Science and Technology Infusion Climate Bulletin
NOAA's National Weather Service
41st NOAA Annual Climate Diagnostics and Prediction Workshop
Orono, ME, 3-6 October 2016

Evaluating CPC's Operational Seasonal Temperature Forecasts: Why Aren't We Beating a Categorically Warm Forecast?

Stephen Baxter
Climate Prediction Center, NOAA/NWS/NCEP, College Park, Maryland

1. Introduction

The U.S. NOAA Climate Prediction Center (CPC) issues operational seasonal outlooks each month for the next 13 overlapping three-month seasons. The current process was implemented in 1995, and is three category in nature (above-, below-, or near-normal). There is also an equal chances (EC) category, in which there is no tilt away from climatological probabilities. Beginning in 2006, the forecast was informed by an objective consolidation of forecast tools detailed in O'Lenic *et al.* (2008). This work was initiated to see whether or not CPC's operational outlooks have adequately incorporated the consolidation forecast in the decade following its implementation.

CPC's seasonal forecast process is informed by various dynamical and statistical tools, including the Climate Forecast System (CFS), the National Multi-Model Ensemble (NMME), canonical correlation analysis, screening multiple linear regression, and constructed analog forecasts. A number of these are evaluated by an objective consolidation process that outputs probabilistic temperature and precipitation forecasts. Even following the implementation of the objective consolidation, the 0.5 month lead seasonal forecast skill has been slow to improve (Peng *et al.* 2012) for both temperature and precipitation. It is known that long term trends provide most of the forecast skill in temperature, with Peng *et al.* (2012) showing that a simple 10-year Optimal Climate Normals (OCN, average anomaly of the most recent 10 years) outperforms CPC's 0.5-month lead seasonal forecast. Only non-EC forecasts are considered in that calculation.

	1995-2005	2006-Present
Non-EC Forecast	21.83	26.93
All Forecasts	10.15	***15.91***
Coverage (%)	47.46	***58.54***

Table 1 Comparison of forecast skill and coverage before and after the implementation of the objective consolidation. Bold figures in the recent period suggest that the difference between the two periods is significant at or beyond the 98% level according to a two-sample *t*-test.

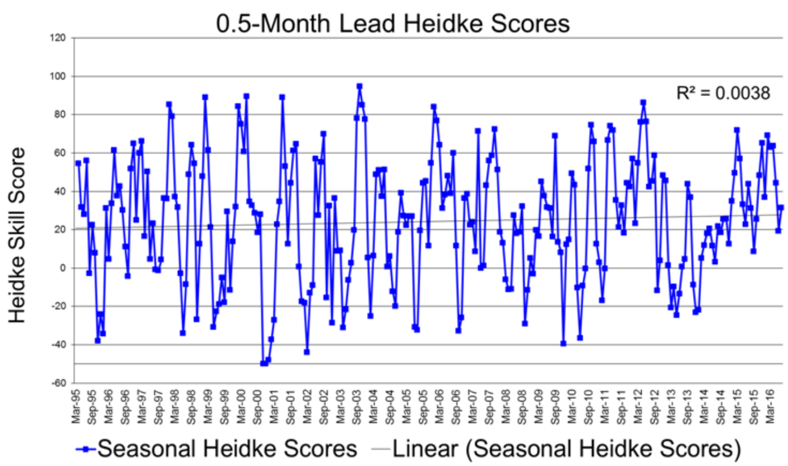

Fig. 1 Time series of official forecast skill for non-EC forecasts from 1995-present. No statistically significant trend is observed. The average over all forecasts is 24.36.

The goal of this study is two-fold:

- Evaluate the performance of CPC's official seasonal temperature forecasts in the decade following implementation of the objective consolidation.
- Consider as a benchmark a categorically warm forecast. This is the simplest assumption given a non-stationary climate.

Correspondence to: Stephen Baxter, Climate Prediction Center, NOAA/NWS/NCEP, 5830 University Research Court, College Park, MD; E-mail: Stephen.Baxter@noaa.gov

2. Methods

CPC's three category (above-, near-, and below-normal) 0.5-month lead seasonal temperature forecasts are evaluated over the CONUS on a 2°x2° grid from 1995-2016 over all seasons, with comparisons made for pre- and post-consolidation time periods. The verifying dataset results from a CPC analysis of Global Telecommunication System (GTS) data. The forecast and verification categories are the terciles derived from the appropriate climatology (*i.e.* the 1961-1990 climatology for forecasts issued in the 1990s).

Since CPC's official Government Performance and Results Act (GPRA) measure is the 48-month running Heidke skill score for non-EC areas of the 0.5-month lead seasonal temperature forecast, skill is assessed here using the Heidke skill score (HSS):

$$HSS = 100 * \frac{H-E}{T-E}$$

where H is the total number of hits, E is the number expected by chance, and T is the total number of forecasts. For a three category system, E=T/3. Heidke scores are also calculated for a categorically warm forecast ('warm dot') for comparison to the official forecast. Linear trends are calculated for the various time series, and statistical significance is assessed using a Student's *t*-test.

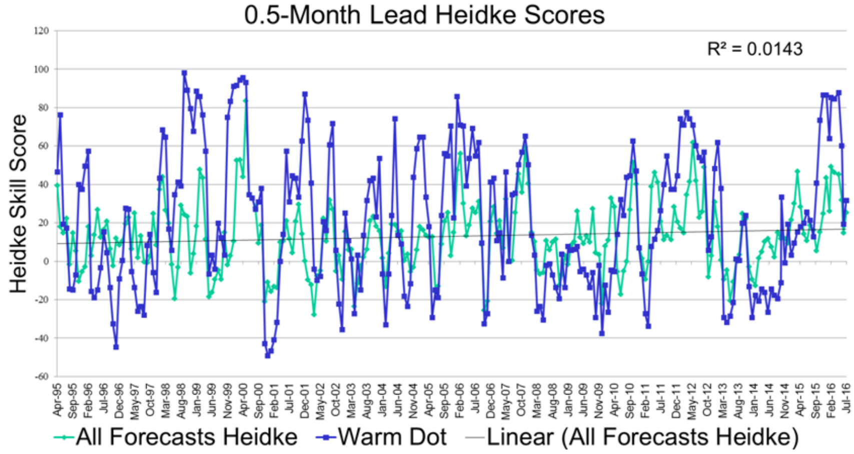

Fig. 2 Time series of official forecast skill from the official forecasts using all grid points (including EC). The time series of skill from a categorically warm forecast is also plotted. There is no signfiicant trend in either. The two time series are correlated at 0.61. The average of all official forecasts is 13.01; the average of a categorically warm forecast is 21.33.

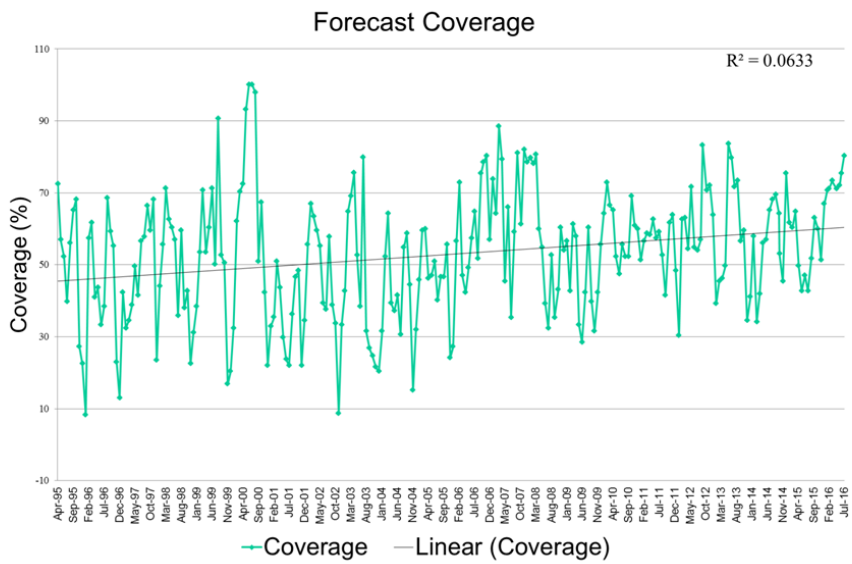

Fig. 3 Time series of forecast coverage from 1995-2016. This is the percentage of non-EC forecasts. There is a significant trend in this time series, owing mostly to a step evident when the consolidation became operational.

3. Results

The time series of official forecast skill (non-EC) from 1995 through summer 2016 is plotted in Figure 1. There is little trend in overall seasonal forecast skill since the implementation of the current forecast process in 1995. However, the implementation of the consolidation in 2006 did lead to a significant increase in forecast coverage and, as a result, all-forecasts skill (Figures 2 and 3). Since the linear trends may conceal differences in forecast skill between the pre- and post-consolidation periods, Table 1 summarizes the epochal differences in forecast skill, showing that there is indeed a statistically significant increase in 'all forecasts' skill and coverage of non-EC forecasts. Note that the non-EC forecast skill and 'all forecasts' skill are perfectly related by coverage, since EC forecasts are by definition correct one-third of the time, contributing a HSS value of 0.

While the consolidation has made a positive impact on the official forecast, a categorically warm forecast performs substantially better than the 0.5-month lead official forecast (Figure 4), and this difference grows with longer leads (not shown). The official forecast appears to add value to a categorically warm forecast over parts of the Northwest and extreme southern portions of the CONUS. This is likely due to skill afforded by the El Niño-Southern Oscillation phenomenon. Analyzing the time series of the difference in forecast skill between the official and categorically warm forecasts suggests that the categorically warm forecast is only outperformed when observations are cold (not shown).

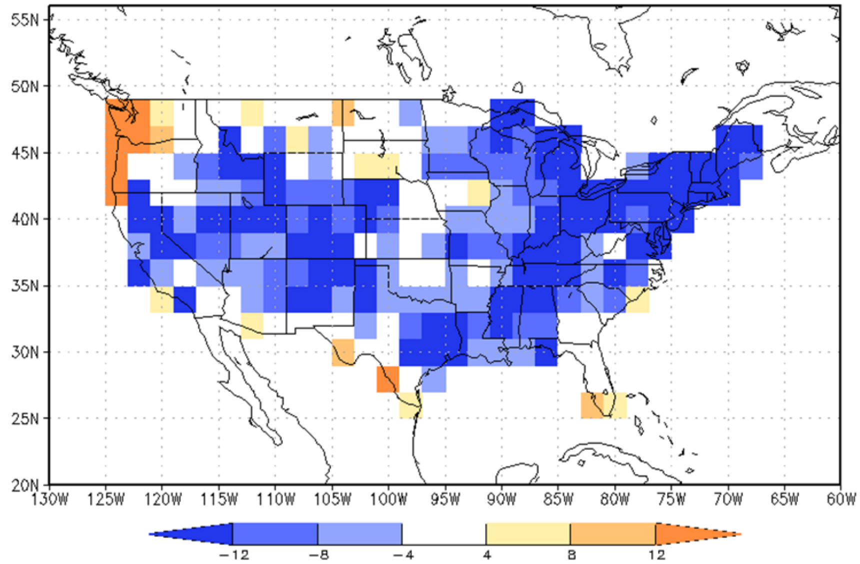

Fig. 4 Spatial distribution of the difference in HSS between the official forecast (all forecasts) over all seasons, and HSS of a categorically warm forecast. Blue colors show where a categorically warm forecast outperforms the official forecast.

4. Conclusions

While a statistically significant linear trend in the 0.5-month lead temperature forecast scores is not found, there is a significant difference in coverage and all-forecast (including the EC category) scores between the early (1995-2005) and late (2006-2016) periods. However, it is found that a categorically warm forecast is substantially better than CPC's seasonal forecasts, at least in a deterministic sense. The spatial distribution of forecast skill suggests that CPC's forecast skill derives almost entirely from trend, with some enhancement from ENSO over the Northwest, especially during spring (not shown).

Because CPC's official 0.5-month lead seasonal temperature forecasts do not adequately capture the long-term trends related to a non-stationary climatology, they can add little value to stakeholders. There are concrete steps that we can take to quickly improve our all-forecast skill scores and improve the value of our seasonal forecast products:

- Eliminate the 'Equal Chances' category from the seasonal temperature outlooks.
- Implement a new consolidation process that more transparently decomposes the seasonal forecast into components related to trends, decadal variability, and interannual variability.
- Explore issuing forecasts relative to a current climatology to isolate potentially predictable patterns of decadal and interannual variability; the available skill may currently be masked by mishandling of long-term trends.

References

O'Lenic, E. A., 2008: Developments in operational long-range climate prediction at CPC. *Wea. Forecasting*, **23**, 496–515.

Peng, P., A. Kumar, M. S. Halpert, and A. Barnston, 2012: An analysis of CPC's operational 0.5 month lead seasonal outlooks. *Wea. Forecasting*, **27**, 898–917, doi:10.1175/WAF-D-11-00143.1.