

## VERIFICATION OF VORTEX-94 FORECASTS

Charles A. Doswell III\*, Robert Duncomb#, Harold E. Brooks\*, Frederick H. Carr#

\* NOAA/ERL/National Severe Storms Laboratory Norman, Oklahoma

# School of Meteorology, University of Oklahoma Norman, Oklahoma

### 1. INTRODUCTION

During the springs of 1994 and 1995, the field data collection phase of a project to study tornadoes and tornadic storms was carried out. This project is described in detail in Rasmussen et al (1994); in that report it is noted that project involved forecasting in two way. First, a forecast was needed in support of the field data collection efforts. Second, some aspects of the forecasting were, in effect, forecasting experiments that are quasi-independent of the project itself. The experimental forecasts were extensions of earlier work (Doswell and Flueck 1986; Jincal et al. 1992) and follow the basic ideas discussed in Doswell et al. (1986).

Operating under the assumptions laid down in these earlier works, we are assuming that verification of forecasts is a critical component in going through a forecasting exercise. The VORTEX experimental forecasts are designed not only how to continue previous efforts to see how well we can forecast various aspects of convective weather phenomena, but also to explore how to issue

probabilistic forecasts for severe convective events. Apart from the pioneering work of Murphy and Winkler (1982), severe weather forecasts in operations are done categorically (or, more properly, dichotomously) instead of probabilistically (or polychotomously). As part of the modernization of the National Weather Service (NWS), it is likely that forecasts issued by the Storm Prediction Center (SPC) in the future will be taking on a probabilistic format. Given a lack of experience with probabilistic forecasting in this context, we felt it was important to attempt probabilistic forecasting. The question naturally arises as to how that verification is to be done. Murphy and Winkler (1987) have developed a framework for verification, based on the joint distribution between forecasts and observations (i.e., the contingency table). Their method is "distributions-oriented" rather than the more traditional "measures-oriented" schemes. Thus, another aspect of the VORTEX verification effort is to compare and contrast distributions-oriented and measures-oriented verification schemes.

In what follows, part 2 will address briefly the forecasts and the data used to verify the forecasts. Part 3 will describe, again briefly, the verification methodology. Selected results will be presented in Part 4, and Part 5 will be devoted to a

---

*Corresponding author address:* Dr. Charles A. Doswell III, National Severe Storms Lab., 1313 Halley Circle, Norman, OK 73069 Internet e-mail: doswell@nssl.uoknor.edu [*no longer valid*]

discussion of these results and their implications for operational forecasting.

## 2. FORECASTS AND DATA

At the time of this study's beginning, only the VORTEX-94 data were available, so this paper is not going to cover the 1995 results. VORTEX forecasts involved many different components, some of which will not be discussed here. The relevant forecasts took two primary forms: area forecasts and gridded contour forecasts. The area forecasts were designed to answer the question of "if" an event was going to occur, and the area forecasts were aimed at answering the question "where" that event was most likely, in the judgment of the forecaster. All probabilities were determined subjectively.

### 2.1 Area forecasts

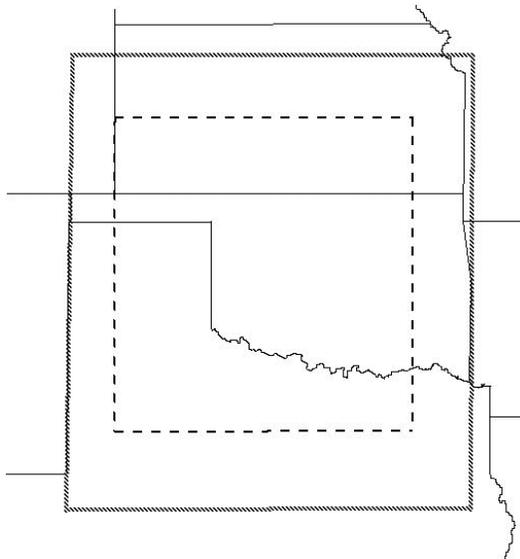


Figure 1. VORTEX forecast area (outer hatched line), as well as the nominal VORTEX operations area (stippled area outlined by dashed line).

The VORTEX forecasting area is shown in Fig. 1. Two types of area forecasts were made: one for the day of the forecast (Day-1) and one for the next day

(Day-2). Day-1 forecasts were valid from 1400 UTC to 0400 UTC the next day, whereas Day-2 forecast valid times were from 1100 UTC the next day to 0400 UTC the day after. For each of these time periods, forecasts were made for the probability of one or more: 1) cloud-to-ground (CG) lightning flashes (denoted "L"), 2) severe storms (denoted "S"), 3) tornadoes (denoted "T"), and 4) "targetable" storms (denoted "TS) within the VORTEX forecast area. A "targetable" storm is defined as one which a) produces a tornado, b) prompts an NWS-issued tornado warning, or c) is in fact targeted by the VORTEX field experiment. Thus, the occurrence of one or more of these events within the space-time volume in which the forecasts are valid constitutes a "hit" for that forecast; if no such occurrence is observed, this is a "non-event" for that forecast.

### 2.2 Contour forecasts

The other forecast type consisted of probability contours. These contours defined regions, inside which the probability of an event was considered to be constant and equal to the contour value. These contours were hand-drawn by the forecasters on paper forecast forms, but for verification purposes, the probabilities were assigned to grid points after the experiment was over. The "grain size" associated with the contours was the Manually-Digitized Radar (MDR) grid (see Fig. 2); each MDR box in the VORTEX forecast area was assigned a value based on the converted probability contours.

These forecasts were only issued for Day-1, and covered the same valid time as the area forecasts. There were four types of contour forecasts: 1) CG light-

ning, 2) targetable storms, 3) tornadoes, given that there was a tornado somewhere within the VORTEX forecast area (called "tornado given a tornado," or TGT), and 4) tornadoes, given that a CG lightning flash occurred in that MDR box (called "tornado given lightning," or TGL). The latter two are obviously conditional probabilities.

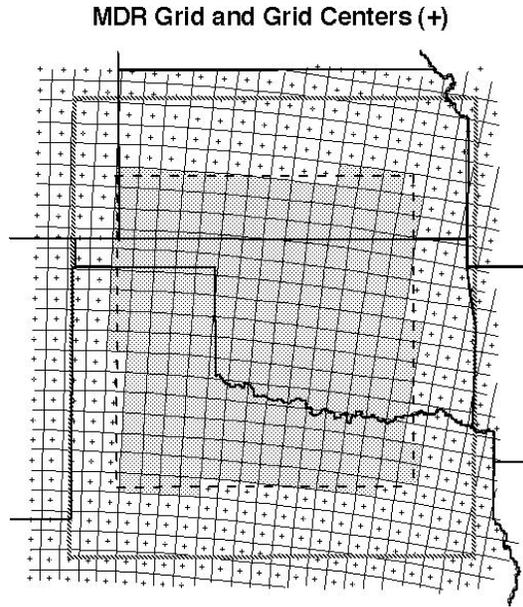


Figure 2. Manually Digitized Radar (MDR) grid box centers (+ signs) superimposed on the VORTEX forecast area (enclosed by the hatched line). MDR boxes considered to be in the VORTEX ops area are stippled..

### 2.3 Verifying data

Verifying data for the experiment consisted of 1) CG lightning strike data from the National Lightning Detection Network, 2) the SPC (formerly National Severe Storms Forecast Center) log of severe thunderstorm events, 3) Storm Data reports of tornadoes, 4) NWS tornado warnings. Also included were the observations of the VORTEX field teams.

## 3. VERIFICATION METHODOLOGY

<i>Probability (%)</i>	<i>H</i>	<i>N</i>	<i>#fcsts</i>
0	0	2	2
2	0	2	2
5	0	3	3
10	2	2	4
20	3	0	3
30	1	0	1
40	3	1	4
50	7	1	8
60	2	0	2
70	4	1	5
80	6	0	6
90	2	0	2
95	7	0	7
98	2	0	2
100	25	0	25
<b>Totals</b>	<b>64</b>	<b>12</b>	<b>76</b>

Table 1. Contingency table for Day-1 Lightning (L) area forecasts; "H" denotes hits and "N" denotes misses. Numbers in the table correspond to VORTEX area forecast days.

<i>Probability (%)</i>	<i>H</i>	<i>N</i>	<i>#fcsts</i>
0	407	9042	9949
1	264	4730	4994
10	563	3387	3950
20	1019	3134	4153
40	1169	2195	3364
60	1091	1694	2785
80	551	646	1197
90	838	432	1270
99	660	174	834
<b>Totals</b>	<b>6562</b>	<b>25434</b>	<b>31996</b>

Table 2. As in Table 1, except for L-contour forecasts. Numbers in the table correspond to MDR box forecast days. In the VORTEX forecast area there were 421 MDR boxes and there were 76 forecast days, for a total of 31,996 MDR box days during the experiment.

In order to compare the measures-oriented approach with a distributions-oriented scheme, we basically did both types of verification, comparable to what

has been done by Brooks and Doswell (1996). Space does not permit a very detailed description of the methodology; consult the references for more details.

We calculated some typical summary measures (e.g., Brier scores, POD, a skill score based on the sample climatology, etc.), as well as some atypical ones (e.g., a reliability measure, a discrimination measure, etc.) For both the area and the contour forecasts, contingency tables were developed (see e.g., Doswell et al. 1990) and the verification proceeds from the numbers in the tables. A sample contingency table for the area forecasts (for CG lightning) is shown in Table 1.

For the contour forecasts, the tables are somewhat smaller (Table 2), since there were fewer forecast probability categories. Table 3 is an example of the contingency table for the conditional probabilities used in the contour forecasts.

<i>Probability (%)</i>	<i>H</i>	<i>N</i>	<i>#fcsts</i>
0	16	2235	2251
1	8	2005	2013
10	16	1068	1084
20	20	689	709
40	24	443	467
60	10	28	38
80	0	0	0
90	0	0	0
99	0	0	0
Totals	94	648	6562

Table 3. As in Table 2, except for TGL-contour forecasts. During the VORTEX forecast area there were 6562 MDR boxes that met the condition of having CG lightning in them.

#### 4. RESULTS

Space simply does not permit a very extensive presentation of the results. A

sample of the summary measures is given in Tables 4 and 5.

<i>Score</i>	<i>L</i>	<i>S</i>	<i>T</i>	<i>TS</i>
Hit freq.	.84	.65	.36	.42
Avg fcst	.68	.45	.19	.42
Bias	-.16	-.20	-.16	-.18
Brier Score	.12	.18	.21	.22
Skill Score	.13	.20	.09	.12

Table 4. Area forecast summary measures.

<i>Score</i>	<i>L</i>	<i>TS</i>	<i>TGT</i>	<i>TGL</i>
Hit freq.	.21	.007	.008	.014
Avg fcst.	.23	.04	.10	.07
Bias	.02	.03	.09	.07
Brier Score	.14	.01	.02	.01
Skill Score	.14	--	--	--

Table 5. Contour forecast summary measures.

The area forecasts exhibit underforecasting (a negative bias); forecasters underestimated the frequency of events within the VORTEX domain. The Brier scores (basically, the mean square error of the forecasts) indicate that lightning forecasting is best, with the best skill over the sample climatology exhibited by the severe storm forecasts. On the other hand, the contour forecasts show persistent overforecasting. The Brier scores seem to suggest that the contour forecasts are better than the area forecasts, but this is an illusion resulting from the relative rarity of events within MDR boxes. Even on a relatively active day, only a few MDR boxes are affected, so the scores are dominated by the large number of correct forecasts of non-events.

The distributions-oriented approach is exemplified here by two figures, both for the area forecasts (space does not permit a more extensive presentation). The first

(Fig. 3) shows the so-called Reliability Diagram for Day-1 lightning area forecasts. The Reliability Diagram is relatively well-known. Underforecasting can be seen clearly in the forecast probability range from about 0.2 to 0.6. Note, however, that there are relatively few forecasts in this range.

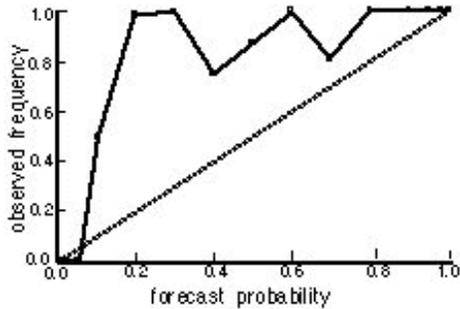


Figure 3. Reliability diagram, area lightning forecasts; hatched line is the theoretical "perfect" reliability line, where observed frequency equals forecast probability.

Another diagram, much less well-known than the Reliability Diagram is the so-called Discrimination Diagram. Again we show the result for Day-1 Lightning (Fig. 4).

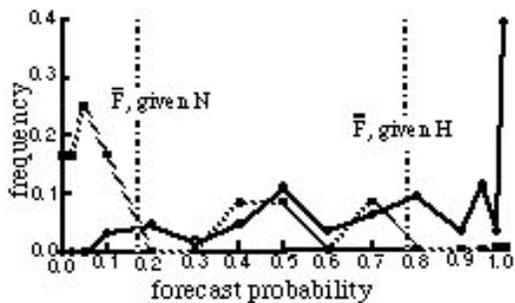


Figure 4. Discrimination Diagram, area lightning forecasts; solid line is the distribution of forecasts given H, the hatched line is the forecast distribution given N, and vertical hatched lines are the two mean forecasts, given N and H.

In this diagram, what is shown is the distribution of the forecasts, given the events (H or N). Quality forecasts

should have a relatively high mean forecast given H, whereas the mean forecast given N should be relatively low. Moreover, the distribution of forecasts given H should show increasing frequency as the forecast probability increases, whereas the frequency should decrease as the forecast probability increases for forecasts given N. Figure 5 does in fact exhibit these characteristics, albeit with considerable noise.

## 5. DISCUSSION

Again, space does not permit an extensive presentation. The outcome of the VORTEX-94 verification suggests overall forecast quality comparable to other, similar experiments we have referenced. However, the verification has revealed a number of flaws in the design of the VORTEX forecasting experiment:

- 1) the MDR box is too small a "grain size" given the uncertainties in forecasting for convection,
- 2) there seem to have been too many forecast probability categories and the distribution of the forecast probability categories is going to require further thought,
- 3) forecasters need to be aware of the climatological frequencies of the events they are forecasting,
- 4) feedback of the verification results to the forecasters needs to be immediate, and
- 5) training in probability forecasting is important to being able to interpret the feedback.

Interested forecasters should consult Murphy (1993) for an insightful discussion about how to evaluate forecasts.

Although the abbreviated results given here cannot be considered compelling evidence, we have found the distributions-oriented verification approach to be far more revealing of the strengths and weaknesses in the forecasts, as Murphy's (1991) presentation suggests. It is our belief that operational verification should be done in this mode, rather than the traditional measures.

In the future, we intend to conduct a similar study of the VORTEX-95 forecasts and to merge the results. Since 1995 had considerably more activity than 1994, it is possible that some of the problems with our study, related to small samples in some event categories, can be alleviated.

*Acknowledgments:* We appreciate the efforts of Charlie Crisp (NSSL) and Philip Bothwell (SPC) in assisting with the preparation of the verification data. The VORTEX forecasters contributed unselfishly of their time to participate in this experiment, and only their numbers preclude naming them all. We also are grateful to their respective agencies for allowing them to participate in the experiment.

## 6. REFERENCES

Brooks, H. E., and C. A. Doswell III, 1996: A comparison of measures-oriented and distributions-oriented approaches to forecast verification. *Wea. Forecasting*, **11**, 288-303.

Doswell, C. A. III, R. A. Maddox and C. F. Chappell, 1986: Fundamental con-

siderations in forecasting for field experiments. Preprints, *11th Conf. Wea. Forecasting and Analysis* (Kansas City, MO), Amer. Meteor. Soc., 353-358.

\_\_\_\_\_, and J. A. Flueck, 1989: Forecasting and verifying in a field research project: DOPLIGHT '87. *Wea. Forecasting*, **4**, 97-109.

\_\_\_\_\_, R. Davies-Jones, and D. Keller, 1990: On summary measures of skill in rare event forecasting based on contingency tables. *Wea. Forecasting*, **5**, 576-585.

Jincai, D., C. A. Doswell III, D. W. Burgess, M. P. Foster and M. L. Branick, 1992: Verification of mesoscale forecasting made during MAP '88 and MAP '89. *Wea. Forecasting*, **7**, 468-479.

Murphy, A. H., 1991: Forecast verification: Its complexity and dimensionality. *Mon. Wea. Rev.*, **119**, 1590-1601.

\_\_\_\_\_, 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **11**, 3-20.

\_\_\_\_\_, and R.L. Winkler, 1987: A general framework for forecast verification. *Mon. Wea. Rev.*, **115**, 1330-1338.

Rasmussen, E.N., J.M. Straka, R. Davies-Jones, C.A. Doswell III, F.H. Carr, M.D. Eilts and D.R. MacGorman, 1994: Verification of the origins of rotation in tornadoes experiment: VORTEX. *Bull. Amer. Meteor. Soc.*, **75**, 995-1006.