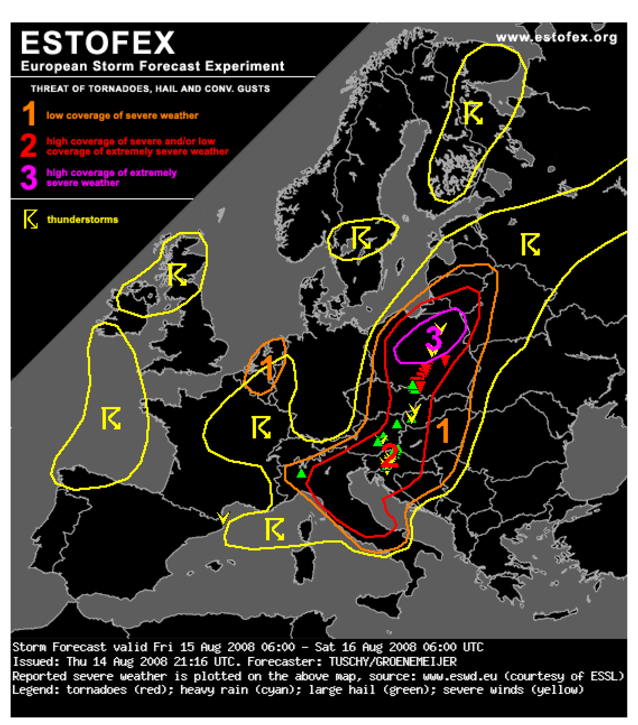
Evaluation of ESTOFEX forecasts: Severe thunderstorm forecasts

H. E. Brooks[1], T. E. Thompson[2], C. M. Shafer[2], C. S. Schwartz[2],
P. T. Marsh[2], A. Kolodziej[2], N. Dahl[2], and D. Buckey[2]
[1] NOAA/National Severe Storms Laboratory, Norman OK
[2] School of Meteorology, University of Oklahoma, Norman, OK

What is ESTOFEX?

The European Storm Forecast Experiment (ESTOFEX) is an initiative of a team of European meteorologists, meteorology students and trained enthusiasts, who intend to learn how to forecast severe convective storms in Europe. The forecasts resemble the categorical Storm Prediction Center forecasts, with a general thunder line and three levels of threat. Their goals are to issue daily forecasts of convective weather in Europe, improve understanding of European severe weather, use the new European Severe Weather Database (ESWD) (Dotzek et al. 2009), and verify their forecasts using lightning data and ESWD reports. See http://www.estofex.org/ for more information and their daily forecasts.

ESTOFEX Forecasts and Data



An example of an ESTOFEX forecast is shown at left. The forecast period covers 24 hours, starting at 06 UTC, and the forecast is usually created the evening before. The yellow contour indicates general thunder and the other contours are level 1 through level 3 threats. The colored symbols indicate severe weather occurrences from the ESWD (red-tornado, green-hail, winds-yellow). This forecast has a level 3, which did not occur in our period of analysis. This forecast is for a day in August when an F3 tornado occurred in southern Poland and a total of 8 people were killed with the severe weather.

The ESTOFEX forecasters provided us with 553 days of forecasts from late April 2006 to December 2007 made by 5 different forecasters. On occasion, updates to the original forecast were issued. We chose to use only the first outlook. Also, in a small number of cases, two forecasters' names were listed on the forecast. We gave credit to the first forecaster listed. The forecasts were put onto a 0.5x0.5 lat/lon grid and reports from the ESWD were put on the same grid.

Big Questions

1. Can we qualitatively say anything about the forecasts?
2. Are the forecasters different or do they act like a unit?
3. Can we see changes during the period of the forecasts?

ACKNOWLEDGMENTS

The authors would like to thank the ESTOFEX forecasters for providing the data and background information. The initial analysis was done as part of a graduate course at the University of Oklahoma School of Meteorology in the spring of 2008.

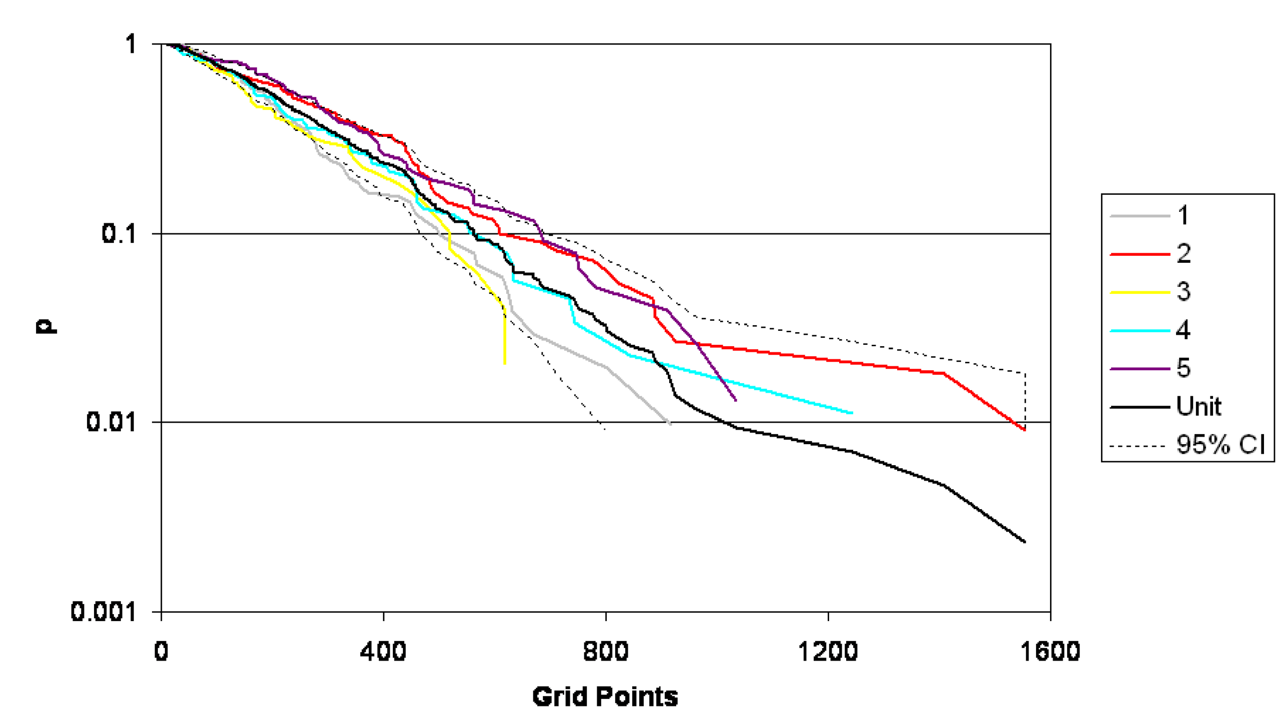
References:

Dotzek, N., P. Groenemeijer, B. Feuerstein, and A. M. Holzer, 2009: Overview of ESSL's severe convective storms research using the European Severe Weather Database ESWD. *Atmos. Res.*, in press.

Roebber P. J., 2009: Visualizing Multiple Measures of Forecast Quality. *Wea. Forecasting*, in press.

Forecast Areas

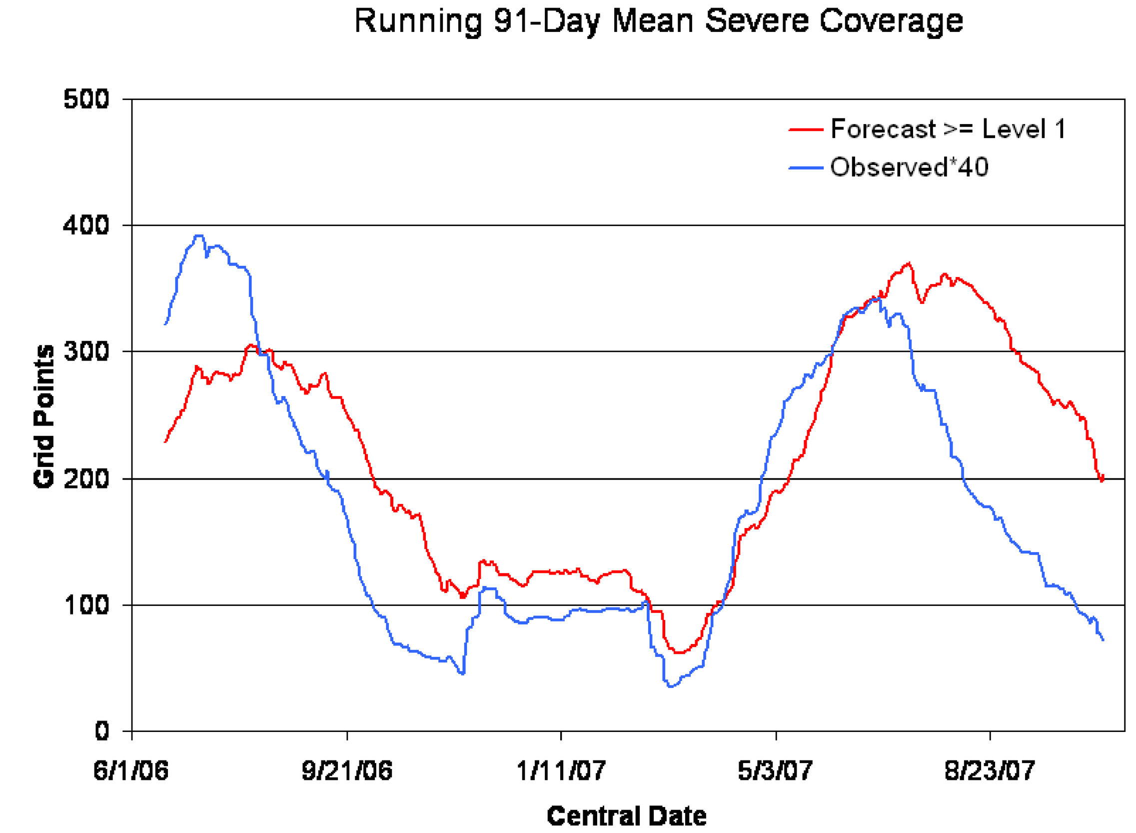
Probability of Exceedance of Forecast Areas From Forecasters
(Grid Points)



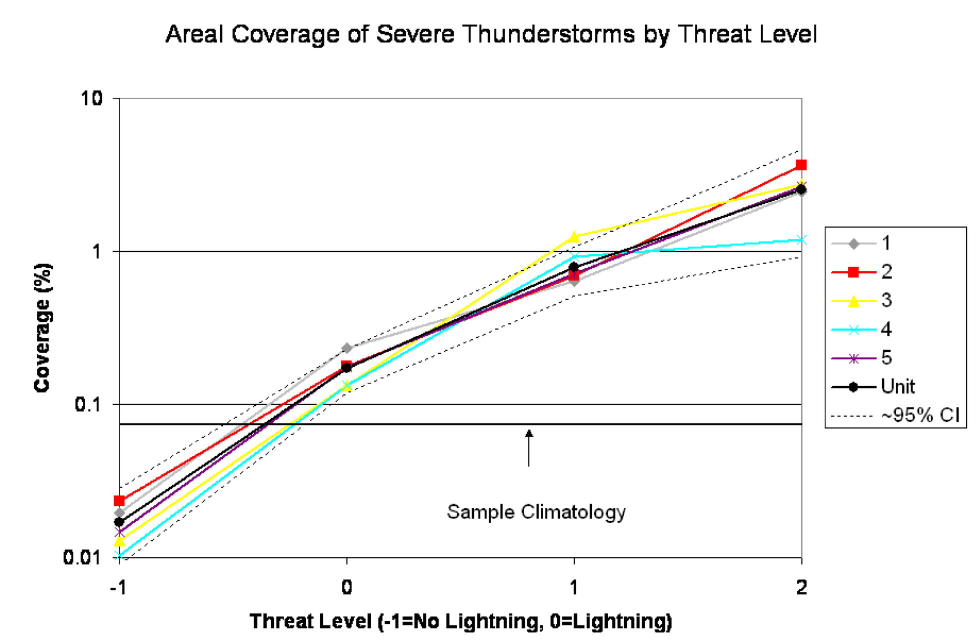
The figure above shows the distribution of forecast areas of at least level 1 threat from the five forecasters, as well as the unit average. The dashed lines are an estimate of the 95% confidence interval (CI) on the unit distribution, assuming that the forecasters worked the same distribution of days. As a result, it is likely an underestimate of the true variability. The distribution for each forecaster falls in or near this distribution, offering support to the notion that the forecasters act much more like a unit than separate forecasters.

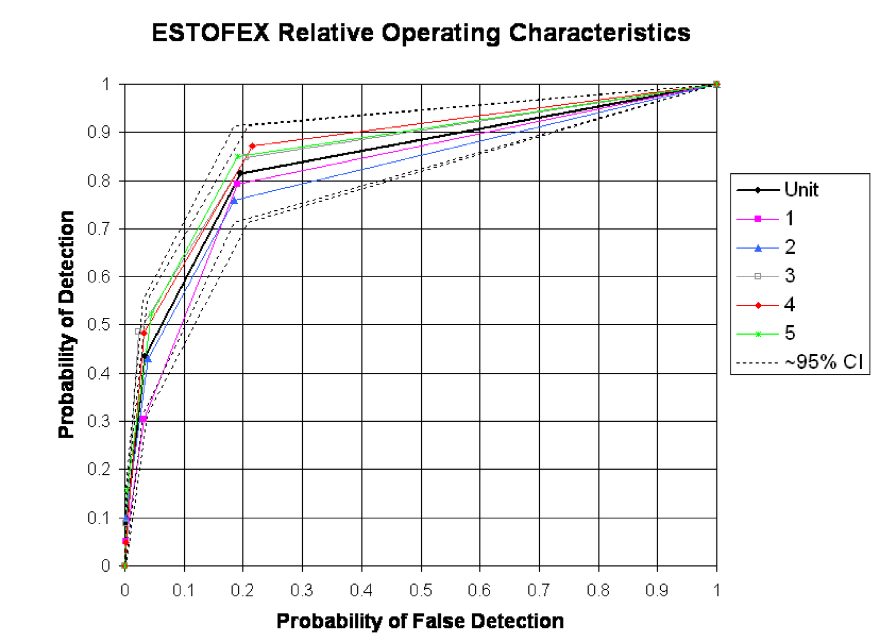
Below, we see a 91-day (roughly providing a continuous look at the season) running mean of forecast and observed severe thunderstorm coverage (observed area multiplied by 40). The big signal is the annual cycle, with greater coverage in the warm season. Note that the two warm seasons have similar observed coverage, but that the 2007 warm season forecast areas were larger, leading to a larger forecast bias. Also, note that the large forecast areas persist later in the season than the observed area. This is qualitatively similar to what is seen from the SPC forecasts in the US where a large area of scattered severe reports are frequently seen in late summer.

Running 91-Day Mean Severe Coverage



Forecast Performance

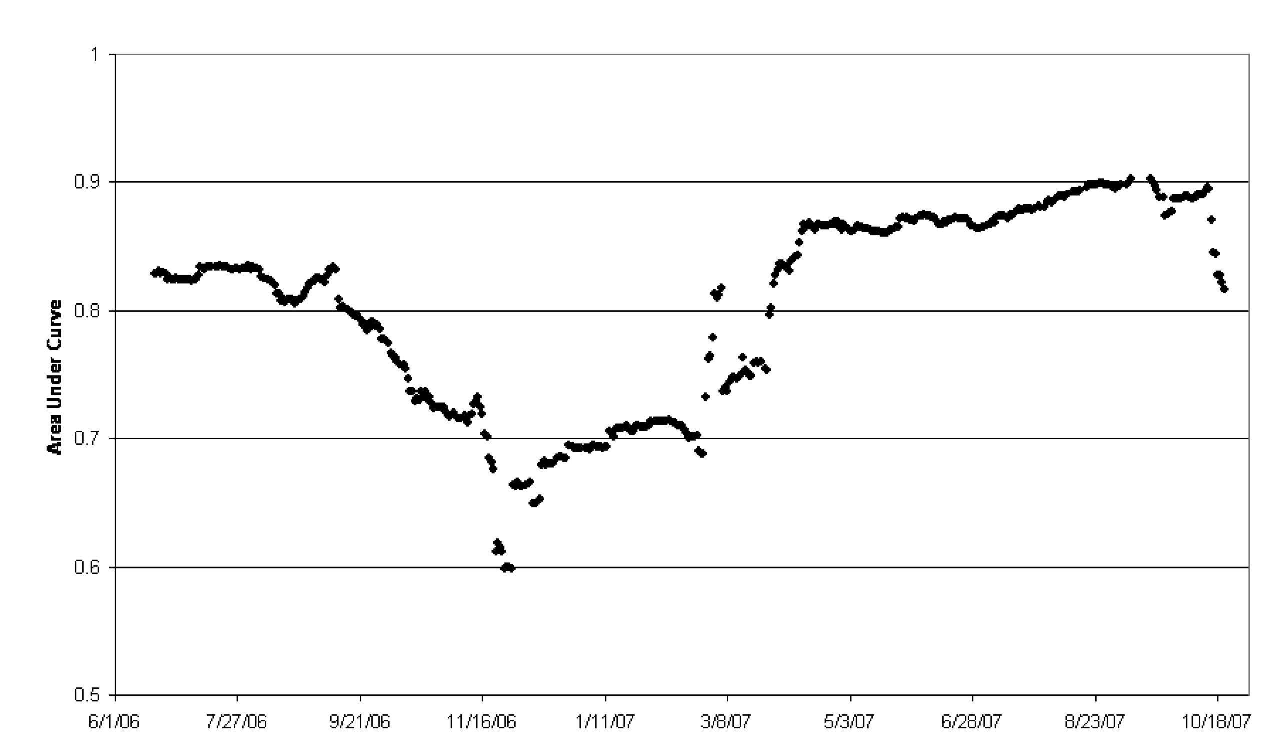




The figures above forecast performance from the individual forecasters and unit as a whole. The left panel shows fractional coverage of severe weather by threat area (no lightning=-1). Fractional coverage increases by an order of magnitude from general thunder up to level 2 threat, with level 2 associated with 40 times the sample climatology. The right panel shows relative operating characteristic curves associated with the forecasters. The area under the unit curve is 0.84, equivalent to the difference between two hypothetical normal distributions separated by 1.4 standard deviations. The dashed lines represent conservative estimates of the CI. (There are two sets of lines on the ROC because of different approaches to finding the CI.)

Below, we see a 91-day running mean of the area under the curve. Forecast performance is much worse in the cold season when coverage of severe storms is less. The second warm season is much higher than the first, an encouraging sign for improvements.

Area Under ROC Curve
(Running 91-day Mean)



Results

1. Forecasts are of a reasonably high quality
2. Little difference in performance between forecasters
3. Forecasts were better in 2007 than in 2006