# 4B.1 PRELIMINARY INVESTIGATION INTO LIGHTNING HAZARD PREDICTION FROM HIGH RESOLUTION MODEL OUTPUT

Stuart D. Miller, Jr.[1,2], Gregory W. Carbin[3], John S. Kain[4], Eugene W. McCaul, Jr.[5], Andrew R. Dean[3], Christopher J. Melick[3], and Steven J. Weiss[3]

[1]NOAA/OEd/Ernest F. Hollings Undergraduate Scholarship Program
and
[2]Department of Soil, Environmental, and Atmospheric Science, University of Missouri, Columbia, MO

[3]NOAA/NWS/NCEP/Storm Prediction Center, Norman, Oklahoma

[4]NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

[5]Universities Space Research Association, Huntsville, Alabama

## 1. INTRODUCTION

The Storm Prediction Center (SPC) currently issues Experimental Enhanced Resolution Thunderstorm (ENHT) outlooks based, in part, upon probabilistic guidance from the National Centers for Environmental Prediction (NCEP) Short Range Ensemble Forecast (SREF). There is currently no method for the explicit prediction of lightning using physical means in operational deterministic models, but one has recently been implemented in the National Severe Storms Laboratory (NSSL) realtime configuration of the Weather Research and Forecasting (WRF) model. This configuration (hereafter NSSL-WRF) is initialized daily with 0000 UTC data for 36 h forecasts (e.g., Kain et al. 2010). The lightning prediction algorithm is known as Lightning Threat 3 (McCaul et al. 2009) and will be referred to as Flash Rate Density (FRD) for the purposes of this paper. To develop the FRD algorithm, McCaul et al. (2009) used the WRF, version 2.1.2 on a 2- km native grid, to simulate several convective cases in the Northern Alabama region. Simulations that failed to appropriately depict the convection as it actually occurred were discarded as the goal was to calibrate FRD, not evaluate model performance. FRD attempts explicit prediction of lightning flash-rate density, expressed in flashes $(5\ \text{min})^{-1}\ \text{km}^{-2}$ by blending graupel flux at -15°C with vertically integrated ice content. Each term was calibrated separately. The graupel flux term, $F_1$ is expressed as

$$F_1 = f\left[(wq_g)_m\right], \tag{1}$$

where f is a calibration factor, $w$ is vertical velocity, $q_g$ is graupel mixing ratio, and the subscript $m$ denotes evaluation at the -15°C level. This level was chosen to capture the mixed-phase region of convective clouds. The vertically integrated ice term, $F_2$, is expressed as

$$F_2 = h\left[\int \rho\left(q_g + q_s + q_i\right) dz\right], \tag{2}$$

where $\rho$ is the local air density and $q_g$, $q_s$, and $q_i$ are mixing ratios of graupel, snow, and ice, respectively, and h is a second calibration factor. Both the McCaul et al. (2009) and NSSL-WRF implementation use the WSM6 microphysics scheme (Hong et al. 2004; Hong and Lim 2006) to predict hydrometeor mixing ratios but the concepts should be applicable to any microphysical parameterization.

In order to find $f$ in (1), McCaul et al. (2009) examined each of the operative simulations and plotted maximum values of $F_1$ and lightning flash rate from the North Alabama Lightning Mapping Array (LMA). From this, a linear regression with its intercept at the origin and a slope of 0.042 was obtained. This resulted in $f$ equaling 0.042. Using the same method with $F_2$ to calculate $h$ in (2) generates a value of 0.20.

Furthermore, the simulations performed by McCaul et al. (2009) indicated that $F_1$ produced good temporal resolution for lightning occurrence, while underforecasting its areal coverage. On the

[1] *Corresponding author address:* Stuart Miller, 302 ABNR, Columbia, MO 65211, sdmvd6@mail.missouri.edu

other hand, $F_2$ yielded better lightning coverage since it accounted for anvil cirrus and was not focused solely on updraft regions. Consideration of these factors led to a blended threat, $F_3$ (or FRD), which is a weighted combination of (1) and (2).

$$FRD = r_1 F_1 + r_2 F_2 , \quad (3)$$

where $r_1 = 0.95$ and $r_2 = 0.05$. These values were determined by testing the effect of different weight combinations on peak flash-rate densities and coverages in the simulations and again comparing to the LMA to find the most favorable correlation. The final choice of the weights was dictated by two simple factors: 1) the graupel flux weight had to be large in order to retain realistic time variations; 2) the vertically-integrated-ice factor had to be large enough so that areal coverage values, appropriately thresholded, do not show unstable behavior, which is possible when the threshold value becomes very small.

The aim of this study is to provide a preliminary verification of FRD in the NSSL-WRF. However, some complications arise when considering the consequences of examining its performance in this particular context. For example, since FRD was calibrated by McCaul et al. (2009) in a version of the WRF with a finer grid spacing than the NSSL-WRF (grid point spacing of 4 km), its calibration factors are not quantitatively applicable to the NSSL-WRF. Also, FRD was calibrated over Northern Alabama, which has a specific climate with an associated typical range of flash rate intensity and areal coverage that may differ from other regions. In addition, $F_1$ and $F_2$ are computed in the NSSL-WRF using hourly-maximum values (see Kain et al. 2010) of graupel flux and vertically integrated ice, rather than instantaneous values at output time. Finally, the North Alabama LMA used in calibration detects total lightning. In order to verify FRD in the NSSL-WRF domain, the National Lightning Detection Network (NLDN) had to be used, which primarily detects cloud to ground (CG) flashes. These are severe limitations that preclude quantitative assessments of lightning intensity until additional calibration is performed. Thus, this preliminary evaluation focuses on model predictions of where and when lightning will occur rather than how intense it will be.

Section 2 provides details on the steps taken to evaluate FRD with NLDN data as well as its performance compared with the calibrated cloud physics thunder parameter (CPTP) from the SREF (Bright et al. 2005; Bright et al. 2009) and the SPC ENHT outlooks. Section 3 reveals the results of these verification methods and their interpretation. Section 4 contains a summary of the results, section 5 lists references, and section 6 provides all figures.

## 2. METHODOLOGY

### 2.1 *Data Acquisition and Manipulation*

FRD has been available from the NSSL-WRF (initialized daily at 0000 UTC) since 12 March 2010 and a preliminary three month data set was available when this study was conducted. The dates included in this study are 12 March 2010 - 11 June 2010, with the exception of 20 March, 12 April, and 25 April due to gaps in data. In order to avoid the daily overlap of data for forecast hours 0 - 12 and 24 - 36, only forecast periods ending at hours 13 - 36 were examined. A CONUS mask was applied to both forecast (NSSL-WRF) and observed (NLDN) data.

Hourly NLDN data was mapped in binary format to the nearest NSSL-WRF 4 km grid point, i.e., points with one or more lightning strikes were assigned a value of 1, while all others remained zero.

Next, a threshold value for FRD was selected such that the frequency bias, aggregated over the entire time period, was approximately 1, i.e., the number of FRD values exceeding this threshold was approximately equal to the number of NLDN values greater than zero. This threshold was determined to be 0.55 flashes $(5\ \text{min})^{-1}\ \text{km}^{-2}$. The FRD grids were then converted to binary values on the basis of this threshold. Using the binary data, Gilbert Skill Score (GSS) was calculated for the hourly FRD data on the NSSL-WRF grid for the sample period. It was found to be 0.03 – very low, but not surprising considering the fact that high-resolution deterministic grid-point output was being assessed. The difficulty of obtaining a high GSS in this situation is substantial because lightning occurs on such small scales and model forecasts nearly always have mesoscale and stormscale displacement errors.

To account for these displacement errors, a two-dimensional Gaussian probability distribution function (PDF) was applied to each point on the FRD binary grid. This process created a field with values between 0 and 1 that can be considered a spatial probability field.

Specifically, the probability value at any point, x, is given by

$$P(x)= \sum_{n=1}^{N} \frac{b_n}{2\pi\sigma} \exp\left(-\frac{d_n^2}{2\sigma^2}\right), \quad (4)$$

where $N$ is the total number of grid points, $b_n$ is the binary value at point $n$, $d_n$ is the distance to the $n^{th}$ grid point, and $\sigma$ is the standard deviation of the distribution. For this study, a $\sigma$ value of 120 km was used. These probability fields were verified for reliability and resolution using reliability diagrams and Relative Operating Characteristic (ROC) curves.

A major goal was to compare FRD performance to that of the CPTP and ENHT outlooks. The latter do not include a continuous range of probability values from 0 - 100%, rather, the ENHT product includes only the probability of lightning occurrence at 10%, 40%, and 70%. The ENHT outlooks are issued five times daily and are valid as follows:

0600 UTC issue: Forecast valid 1200 - 1600 UTC
Forecast valid 1600 - 2000 UTC
Forecast valid 2000 - 0000 UTC

1300 UTC issue: Forecast valid 1600 - 2000 UTC
Forecast valid 2000 - 0000 UTC
Forecast valid 0000 - 0400 UTC

1700 UTC issue: Forecast valid 2000 - 0000 UTC
Forecast valid 0000 - 0400 UTC
Forecast valid 0400 - 1200 UTC

2000 UTC issue: Forecast valid 0000 - 0400 UTC
Forecast valid 0400 - 1200 UTC

0100 UTC issue: Forecast valid 0400 - 1200 UTC

For this preliminary study, the focus is on the 1300 UTC ENHT outlook and the forecast valid 2000 - 0000 UTC. Furthermore, for comparison with the NSSL-WRF based forecasts, we concentrate on the 0300 UTC initialization of the SREF. This initialization is most relevant to the 0000 UTC NSSL-WRF because at these times both modeling systems benefit from the broad spectrum of observational data available at 0000 UTC. In addition, the raw hourly FRD and NLDN data are aggregated for the 4-h 2000 - 0000 UTC period for this comparison. Finally, the aggregated binary FRD and NLDN data are mapped to the 40-km NCEP 212 grid to be consistent with the other two datasets, and the FRD-based probabilities are recomputed on this grid using the same parameters in the Gaussian PDF.

ENHT outlook grids are not archived for more than 30 days, so precise grid point calculations could not be made to determine their reliability and ROC curve. The SPC does, however, archive real-time verification of ENHT outlooks in the form of reliability diagrams and ROC curves. The number of grid points contained within each probability bin is listed on the diagrams, so by using the observed relative frequency of lightning, the number of verifying grid points with observed lightning within each bin could be calculated for the study period. It was admittedly not as precise a method as the authors preferred, since it required estimation of observed relative frequencies from a graphical product, but it was the only option available if FRD was to be compared to these operational forecasts.

### 2.2 *Plot Generation*

Once all data were in their desired forms, time series of both hourly and average hourly FRD percent domain coverage vs. percent domain coverage of observed lightning from both sets of binary grids (at 4-km grid spacing) were plotted. (Figs. 1 and 2). Using the hourly FRD probabilities on the native grid along with the binary observed flashes, the reliability and ROC curve using 0 - 1%, 1 - 5%, 5 - 10%, 10 -15% . . . 95 - 100% probability bins for FRD could be calculated over the entire period. (Figs. 3 and 4)

The reliabilities and ROC curves were also plotted for the CPTP, ENHT outlooks, and the 4-hourly FRD data on the 212 grid. They were plotted for the 2000 - 0000 UTC period issued in the 1300 UTC ENHT outlook with probability bins of 0 - 10%, 10 - 40%, 40 - 70%, and 70 - 100%. (Figs. 7 and 8) This was done to remain consistent with the format in which the SPC displays its real-time verification of the CPTP and ENHT outlooks. Plots of the reliability and ROC curve were then created for the 4-hourly FRD data along with the CPTP data using 5% probability bins. (Figs. 5 and 6)

## 3. RESULTS

### 3.1 *Hourly FRD Data*

It was found that the hourly FRD data on the NSSL-WRF grid corresponded reasonably well with NLDN lightning over the sample time period. When grid coverage of binary NLDN and FRD data is plotted in time-series format for all hours a rough correlation is evident (Fig. 1). When these same data are plotted as average hourly values for each field (all days included) an even clearer correspondence emerges (Fig. 2). Most notably, the diurnal cycle of lightning activity appears to be predicted remarkably well.

Once the skill of FRD in regard to timing of lightning was established, its skill with respect to placement was assessed using reliability diagrams and ROC curves. They can be seen for the hourly, 4-km gridded data in figs. 3 and 4. The reliability was relatively good from 0 - 1% to 15 - 20%, but noticeable overforecasting occurs from 20 - 25% to 30 - 35%. (Fig. 3) There are no forecast guidance probabilities above 35%. These data resulted in a ROC curve area of 0.920 (Fig. 4), indicating very good resolution between events and non-events.

### 3.2 *Summed FRD Data*

The daily 4-hour, 40-km FRD data discussed in section 2.1 were scored next over the sample period for the hours 2000 - 0000 UTC. The ROC curve area (using 0 - 1%, 1 - 5%, 5 - 10%, 10 -15% . . . 95 - 100% probability bins) for the 4-hour CPTP from the SREF was 0.914 compared to 0.906 for FRD. (Fig. 6) Reliability for the CPTP evaluated with these bins was very good up to about 55%, above which modest overforecasting was evident, with somewhat erratic behavior at the highest probabilities. (Fig. 5) The FRD also exhibited very good reliability and was perhaps somewhat better than the CPTP above about 55%.

Finally, the reliabilities and ROC curves for FRD, the CPTP, and ENHT outlooks were computed over the sample period. As stated in section 2.1, because of the broad probability intervals associated with the ENHT outlooks, calculations for all three were made using 0 - 10%, 10 - 40%, 40 -70%, and 70 - 100%. In terms of ROC curve area, the human forecasts (ENHT) scored the highest with area of 0.910. This was followed by FRD with an area of 0.873 and the CPTP with an area of 0.870. (Fig. 8)

Since ENHT outlook probabilities are discontinuous, there is an acceptable reliability range in which the forecasts can fall. This range is represented by step function bounds illustrated by the red lines in Fig. 7. As can be seen, all three probabilistic forecasts lie in these acceptable bounds from 0 - 10% to 40 - 70%. In the 70 - 100% bin, both CPTP guidance and ENHT outlooks provide a slight overforecast while FRD remains just within the acceptable reliability range.

## 4. CONCLUSION

These preliminary results show that when spatial uncertainty is represented using a Gaussian PDF, the FRD algorithm in the NSSL-WRF appears to perform quite well at predicting placement, and timing of lightning, when evaluated on larger time and space scales, i.e. a 4-h time window and a 40 km grid. In the more demanding context of a 1-h time window and the native 4-km NSSL-WRF grid, verification scores are lower, but skill is still evident.

In general, the FRD-based lightning forecasts had skill levels comparable to the SPC's ENHT product and the SREF-based CPTP. These results are very encouraging because they come from a preliminary assessment of an uncalibrated FRD-based product. Thus, the FRD-based output shows promise as an additional tool to supplement current forecast guidance for lightning prediction.

## ACKNOWLEDGEMENTS

The lead author is appreciative of the Ernest F. Hollings Undergraduate Scholarship Program for providing the funds to make this work possible. We are also grateful to Scott Dembek (CIMMS/NSSL) and Jonathan Case (USRA/SPoRT) for implementing the FRD-related algorithms in the NSSL-WRF forecasting system. Finally, we would like to thank Jill Hardy (NOAA/OEd/Hollings Scholarship Program) for providing an efficient method for evaluating ROC curves.

## 5. REFERENCES

Bright, D.R. and J.S. Grams, 2009: Short Range Ensemble Forecast (SREF) calibrated thunderstorm probability forecasts: 2007-2008 verification and recent enhancements. Preprints, *Conf. on Meteorological Applications of Lightning Data*, Phoenix, AZ, Amer. Meteor. Soc., CD-ROM, 6.3

Bright, D.R., M.S. Wandishin, R.E. Jewell, and S.J. Weiss, 2005: A physically based parameter for lightning prediction and its calibration in ensemble forecasts. Preprints, *Conf. on Meteorological Applications of Lightning Data*, San Diego, CA, Amer. Meteor. Soc., CD-ROM, 4.3

Hong, S.-Y., and J.-O. J. Lim, 2006: The WRF single-moment 6-class microphysics scheme (WSM6). *J. Korean Meteor. Soc.*, **42**, 129–151.

Hong, S.-Y., J. Dudhia, and S.-H. Chen, 2004: A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Mon. Wea. Rev.*, **132**, 103–120.

Kain, J. S., S. R. Dembek, S. J. Weiss, J. L. Case, J. J. Levit, and R. A. Sobash, 2010: Extracting unique information from high resolution forecast models: Monitoring selected fields and phenomena every time step. *Wea. Forecasting*, **25**, 1536-1542.

McCaul, E. W., S. J. Goodman, K. M. LaCasse, D. J. Cecil, 2009: Forecasting Lightning Threat Using Cloud-Resolving Model Simulations. *Wea. Forecasting*, **24**, 709–729.

## 6. FIGURES

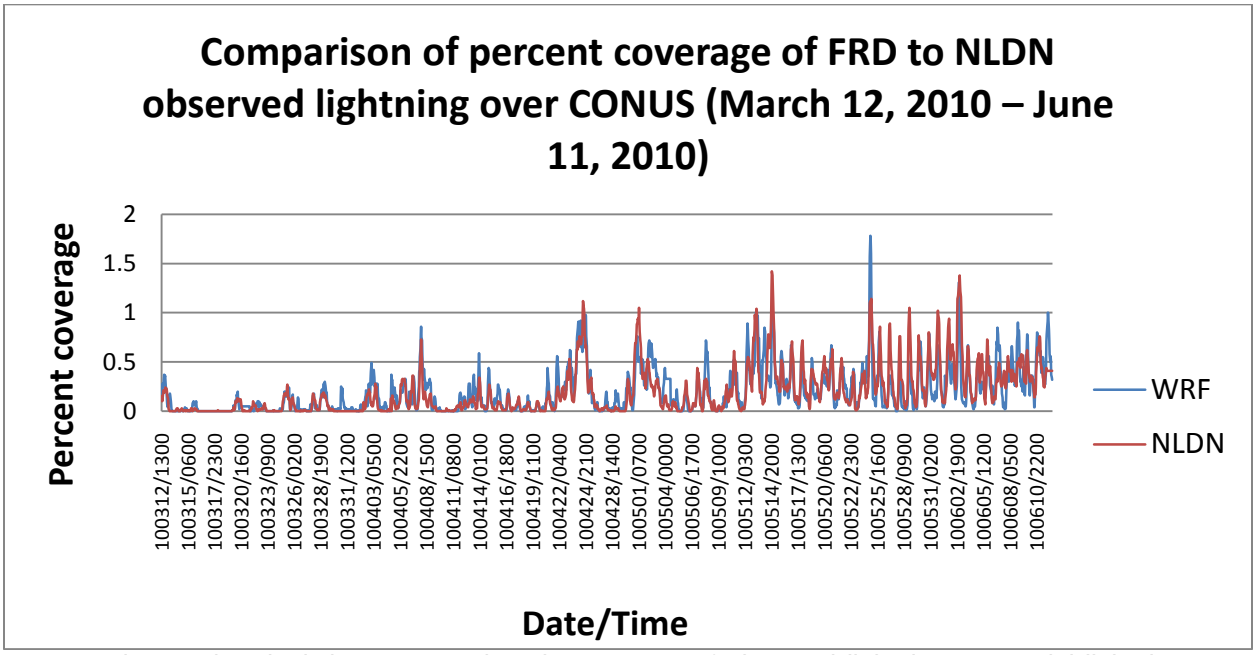


**FIG. 1.** Time series depicting percent domain coverage of observed lightning vs. model lightning over CONUS.

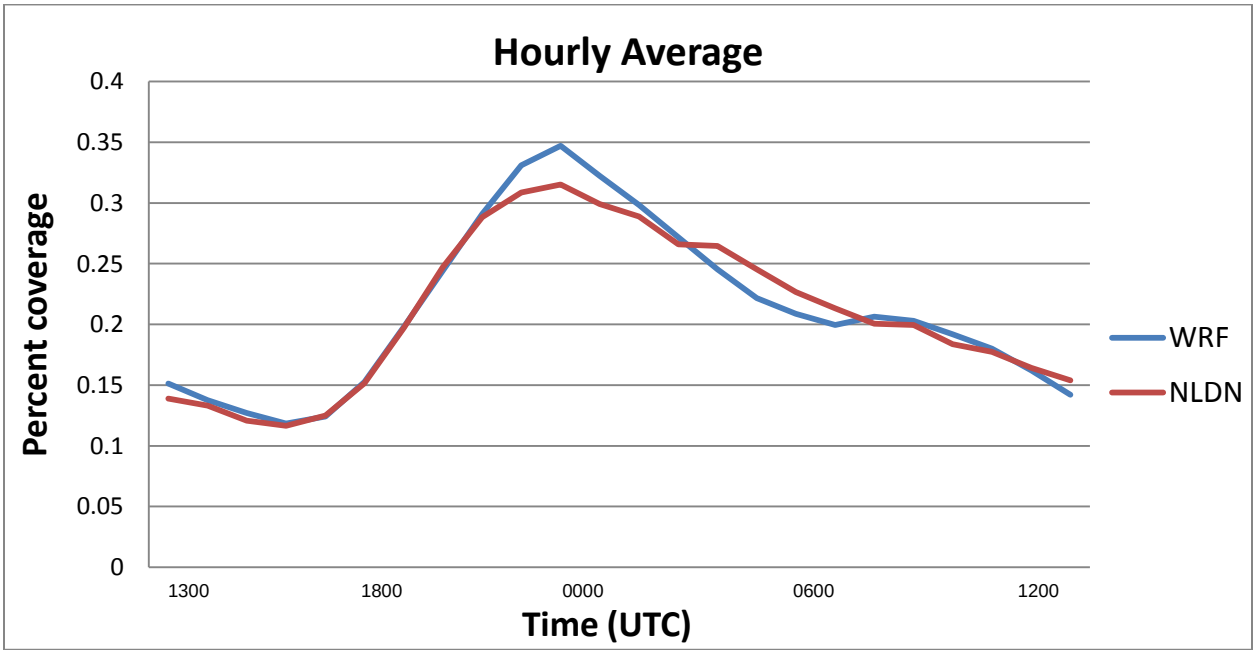


**FIG. 2.** As in Fig. 1., except average hourly values for the sample time period are presented.

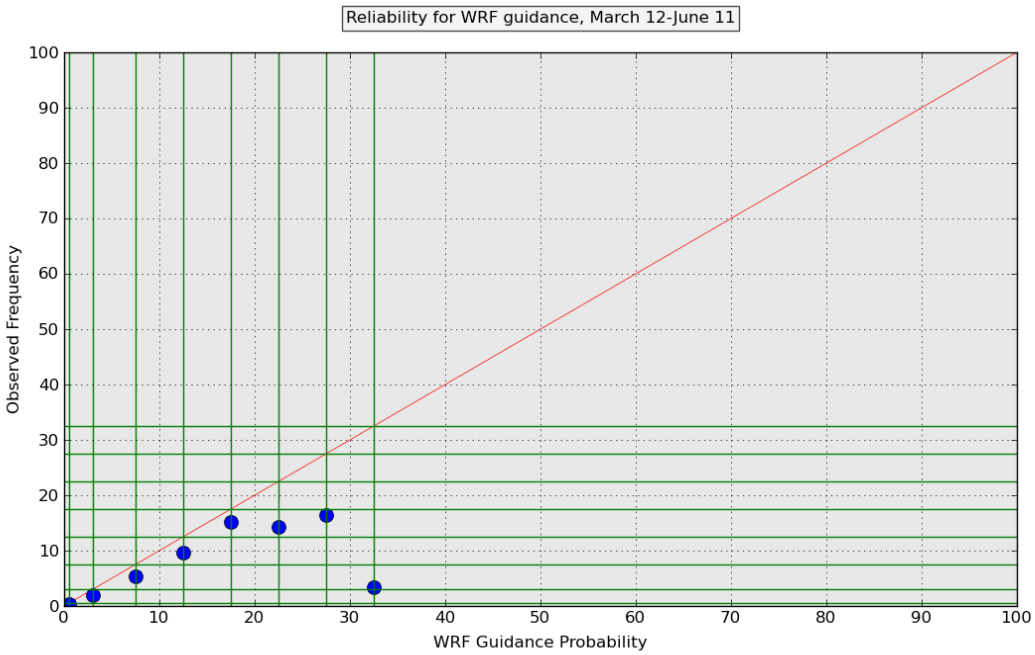


**FIG. 3.** Reliability diagram depicting guidance probability on the abscissa and observed relative frequency on the ordinate. The diagram was computed using 0 - 1%, 1 - 5%, 5 - 10%, 10 - 15% . . . 95 - 100% probability bins. The red line is the perfect skill line.

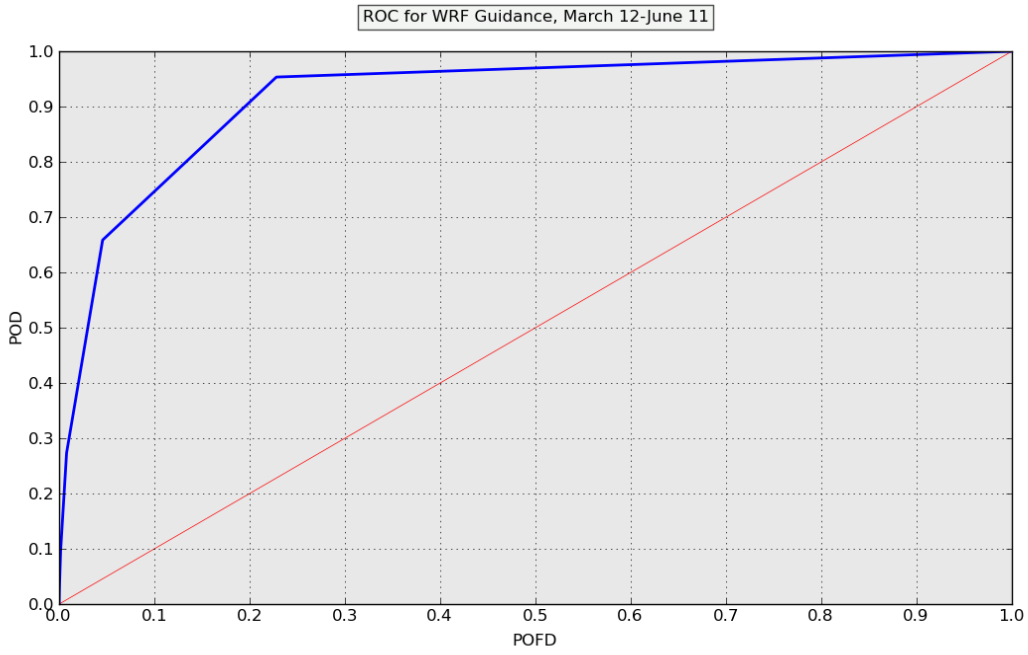


**FIG. 4.** ROC curve for FRD guidance depicting probability of false detection (POFD) on the abscissa and probability of detection (POD) on the ordinate. The diagram was computed using 0 - 1%, 1 - 5%, 5 - 10%, 10 - 15% . . . 95 - 100% probability bins. The red line is the no skill line.

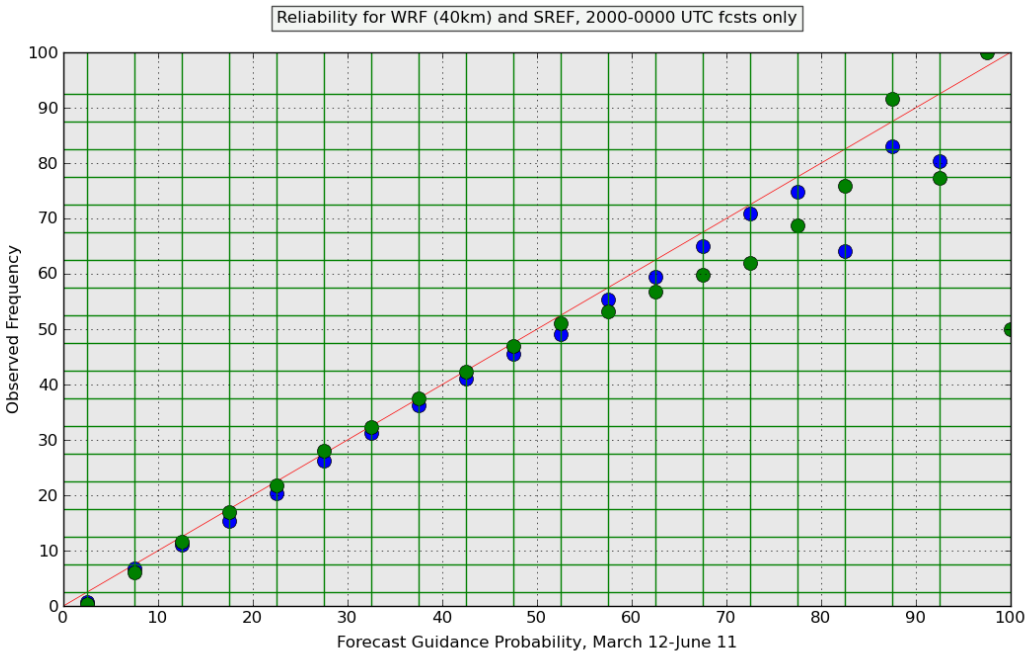


**FIG. 5.** As in Fig. 3, except comparing FRD (blue) to CPTP (green). Both data are valid daily from 2000 - 0000 UTC over the forecast period and mapped to the NCEP 212 grid.

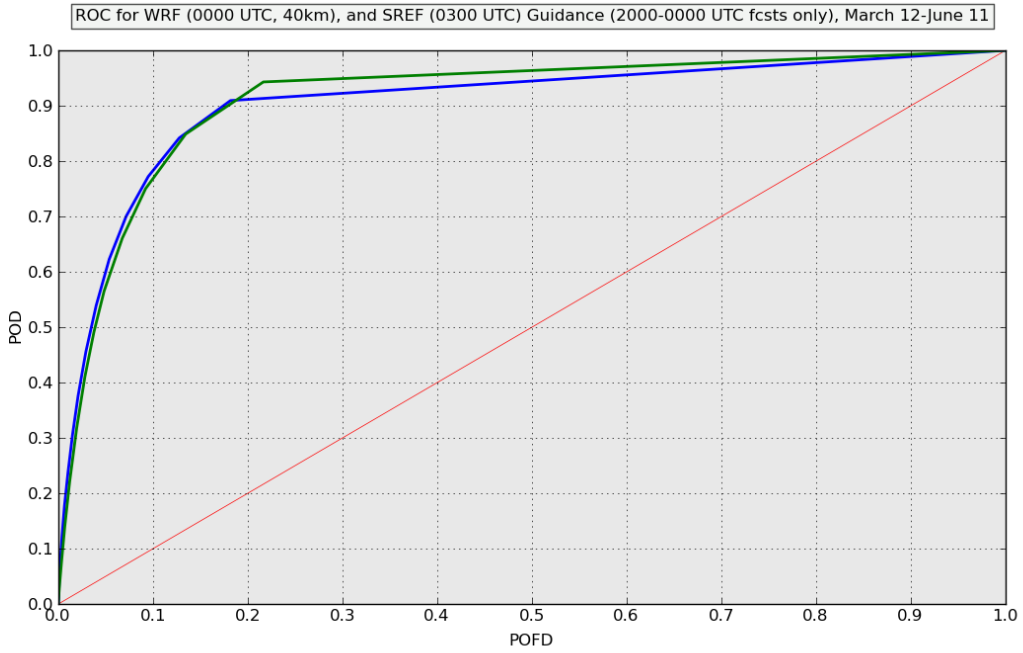


**FIG. 6.** As in Fig. 4, except comparing FRD (blue) to CPTP (green). Both data are valid daily from 2000 - 0000 UTC over the forecast period and mapped to the NCEP 212 grid.

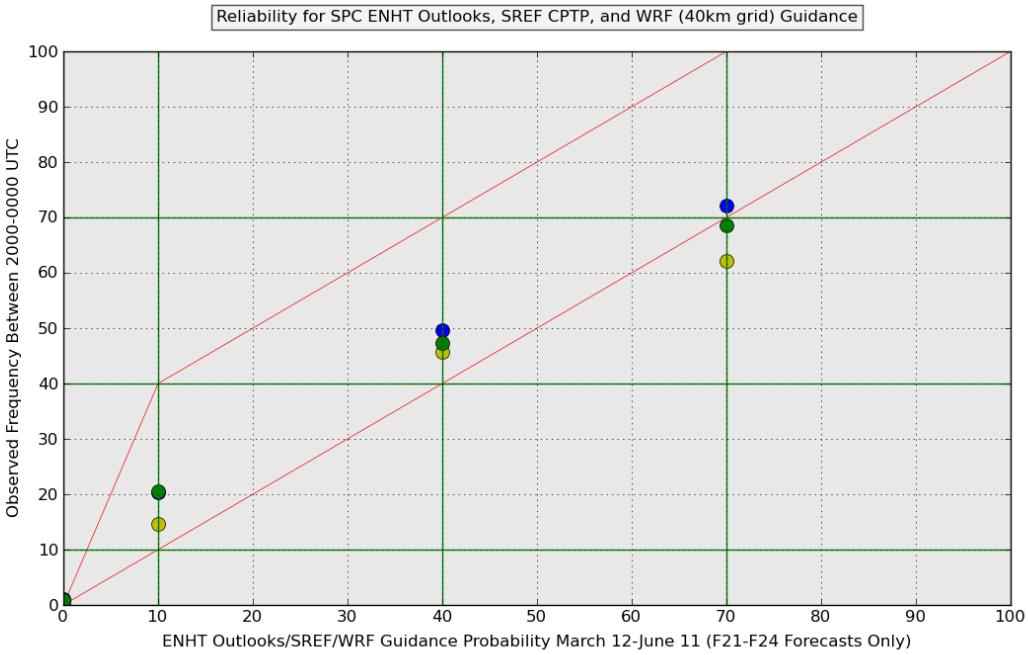


**FIG. 7.** As in Fig. 5, except comparing ENHT outlooks (yellow), FRD (blue), and CPTP (green). The probability bins used for computation were 0 - 10%, 10 - 40%, 40 - 70%, and 70 - 100%. The red line in addition to the perfect skill line is a "step-function" line. The area between the perfect skill line and the "step-function" line defines an acceptable reliability range.

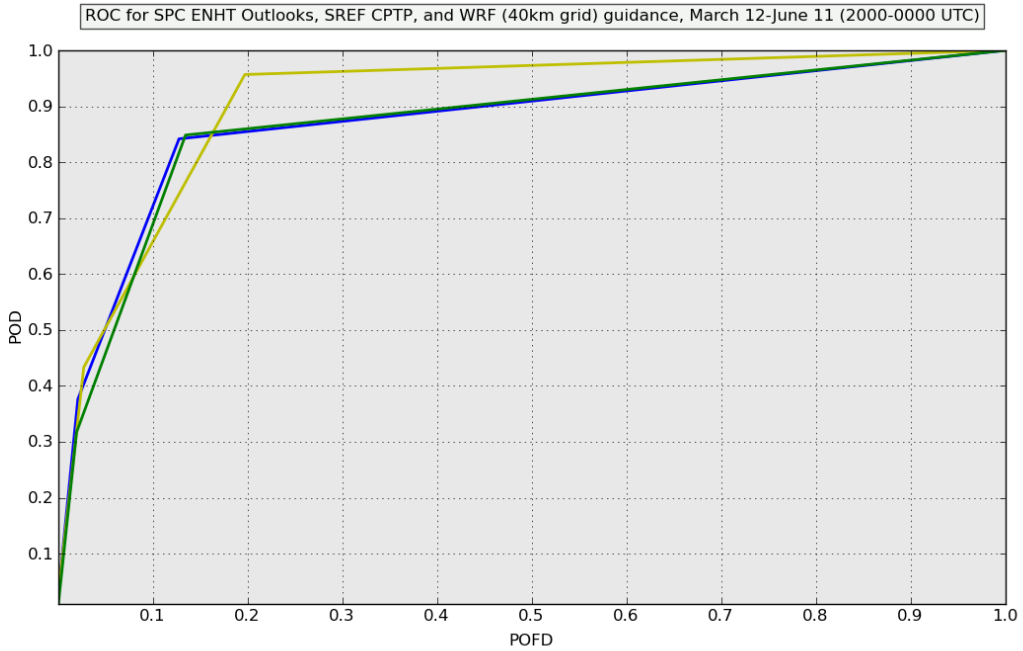


**FIG. 8.** As in Fig. 6, except comparing ENHT outlooks (yellow), FRD (blue), and CPTP (green). The probability bins used for computation were 0 - 10%, 10 - 40%, 40 - 70%, and 70 - 100%.