# Verification of Ensemble River Forecasts at Middle Atlantic River Forecast Center


**Andrew Philpott, Patricia Wnek, and Peter Ahnert**
NOAA/National Weather Service
Middle Atlantic River Forecast Center, State College, PA
Andrew.Philpott@noaa.gov

**James D. Brown**
NOAA/National Weather Service
Office of Hydrologic Development, Silver Spring, MD
and University Corporation for Atmospheric Research



## Abstract

Short Range Ensemble Forecast (SREF), Global Ensemble Forecast System (GEFS), and North American Ensemble Forecast System (NAEFS) precipitation and temperature simulations were used to generate ensemble river forecasts. These were studied using the Ensemble Verification System (EVS) provided by the Office of Hydrologic Development. Box plots of ensemble error, mean error of the ensemble mean, reliability diagrams, and other tools in EVS provided some verification results, particularly for the SREF ensembles for which 4200 forecasts over 3 years were available. This study confirms that these Meteorological Model-based Ensemble Forecasting System (MMEFS) forecasts provided online by the National Weather Service (http://www.erh.noaa.gov/mmefs/) provide useful information about 3-7 day flood risk, in spite of some bias and underspreading.

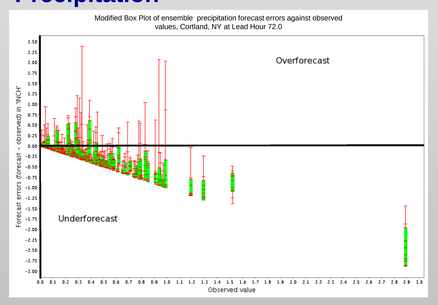

## Forecasts Available:

SREF 12/1/2008 – 11/30/2011 (3 years)
NAEFS and GEFS 10/1/2010 – 11/30/2011 (1 year)

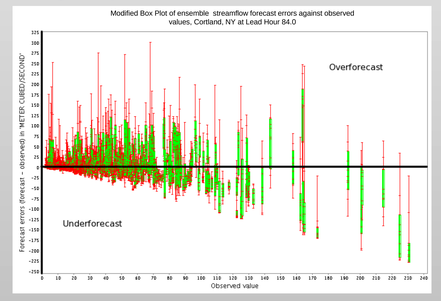
## Box Plots showing conditional bias upon the observed value

Underforecasting bias on larger observed events, both in precipitation inputs and resulting flow ensembles.

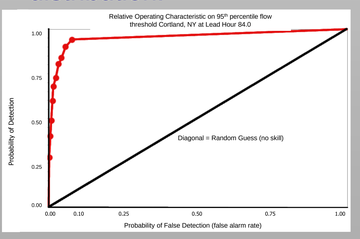
### Precipitation

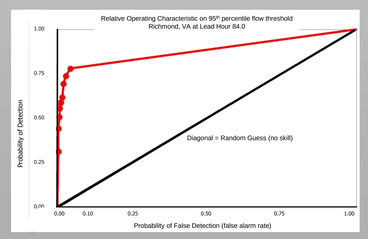


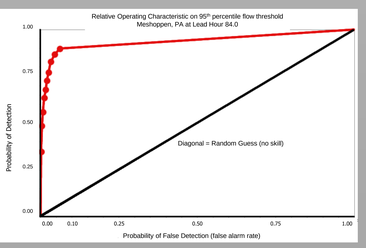

### River Flow

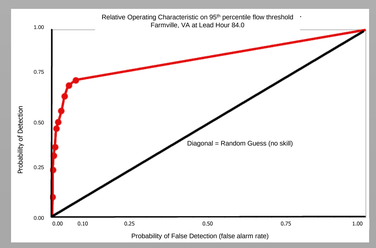



Forecast discrimination at lead time 84 hours on the 95th percentile flow. The points on the plot represent the probability of detection and probability of false detection for evenly-spaced quantiles of the forecast probability distribution.



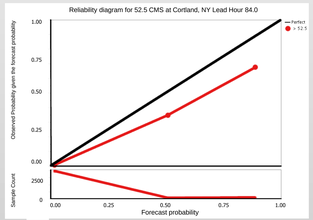



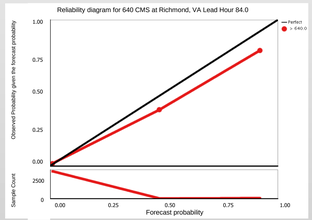




## Reliability Diagrams:

These reliability diagrams evaluate how well the ensemble forecast predicts the chance of exceeding the 95th percentile flow.

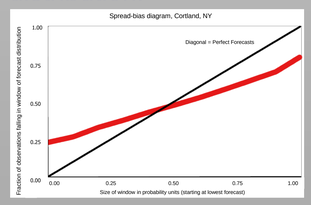


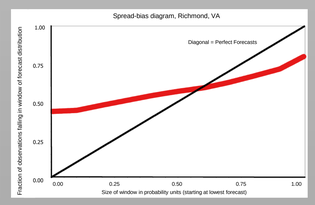




The vast majority of forecasts are for near 0% exceedence probability. In the few hundred cases where a greater chance was forecast, this chance tended to be slightly too high. This indicates either positive bias or underspreading. However, with a relatively small sample size, sampling uncertainty could be involved.

## Spread-Bias Diagrams on All Flows:

The following diagrams, similar to a Talagrand Diagram, show how often the observations fall within each portion of the ensemble forecast distribution.



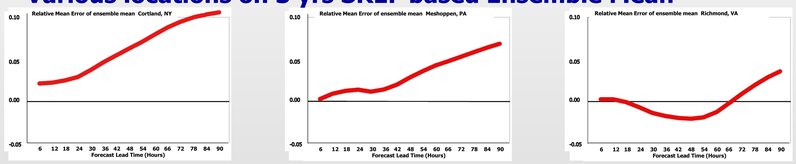




The bias on the median forecast is positive at Richmond and slightly negative at Cortland. The deviation of these forecasts from the “perfect” line is due to underspread forecasts and conditional bias of the forecasts.

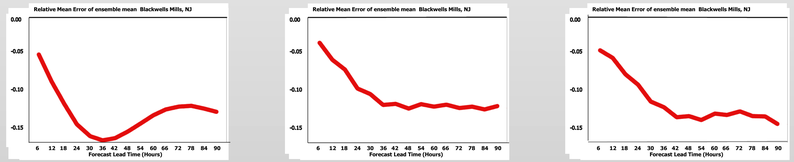
## Bias of Ensemble Mean

These are plots of relative mean error, which is the mean error divided by the mean observed value.
The streamflow forecasts have dramatically different biases in the ensemble mean at different locations due to the different hydrologic regimes present, despite similar meteorological forcings .
There is not much difference in the biases on the ensemble mean streamflow forecasts forced by the three different meteorological ensembles. NAEFS & GEFS also go out to 168 hours lead time.

### Various locations on 3 yrs SREF based Ensemble Mean




### Comparison of ensemble means from 1 yr, SREF, NAEFS, and GEFS




## Conclusions:

- **Over 1 year, GEFS, NAEFS, and SREF based river forecasts verified similarly, especially in the ensemble mean. Forecast spreads differed, with GEFS having the least spread on average (not shown).**
- **Precipitation forecast ensembles have a conditionally negative bias, with large events underforecast.**
- **This creates a conditionally negative bias in the river ensembles as well, but basin-specific hydrologic modeling uncertainties also strongly affect the results at each forecast point, creating different biases at different locations.**
- **The probability of detection on the 95th percentile flow is much higher than the probability of false detection, which indicates good discrimination between events and non-events.**
- **Underspreading and bias in the forecasts yield errors in reliability of these forecasts; they are too certain about the chance of the threshold flow being exceeded . However, future work should assess how much these apparent errors are affected by sampling uncertainty.**
- **A post-processing method, to increase spread and correct point-specific biases, could potentially improve the forecast reliability. This would attempt to account for hydrologic modeling uncertainties. Only precipitation/temperature inputs are varied in MMEFS.**