Evaluating Probabilistic Precipitation Forecasts from Convection-Allowing NWP Models as Guidance for Flash-Flood Forecasts

***Jill Hardy**: (jilldhar@usc.edu) NOAA/NSSL, Norman, OK*
***Jack Kain**: NOAA/OAR/NSSL, Norman, OK*
***David Novak**: NOAA/NWS/NCEP/HPC, Camp Springs, MD*
***J.J. Gourley**: NOAA/OAR/NSSL, Norman, OK*
***Matt Pyle**: NOAA/NWS/EMC, Camp Springs, MD*

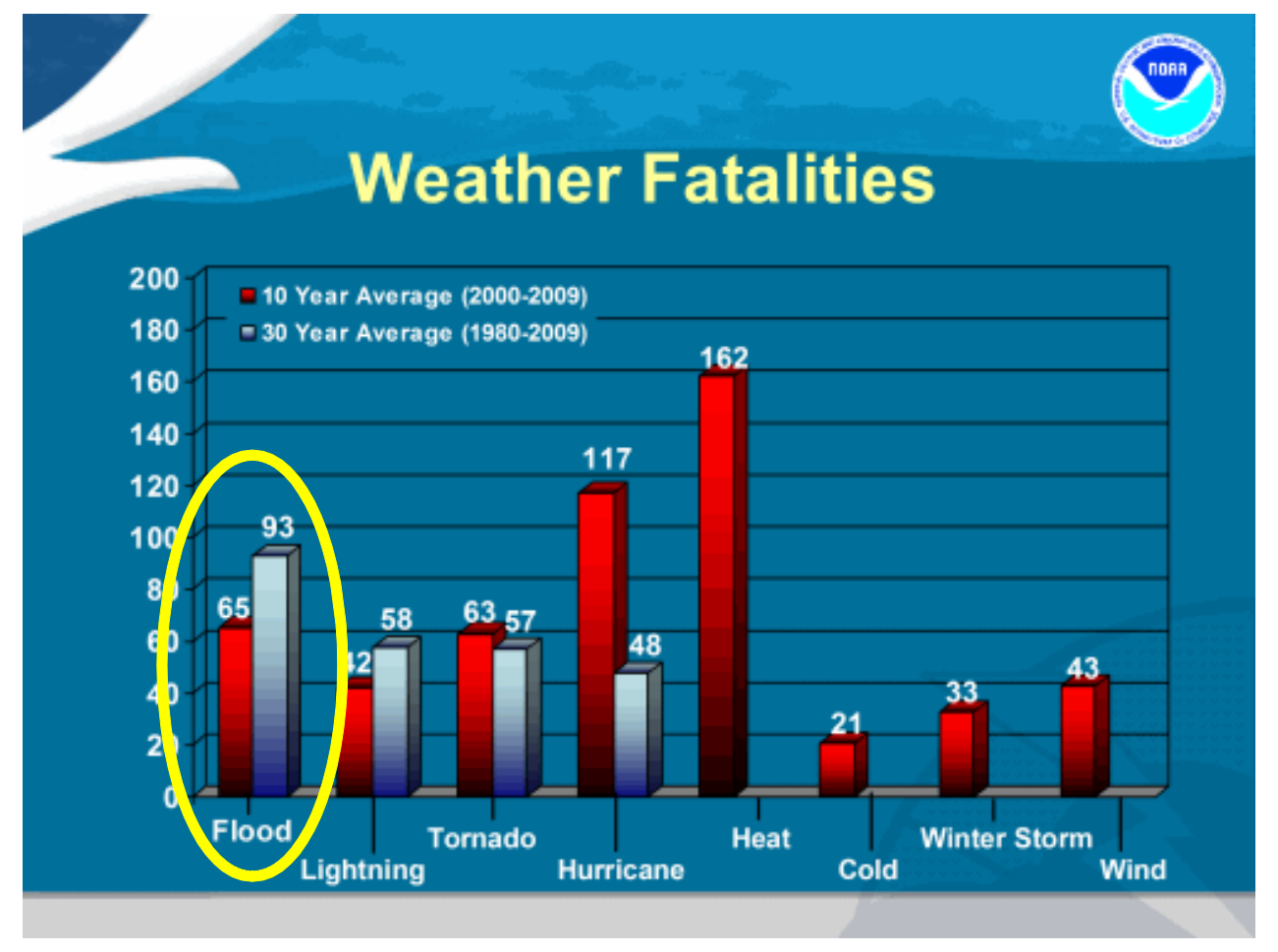
Abstract

Scientists at the NOAA/OAR/National Severe Storms Laboratory (NSSL) have been using a CONUS-scale, convection-allowing (4 km grid spacing) configuration of the Weather Research and Forecasting (WRF) model for daily real-time forecasting since 2006 (hereafter NSSL-WRF). Output from these forecasts has been provided to the NOAA/NWS/NCEP/Storm Prediction Center (SPC) and numerous local NWS offices, and has proven to be very useful as guidance for the prediction of severe weather (e.g., tornadoes, large hail, damaging winds). Over the last year, in collaboration with scientists and forecasters from the NOAA/NWS/NCEP/Hydrometeorological Prediction Center (HPC), we have begun to investigate the potential utility of these forecasts as guidance for the prediction of heavy rainfall.

In the first stage of this work, quantitative precipitation forecasts from the NSSL-WRF are compared to corresponding output from two additional realtime 4-km forecasts. Both of these are generated at NCEP/EMC; one is provided to the SPC on a daily basis like the NSSL-WRF and the other is under development for similar applications. This initial comparison indicates that the NSSL-WRF forecasts compare favorably to the forecasts from the 4-km EMC configurations. Thus, the NSSL-WRF forecasts are explored further. Like coarser resolution, convection-parameterizing models that are run operationally by the NWS, the NSSL-WRF configuration typically suffers from placement errors in predicting the location of heavy rainfall. Compared to the coarser resolution models, however, the NSSL-WRF is relatively accurate in predicting the frequency of heavy precipitation amounts. For example, its long-term frequency bias for the 1 in./6 hr. threshold is about 1.1, compared to about 0.5 for the operational NAM model. Thus, we have been exploring post-processing strategies that allow us to take advantage of the demonstrated skill in predicting the occurrence of heavy rain events, while providing meaningful quantitative expressions of the uncertainty in predicting the specific location where these events will occur.

This work explores application of a simple post-processing algorithm that introduces an estimate of spatial uncertainty in the prediction of heavy rainfall. The algorithm uses a two-dimensional Gaussian probability distribution function to define a regional "neighborhood" of non-zero probabilities for an event predicted by the model. The resulting guidance products show good reliability and resolution in preliminary testing and calibration holds promise for even better predictive skill with this technique. The potential value of these forecasts as guidance for flash flood prediction will be discussed.

Figure 1. Graph showing 10-year and 30-year averages of weather-related fatalities from 2000-2009 and 1980-2009, respectively. Flooding is circled in yellow to show that it has the highest 30-year average, thus, being the deadliest over the long term. Source: NWS.



1. Introduction
Why is this important?

Flash flooding is the #1 weather-related killer (Figure 1). It is very difficult to anticipate. Forecasters and researchers can use the results of this project to work towards increasing public safety.

We need better prediction of flooding. Warm season precipitation has low threat scores (Figure 2). Unfortunately, the majority of flooding occurs during the warm season.

Traditional numerical guidance has shortcomings. It is low resolution. For example, the NCEP/EMC NAM model uses 12-km grid spacing, requiring parameterization of thunderstorms. Due in part to this limitation, it forecasts heavy precipitation events poorly (Fig. 3), suffering from both displacement errors and under-forecasting of event frequency.

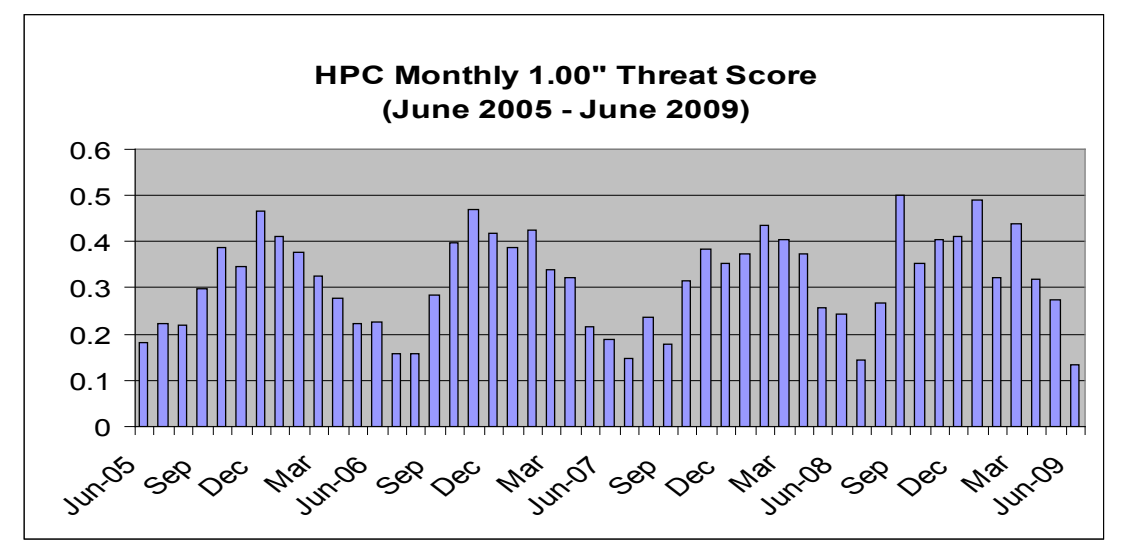


Figure 2. Graph showing HPC monthly 1" threat scores. Note the dips in scores during the warmer months, where we would ideally have the best scores. Source: David Novak, HPC.

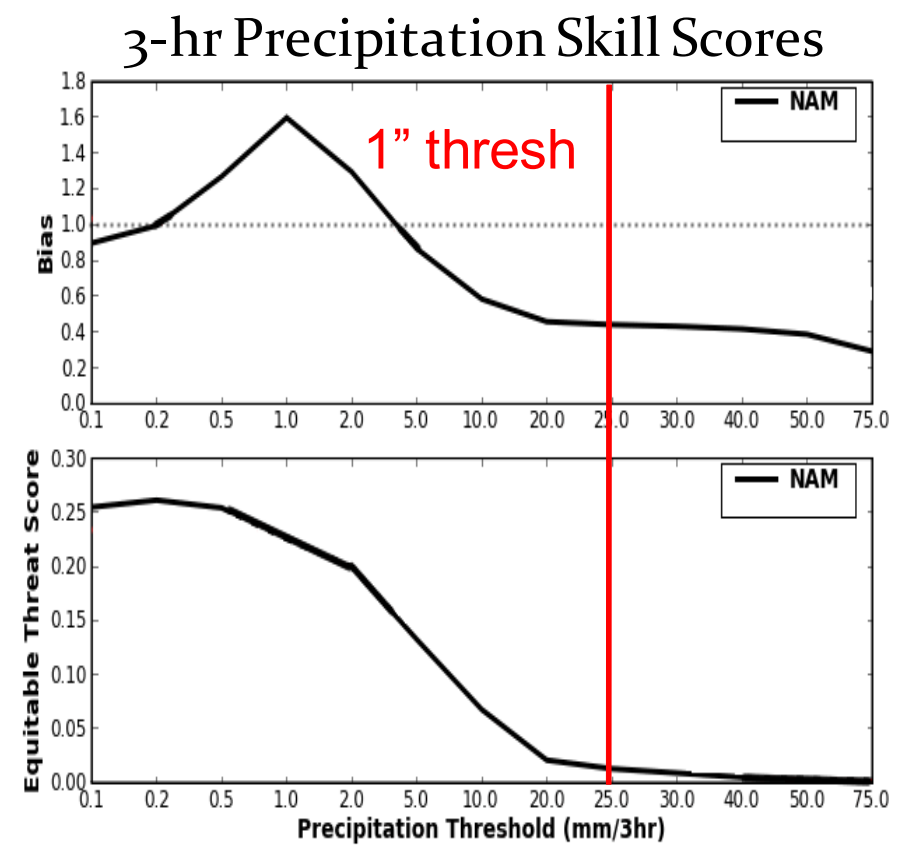


Figure 3. Graph showing 3-hour precipitation skill scores using the North America Model (NAM), for a two-year period ending April 2009. Note the scores at the 1" threshold and above. For the BIAS, coverage is less than one-to-one. For the GSS, the degree of overlap is very small. Source: Craig Schwartz.

2. Background
Model Information

National Severe Storms Laboratory (NSSL) and Environmental Modeling Center (EMC) models
--Experimental configurations of the Weather Research and Forecasting (WRF) model
--4-km grid spacing
--Primary convection-allowing models
--NSSL and EMC1: incorporated into SPC data stream since at least 2006 (WRF-ARW and WRF-NMM)
--EMC2: tested in real-time since June 2009 (WRF-NMM)

Quantitative precipitation estimations (QPEs)
--NCEP Stage IV 240 grid
--4.7-km grid for observations

All model data was interpolated to the Stage IV grid before being verified.

Storm Prediction Center (SPC) forecasters have been treating these models like operational models since 2006, but the model forecasts have never been evaluated systematically.

3. Deterministic Verification
Preliminary quantitative precipitation forecasts (QPFs) analysis using traditional scores

The two traditional verification statistics used were the frequency bias (FBIAS) and the Gilbert Skill Score (GSS), in Figures 4 and 5, respectively. Each statistic was created from a 14-month period of precipation data from April 2009 to May 2010. Several different thresholds were used ranging from 0.1mm to 75.0mm.

Conclusion:
The current operational NAM model underforecasts (FBIAS <1) and has a displacement error (low GSS), as shown in Figure 3. For the NSSL-WRF run, in particular, we still have the displacement error (low GSS), but we are more successful at predicting whether an event would occur or not (FBIAS close to 1). So can we correct for errors in displacement?

As with the NAM model, for all three model runs, we still have a discrepancy between the warm and cold seasons, as well as, with the lower and higher thresholds. The warm seasons and the higher thresholds still produce more variable and less skillful results.

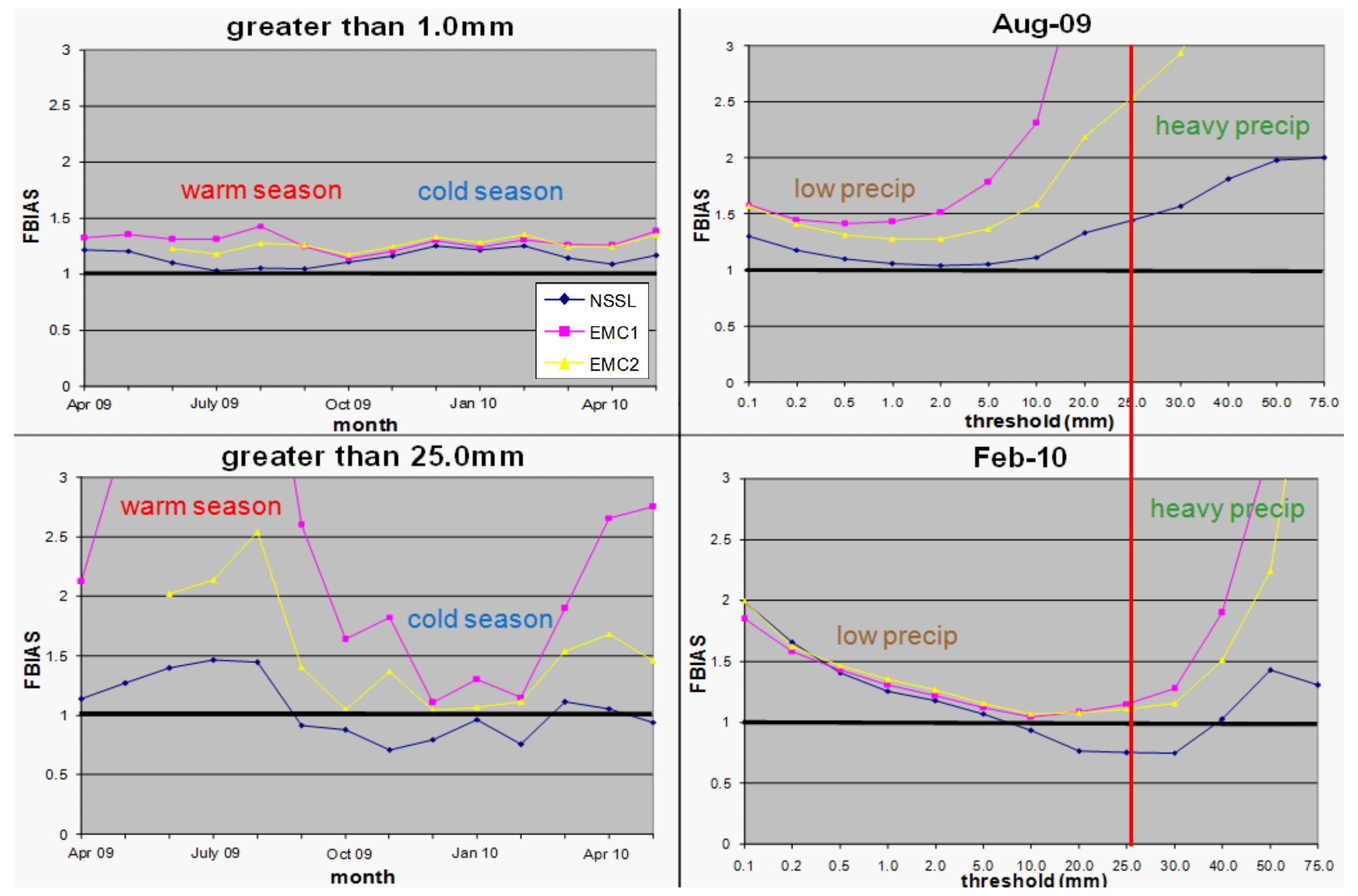


Figure 4. Graphs of the frequency biases (FBIAS). The left side shows two thresholds (top: 1.0mm, bottom: 25.0mm) by month. The right side shows a warm month (top: August) and cold month (bottom: February) with all thresholds. The 25.0mm (~1") threshold is highlighted with the red line. For all graphs, the ideal score of 1 is bolded.

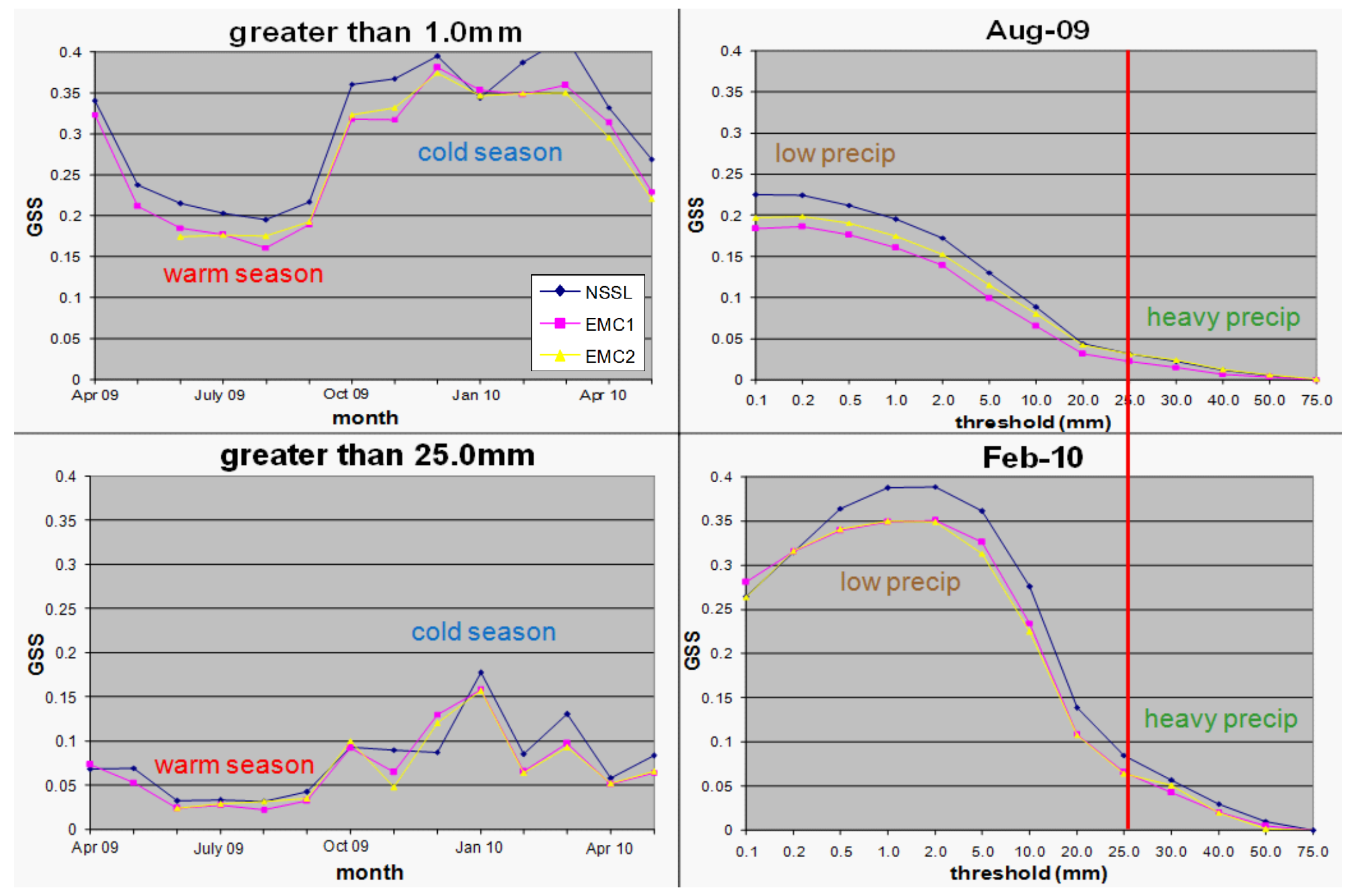
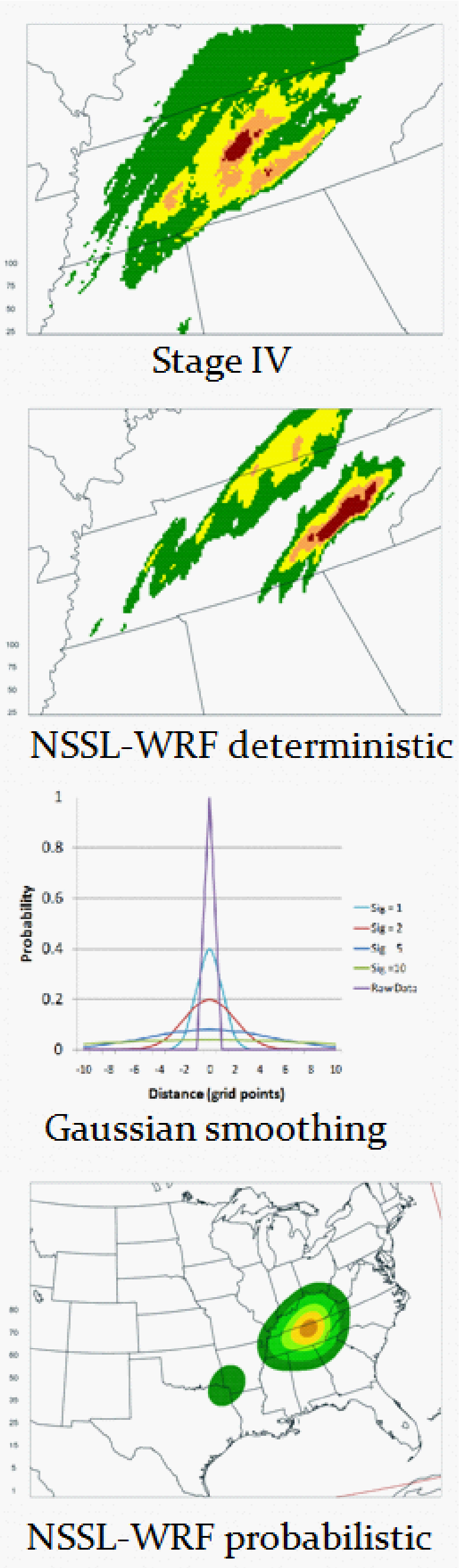


Figure 5. Graphs of the Gilbert Skill Score (GSS). The left side shows two thresholds (top: 1.0mm, bottom: 25.0mm) by month. The right side shows a warm month (top: August) and cold month (bottom: February) with all thresholds. The 25.0mm (~1") threshold is highlighted with the red line. Higher scores are ideal.

4. Probabilistic Verification
Procedure

First, we generated a probability field from the deterministic forecasts. For this project, I chose the 1 in./6 hr. threshold because it is commonly used for flash flood analysis. I applied a Gaussian smoother (Figure 6c) with a standard deviation of 30 grid points to binary (yes/no) fields in deterministic forecasts.

The result of this process is shown in an example of the Nashville, TN flooding event on May 1-2, 2010 (Figure 6). This is a comparison of the observed rainfall (Fig. 6a), the NSSL-WRF deterministic forecast (Fig. 6b), and the NSSL-WRF probabilistic forecast (Fig. 6d). There is a displacement error between the Stage IV observed data and the NSSL-WRF deterministic forecast. The probabilistic forecast smoothes out the single deterministic forecast and produces the probability field. From this probability field for this particular threshold of 1 in./6 hr., we created verification statistics.



Stage IV

NSSL-WRF deterministic

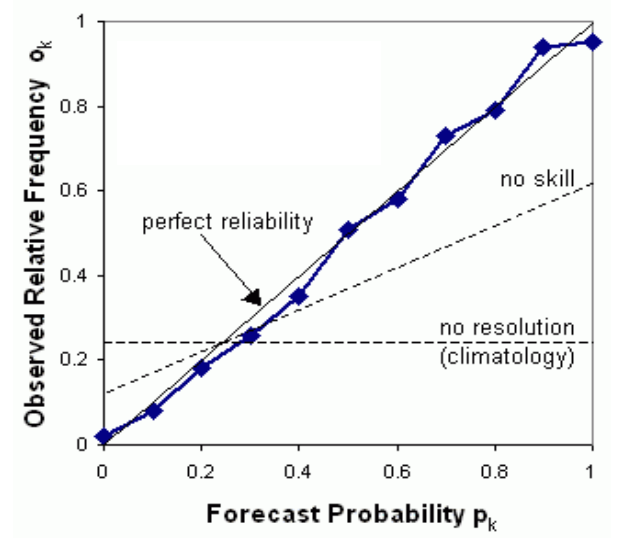
Gaussian smoothing

NSSL-WRF probabilistic

Figure 6. Maps of the flooding event in Nashville, TN from May 1-2, 2010. a) Stage IV observational data. b) NSSL-WRF deterministic forecast. c) Gaussian smoothing applied to the deterministic forecast. d) NSSL-WRF probabilistic forecast of the 1 in./6 hr. threshold. Note the displacement error between the observed rainfall and deterministic forecast.

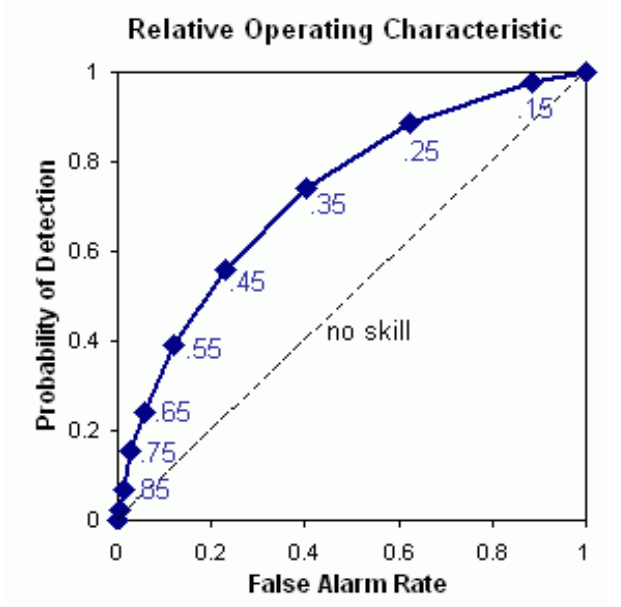
4. Probabilistic Verification (cont)
Measures of Reliability and Resolution

Reliability Diagram
--How well do the predicted probabilities of an event correspond to their observed frequencies?
--Ideal scores run along the "perfect reliability" line.



From Figure 7:
--All 3 model runs did a good job at staying near the diagonal at lower probabilities, but became less skillful at higher forecast probabilities.
--The NSSL-WRF model was the most consistent.

Relative Operating Characteristic (ROC) Curve
--What is the ability of the forecast to discriminate between events and non-events?
--POD vs POFD (False Alarm Rate)
--Area under the curve is commonly used as a measure of discrimination capability. The more area, the better the forecast.



From Figure 8:
--All 3 model runs had very high areas under the curve, showing better forecast results.
--The NSSL-WRF model did slightly better than both of the EMC model runs.

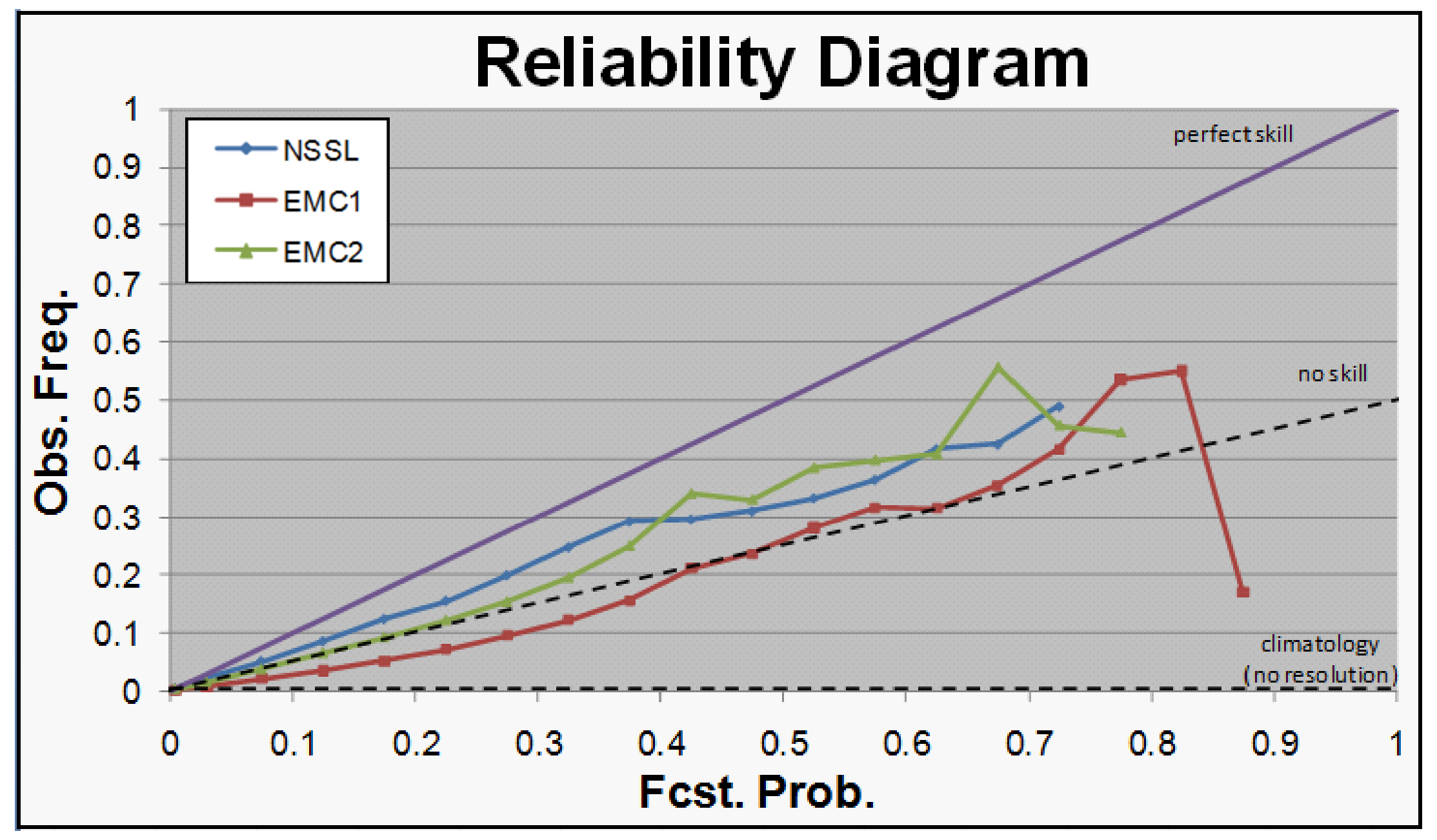


Figure 7. Reliability diagram for the probabilistic verification of the 1 in./6 hr. threshold. Perfect skill is denoted by the purple diagonal line. Above the line, the model underforecasted. Between the skill and no skill line, the model overforecasted. Below the no skill, dashed line there is no way to skillfully resolve the forecast. There is no distinction between the forecast and chance related to the climatology.

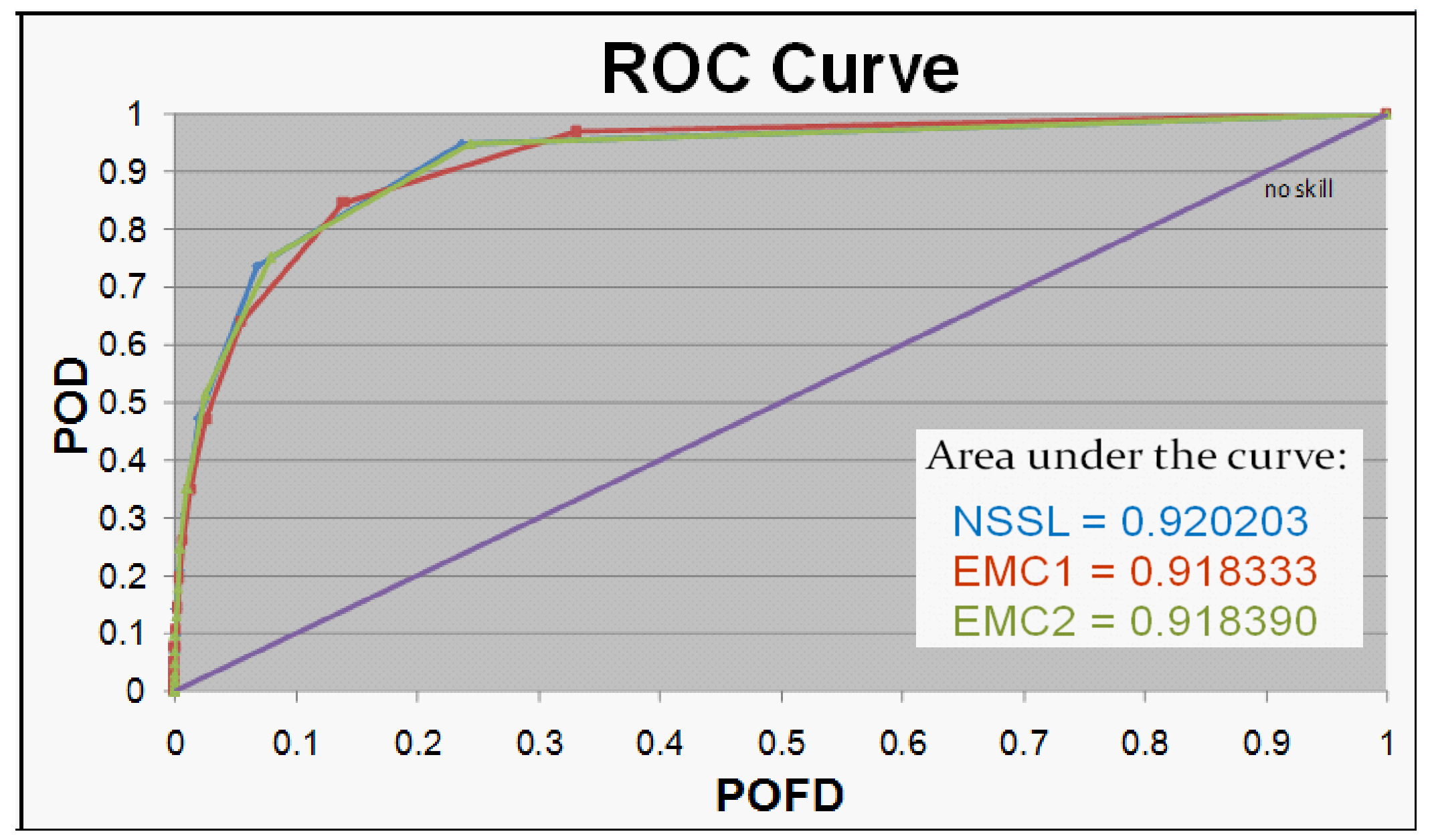


Figure 8. Relative Operating Characteristic curve for the probabilistic verification of the 1 in./6 hr. threshold. The more area under the curve denotes higher skill. Note the areas listed. Legend is the same as Figure 7.

5. Conclusions

The NSSL-WRF and a new experimental version of the 4-km WRF-NMM (run by EMC) both appear to provide valuable guidance for the prediction of heavy rainfall.

This experimental method for producing probabilistic information from deterministic model solutions shows great promise, producing forecasts with very good reliability and resolution.

References

Baldwin, Michael E., John S. Kain, 2006: Sensitivity of Several Performance Measures to Displacement Error, Bias, and Event Frequency. Wea. Forecasting, 21, 636–648.

Fritsch, J. Michael, R. E. Carbone, 2004: Improving Quantitative Precipitation Forecasts in the Warm Season: A USWRP Research and Development Strategy. Bull. Amer. Meteor. Soc., 85, 955–965.

Kain, John S., and Coauthors, 2008: Some Practical Considerations Regarding Horizontal Resolution in the First Generation of Operational Convection-Allowing NWP. Wea. Forecasting, 23, 931–952. doi: 10.1175/WAF2007106.1

Roberts, Nigel. An Investigation of the Ability of a Storm Scale Configuration of the Met Office NWP Model to Predict Flood-producing Rainfall. Tech. no. 455. Rep. no. 150. Crown, 2005. Print.

Roebber, Paul J., 2009: Visualizing Multiple Measures of Forecast Quality. Wea. Forecasting, 24, 601–608. doi:10.1175/2008WAF2222159.1

Acknowledgements

The research was performed under appointment to the National Oceanic and Atmospheric Administration Ernest F. Hollings Undergraduate Scholarship Program administered by Oak Ridge Associated Universities through a Cooperative Grant sponsored by the National Oceanic and Atmospheric Administration. I thank Jack Kain for being my mentor throughout the summer and helping me along every step of the internship. Thank you to David Novak, J.J. Gourley, and Matt Pyle for the help you provided as needed throughout the project. And a thanks to Keli Tarp and all the other interns at the NWC this summer for making it a great experience.