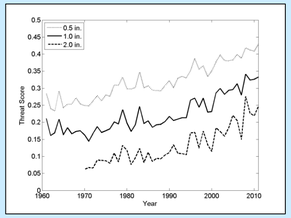
Extreme Quantitative Precipitation Forecast Performance Benchmarks over 11 Years

E. Sukovich[1,2], F. M. Ralph[1], D. Novak[3], F. Barthold[3,4], and D. Reynolds[1,2]

[1]NOAA/Earth System Research Laboratory, Boulder, CO; [2]University of Colorado, CIRES, Boulder, CO; [3]NOAA/NWS/Hydrometeorological Prediction Center, College Park, MD ; [4]I.M. Systems Group, Inc., Rockville, MD

Motivation

- Many key end-users of quantitative precipitation forecasts (QPF) require accurate forecasts (e.g., location, timing, and amount) of extreme precipitation events.
- The current QPF evaluation method (i.e., > 1 in 24 h^{-1} threat score) is sub-optimal for extreme events, which tend to occur less frequently and over smaller areas than weaker precipitation events.



Annual threat scores for the HPC's 0.50-, 1.0-, and 2.0-in forecasts for day-1 from 1961 through 2011.

Objective

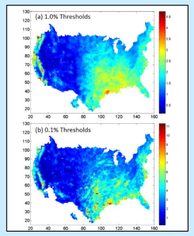
- To baseline and analyze extreme QPF performance over 11 years (2001-2011) using regional extreme precipitation thresholds for lead times of 24 h (day-1), 48 h (day-2), and 72 h (day-3).

Data and Methodology

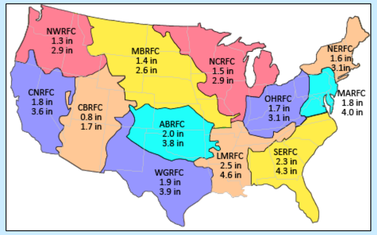
- Data was obtained from the National Precipitation Verification Unit (NPVU)
  - Forecast data: NCEP/HPC's 32-km gridded QPFs
  - Verification data: NWS/River Forecast Centers (RFC) Stage IV data gridded to 32-km
- Analysis period: January 1, 2001 to December 31, 2011
- Verification software: Developmental Testbed Center (DTC) Model Evaluation Tools (MET)

Extreme Precipitation Events

- 99th and 99.9th percentile event thresholds (i.e., top 1.0% and top 0.1%) of all precipitation events from 2001 to 2011



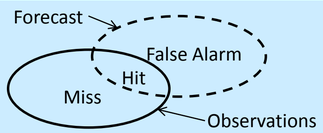
Extreme thresholds by 32-km grid point



Regional thresholds by RFC; upper/lower number is the top 1.0%/0.1% of precipitation events

QPF Performance Measures

Events	Observed (O)	Not observed
Forecast (F)	Hit (H)	False alarm (FA)
Not Forecast	Miss (M)	Correct rejection

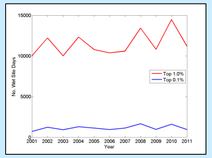


- Probability of detection (Hit rate)
  POD = H/(H+M)
- False alarm ratio
  FAR = FA/(FA+H)
- Critical Success Index (Threat Score)
  CSI = H/(H+M+FA)
- Mean absolute error
  MAE = $\frac{1}{n}\sum_{i=1}^{n} | Fi - O_i |$
- Bias
  Bias = QPF/QPE

National Extreme QPF Performance

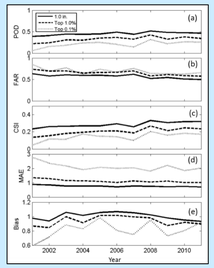
- National performance metrics were calculated by applying regional extreme precipitation thresholds to each RFC, aggregating the results over the calendar year, and then computing the various metrics.

Distribution of extreme events by year



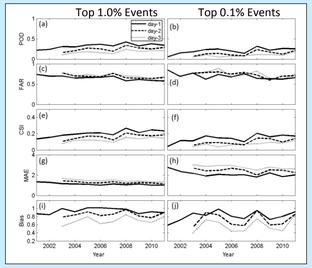
- Top 1.0% events: Minimally 10,000 events per year
- Top 0.1% events: Minimally 700 events per year

By precipitation threshold



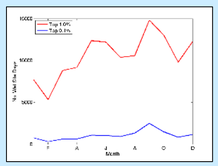
- POD, FAR, and CSI values for extreme events show lower skill than the 1.0 in 24 h^{-1} (GPRA) events.
- POD, FAR, and CSI values have incrementally improved over the 11 year period for all 3 thresholds.
- MAEs have improved (i.e., decreased) over the 11 year period for all 3 thresholds.
- Bias values indicate the extreme events tend to be underforecast.

By lead time

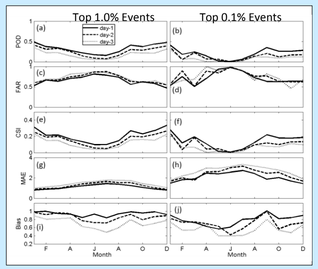


- POD, FAR, and CSI values decrease in skill with longer lead time.
- POD, FAR, and CSI values have improved from 2001 to 2011 for all 3 lead times.
- At times, the day-3 QPF skill exceeds the skill of the day-2 QPFs for the 0.1% events.

By month/season



- Cool season has less events; warm season has more
  - Minimum number of events in February
  - Maximum number of events in September.



- POD, FAR, and CSI values decrease in skill during warm season months (JJA).
- POD, FAR, and CSI values increase in skill during cool season months (DJF).
- At times, the day-3 QPF skill exceeds the skill of the day-2 QPFs.

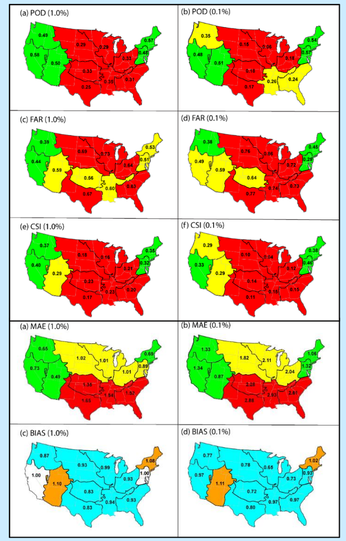
Acknowledgements

This research was supported by NOAA's Hydrometeorological Testbed (HMT) program and the United States Weather Research Program (USWRP).

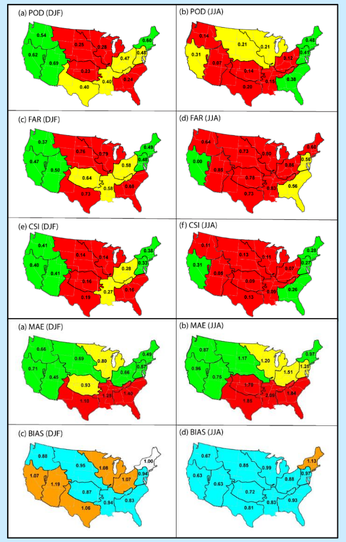
Regional Extreme QPF Performance

- Performance metrics were calculated by applying extreme precipitation thresholds to each RFC region, aggregating the results over the calendar year, and then computing the various metrics for 5 years (2007-2011.)
- Relative regional performance was compared by terciles. RFCs in the top tercile are green, in the middle tercile yellow, and the bottom tercile red.
- Bias is color-coded with blue as underforecast (bias<1), white unbiased (bias = 1), and orange overforecast (bias>1).

By precipitation threshold



By season



For 1.0% and 0.1% events, western and northeastern RFCs (CNRFC, CBRFC, NWRFC, MARFC, NERFC) are more likely to have:

- Higher skill
- Lower error
- Bias = 1 (unbiased) or Bias > 1 (overforecast)

During the cool season (DJF),

- Skill scores are higher
- Western and northeastern RFCs have higher skill
- More extreme events are biased high (overforecast)

During the warm season (JJA),

- Skill scores are lower
- Most extreme events are biased low (underforecast)

Summary

- Extreme precipitation events are objectively defined (i.e., 99th and 99.9th percentile thresholds) of all precipitation events for 2001-2011 by RFC region.
- Multiple verification metrics (POD, FAR, CSI, MAE, bias) are used to analyze extreme QPF performance.
- Extreme QPFs are found to have incrementally improved since 2001.
- Extreme QPF performance tends to be lower with longer lead time and larger precipitation thresholds.
- Extreme QPF performance is lower during JJA and higher during DJF.
- Extreme QPFs in the western and northeastern RFCs perform better relative to the other RFCs regions.