Updraft Helicity as a Forecast Parameter

S. M. Hitchcock[1,2], P. T. Marsh[2,3], H. E. Brooks[3], and C. A. Doswell III[4]

[1]National Weather Center Research Experiences for Undergraduates Program
[2]School of Meteorology, University of Oklahoma, Norman, OK
[3]NOAA/National Severe Storms Laboratory, Norman, OK
[4]Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, OK

ABSTRACT

Improved science and technology has created the opportunity to explore the impacts of different model diagnostic fields as indicators of convection developed in high-resolution numerical models. Indication of the success of different diagnostic fields has been discussed (Kain et al. 2008, Sobash et al. 2008). Updraft helicity (UH) has shown a particular ability to identify supercell-like structure in convection allowing model observed locations. UH will be examined to determine the best integration layer over which to calculate UH.

Output of updraft helicity over different layers from the convection allowing 4-km National Severe Storms Laboratory- Weather Research and Forecasting Radar (NSSL-WRF) Advanced Research WRF (ARW) from the Spring Experiment 2008 was compared to Storm Prediction Center (SPC) storm reports using contingency tables. Verification measures (Probability of Detection, False Alarm Ratio, Critical Success Index, bias) were calculated from the contingency tables and used to create several visual comparisons. These include Relative Operating Characteristic curves (ROC) (Mason 1982), and Performance Diagrams (Roebber 2008), as a comparison of different depth's success as a forecast parameter.

1 . INTRODUCTION

Finding forecast verification techniques appropriate for rare severe weather events, is not a new challenge. From as early as Finley's publication entitled "Tornado Predictions" in 1884, scientists have been working to understand the forecast verification of severe weather events. These processes, however challenging, are essential to the protection of life and property in a severe weather event.

Severe weather event verification in the United States has been based on several measures of skill. Emphasis has been placed on increasing the Probability of Detection (POD) while limiting the False Alarm Rate (FAR). The relationship between

Corresponding author address:
Stacey Hitchcock
2730 Chautauqua
Traditions Q206
Norman, OK 73072
Email: Stacey.M.Hitchcock-1@ou.edu

POD and FAR is such that, in order to achieve this, improvements to science and technology must be made (Brooks 2004). Considering the implications that a missed detection has on the protection of life and property, increasing POD is given priority; a False Alarm costs significantly less than a missed detection.

As technology has improved, numerical weather prediction (NWP) models have also improved. Models are now capable of simulating deep convective storms owing to smaller grid spacing over larger domains. This has allowed models to operate on the time and space scales appropriate for the Storm Prediction Center (SPC). The SPC outlook time scale is 24 hours while the space scale for verification purposes is for a radius of 40-km. These convection-allowing numerical models permit grid scale processes to develop storms instead of parameterizing them. Previously, in models with coarser resolution, relatively crude convective parameterization left forecasters to base their forecasts on environmental conditions, rather than explicit model forecast of thunderstorms.

Updraft helicity (UH) as a new diagnostic field in model forecasts was introduced by Kain et al. (2008). During the Spring Experiment 2005 he found that "models generated storms containing localized UH maxima under environmental conditions that produced observed supercells" (Kain et al. 2008). UH maxima from 2-5 km was proven useful in severe event forecasts due to its unique ability to detect simulated convection (Sobash et al. 2008).

This inspires the question: Is there a better integration layer than the 2-5 km layer that is currently used in model forecasts?

2. BRIEF HISTORY OF VERIFICATION

After Finley's publication entitled "Tornado Predictions" was released in 1884 and claimed a success rate > 95%, many subsequent articles were published (cited in Murphy 1995). Finley's method of verification included the success of predicting tornado events as well as non-tornado events in his calculation of "Percentage of Tornado Predictions" (Finley 1884). The articles that followed created new methods, and critiqued Finley's methods, as well as each other's methods. Gilbert explained that including the successful prediction of non-events in a forecast does not show the true skill of a forecaster because one could predict no tornadoes for an entire year and still have 95% success (Gilbert 1884). The issue becomes more complicated when considering that the cost of a missed detection can be significantly higher than that of a false alarm.

Even after more than a century of new verification techniques, studies, and technology, scientists are still working to understand severe event forecasting.

3. DATA AND METHODS

The dataset used is output from the 4-km grid length Weather Research and Forecasting (WRF) model - Advanced Research WRF (WRF-ARW) run at the National Severe Storms Lab (NSSL) during the 2008 Hazardous Weather Testbed (HWT) Spring Experiment (SE2008). SE2008 was held from April 16 – June 8, 2008. The model was initialized daily at 00 UTC and ran out to forecast hour 36. However, 16 May was omitted because on that day, the model only ran out to forecast hour 30. This left 52 complete model forecasts. We only evaluated forecast hours 12-36 which correspond to the Storm Prediction Center (SPC) convective day (12 UTC – 12 UTC).

Updraft helicity was calculated over all combinations of continuous layers ranging 1-2 km to 1-6 km for the top of each hour (Top of the Hour

Forecasts). Maximum updraft helicity was calculated over the hour integrated over the 2-5 km layer (3 km depth).

The verification data set consisted of severe weather observations from the publication *Storm Data*. A radius of 40-km was searched around each observed report, and all grid points within that radius were considered as part of the "yes" observation. A 40-km radius was chosen so that our verification radius matched that of the SPC.

Thresholds were chosen for each field based on the range of values of each field. The threshold values used were integers that ranged from 0 (all grid points were yes forecasts) to the maximum integer value of the particular field.

Contingency tables were constructed by comparing grid points of the model forecast to the verification data set for each threshold of a given field. The values a, b, c, and d (see Table 1) were determined from model forecasts and events. Several measures of verification were then calculated:

Table 1. The 2x2 contingency table (Reproduced from Brooks 2004)

Forecast		Observed		
		Yes	No	Sum
	Yes	a	b	$a+b$
	No	c	d	$c+d$
	Sum	$a+c$	$b+d$	1

- Probability of Detection (POD) = $a/(a+c)$
- False Alarm Ratio (FAR) = $b/(a+b)$
- Probability of False Detection (POFD) = $b/(b+d)$
- Critical Success Index (CSI) = $a/(a+b+c)$
- Bias = $(a+b)/(a+c)$

Relative operating characteristic (ROC) curves (Mason 1982) plots POD against the POFD at different decision thresholds (Fig. 1). On a ROC diagram, a theoretically perfect forecast is the far upper left corner (POD=1 and FAR=0). This means that larger areas under the curve correspond to better forecasts. Rather than displaying a ROC diagram for every individual day during the period, the area under each ROC curve was calculated for each day of each field and the distributions were plotted in the form of a box and whisker diagram as a function of data. This allows a comparison of the overall performance of different layers over the entire time span.

Finally, Performance Diagrams (Roebber 2009) were also created. The Performance Diagram is a summary diagram for the results of contingency tables. CSI and bias were derived in terms of SR=1-

FAR, and POD (Roebber 2009) in equations that can be solved for POD.

$$CSI = 1/((1/SR) + (1/POD) -1) \qquad (4)$$

$$bias = POD/SR \qquad (5)$$

The values of POD can then be plotted against SR to show values of CSI and bias. This allows a user to quickly visualize the verification measures of various forecast fields in one diagram.

Using the 2x2 contingency table values, POD and SR were calculated for each of the fields. POD was then plotted against SR in a diagram constructed using the technique above.

On a Performance Diagram, a theoretically perfect forecast (POD=1, FAR=0) is located in the upper right hand corner of the graph or (1, 1). Here, there is no bias, and the CSI is at a maximum.

4. RESULTS

The model forecasts of UH were able to more accurately predict severe weather events as the depth of the integrated layer increased. As depths were increased, the median values of the area under the ROC curves increased (Fig. 2). Since the area under the ROC curve is an indication of performance, a larger ROC curve area represents a better model performance. The maximum UH field had the highest median area under the ROC curve and therefore the best performance. However, the spread, or variability, of the ROC areas increase as the depth of the integrations increase. It is also notable that the field of maximum UH has the largest variability. A smaller variability indicates greater consistency and more precision.

This improvement of the model forecast with depth can also be seen in a Performance Diagram (Fig. 3). Maximum UH (calculated over the 3-km depth of layers 2-5) clearly has a higher POD than each of the corresponding thresholds in the top of the hour forecast. At the same time, the false alarm ratio for the maximum over the hour forecast showed little difference from that of the top of the hour forecast. UH also exhibits a higher CSI when compared to the top of the hour forecast for the same depth (Fig. 4). In fact, Maximum UH performs significantly better than all of the top of the hour forecasts (Fig. 5)

5. DISCUSSION

The model forecast showed improvement as the depth of the integrated layer increased, but at the expense of increased variability. The maximum UH field generally had the highest success in forecasting events, but with a greater variability than all of the other fields examined. The comparison between the maximum UH field and the top of the hour forecasts for the same depth enhanced the result that top of the hour forecasts do not do as well as forecasts that are based off of the maximum value computed over the entire hour. For this reason, all users would tend to pick the maximum UH forecast over the top of the hour forecasts, due to the increase in POD at every threshold with little to no change in FAR.

Since the points on the Performance Diagrams were created based on a series of thresholds that were examined, it is clear that different thresholds have different CSI and bias. It is then possible for a user to pick a particular threshold that exhibits his or her desired CSI and bias. For example, due to the cost of a missed detection, many users would favor thresholds that favor POD, even if it means an increase in FAR. A user that has significant costs associated with false alarms that outweigh the cost of a missed detection would prefer a threshold that favors a low FAR.

Different users of the model output may also want to consider examining levels of performance and variability that are tailored to their own purpose. Even though the maximum UH showed the highest performance as a forecast parameter, its variability may prevent it from being the "best" field.

Given that the model forecasts show improvement as depth increases, and the current value of UH used in models is the maximum calculated over the 3-km depth of layers 2-5, it would be worthwhile to examine the maximum values over the 5-km depth of layers 1-6 as well as at other depths.

REFERENCES

Brooks, Harold, 2004: Tornado-Warning Performance in the Past and Future: A Perspective from Signal Detection Theory. *Bulletin of the American Meteorological Society,* **85**, 837-843.

Finley, John, 1884: Tornado Predictions. *American meteorological Journal,* **1,** 85-88.

Gilbert, G. K., 1884: Finley's tornado predictions. *American Meteorological Journal,* **1**, 166-172.

Kain, J. S., and Coauthors, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting*, **23**, 931–952.

Mason, I., 1982: A model for assessment of weather forecasts. *Australian Meteorological Magazine.*, **30,** 291-303.

Murphy, A. H., 1996: The Finley Affair: A signal Event in the History of Forecast Verification. *Weather and Forecasting,* **11,** 3-20.

Roebber, P. J., 2009: Visualizing Multiple Measures of Forecast Quality. *Weather and Forecasting,* **24**, 601-608.

Sobash, R., D. R. Bright, A. R. Dean, J. S. Kain, M. Coniglio, S. J. Weiss, and J. J. Levit, 2008: Severe storm forecast guidance based on explicit identification of convective phenomena in WRF-model forecasts. *Preprints*, 24th Conf. on Severe Local Storms, Savannah, GA, Amer. Meteor. Soc., 11.3. [Available online at http://ams.confex.com/ams/pdfpapers/142187.pdf.]
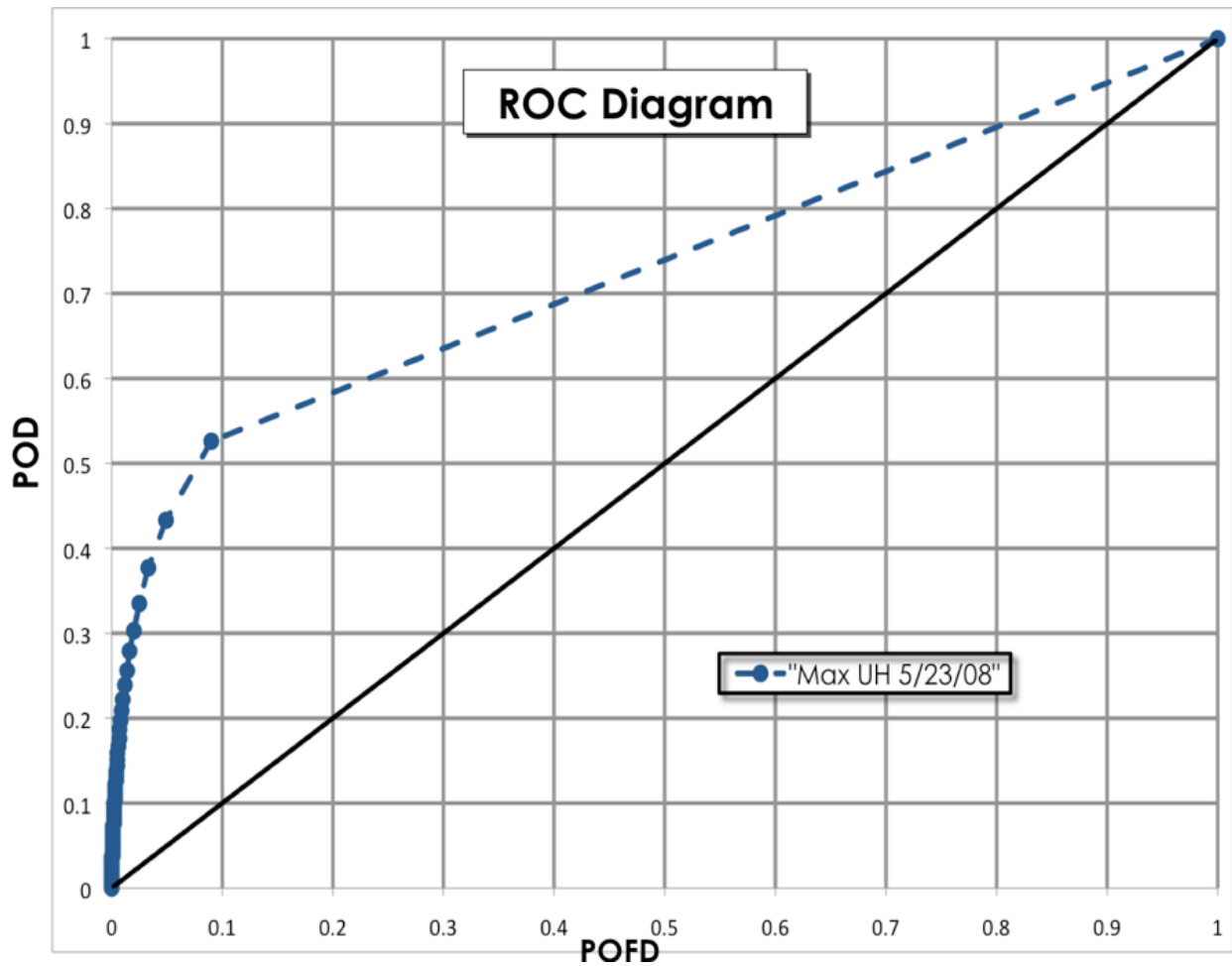
Figure 1. ROC diagram for Maximum UH on May 23, 2008. POD and POFD were calculated based on contingency table values and plotted as POD over POFD. Each individual point represents a different threshold. In this figure, thresholds are integers that range from 0 – 509. The solid black line represents the line of no skill. Points that lie above the line should have some indication of skill.
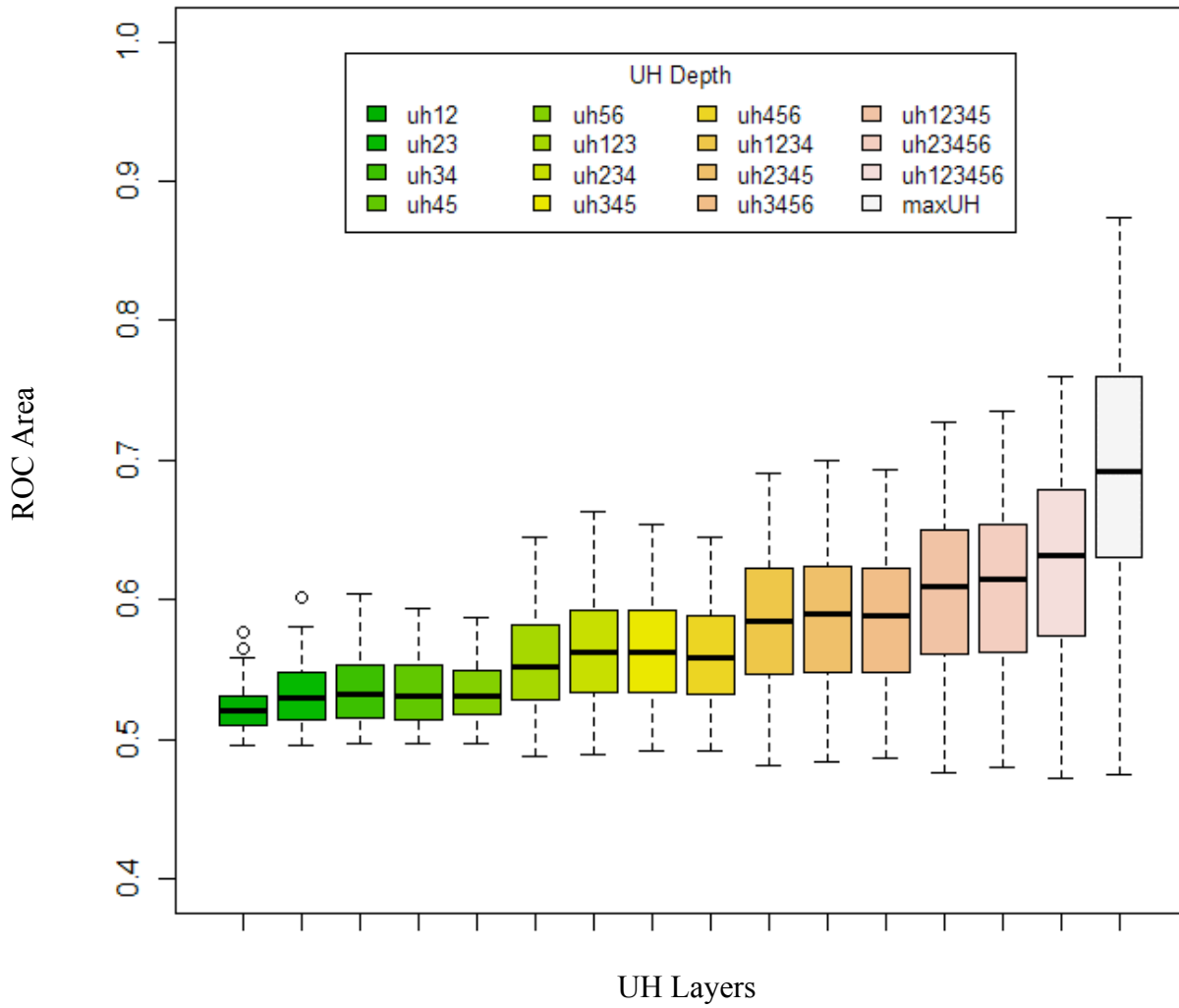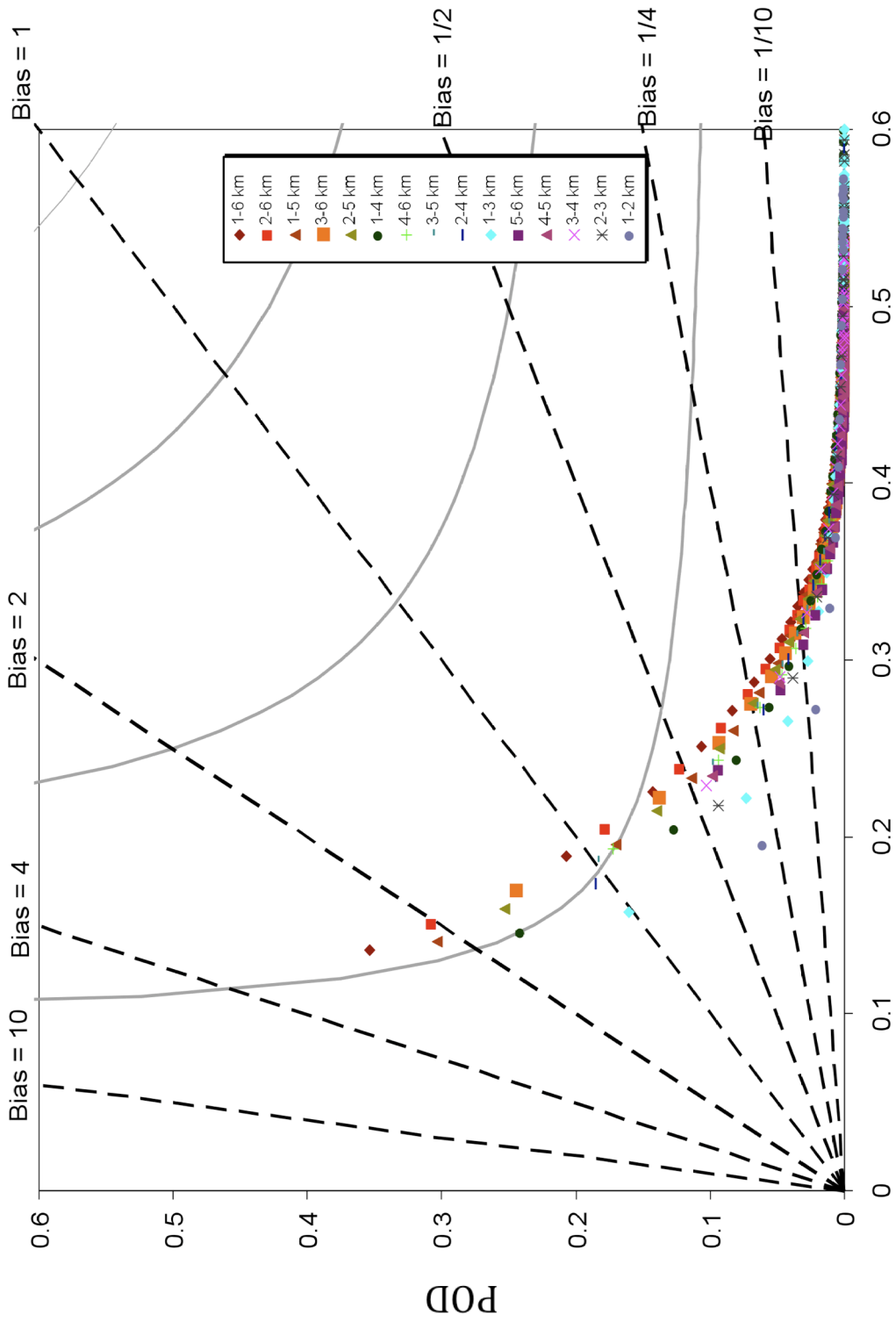
Figure 2. Box and Whisker plot of the distribution of ROC areas for each field.

Figure 3. Performance Diagram: Comparison of Layers. A Performance Diagram plots the POD as a function of SR (1-FAR). Lines of CSI are represented by the solid grey lines, while bias is the dashed black line. In this diagram, the top of the hour forecasts are plotted to show the trend in forecast performance as the depth increased.
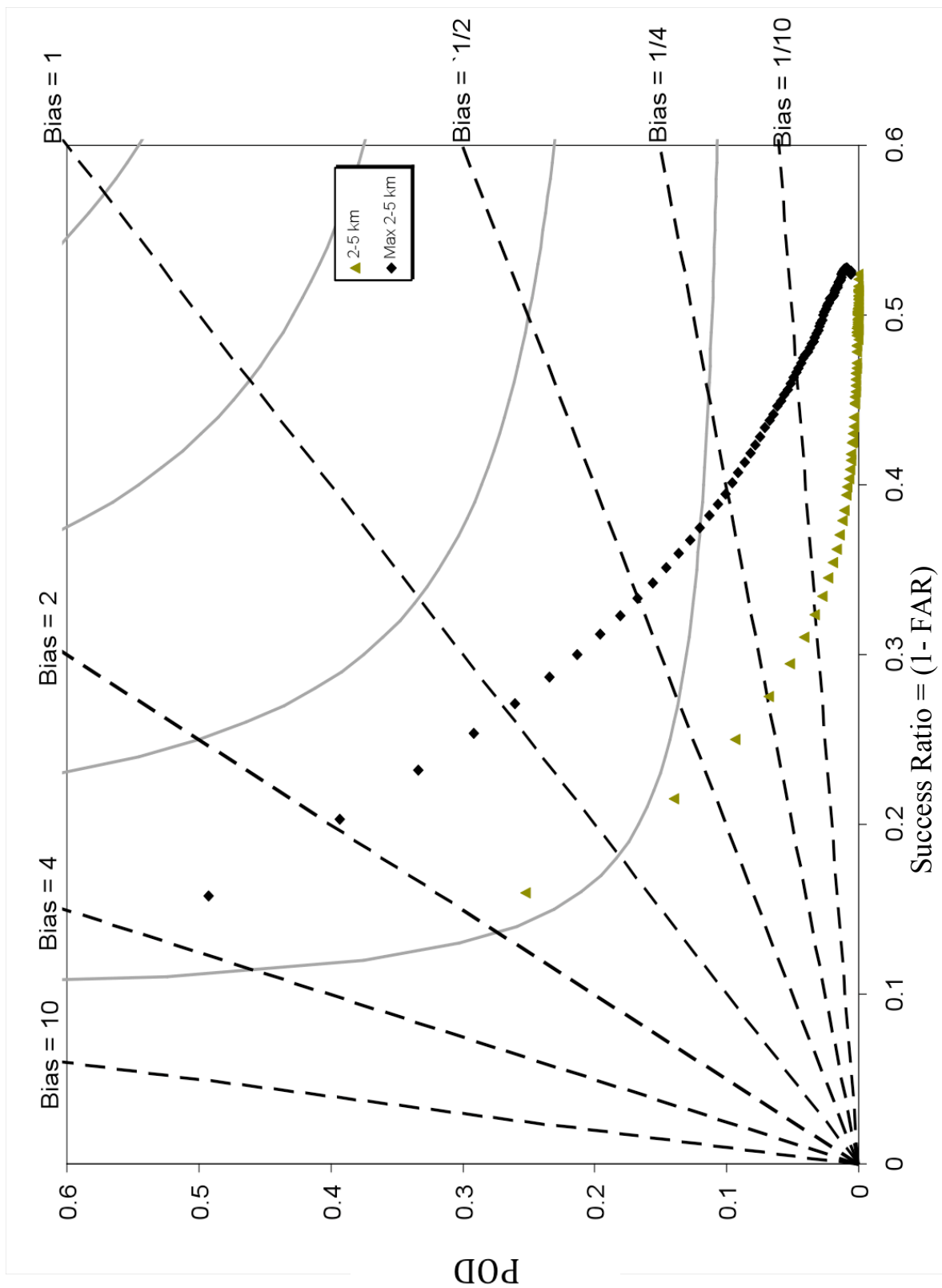
Figure 4. Performance Diagram: Max Compared to Top of the Hour Forecast. Same as in figure 3, but here only maximum UH and the top of the hour forecasts of the same depth (2-5 km) are plotted.
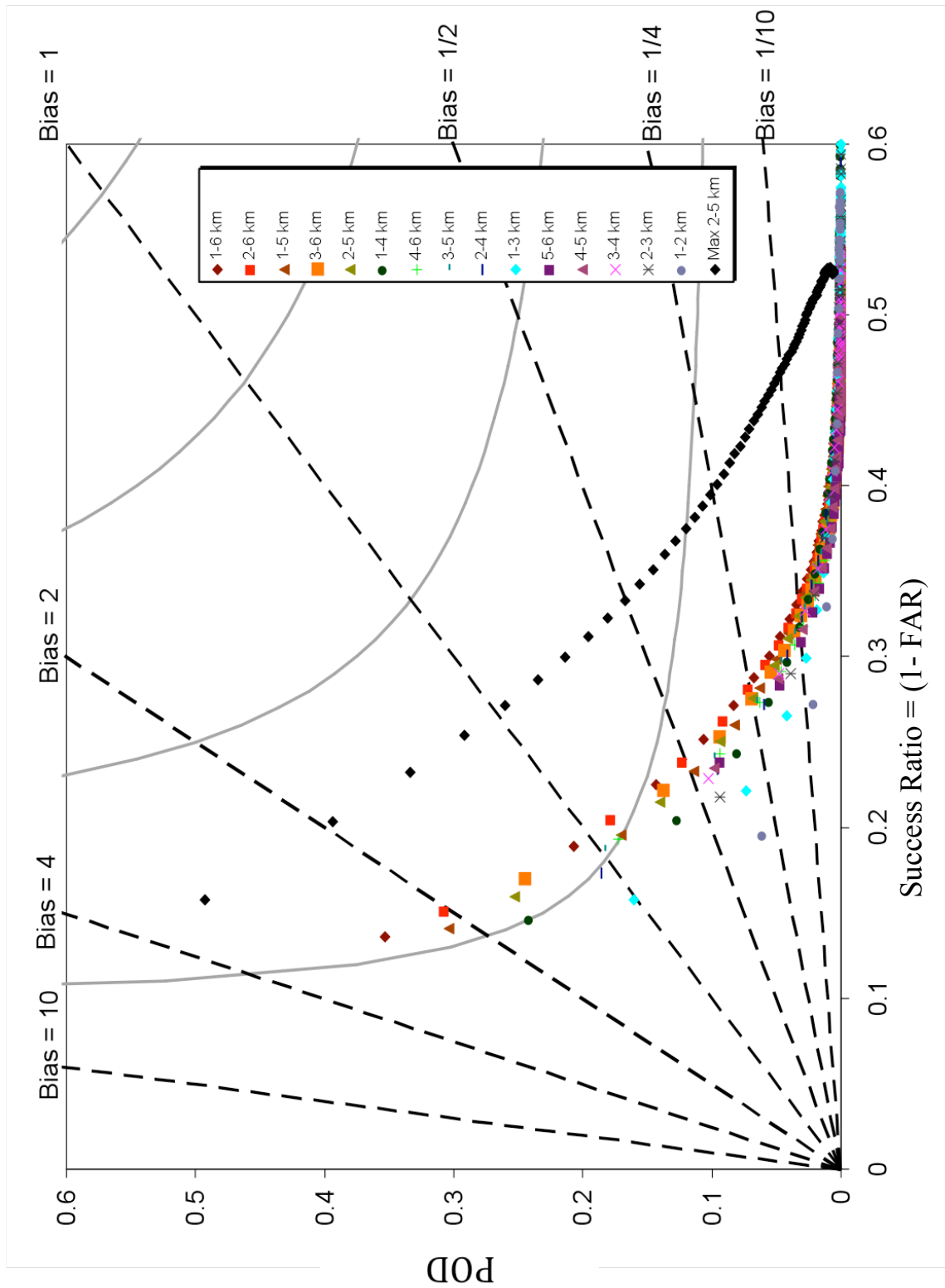
Figure 5. Performance Diagram: All Fields. Same as figures 3 and 4, but this diagram compares all top of the hour forecasts to the maximum over the hour forecast.