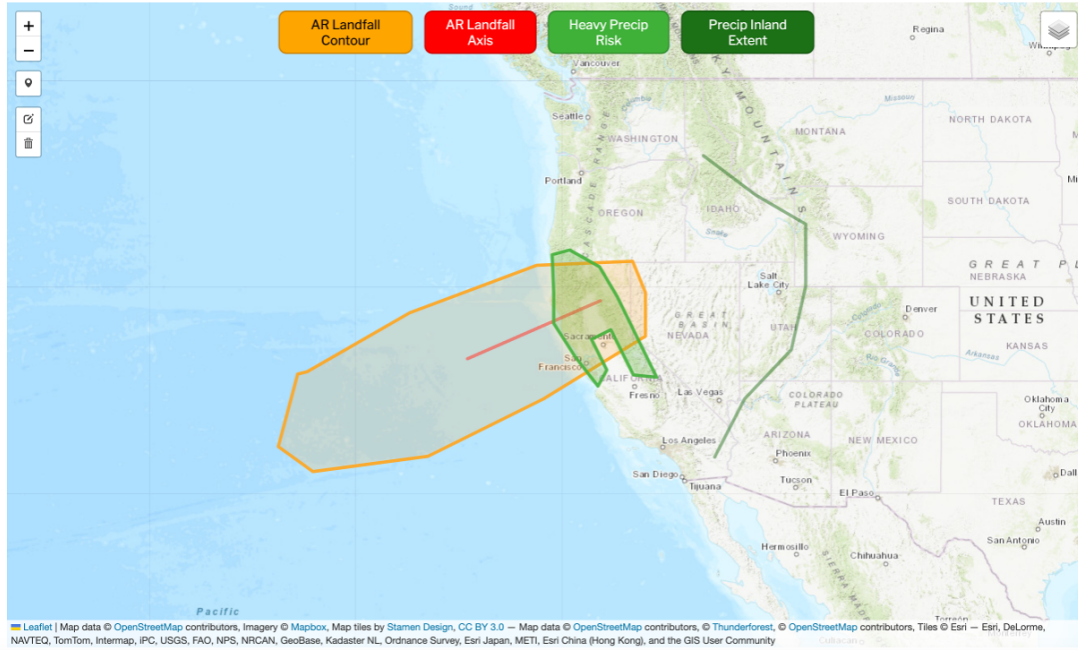


Figure 7: Map-based drawing activities used for both weeks of PEAR, with a sample forecast shown for reference.

Participants were tasked with providing feedback for these activities and their forecast process in the forecast activity and forecast verification surveys, and were also free to provide overall thoughts on the forecast and the model data they looked at after completing each forecast activity. These activities were repeated for each case at day 5, day 3, and day 1 forecast lead times. It should be noted that only the AR-AFS is run out to 120 h forecast lead time, so the day 5 activity only had the AR-AFS available through the middle of the forecast period.

2.7 Verification Methods

Given that this is the first iteration of PEAR, and most of the forecast activities were subjective, most of the verification results will consist of a subjective evaluation of the focus group activities and comparison of participant forecast activities. Quantitative verification consists of object-based verification graphics using the Developmental Testbed Center’s Model Evaluation Tools (MET) Method for Object-Based Diagnostic Evaluation (MODE). MODE was applied to the AR-AFS and UFS-AR for both case studies, with verification from the Unrestricted Mesoscale Analysis

(URMA). Forecast activities involving precipitation amounts were compared to the Multi-Radar Multi-Sensor (MRMS) Gauge Corrected Quantitative Precipitation Estimate (QPE), though it should be noted these estimates can be detrimentally impacted by spurious localized QPE maxima.

3 Results

The results section will primarily focus on the day 1 focus group activities. As most of the forecast activities done were subjective, a limited objective results section will evaluate the models used for the activities, followed by a subjective results section primarily evaluating the participant results from the forecast activities.

3.1 Focus Group

For the sake of brevity, not all of the focus group questions will be discussed in-depth in this section.

3.1.1 AR Forecasting by Lead Time

The first question asked of participants was “I get concerned about an AR event at day (blank) lead time”. The respondent answers for both weeks are shown in Fig. 8. Generally, the majority of participants selected between 5 and 8 day lead times, with the needs of partners generally being cited as a motivating factor for their answers. One of the participants from WFO Juneau, AK noted that southeast Alaska depends on vessels for commerce and supplies, as it is not connected to the Alaskan highway system, and longer lead times help to adequately prepare. Some participant comments that exemplify how their answers depend on partner needs include:

- “Shocking number of partners make [decisions at] lead times longer than I envisioned”
- “Many people have a regular weekend ... we start messaging before the weekend in case people don’t receive that message”

An interesting implication from this discussion and subsequent questions is the discrepancy from the forecaster perspective where technological and practical limitations exist for extended range forecasting, and the partner perspective where even a low-probability forecast scenario at extended lead times is useful information for them if it would lead to substantial impact on their operations. In the Alaska example, one might speculate even a low probability of a major AR at day 6 lead time may necessitate plans to reroute commercial shipping. This may conflict with forecaster capabilities, as one participant noted their answer of day 5 lead time was driven by their software limiting their QPF to only day 5 lead time.

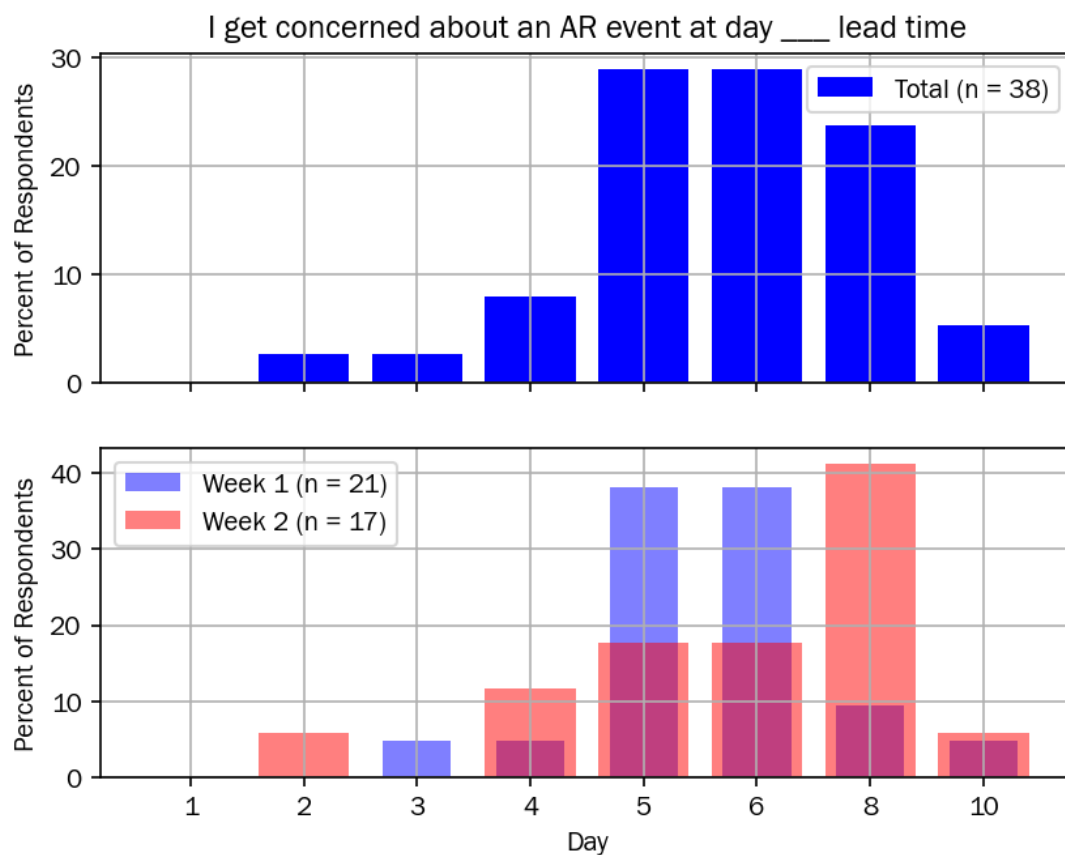


Figure 8: Combined results (top) and results partitioned by week (bottom) for the question “I get concerned about an AR event at day (blank) lead time”, shown as percent of respondents.

The next two questions of interest addressed the lead times at which participants think about WFO-level impactful Quantitative Precipitation Forecast (QPF)

(Fig. 9) and specific location QPF (Fig. 10). Generally, participants think about WFO-level impactful QPF at longer lead times than specific location QPF, with a mode of day 5 lead time for both groups for the former question, and a mode of day 3 lead time for both groups for the latter question. Yet in spite of these answers, there is still a dilemma between wanting to get information out to partners as early as possible, while feeling low confidence in the forecast details at those longer lead times. This can put extra pressure on forecasters to provide deterministic details or narrow probabilistic ranges they may not feel particularly confident in. Some participant comments that exemplify this dilemma include:

- "Confidence matters in what partners are worried about, they don't want us to wait for 50% confidence ... 20-30% confidence level at day 6."
- "I've heard [that] partners want to know about it, 7-10 days in advance. They want to start moving resources. I've heard from partners that even if the event has low confidence, they want to have that forecast on the table."
- "For specific locations, I need high res guidance and it's still jumping all over the place ... even during the event"
- "As I keep hearing about all of this I'm getting more annoyed that I have to put deterministic QPF out for longer lead times because I have to think about generalities and individual locations, even though we don't have a lot of time to do so and guidance isn't very trustworthy"

3.1.2 Model and Tool Utility

In week 1, participants were asked about the importance of IVT in their forecasting process. Following more discussion than anticipated, the focus group activity in week 2 added a question about the importance of PWAT in forecasting ARs to directly compare the two (Fig. 11). While participants agreed that both are important forecast metrics, an overwhelming majority strongly agreed that IVT is important for forecasting ARs, while responses regarding the importance of PWAT were more divided. Dynamically, this makes sense as precipitation is generally associated with moisture flux convergence, rather than the magnitude of instantaneous vertically integrated moisture, and forecasters especially in the

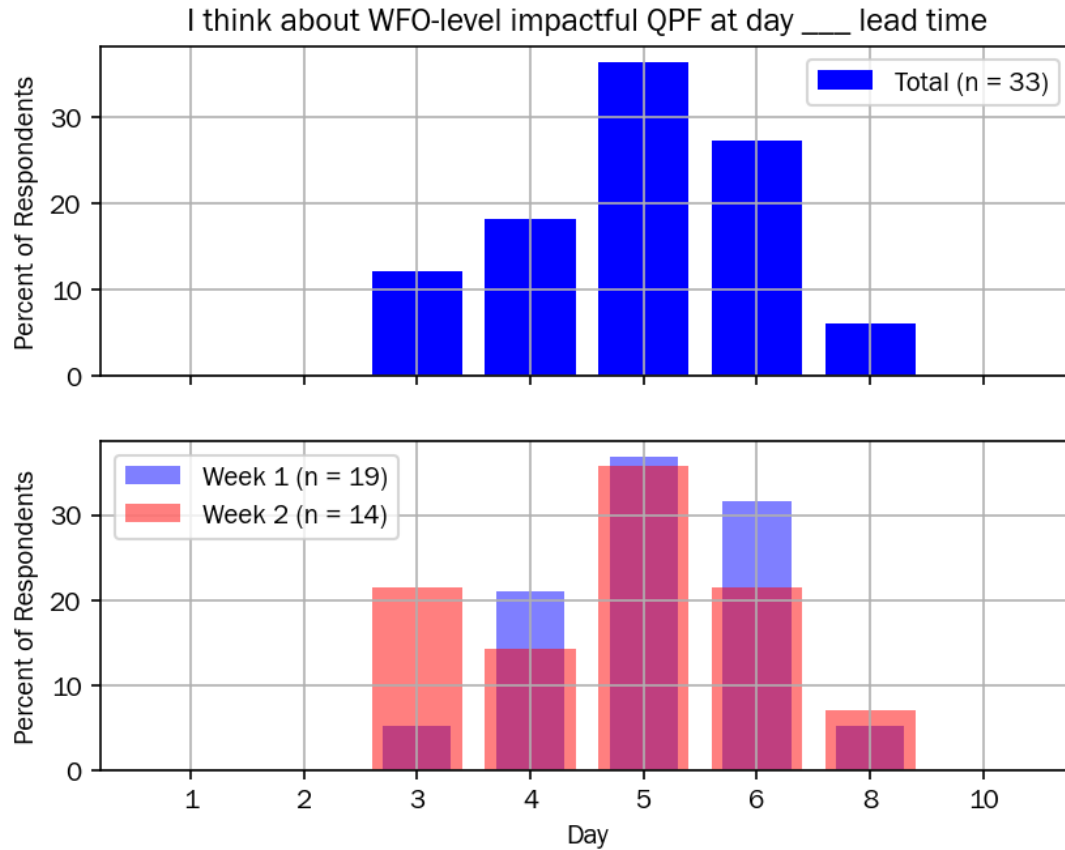


Figure 9: Combined results (top) and results partitioned by week (bottom) for the question “I think about WFO-level impactful QPF at day (blank) lead time”, shown as percent of respondents.

western US have had frequent exposure to IVT as part of AR literature and forecasting. Participant comments reflect these points as well:

- “PWAT is taking out the wind [from IVT]. Not getting as much of the picture. You can have all the PWAT, instability, winds perpendicular to mountain ... it won’t lead to precipitation amounts”
- “I think I have a lesser comprehension of how PWAT varies in the West and in the higher terrain. I struggle to truly identify what PWATs will be impactful in the region, as opposed to IVT and IVT anomalies”

Finally, participants were asked a hypothetical question about whether or not they would use a CAM for forecasting ARs at day 6 lead time. The wording was

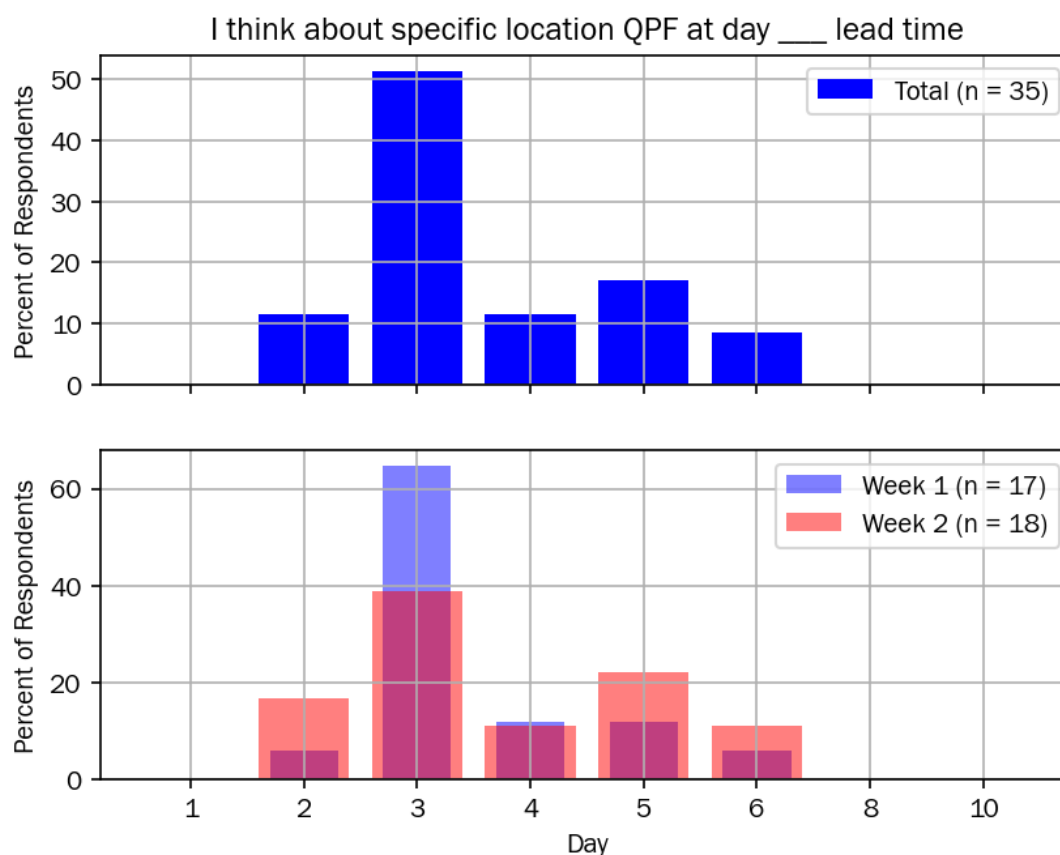


Figure 10: Combined results (top) and results partitioned by week (bottom) for the question “I think about specific location QPF at day (blank) lead time”, shown as percent of respondents.

slightly changed for week 2 to more clearly emphasize likelihood of using CAMs at day 6 lead time after reflecting on the discussion from week 1. The following percent options were presented: 0%, 25%, 49%, 51%, 75%, and 100%. The goal of presenting 49% and 51% separately was to motivate participants on the fence to hedge towards a side. Participant responses (Fig. 12) significantly varied, but with a majority leaning against it. Primary concerns included that deterministic solutions at longer lead times can more easily diverge from ensemble means and lead forecasters into erroneous solutions, or worrying about messaging about an event too early then having to backtrack on the forecast. A comment that came up multiple times in the subsequent discussion is thinking of day 6 CAMS as “another

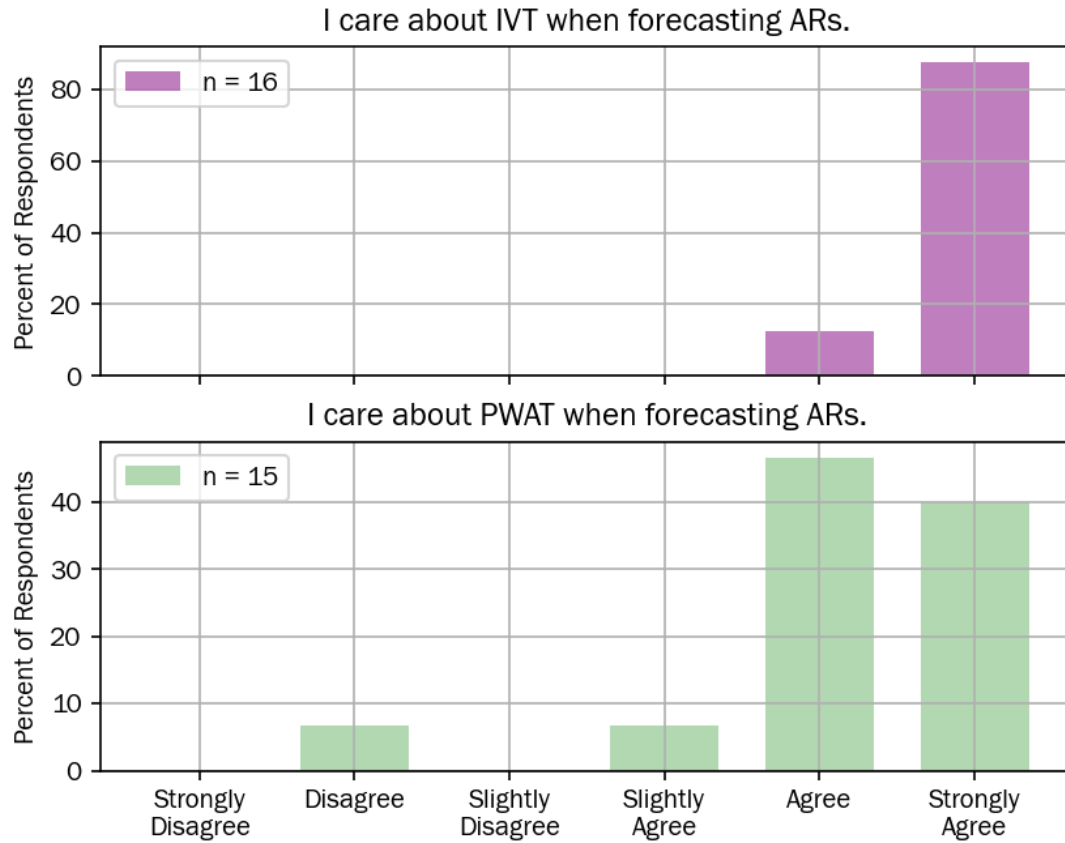


Figure 11: Results from the week 2 participant group for the question “I care about IVT when forecasting ARs” (top), and “I care about PWAT when forecasting ARs” (bottom), shown as percent of respondents.

tool in the toolbox.” Interestingly, there were split perceptions of whether this is a good or a bad thing. The following two quotes especially exemplify this contrast:

- “I would think of it more as another tool in the toolbox ... just like any other model, you can’t take everything at face value but it’s not bad to have”
- “When I hear about adding new tools to the toolbox, it makes me worried because we already have too many. At what point is it too much?”

Participants often brought up their preference for ensembles for extended range forecasting throughout the focus group activities, as well as during the discussion segment for this question. Tying into the discussion about “tools in the toolbox,” when there are too many models available for a forecaster, techniques such as

clustering and ensemble probabilities become useful to condense the available information into practical use for forecasting. However, given the limited time to analyze data in live operational settings, this may come at the cost of an excess focus on probabilities at the expense of synoptic-dynamic and mesoscale analysis of the underlying factors driving ensemble probabilities. It can be speculated that a net benefit for forecasters may come from having both ensemble probabilities and representative members of each cluster easily accessible within their workflow.

Interestingly, forecasters that have access to the West-WRF ensemble for their forecasting already have a day 6 lead time CAM available for their forecast workflow. The juxtaposition between that fact and the responses to this question is worthy of future analysis in future PEARs.

3.2 Forecast Activities

The following section will review the participant results from the forecast activities for both cases, with both subjective evaluation and objective verification results.

After the conclusion of the focus group activities, the remainder of PEAR was focused on conducting forecast activities for the two selected case studies. These cases were selected due to their significant impact on the West Coast with prolonged heavy precipitation, while still having different synoptic and hydrological precursor environments. The first case came in the midst of an active series of ARs and was embedded in a strong west-southwesterly upper level jet, while the second case followed a dry spell along the West Coast and was associated with the interaction of two synoptic-scale troughs, adding an extra layer of forecast uncertainty in the medium range. The differences between these cases offer a glimpse of how forecasters approach forecasting ARs with medium-range uncertainty but with a high-end ceiling for impacts.

3.2.1 Case 1

At day 5 lead time, all major models available showed the AR impacting the West Coast, but were generally too slow and too far north with the heavy

Week 1: Would you start to use CAMs for forecasting ARs at day 6?
Week 2: How likely are you to start to use CAMs for forecasting ARs at day 6?

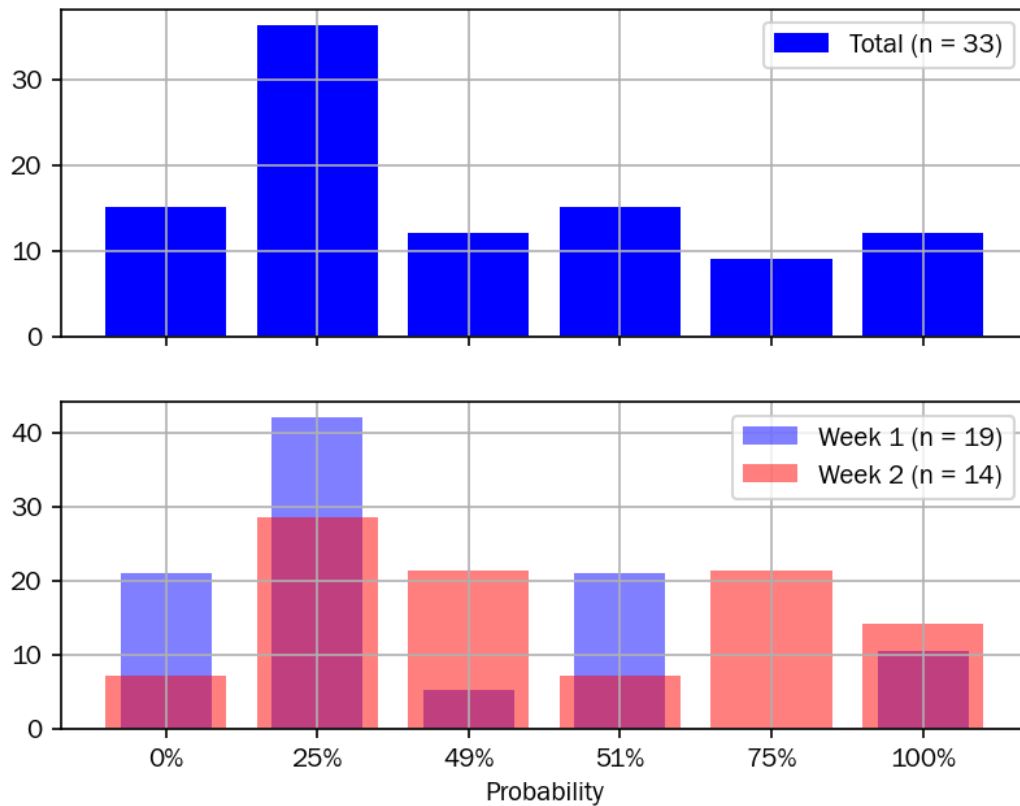


Figure 12: Combined results (top) and results partitioned by week (bottom) for the question “Would you start to use CAMs for forecasting ARs at day 6?” (week 1), and “How likely are you to start to use CAMs for forecasting ARs at day 6?” (week 2), shown as percent of respondents.

precipitation during the 24-hour window ending 1200 UTC 9 January 2023. Even with the caveat that the West-WRF and AR-AFS did not fully cover the 24 h forecast period at day 5 lead time, the AR-AFS was still too far north compared to the West-WRF and to observations. At day 3 lead time, when all models fully covered the 24 h forecast period, there was a southward shift in the forecast maximum precipitation corridor. While it is expected that the GFS was too low with its maximum 24 h QPF (5.55”) compared to the MRMS analyzed 24 h QPE maximum (14.05”) given its coarse resolution, even the AR-AFS was too low with maximum QPF (7.28”), while West-WRF was much closer to MRMS (12.29”). Day 1 lead time brought a continued southward shift in the maximum

QPF swath, and while the forecast maximum QPF for both GFS (7.96") and AR-AFS (9.83") increased compared to day 3 lead time, both models were both still too low compared to the MRMS analysis.

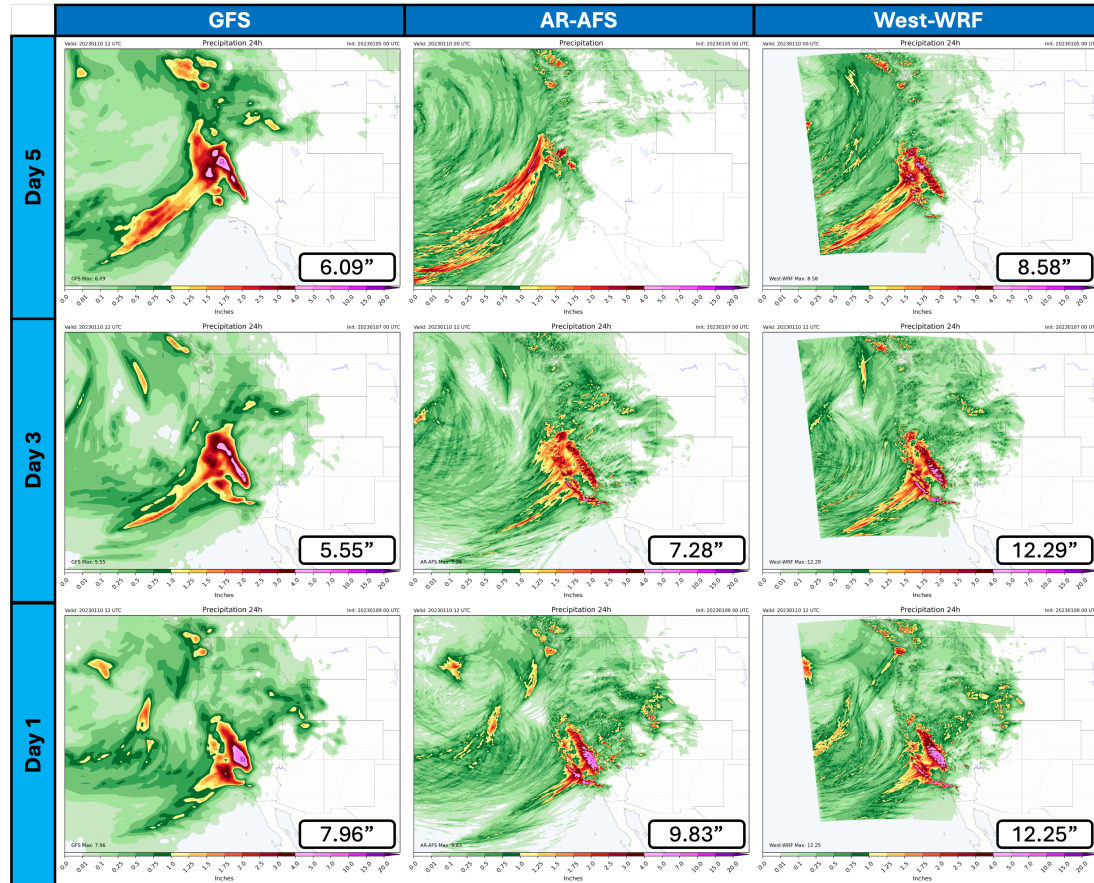


Figure 13: Panel plots showing the forecast 24 h QPF ending 1200 UTC 9 January 2023 (except AR-AFS and West-WRF at day 5 lead time, ending at 0000 UTC 9 January 2023, at their maximum forecast lead times), at 5, 3, and 1 day forecast lead times.

Participants were tasked with reviewing the model data provided at each forecast lead time and making a forecast. As noted in section 2.4, an error in the IVT calculation resulted in erroneous IVT values for case 1, though this largely did not affect the overall location of the IVT plume. With that caveat in mind, the participant forecast ensemble for the AR landfall contour as a function of forecast lead time (Fig. 14) largely followed the model QPF trends by forecast lead time previously discussed. At day 5 lead time, the consensus forecast AR landfall

contour was too far north, while gradually adjusting southward at days 3 and 1 lead time.

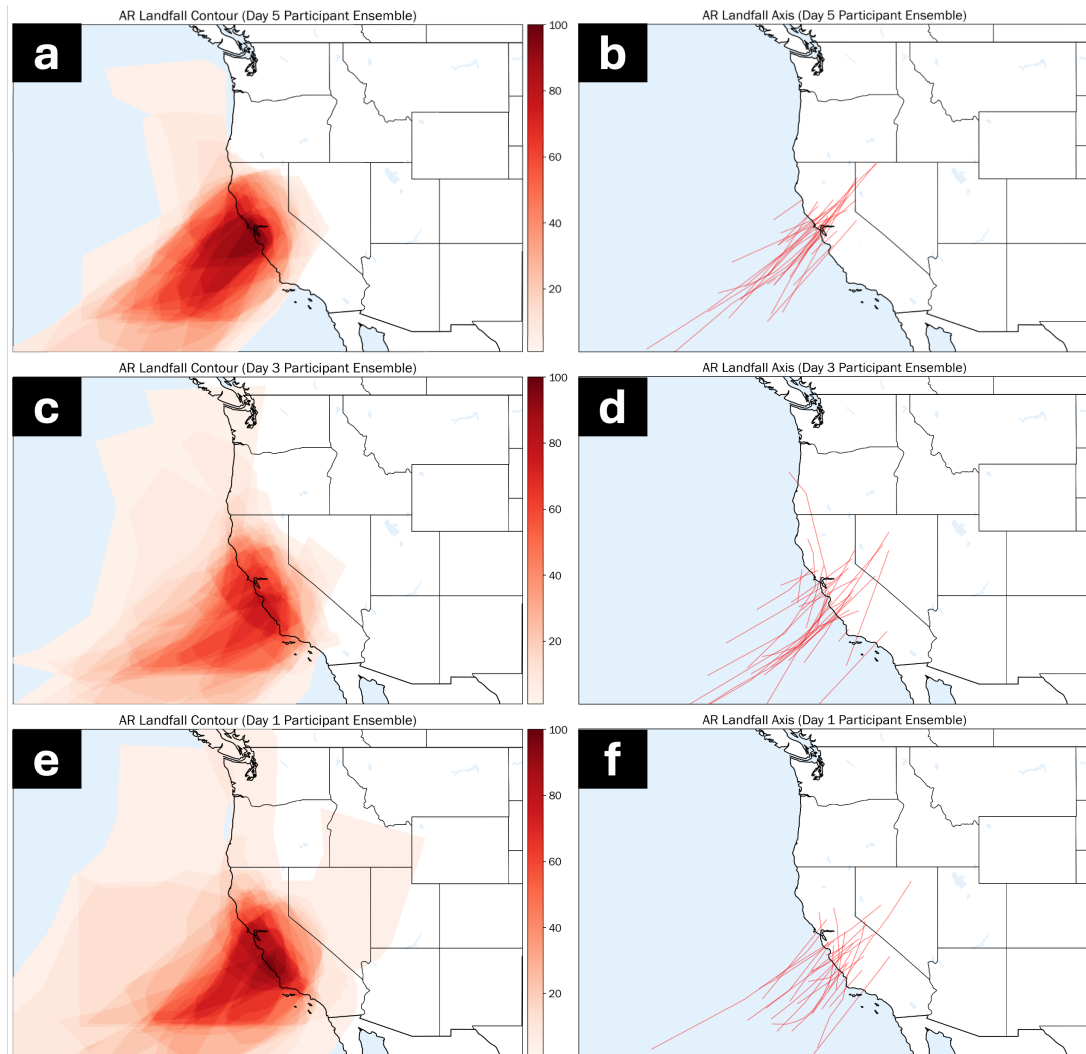


Figure 14: Participant forecast ensemble for the case 1 for (a, c, e) AR contour at landfall (percent of forecasts), and (b, d, f) AR landfall axis, for (a, b) day 5 lead time, (c, d) day 3 lead time, and (e, f) day 1 lead time.

The same forecast trends extend to the participant forecast heavy precipitation risk contour (Fig. 15), which shifted southward with decreasing lead time. Notably, while all participants highlighted mountainous terrain in their contoured area, some highlighted valleys in their contours while others did not. Based on participant discussion in the verification session, the subjectivity in how to define "heavy

precipitation” influenced whether to include the valley in the contour or not, taking into account factors such as local climatology and modeled QPF amounts.

One of the participants noted they felt less confident in their day 3 forecast than the day 5 forecast. A topic of discussion that arose from this question is whether a systematic bias in forecast AR position exists, in the context of the southward trend seen in this case, and participants noted they are not aware of any such systematic bias.

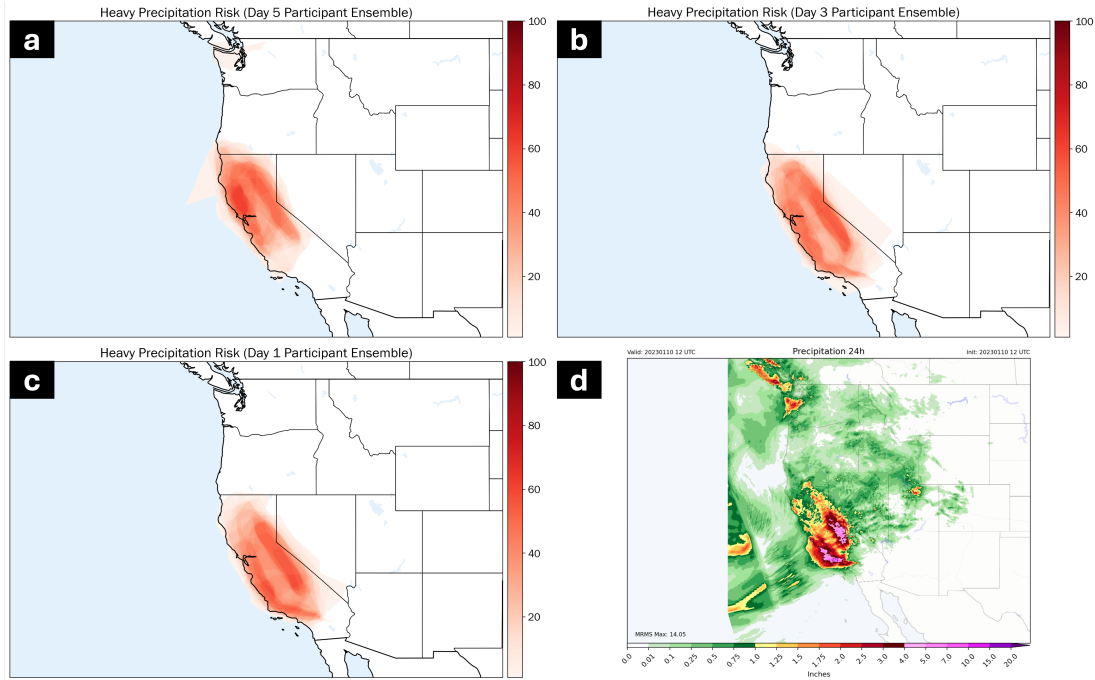


Figure 15: Participant forecast ensemble for case 1 for the heavy precipitation risk contour, at (a) day 5 lead time, (b) day 3 lead time, and (c) day 1 lead time. (d) MRMS analyzed QPE for the 24 h forecast period.

Participants were also tasked with forecasting the maximum QPF in the forecast domain over the 24 h forecast period (Fig. 16). At day 5 lead time, the majority of participants predicted maximum QPF between 6” and 7”, with a forecast mean of 7.15”. As the 00 UTC cycle CAMs at that lead time did not fully cover the forecast period, it is likely that this constraint, in addition to caution in forecasting high-end deterministic QPF amounts at longer lead times, kept forecast QPF totals lower than had CAMs extended farther out, or had the 12 UTC cycle AR-AFS been provided to participants for this case study. At day 3 lead time,

the forecast mean QPF increased to 8.71", but with a very large spread ranging from 3" to 13". By day 1 lead time, with the addition of the UFS-AR, the forecast mean QPF further increased to 10.52", while notably no participant predicted a maximum QPF below 8". Despite the trend towards higher maximum QPF with decreasing lead time, every participant predicted a lower QPF maximum compared to MRMS analysis.

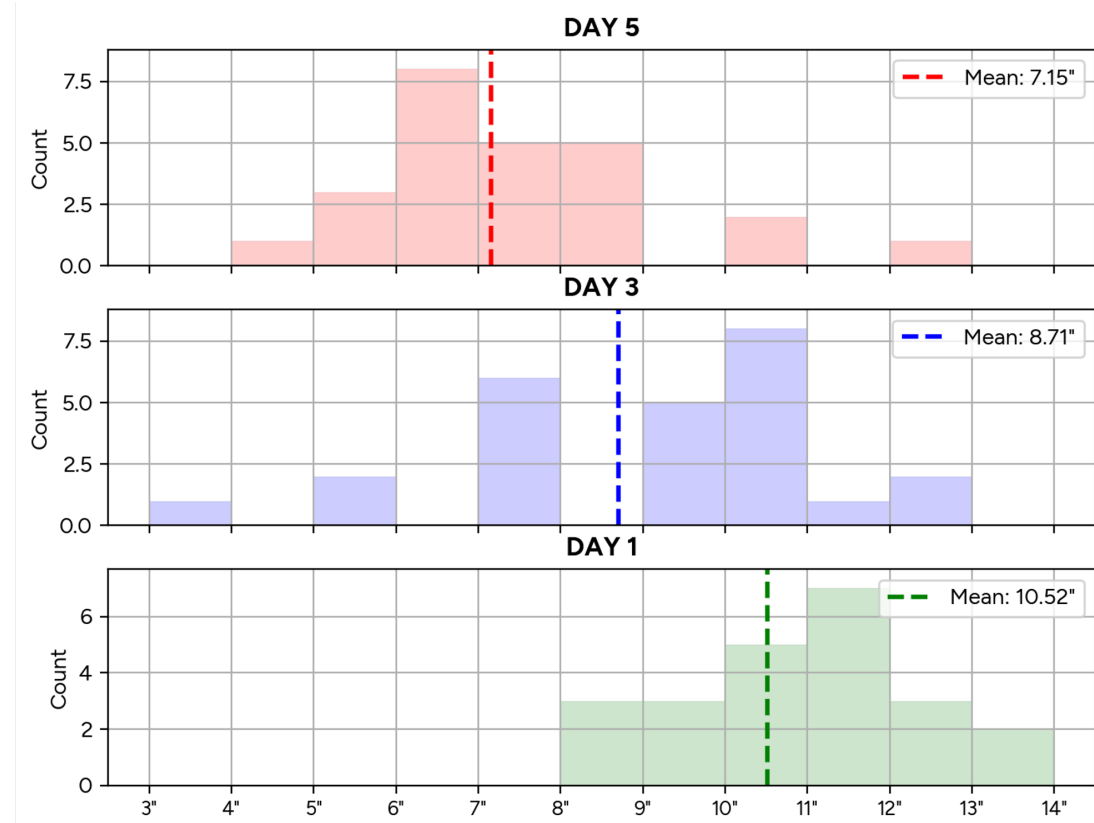


Figure 16: Participant forecast distribution from case 1 for the maximum 24 h QPF in the western U.S. domain at 5, 3, and 1 day lead times. The forecast mean is shown in the dotted line.

3.2.2 Case 2

In between case 1 and 2, the aforementioned IVT calculation error was fixed, allowing for a direct comparison of IVT magnitude and location between all models, and 12 UTC cycles were added for the GFS and AR-AFS. This change meant that

AR-AFS was now available to participants for the entire 24 h forecast period at day 5 lead time.

Comparing the GFS and AR-AFS IVT at different forecast lead times (Fig. 17), the primary uncertainty at day 5 lead time was the evolution of the cutoff low in the North Pacific and the potential for a tropopause vortex off the Northwest Pacific coast. While both 00 UTC GFS and AR-AFS failed to depict the AR making landfall in the West Coast at day 5 lead time, the AR-AFS did correctly identify the presence of the Northwest Pacific tropopause vortex, and consequently had a better depiction of the AR orientation relative to the coast than the GFS. Both models better captured the AR making landfall in their 12 UTC cycles at day 5 lead time (not shown).

By day 3 lead time, both GFS and AR-AFS had a good handle on the synoptic evolution, but differed with respect to the location, timing, and amplitude of both the North Pacific cutoff low and the tropopause vortex. These differences in turn affected the magnitude and poleward extent of the IVT plume. The AR-AFS had a stronger and farther north tropopause vortex, and accordingly a stronger AR extending farther north into the Northwest US, than the GFS. Both models adjusted towards a weaker, farther south, and more progressive tropopause vortex at day 1 lead time, leading to a southward shift in the IVT plume compared to earlier forecasts.

Repeating the AR landfall contour activity from week 1 (Fig. 18), the trends in participant AR landfall contours with lead time also generally correspond well to the GFS and AR-AFS trends in the IVT contour by lead time. While the participant consensus consistently included California in the contour, some participants extended their contour farther north into the Pacific Northwest at days 5 and 3 lead time, before trimming back the northern extent at day 1 lead time. Regarding the AR landfall axis, the aforementioned day 5 synoptic uncertainty manifested in different angles for the AR landfall axis, with the consensus favoring a SW to NE oriented axis while some participants drew a more zonally oriented axis. Nearly all participants converged to a SW to NE landfall axis by day 3 lead time.

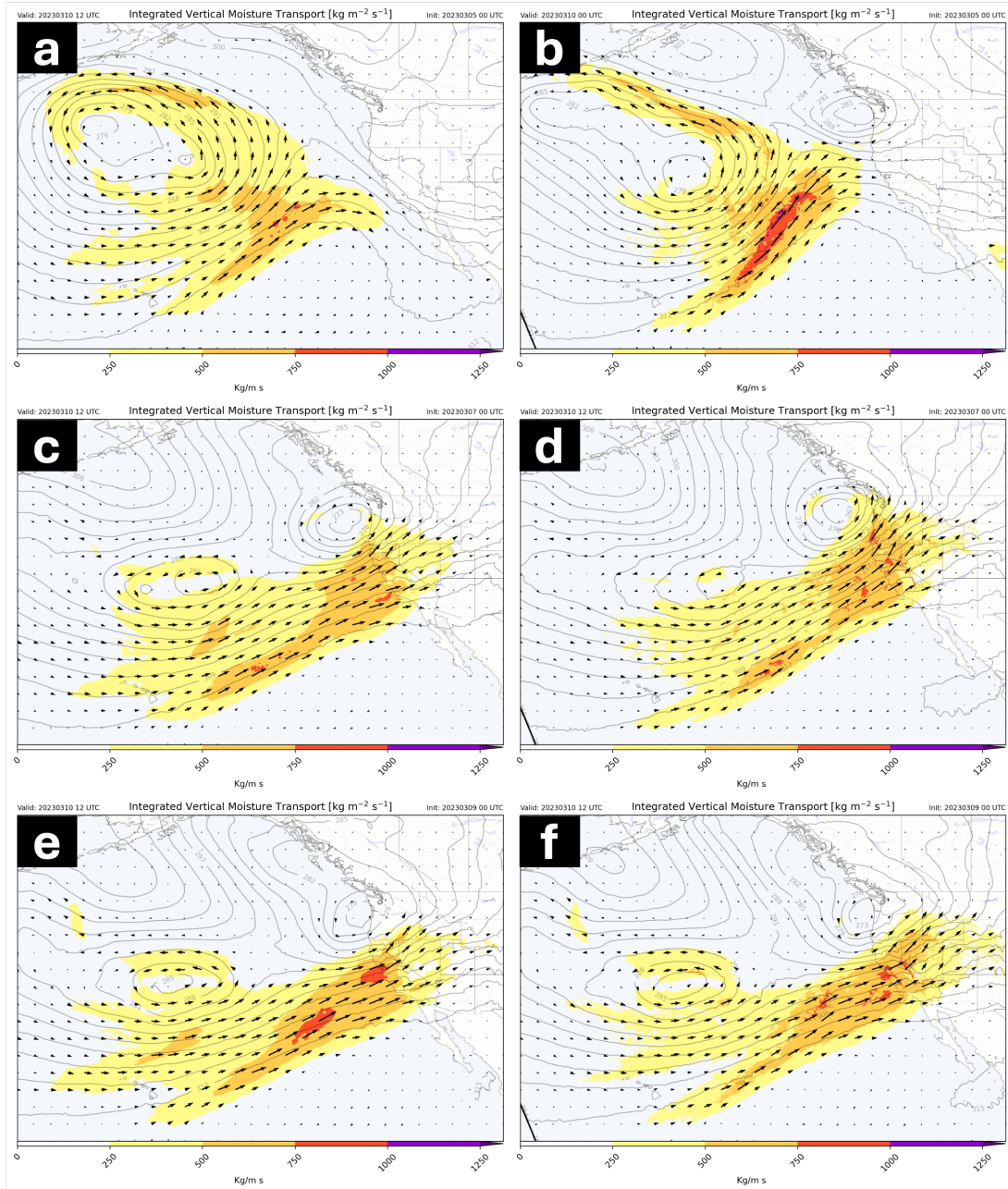


Figure 17: GFS (left; a, c, e) and AR-AFS (right; b, d, f) forecast IVT valid at 1200 UTC 10 March 2023 (except for panel b, valid at 0000 UTC 10 March 2023, at the maximum AR-AFS forecast lead time), for day 5 lead time (a, b), day 3 lead time (c, d), and day 1 lead time (d, e).

Given the loose definition of an AR landfall contour, participants used different criteria to draw their contours. Comments regarding the contours drawn included

"I was pretty broad with my contour because of the model spread", "I used a combo of 250 IVT and precipitation amounts and rates", and "It was pretty broad for an IVT plume and for me I treated it like a weighted average." A topic of discussion during the day 3 verification activity was the presence of two IVT maxima separated in space and time, which resulted in some participants drawing two separate AR landfall contours. As one participant noted, "The reality in meteorology is that things are usually going to be more sloppy than textbook cases and this is one of many examples of that." Another participant noted that the operational WPC and Ocean Prediction Center (OPC) surface analysis included multiple cold fronts, potentially indicative of separate moisture plumes.

The participant forecast heavy precipitation risk contours (Fig. 19) also highlighted the uncertainty in the northward extent of heavy precipitation. As with case 1, participants varied in whether to incorporate the valleys in their contours or not. In addition to the subjectivity element of the forecast noted by participants in week 1, some participants included the valleys in their contour due to IVT extending into the valley through the San Francisco Bay Area gap and concerns for flooding in urban areas. Regarding the northward extent, one participant noted the following: "I was more conservative with how much I drew and I kind of focused on the NW CA area where the focus is. I think that's because I know that when I forecast at these lead times, I can always expand that later ... when we draw these big broad areas at times, they aren't shrunk like they need to be once we get further into the event and more clear about where it is. So for this one I didn't go very far north and I may have gone further north on Day 3."

With the GFS and AR-AFS fully available through the 24 h forecast period at day 5 lead time, and the outer domain for the West-WRF also covering the full period, participants had sufficient information to make a maximum QPF forecast for the full period at all forecast lead times. The resulting distribution (Fig. 20) shows that as with week 1, the participant distribution at day 5 lead time was on the lower end, with a mean of 4.55". The forecast mean increased to 7.33" at day 3 lead time, and 8.42" at day 1 lead time. Unlike case 1, there was a pronounced skewness towards higher-end precipitation totals, with one participant at day 1 lead time forecasting as much as 18" of rain.

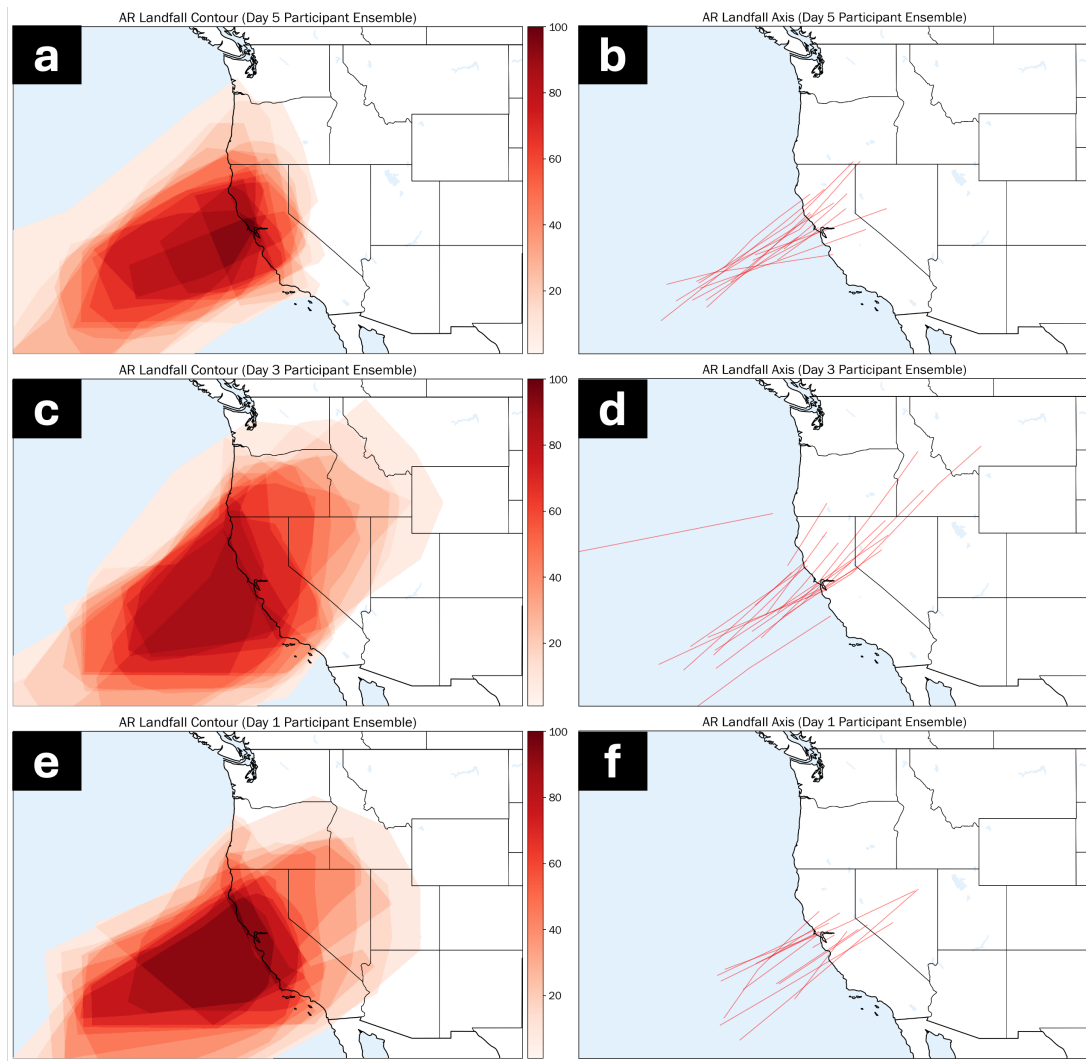


Figure 18: Same as Fig. 14, but for case 2.

The participant ensemble forecast bias differed from case 1 as well, as most participants over-predicted 24 h QPF during the forecast period compared to MRMS analysis. It should be noted that as this forecast period covered only the beginning of the AR event, and higher precipitation amounts occurred beyond the forecast period, it is possible that the higher forecast amounts at later lead times may have contributed to participants forecasting higher totals than what was observed. The participant mean maximum QPF at day 1 lead time was also higher than the 00 UTC cycle data for the GFS (4.05"), AR-AFS (approximately 5"), West-WRF (7.26"), and the UFS-AR (8.26").

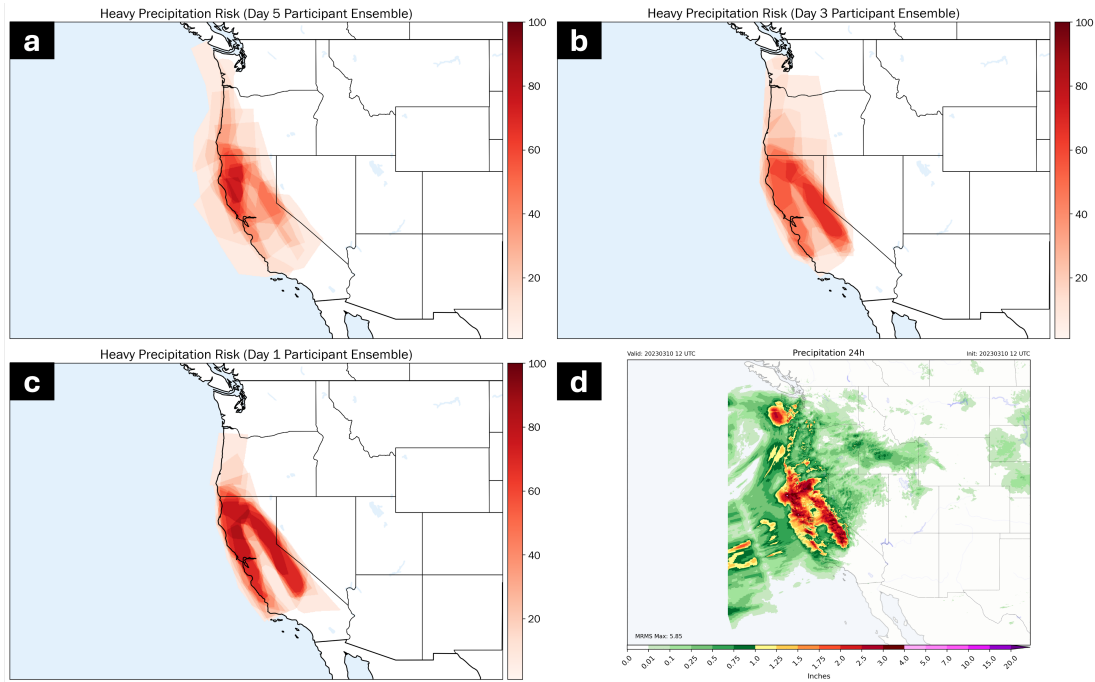


Figure 19: Same as Fig. 15, but for case 2.

3.2.3 Additional discussion

During the verification section, participants were provided MODE plots for days 5, 3, and 1 day lead time for the AR-AFS. For case 2 (Fig. 21), the northward displacement in the AR core and too much QPF off the Pacific Northwest coast at day 5 lead time match the similar participant biases in AR landfall location and heavy precipitation risk contours at those lead times, before both adjusted southward with decreasing lead time. In this particular case, the implication of this bias is under-forecasting the AR impacts at medium-range lead times. Given the limited deterministic data evaluated for this PEAR experiment, it is possible an ensemble may have better indicated the potential for more significant impacts over land, even if only a relatively low-probability solution in the ensemble.

Regarding the data and tools provided, participants noted that the information provided as part of this experiment was limited compared to the data and analyses that would have been made in an operational setting. While this is an understandable limitation of operating in a pseudo-operational setting in a testbed space, adding more maps to help participants better evaluate the experimental

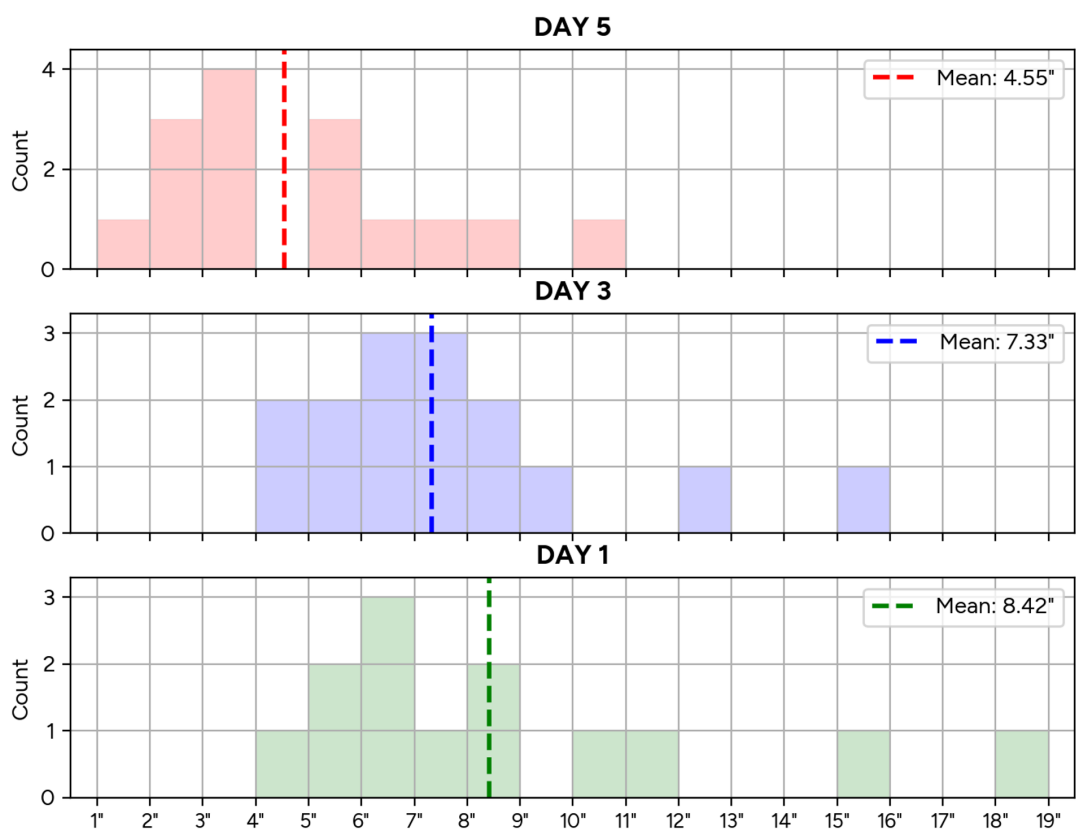


Figure 20: Same as Fig. ??, but for case 2.

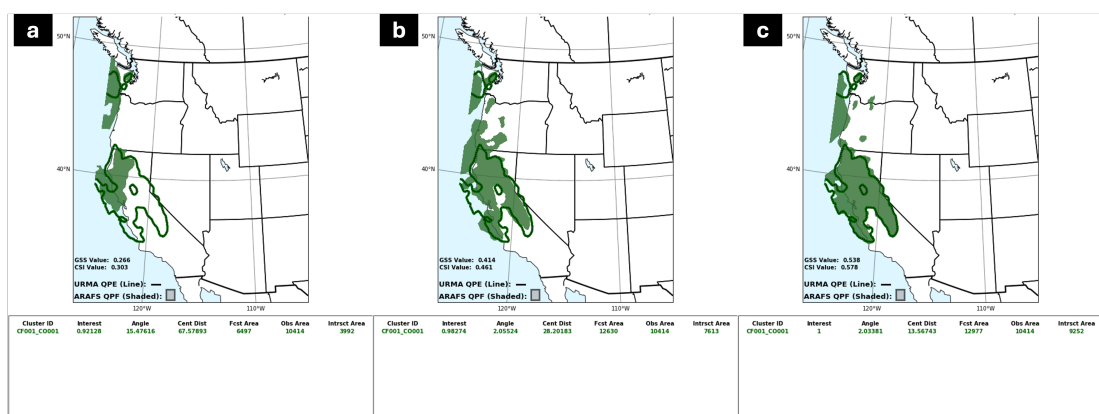


Figure 21: AR-AFS 1.0" precipitation objects (shaded) and URMA analyzed 1.0" precipitation objects (fill) created with MODE for (a) day 5 lead time, (b) day 3 lead time, and (c) day 1 lead time.

models for AR applications will be of use going forward. Several participants noted

that evaluating barrier jets, elevated stability, and mountain waves would have potentially led to a more accurate QPF.

An unexpected result of the forecast activity was discovering a correlation between participant forecasts for maximum 6 h and 24 h QPF for the forecast period. On average, the forecast 24 h maximum QPF was approximately twice the forecast 6 h maximum QPF. Upon further discussion of this finding in the forecast verification section of PEAR, it was discovered that this result was unintentional on the participants' end.

Another segment of the text-based forecast activities asked participants to subjectively write a significant IVT value of their choice, and to explain the reasoning for choosing their IVT value in the forecast activity survey. Although this question was asked of participants for all forecast lead times, the responses were consistent throughout all forecast lead times for each week, with an average of approximately $650 \text{ kg m}^{-1}\text{s}^{-1}$ for week 1 and $600 \text{ kg m}^{-1}\text{s}^{-1}$ for week 2. As the question was open-ended, participants' selected values were based on different reasons. Some selected $250 \text{ kg m}^{-1}\text{s}^{-1}$ as it is often used as a lower bound for ARs, and others selected $750 \text{ kg m}^{-1}\text{s}^{-1}$ as this was noted to anecdotally correspond to higher-end ARs and more significant precipitation events.

4 Summary and Conclusions

The first PEAR experiment provided insight into forecaster approaches to forecasting ARs and communicating their impacts, in addition to subjective evaluation of the utility of the AR-AFS through two case study activities. The addition of an AR experiment to the existing HMT suite of activities brought forecaster perspectives of forecast methods and communication previously not covered under recurring HMT experiments, and provided a baseline from which future PEAR experiments can evolve to better evaluate the utility of the AR-AFS in a pseudo-operational setting and its benefits with regards to communicating forecast hazards to partners and stakeholders.

4.1 Utility of AR-AFS

Multiple participants noted they felt AR-AFS added value to their forecast process, as its high resolution better resolved fine details such as topographic influence on precipitation extrema and distributions, while noting it added most value when used in conjunction with other CAMs such as the West-WRF. Especially at day 1 lead time, the addition of the UFS-AR data meant participants had 3 deterministic CAMs to reference, which they noted increased their confidence in areas where all 3 CAMs were in general agreement. As one participant noted, “I tried to make a mini-ensemble approach and go with what I saw as the median amount given the guidance I had.”

Aggregate verification statistics for the 2022–2023 retrospective West Coast AR season show that on average, the AR-AFS has a lower Critical Success Index (CSI) than other models evaluated and WPC forecasts for low QPF thresholds, but outperforms the global models for high-end QPF thresholds (Fig. 22). On average, participants felt the AR-AFS better handled precipitation extremes than the GFS for both cases (Fig. 23). Model utility comes not just from verification, but also from forecaster confidence in its output. To that end, participants were asked in the forecast activity survey after completing their forecasts if they felt the AR-AFS made them feel more confident in their 24 h QPF forecast, which the majority agreed it did (Fig. 24). Forecast activity surveys showed some variations in the participant reasoning; one noted they leaned more towards the AR-AFS at day 5 due to its higher resolution, while another said they leaned against AR-AFS because they felt it was too high, and didn’t want to message forecast precipitation amounts that were too high then have to walk it back at shorter lead times.

A main takeaway was the importance of medium-range synoptic scale variability on the days 5 and 3 lead time forecasts. While the AR-AFS and West-WRF provided increased resolution at these lead times, differences in mesoscale structures were outweighed by synoptic-scale variability in trough interactions and AR location and magnitude. Case 2 was especially a strong example of this importance at day 5 lead time. On average, week 2 participants said the AR-AFS made them feel more confident when forecasting the AR landfall location (Fig. 25a) when

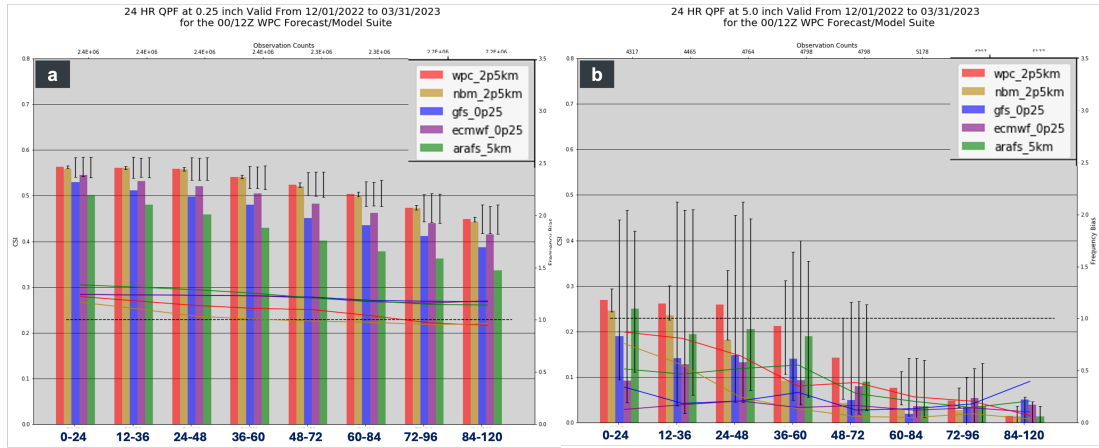


Figure 22: AR-AFS CSI skill scores for (a) 0.25 inch and (b) 5.0 inch thresholds.

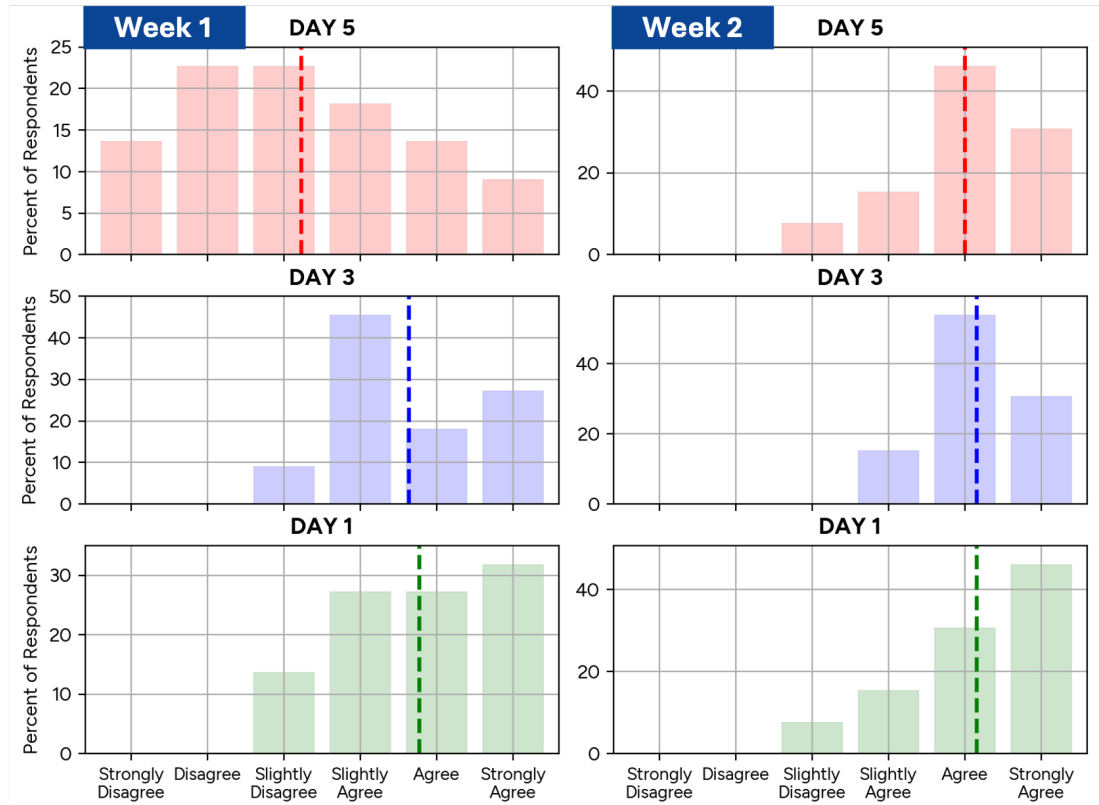


Figure 23: Participant responses to the survey question “AR-AFS better handles precipitation extremes than the GFS.”

completing their forecast activity surveys, but in the subsequent verification survey at the end of the experiment, the average response was between neutral and slightly

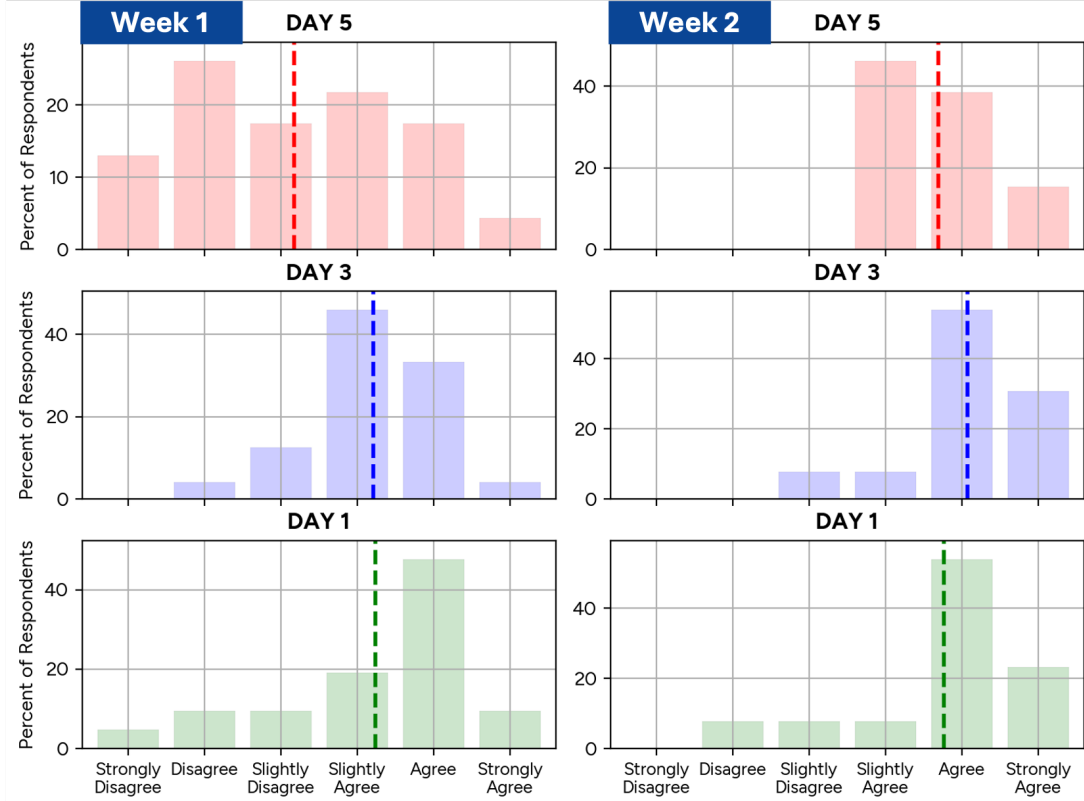


Figure 24: Participant responses to the survey question “AR-AFS made me more confident in Maximum 24-hour QPF totals when making my forecast.”

disagreeing that the AR-AFS better handled the AR landfall location than the GFS (Fig. 25b).

These results indicate a clear discrepancy in forecaster confidence and performance evaluation for the AR-AFS between QPF and AR landfall. While the higher horizontal resolution increases forecaster confidence in the QPF due to better resolving mesoscale phenomena and orographic QPF enhancement and suppression, a higher resolution model may not necessarily resolve the synoptic-scale placement and magnitude of the AR better at medium-range lead times. To that extent, the net benefit from the AR-AFS at these lead times may be more due to better resolving potential high-end precipitation amounts.

A frequent comment that arose from forecasters was the importance of ensembles in medium-range forecasting, especially in context of the increased emphasis on

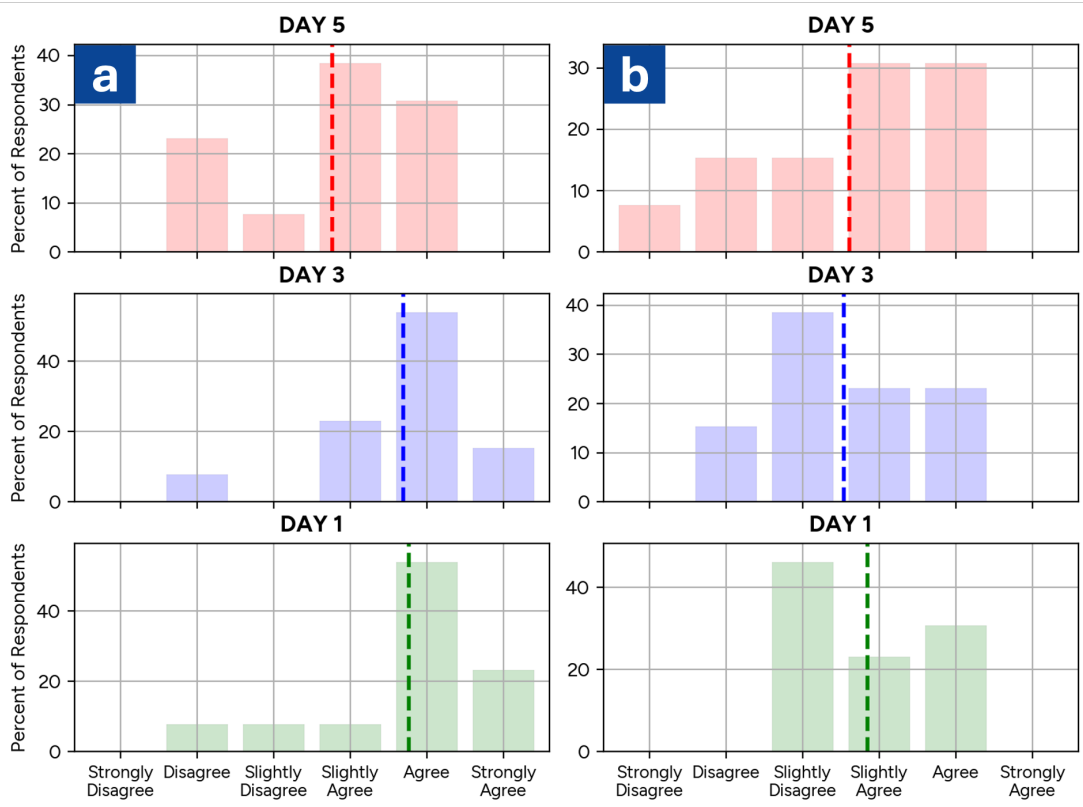


Figure 25: Participant responses to the survey question (a) “AR-AFS made me more confident in AR landfall location when making my forecast”, and (b) “AR-AFS better handles the AR location than the GFS.”

operational probabilistic forecasting, and that having a CAM ensemble would have added great value in their forecast process. As one participant noted, “[AR-AFS] is needed to be used with other models to help with confidence and there needs to be an ensemble element to the AR-AFS to make it much more useful in the forecast process.” While this holds true in an operational setting where there is an increasing focus on probabilistic forecasting, there is still value in deterministic CAM evaluation in forecast experiments. Given the limited time in forecast testbeds, spending too much time looking at model probabilities can mean less time spent evaluating the synoptic environment and differences between each model, less time understanding the mesoscale variability, and what physically results in each model producing its simulated QPF. These are important components of subjective model evaluation that can help identify potential issues or systematic biases in models

that aggregate verification statistics can miss, which in turn can be provided to modelers to address.

4.2 Activity Feedback and Future Objectives

Participants generally found the focus group activities informative, though some noted they felt it lasted too long and their interest waned over time. This is an understandable limitation of a first-time experiment where gathering information on forecaster practices and perspectives is as important as the forecast activities themselves. As the focus group activities from this PEAR yielded valuable information on how forecasters approach AR forecasting, the next PEAR experiment will have a much shorter initial focus group session, intended to follow up on a few outstanding questions from the results of this PEAR, to allow for more time spent on evaluating models and conducting new forecast activities.

The experimental drawing activities received a generally good reception, but with useful suggestions for improvement in future PEARs. Participants found the “heavy precipitation risk” activity to be a useful attempt at generalized risks without being constrained to specific precipitation amounts, though a frequent comment was that defining an “AR contour” was too vague. Similarly, the text-based activities for AR landfall timing were found to be too vague. When asked for suggestions to improve the forecast activities, participants generally favored using specific IVT values, such as 250 or $500 \text{ kg m}^{-1}\text{s}^{-1}$, as the defining threshold. Most participants agreed that the “inland precipitation extent” activity was the least useful for a multitude of reasons, with the most common reason cited being the vague definition of what quantitatively counts as the inland extent of precipitation associated with an AR. Another participant noted “All of the activities got me framing AR forecasting in a way that I hadn’t really thought of in operations.”

When asking participants for feedback on potential PEAR activities in the future, several suggested an interest in conducting more Impact-Based Decision Support Services (IDSS)-based activities, focusing on communicating forecast AR impacts. A suggestion brought up which received support in the group discussion was to expand on previous FFaIR activities which included a group consensus

forecast activity to conduct individual and group mock forecast briefings. Such an activity may help gain better insight into how participants use model data to identify the areas of greatest concern. This suggestion was incorporated into the planning for the next PEAR to be held in Spring 2025. Participants also emphasized the importance of forecast timing activities, as one had noted “[the] biggest communication issue at WFO level when interpreting ensembles is timing. People would much rather adjust plans rather than cancel ... high resolutions models can really help with giving a realistic picture regarding timing of impacts.”

There were generally few significant differences between models at day 1 lead time. This is a noticeable difference from the FFaIR experiment at HMT which primarily evaluates deterministic CAMs at day 1 lead time. It is expected that CAMs will exhibit greater differences for summertime warm convection cases with different convective modes and forcing mechanisms than cold season ARs with strong synoptic forcing and geographically fixed regions of mesoscale QPF enhancement due to orographic ascent. It can be argued that evaluating CAMs at day 1 lead time for ARs may still be of some value for comparing systematic biases with the exact magnitude of QPF enhancement due to orographic ascent and QPF suppression downstream due to rain shadows, although from a forecast experiment perspective this may not be the best use of the participants’ expertise and time. Given the previous discussion about the importance of medium-range synoptic scale uncertainty in AR forecasting, future PEAR activities will place a greater emphasis on days 5 and 3 forecast lead times, where more significant differences in the synoptic scale drive forecast uncertainty and which are outside of current CAM forecast lead times. Recalling the first focus group question where participants generally identified 5- to 8-day lead times as the time frame they start to get concerned about an AR event, the desired lead times for future PEAR activities better target these lead times where decisions are first being made while still being within the 120 h range of available model data.

A topic that came up during discussion segments was the potential utility of artificial intelligence (AI) emulator models. Preliminary evaluations of AI models show poor performance with QPF amounts, in part due to their relatively coarse resolution, which is especially detrimental for QPF forecasting over complex terrain

such as is the case with West Coast ARs. However, these AI models show substantial promise for medium-range synoptic variability. Building off of these discussions, future PEAR experiments will incorporate multiple AI models into the suite of available maps for participants to evaluate. Scenarios are being evaluated for how to best leverage AI models to complement forecasting with and providing feedback for NWP models, such as having participants begin their forecast activities with NWP models, primarily targeting QPF forecasting and obtaining a rough initial idea of synoptic-scale AR placement and magnitude, and later evaluate AI models to determine if there are substantial synoptic-scale differences.

References

- 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>.
- Ralph, F. M., P. J. Neiman, and G. A. Wick, 2004: Satellite and caljet aircraft observations of atmospheric rivers over the eastern north pacific ocean during the winter of 1997/98. *Mon. Wea. Rev.*, **132**, 1721–1745, [https://doi.org/10.1175/1520-0493\(2004\)132<1721:SACAOO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1721:SACAOO>2.0.CO;2).
- Ralph, F. M., 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>.
- Ralph, F. M., and Coauthors, 2020: West coast forecast challenges and development of atmospheric river reconnaissance. *Bull. Amer. Meteor. Soc.*, **101**, E1357–E1377, <https://doi.org/10.1175/BAMS-D-19-0183.1>.
- Trojniak, S., J. Correia, Jr., W. M. Bartolini, and T. Burg, 2024: 2024 flash flood and intense rainfall (ffair) operations plan, published online at https://www.wpc.ncep.noaa.gov/hmt/Reports/FFaIR/2024FFaIR_OpsPlan.pdf. If missing please contact WPC.