Evaluating the usefulness of spatial forecast verification methods for developing a short-term hydrometeorological prediction system

Barbara Brown, Amanda Anderson,
Randy Bullock, and Eric Gilleland
NCAR, Boulder, CO USA

ECAM/EMS
September 2013

Short-Term Explicit Prediction (STEP) Project

Developing a Quantitative Precipitation Estimation, Nowcasting and Streamflow Prediction System for the Rocky Mountain Front Range

Project Focus

Prediction of localized heavy precipitation and flooding events

Project goals

- Improve prediction of heavy rainfall, flash floods and streamflow in Colorado Front Range
- Develop optimal mix of observational systems and NWP for real-time nowcasting and forecasting of heavy rainfall and streamflow
- End-to-end hydrometeorological nowcast system
- Test end-to-end hydrometeorological system in real-time in 2014

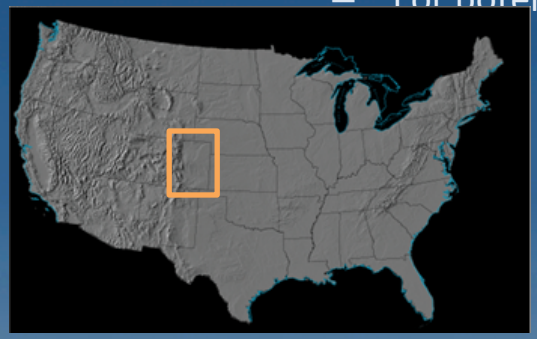
Evaluation System Goals

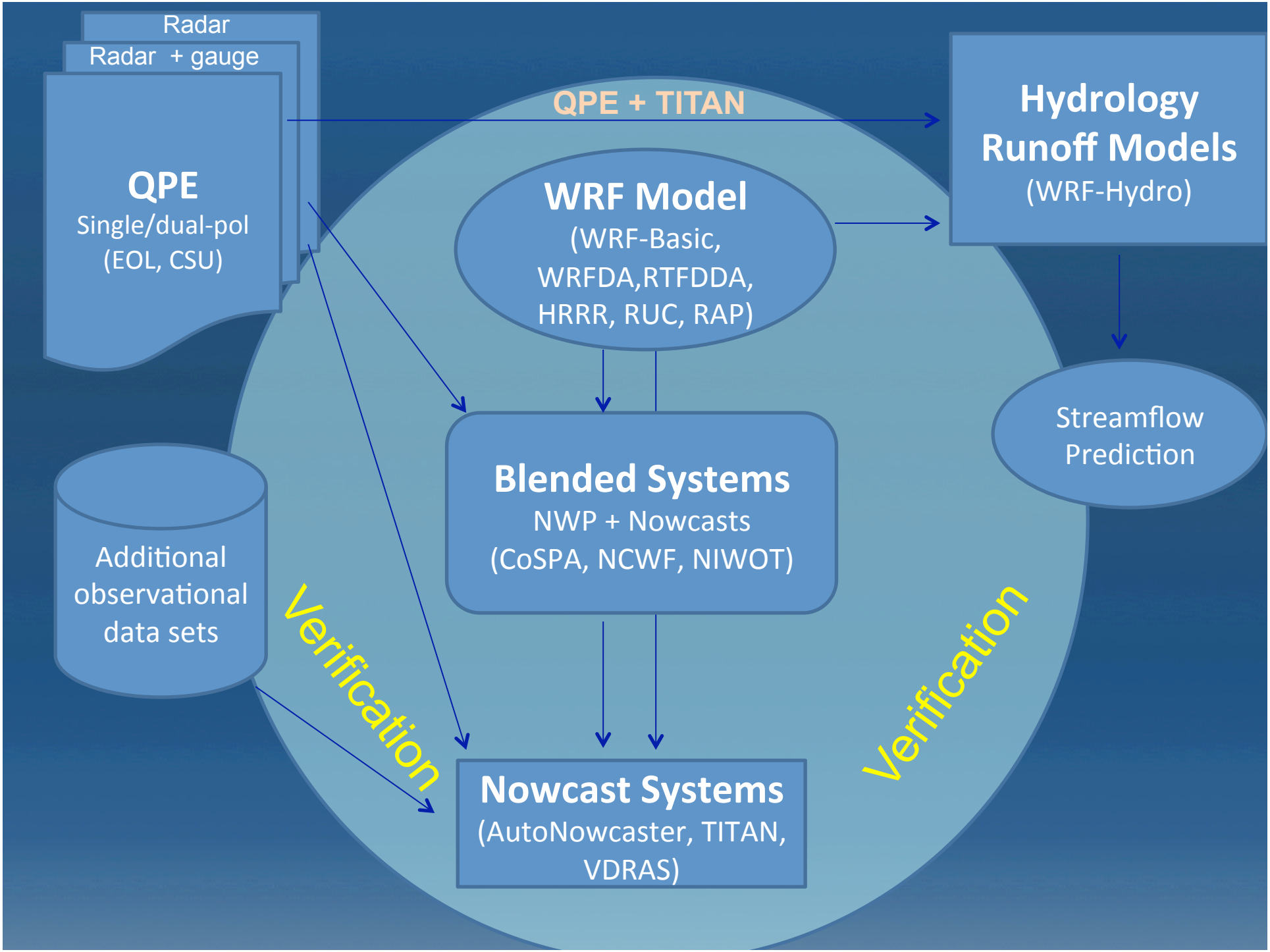
Long-term

- Integrated real-time evaluation system that provides ongoing, meaningful answers to the core questions about forecast quality

Short-term

- Test a variety of methods on case studies and other forecasts of opportunity
- Determine which approaches provide the most useful information for the forecast developers and to answer important questions about performance
  - Of individual components
  - For potential end users




Radar
Radar + gauge
QPE
Single/dual-pol
(EOL, CSU)
QPE + TITAN
Hydrology
Runoff Models
(WRF-Hydro)
WRF Model
(WRF-Basic,
WRFDA,RTFDDA,
HRRR, RUC, RAP)
Streamflow
Prediction
Blended Systems
NWP + Nowcasts
(CoSPA, NCWF, NIWOT)
Additional
observational
data sets
Verification
Verification
Nowcast Systems
(AutoNowcaster, TITAN,
VDRAS)

Targeted approaches

Diagnostic
Not focused on single measure summaries

Why? Provide information that can answer the questions of interest

Spatial
Take into account spatial attributes of phenomena; separate skill according to scale

Why? Provide information about performance relative to particular spatial attributes (coverage, intensity distribution, etc.)

Spatial-temporal
Consider temporal as well as spatial errors

Why? Better determine sources of errors

Conditional
Consider differences in performance as function of synoptic or other variables

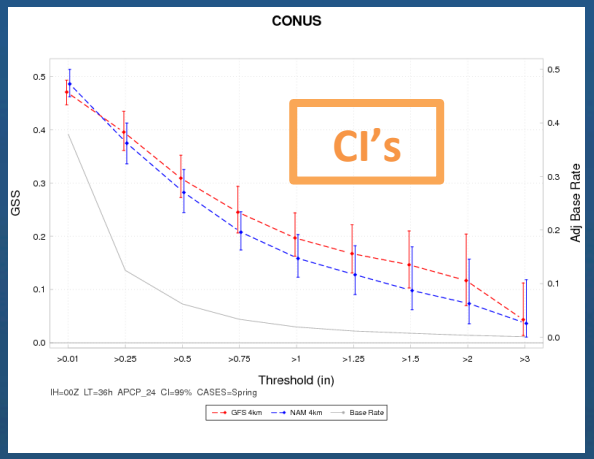
Why? Document when system performs best/worst

Overall measures of (comparative) performance
Measures that document significance of improvements over other systems

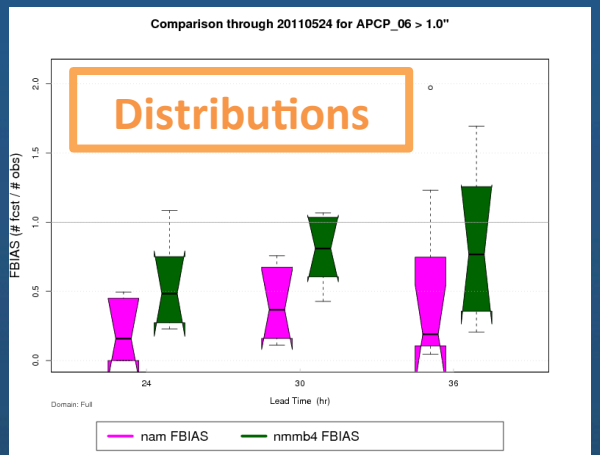
Why? Document improvements; marketing

Example approaches

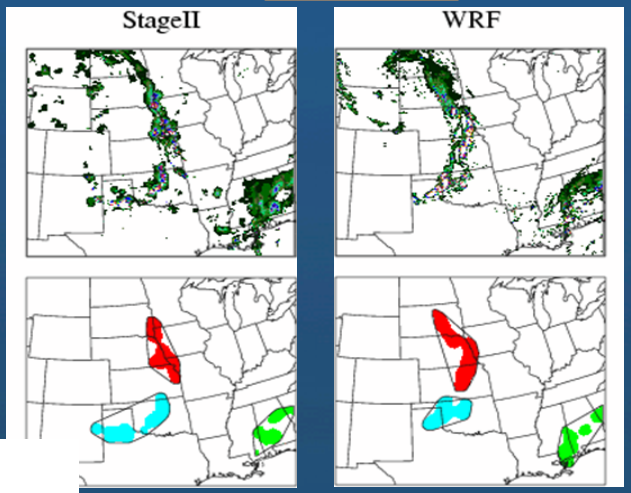
MODE

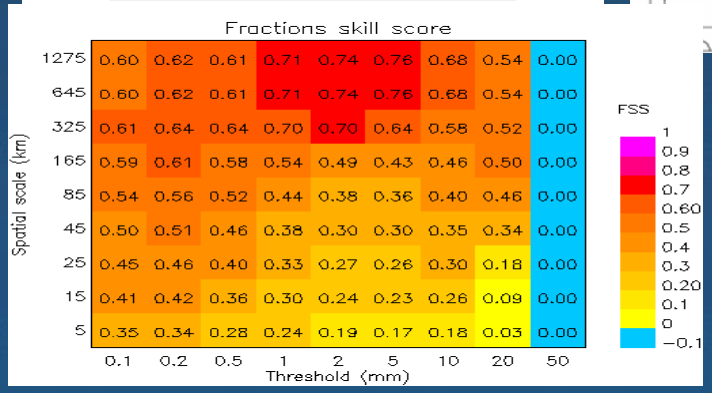


CI's



Distributions





good performance

Quilt plots

poor performance

Perf Diagram

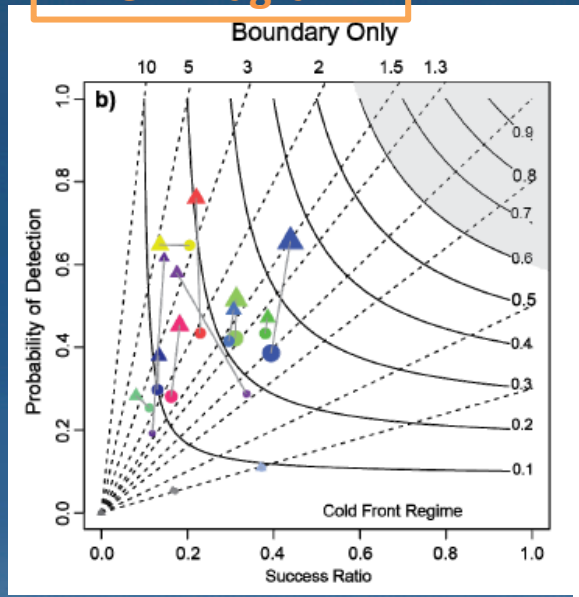
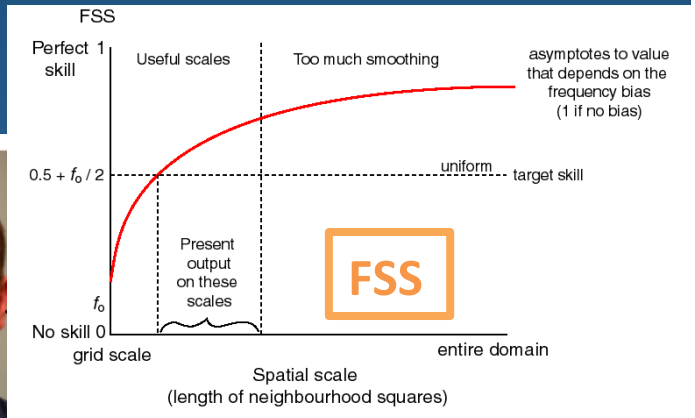
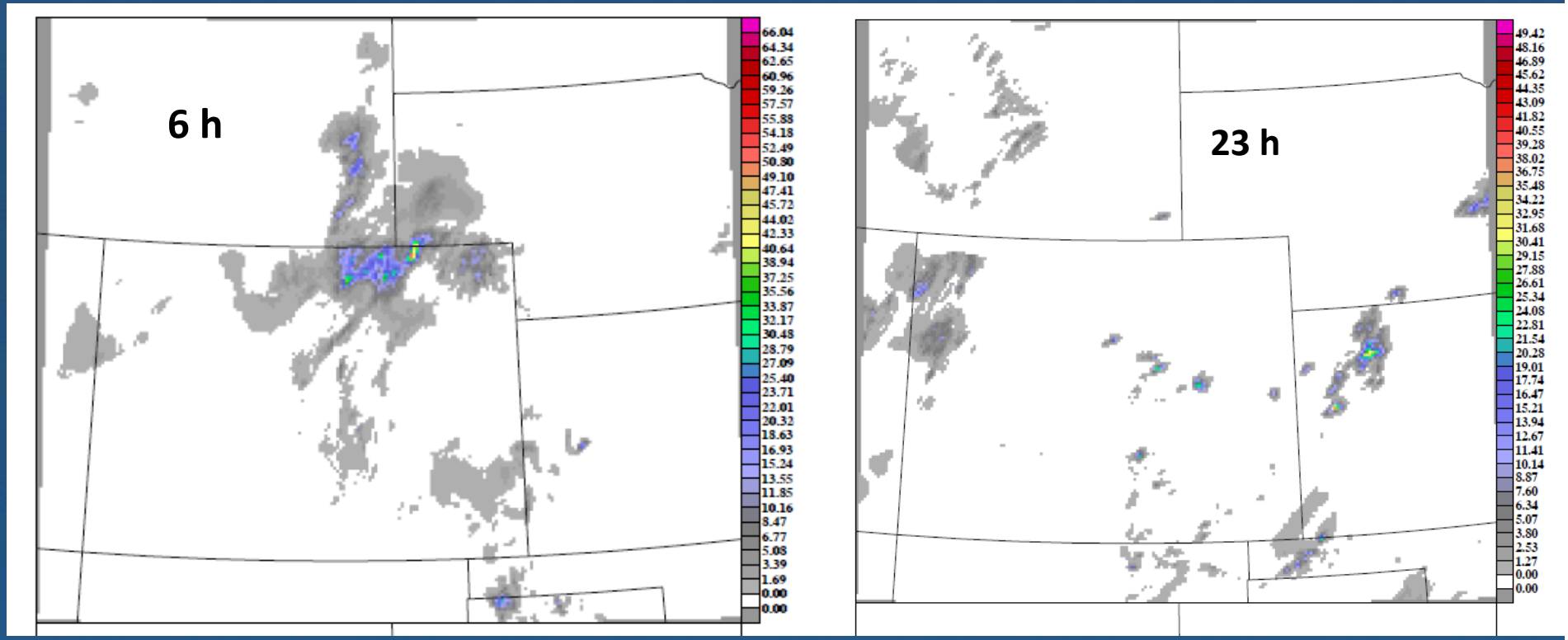


Image Warping



FSS

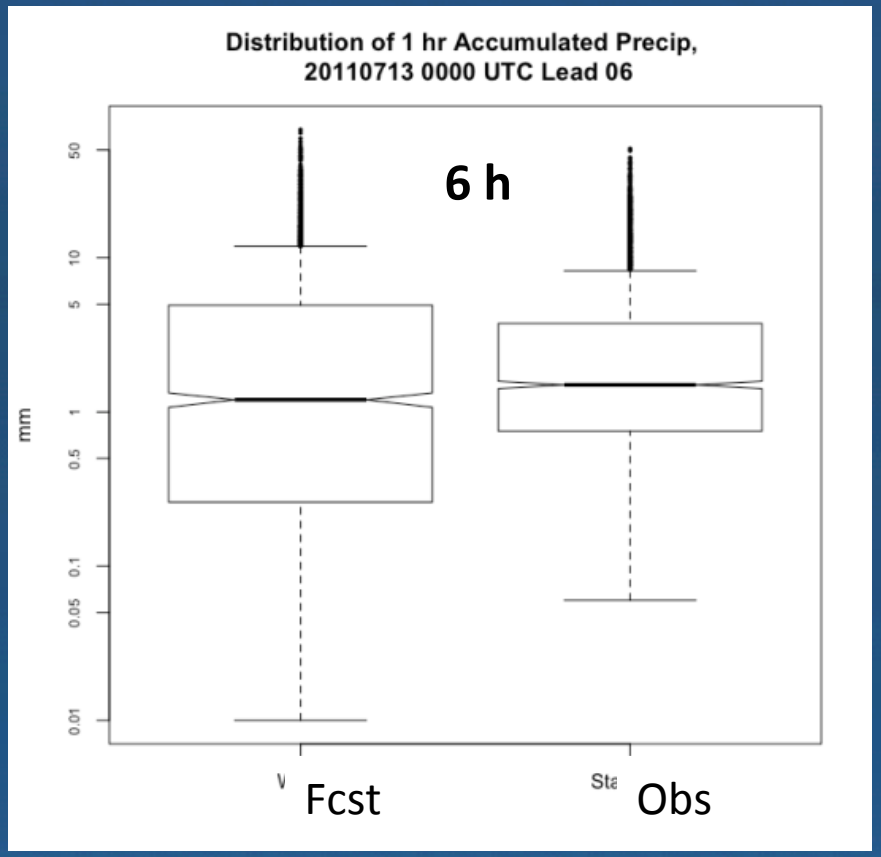
July 13-14 Case: 4-Mile Canyon Flood
Stage IV observations

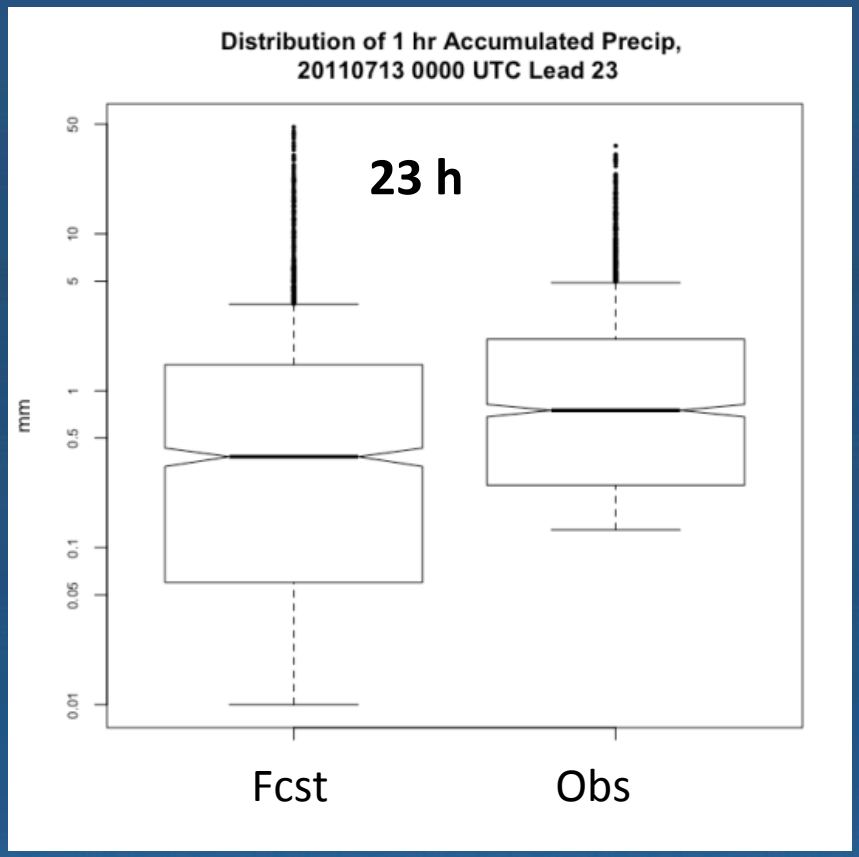


Focus: Baseline WRF model runs (no assimilation)

Stage IV (Radar + Gauge) 4-km gridded fields used for verification

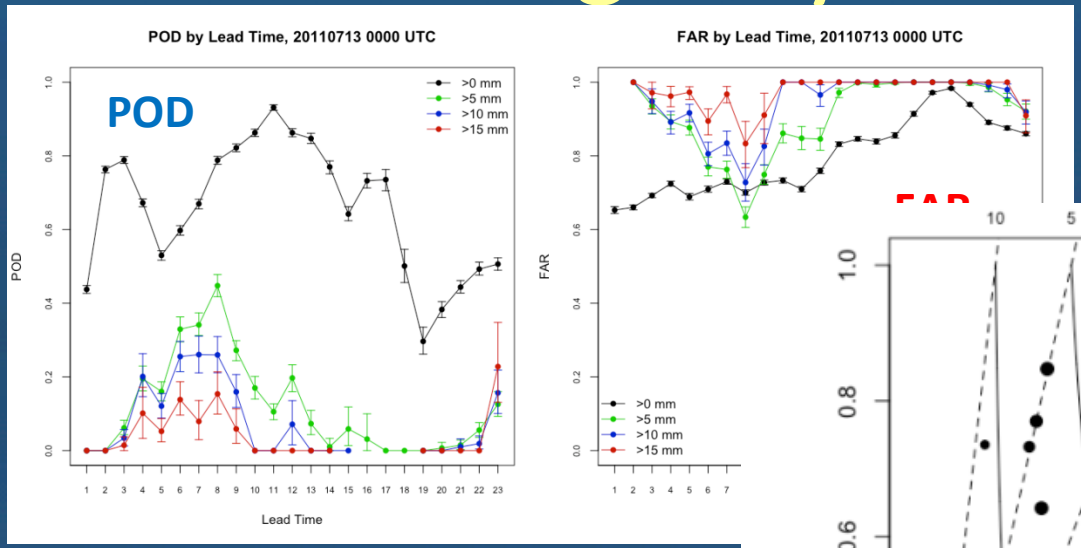
Precipitation distributions

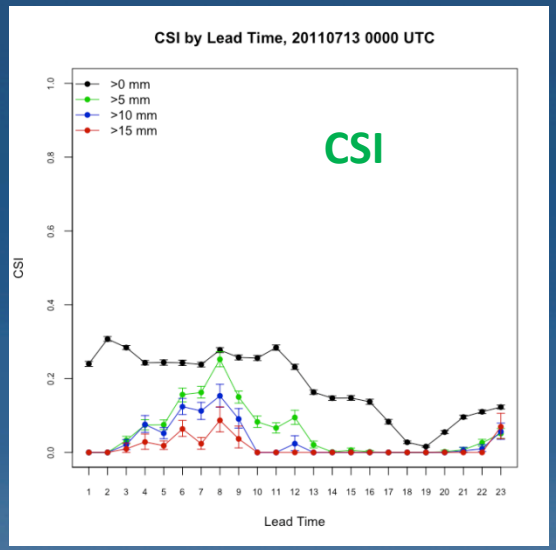


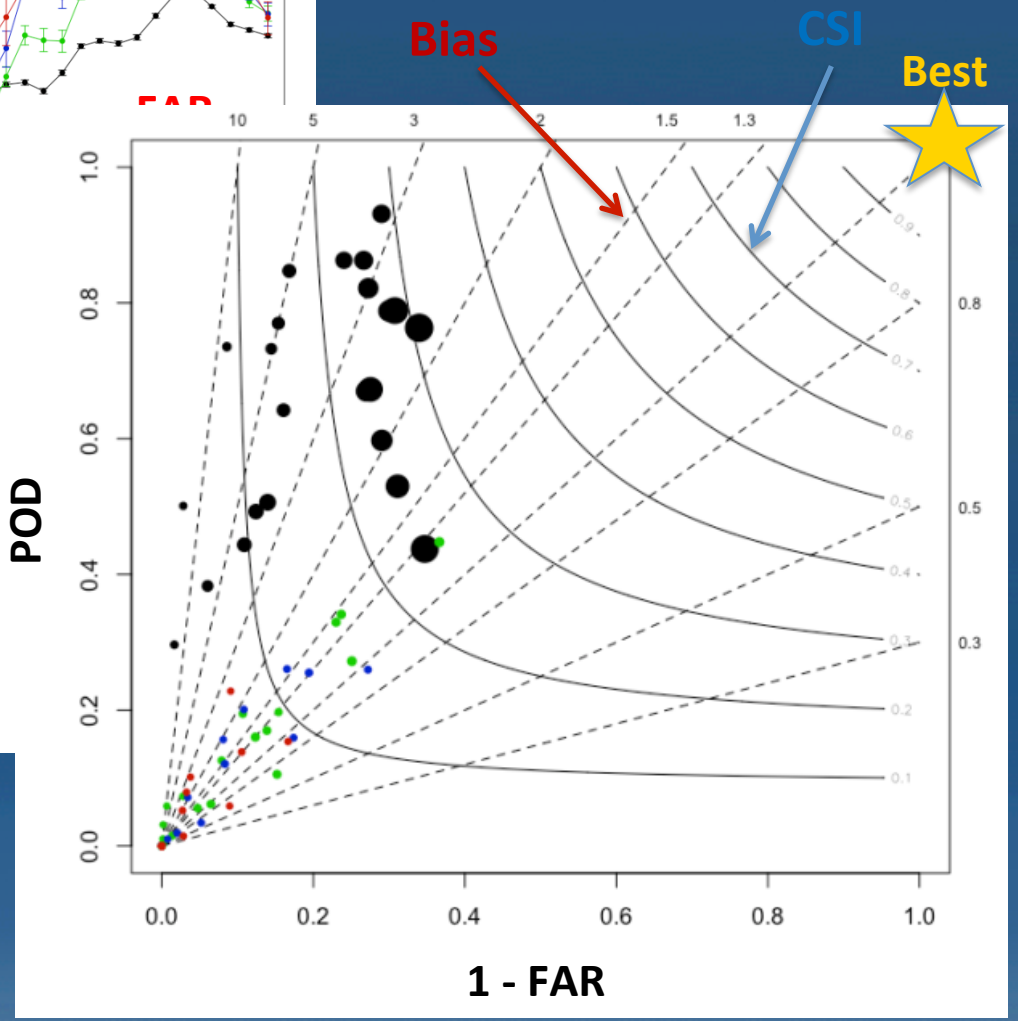


Distributions are significantly different from each other
Overall, WRF overforecasted small precipitation amounts and underforecasted large amounts

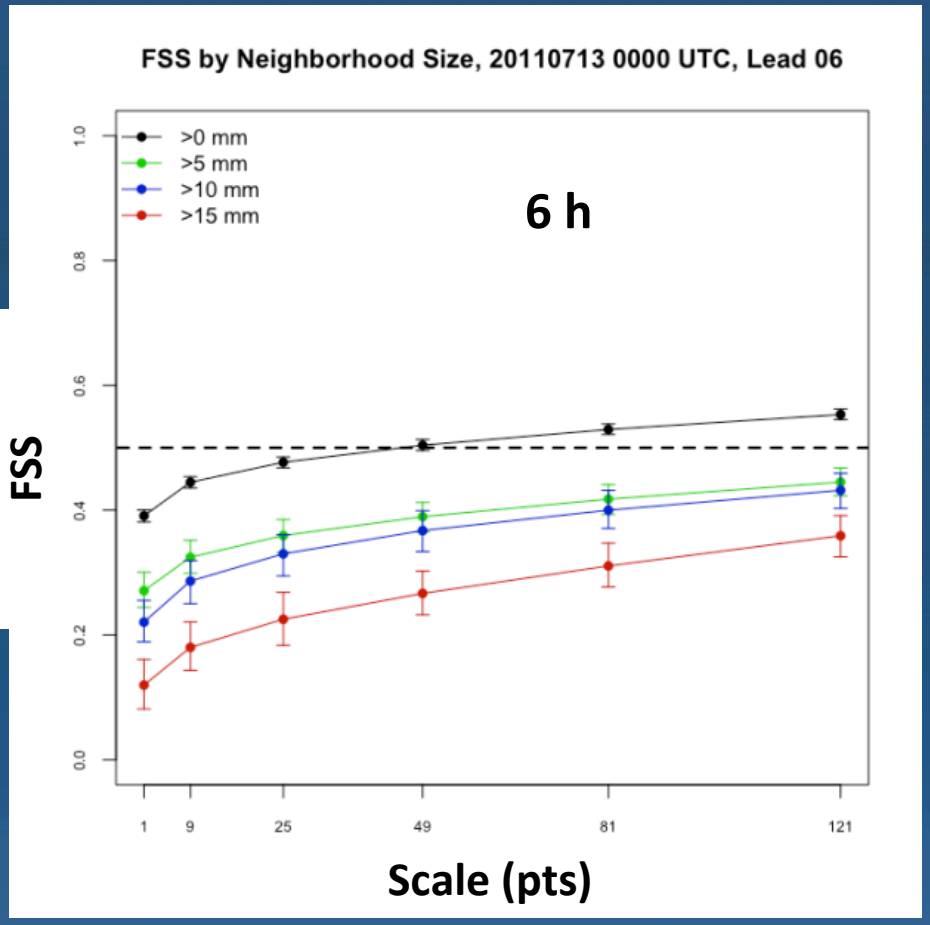
Contingency Table statistics

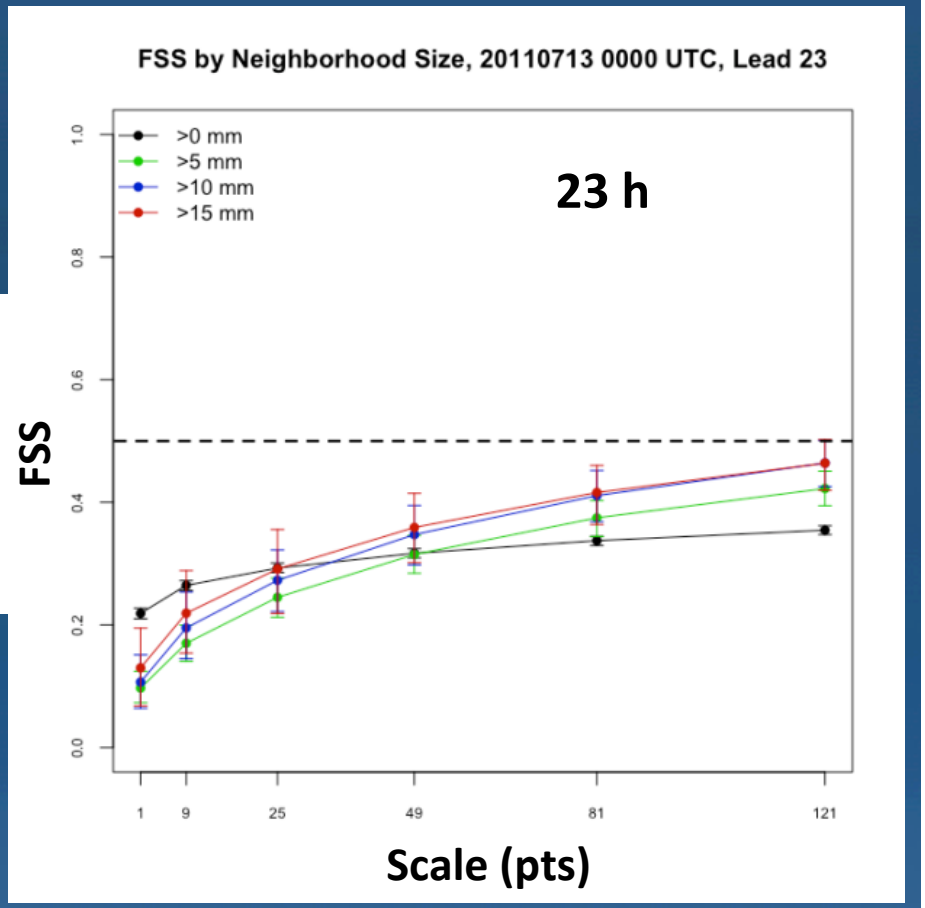






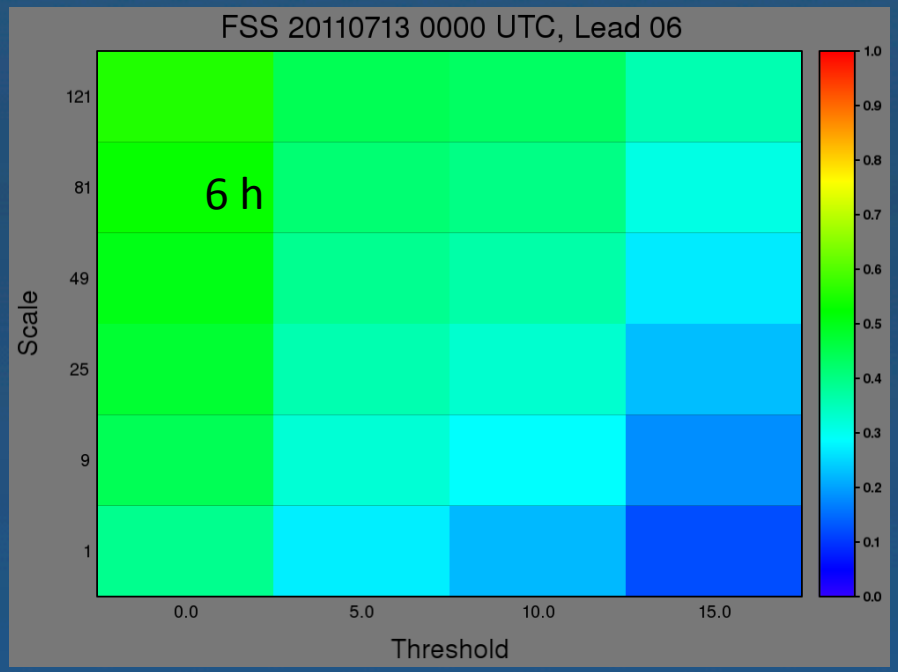
Fraction Skill Score (FSS): Skill as a function of spatial scale



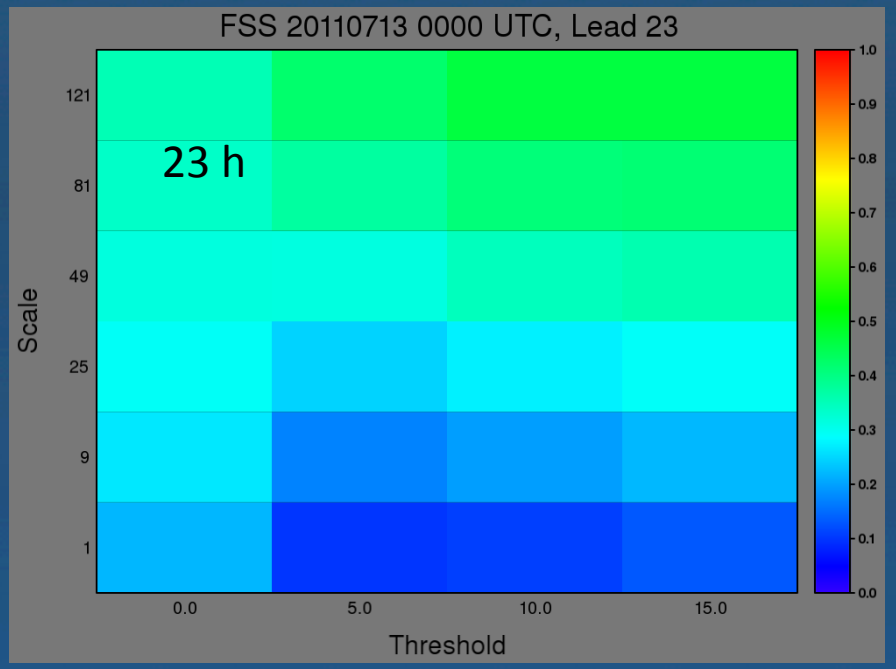


FSS indicates little usable skill except for lowest threshold and largest scale

Fractions Skill Score (FSS): Quilt Plots

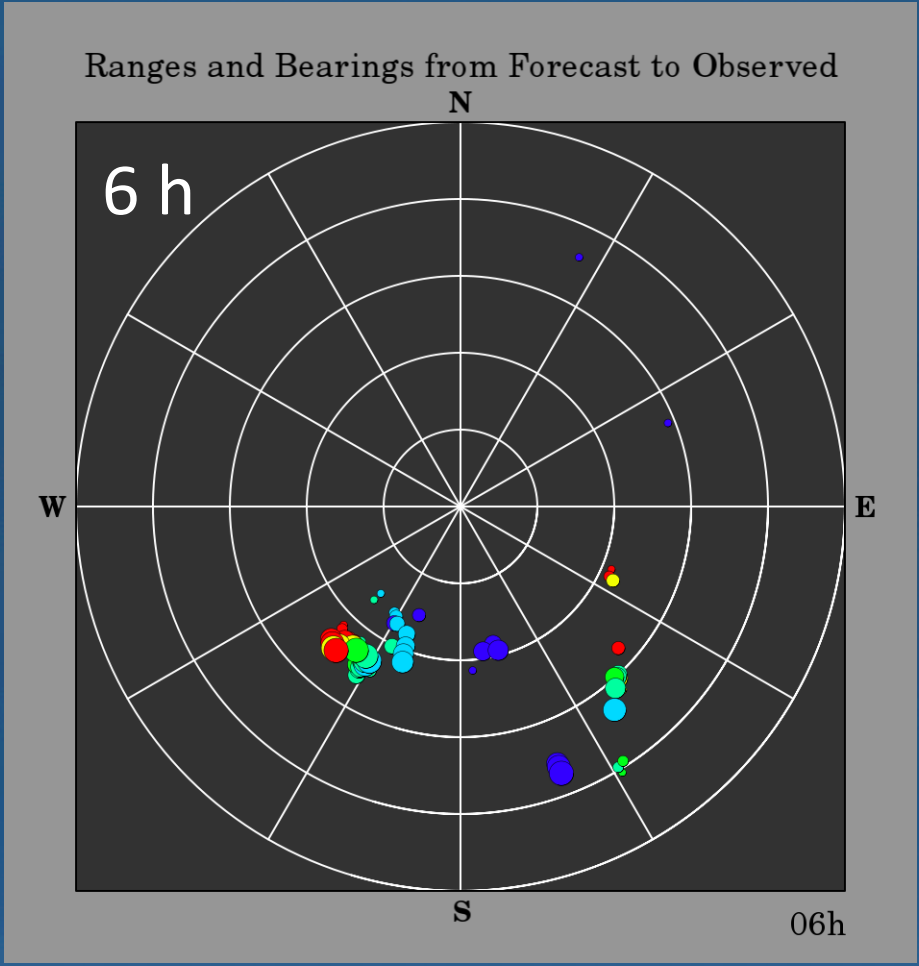


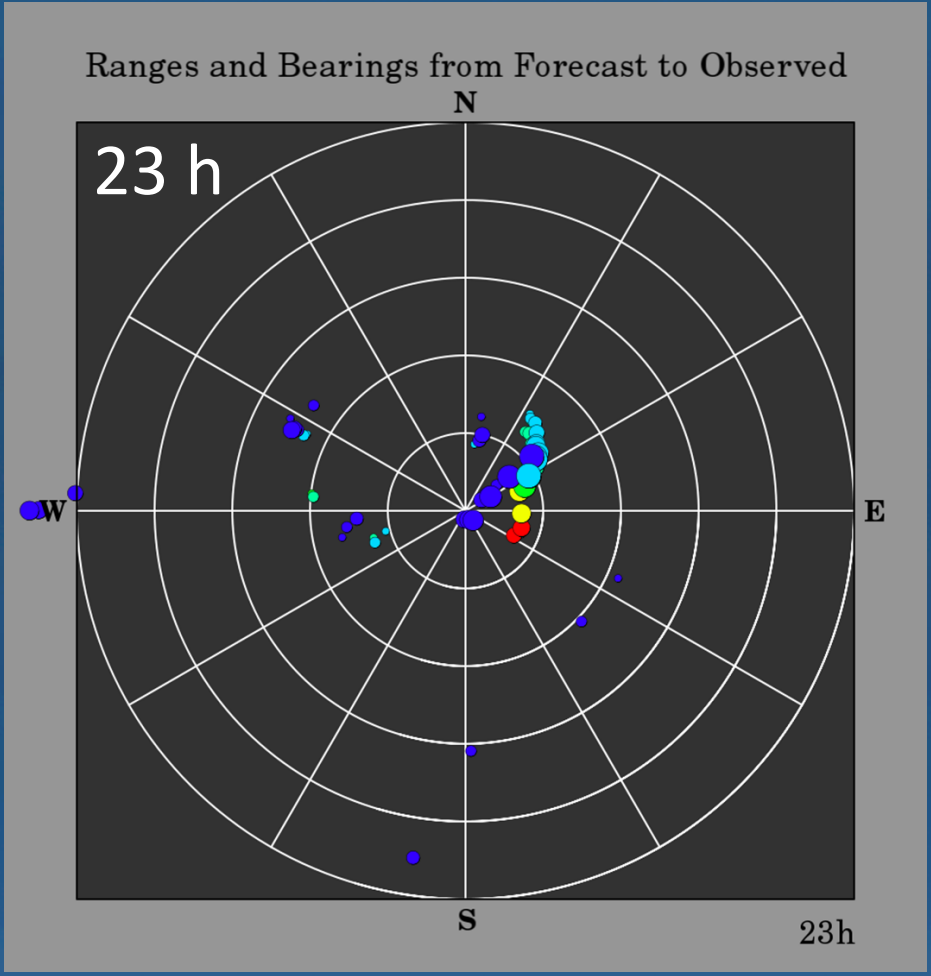
Skill best for lowest thresholds



Skill best for higher thresholds

MODE[1] Centroid differences

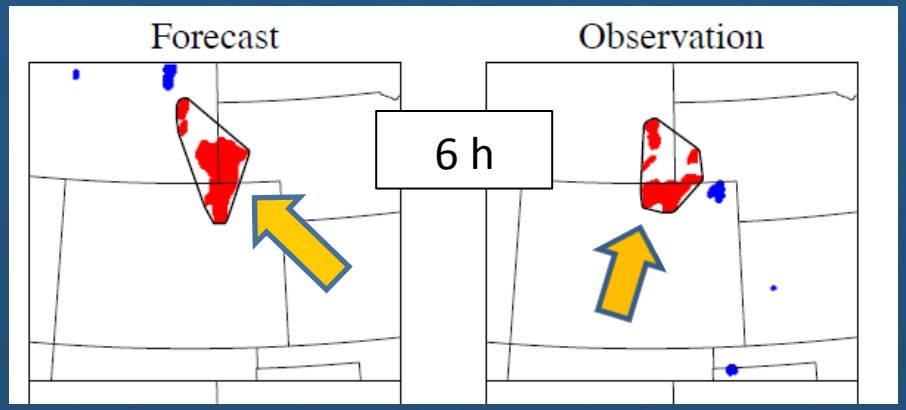




Very different patterns of centroid errors at 6 h and 23 h

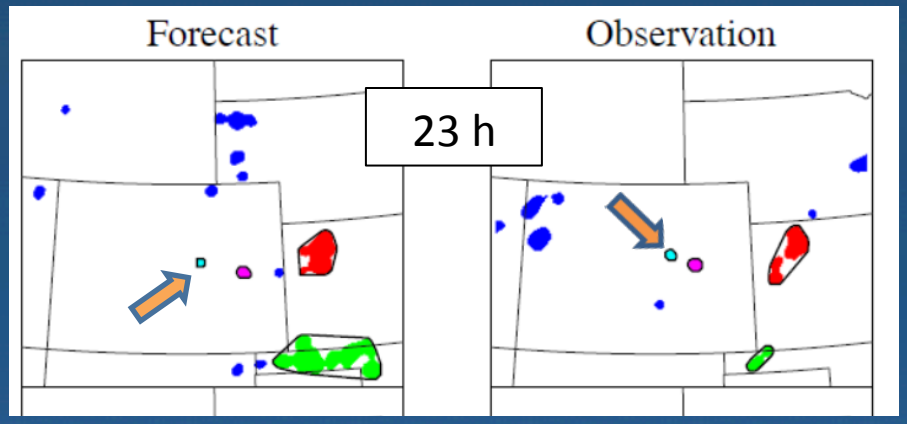
[1]MODE = Method for Object-based Diagnostic Evaluation

Example MODE results



Area ratio	1.5
Intersection area / Obs area	0.51
Centroid distance	31 km
50th percentile intensity ratio	0.98
90th percentile intensity ratio	1.10

Convolution radius = 2 grid boxes; Threshold = 4 mm h^{-1}



Area ratio	0.70
Intersection area / Obs area	0.06
Centroid distance	19 km
50th percentile intensity ratio	1.14
90th percentile intensity ratio	0.78

Results show (1) forecasts have some skill in capturing these events and (2) in which aspects the forecasts need improvement

Ex: 90th percentile of precipitation; storm placement/timing

Summary and future work

Each approach provides different kinds of information

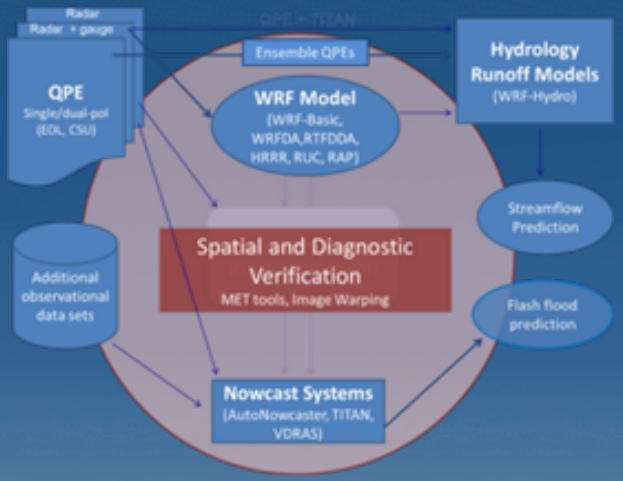
Application of different approaches shows appeal and benefits of *diagnostic methods*

Examples:

- Small amounts of precipitation forecast too frequently; large amounts too infrequently
- Forecasts are poor according to traditional stats except when there is a lot of precip (i.e., coverage)
- Little skill except for when extensive smoothing applied
- Generally, forecast objects are offset from observed, especially at 6 h
- But, overall, forecasts may have some skill… according to MODE results

Next steps:

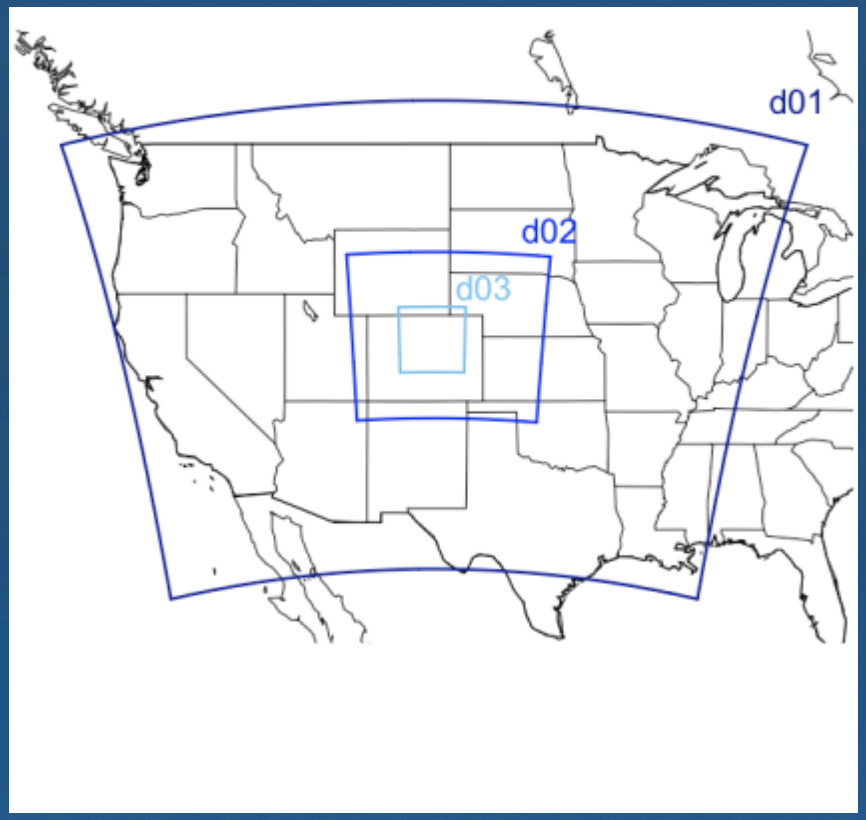
- *Further interactions with forecast developers regarding utility of different approaches*
- *Comparison of baseline and "enhanced" approaches*
- *Evaluation of additional aspects of the overall system (Nowcasts, Streamflow)*
- *Further consideration of approaches from perspective of system developers*
- *Development of near-real-time system for 2014*



END

EXTRA SLIDES

Domains


d01
d02
d03


Alliance
Scottsbluff
Rawlins
Cheyenne
Sidney
Red Mountain Open Space
Fort Collins
Sterling
Loveland
Greeley
Boulder
Westminster
Glenwood Springs
Lakewood
Denver
Centennial
Colorado
Colorado Springs
Canon City
Pueblo
La Junta
Lamar
Alamosa
Durango
Google
Map data ©2013 Goog
42
41
40
39
38
-108
-106
-104
-102
lon