Evaluation of Ensemble Precipitation Forecasts at Environmental Modeling Center (EMC) of NCEP

Yan Luo and Yuejian Zhu

NOAA/NWS/NCEP/EMC, 5830 University Research Court, College Park, MD 20740

Introduction

A verification framework for evaluating the precipitation forecast performance of a global ensemble system was developed at NECP/EMC. The verification framework was initially applied to the case of NCEP global ensemble forecast system (GEFS) and over CONUS. The 24-hour accumulated precipitation forecasts out to 15 days are verified against the Climatology-Calibrated Precipitation Analysis (CCPA). CCPA is a high resolution data set of precipitation analysis generated by statistically adjusting Stage IV toward CPC gauge-based analysis. A variety of performance metrics have been used to evaluate the accuracy and probabilistic skill of precipitation forecast. The framework produces not only deterministic and categorical scores including traditional error statistics and contingency table statistics for the ensemble mean, but also probabilistic skill scores such as Brier Skill Score and reliability diagram. The forecast skill is calculated relative to a reference from a 10-year CCPA daily climatology data. The framework has been also expanded to evaluate precipitation forecasts of a global ensemble system from several other world forecast centers. In this presentation, the framework is demonstrated from its application to the routine evaluation of contemporary operational medium-range ensemble forecasts from these centers since 2013 and the assessment of changes in ensemble performance from every GEFS upgrade. Based on available observational and forecast archives, an inter-comparison study of these centers' forecast performance is conducted using the framework for a period of 2013-present. The verification methods included in the framework are found to give useful and detailed information about the forecasts performance.

Motivations

Provide a comprehensive overview of operational ensemble model performance over CONUS

- ❑Help forecasters and hydrologists understand the forecast accuracy, skill, and reliability of current operational global ensemble precipitation products over CONUS.
- ❑Exploit knowledge of model biases, capabilities, and limitations of an ensemble system that in turn has great potential to improve forecasts

Verification Method

- ❑Daily (24-h) QPF/PQPF
- ❑00 cycle forecast only
- ❑Verified every 24 hours out to 15 days (Day1=12-36h, Day2=36-60h, etc)
- ❑Verify against CCPA
- ❑Daily Climatological Reference: 10 year (2002-2012) CCPA
- ❑Resolution: 1*1 degree gridded
- ❑Domain: CONUS
- ❑Performance metrics
  - ➢RMSE/SPRD, Mean ERR/MAE
  - ➢ETS,TSS, Bias
  - ➢CRPS, BS, BSS, Reliability Diagram/Histogram

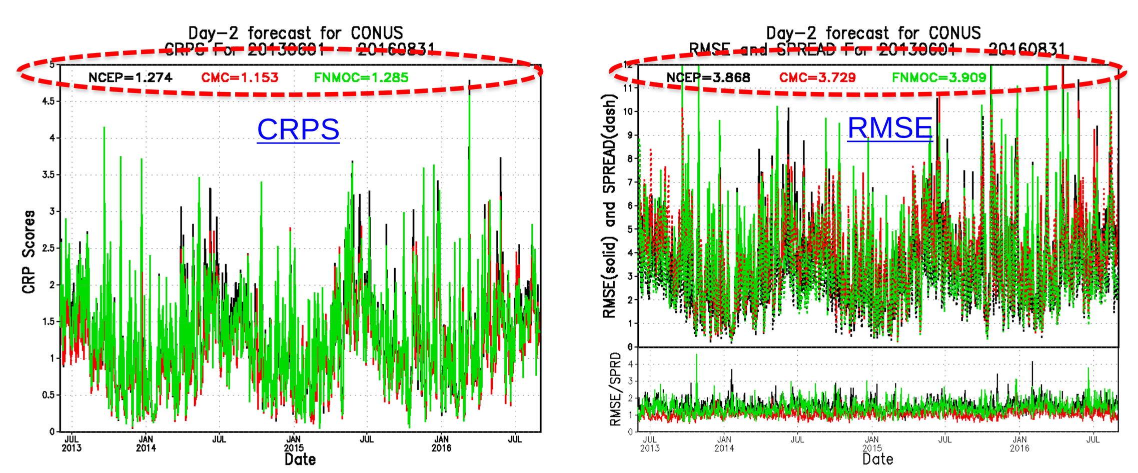
CCPA Dataset

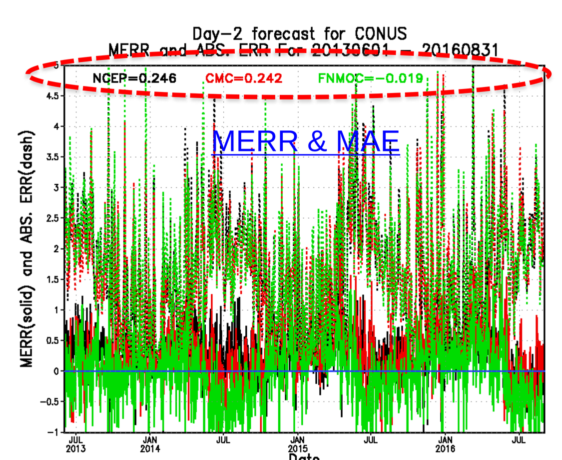
- ❑Climatology-Calibrated Precipitation Analysis (CCPA)
  - ➢A dataset of precipitation analysis, over CONUS at 6h, ~4km resolution
  - ➢Statistical adjustment of Stage IV data toward CPC analysis
  - ➢Simple linear regression at 0.125 degree and 24h accumulation
  - ➢Keep the fine scale structures of Stage IV
  - ➢Closer to CPC Unified Precipitation Analysis, in the sense of climatology
- ❑Application: Provide a proxy of truth for precipitation forecast calibration and downscaling
- ❑Developed and distributed by NCEP/EMC for operation
- ❑First operational implementation on July 13, 2010
- ❑CCPA upgrades:
  - ➢CCPA v2: Add 3-hourly precipitation analysis in July 2011
  - ➢CCPA v3: Update regression coefficients using a longer period (2002-2014) training data in March 2016
- ❑Product period: 2002 - present
- ❑Product grids:
  - ➢HRAP (primary)
  - ➢2.5km & 5km NDGD, 0.125, 0.5 and 1.0 degree resolutions (byproducts)
- ❑CCPA websites:
  - ➢Introduction: http://journals.ametsoc.org/doi/abs/10.1175/JHM-D-11-0140.1
  - ➢Image: http://www.emc.ncep.noaa.gov/gmb/yluo/ccpa/ccpa.php

Verification Activities

- ❑Routine inter-comparison of global ensemble forecasts from world forecast centers (focus of this presentation)
  - ➢NCEP, CMC, and FNMOC individual ensembles
    - NCEP GEFS configuration (http://www.emc.ncep.noaa.gov/GEFS/gmb/Doc-5.1.1_EPS_activities_NCEP_GEFS.pdf)
    - CMC GEPS configuration (http://collaboration.cmc.ec.gc.ca/cmc/CMOI/product_guide/product-pages/image_ens_prog_ensemble-forecasts_gen_e.html)
    - FNMOC ensemble configuration (https://www.fnmoc.navy.mil/wxmap_cgi/index.html#ensemble)
  - ➢Started from summer 2013
  - ➢Update by end of each season
- ❑GEFS upgrade
  - ➢Before implementation, compare operational and experimental/parallel run results
- ❑Evaluation of precipitation post-processing techniques (precipitation calibration)
  - ➢Raw vs. Bias-corrected
- ❑Link to website:
  - ➢http://www.emc.ncep.noaa.gov/gmb/yluo/nuopc_precip.html

Three comparison: NCEP vs. CMC vs. FNMOC – CRPS, RMSE & SPRD, MERR & MAE (3yrs)



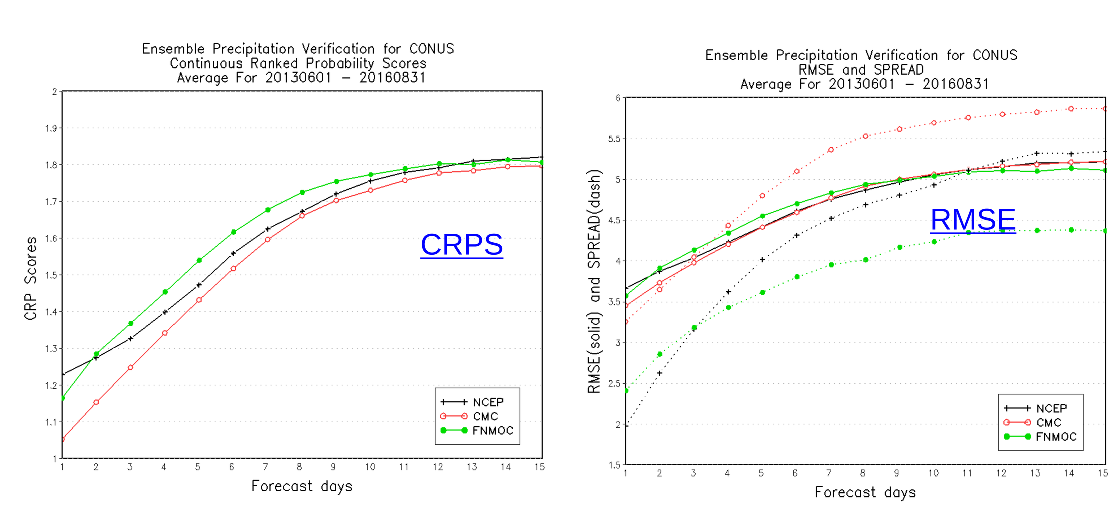


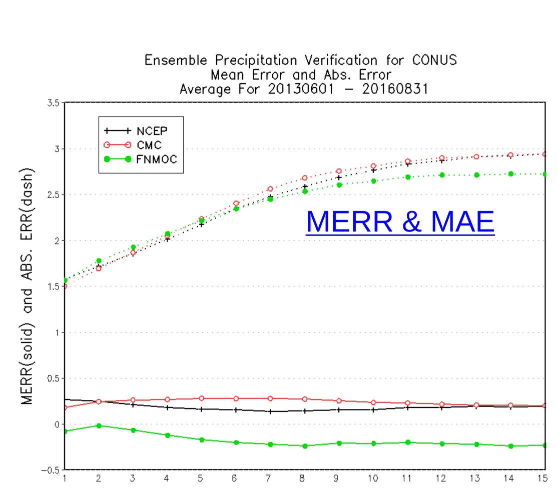
Average numbers are on the top of each plot.

3-years time series for precipitation accumulated on f36-f60

Distinct seasonal variability in performance

- Large errors in warm seasons, and small errors in cold seasons
- Dry biases in warm seasons, and wet biases in cold seasons





CRPS

- CMC outperforms the other two at all lead times
- NCEP has less errors than FNMOC

RMSE/SPRD

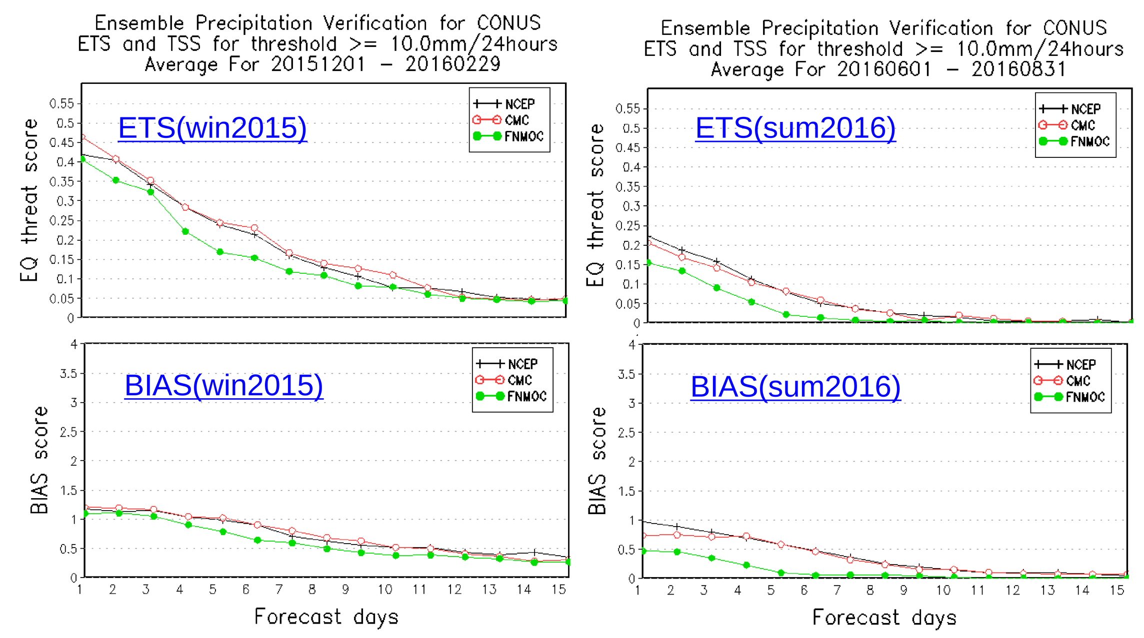
- CMC has the best spread-error relationship
- Both NCEP and FNMOC have far not enough spread and are severely under dispersion; FNMOC has the slowest growth rate

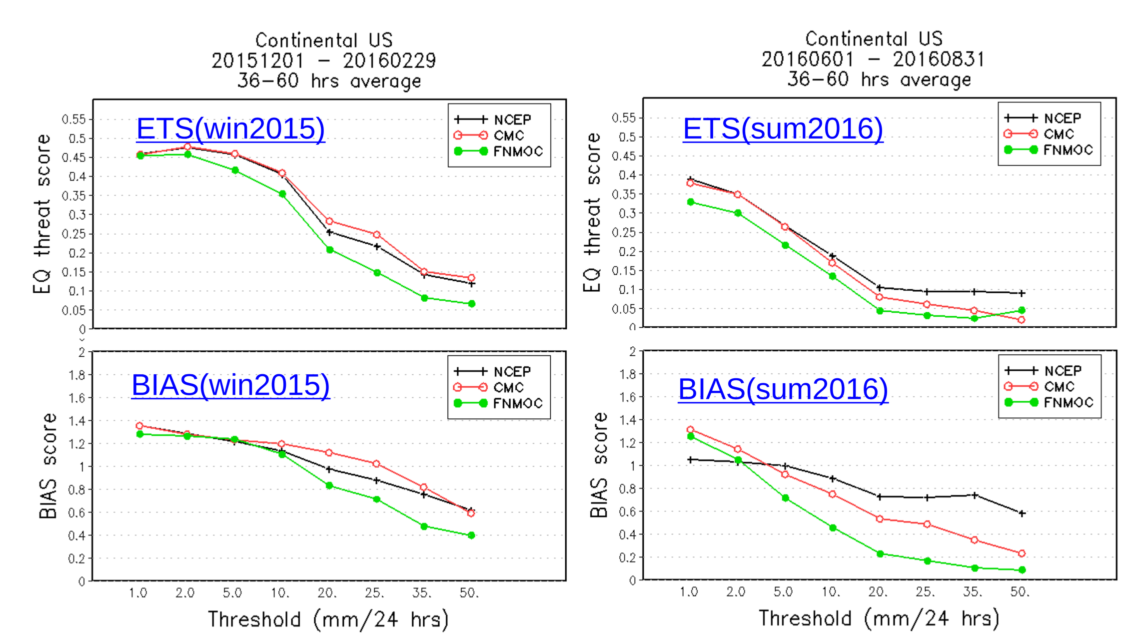
MERR

- Both NCEP and CMC suffer total wet biases
- FNMOC suffers total dry biases

Three comparison: NCEP vs. CMC vs. FNMOC – ETS and Bias

For ensemble mean deterministic verification





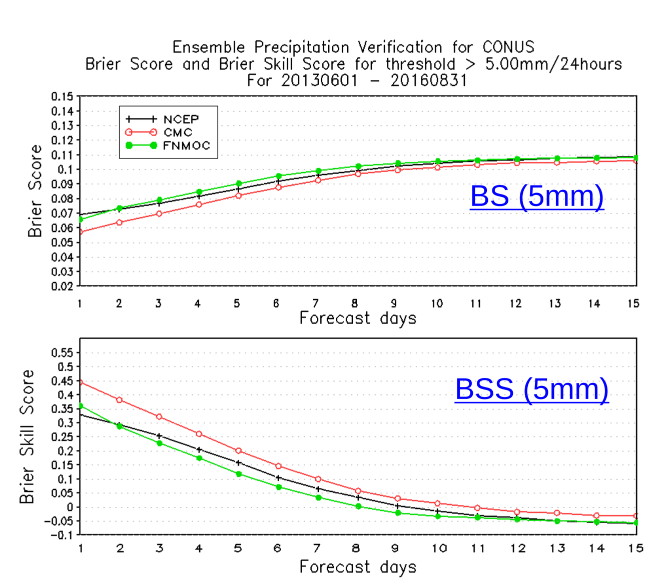
ETS

- High skill in winter, and low skill in summer
- NCEP and CMC are comparable, and more skillful than FNMOC

BIAS

- All have slight over-prediction at low thresholds and under-prediction at high thresholds
- NCEP relatively close to 1.0 at all thresholds, has the best bias scores

Three comparison: NCEP vs. CMC vs. FNMOC – BS, BSS and Reliability Diagram (3yrs)



For probabilistic verification

BS & BSS

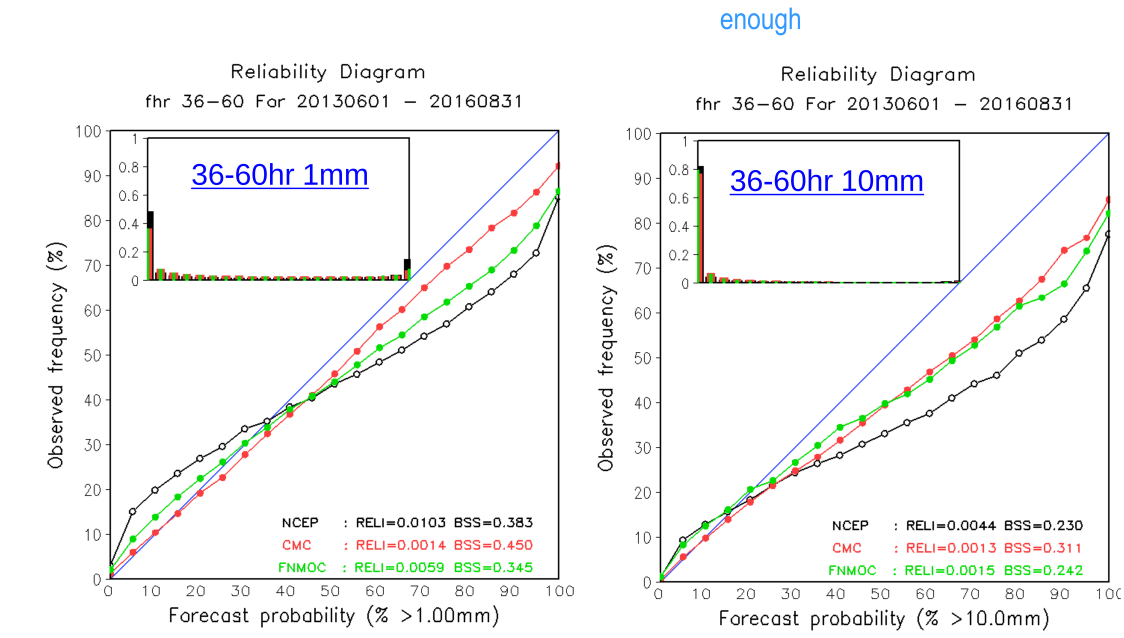
- CMC has the best BS & BSS scores, outperforms the others
- NCEP is relatively better than FNMOC

Reliability Diagram

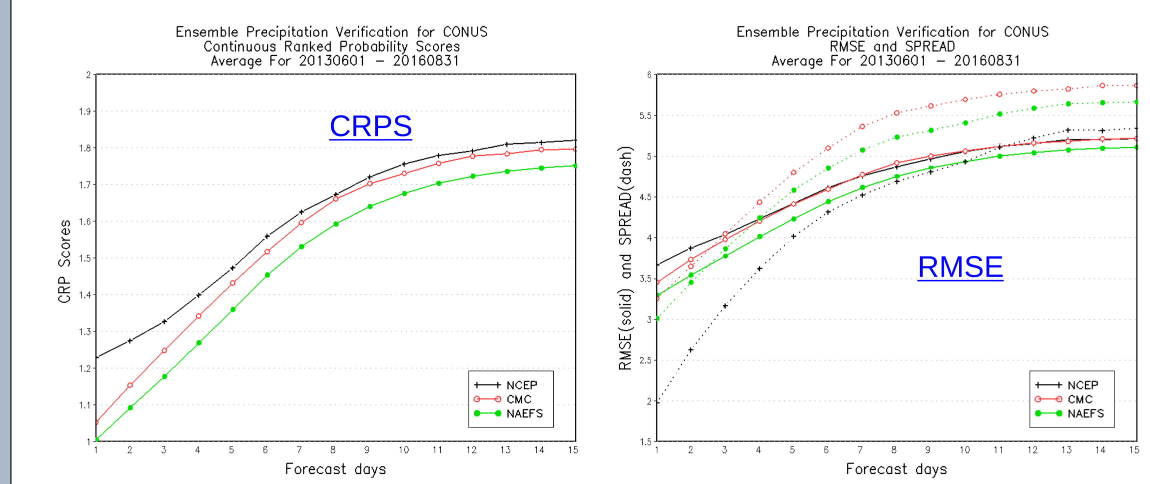
- Overconfidence in general, especially at large threshold
- CMC is more reliable
- NCEP has too little (narrow) spread as the slope of reliability curve much less than 1

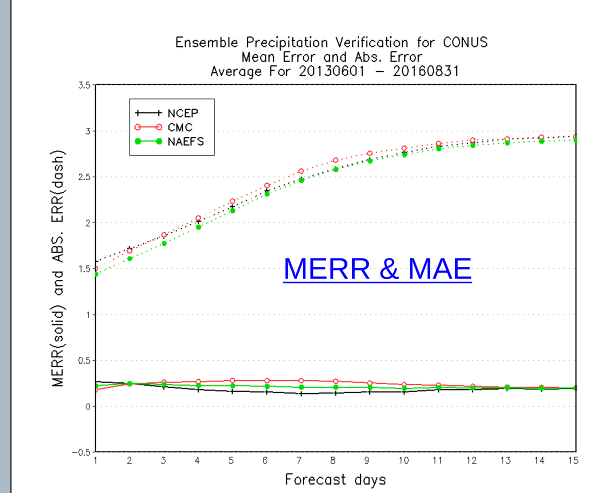
Histogram

- NCEP is sharper
- CMC and FNMOC not sharp enough



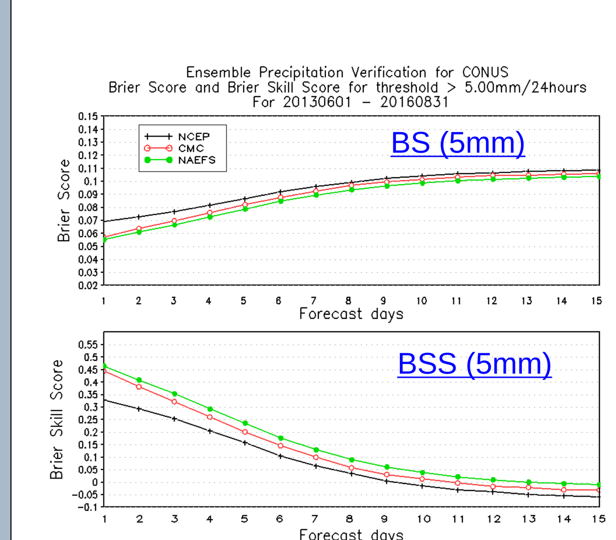
Three comparison: NCEP vs. CMC vs. NAEFS (Combination of NCEP and CMC) – Any benefits from NAEFS?

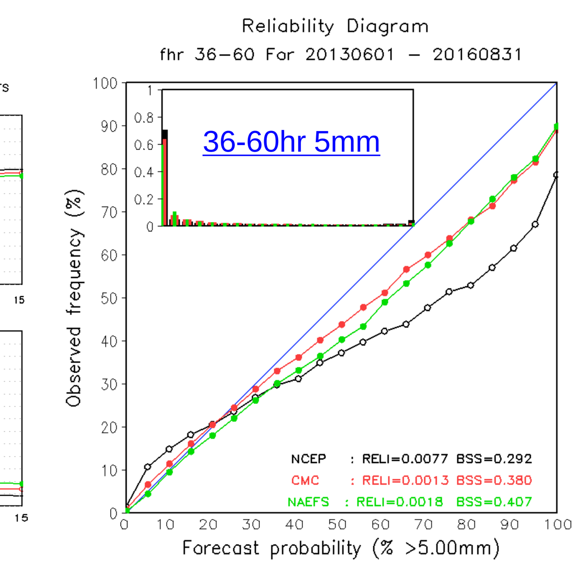




NAEFS

- Improve over individual ensembles
- Reduced errors with smallest CRPS, RMSE, MAE, and BS
- Improved skill with highest BSS and ETS (not shown)
- More reliable on low forecast probability
- Very close to CMC on high forecast probability





Summary

- ❑Routine verification against CCPA gridded analysis for CONUS at one degree resolution
- ❑Individual ensembles and combined are compared
- ❑CMC ensemble has better spread-error relationship. CMC ensemble has better CRPS and Brier scores; NCEP ensemble is sharper
- ❑Combined NCEP and CMC's ensemble (NAEFS) shows a statistical improvement, and is benefit to individual ensembles

Future Plan

- ❑Increase resolution of the verification framework to 0.5° x 0.5°, as more and more half-degree operational forecast products are available now or near future
- ❑Include verification of ECMWF ensemble
- ❑Add side-by-side map comparison for near real-time forecasts
- ❑Enhance verification capabilities with new verification metrics and significant tests
- ❑Continuously contribute to improving ensemble forecasts through verification using this framework; many ongoing efforts on various ensemble techniques are being evaluated
  - ➢Model improvement
  - ➢Increase ensemble size
  - ➢Increase resolution
  - ➢Initial perturbations
  - ➢(Physically based) stochastic parameterizations
  - ➢Multi-model ensembles
  - ➢Statistical post-processing