

Moving from Categorical To Probabilistic Verification Measures At CPC

by Ed O’Lenic, Wesley Ebisuzaki and Scott Handel

NOAA-NWS-Climate Prediction Center

5200 Auth Road

Camp Springs, Maryland

1. Introduction.

Climate Prediction Center’s forecasts have been verified using the Heidke skill score, a categorical score, for decades. This paper describes a new, automated verification system which incorporates probabilistic verification measures, along with the traditional categorical measure.

2. The Verification System

The verification system includes the following components:

- probabilistic verification measures, including ranked probability skill score, rpss, relatively operating characteristics, roc, reliability diagrams (Wilks, 1990),
- makes better use of categorical skill information, including Heidke skill scores and contingency tables,
- unifies all time scales,
- automated,
- allows diagnosis of individual and bulk properties of official forecasts and forecast tools.

3. Example: 6-10 Day Forecasts

Forecasters need continuous feedback on the performance of forecast tools and their own performance so they can adjust their technique and their use of forecast tools. Figure 1 shows a typical example of a continuous record of forecast skill for several tools over a period of about 2 months. “Official” refers to the official forecast produced by the forecaster. “Auto Blend” is a fully automated forecast tool which uses a pre-selected, fixed set of inputs. “Manual Blend” refers to a forecast tool whose inputs are chosen by the forecaster.

Figure 2 shows a reliability diagram for the above category for the forecast made on 27 May, 2002. Perfectly reliable forecasts would have a black curve on top of the diagonal in the left hand graphs.. These graphs indicate that low probabilities are over predicted and high probabilities are under predicted for below and above.

Figure 3 shows the ranked probability skill score (red line) for 6-10 day temperature forecasts from January through mid-June, 2002. The blue line is the score which would have been obtained had the categorical forecast been perfect.

4. Example: Seasonal Forecasts

Skill feedback is even more essential in seasonal. Figure 4 shows a contingency table for canonical correlation analysis (CCA) forecast of U.S. surface temperature for December-January-February 1997-98. Among the useful data in this table are that 84 percent of the observations were above normal, and none were below, while the CCA forecast was split 51% to 46% between below and above. The CCA had a Heidke skill

score of only 15 for this forecast - not a terrible score, but far below what the official forecast achieved.

Figure 5 shows the RPSS for CPC official ½ month lead seasonal temperature forecasts from 1995 to 2002. As in Figure 4, the red line is the RPSS, while the blue line is the RPSS the forecasts would have achieved if the categorical forecasts had been perfect. This is a measure of the maximum achievable RPSS, given the probability forecast that was made.

Figure 6 uses the same conventions as Figure 2. This seasonal temperature forecast over-predicted the below category and under-predicted the above category. The above category is predicted with greater reliability than was the case for the 6-10 day forecast in Figure 2.

5. Summary and Future Plans

CPC's new verification system allows forecasters to adjust their use of forecast tools to suit the skill of the forecast tools, to monitor the performance of individual forecasters and to critically evaluate the performance of forecast tools to facilitate improvements to the tools.

Future plans include

1. Pooling forecast statistics by week, season and month,
2. Developing an annual cycle of forecast skill,
3. Compile statistics for 8-14 day forecasts,
4. Relate forecast statistics to modes such as ENSO, MJO, NAO, etc...

Reference:

Wilks, D.S., 1995. *Statistical Methods in the Atmospheric Sciences: an Introduction*, International Geophysics Series, **Vol. 59**, Academic Press, 464 pp.

D+8 Heidke Skill Scores (Temp.)
Monthly Time Series

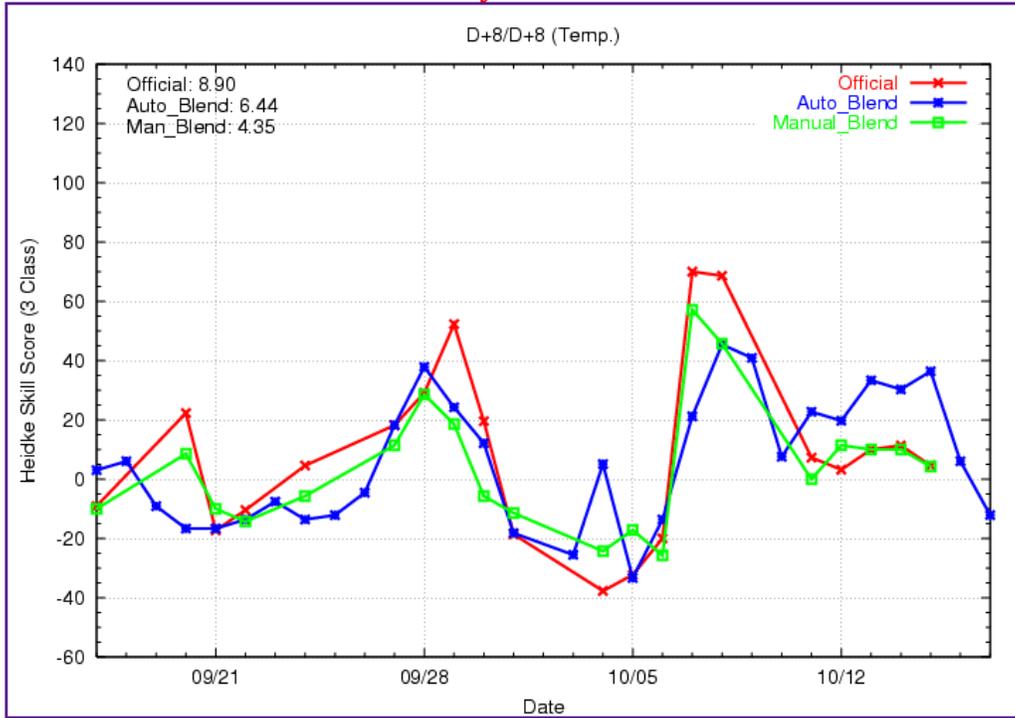


Figure 1. Time series of Heidke skill scores for 3 forecast methods.

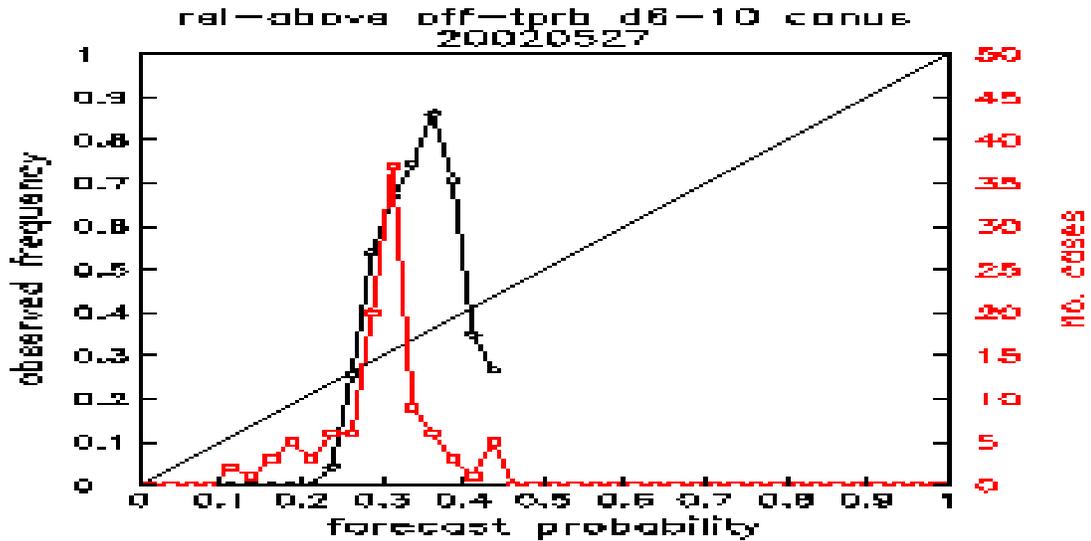


Figure 2. Reliability diagram for the above normal category for the 6-10 day temperature forecast made May 27, 2002.

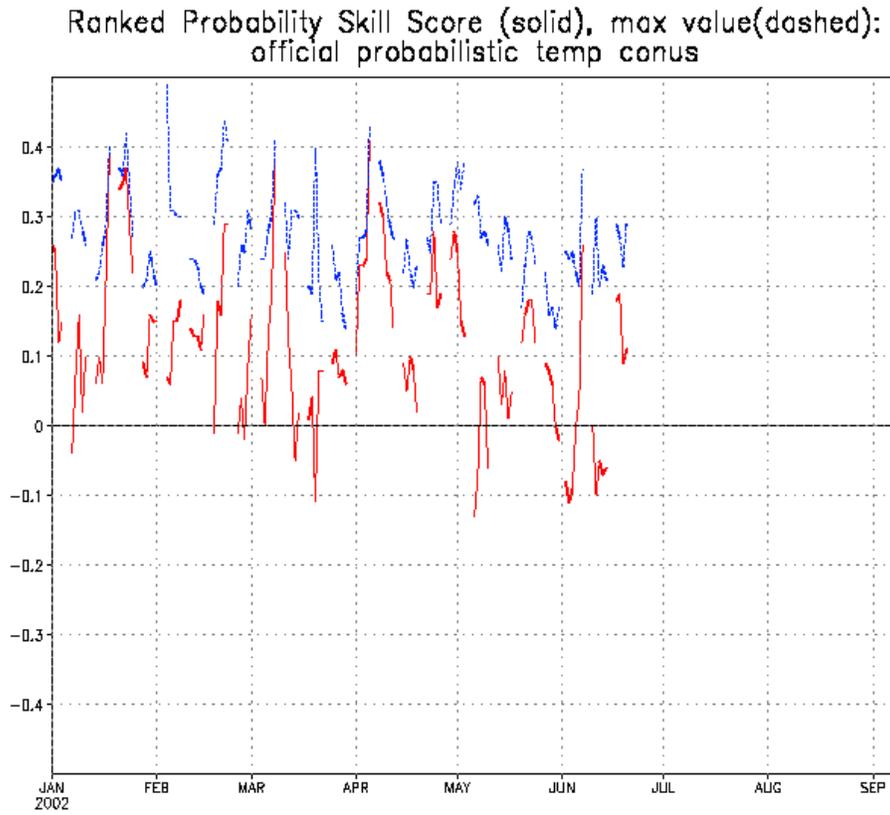


Figure 3. Ranked Probability Skill Score for 6-10 day temperature forecasts.

% contingency table				
		obs		
fcst	0	14	38	51
	0	0	3	3
	0	3	43	46
	0	16	84	100
hitrate: 0.432				
bias ratio (>1 overfcst): -- 0.17 0.55				
Heidke skill score: 0.149				

Figure 4. Contingency Table for 0.5 month lead seasonal mean CCA temperature forecasts over the continental U.S. Rows(columns) from top to bottom(left to right) correspond to below, normal, above categories, respectively.

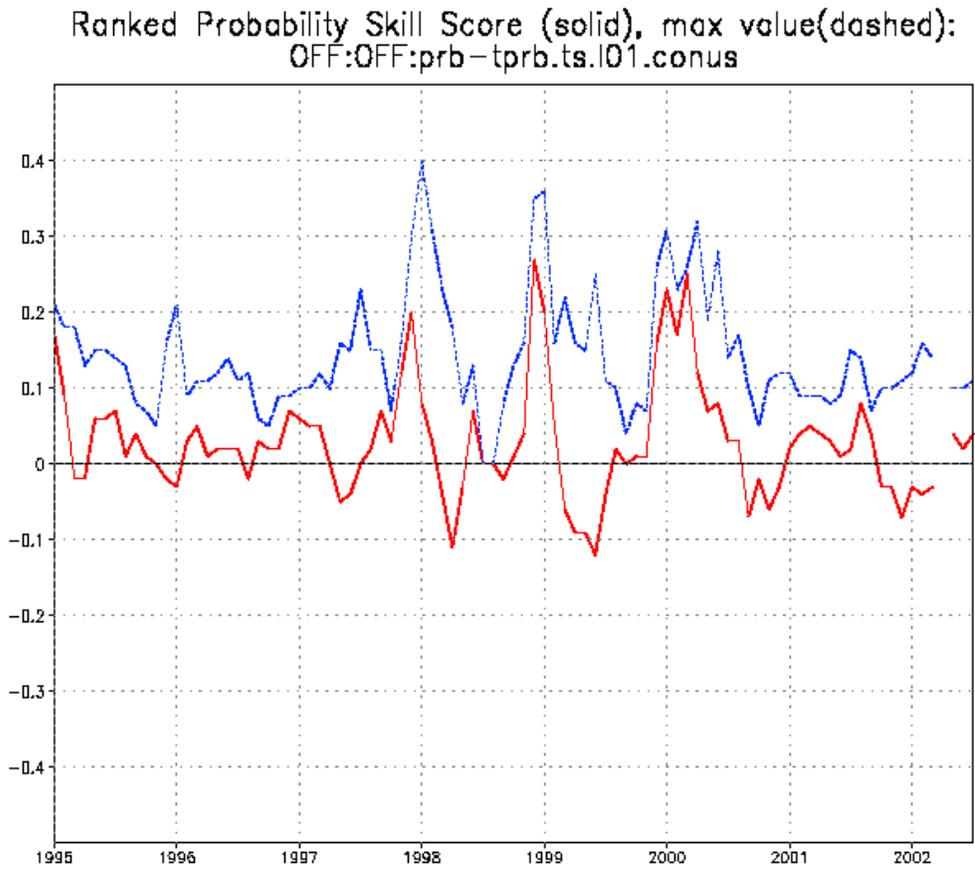


Figure 5. RPSS of official 0.5 month lead temperature forecasts, 1995-2002.

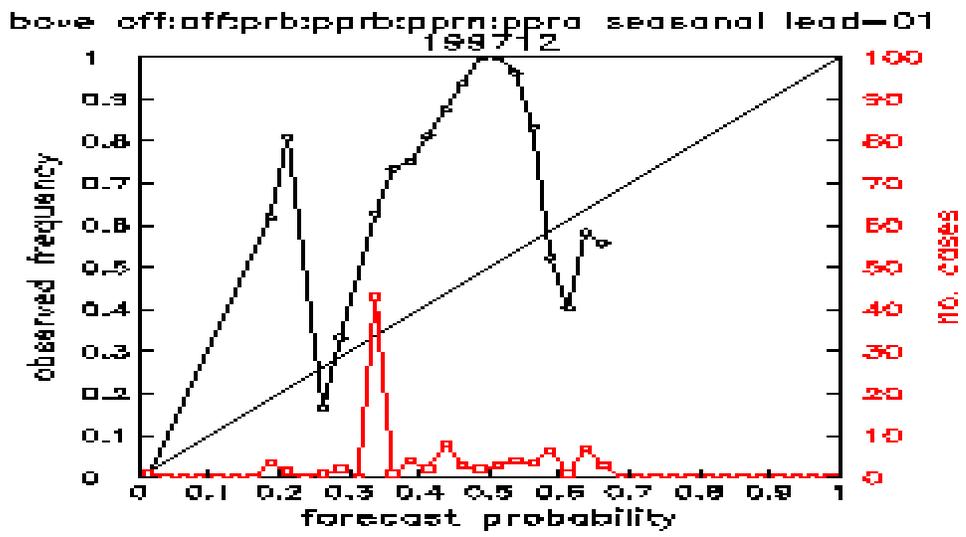


Figure 6. Reliability diagram for above normal category of 0.5 month lead official U.S. temperature forecasts.