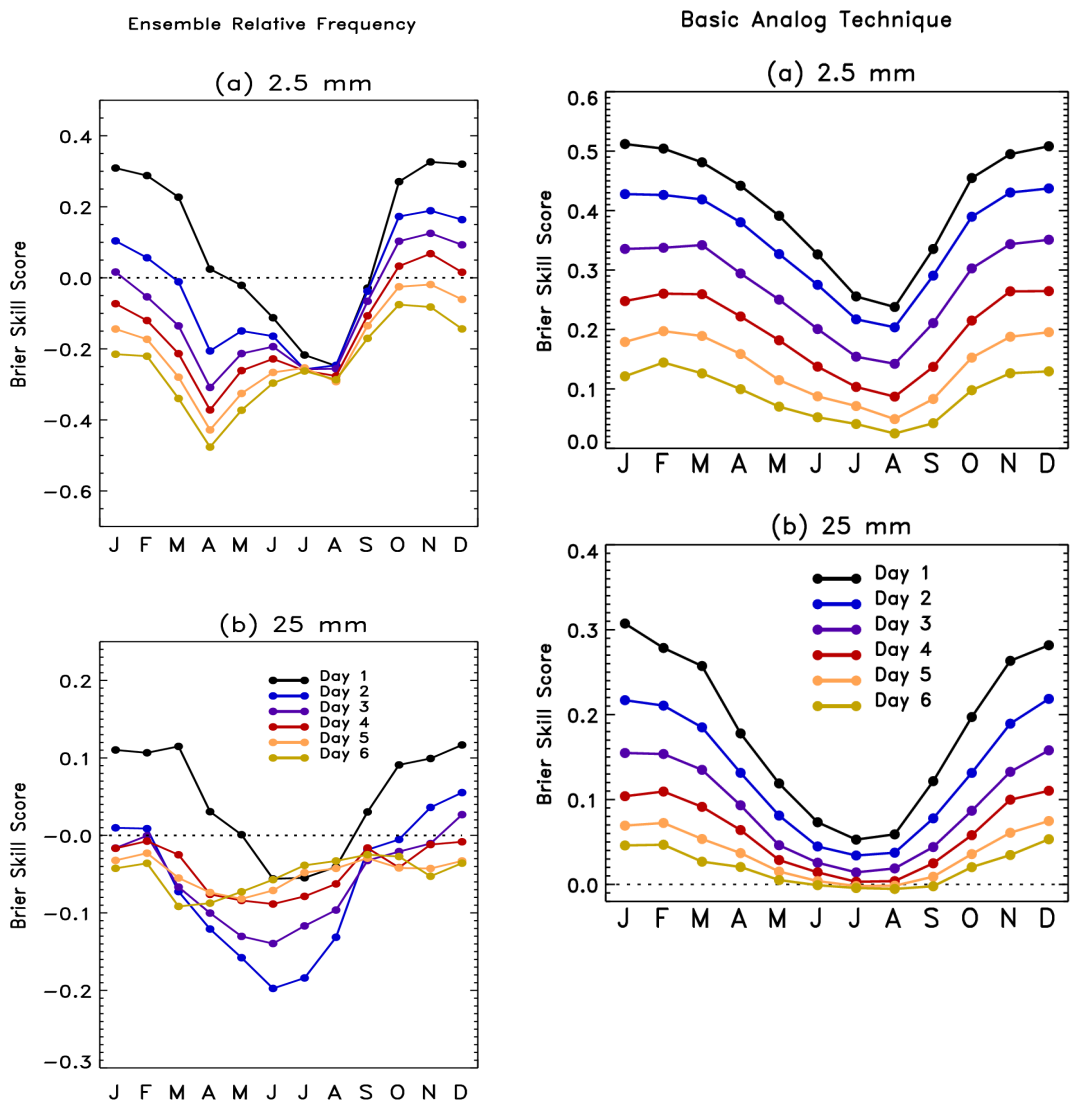
NOAA Earth System Research Laboratory

Verification and validation of precipitation forecasts from the 2nd-generation GEFS reforecast

Tom Hamill
NOAA ESRL, Physical Sciences Division
tom.hamill@noaa.gov

also: Jeff Whitaker, Gary Bates, Don Murray

Why reforecast?



Potential for large improvements in skill from post-processing to remove systematic errors, statistically downscale.

Here, “analog” approach used as discussed in Hamill and Whitaker, MWR, 2006.

But: first-generation reforecast used 1998 T62 GEFS; out of date, not many variables saved.

GEFS reforecast v2 basics

- Seeks to mimic GEFS operational configuration as of February 2012.
- Reforecasts produced every day, for 1984120100 to current (actually, working on 2011 and 2012 now).
- Each 00Z, 11-member forecast, 1 control + 10 perturbed.
- CFSR initial conditions (GSI) + ETR perturbations (cycled with 10 perturbed members). After ~ 22 May 2012, initial conditions from hybrid EnKF/GSI.
- Resolution: T254L42 to day 8, T190L42 from days 7.5 to day 16.
- Fast data archive at ESRL of 99 variables, 28 of which stored at original ~1/2-degree resolution during week 1. All stored at 1 degree. Also: mean and spread to be stored.
- Full archive at DOE/Lawrence Berkeley Lab, where data set was created under DOE grant.

Status

- 00Z reforecasts 1985-2010 completed.
- 2011-12 and archiving real-time GEFS in progress.
- We are a few weeks away from opening our ftp/web site (ESRL, fast access, limited data, + DOE, slow access with full data set) to full public usage.
- We have written some select fields of reforecast data to external hard disks for important customers, e.g., OHD, CPC/EMC in order to jumpstart their product development.
- We will be working to convert our web-based experimental precipitation forecast products over to the new reforecast data set (several months before this is ready; aiming for next cool season).
- Several other groups are working on product development, e.g.,for improved hurricane and tornado forecasts.

Data to be readily available from ESRL

Table 1: Reforecast variables available for selected mandatory and other vertical levels. Φ indicates geopotential height, and an X indicates that this variable is available from the reforecast data set at 1-degree resolution; a Y indicates that the variable is available at the native ~0.5 degree resolution. AGL indicates "above ground level."

Vertical Level	U	V	T	Φ	q	Wind Power
10 hPa	X	X	X	X		
50 hPa	X	X	X	X		
100 hPa	X	X	X	X		
200 hPa	X	X	X	X		
250 hPa	X	X	X	X		
300 hPa	X	X	X	X	X	
500 hPa	X	X	X	X	X	
700 hPa	X	X	X	X	X	
850 hPa	X	X	X	X	X	
925 hPa	X	X	X	X	X	
1000 hPa	X	X	X	X	X	
$\sigma \cong 0.996$	X	X		X		
$\sigma \cong 0.987$	X	X		X		
$\sigma \cong 0.977$	X	X		X		
$\sigma \cong 0.965$	X	X		X		
80m AGL	X,Y	X,Y				X,Y

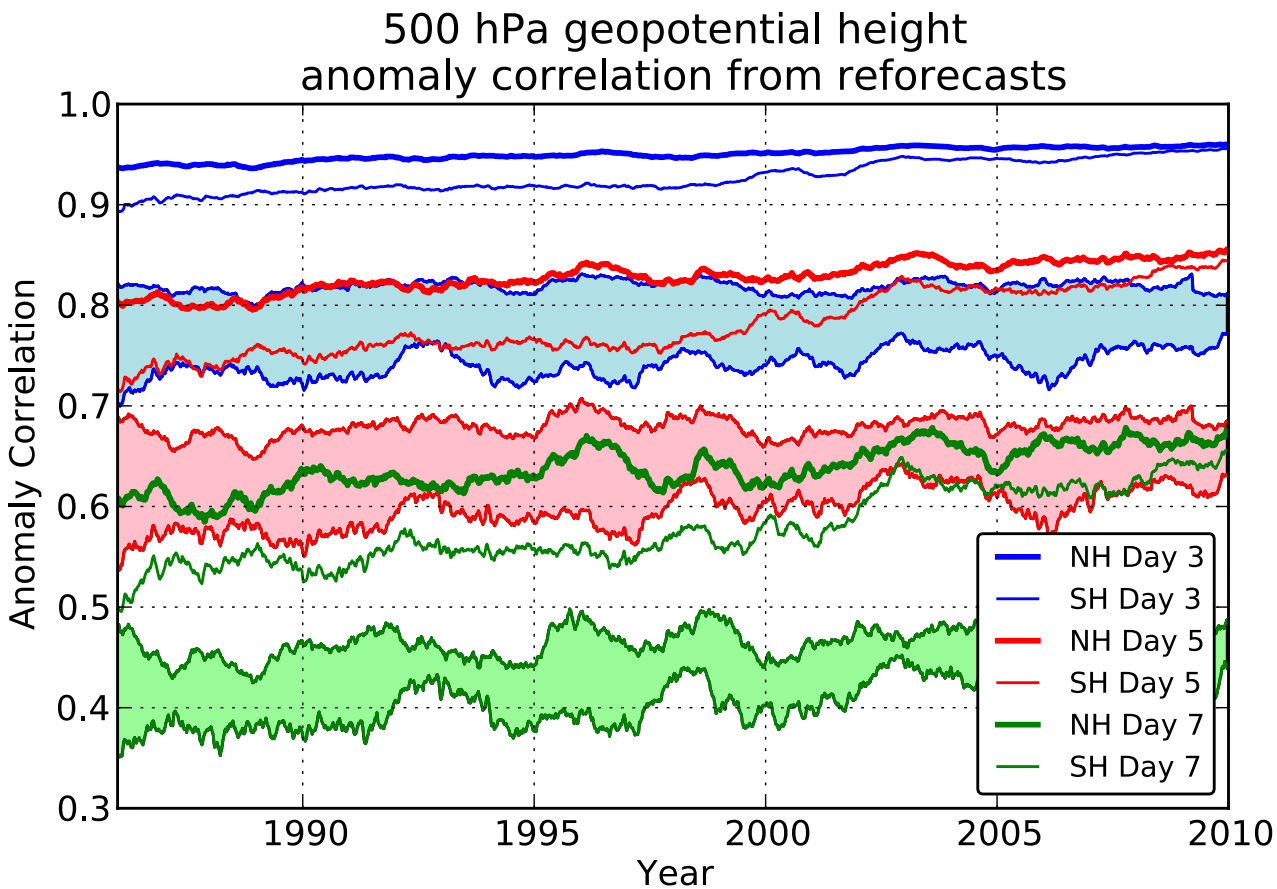
Also: hurricane ATCF files

Data to be readily available from ESRL

Table 2: Single-level reforecast variables archived (and their units). Where an [Y] is displayed, this indicates that this variable is available at the native ~0.5-degree resolution as well as the 1-degree resolution.

Variable (units)
Mean sea-level pressure (Pa) [Y]
Skin temperature (K) [Y]
Soil temperature, 0.0 to 0.1 m depth (K) [Y]
Volumetric soil moisture content 0.0 to 0.1 m depth (fraction between wilting and saturation) [Y]
Water equivalent of accumulated snow depth (kg m^{-2}, i.e., mm) [Y]
2-meter temperature (K) [Y]
2-meter specific humidity (kg kg^{-1} dry air) [Y]
Maximum temperature (K) [Y]
Minimum temperature (K) [Y]
10-m u wind component (ms^{-1}) [Y]
10-m v wind component (ms^{-1}) [Y]
Total precipitation (kg m^{-2}, i.e., mm) [Y]
Water runoff (kg m^{-2}, i.e., mm) [Y]
Average surface latent heat net flux (W m^{-2}) [Y]
Average sensible heat net flux (W m^{-2}) [Y]
Average ground heat net flux (W m^{-2}) [Y]
Sunshine
Convective available potential energy (J kg^{-1}) [Y]
Convective inhibition (J kg^{-1}) [Y]
Precipitable water (kg m^{-2}, i.e., mm) [Y]
Total-column integrated condensate (kg m^{-2}, i.e., mm) [Y]
Total cloud cover (%)
Downward short-wave radiation flux at the surface (W m^{-2}) [Y]
Downward long-wave radiation flux at the surface (W m^{-2}) [Y]
Upward short-wave radiation flux at the surface (W m^{-2}) [Y]
Upward long-wave radiation flux at the surface (W m^{-2}) [Y]
Potential vorticity on θ = 320K isentropic surface (K m^2 kg^{-1} s^{-1})
U component on 2 PVU (1 PVU = 1 K m^2 kg^{-1} s^{-1}) isentropic surface (ms^{-1})
V component on 2 PVU isentropic surface (ms^{-1})
Temperature on 2 PVU isentropic surface
Pressure on 2 PVU isentropic surface

500 hPa Z anomaly correlation
(from deterministic control)



Lines w/o filled colors for second–generation reforecast (2010, T254)

Lines with filled colors for first-generation reforecast (1998, T62).

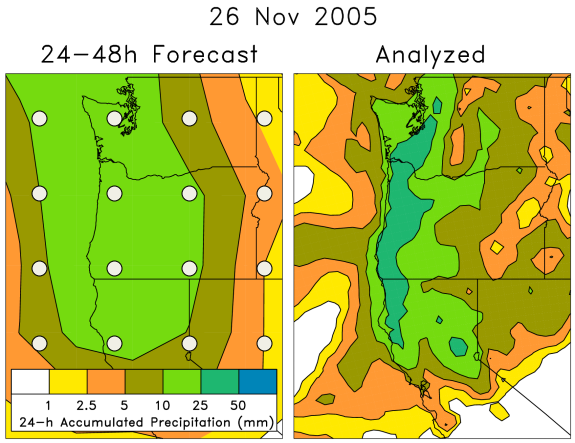
Perhaps a 1.5-2.5 day improvement.

How does the skill of post-processed forecasts change from reforecast v1 to v2?

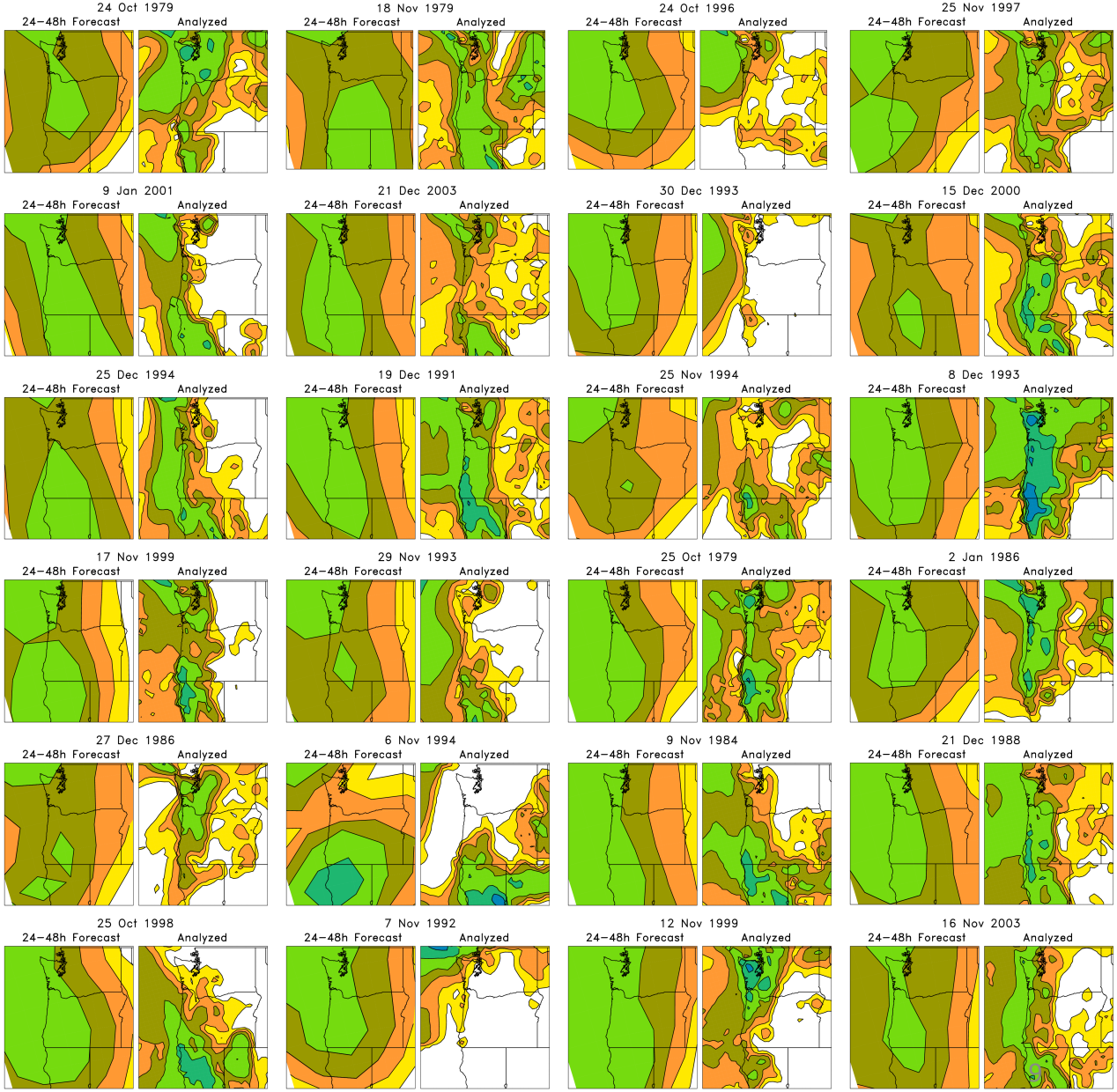
- Let's look at post-processed 24-h precipitation forecast guidance.
- Will use a "rank analog" approach for post-processing (description to follow)

Basic analog technique for statistical downscaling (here, v1)

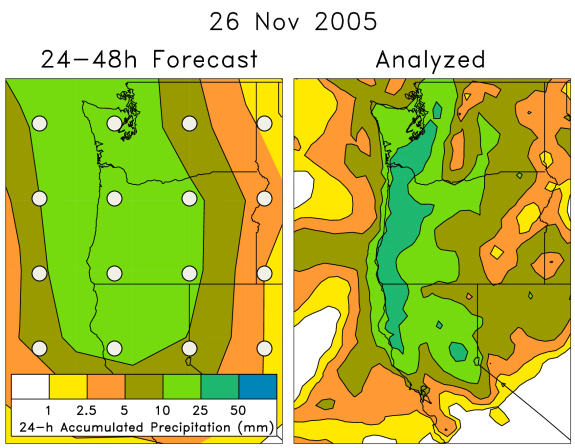
Today's ens. mean forecast + subsequent analyzed precipitation

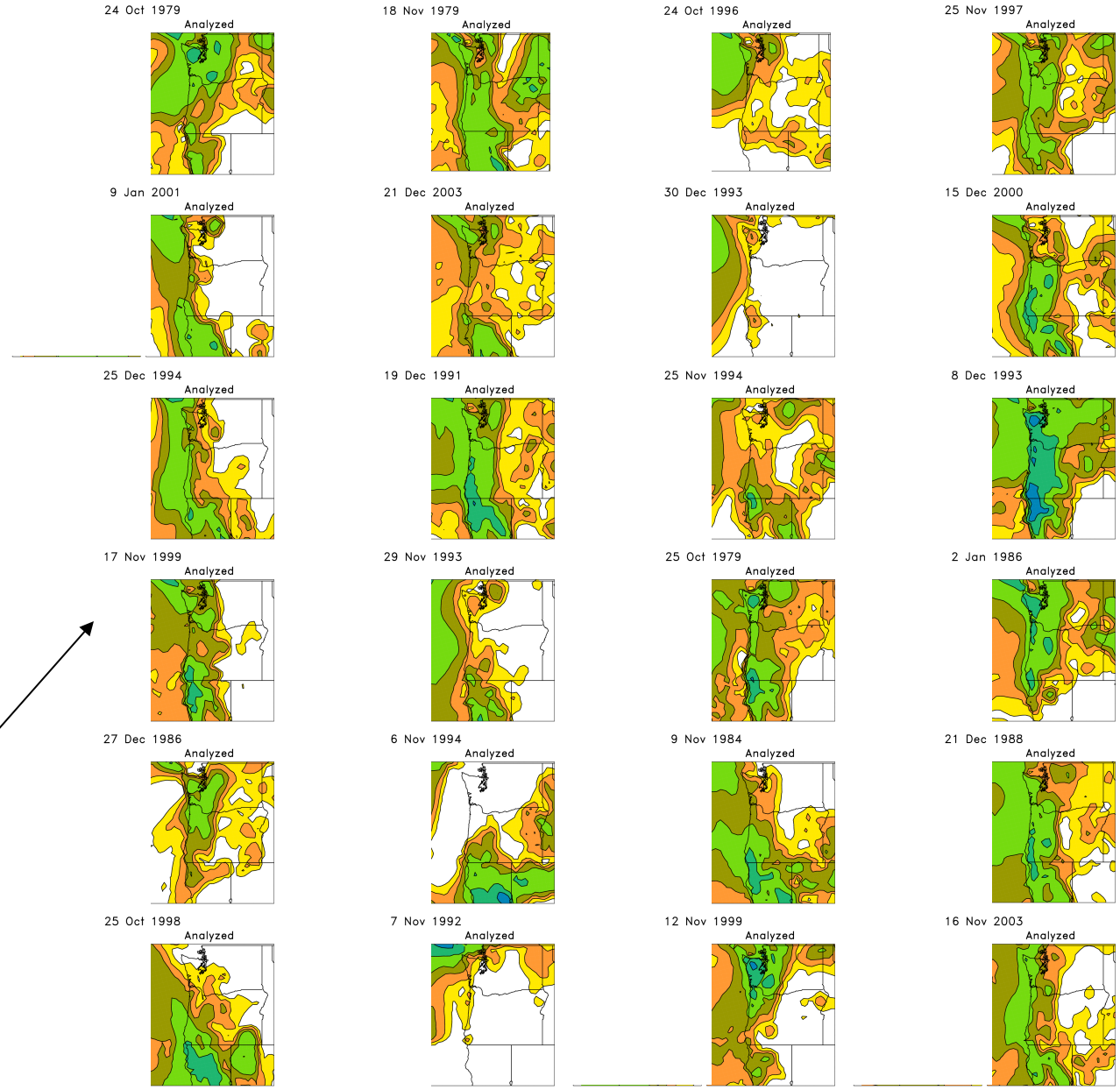


On the left are old forecasts similar to today's ensemble-mean forecast. For making probabilistic forecasts, form an ensemble from the accompanying analyzed weather on the right-hand side.



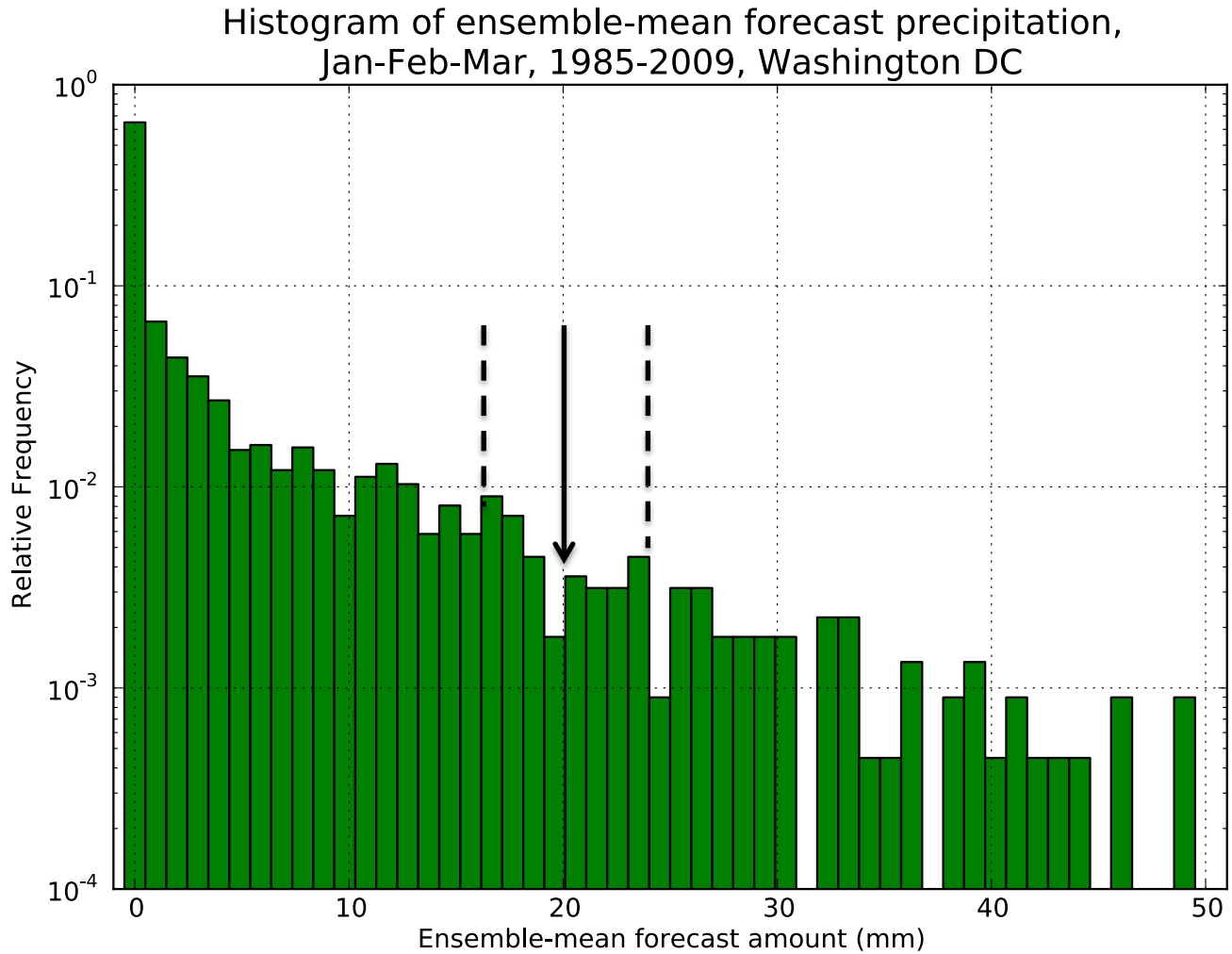
Analog technique for statistical downscaling





Form an ensemble from these

Problem with basic analog reforecast technique



Say today's forecast is for 20 mm. There are more forecasts slightly less than 20 mm than slightly more than 20 mm.

Assuming correlation between forecast and observations, analogs will be biased toward lower precipitation amounts.

“Rank” analog procedure

- Convert precipitation forecast time series to ranks:

today’s

```
x   = [0,   0,    7, 15,   1,  3, 6, 4,   1, 2, 12, 5, 6, 8]

x(r)= [1.5, 1.5, 11, 14, 3.5,  6, 9, 7, 3.5, 5, 13, 8,10,12]
```

“Rank” analog procedure

- Convert precipitation forecast time series to ranks:

today’s

```
x    = [0,   0,    7, 15,   1,  3, 6, 4,   1, 2, 12, 5, 6, 8]

x(r)= [1.5, 1.5, 11, 14, 3.5,  6, 9, 7, 3.5, 5, 13, 8, 10, 12]
```

with standard analog, these would be the two forecasts with the closest values.

“Rank” analog procedure

- Convert precipitation forecast time series to ranks:

today’s

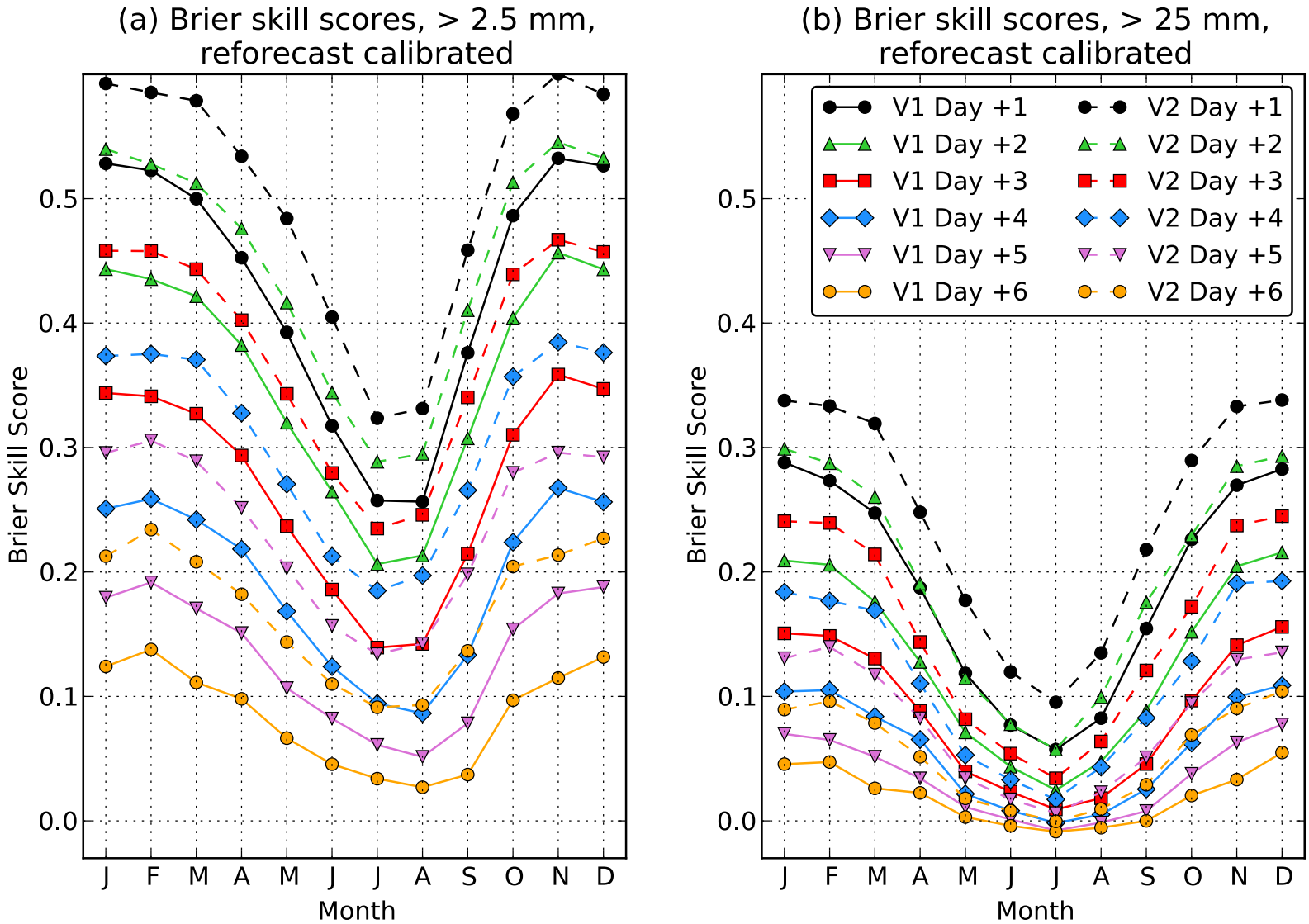
```
x    = [0,   0,   7, 15,   1,  3, 6, 4,   1, 2, 12, 5, 6, 8]

x(r)= [1.5, 1.5, 11, 14, 3.5,  6, 9, 7, 3.5, 5, 13, 8,10,12]
```

with rank analog, these would be the two forecasts with the closest ranks.

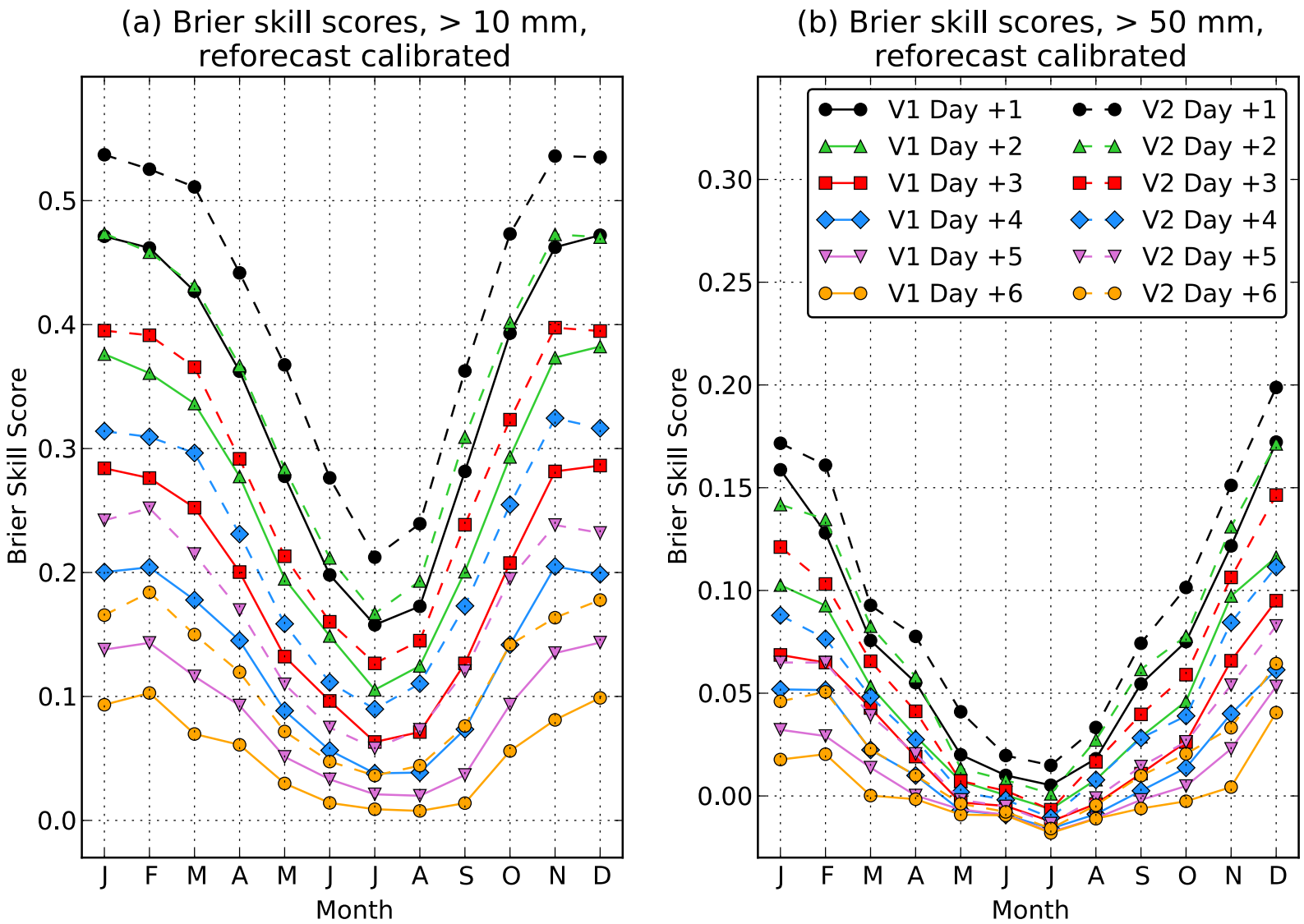
Rank analog calibration details

- 24-h accumulated precipitation, validated on NARR grid (~32 km) over CONUS, 1985-2009.
- Rank analog approach: at each grid point in CONUS, using that grid point and +/- 3 surrounding grid points in N-S, E-W direction, find dates of 75 past forecasts that are closest in average precipitation rank of ensemble mean forecast. Make probabilistic forecasts from analyzed NARR precipitation data on dates of those 75 analogs.
- (Conventionally calculated) Brier Skill Scores, reliability diagrams, etc. NARR again used for verification.

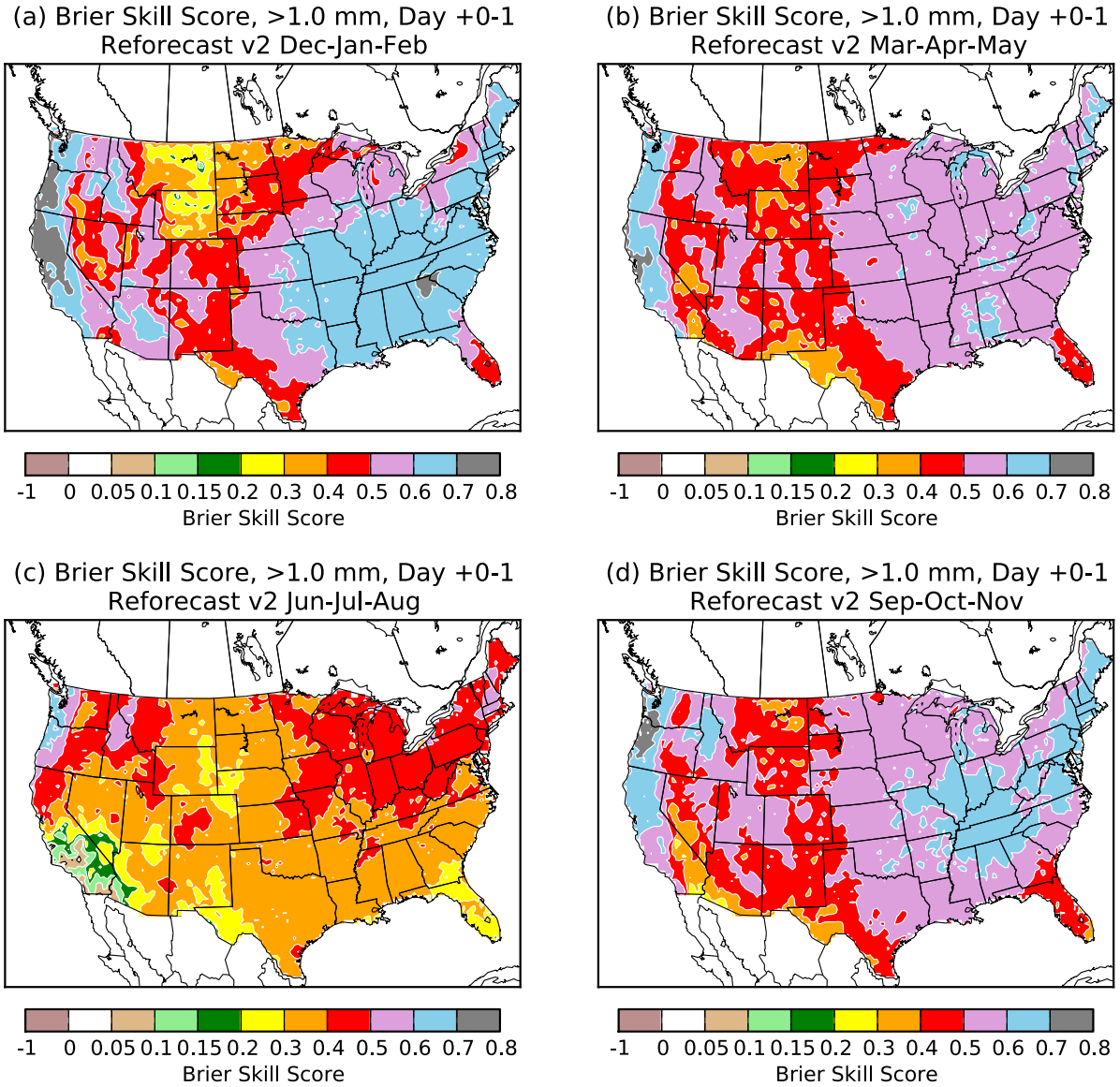
Skill of calibrated precipitation forecasts



Scores, > 10 mm and > 50 mm

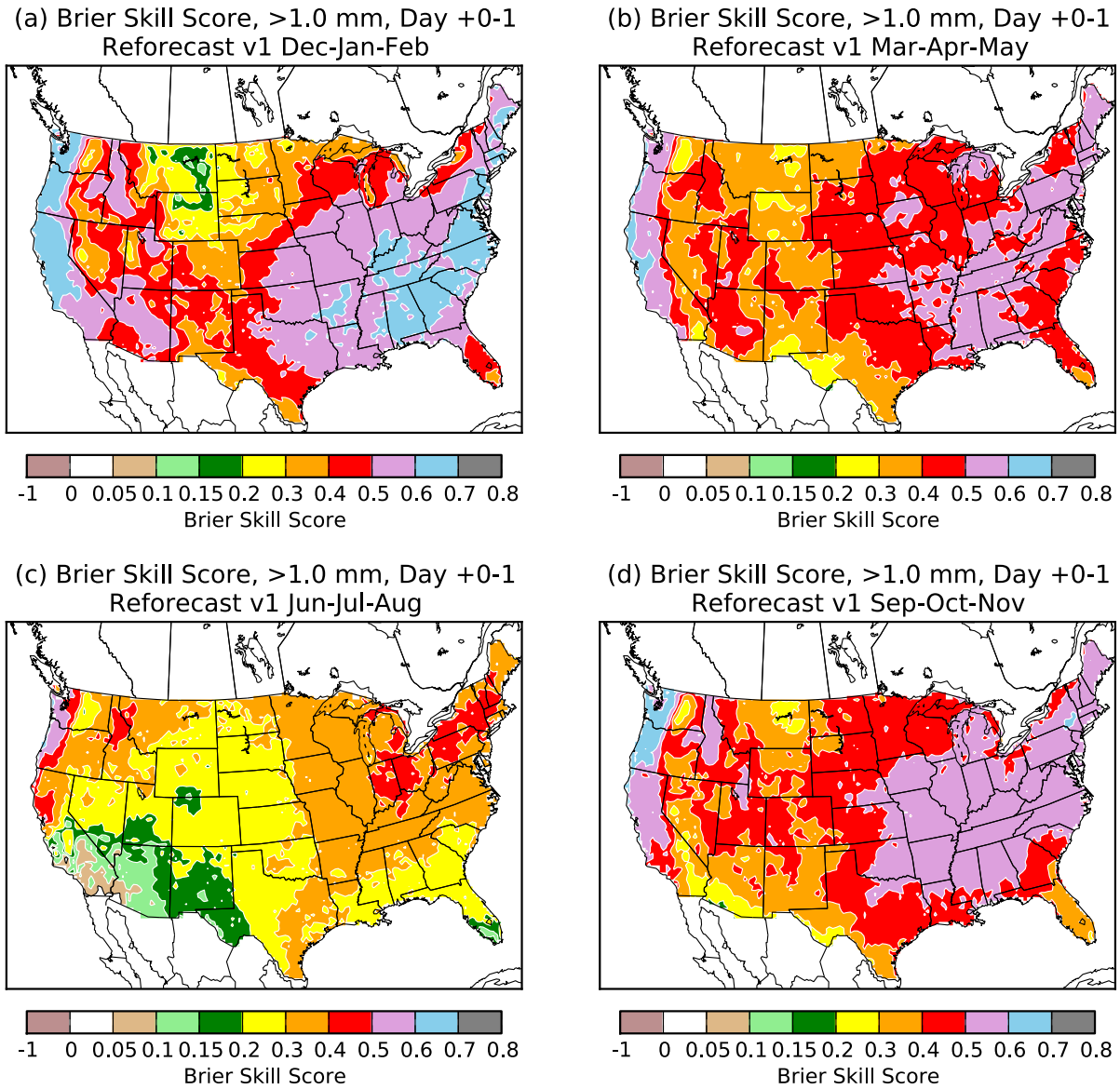


Brier Skill Scores, Day +0-1, 1.0 mm $24h^{-1}$ (reforecast version 2)

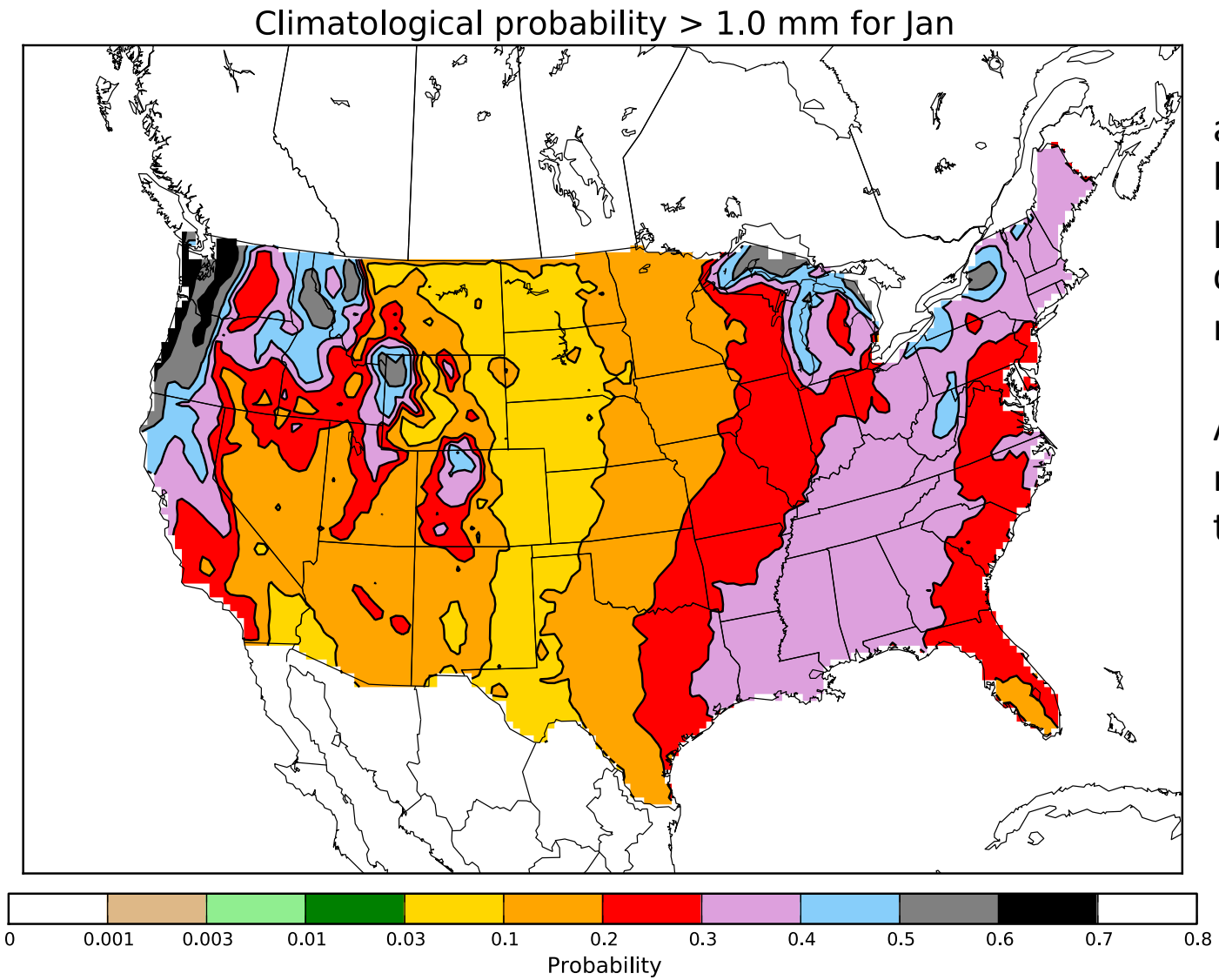


Brier Skill Scores, Day +0-1, 1.0 mm $24h^{-1}$

(reforecast version 1)



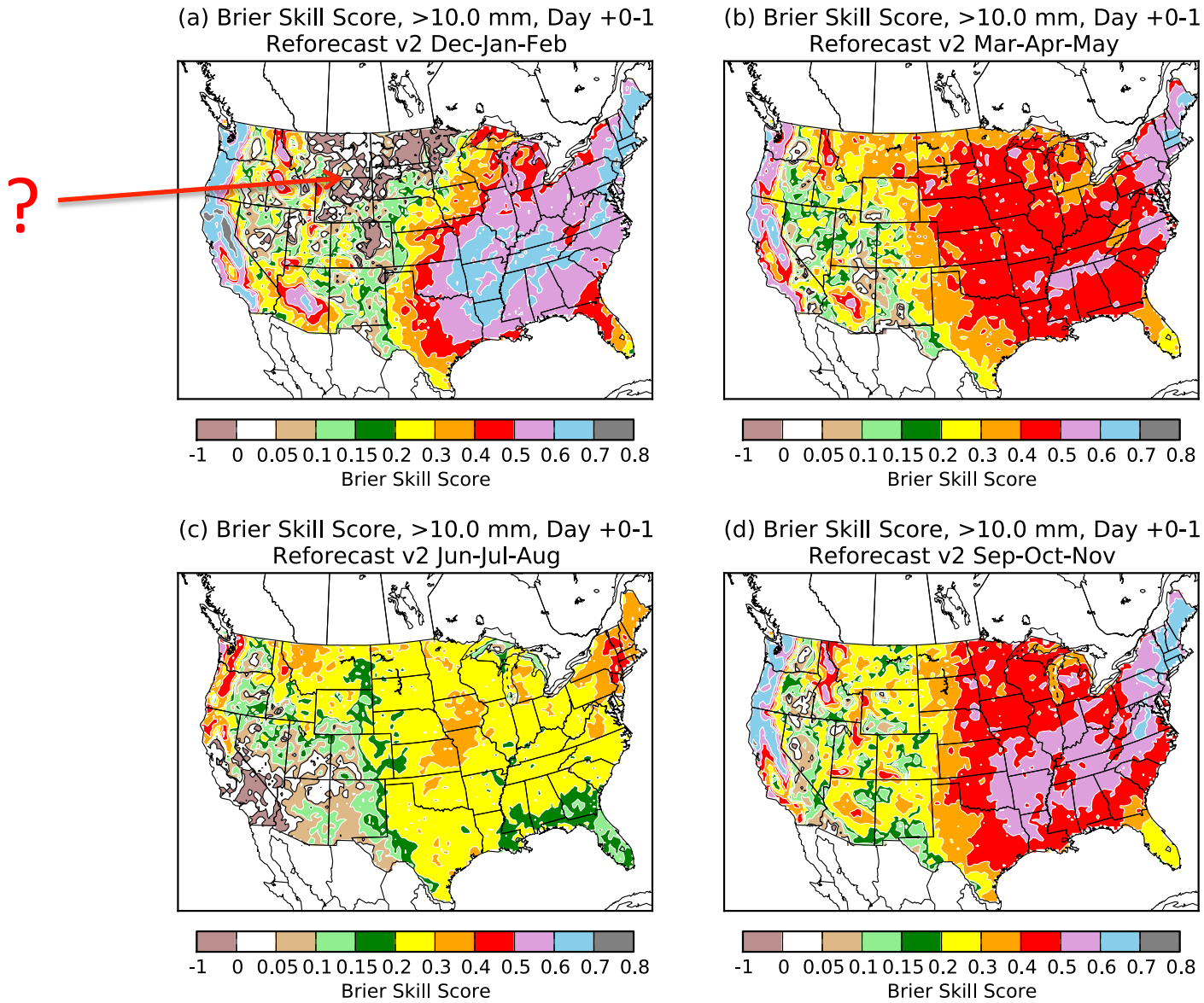
NARR climatology, > 1-mm 24 h^{-1}



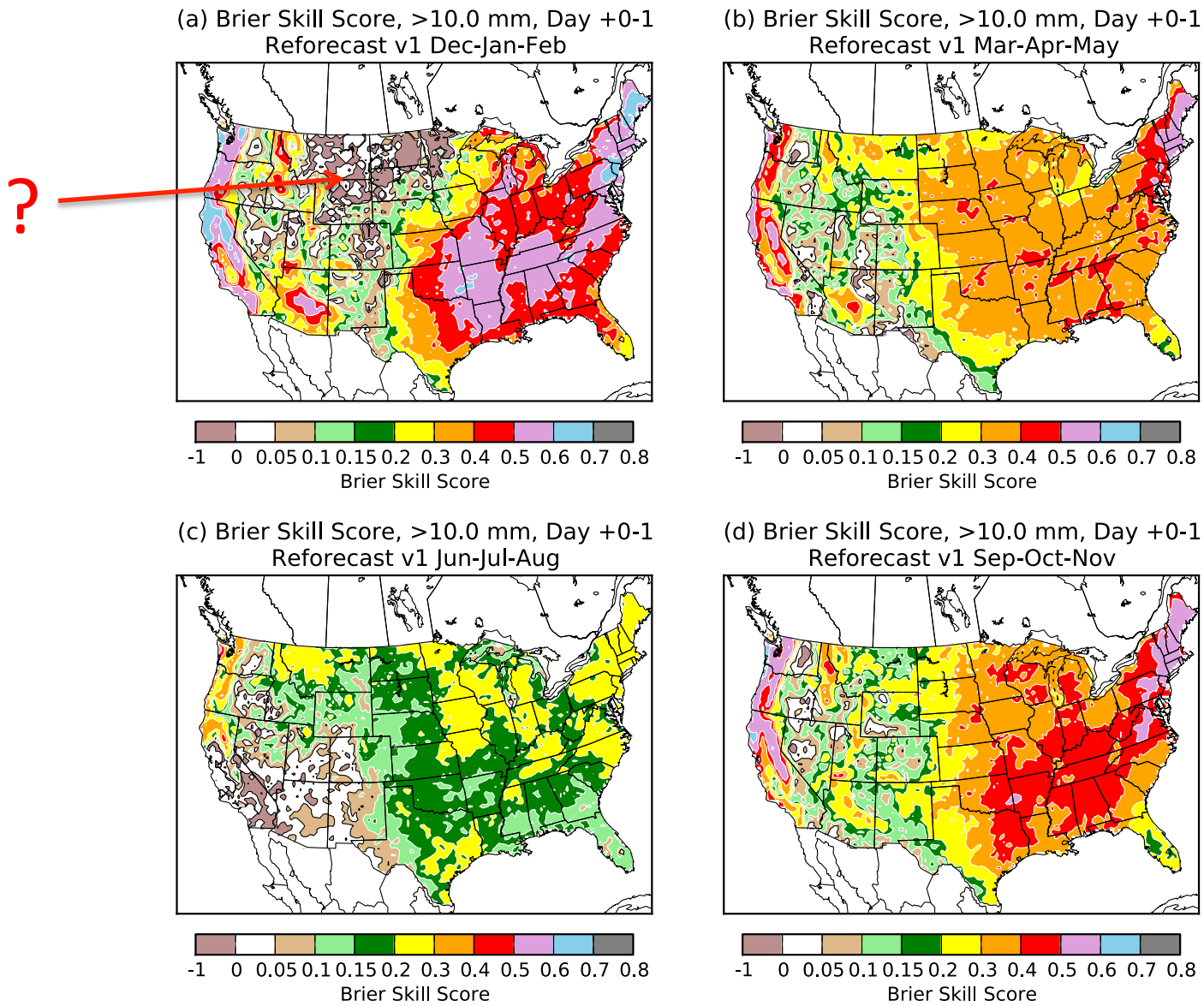
a modest relationship between where precipitation is common and the magnitude of skill.

Also, note relationship with terrain forcing.

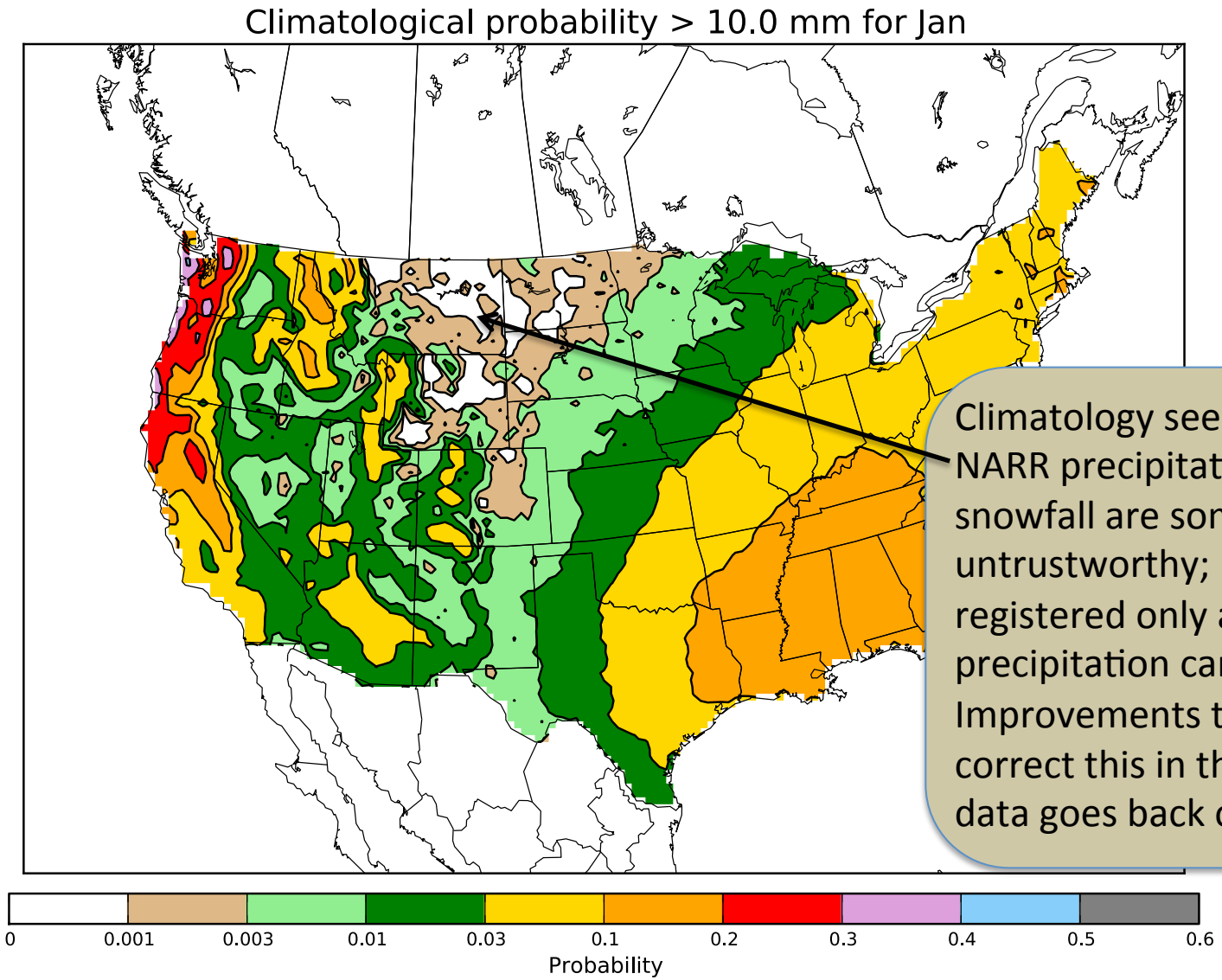
Brier Skill Scores, Day +0-1, 10.0 mm $24h^{-1}$ (reforecast version 2)



Brier Skill Scores, Day +0-1, 10.0 mm $24h^{-1}$ (reforecast version 1)

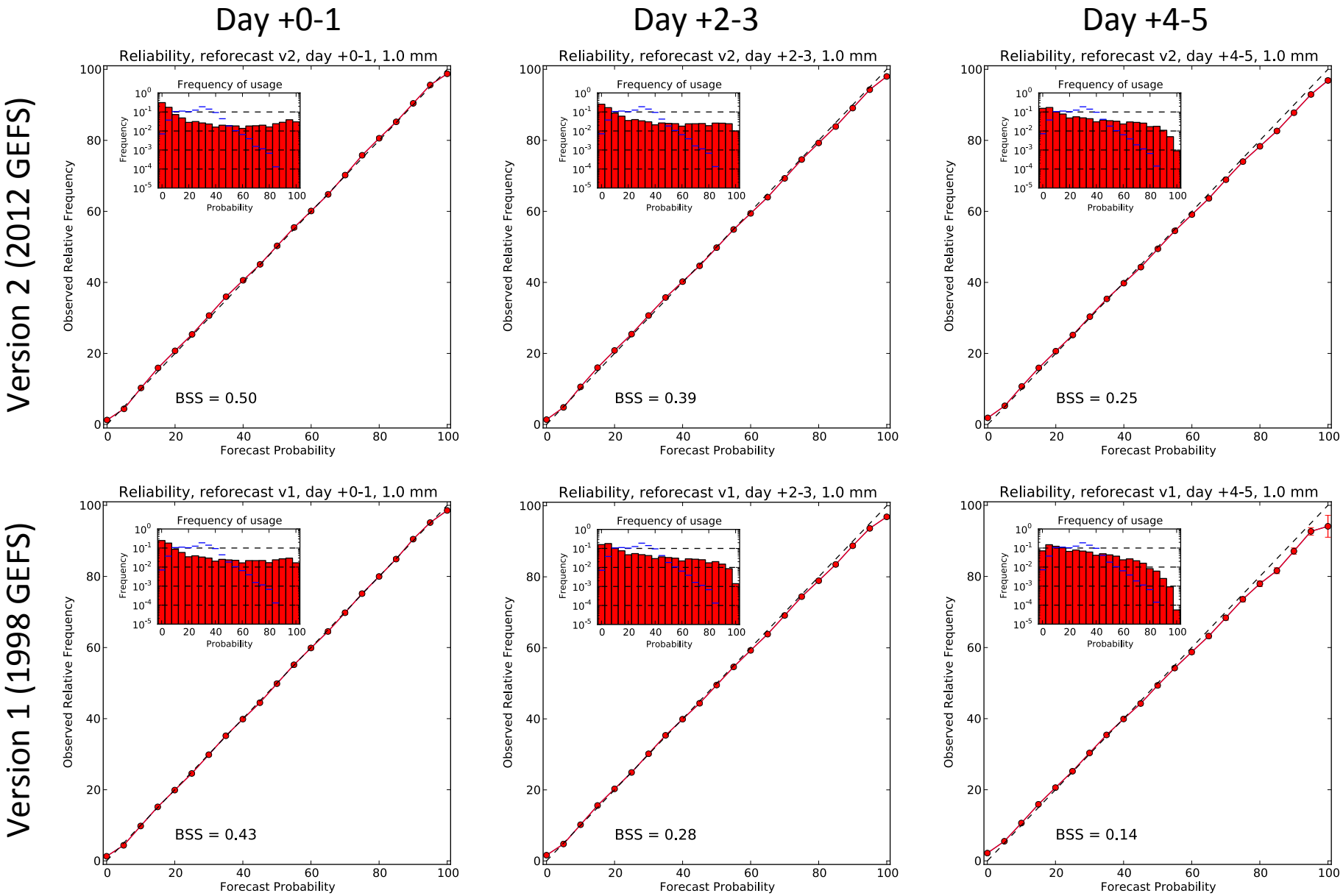


NARR climatology, > 10 mm 24 h^{-1}



Climatology seems anomalously dry. NARR precipitation analyses after snowfall are somewhat untrustworthy; precipitation registered only as snow melts, so precipitation can be off by days. Improvements to Stage IV may correct this in the future, but this data goes back only to 2002.

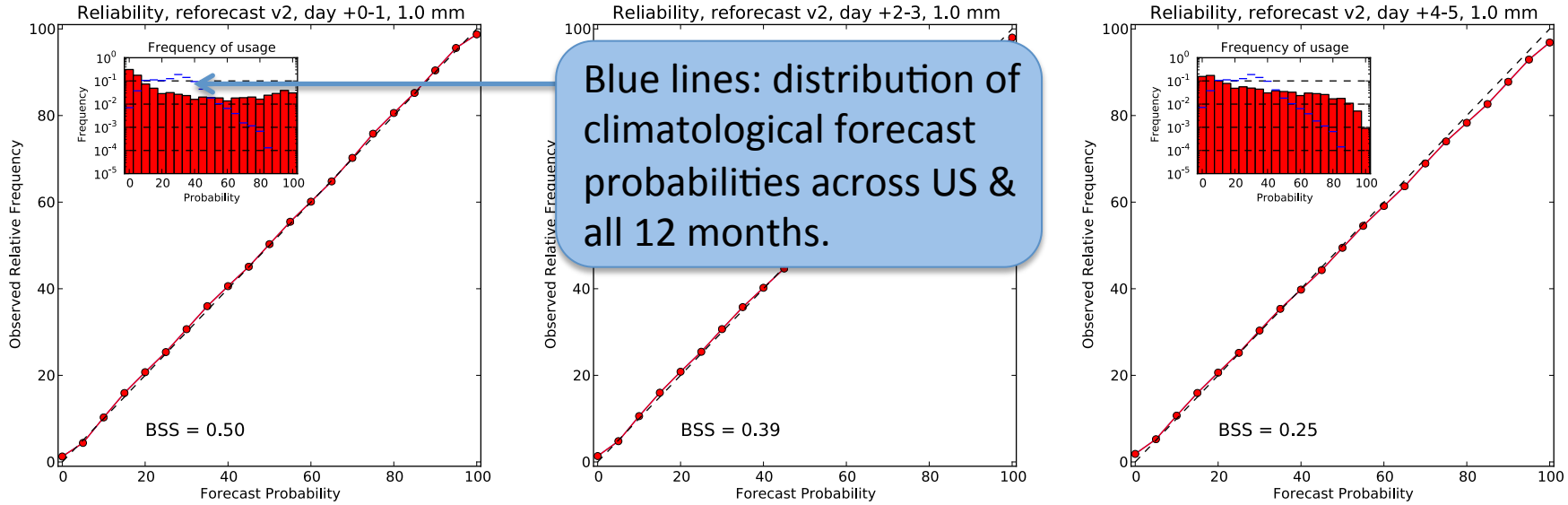
Reliability, > 1 mm precipitation 24 h^{-1}



Reliability, > 1 mm precipitation 24 h^{-1}

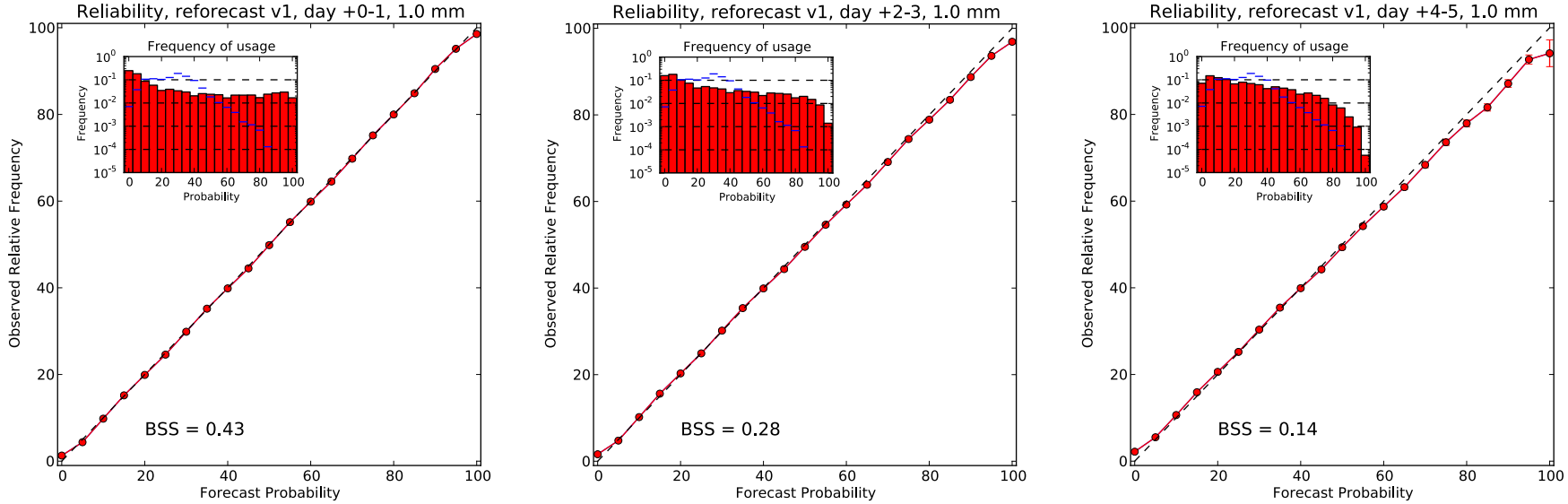
Day +0-1 | Day +2-3 | Day +4-5

Version 2 (2012 GEFS)



Blue lines: distribution of climatological forecast probabilities across US & all 12 months.

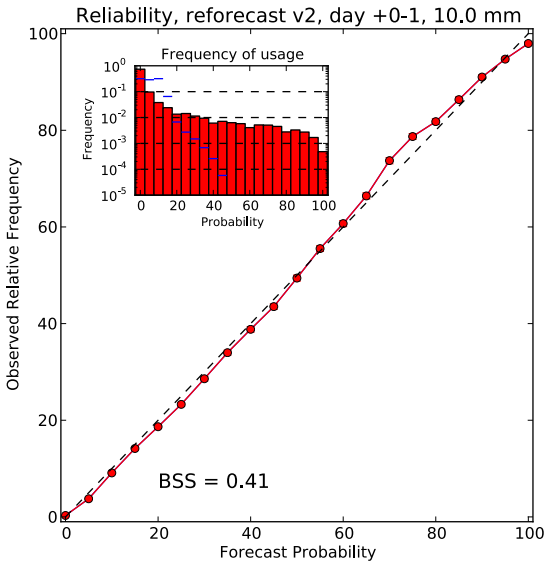
Version 1 (1998 GEFS)

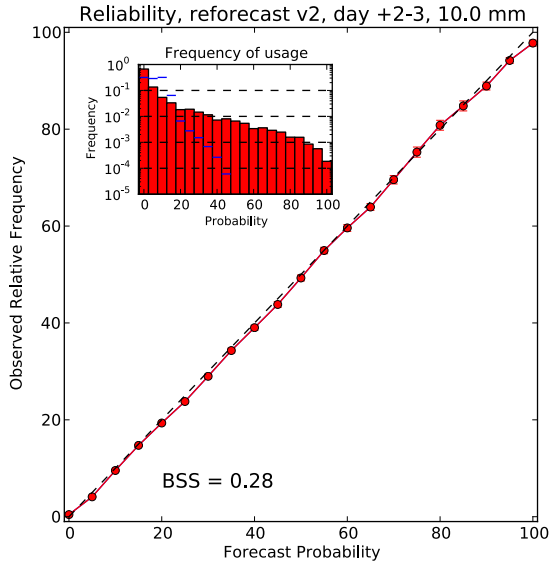


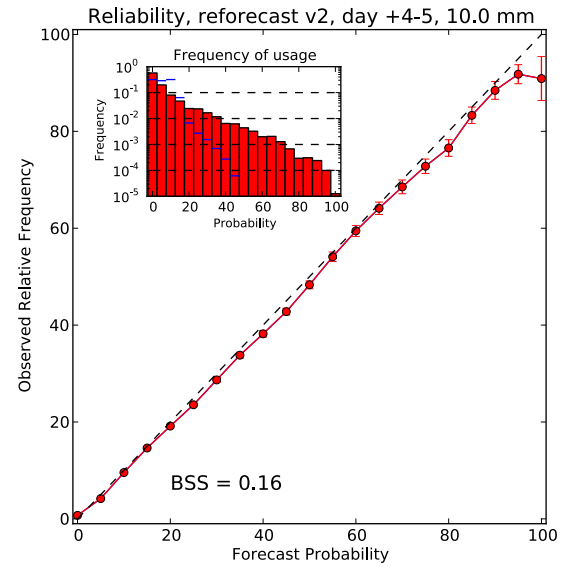
Reliability, > 10 mm precipitation 24 h^{-1}

Day +0-1 | Day +2-3 | Day +4-5

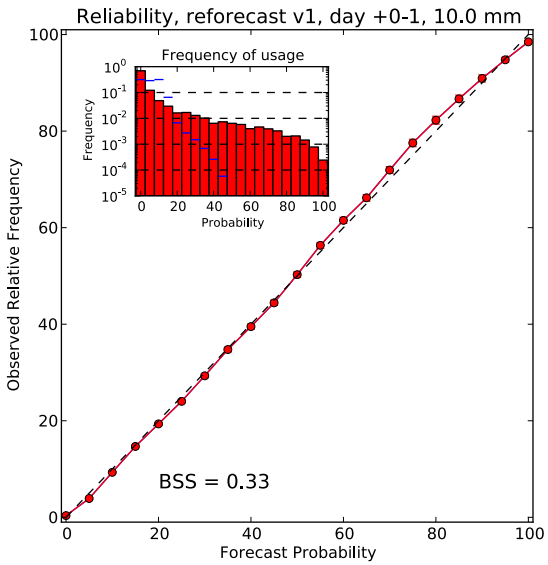
Version 2 (2012 GEFS)

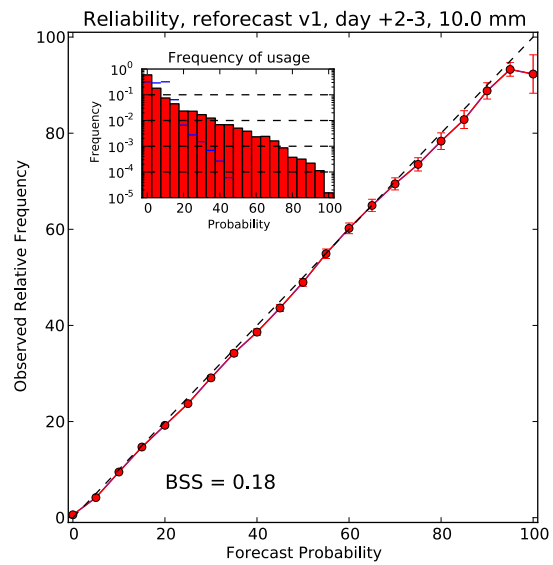


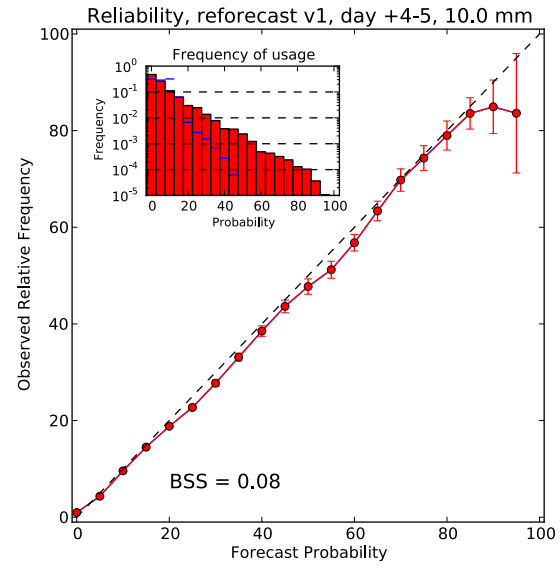




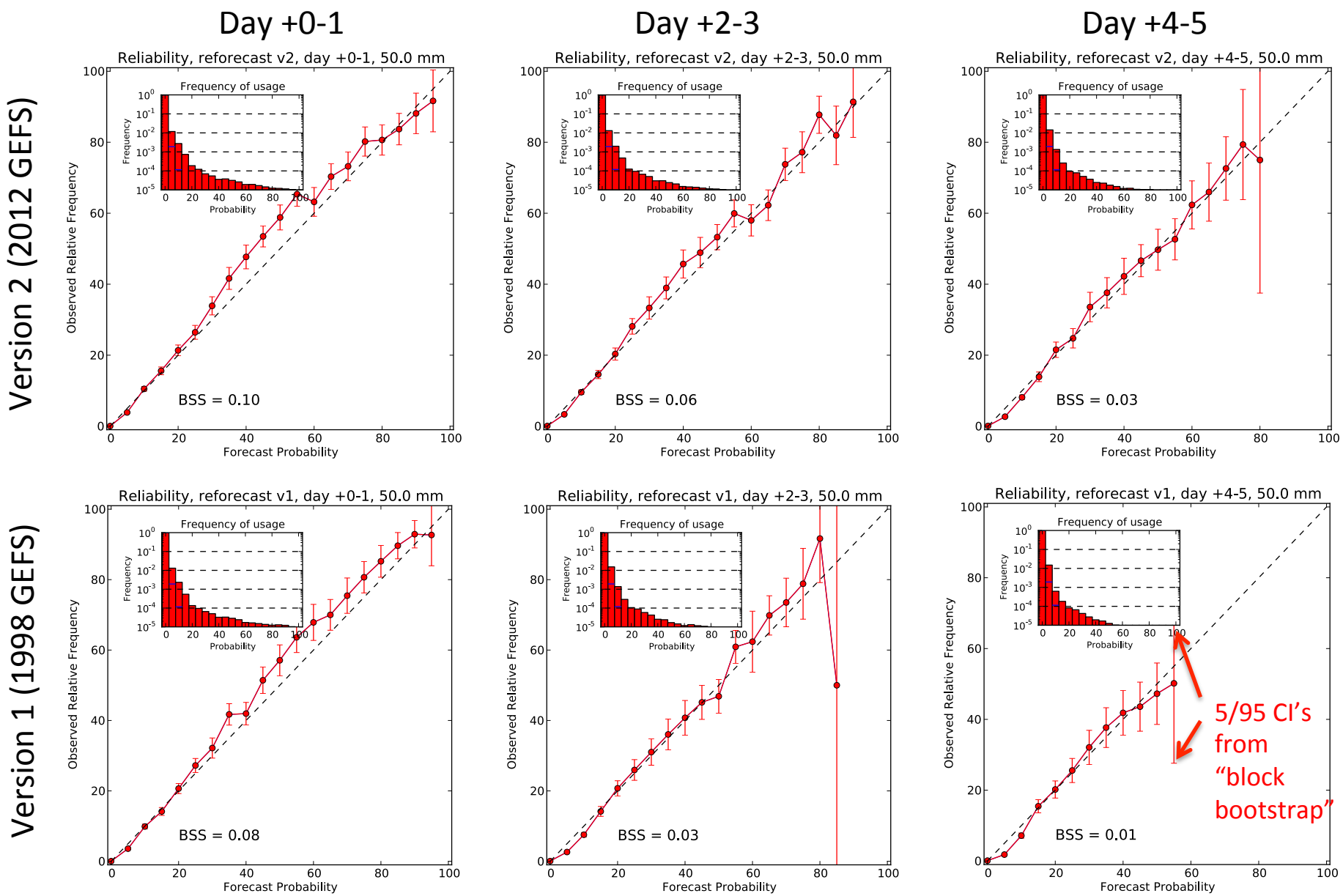
Version 1 (1998 GEFS)



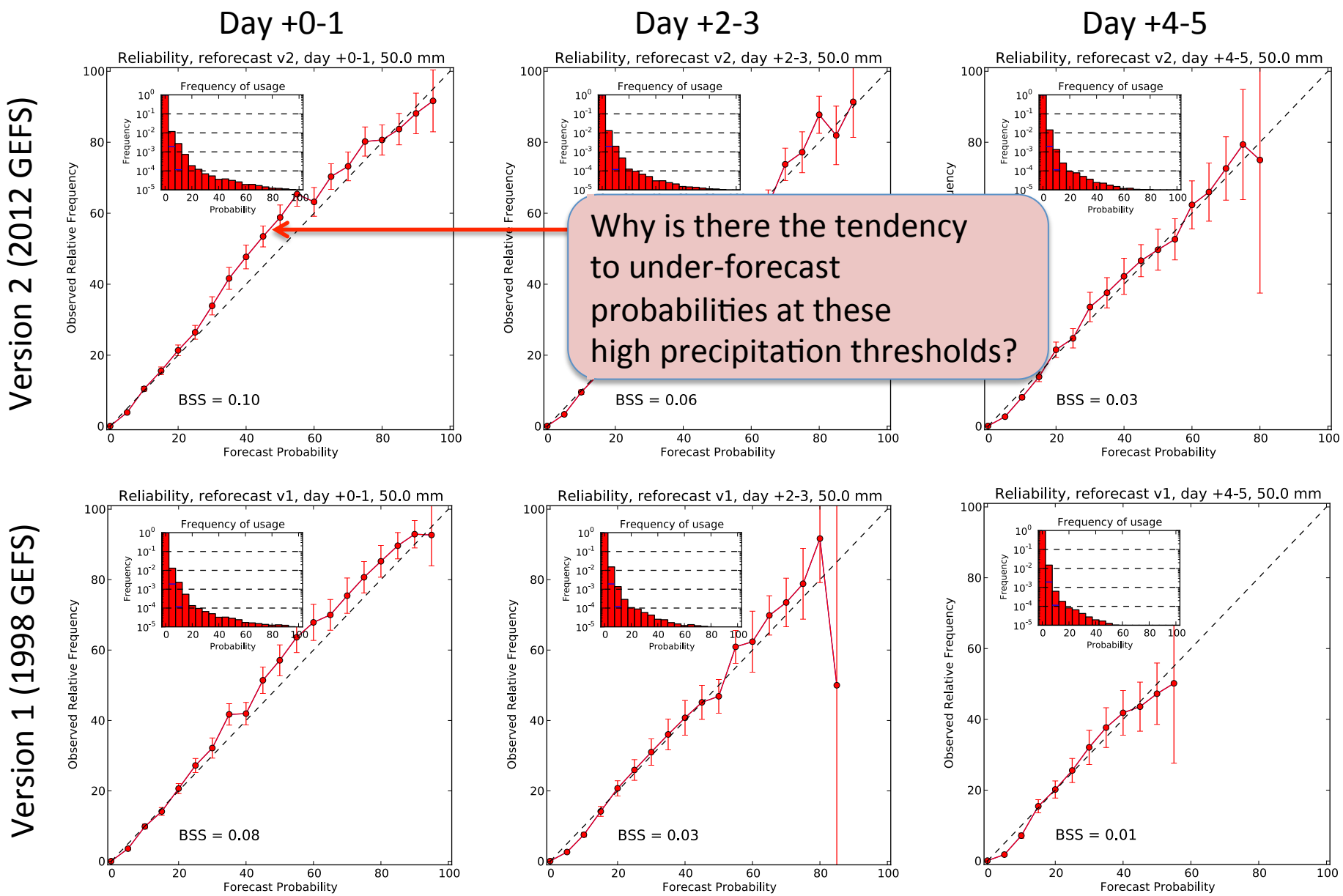




Reliability, > 50 mm precipitation 24 h^{-1}

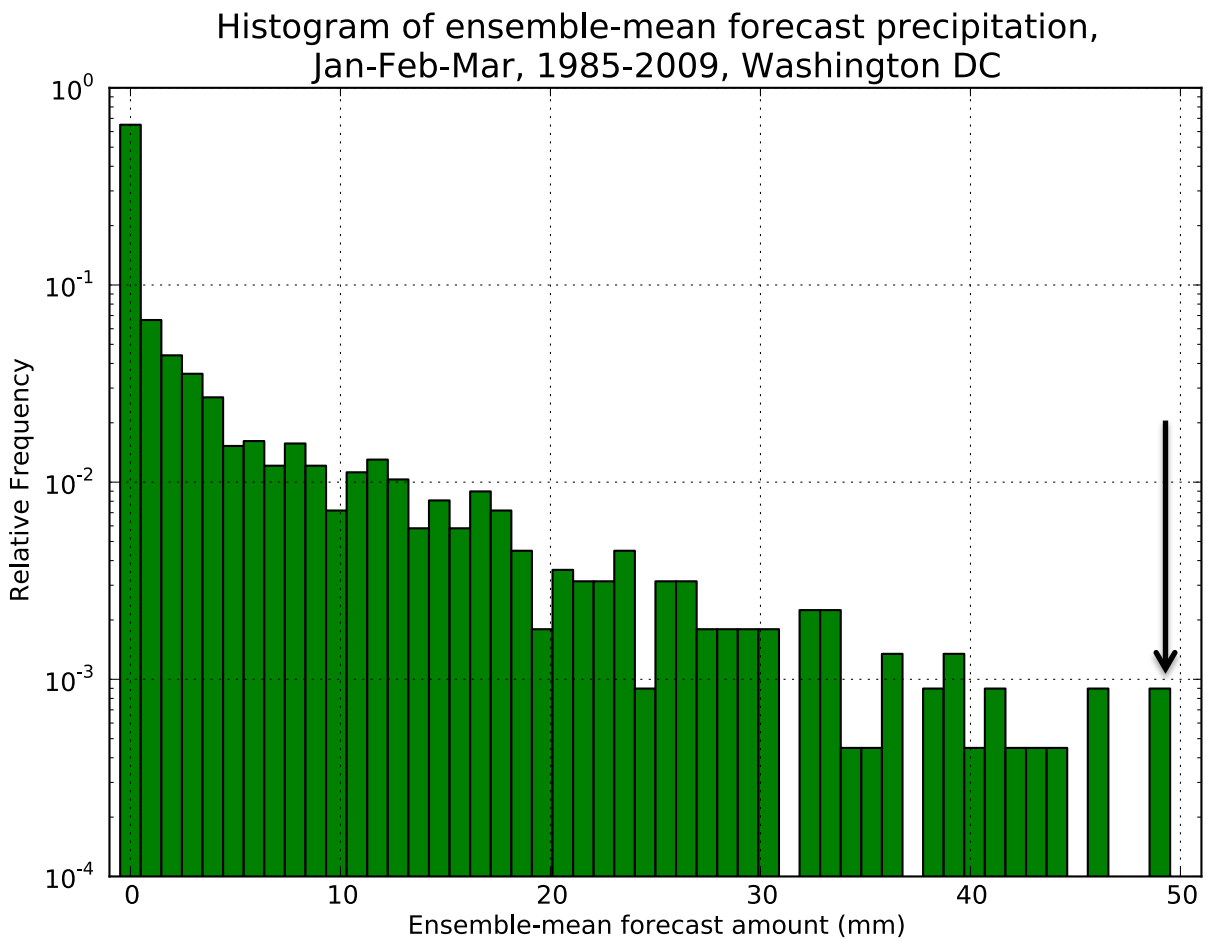


Reliability, > 50 mm precipitation 24 h^{-1}



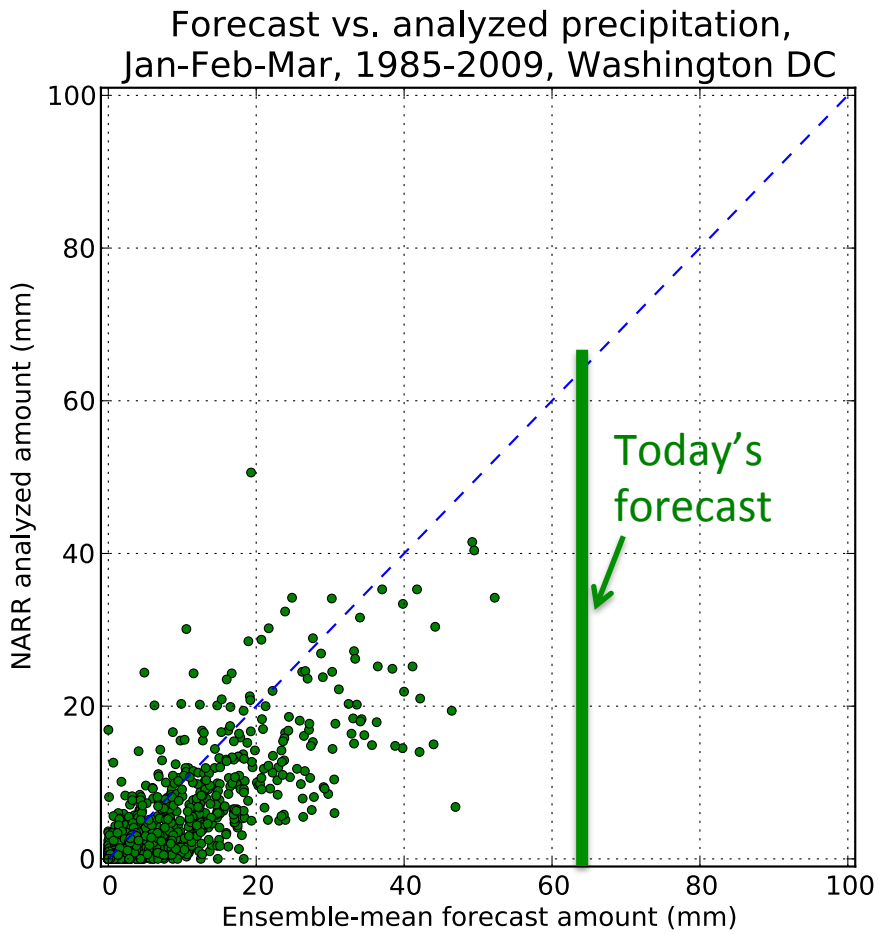
Why is there the tendency to under-forecast probabilities at these high precipitation thresholds?

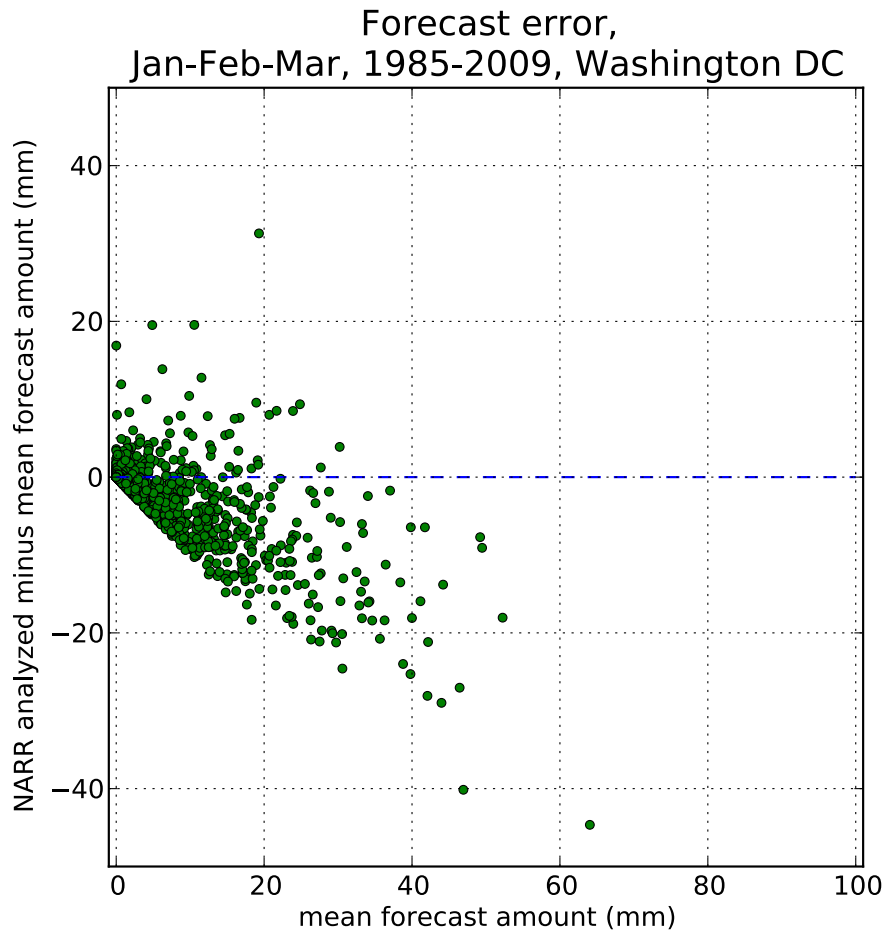
Problem at extreme forecast amounts



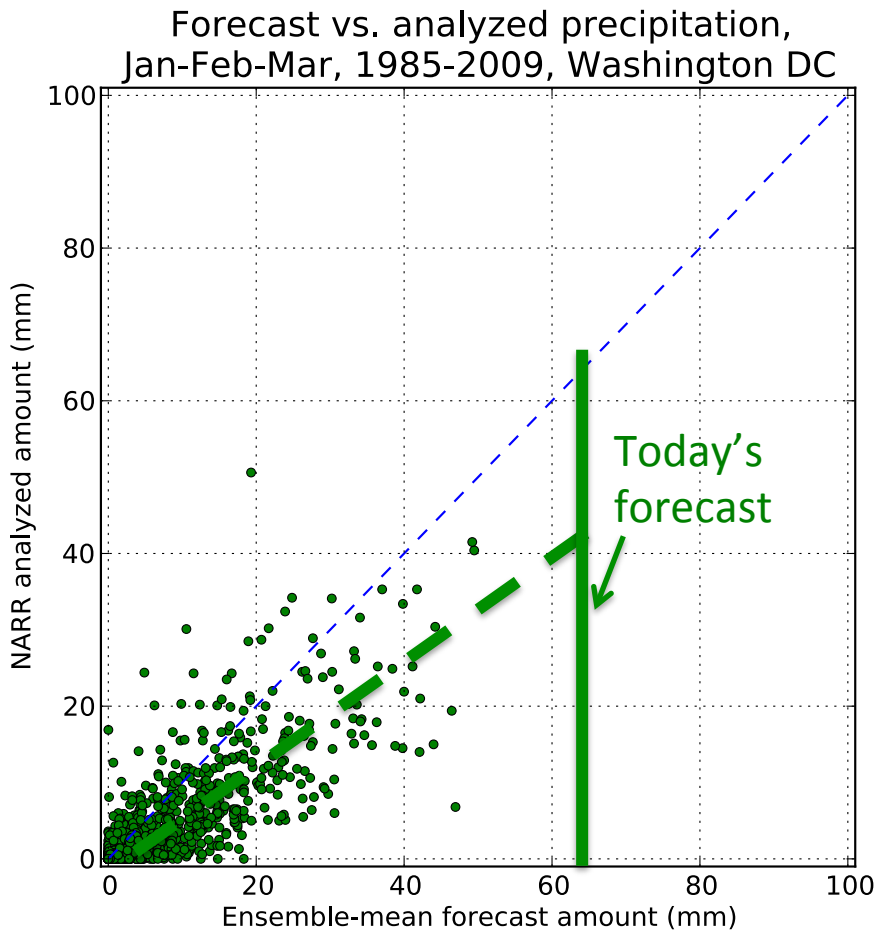
Say today's forecast is for ~ 50 mm, and suppose this is the higher in amount than any forecast in the reforecast time series. The analogs will be comprised of nothing but observed on dates when prior forecasts had lower than 50 mm. Hence there may be a strong tendency to under-forecast.

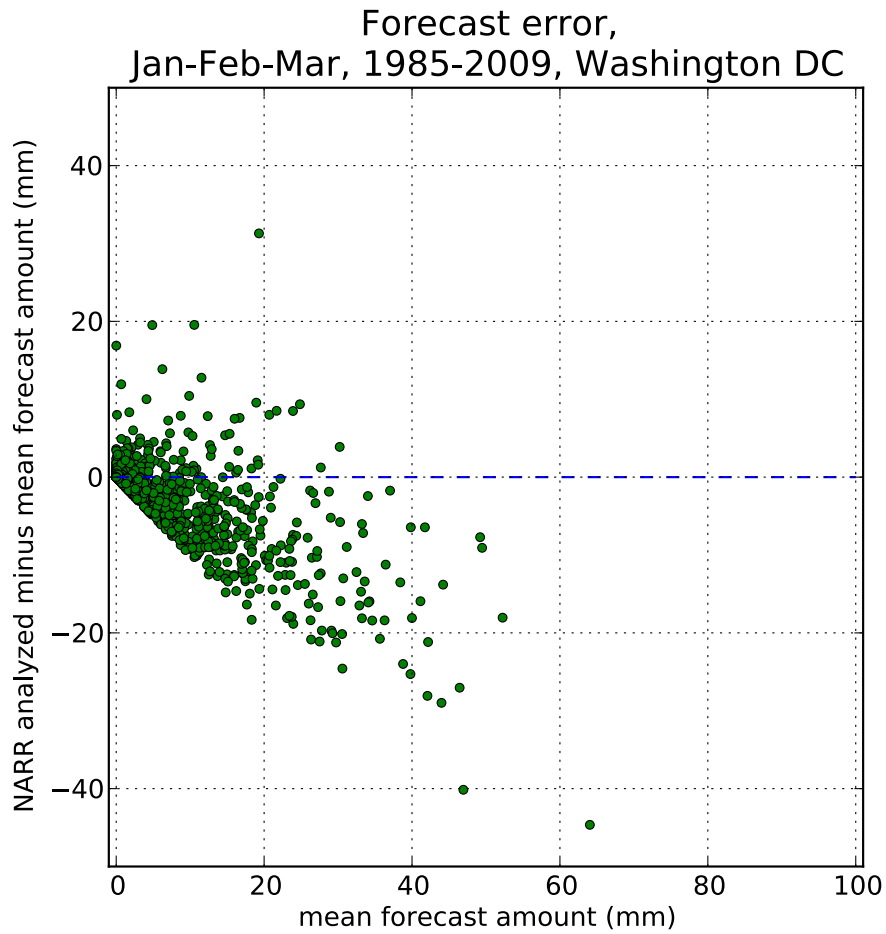
Example of forecast / analysis data





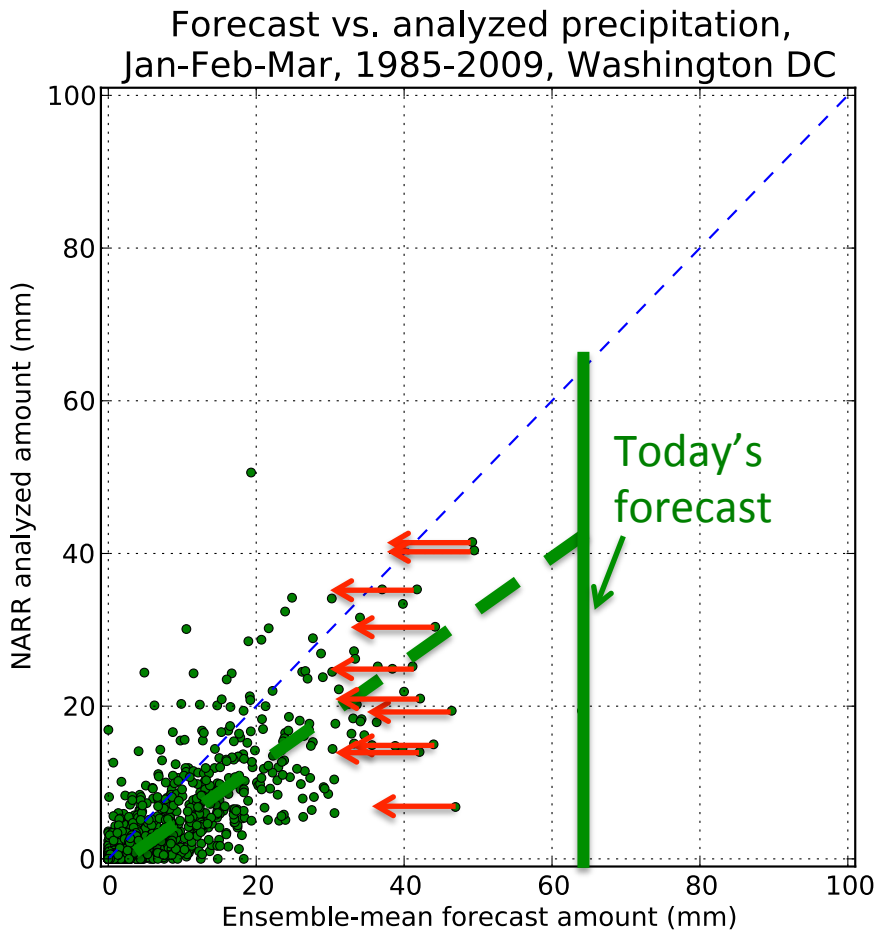
Example of forecast / analysis data

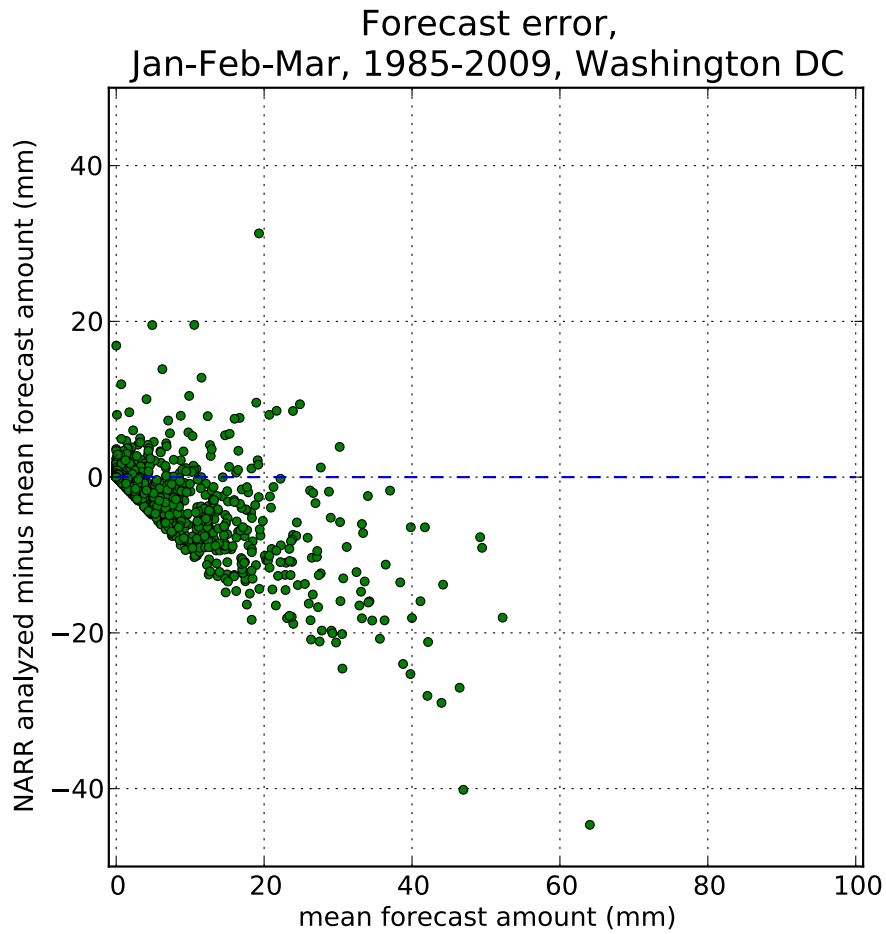




Is THIS an appropriate model for the general relationship between forecast & analyzed?

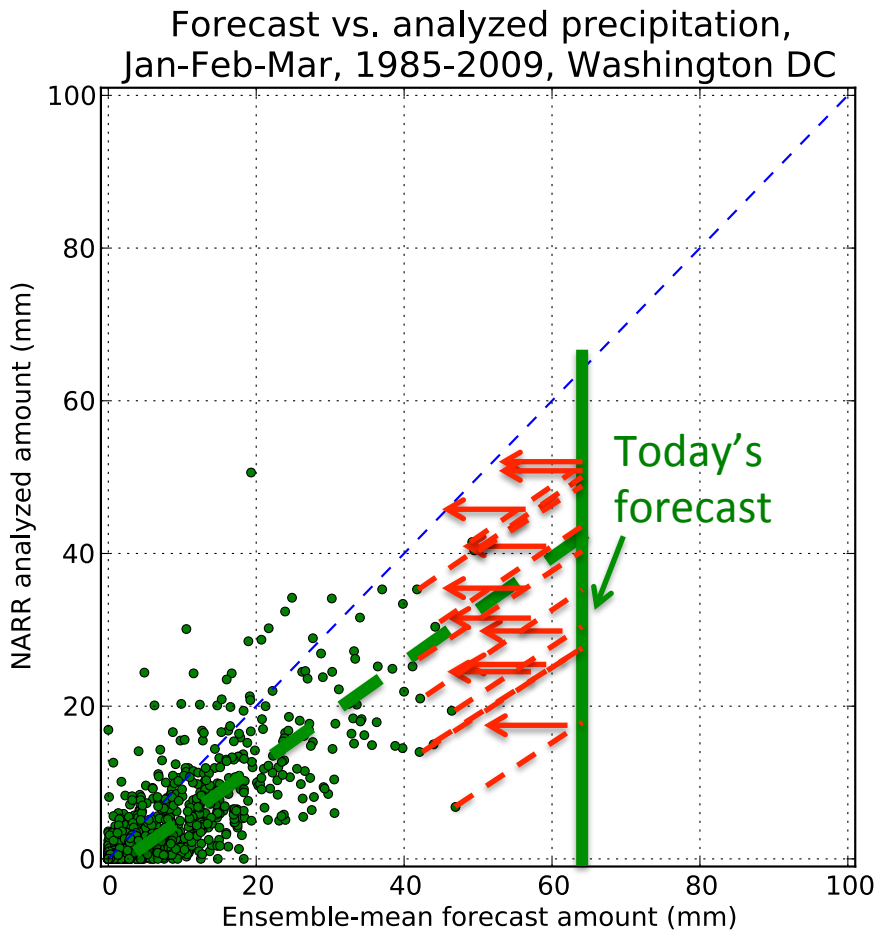
Example of forecast / analysis data

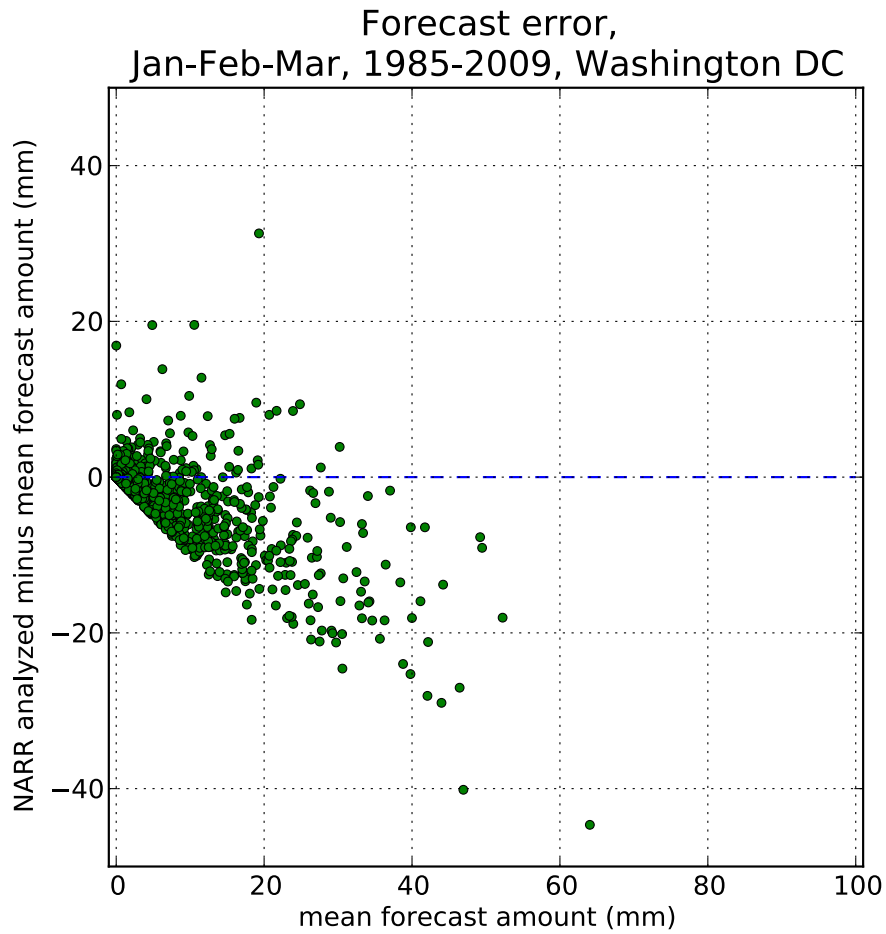




And if so, is this sample of the observed given the forecasts biased too low for today's forecast?

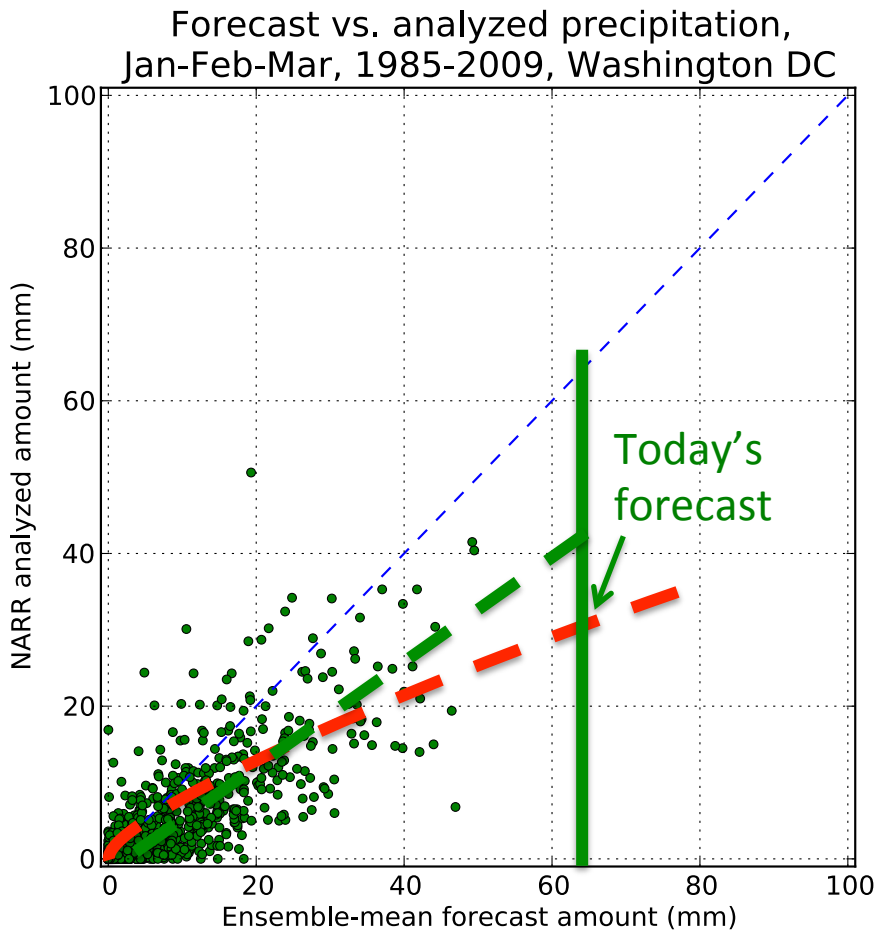
Example of forecast / analysis data

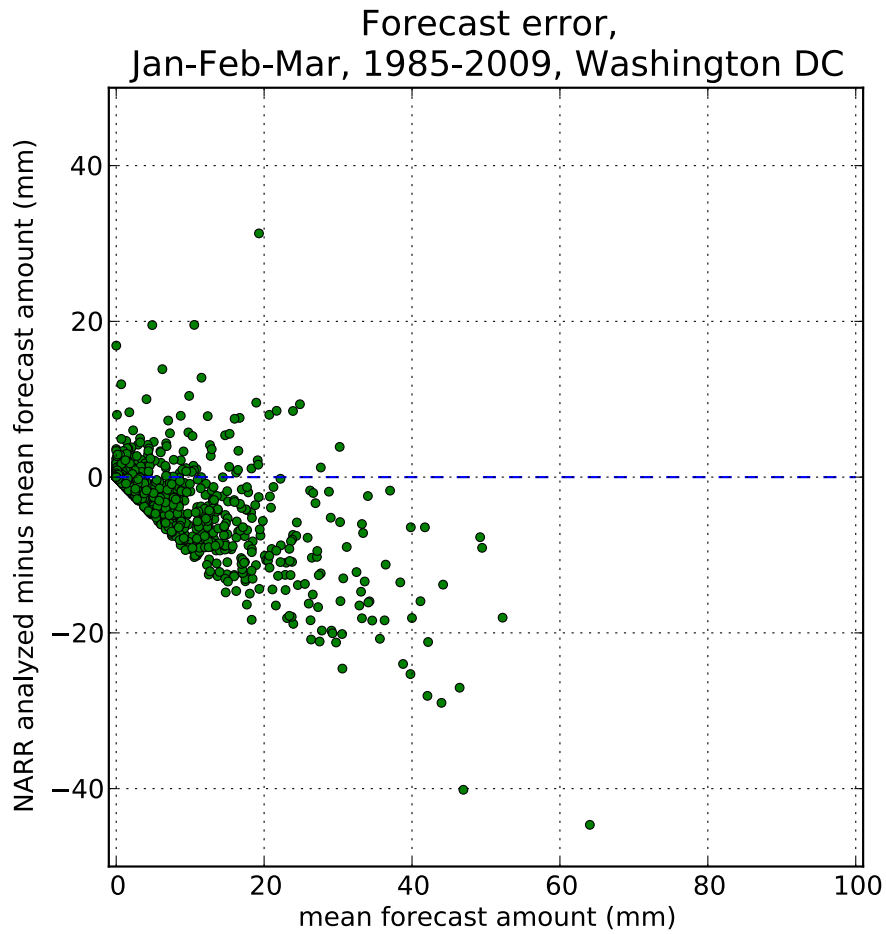




Perhaps this sample of the regression-shifted observed given the forecasts is better?

Example of forecast / analysis data

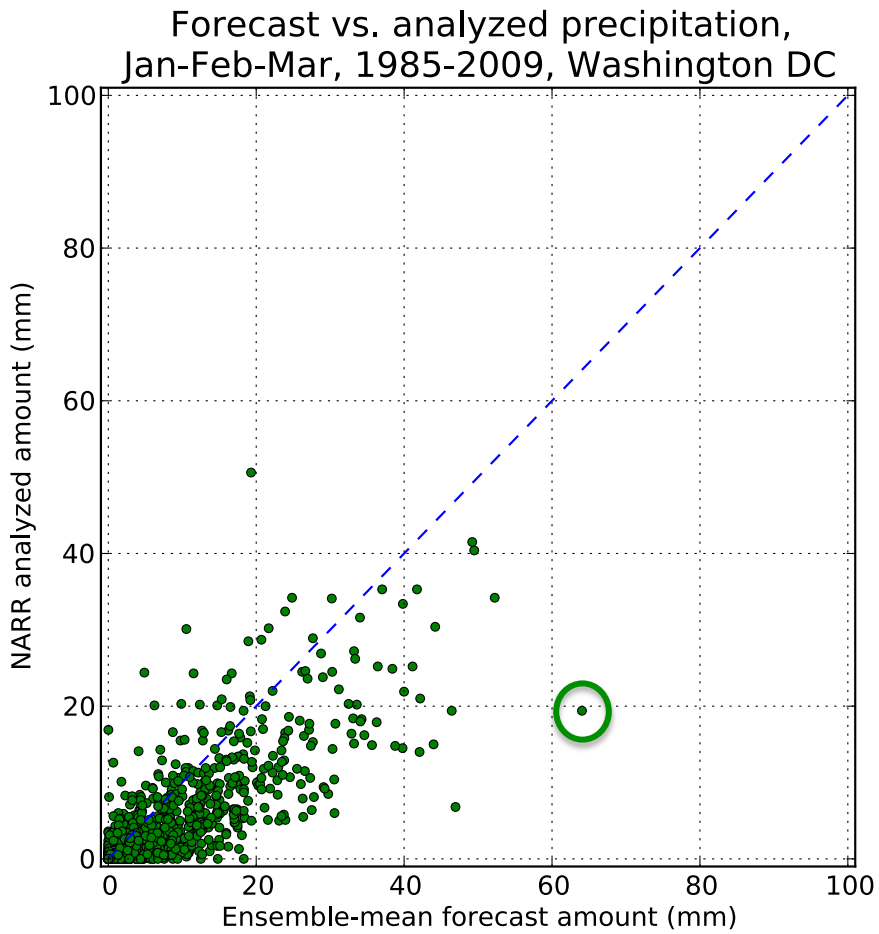


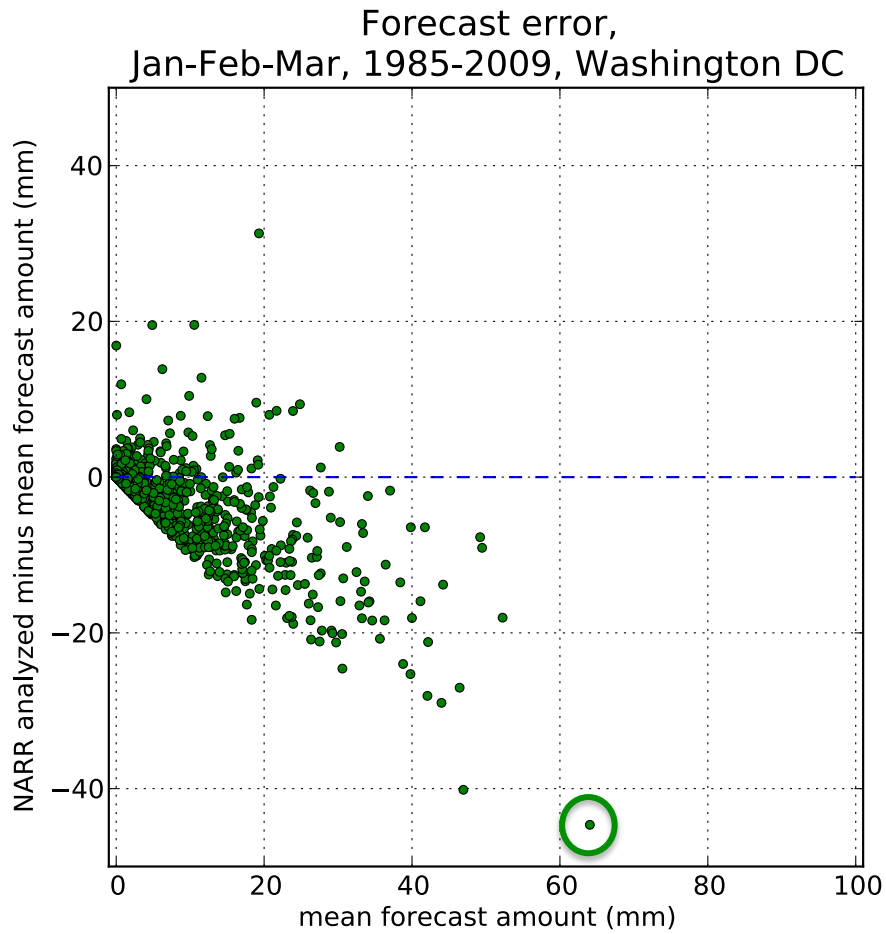


Yet the best regression relationship at high amounts is uncertain. Perhaps THIS instead is a better model than THAT? Highlights a general problem: "extrapolating the regression."

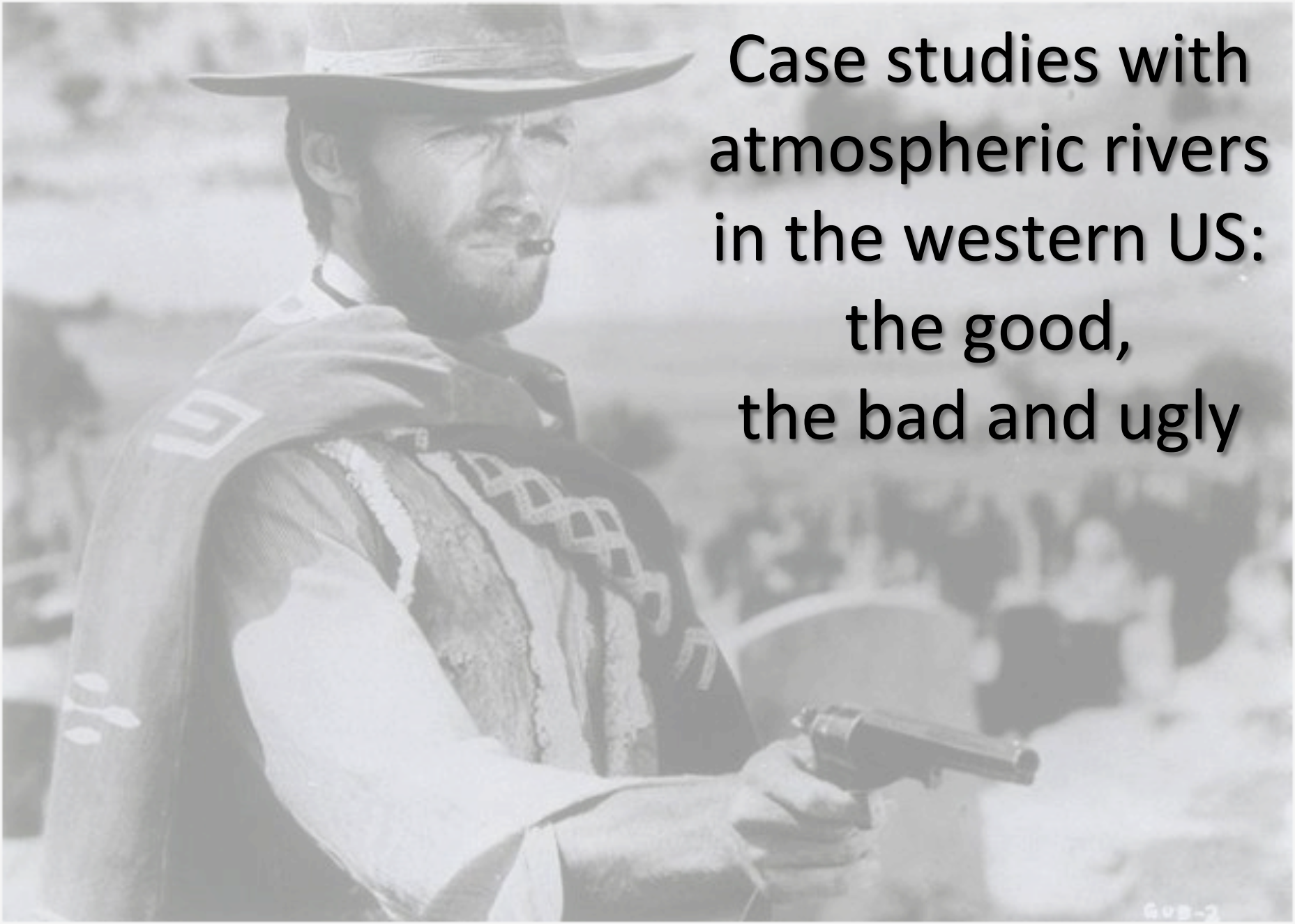
Danger!
One can spend multiple lifetimes fiddling…

Example of forecast / analysis data



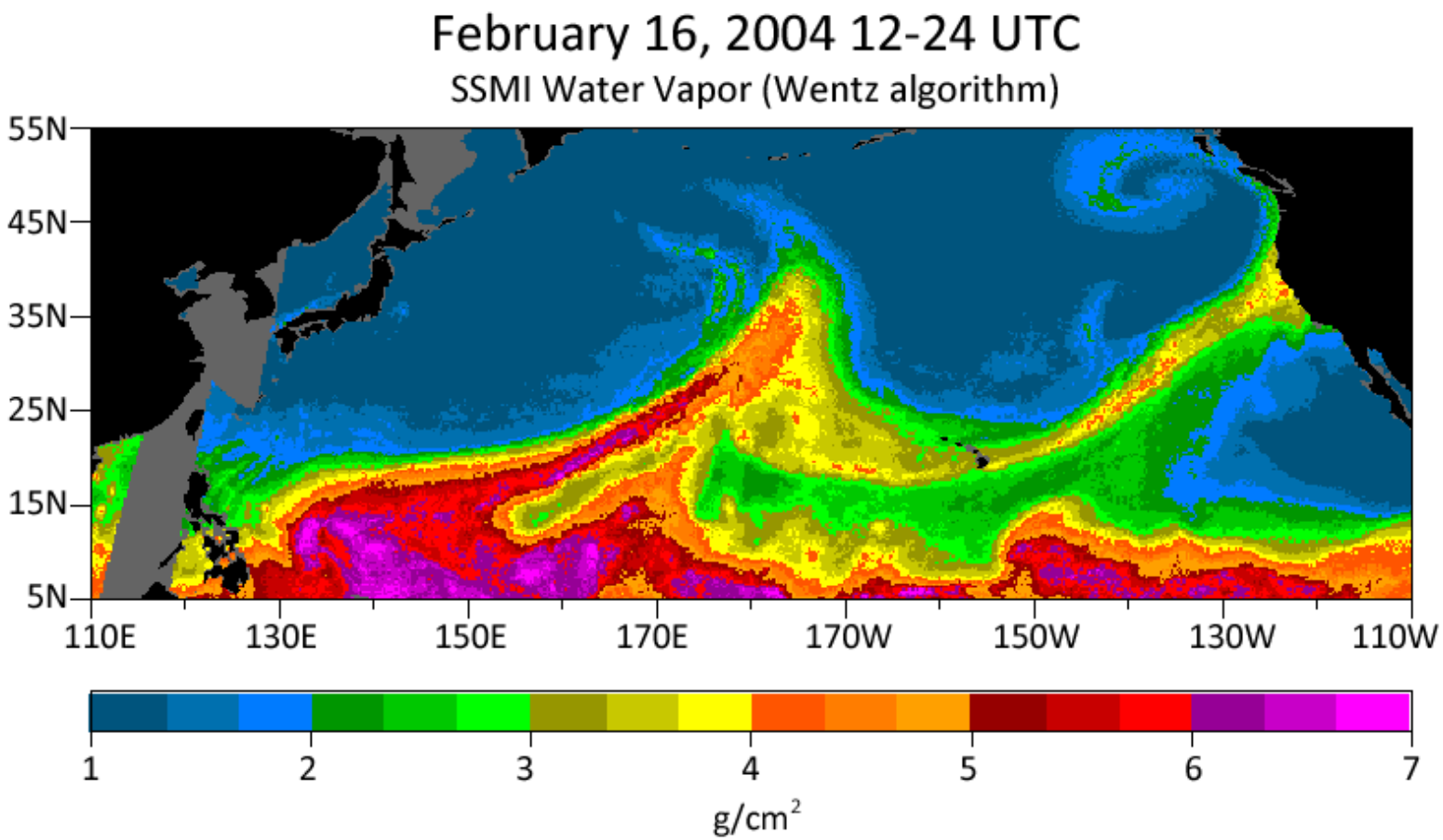


Here is the actual data point. In this circumstance, most seemingly plausible regression corrections would have made the forecast worse.



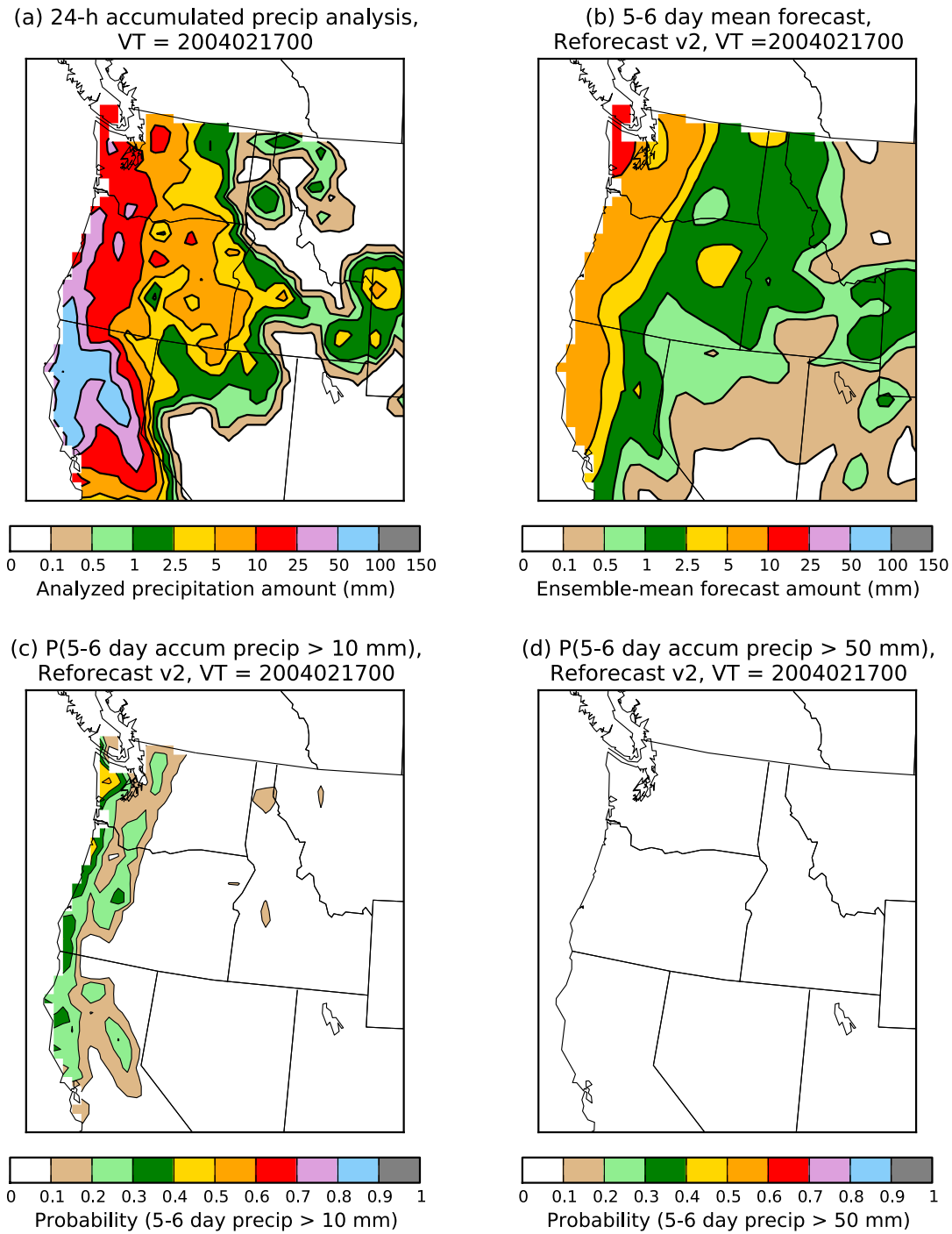
Case studies with atmospheric rivers in the western US: the good, the bad and ugly

The bad and ugly atmospheric rivers case study

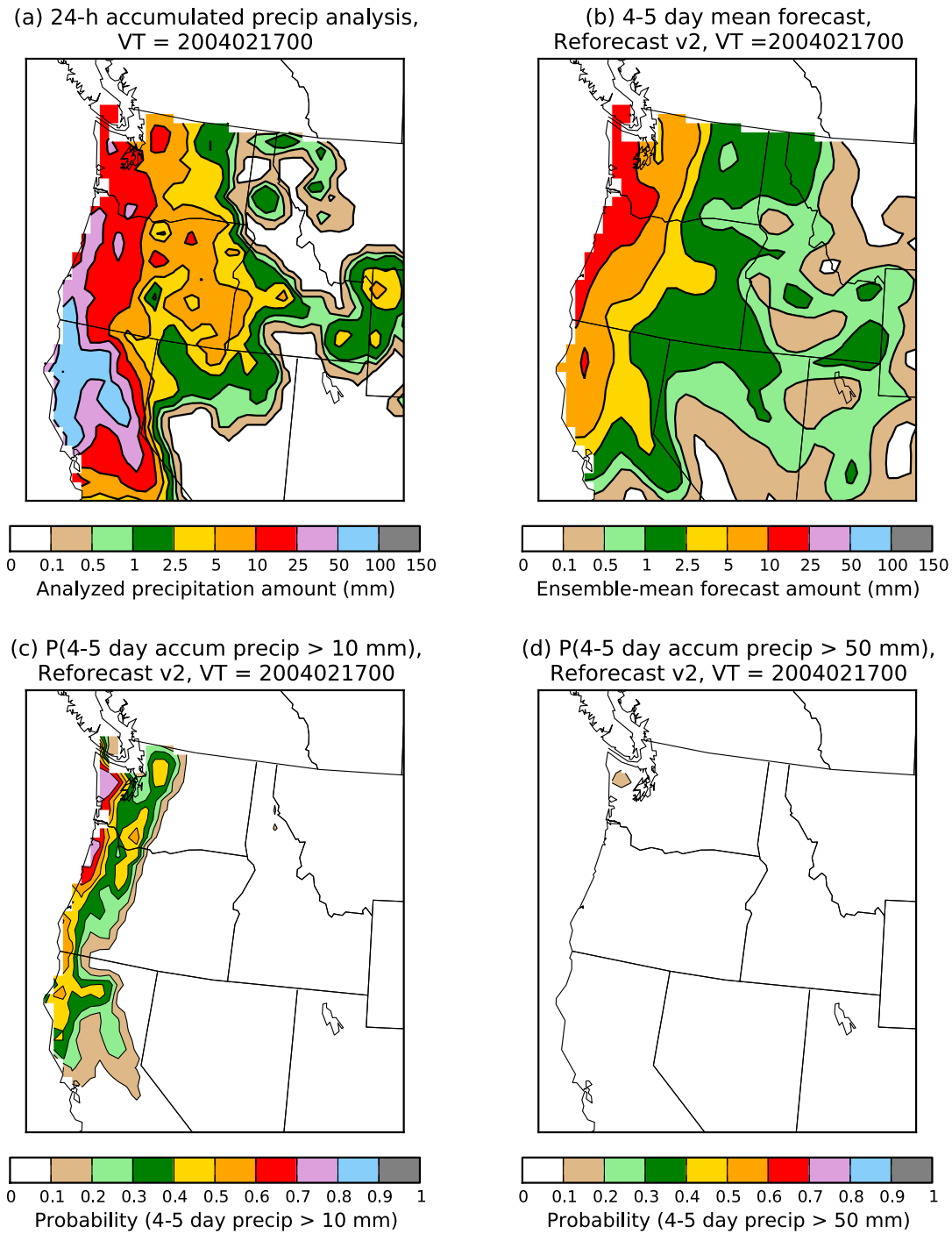


- 10” rain in the coastal mountains, 4-7” in Russian River watershed. Streamflows in top 0.2% of historical observations.

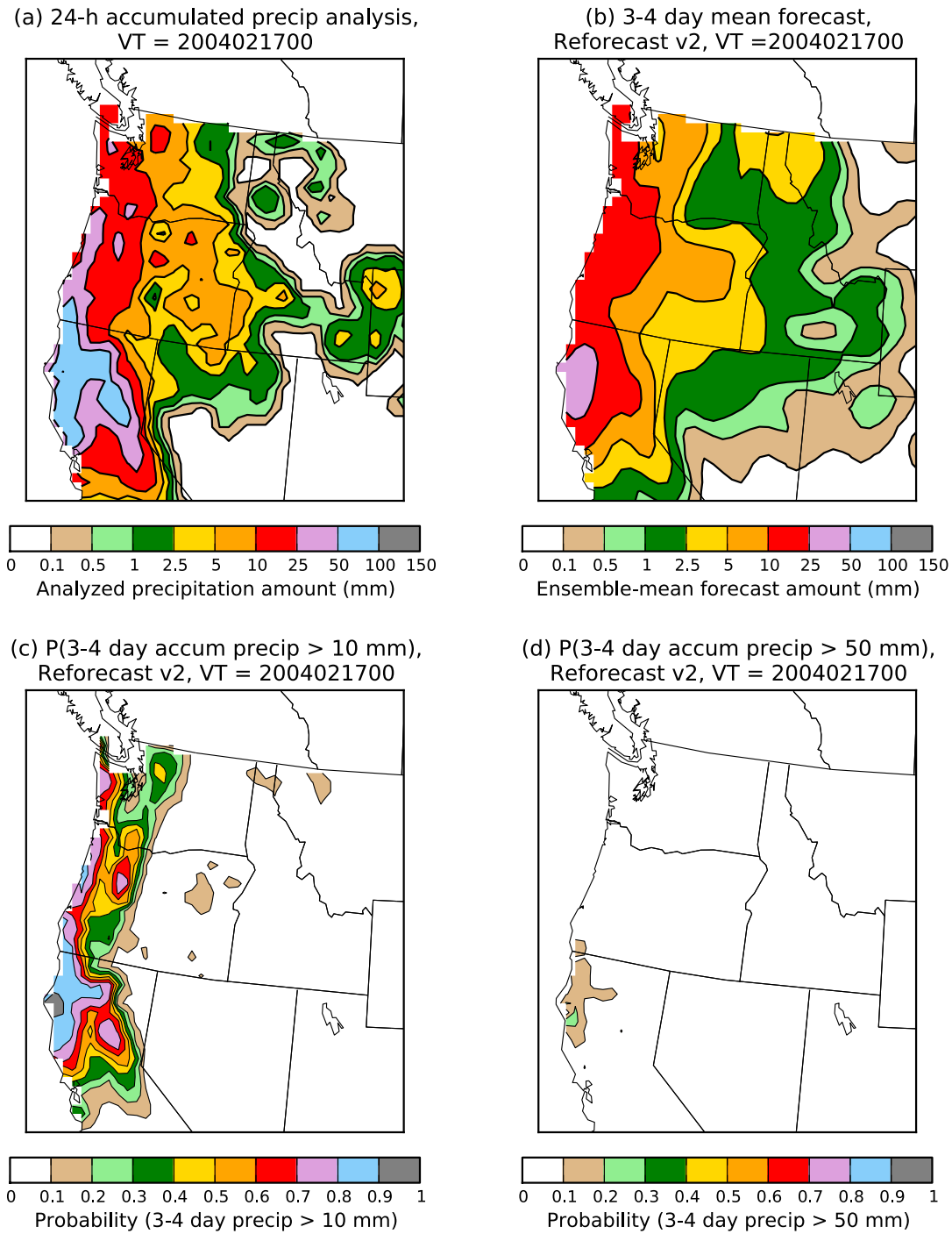
Source: http://www.esrl.noaa.gov/psd/atmrivers/events/



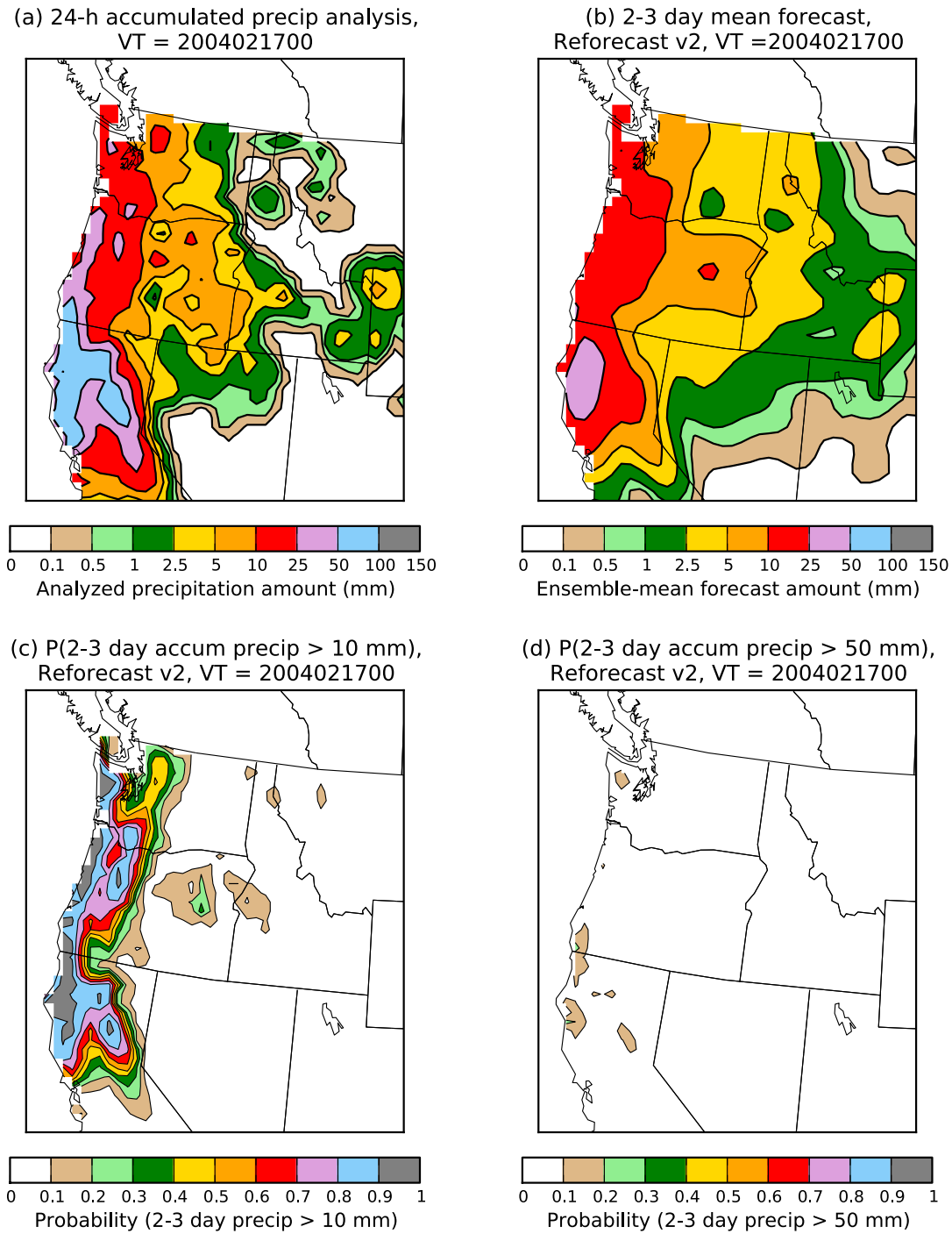
6-day forecast



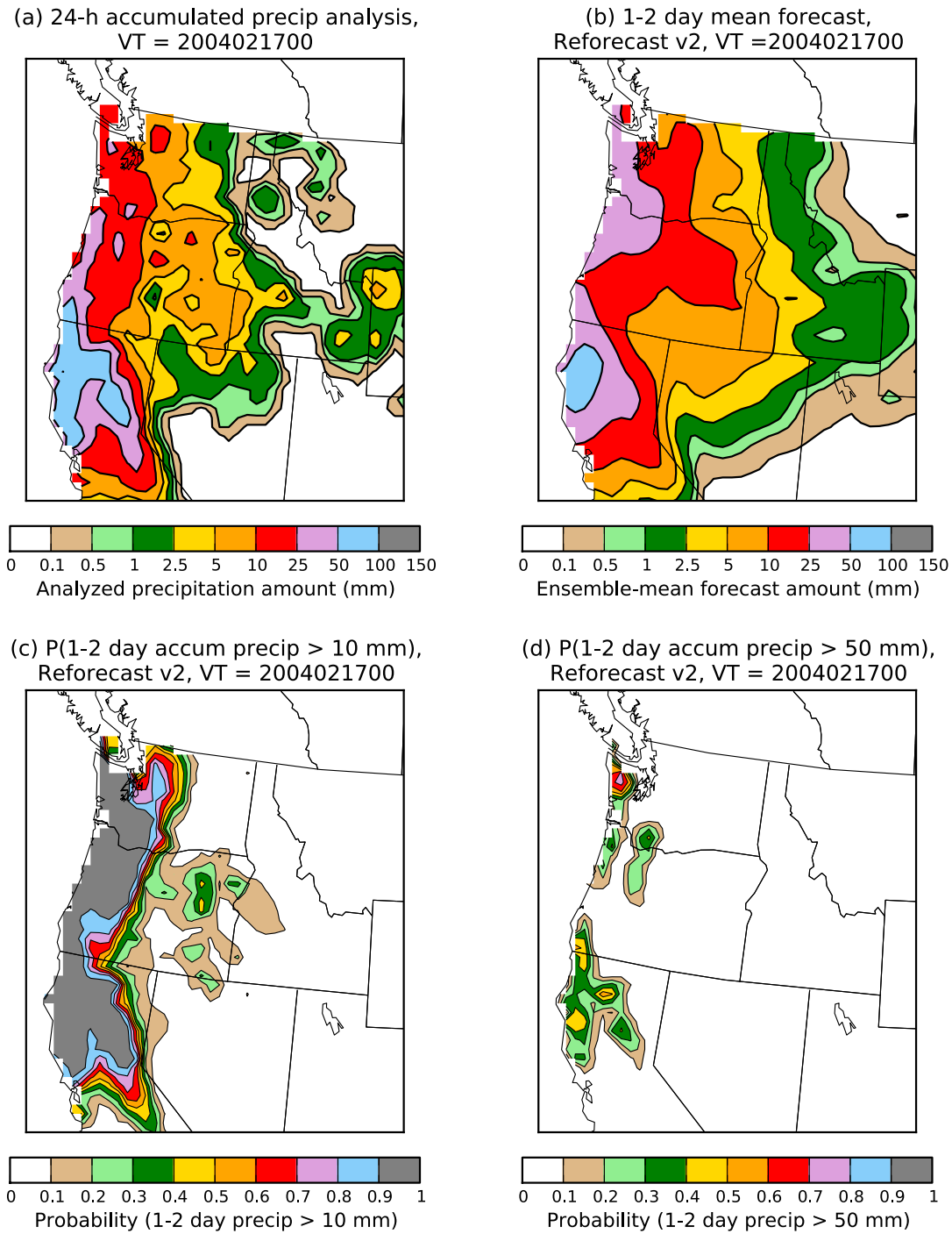
5-day forecast



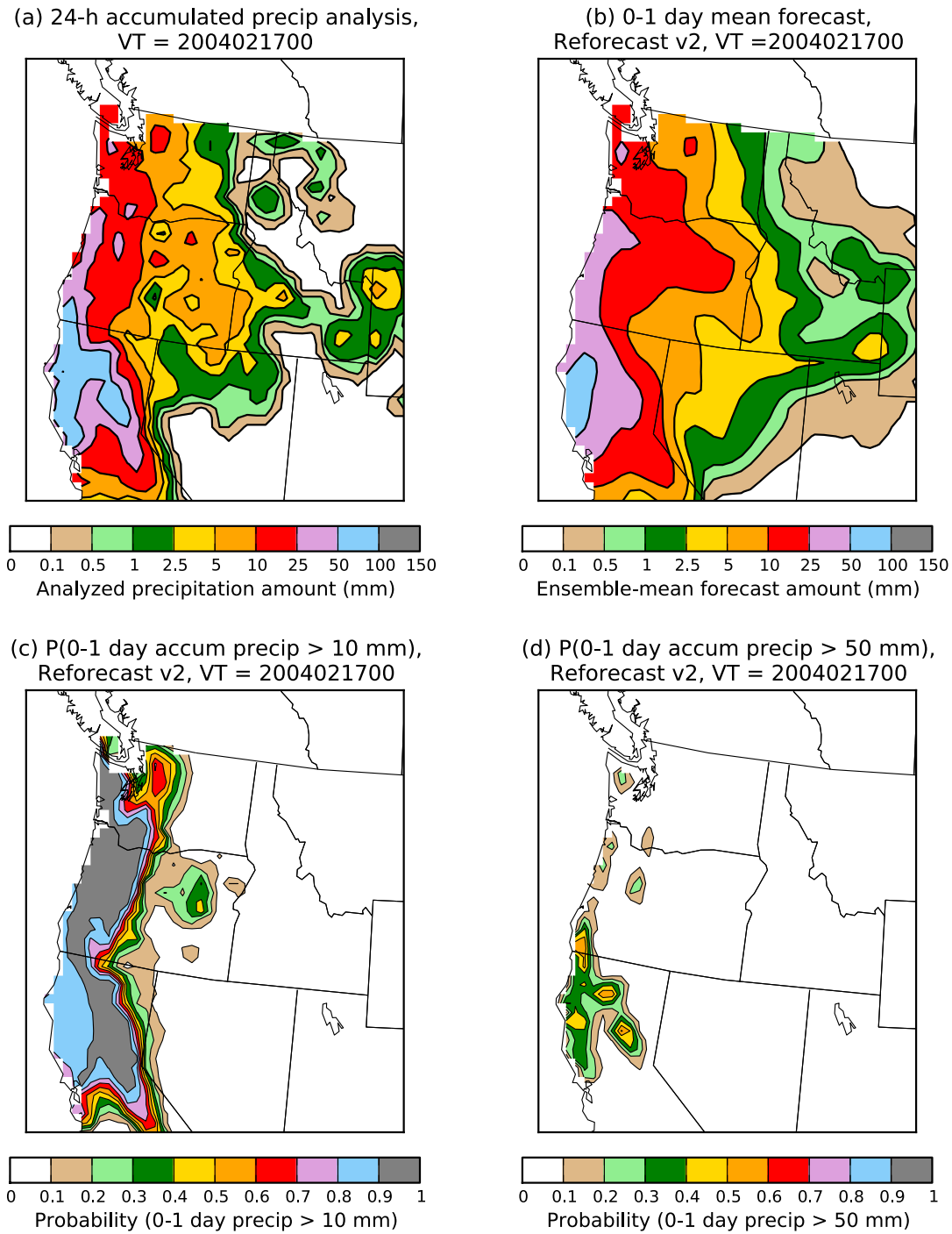
4-day forecast



3-day forecast

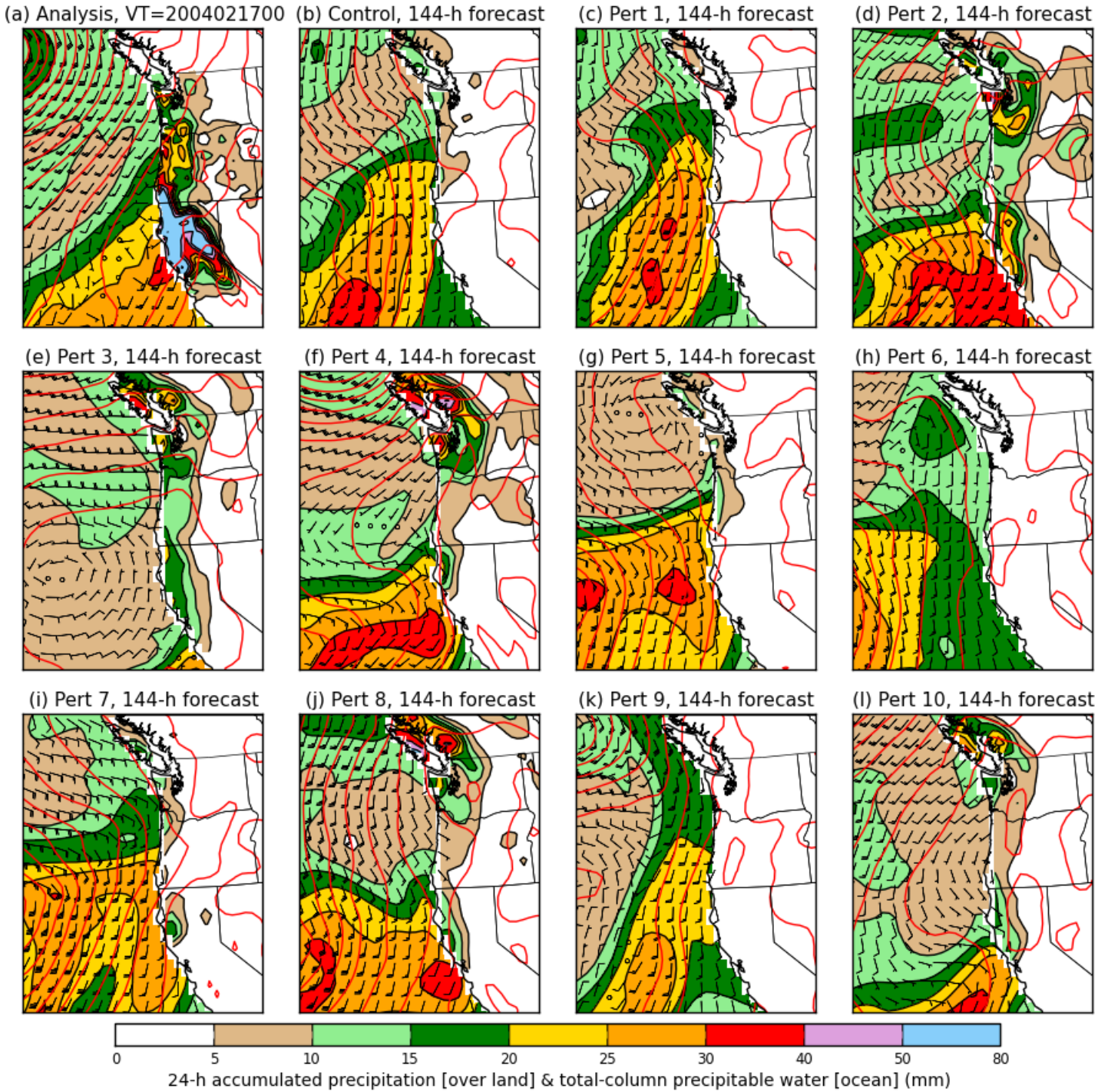


2-day forecast



1-day forecast

Why did it take till so close to the event to start forecasting even moderate probabilities of heavy rain near the Russian River basin in northwest CA?

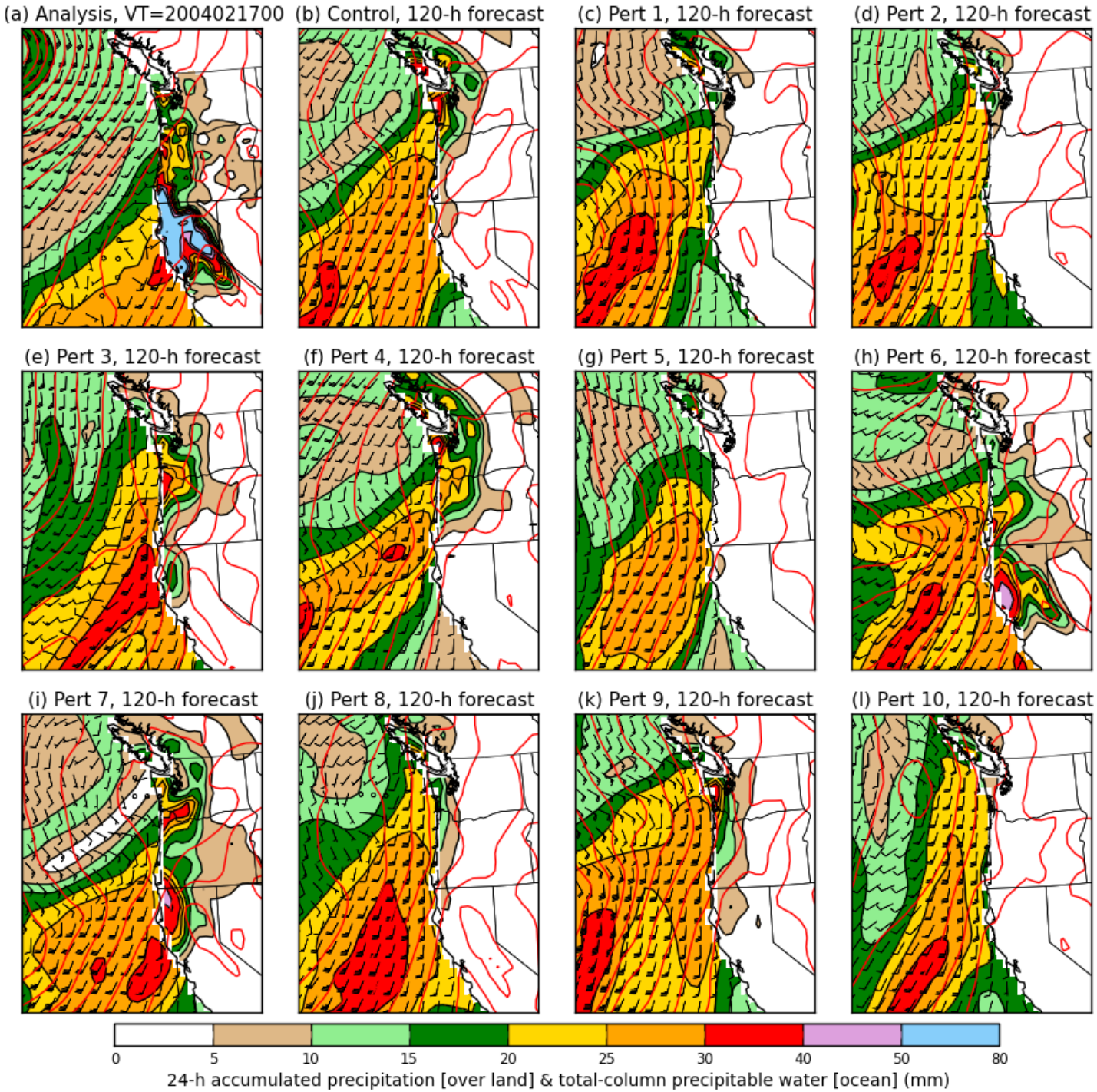


6-day forecast

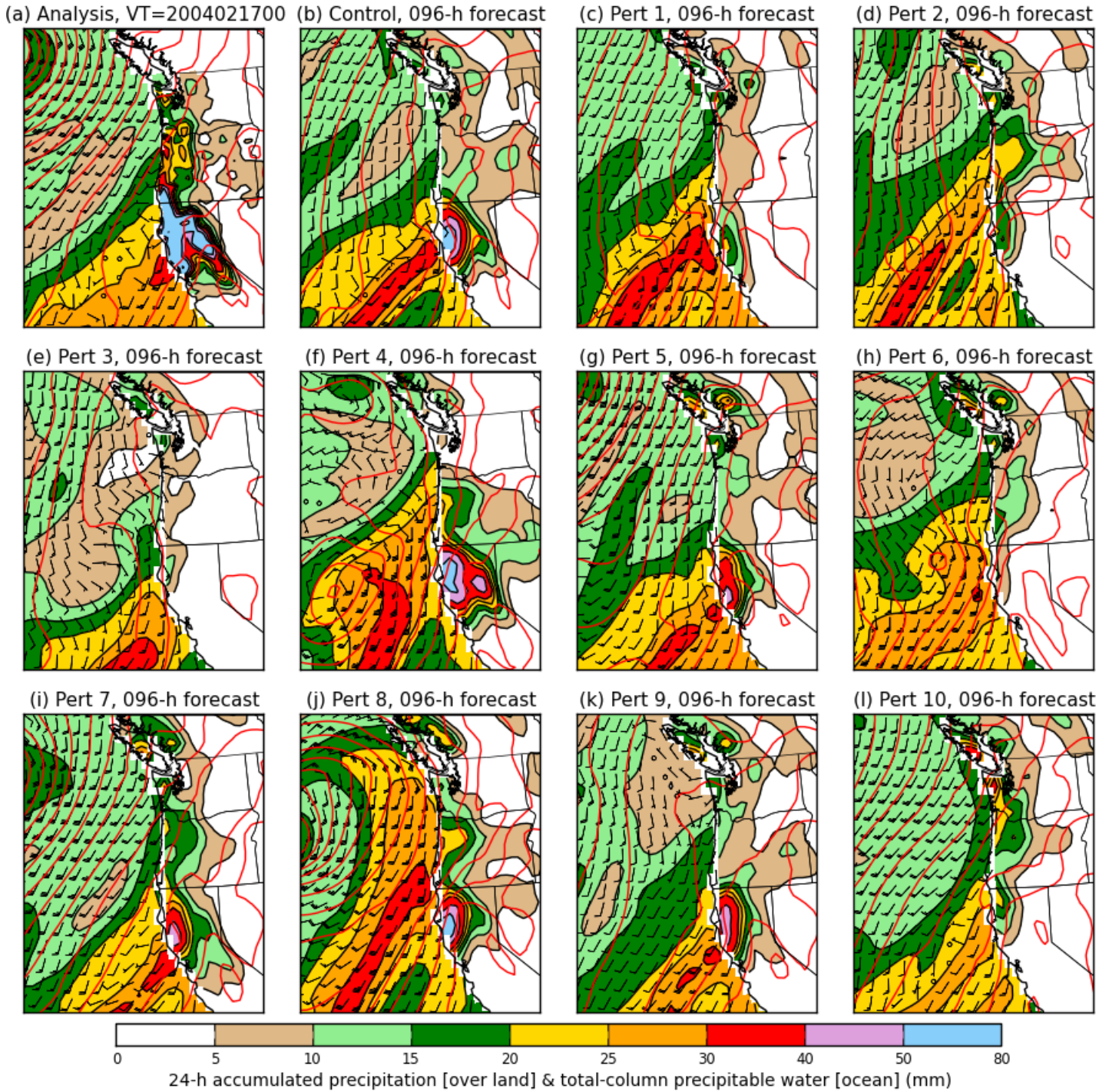
Colors: over the ocean, the total-column precipitable water. Over land, the 24-h accum. precipitation.

Wind barbs for the 925 hPa level.

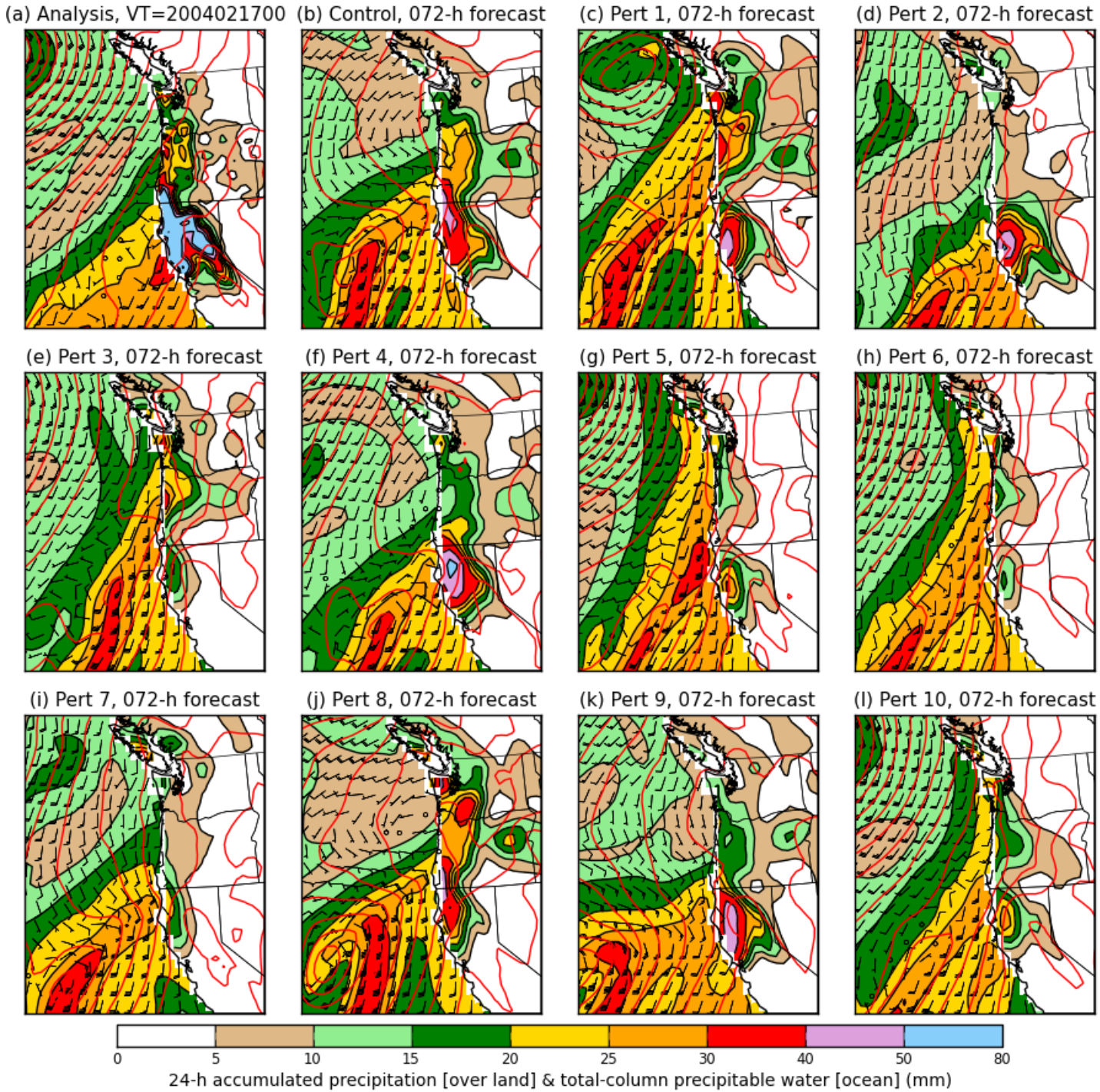
Red contours: mean sea-level pressure.



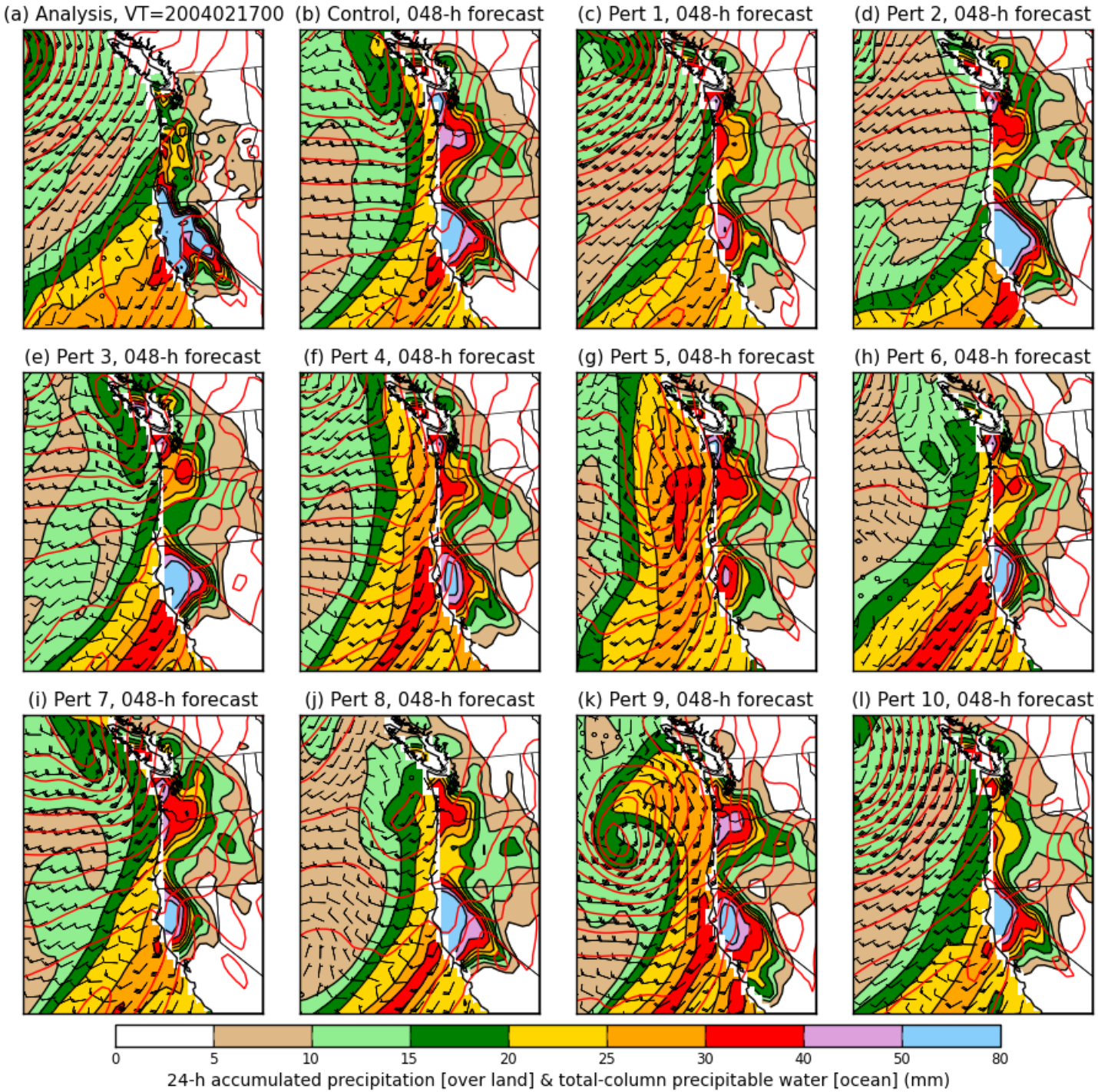
5-day forecast



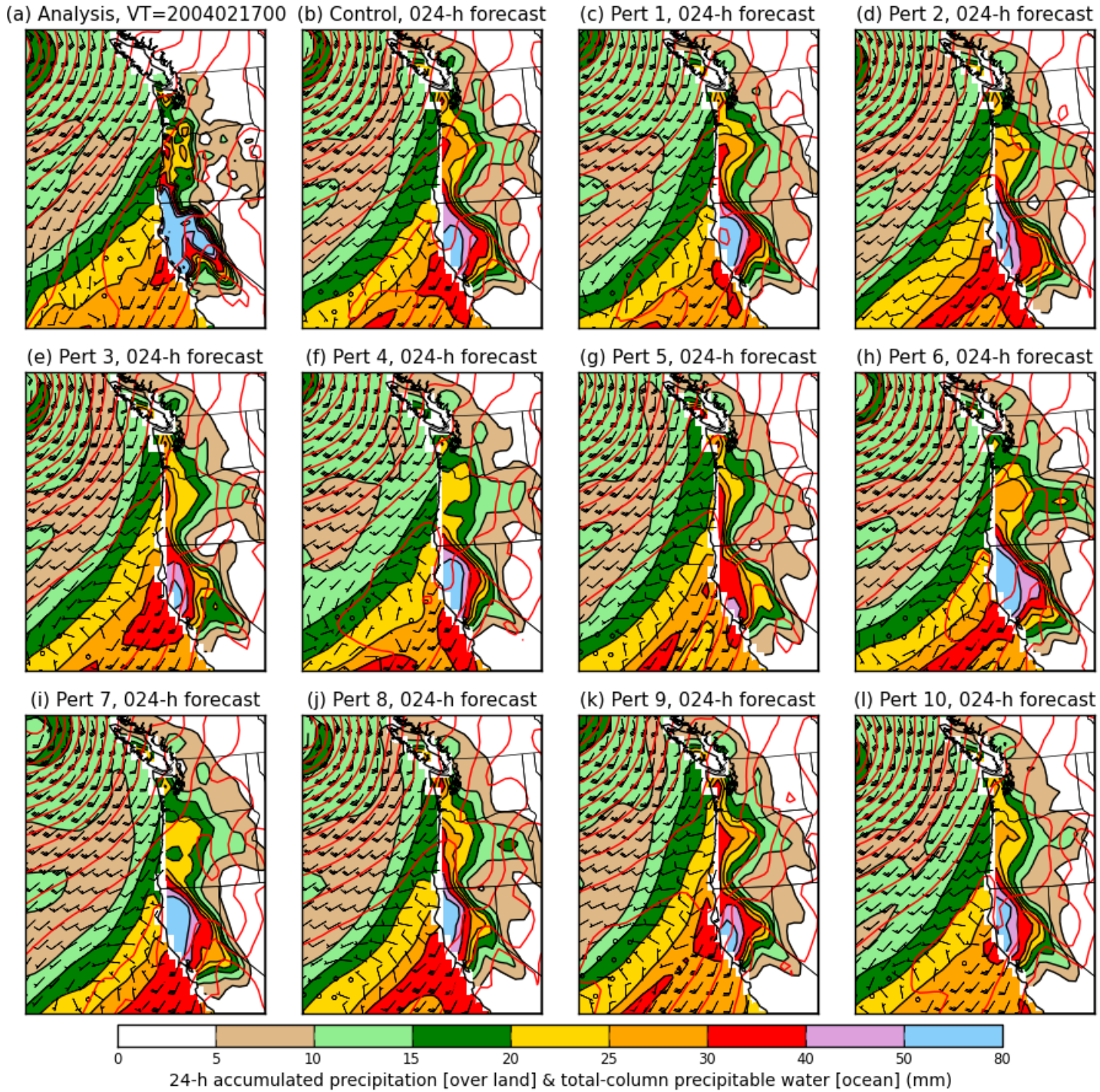
4-day forecast



3-day forecast



2-day forecast

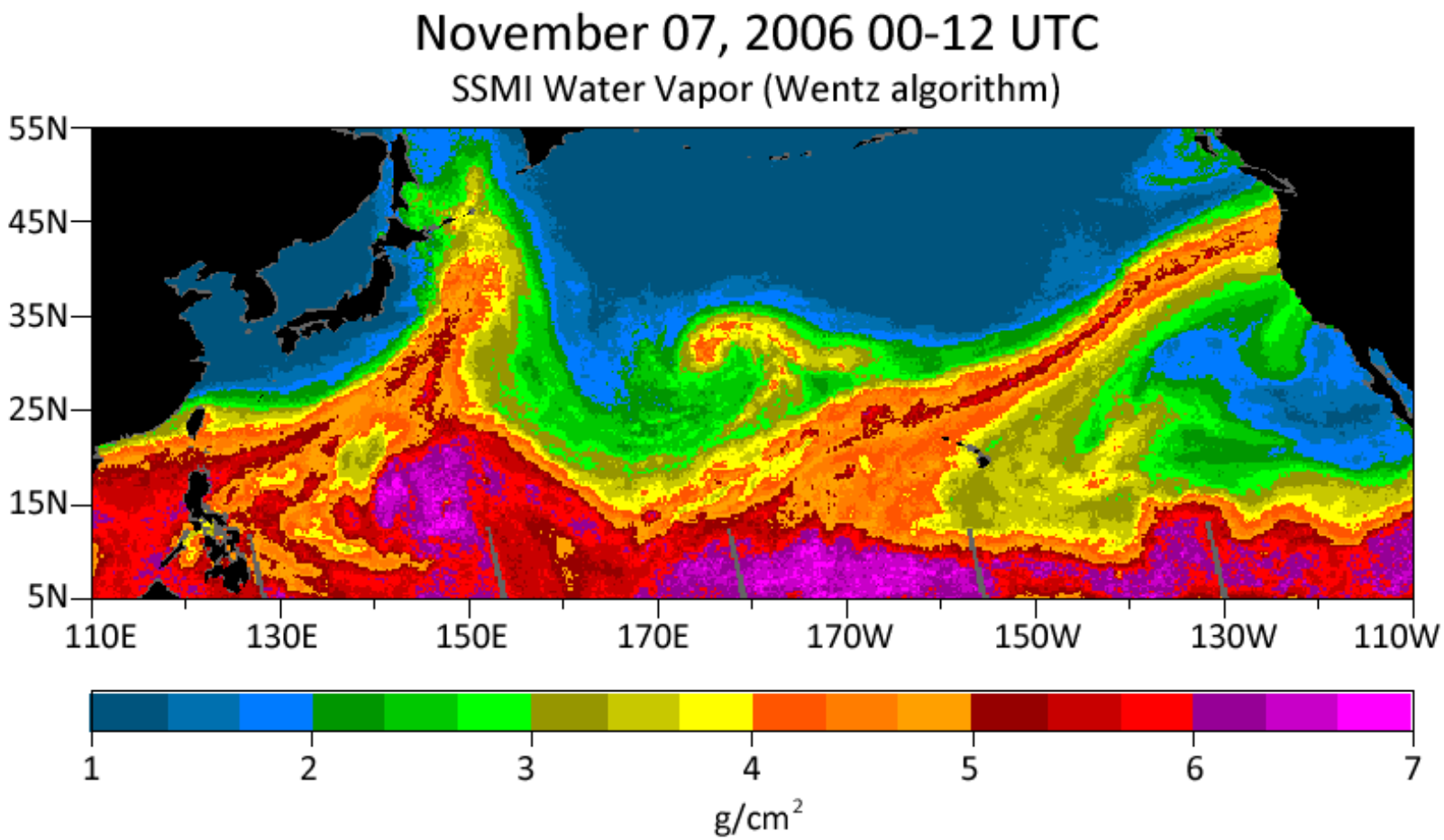


1-day forecast

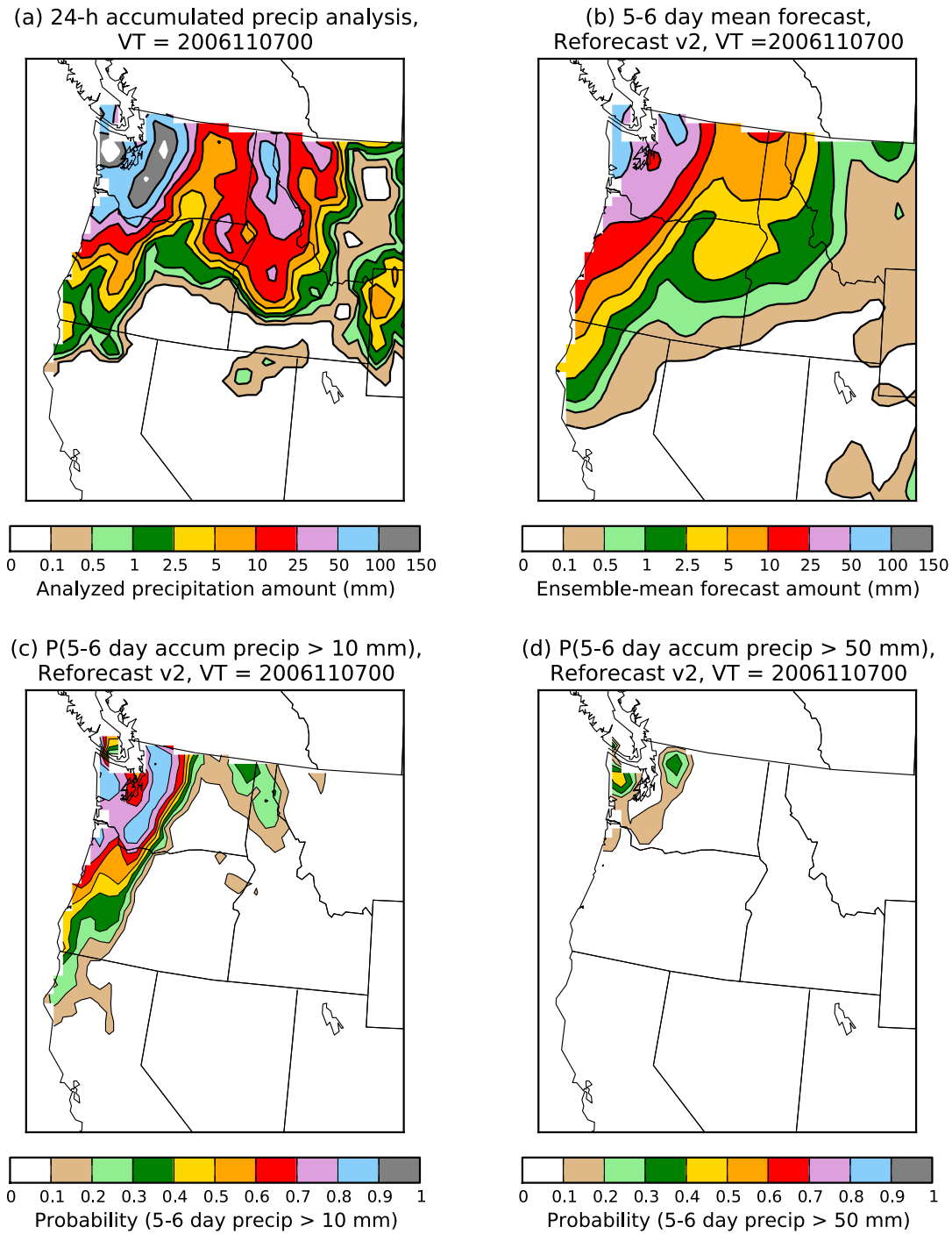
Case study, tentative conclusions

- Statistical post-processing will not be able to correct for everything. In this case, the synoptic-scale predictability was apparently quite low.
- Improvements to post-processed probabilistic forecasts in such a case will require improved ensemble guidance.

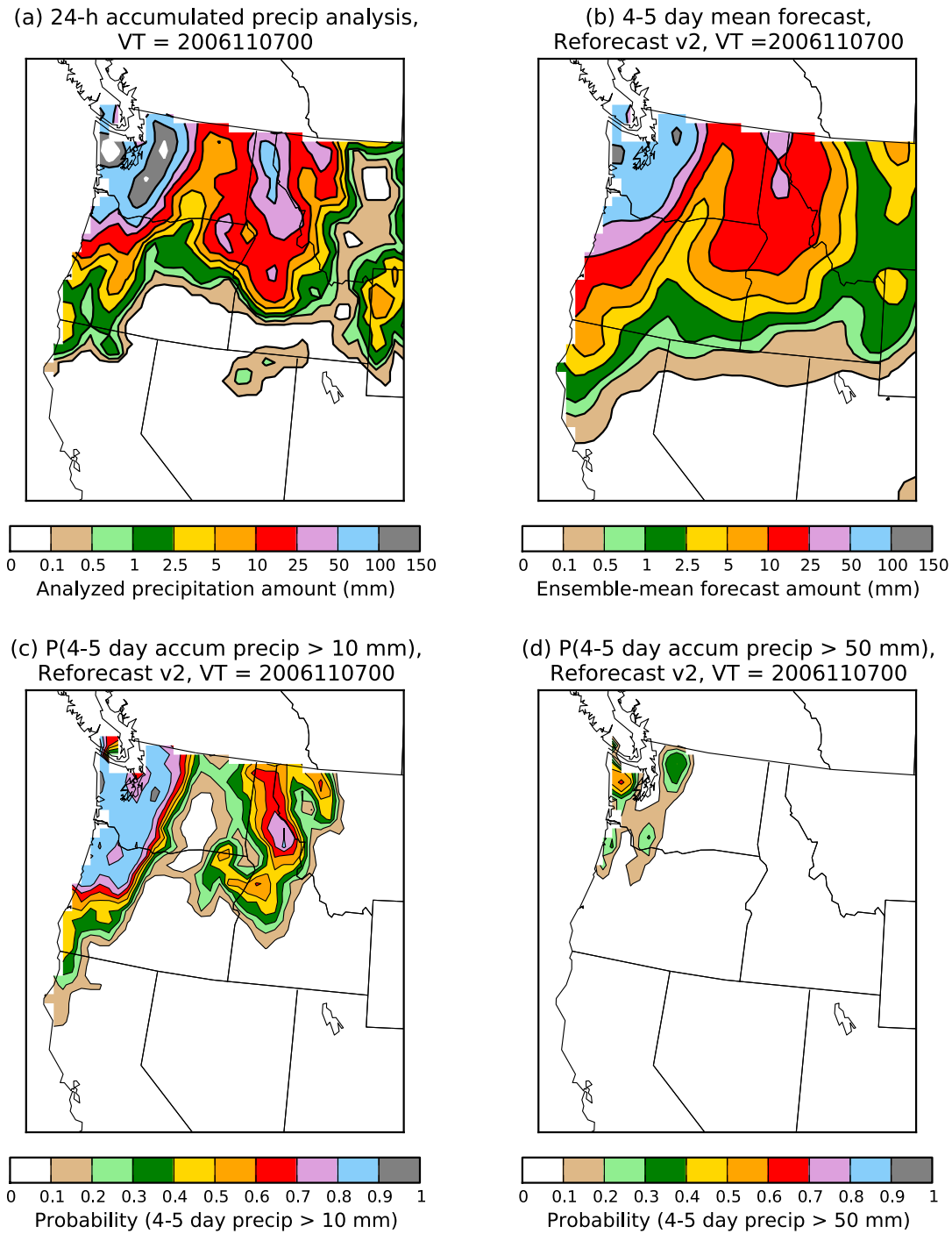
The “good” atmospheric rivers case study:
Nov 2006 Oregon-Washington floods



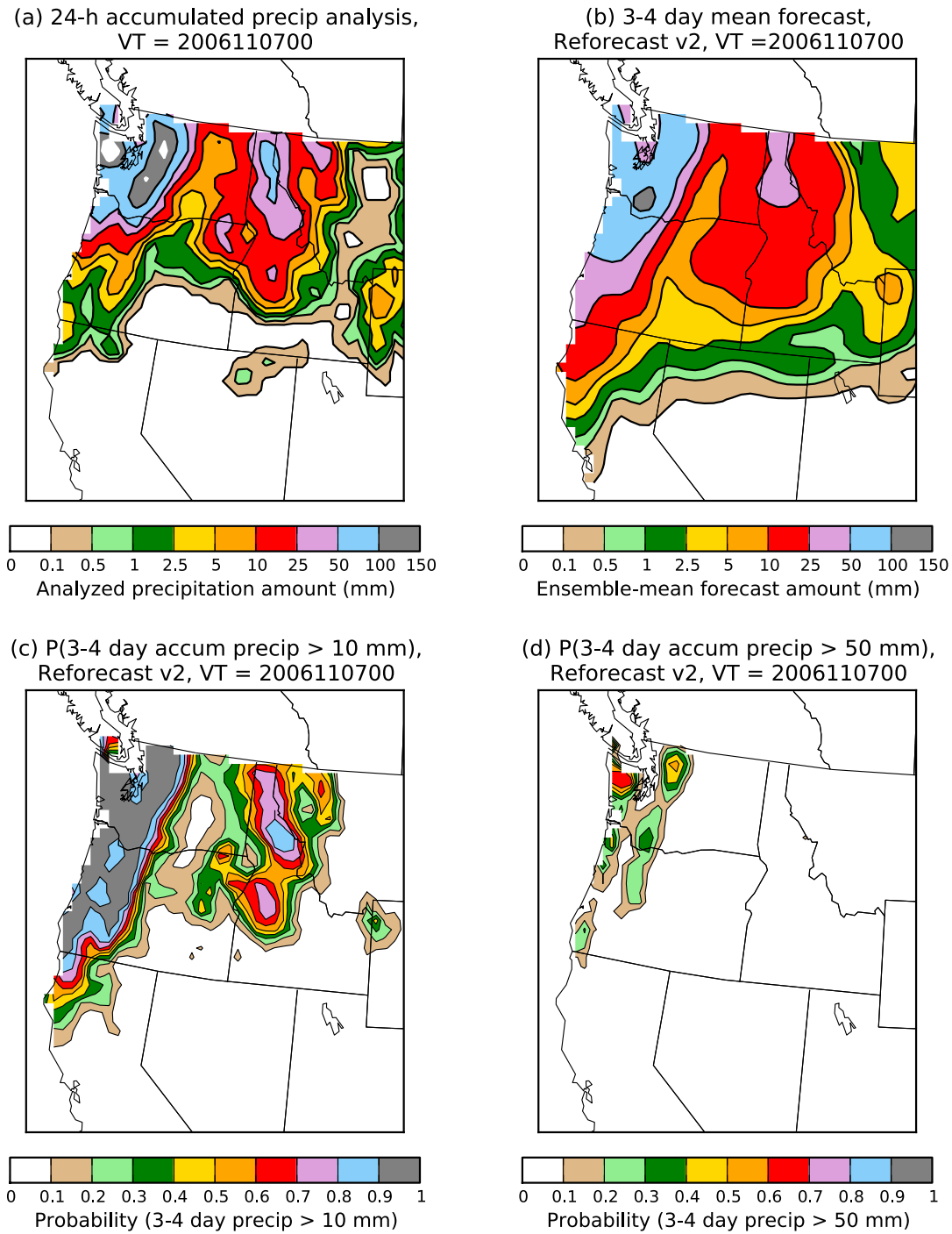
8-20 inches of rain in Cascades; flooded rivers; extensive damage to Mt. Rainier NP.



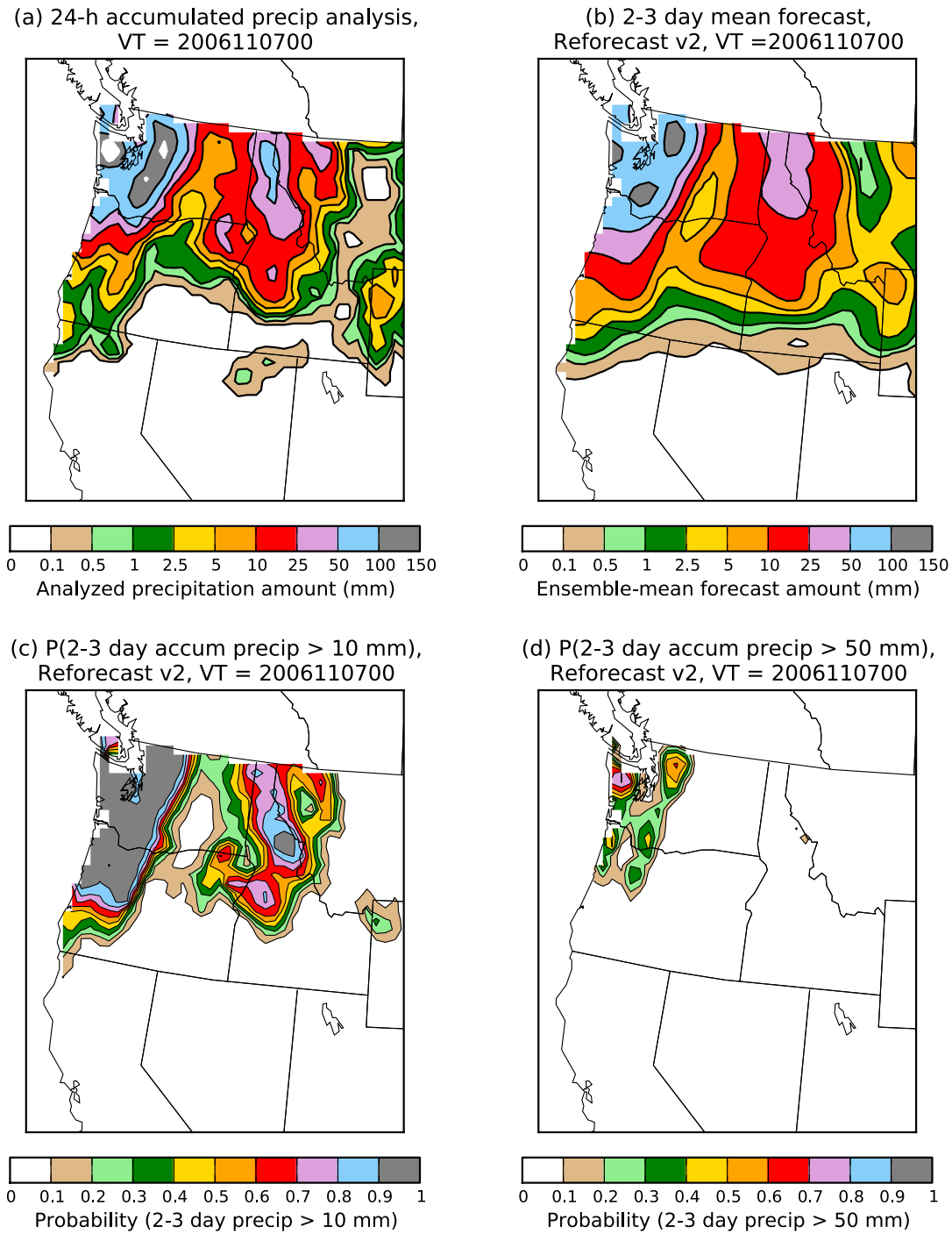
6-day forecast



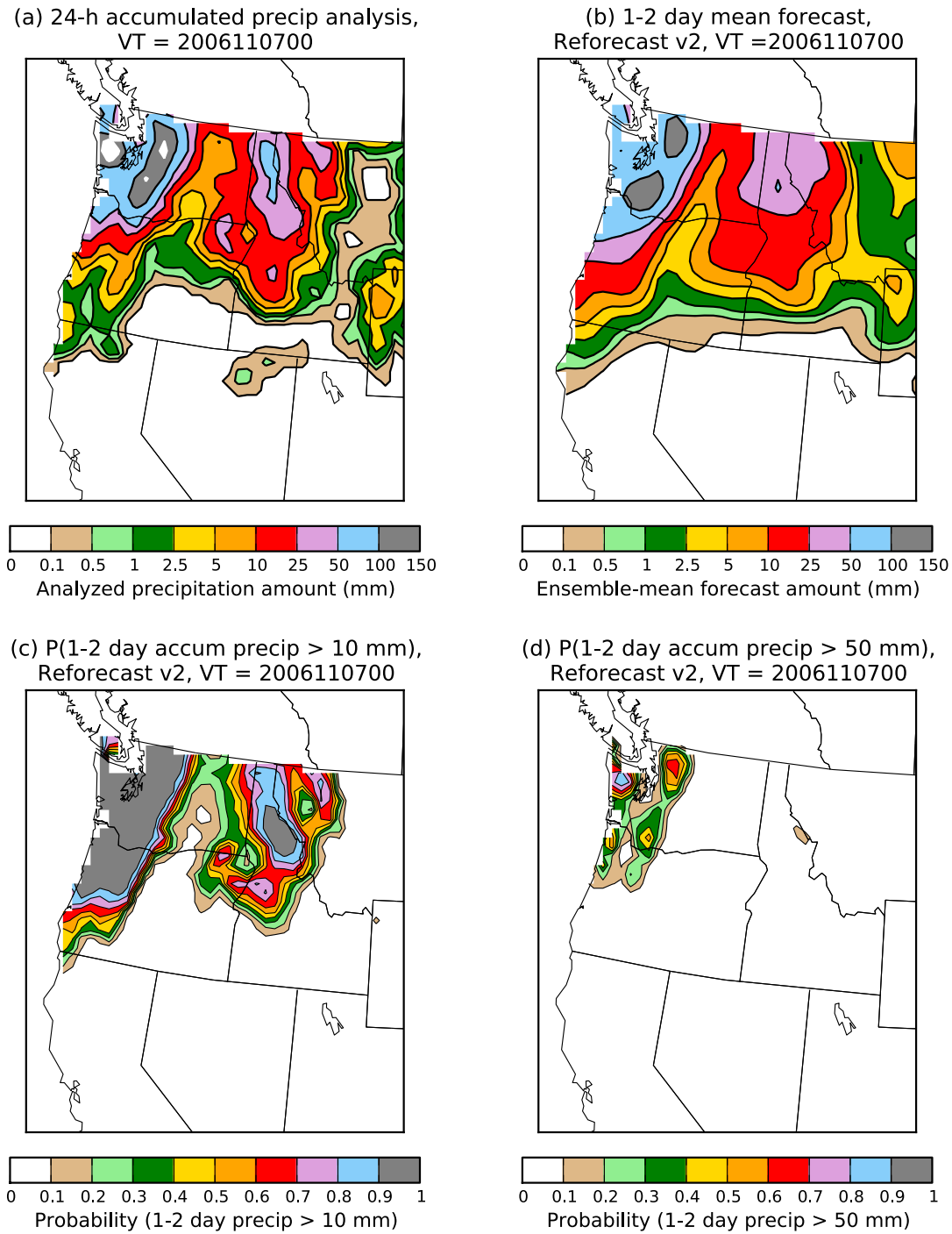
5-day forecast



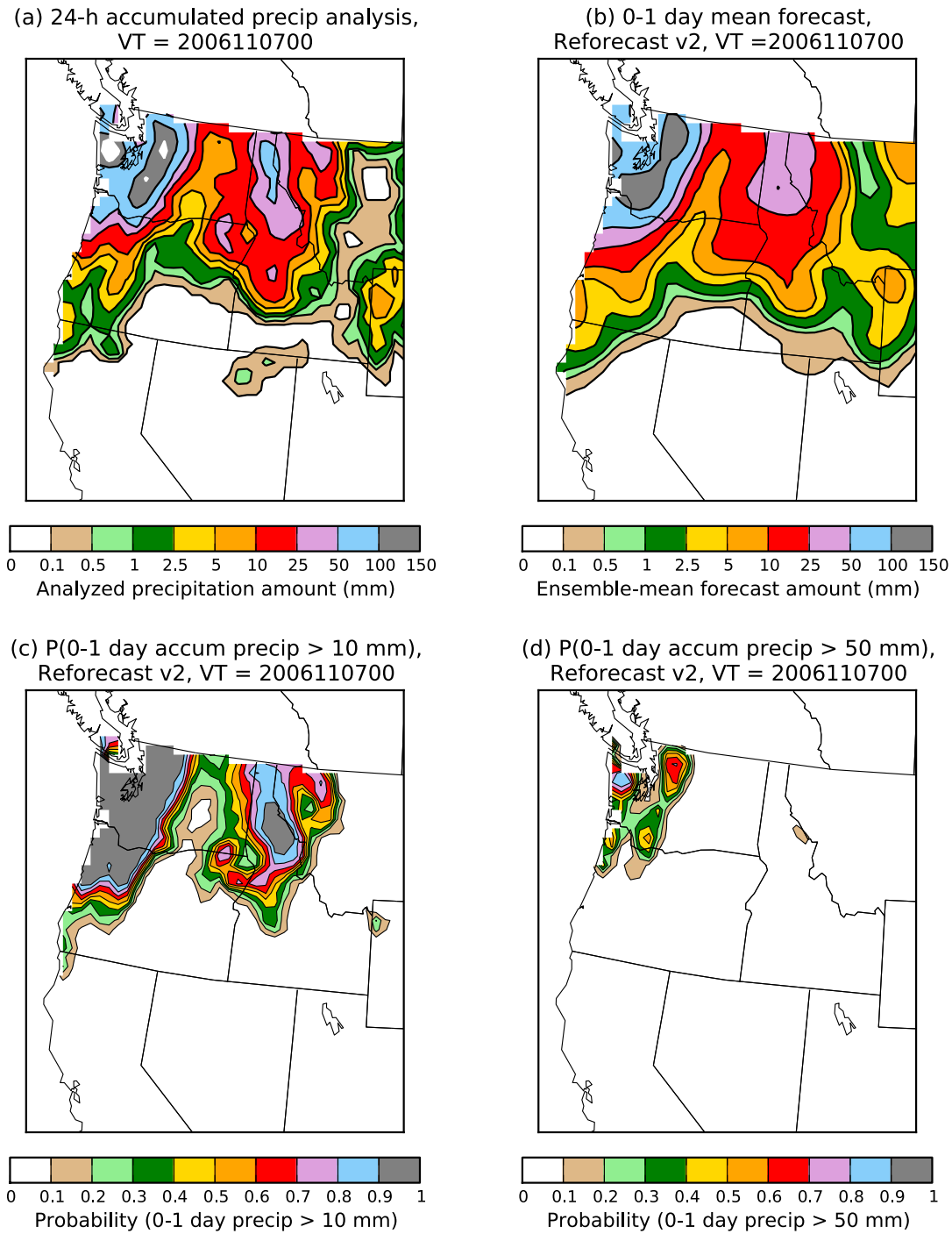
4-day forecast



3-day forecast

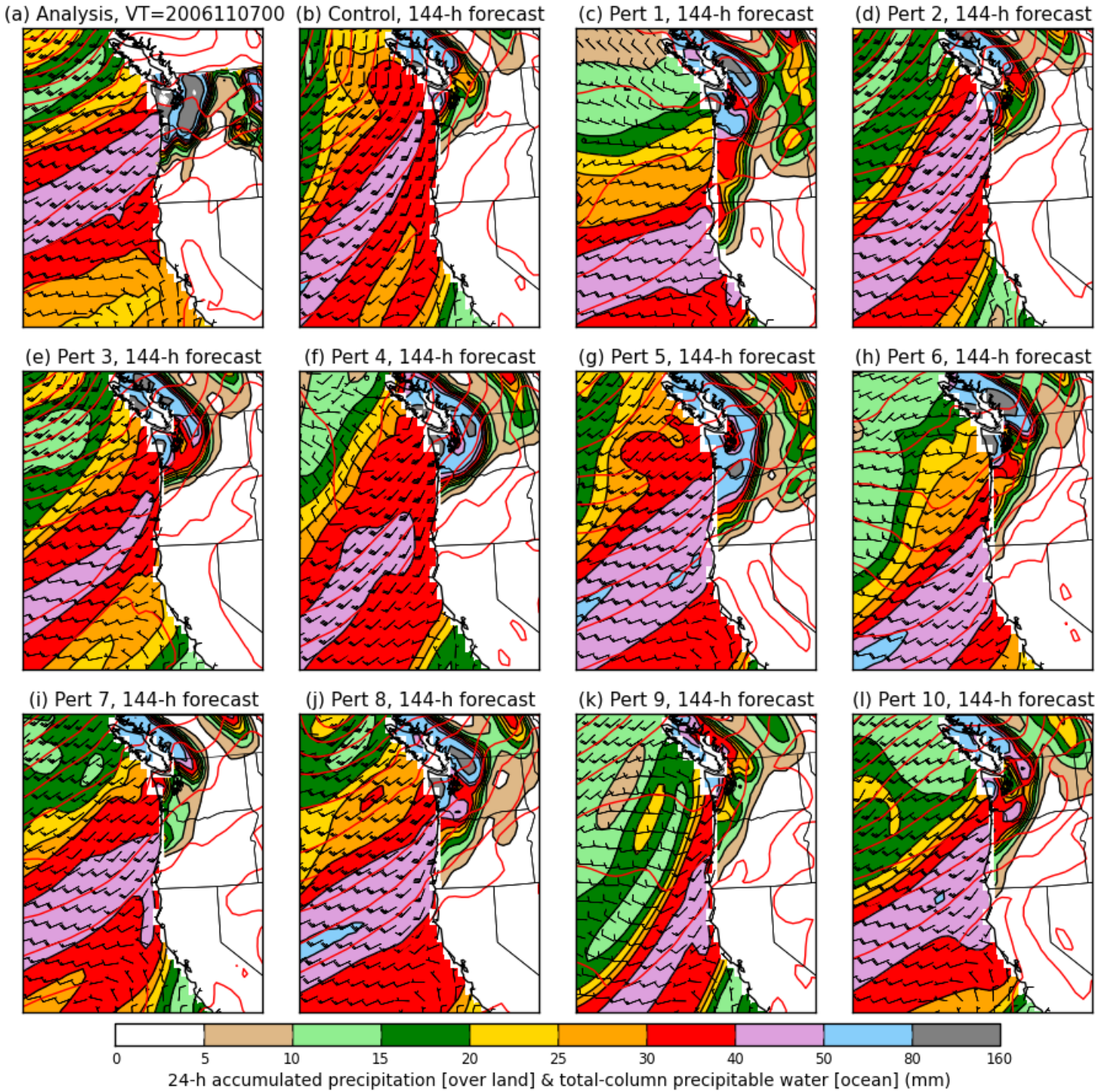


2-day forecast



1-day forecast

Comparatively, this event was much more consistently forecast as a heavy precipitation event even at the long leads.

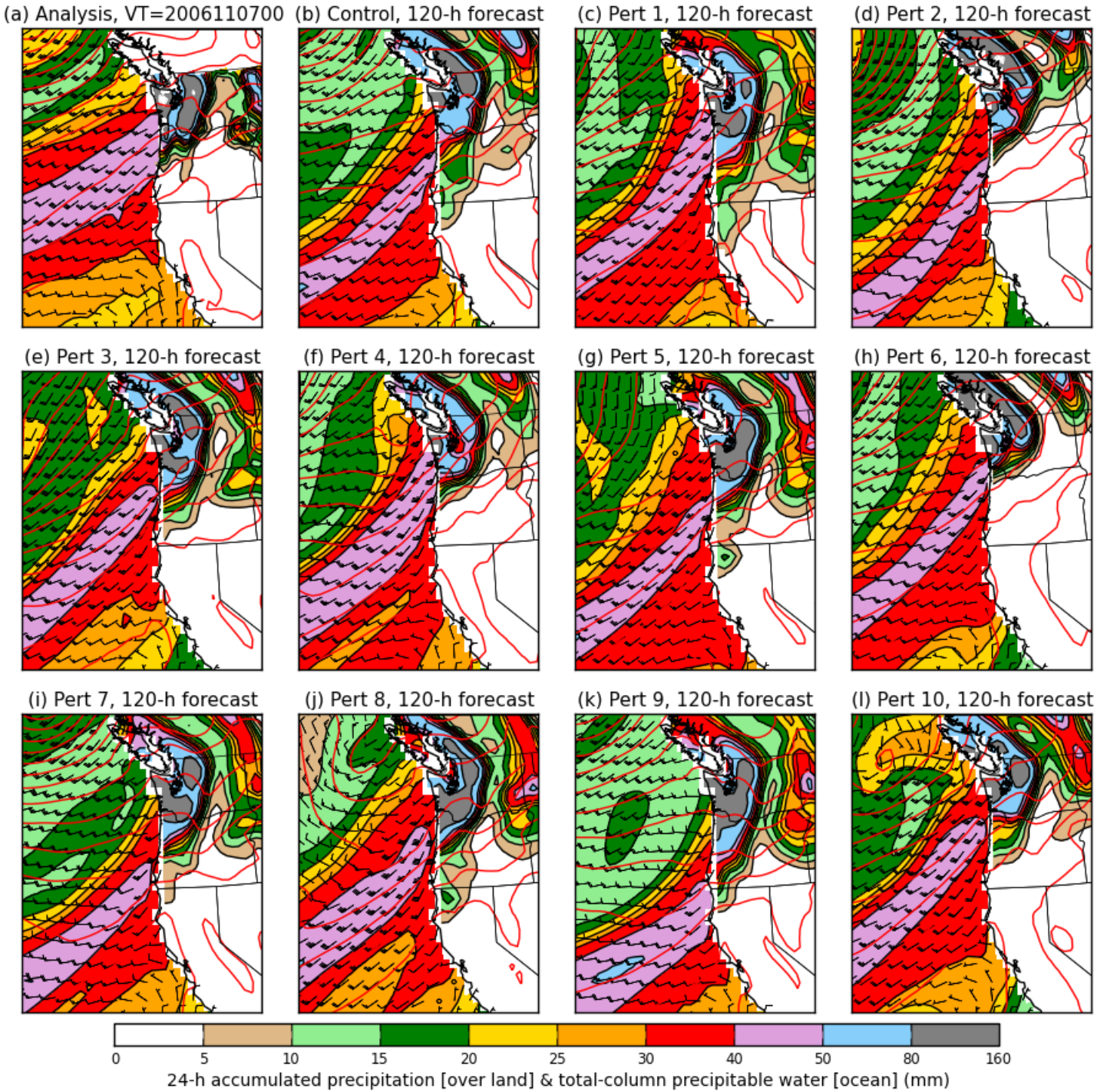


6-day forecast

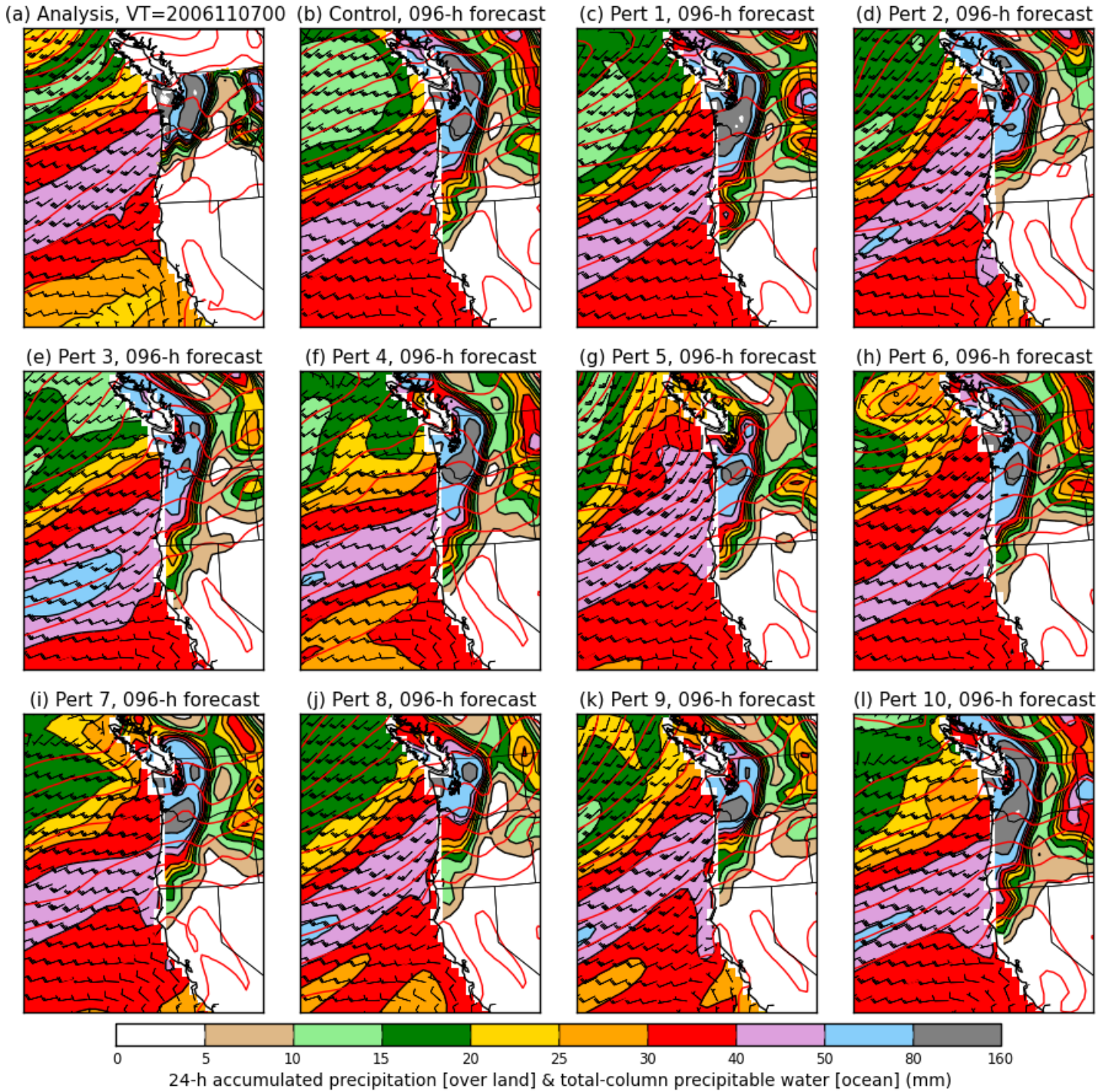
Colors: over the ocean, the total-column precipitable water. Over land, the 24-h accum. precipitation.

Wind barbs for the 925 hPa level.

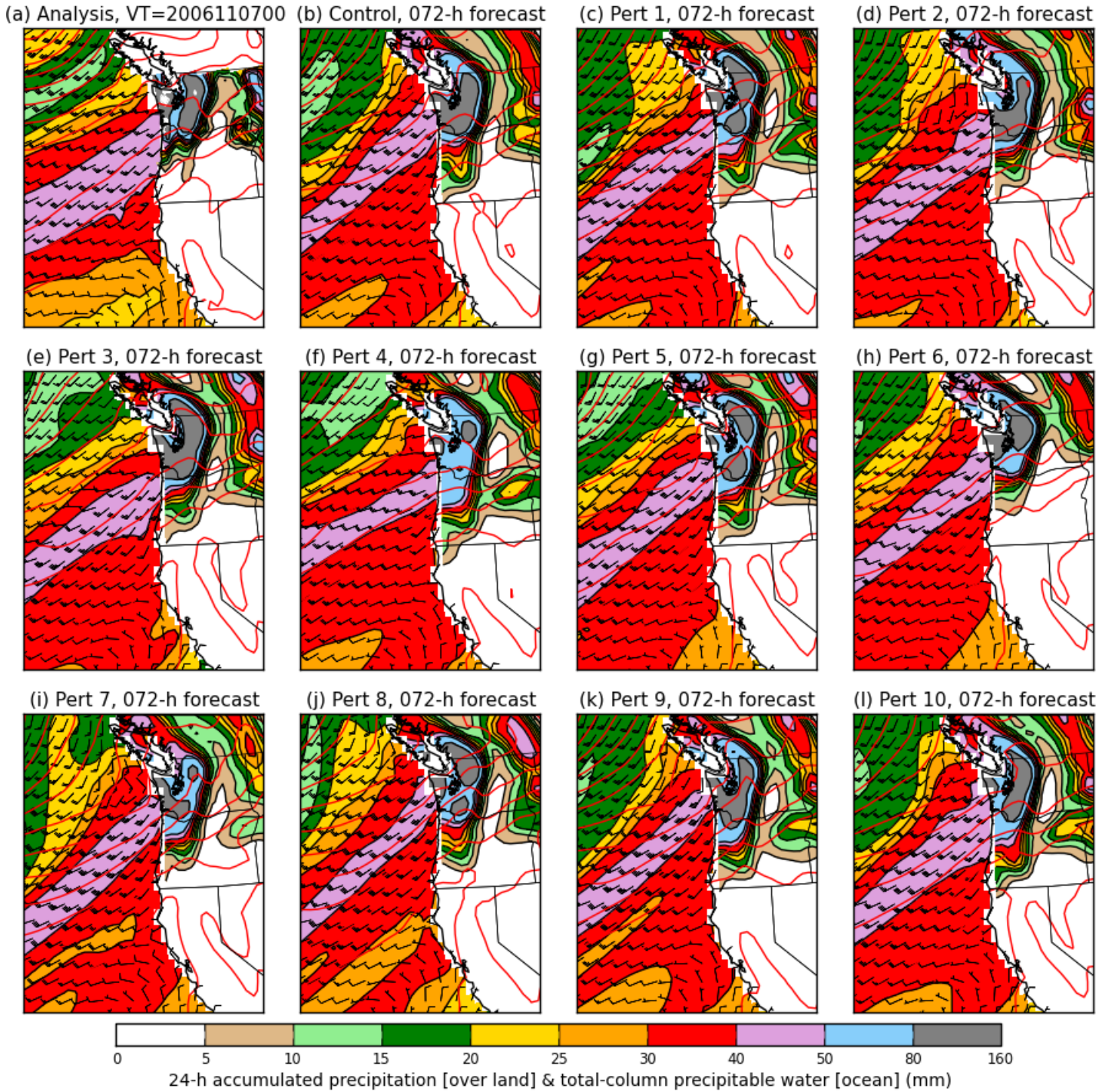
Red contours: mean sea-level pressure.



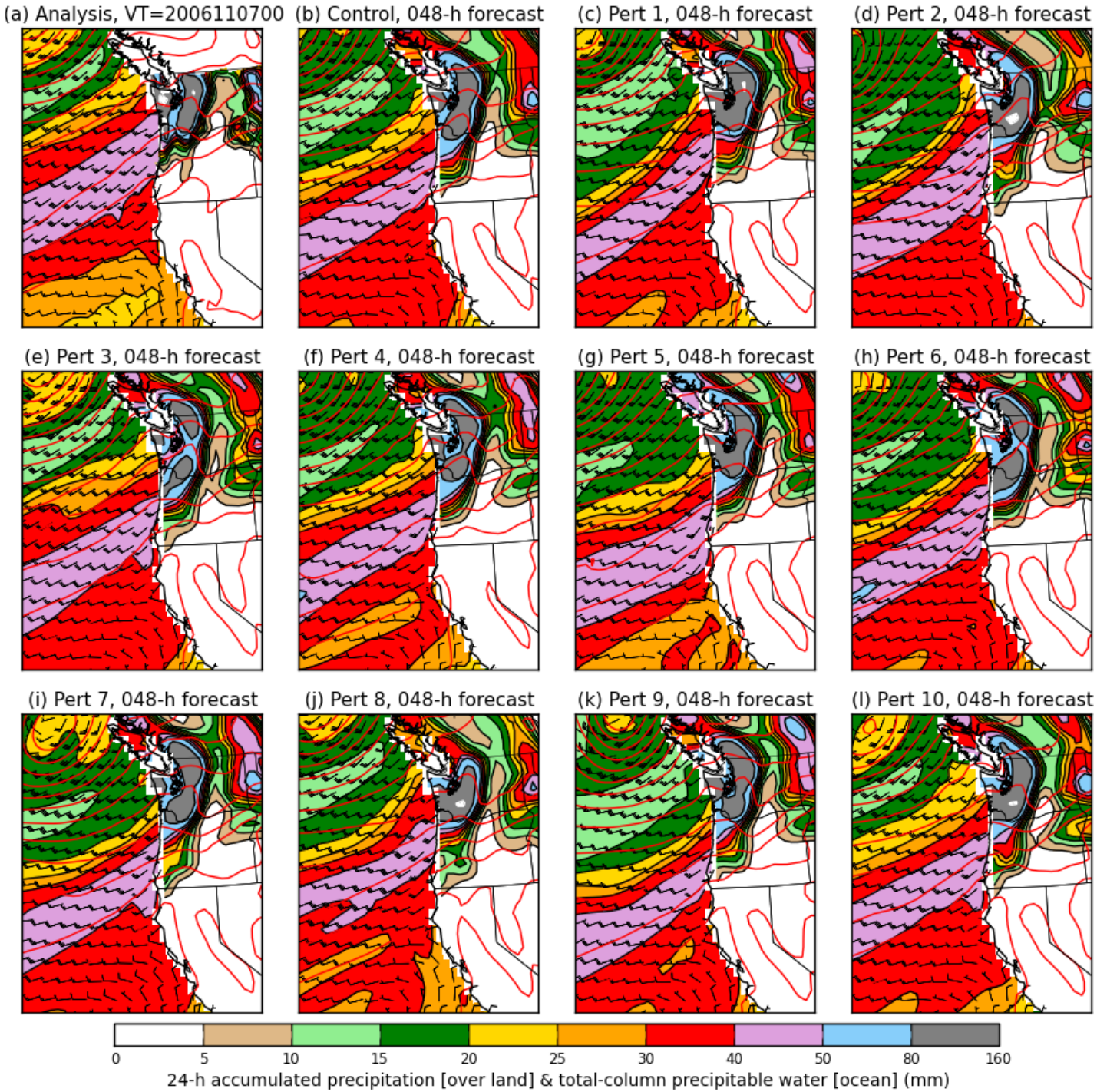
5-day forecast



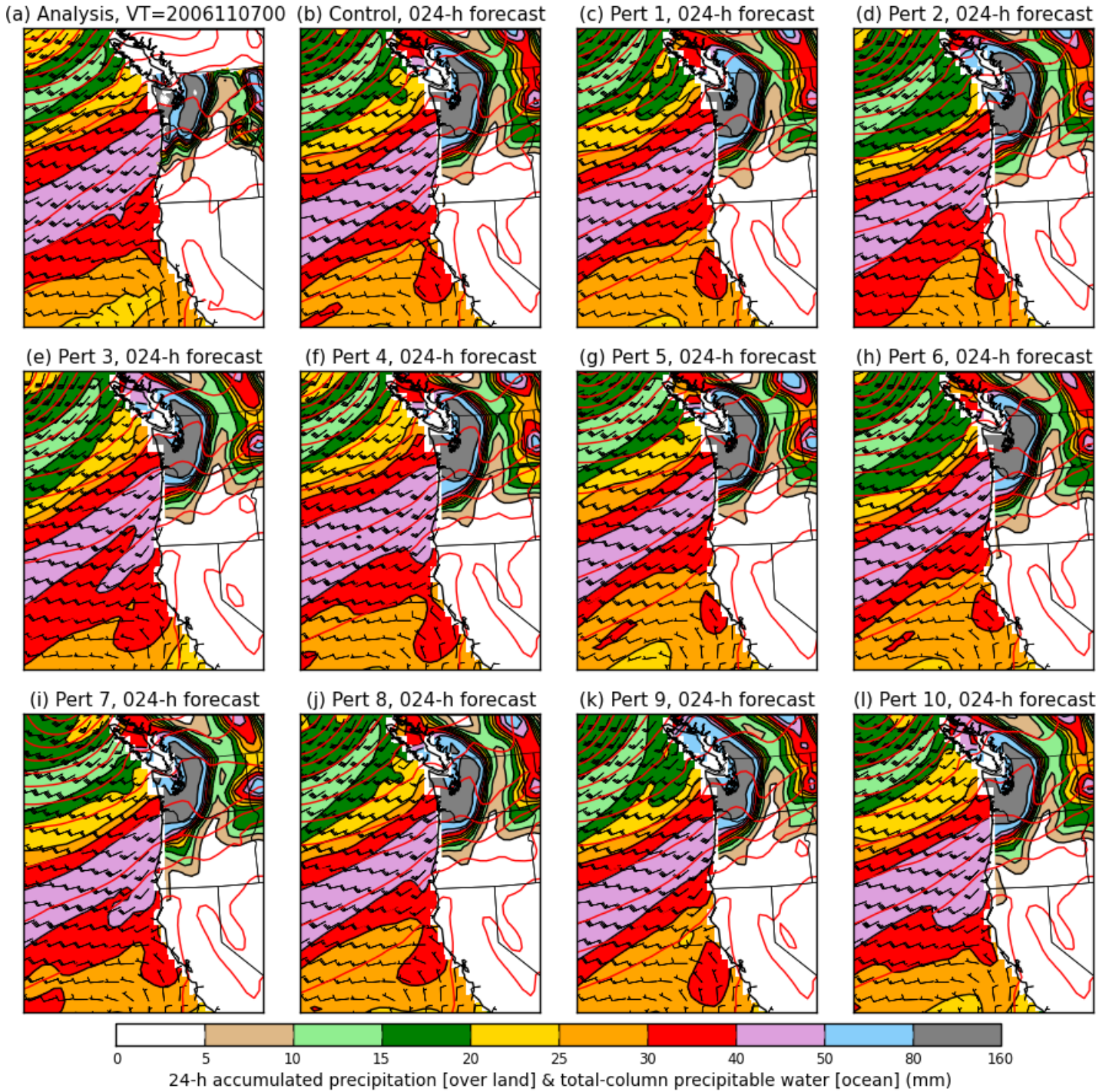
4-day forecast



3-day forecast

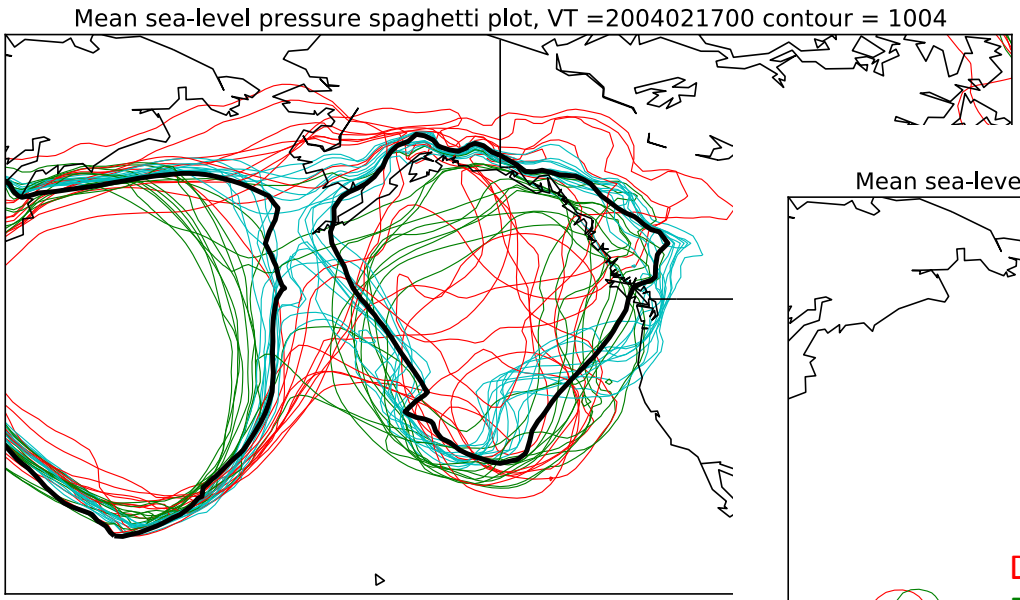


2-day forecast

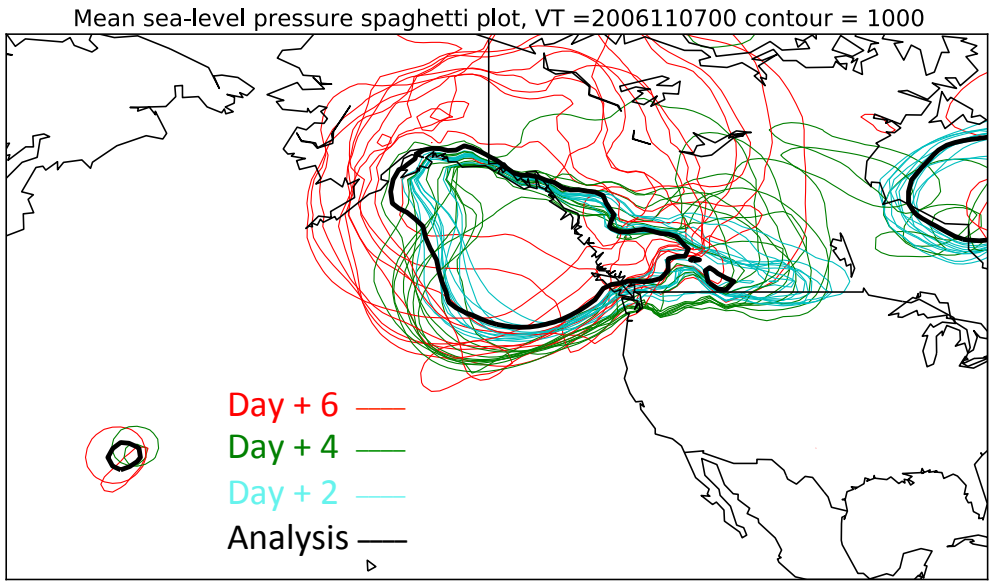


1-day forecast

Spaghetti Westerns



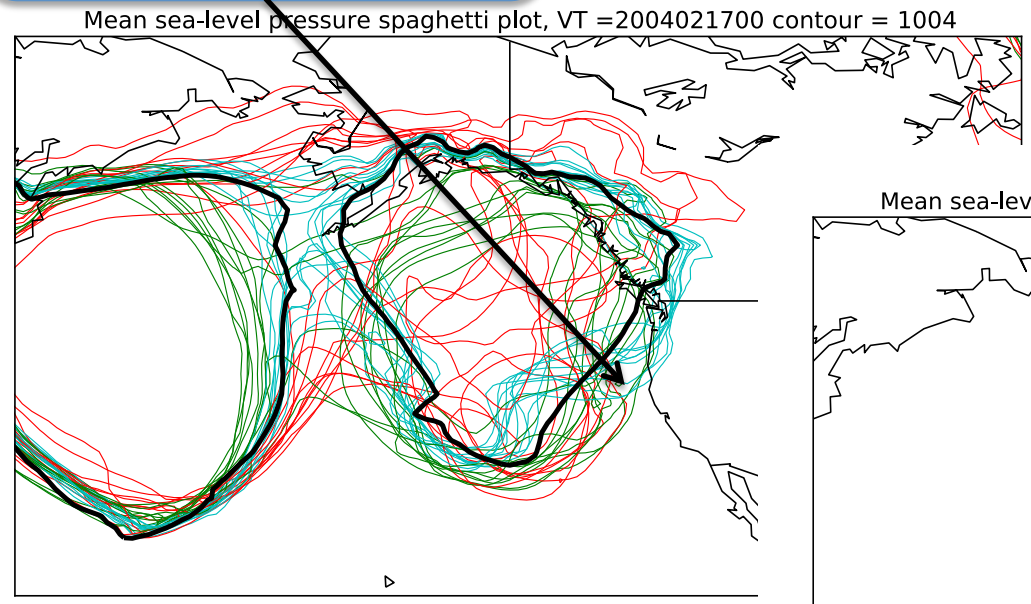
The bad and the ugly



The good

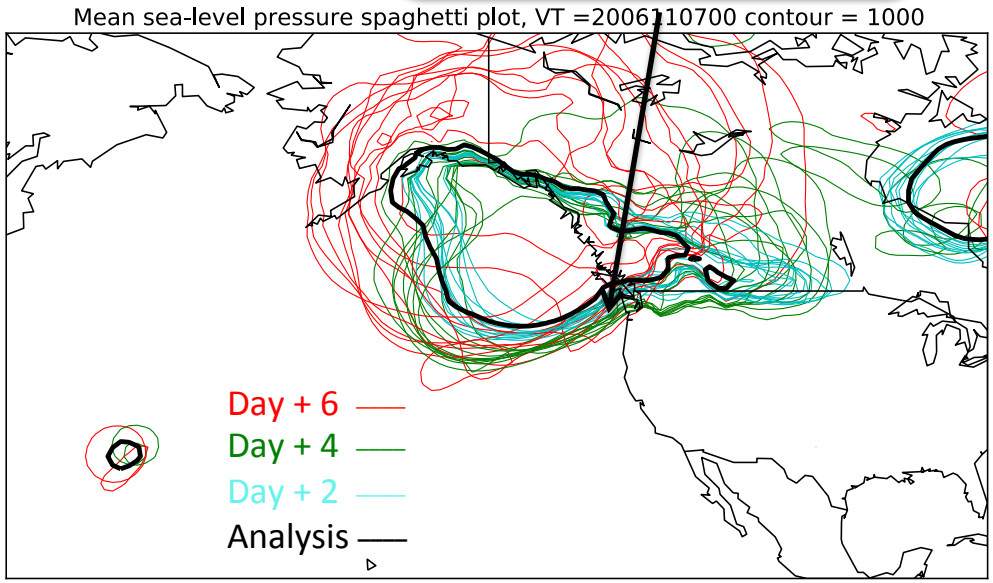
Spaghetti Westerns

Inconsistent forecasts of orientation of isobars and hence wind direction



The bad and the ugly

Consistent forecasts of orientation of isobars and hence wind direction



The good

Fruitful avenues for further near-term work in ESRL/PSD

- Verification on more rivers cases.
- Development of web-based applications to show relevant real-time ensemble forecast data and post-processed information.
- Development of refinements to statistical post-processing algorithms.
  - Appropriate statistical downscaling to ~4 km, and validation of forecasts at this resolution.
  - Use of alternative verification data sets
  - Additional predictors, different methods.

Conclusions

- Reforecasts should help with longer-lead precipitation forecasts in the western US.
- Web-based applications for displaying real-time ensemble data and post-processed data ought to be a priority.

other slides

ESRL interface to data

DOE/Lawrence Berkeley interface

Facilitating regional reforecasting

- With full model states downloaded from DOE site, can initialize models like WRF to produce regional reforecasts.

- [insert WRF picture]

Some known applications in development

- Real-time web page of experimental PQPF forecasts (ESRL, in development to replace v1 web page)
- Development of experimental forecast products for renewable energy sector (Hamill, ESRL).
- 6-10 day and 8-14 day forecasts (Dan Collins, CPC).
- Detecting and correcting bias in hurricane track and intensity (Jiayi Peng, NCEP/EMC, Tom Galarneau, NCAR).
- Objective probabilities for severe weather several days to weeks hence (Francisco Alvarez, St. Louis Univ. + Hamill)
- Improving hydrologic predictions (Haksu Lee et al., NOAA/NWS/OHD).
- Also: BAMS article once published will presumably spur wider usage.