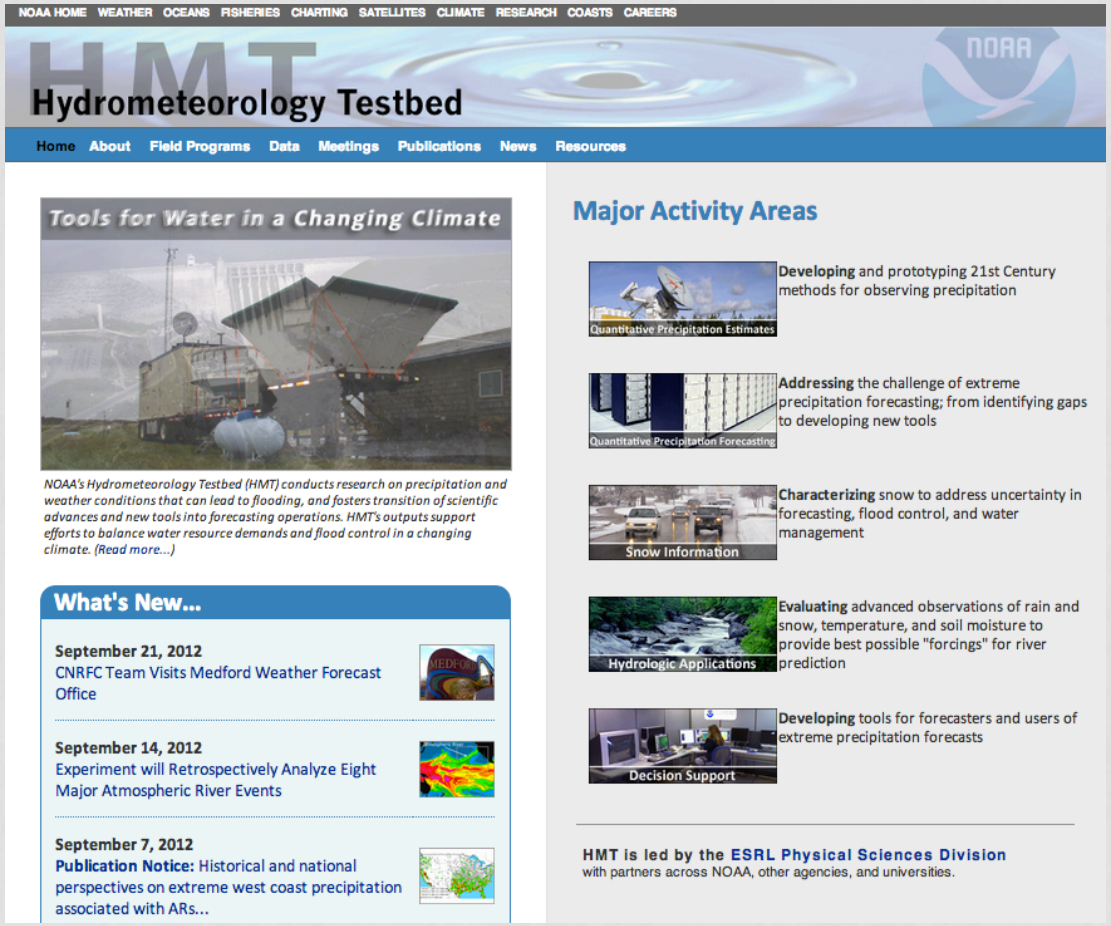
HMT-HPC/ESRL Atmospheric River Retrospective Forecasting Experiment (ARRFEX)

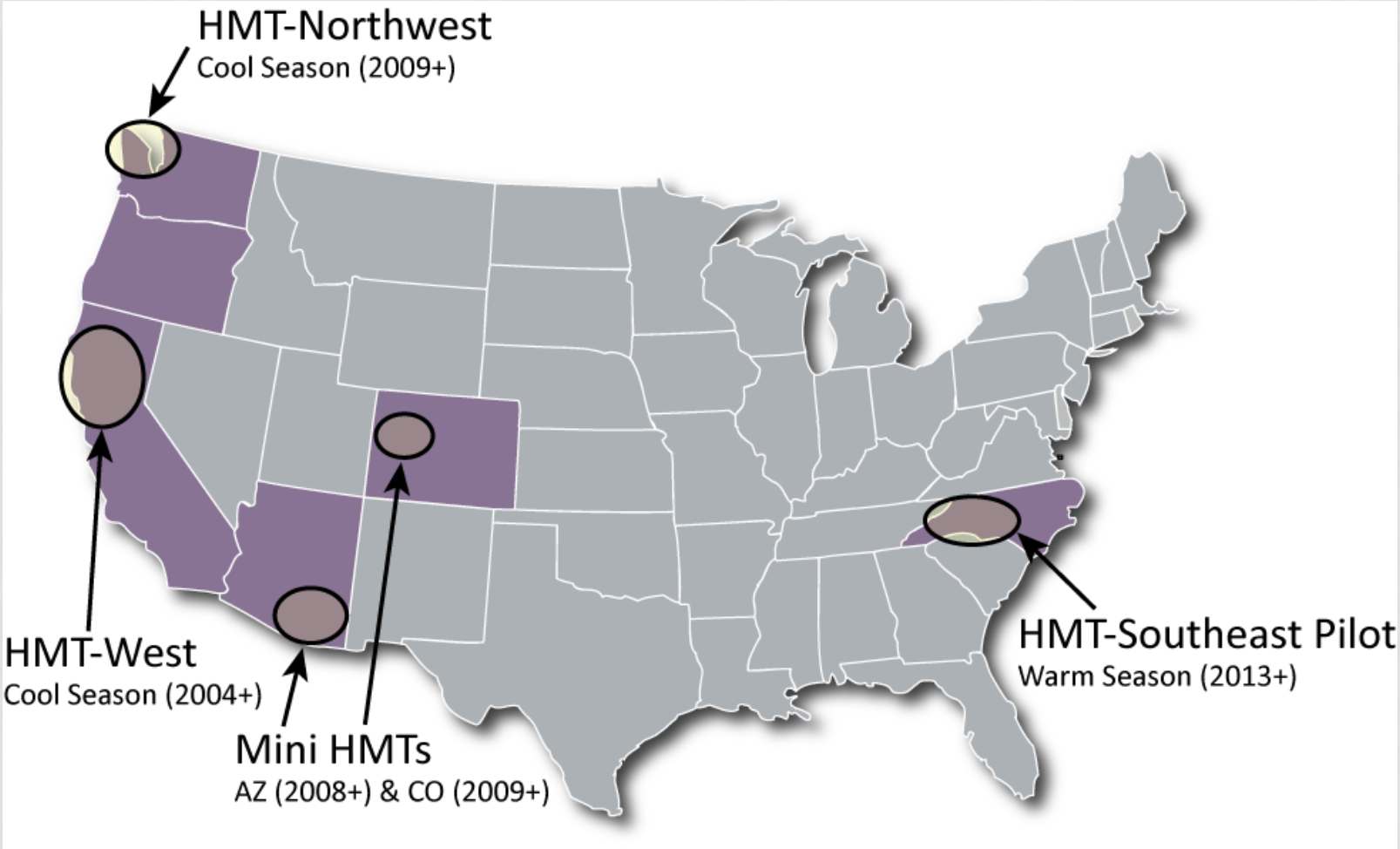
September 17-28, 2012

HYDROMETEOROLOGY TESTBED (HMT)



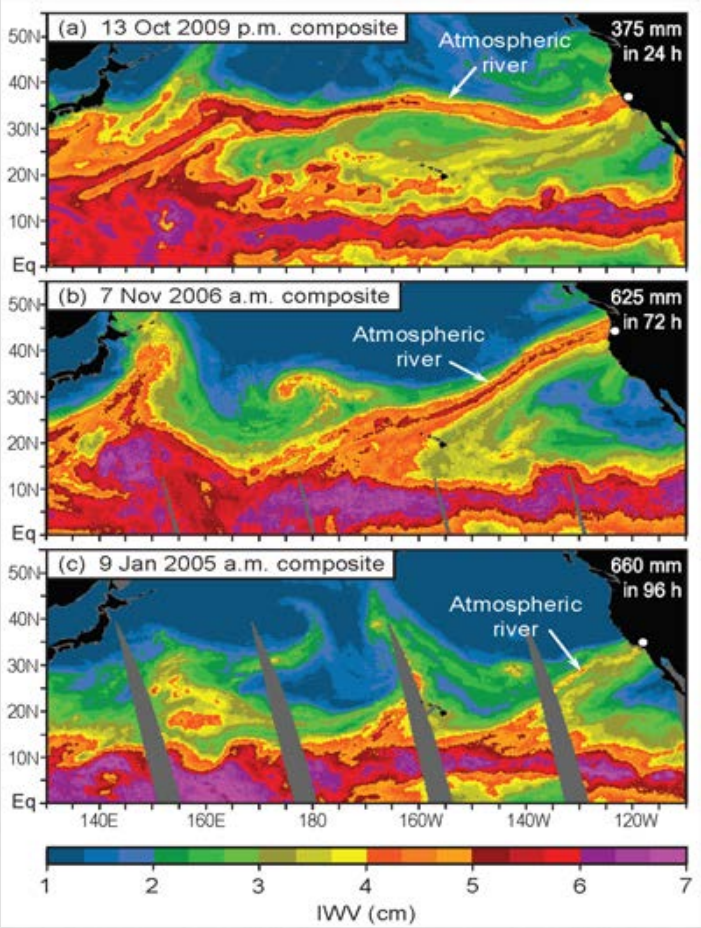
- HMT conducts research on precipitation and weather conditions that can lead to flooding, and fosters transition of scientific advances and new tools into forecasting operations
- 5 major activity areas
  - QPE
  - QPF
  - Snow Information
  - HASP
  - DST
- hmt.noaa.gov

HMT REGIONAL ACTIVITIES



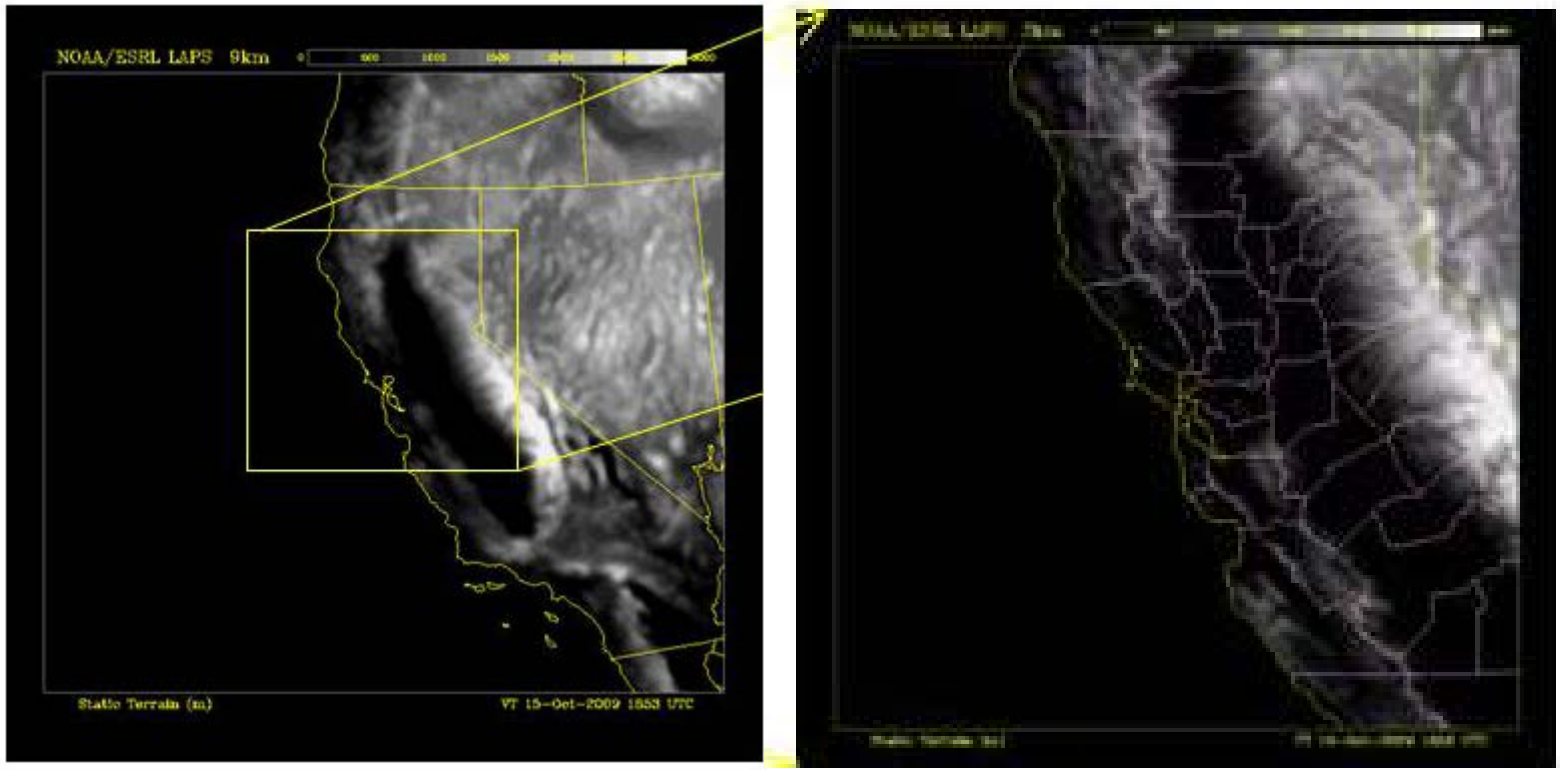
QPF ACTIVITIES IN HMT-WEST USING HMT-ENSEMBLE

- Improve precipitation forecasts (intensity, duration, timing)associated with land falling tropical cyclones along U.S. west coast (Atmospheric Rivers)
  - Improve microphysical representation of orographic rainfall processes
  - Explore probabilistic forecast skill
  - Assess uncertainty of ensemble forecast in comparison to operational ensembles
  - Effort supported through California Department of Water Resources and USWRP



Example Atmospheric River events

HMT-ENSEMBLE EXPERIMENT DESIGN 2009-2010, 2010-2011, 2011-2012



Nested domain:

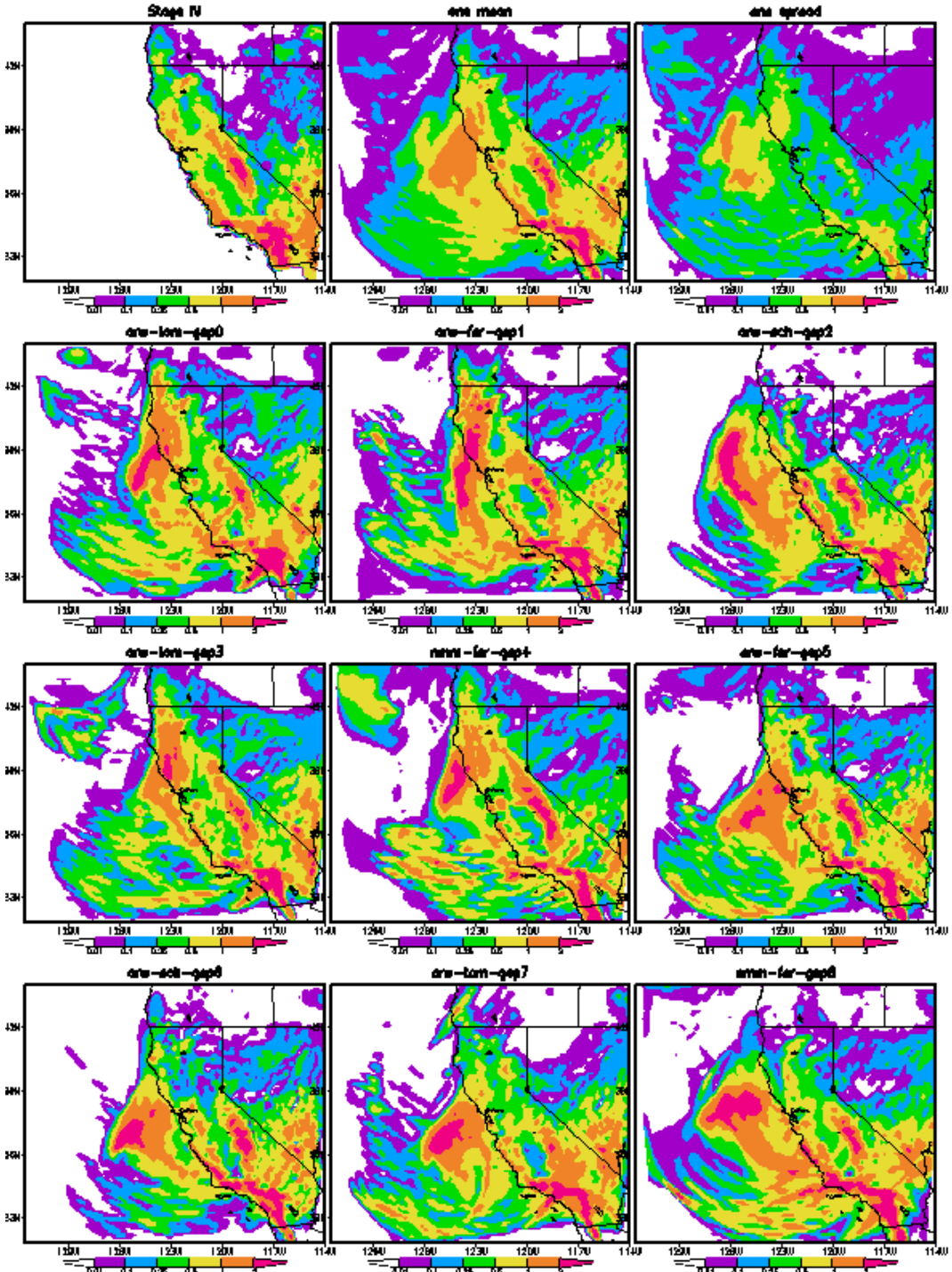
- Outer/inner nest grid spacing 9 and 3 km, respectively.
- 6-h cycles, 120hr forecasts foe the outer nest and 12hr forecasts for the inner nest
- 9 members (listed in the following slide)
- Mixed models, physics & perturbed boundary conditions from NCEP Global Ensemble

QPF

Example of 24-h QPF
9-km resolution

9 members:
ARW-TOM-GEP0
ARW-FER-GEP1
ARW-SCH-GEP2
ARW-TOM-GEP3
NMM-FER-GEP4
ARW-FER-GEP5
ARW-SCH-GEP6
ARW-TOM-GEP7
NMM-FER-GEP8

- WRF-ARW and WRF-NMM
- WRF-ARW runs: Ferrier, Schultz, Thompson microphysics
- WRF-NMM run: Ferrier microphysics



ARRFEX

- Goal: bring researchers and operational forecasters together and evaluate impact of new “experimental” models and tools for forecasting AR events
- Hosted by Hydrologic Prediction Center (HPC)
  - HPC forecasters
  - Environmental Modeling Center (EMC)
  - Eureka Weather Forecast Office (WFO)
  - Northwest River Forecast Center (NWRFC)
  - ESRL Physical Sciences Division (PSD)
  - University of Utah

Experiment Goals

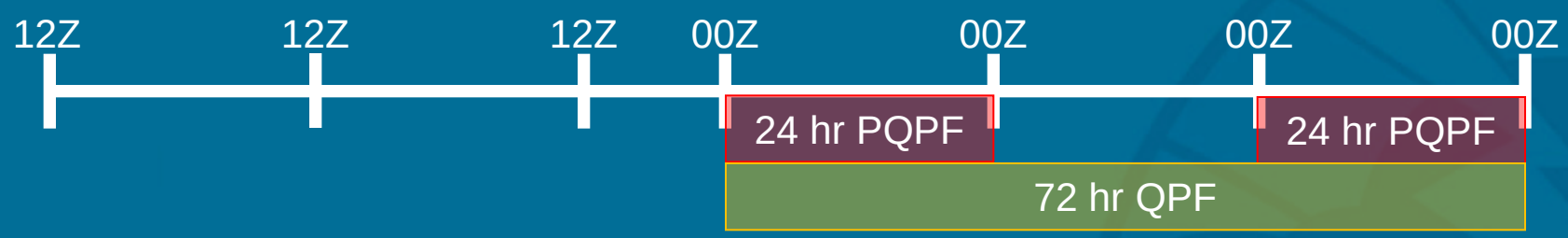
- How can forecasters add value to extreme precipitation forecasts at mid-range (3 – 5 day) timeframes?
- Do probabilistic QPF (PQPF) products provide a viable alternative for increasing awareness of extreme precipitation events?
- Do the high-resolution, multiple-physics schemes of the HMT-ensemble benefit extreme precipitation forecasting?
- Do the experimental datasets (HMT-ensemble, multi-model ensemble, reforecasting data) improve extreme precipitation forecasts?
- What are the biases/strengths/weaknesses of current model guidance in terms of AR forecasting?

Experiment Activities

- Issue 24 h probability of exceedance forecasts for 3” of precipitation at 5 day (120 hour) and 3 day (72 hour) lead times
  - 10% and 40 %
- Issue 72 h total QPF forecast for Days 1-3
  - 4”, 8”, >12”
- Issue an AR-duration (start/stop) forecast for specific West Coast locations
  - 6 hour windows
- Explore uses of known AR forcing mechanisms for use in precipitation forecasting
- Subjective model and forecast evaluation

Numerical Model Guidance

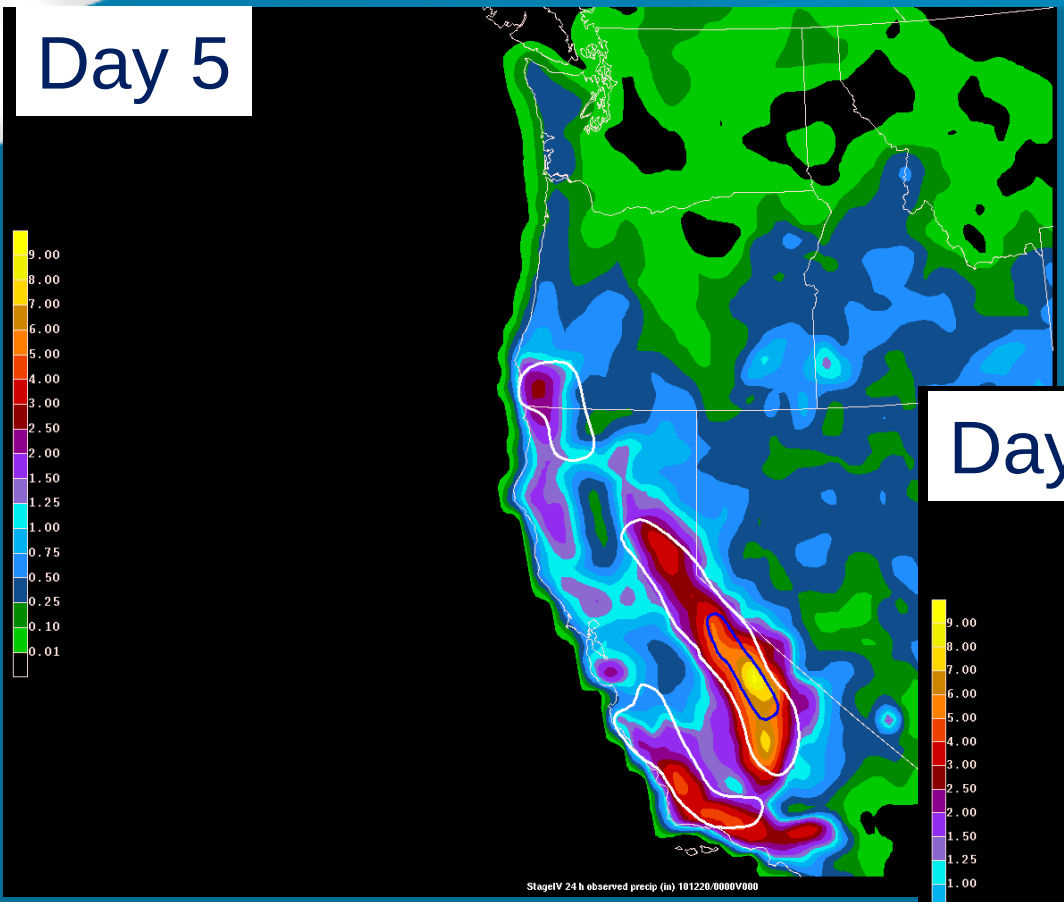
Provider	Model	Resolution	Availability
NCEP	GFS	1.0 deg	180 h
NCEP	GEFS	70 km	180 h
ECMWF	ECMWF	1.0 deg	196 h
ECMWF	ECENS	1.0 deg	196 h
NCEP	NAM	32 km	84 h
CMC	CMCE	1.0 deg	196 h
ESRL/GSD	HMT-Ensemble*	9 km	84 h
ESRL/PSD	Reforecast	32 km	96 h
NCEP/ECMWF/ CMCE	MMENS	70 km	180 h

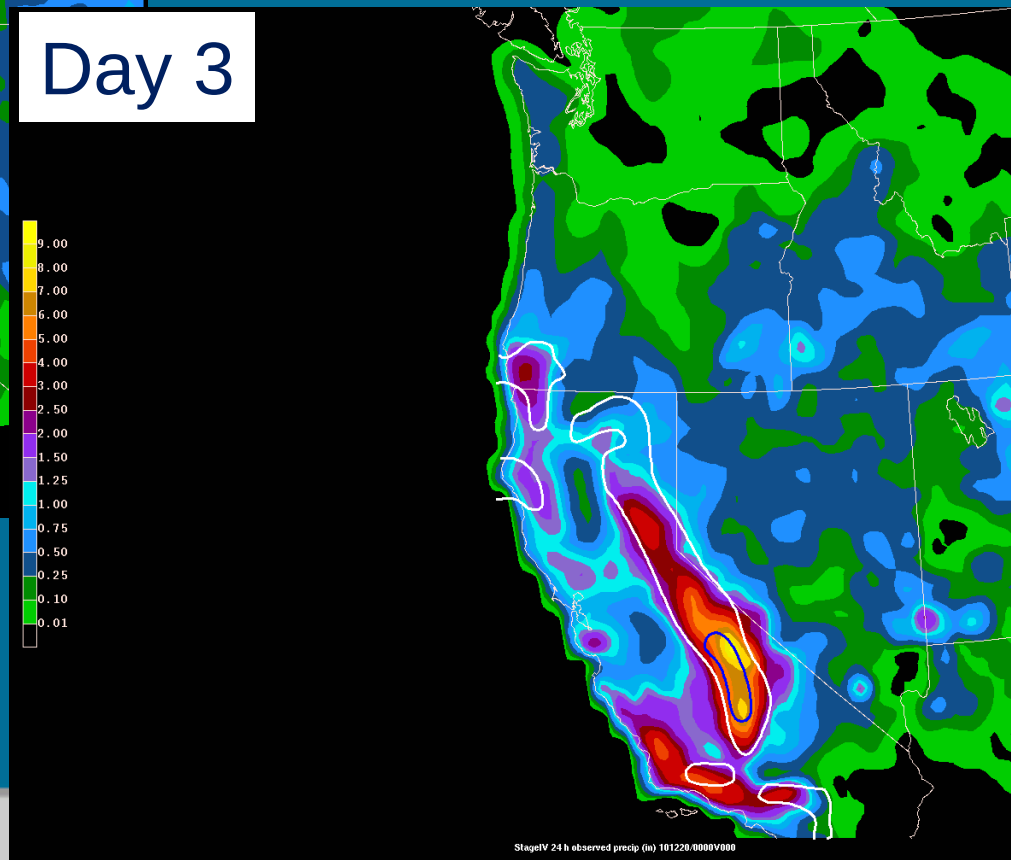


Subjective Verification

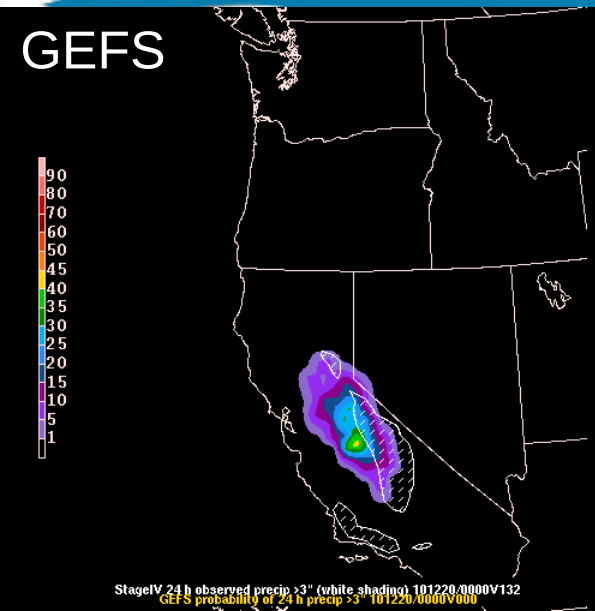
- Compare model QPF and user-generated forecasts to:
  - Stage IV: 32 km or 4 km
  - Atmospheric River Observatory
- Analyze model performance as a method to systematically assess strengths/weaknesses and biases regarding QPF forecasts
  - Does model appear to be consistently better with precip location? Amount? Timing?
  - Does the experimental guidance demonstrate more skill?
- Comparison to archived HPC QPF:
  - Does the inclusion of the experimental guidance appear to lead to better forecasts?
- Evaluate new methods of AR forecasting
  - Can model representations of known AR forcing mechanisms be used as indicators of heavy precipitation events?
  - Can they be used to help identify location and/or timing?
  - Do the experimental datasets provide valuable information to forecasters?

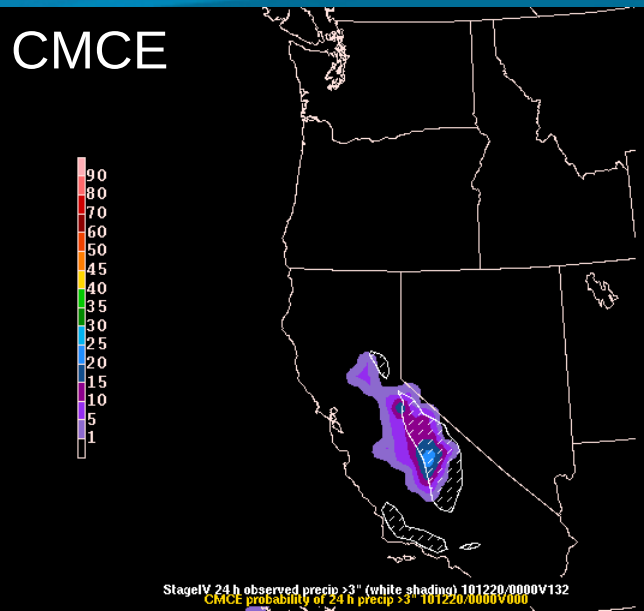
Day 5/3 24 hr probability > 3”

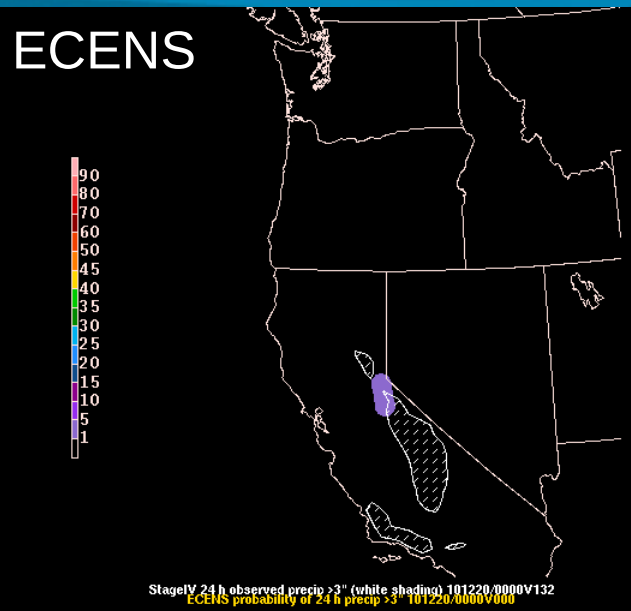


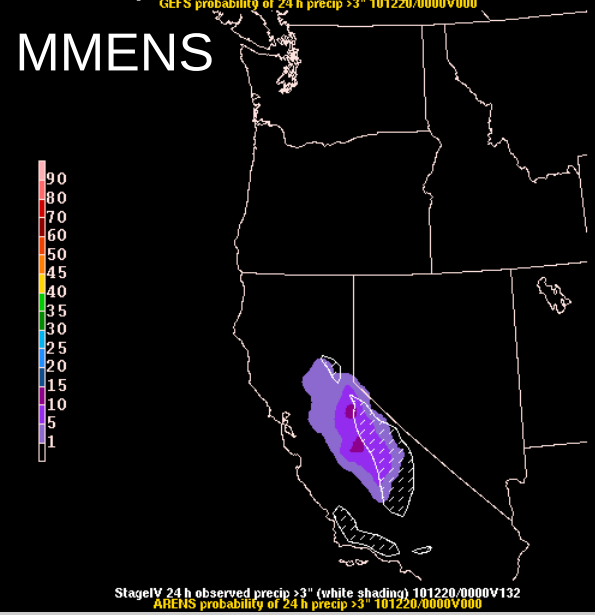


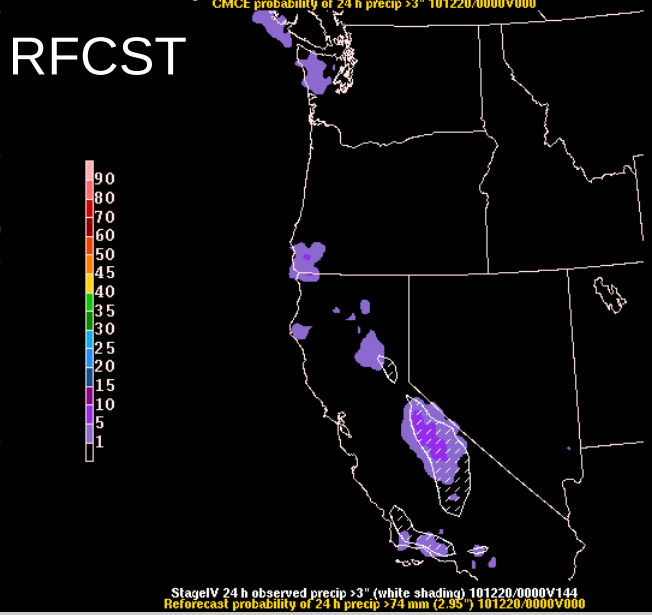
Verification: Day 5



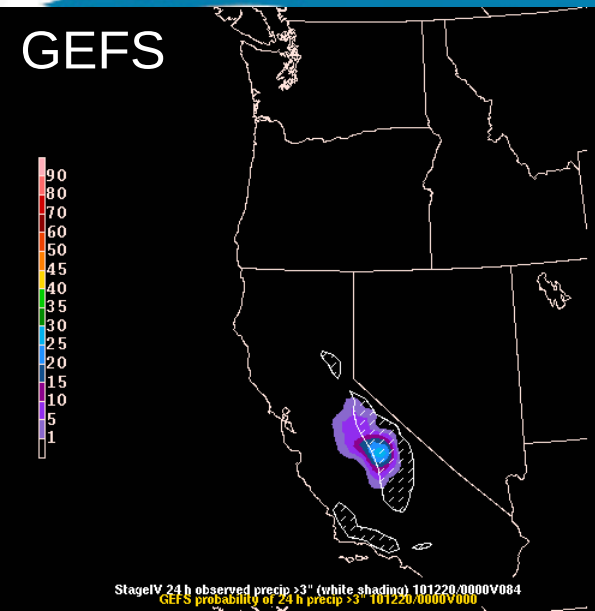


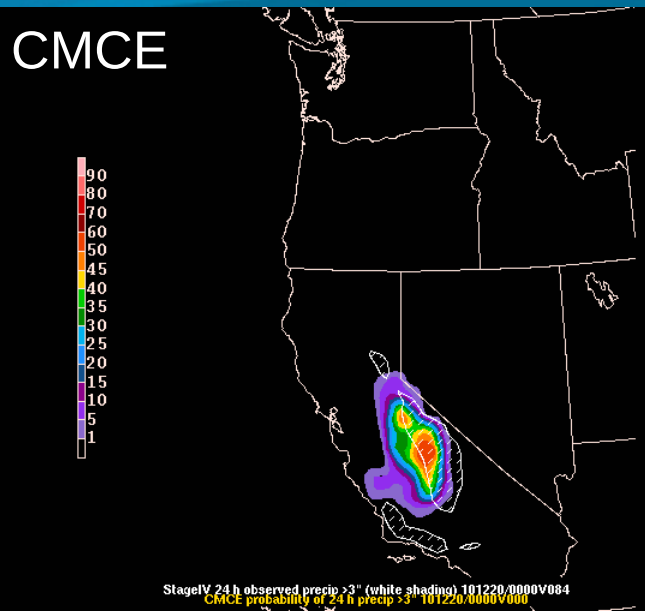


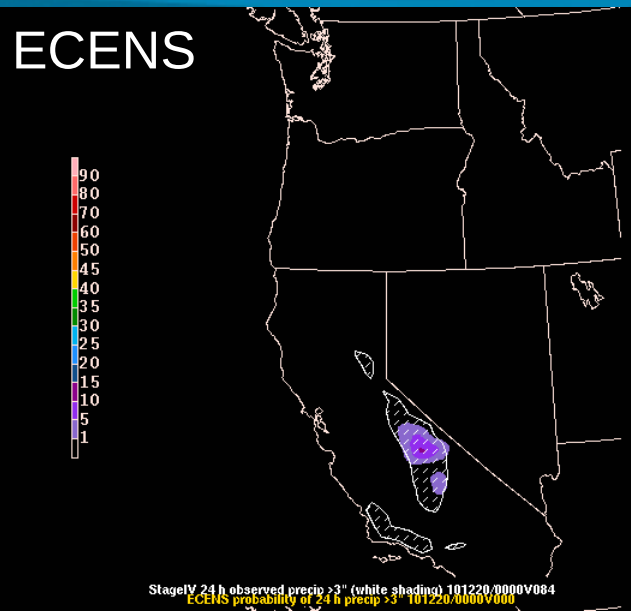


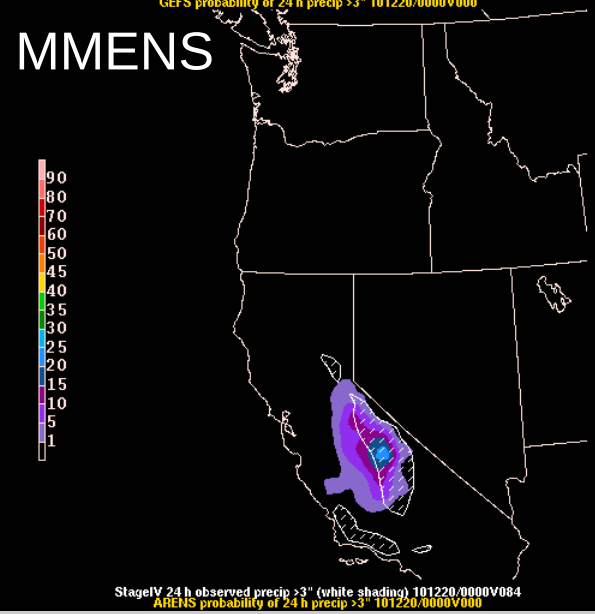


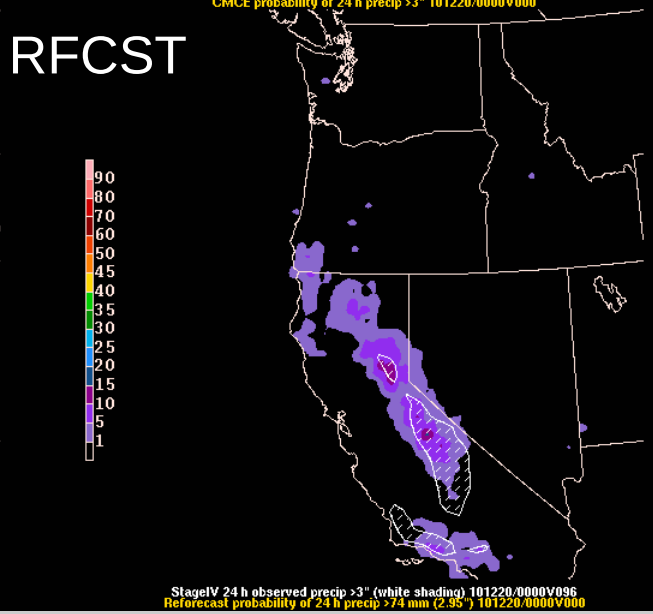
Verification: Day 3

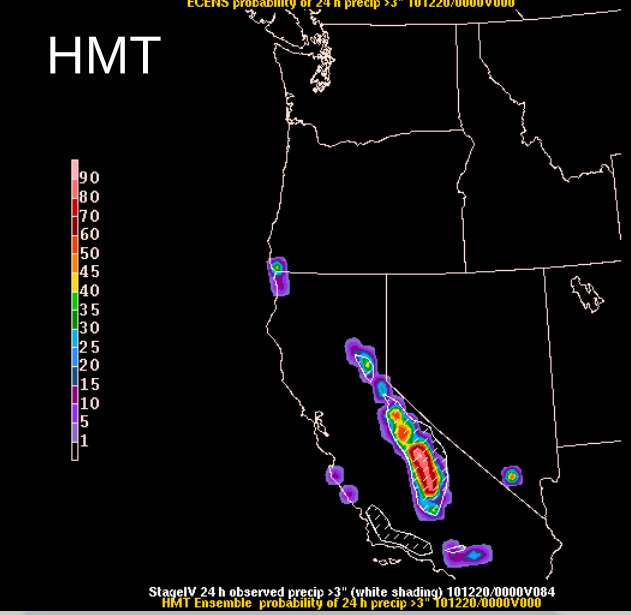












Probability Subjective Verification

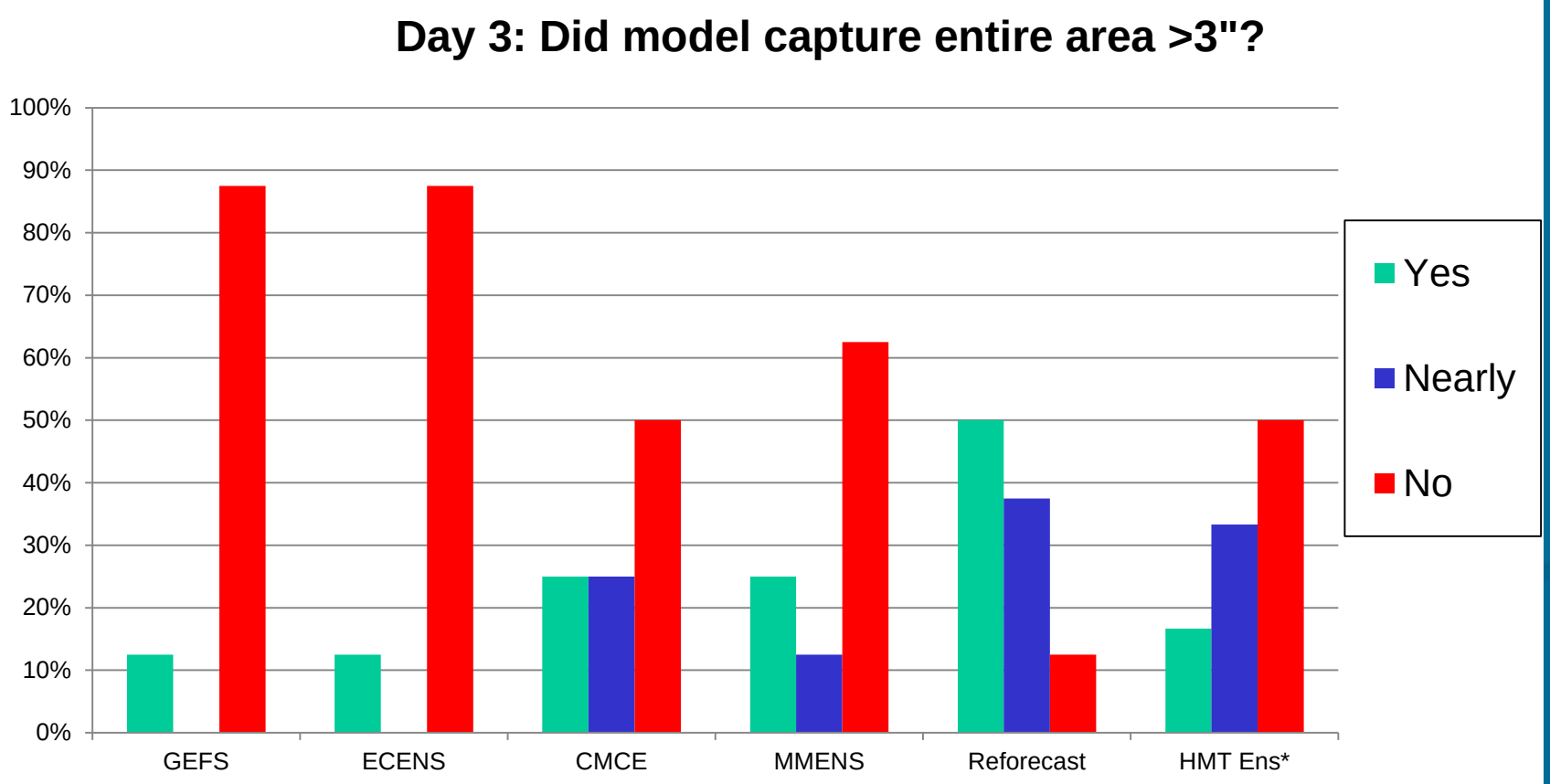
Day 5

- Rate your forecast (Good, Fair, Poor)
  - Explain
- For each of the model probability forecasts, did the probability field capture ALL of the area observed to receive > 3"?
- In your opinion, how did the experimental guidance perform compared to the ECENS?
  - Explain

Day 3

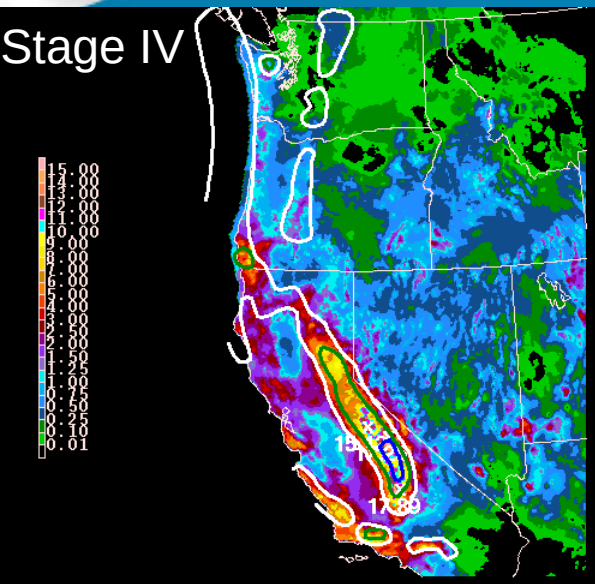
- Rate your forecast (Good, Fair, Poor)
  - Explain
- For each of the model probability forecasts, did the probability field capture ALL of the area observed to receive > 3"?
- In your opinion, how did the experimental guidance perform compared to the ECENS?
  - Explain
- In your opinion, which ensemble probability forecast provided the most useful guidance? Which provided the least? Why?

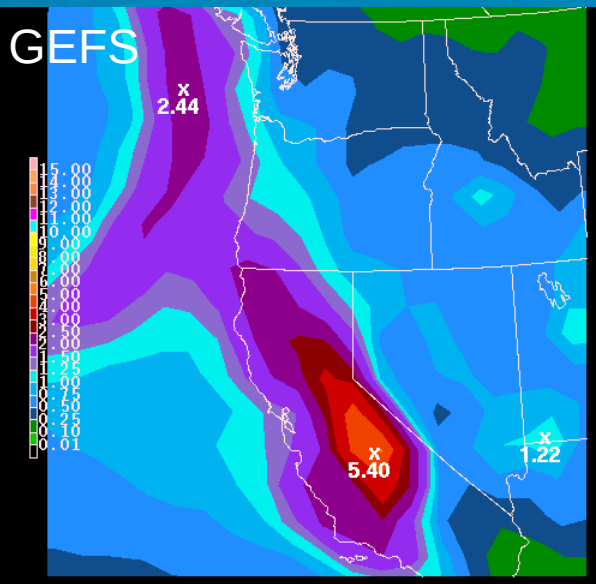
Probability Subjective Verification

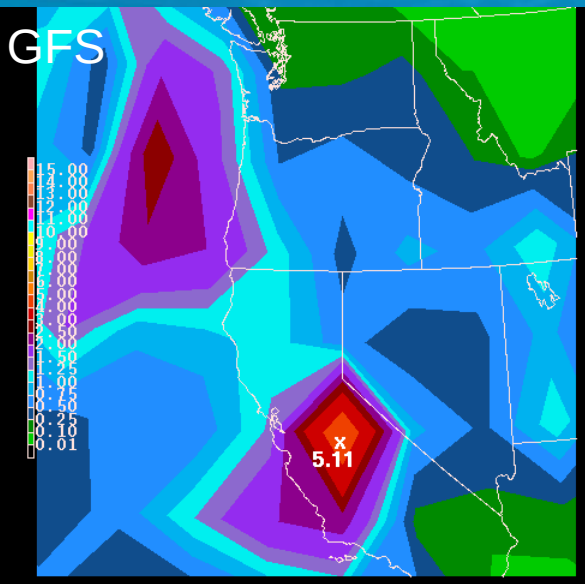


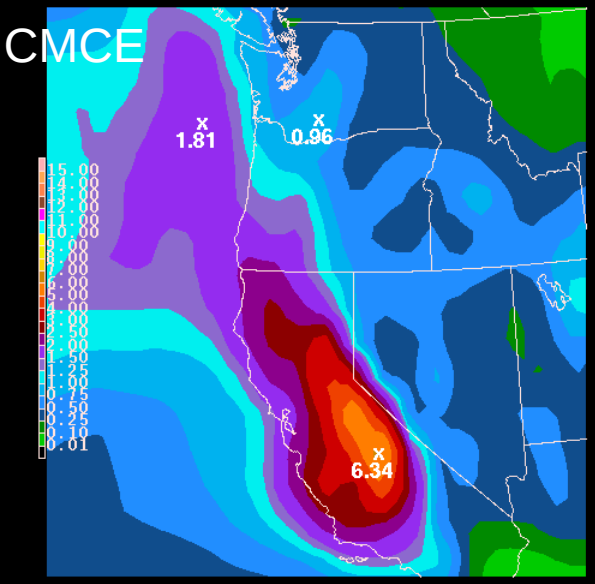
→ Reforecast deemed 'most helpful' in 6 cases (CMCE: 1, HMT: 2)
→ Reforecast and HMT-ENS consistently better than ECENS (when applicable)

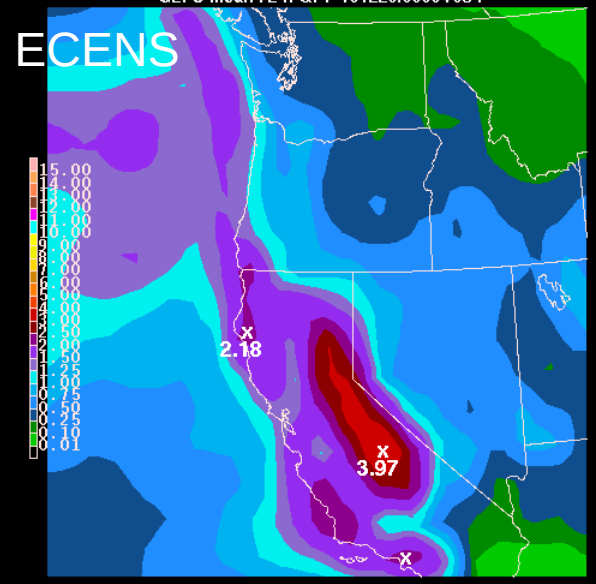
Verification: 72 h QPF

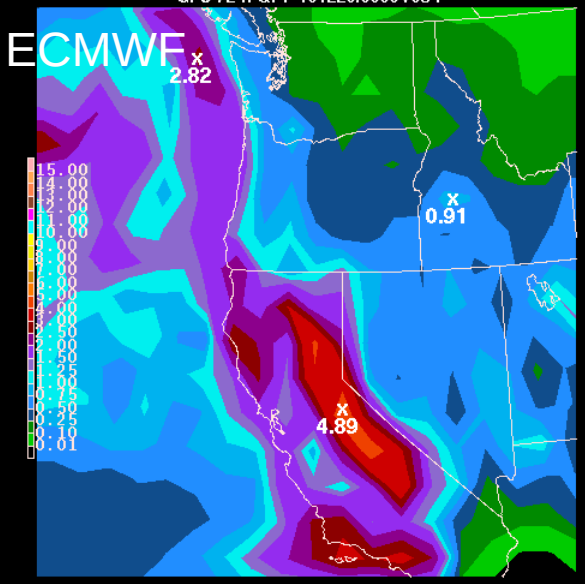






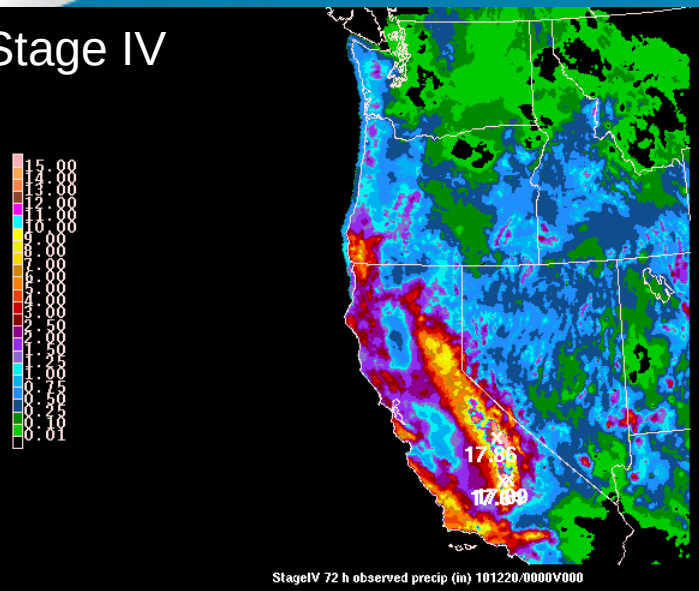




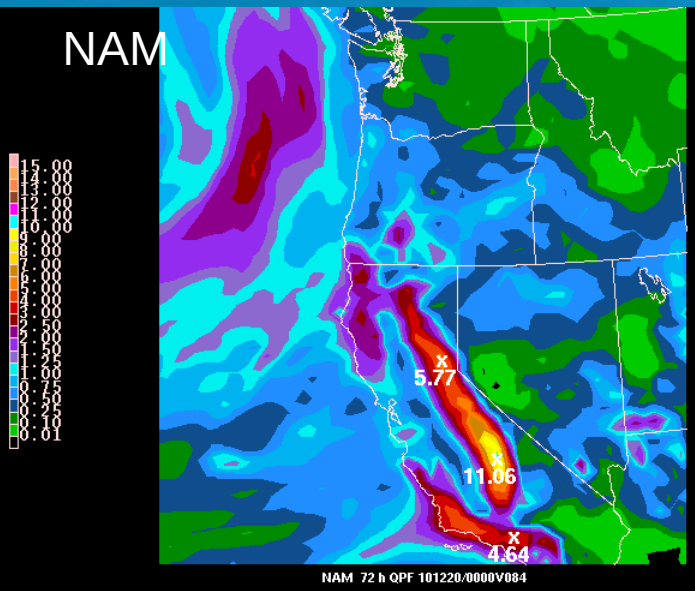


Verification: 72 h QPF

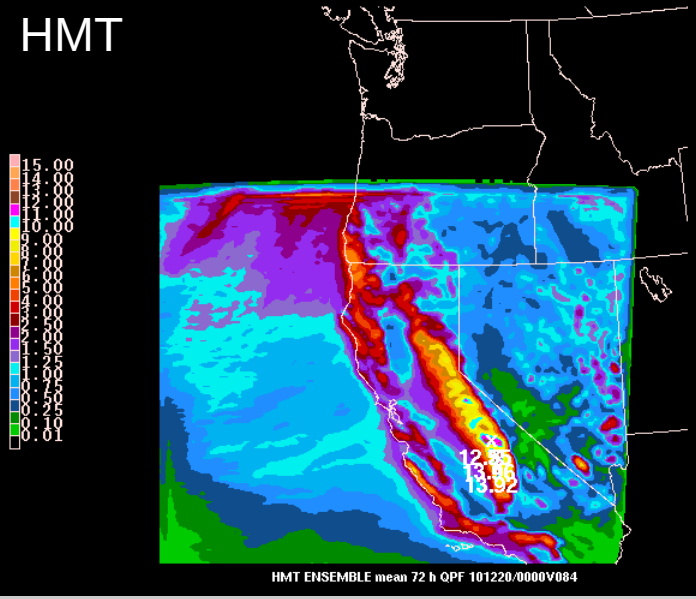
Stage IV



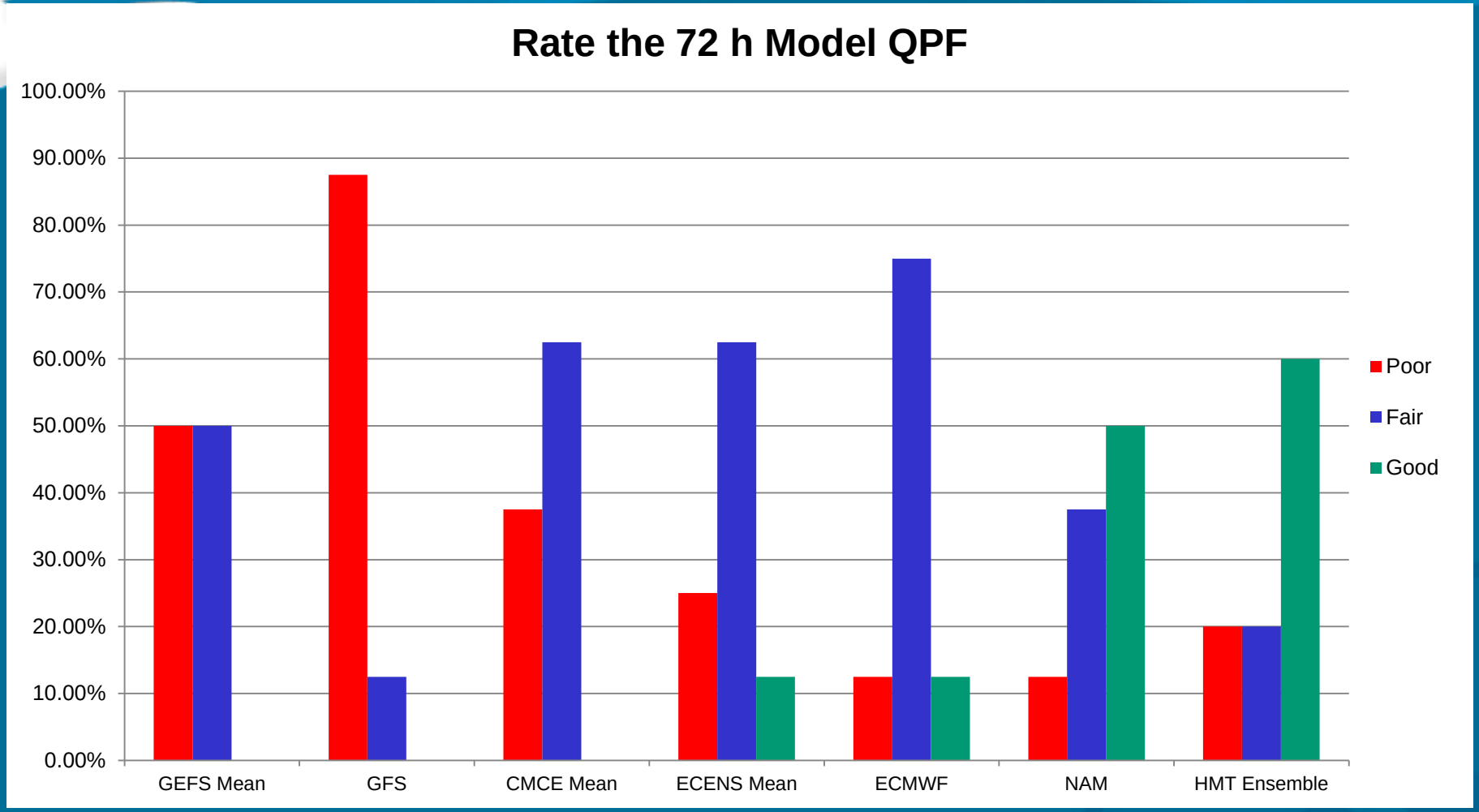
NAM



HMT

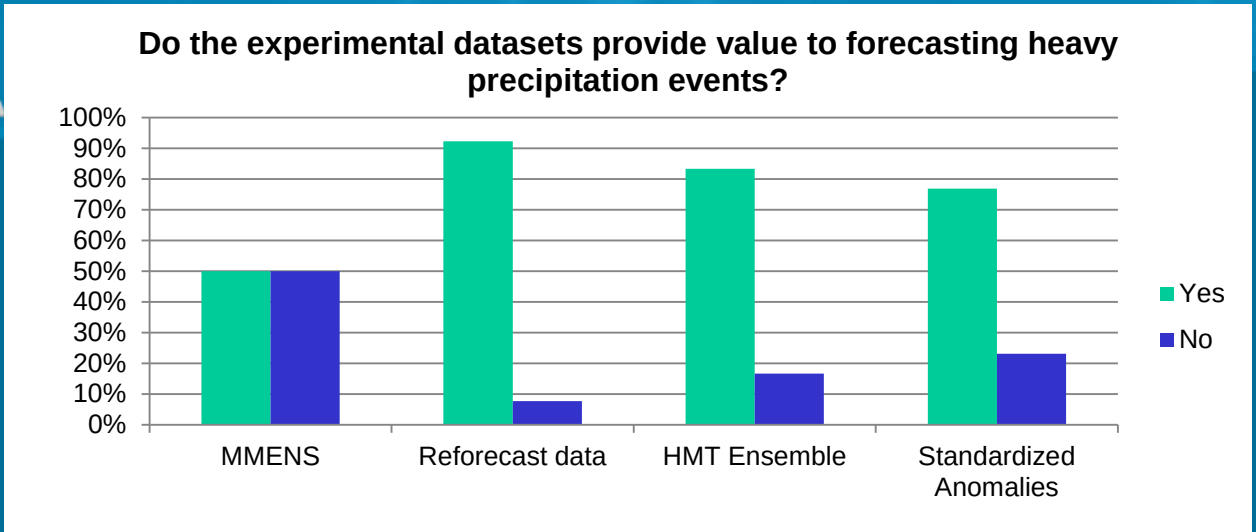


72 h QPF Subjective Verification



→ User forecasts were deemed and 'improvement' over the archived HPC forecast in 7 out of 8 cases.

Experiment Summary



- "Reforecast dataset is very useful."
- "The best guidance was the reforecast data…this is a big improvement over the raw member forecast…I hope this can be implemented operationally at HPC asap."
- "Reforecast seems like a very nice way to try and squeeze more utility out of climatological data…I see good performance in the precip patterns... but have no good feel for the magnitude."
- "…reforecast data seems to miss both the magnitude and distribution in general of the extreme nature of the QPF near the focus (AXIS) of the AR (showing a more dispersed solution)…"

- "The HMT ensemble was about the best in accurately predicting the magnitude and placement of heavy precipitation. The only drawback of this product…is that its domain is too small."
- "…the HMT…was by far the most superior of the guidance we interrogated during this experiment with high-resolution data over the favored topography."
- "…HMT ensemble members clustered well, and the maxima were all false alarms."
- "In particular, the HMT and Reforecast information was very helpful and generally more accurate."

Experiment Summary

- Higher-resolution data is very beneficial, especially in West Coast/terrain driven events
- Surprised by NAM/HMT performance
- Reforecast (PQPF) and HMT-ensemble data largely considered the best guidance, GFS/GEFS/ECENS rated as the worst
  - HMT could be *too* wet (?)
  - Resolution of global ensembles a detriment
  - PQPF seems to be a worthwhile way to explore extreme QPF at mid-range lead times:
    - "The primary thing the PQPF offers is an enhancement in situation awareness showing the threat for possible heavy precipitation in the medium range time scales."
    - "Currently, the viability seems limited, but I think the potential is huge and it's absolutely where we need be going with our precipitation forecasting"
  - Several participants noted that their time in the experiment was beneficial: *interaction, discussion, training, additional insight, product development, etc….*

Next Steps

- Victor Stegemiller (NWRFC) office training session (October 2012)
- Tom Wright (Medford WFO) forecaster training seminar (November 2012)
- Western Region webinar (tentatively mid-November 2012)
- AMS annual meeting oral presentation (January 2013)

Other

- ARRFEX summary report
- Already had ARRFEX data and result requests (Tom, Victory, Jon, Environment Canada)

Backup Slides

Additional Experimental Guidance

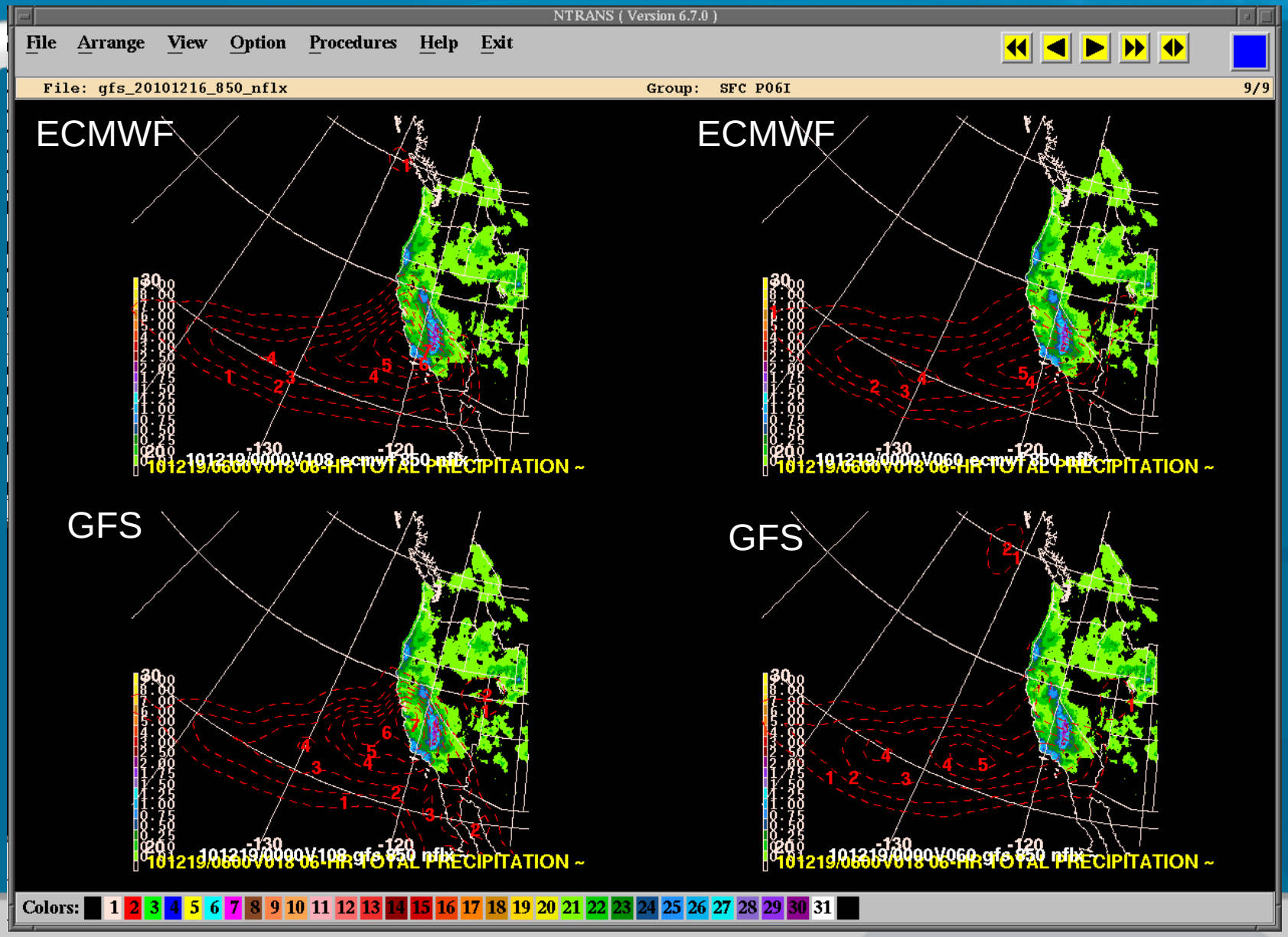
- Multi-Model Ensemble
  - 90 members: 20 CMCE, 20 GEFS, 50 ECENS
  - 70 km resolution
  - Provides a more realistic “true” probability (Hamill, 2012)

- Standardized Anomalies
  - Computed from a 1948-2011 climatology of NCAR/NCEP Reanalysis data
  - Model output interpolated onto the 2.5 degree Reanalysis grid
  - 15 day centered averages used to filter climate data and make anomalies (standard deviations)

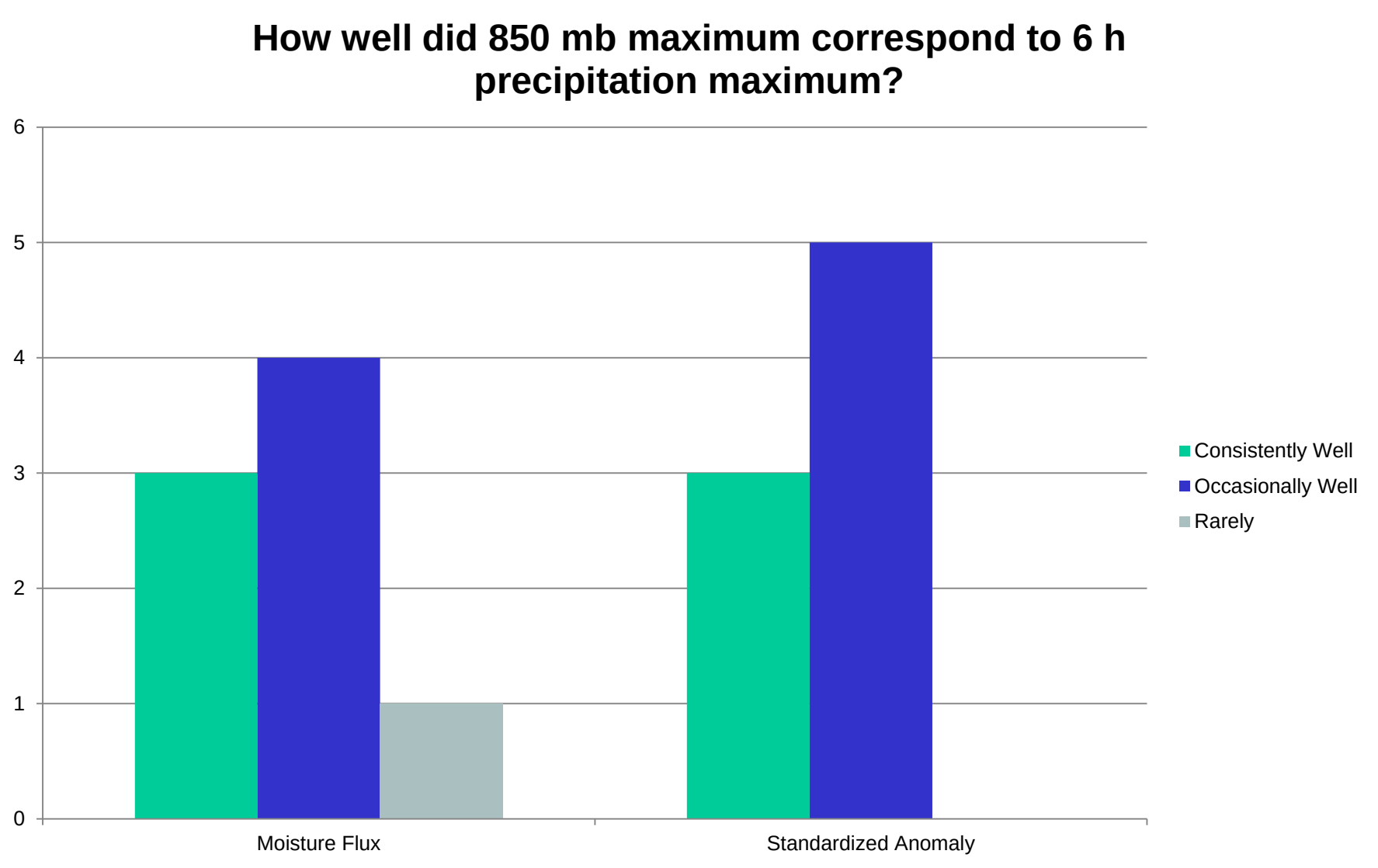
Point Forecast Verification

- When do you forecast the precipitation to start and stop at the given location?
- When do you forecast the heaviest precipitation to start and stop at the given location?
- What is the confidence level of this forecast?
- When do the observations show that the precipitation (as well as the heaviest precipitation) started and stopped at the given location?
- How well do the locations of the maxima of moisture flux correspond the locations of the 6-hour precipitation maxima?
- How well do the locations of the maxima of the standardized anomalies of moisture flux correspond the locations of the 6-hour precipitation maxima?
- Were there any large changes in the guidance compared to the initializations from 2 days earlier? (e.g. do the models show run-to-run consistency?)

Point Forecast Verification



Point Forecast Verification

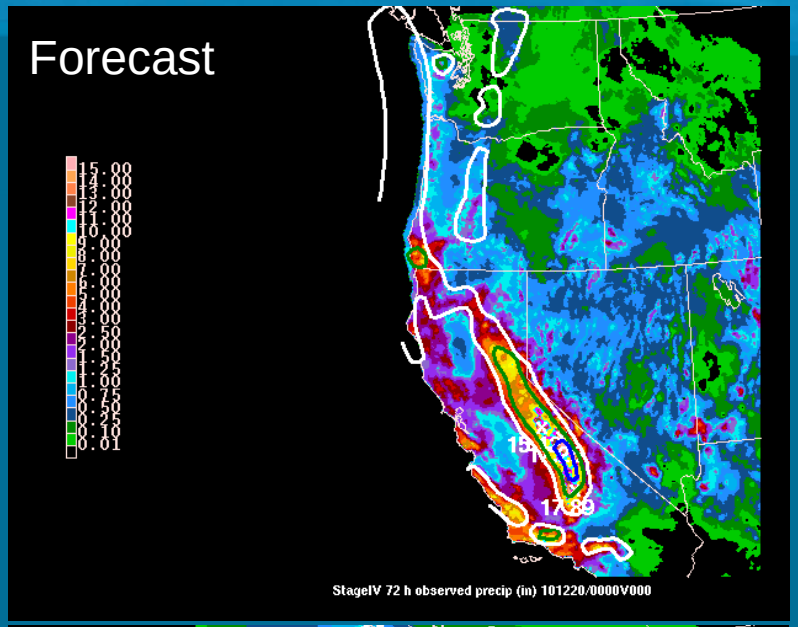


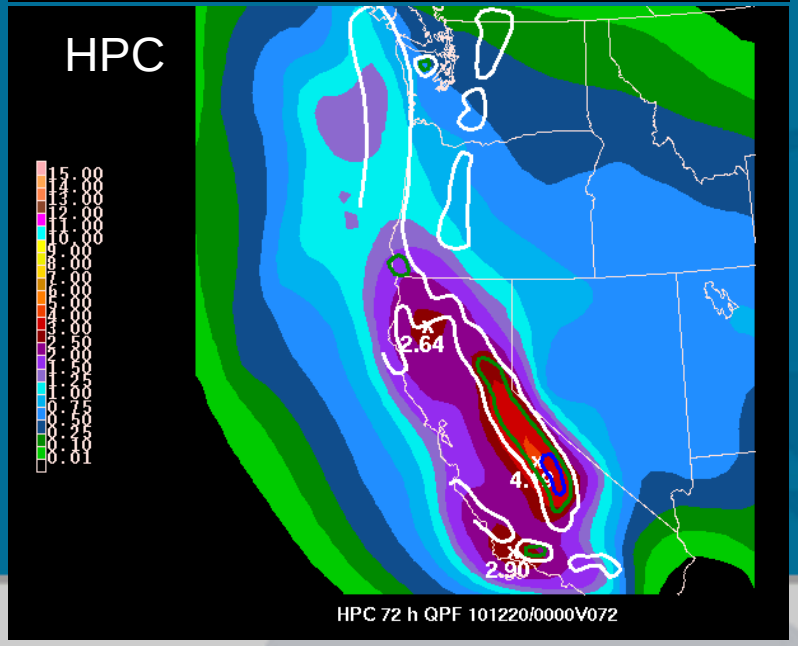
Daily Schedule

- **8:00 am (Monday only)** – Orientation
- **8:30-9:45 am** – Create 24-h PQPFs (00Z to 00Z) for 5 day and 3 day lead times
- **10:00-11:00 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 (00Z to 00Z) 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 at specific locations
- **4:15-4:30 pm** – Group discussion and/or exit questions

72 h QPF Subjective Verification

- Rate your forecast (Good, Fair, Poor)
  - Explain
- Does the group's forecast appear to be an improvement of or a degradation of (e.g. quantitatively or spatially) of the archived HPC forecast?
  - Explain
- Rate the QPF output of each model forecast (Good, Fair, Poor)





Experiment Summary

- In your opinion, did the experimental guidance provide benefit to the forecasting of heavy precipitation events at the mid-range (3 and 5 day) time frame?
- What is your opinion of the higher resolution guidance (HMT-ensemble and NAM) in terms of forecasting precipitation in atmospheric river events?
- Was there a model or guidance product that you felt performed the best during the experiment? Worst?
- What is your opinion of the viability of PQPF in terms of forecasting heavy precipitation events at mid-range lead times? Do they provide any additional benefit compared to the traditional deterministic QPF products?
- Do you feel using known AR characteristic parameters, such as precipitable water or moisture flux, provide benefit to forecasting precipitation timing, location and amount?
- In your opinion, what is the most crucial issue with forecasting West Coast heavy precipitation events?

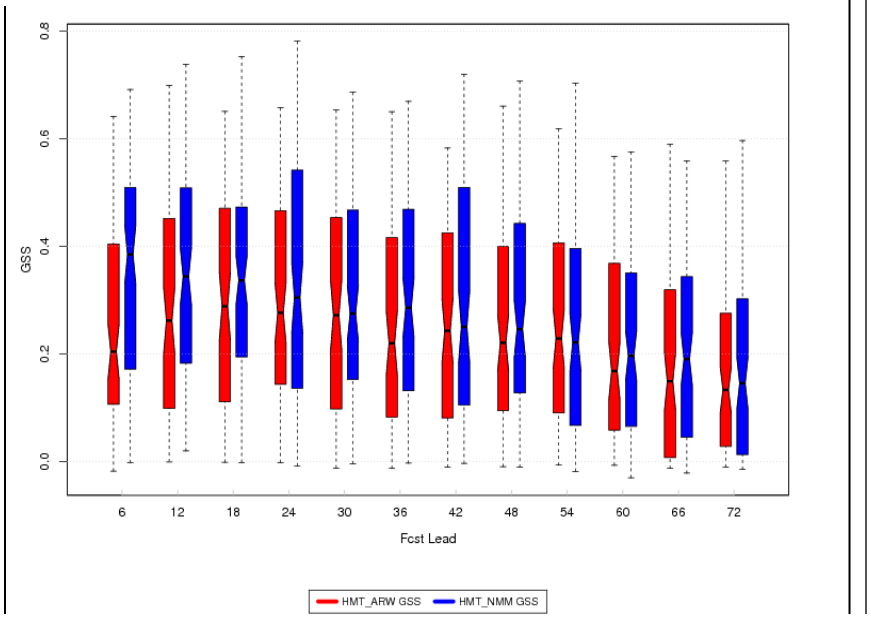
Next Steps

- Analyze ARRFEX participant survey results (Lead: Tom Workoff)
- Conduct objective and spatial verification of the ARRFEX data (Lead: Ellen Sukovich)
  - Probabilities (including reforecast HMT-Ensemble)
  - 24 h QPF (including reforecast mean and each HMT ensemble member)
  - 72 h QPF
- Continue to investigate the reforecast dataset for AR events and precipitation (Lead: Ben Moore)
  - Differing timeframes for QPF (e.g. 72 h and 6 h)
  - Use other parameters as analogs
- Investigate various model parameters and QPF duration, location, and intensity (Lead: Tom Workoff)
- Study MJO and AR relationships (Lead: Tom Workoff & Mike Bodner)
  - CPC plots → Cases from ARRFEX show a tendency to occur in MJO phases 5-6-7

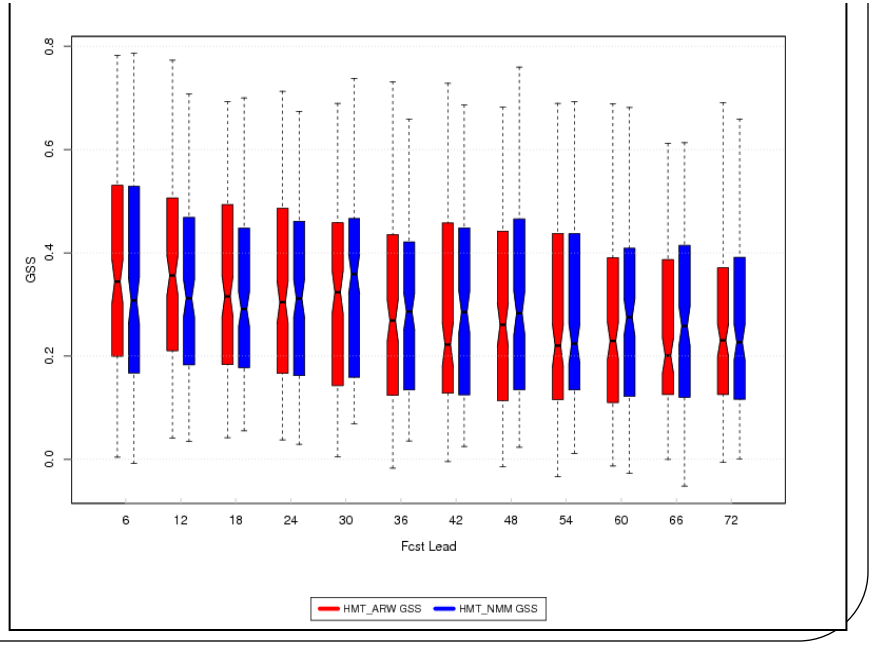
Results – Hot Start Implications

- During the 2011-2012 HMT winter exercise, ARW ensemble model members (red) were hot-started using LAPS during the period March 15-April 15, but were not hot-started before that.
- The NNM members (blue) were not hot started during either period.
- To assess the impact of this LAPS feature, we compare verification results during the later period with an earlier month with a similar rainfall regime.

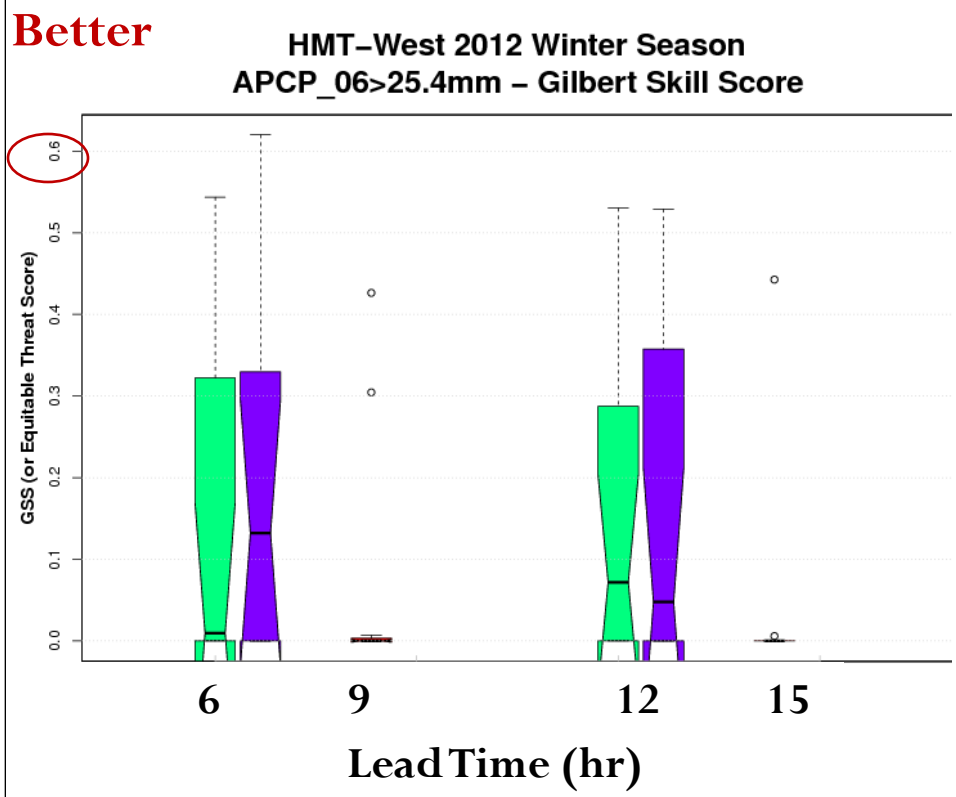
GSS 1/15-2/15 (w/o hot start)

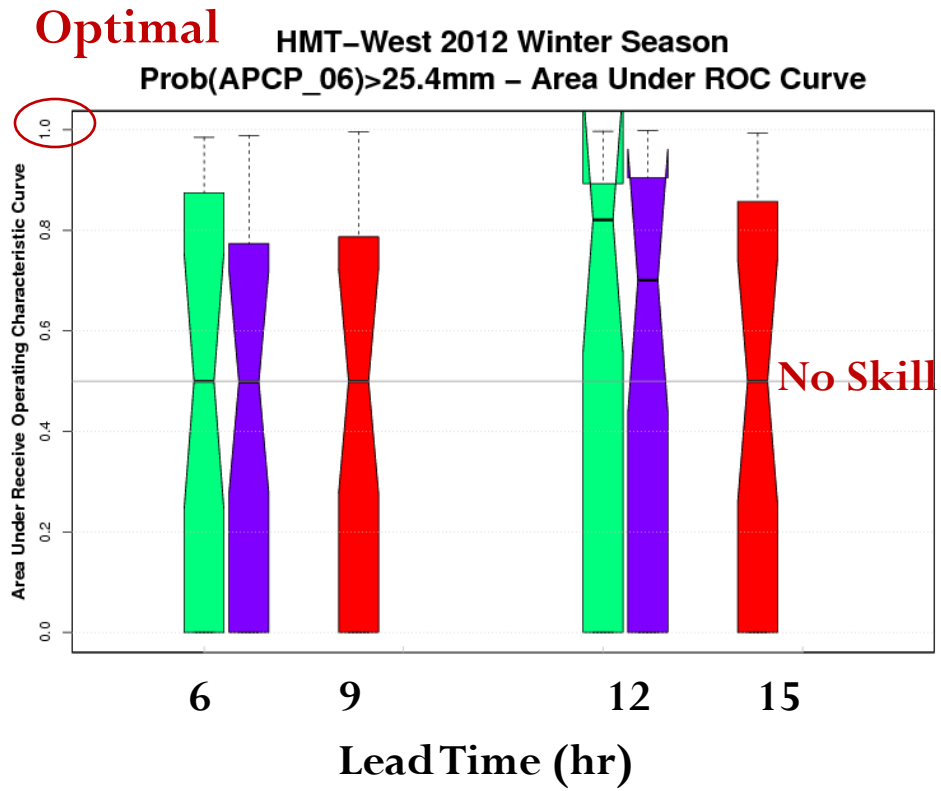


GSS 3/15-4/15 (with hot start)



Results - QPF





AFWA (4km - 10 member) **HMT/GSD** (3km - 9 member) **SREF** (32km - 21 member)

- **6hr Accum Precip > 1"** - All scores are low – partially due to sample-size but SREF (32km) shows very little skill whereas HMT & AFWA (3 & 4km) ensembles can score as high at 0.6
- **Prob (6hr Accum Precip) > 1"** - All scores are low at 6hr lead time – There are differences in the median AFWA and SREF values at 12 hr leads that may be significant