Trends in Precipitation Forecasting Skill at the California Nevada River Forecast Center

Alan Haynes
Service Coordination Hydrologist
California Nevada River Forecast Center
Sacramento, CA

Letitia Soulliard
National Precipitation Verification Unit
Office of Climate, Water, and Weather Services
Camp Springs, MD

Introduction

The California Nevada River Forecast Center (CNRFC) is one of 13 National Weather Service (NWS) River Forecast Centers in the United States. In support of fulfilling the NWS hydrologic mission of protecting lives and property, the CNRFC provides river and flood forecasts for California, Nevada, and a portion of Southern Oregon. Cool season rainfall runoff is the primary driver for high flows in many of the CNRFC's watersheds, a significant portion of which have response times less than 12 hours. Thus, CNRFC river forecasts depend heavily on quantitative precipitation forecasts (QPF).

In October of 2000, the National Precipitation Verification Unit (NPVU) was established by the Office of Climate, Water, and Weather Services (OCWWS) and is located at the Hydrologic Prediction Center (HPC) for the purpose of collecting, displaying and archiving QPF metrics. Verification metrics from the NPVU are available since November 2000 for the CNRFC. This paper attempts to ascertain the QPF verification trends that can be distinguished from this rich set of data. Given the importance of QPF in forecasting rivers at the CNRFC, understanding QPF verification trends may help the CNRFC tune the river forecasting process to generate overall improvements in their forecasts. Additionally, it is important to quantify the improvement, if any, that the human forecaster provides in the QPF inputs to the hydrologic model. From the CNRFC perspective, the three most important questions to answer are: How has the CNRFC forecast skill changed over time? Is value being added at each stage in the process? What is the Bias?

Verification Metrics

Threat Score and Bias are used in this paper to evaluate and discuss the questions presented. Threat Score is a metric that describes how accurately a forecast threshold matches the observations both spatially and temporally, while Bias is a metric that reflects how much area a forecast covers for a given threshold verses the area covered by the observations for that same threshold:

area forecast = A_f,

area observed= A_o,
area correctly forecast= A_c,

Threat Score = $A_c/(A_f + A_o - A_c)$,

Bias = A_f/A_o.

Thus, the Threat Score for a given threshold amount represents the ratio of the area where the forecast matches the observations over the area covered by either the forecast or the observations. A perfect score of 1 results when the forecast area and location match the observed area and location.

Regarding Bias in this case, just the size of the area covered matters. A perfect Bias will have the forecast and observations equal in area and the score will equal 1. A wetter or high Bias (greater than 1) is one where the forecast area for the threshold is larger than the observed area. A dryer or low Bias (less than 1) is one where the forecast area for the threshold is smaller than the observed area. A perfect Threat Score by definition must have a perfect Bias, but a perfect Bias does not mean a perfect Threat Score since the Bias does not look at how well the areas match up in location, just in size.

The CNRFC is interested in long term trends in QPF skill for conditions that are likely to impact river flows. Thus, the focus is on cool season precipitation where areal amounts exceed 0.50 inches in any six-hour period on a 32km HRAP grid at 6hr intervals. The GFS and RFC forecasts and the observations have both been remapped from their original grids to this verification grid. The HPC grid is native to the verification grid. The verification metrics are readily accessible for the "cold season" from the NPVU (October through March, except November through March for 2000) and for categorical amounts exceeding 0.50 inches in six hours in the form of Threat Score. This "cold season" coincides reasonably well with the Mediterranean climate that predominates over California, where most of the annual precipitation falls during the months from October to April.

For the purposes of this paper, the QPF process represents a cascade starting from the numerical model guidance availability to HPC forecasters and ending with the CNRFC forecasters. Consequently, there is a temporal disconnect between all three sources of QPF, which is something that should be considered when comparing the QPF metrics. The GFS reflects a 0000 UTC forecast basis, while it is typically available to the field beginning around 0300 UTC. The HPC forecast, which is based on the 0000 UTC GFS, is considered to be anchored at 1200 UTC and is published for the CNRFC's use around 1000 UTC. The CNRFC forecast is also considered to have a 1200 UTC basis, although it is published near 1500 UTC.

Currently, the verification metrics that are easily accessible from the NPVU are lumped over the entire CNRFC domain. Although it would be useful to examine long term trends for smaller domains with similar hydrologic characteristics (e.g., the California North Coast), it is not necessary for the general purposes of this study.

Analysis and Discussion

How has the CNRFC forecast skill changed over time? Figure 1 shows the Threat Score computed over each cool season for the CNRFC, HPC and the Global Forecast System (GFS) numerical model. The scores are for events defined by occurrences where either the observations or forecasts equaled or exceeded 0.50 inches in a six-hour period for all forecasts in day 1 (F06, F12, F18, and F24) over the cool season only (Oct-Mar). A linear fit to the data is annotated for each of the forecast sources. Also plotted in figure 1 is the number of observations for each season, to highlight periods with relatively low sample sizes, such as in the 2000, 2006 and 2008 seasons.

There are several things that stand out in figure 1. First, it is very clear that seasonal skill in the CNRFC forecasts closely parallels the skill in the HPC forecasts, upon which they are based. Likewise, both the CNRFC and HPC forecast skill closely parallel the performance of the GFS model. Earlier studies (Olson et al. 1995) pointed out that the increase in forecaster QPF accuracy over the years was more dependent on numerical model guidance than on the number of years of experience by the forecaster. Reynolds (2003) articulated the value of the forecaster in QPFs. He looked at long-term trends in the day-1 manual QPF Threat Scores over a 37-year period using the 1.0 inch threshold. Reynolds asserted that it is not only appropriate to attribute the slow but steady rise in the manual forecast accuracy to a slow and steady rise in the accuracy of the numerical models but that any improvement beyond that produced by the models can be defined as the accuracy of the forecasters relative to the accuracy of the models.

We can now consider whether value is being added at each stage in the process. Figure 1 shows that for each season the HPC skill is higher than the GFS skill and the CNRFC skill is higher than the HPC skill, affirming that each step in the QPF process adds value. Also, although it is apparent that improvement in QPF accuracy is slow and steady, the rate of change in improvement is evidently higher for the CNRFC forecast than for the HPC and GFS forecasts. The CNRFC has invested heavily in its Hydrometeorological Analysis and Support (HAS) function, dedicating three meteorologists who focus solely on the forecast precipitation and temperature inputs to the hydrologic model. This investment appears to be validated, not only due to the improvement in accuracy of the CNRFC forecasts over the HPC and GFS forecasts, but also as distinguished in the subtle, yet significant difference in the rate of improvement in the forecast. There has been little turnover in the HAS forecaster ranks at the CNRFC over the past 10 years. The higher rate of improvement for CNRFC forecasters could be attributed to the development of a shared experience base, given the concentration on a limited geographic domain (i.e., the CNRFC area of responsibility) and the focus on a narrow set of meteorological variables (i.e., temperature and precipitation). Additionally, the availability of NPVU verification metrics may play an important role in providing feedback to the CNRFC HAS forecasters.

There are a couple of caveats when considering the general attribution of added value by the HPC and CNRFC forecasters. First, there is a temporal lag introduced in the cascading process, i.e., the CNRFC forecasters are operating closer in time to the forecast events than both the GFS and HPC, which gives them an advantage. This advantage is especially significant in the first period, since the CNRFC forecast is actually issued a couple of hours after the stated forecast initial time. Furthermore, additional guidance beyond the GFS is available to the forecaster. For example, Novak (2010) argues that much of the HPC QPF skill is related to their access to the very skillful ECMWF deterministic and ensemble guidance. HPC makes nearly a 30% improvement over the GFS, but only a 5-10 % improvement on the ECMWF. It should also be noted that the verification metrics from the proprietary ECMWF model are not tracked by the NPVU.

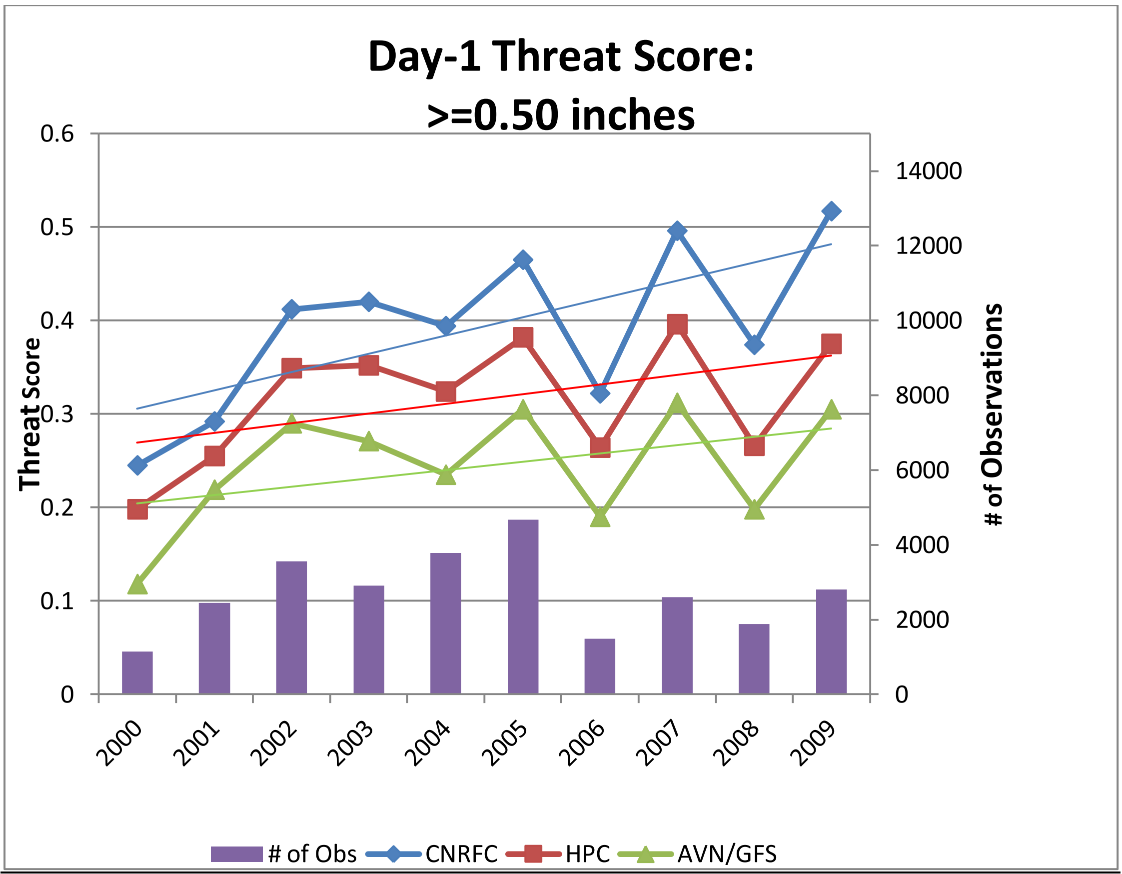

Fig 1. A comparison of CNRFC, HPC and GFS day-1 seasonal (Oct-Mar) Threat Scores. The linear trends for each source are annotated and the number of observations is plotted in bar graph form for each season.

So how can the aggregate value added by forecasters at HPC and the CNRFC be quantified? Reynolds (2003) related forecaster improvements made to the numerical guidance QPF to the number of years of model development required to reach the same level of accuracy. Looking at a plot of GFS forecast skill for day 1, (figure 1), considering a linear fit to the data and a reference year of 2000, it would take 12.4 years for the GFS accuracy to match the 2000 CNRFC accuracy, while it took 4.5 years for the HPC forecasts to match the CNRFC accuracy in 2000.

Alternatively, if we now juxtapose the CNRFC plots of Threat Score for days 2 and 3 with that of the GFS for day 1 (figure 2), we see that the CNRFC forecast provides well over one day of lead time in accuracy compared to the GFS, actually closer to two days of lead time improvement.

Examining the linear trends for the GFS model for the day-1, day-2 and day-3 forecasts (figure 3), the pace works out to 5.9 years for the day-2 accuracy to match the day-1 accuracy, while it takes 10.4 years for the day-3 forecast to match the accuracy of day-1. From this view, we can say that the CNRFC forecast provides close to 10 years of GFS model development. The question is sometimes posed as to whether it is worth the cost to manually modify the QPF given that model accuracy continues to improve, especially considering that numerical model improvements have become more frequent, making it harder for forecasters to track the Biases and potentially add value to the forecasts. Reynolds (2003) builds on the notion put forth by Doswell and Brooks (1998) that the benefit of the forecaster to the user community must be quantified and he demonstrated a 14-year improvement in QPF accuracy for HPC forecasts over the numerical models. Here we demonstrate that HPC and CNRFC human forecasters provide roughly a decade

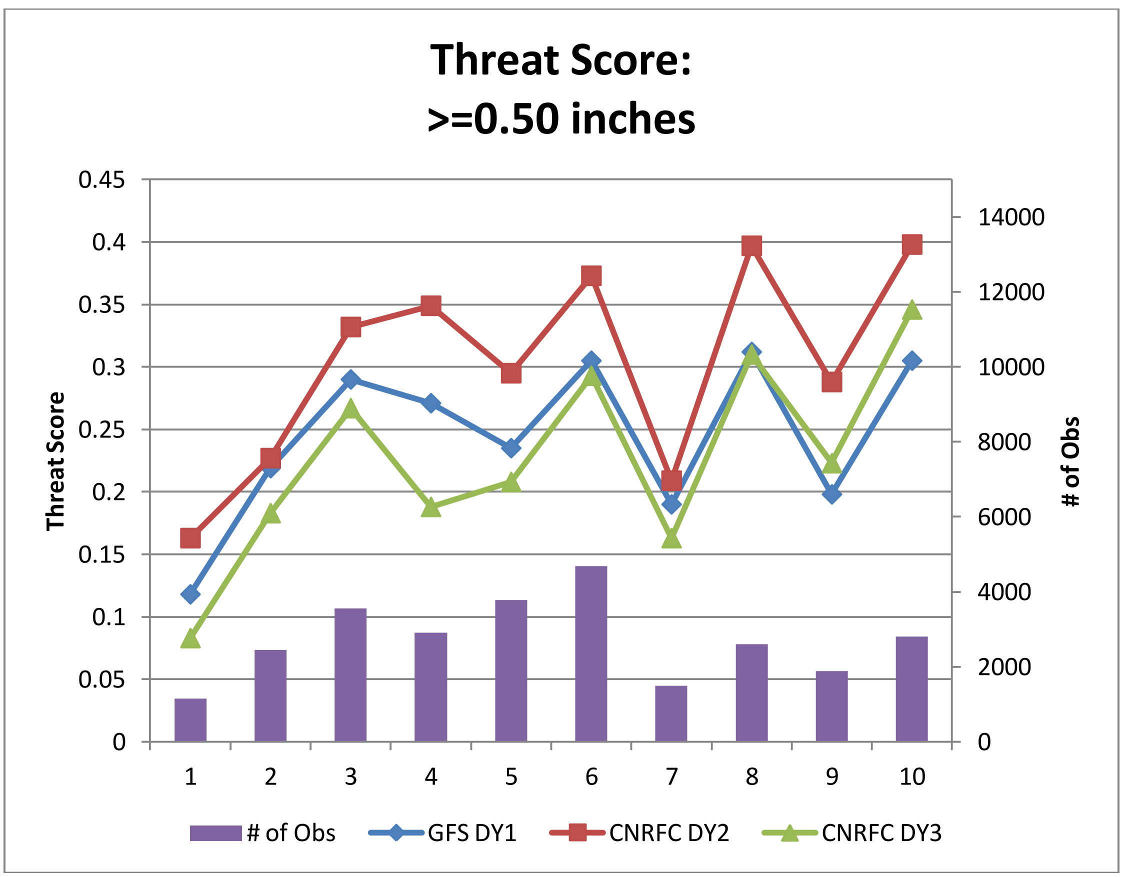

Fig 2. A comparison of GFS day-1 and CNRFC day-2 and day-3 seasonal (Oct-Mar) Threat Score. The linear trends for each day are annotated and the number of observations is plotted in bar graph form for each season.

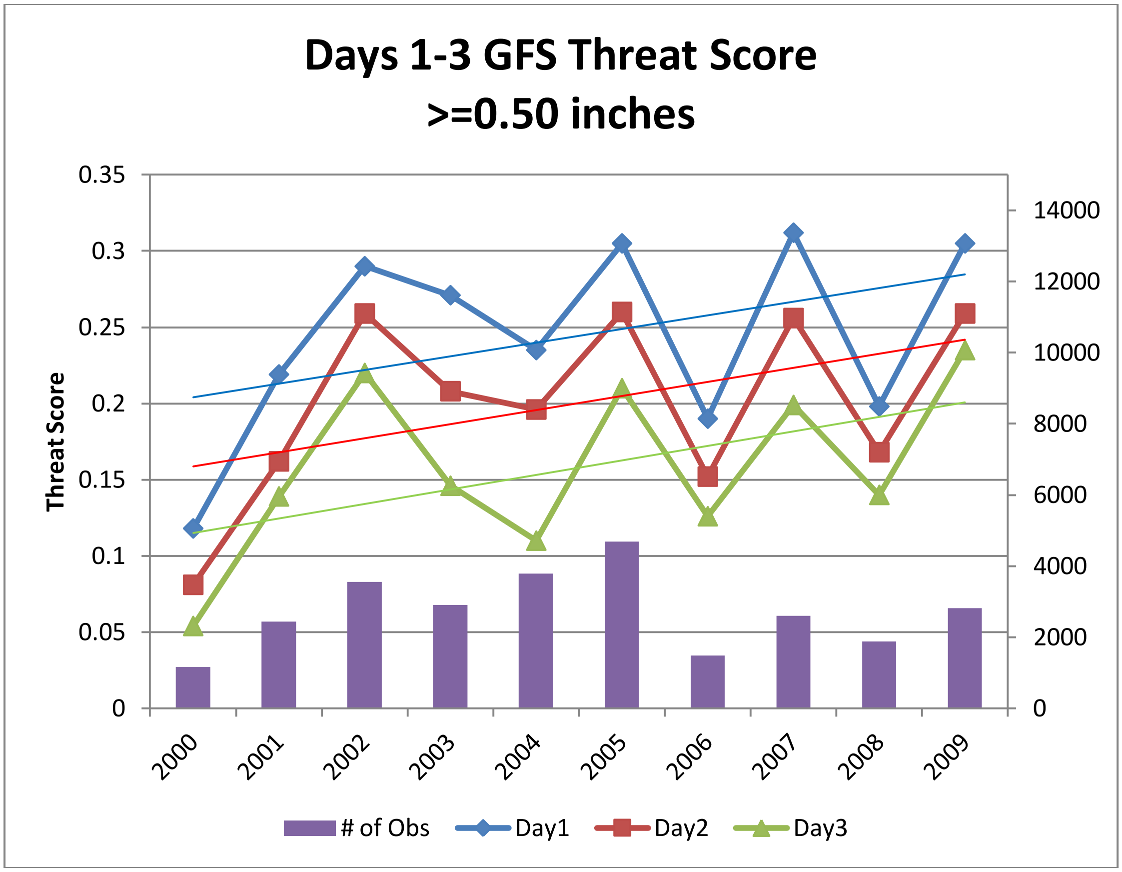

Fig 3. GFS days 1-3 seasonal (Oct-Mar) Threat Scores. A linear fit to the data is annotated for each day. The linear trends for each day are annotated and the number of observations is plotted in bar graph form for each season.

worth of model development time, or about two days of improved lead time over the GFS model. This validates the current paradigm of QPF production where HPC adjusts model-produced QPF at a broad scale followed by further refinement by the CNRFC HAS unit.

Figure 4 is similar to figure 1, but shows only the CNRFC scores for Days 1, 2, and 3. Again, if we consider a reference year of 2000 and examine the linear trends, what we find is that the CNRFC day-2 forecast accuracy took about 5.7 years to reach the day-1 accuracy, while the day-3 accuracy took about 10 years to reach the reference day-1 accuracy. So just looking at the trends internally to the CNRFC, we can conclude that the CNRFC gains about one day of lead-time's worth of QPF accuracy about every six years. It also appears that relatively low sample size negatively impacts forecast accuracy, such as observed in the 2006 and 2008 seasons.

One final piece of statistical information to consider is Bias. Are the CNRFC forecasts Biased? Figure 5 plots the seasonal Biases for the CNRFC, HPC and GFS day-1 QPFs. One notable feature is that the GFS forecasts are remarkably unBiased. The CNRFC forecasts have a dry Bias for the categorical events greater than or equal to 0.50 inches (0.87) while the HPC Bias is 1.08 and the GFS 1.03. However, when considering all categories (figure

6), the seasonally averaged CNRFC Bias moves close to one (1.08) while both HPC and the GFS show wet Biases (1.20 and 1.30 respectively).

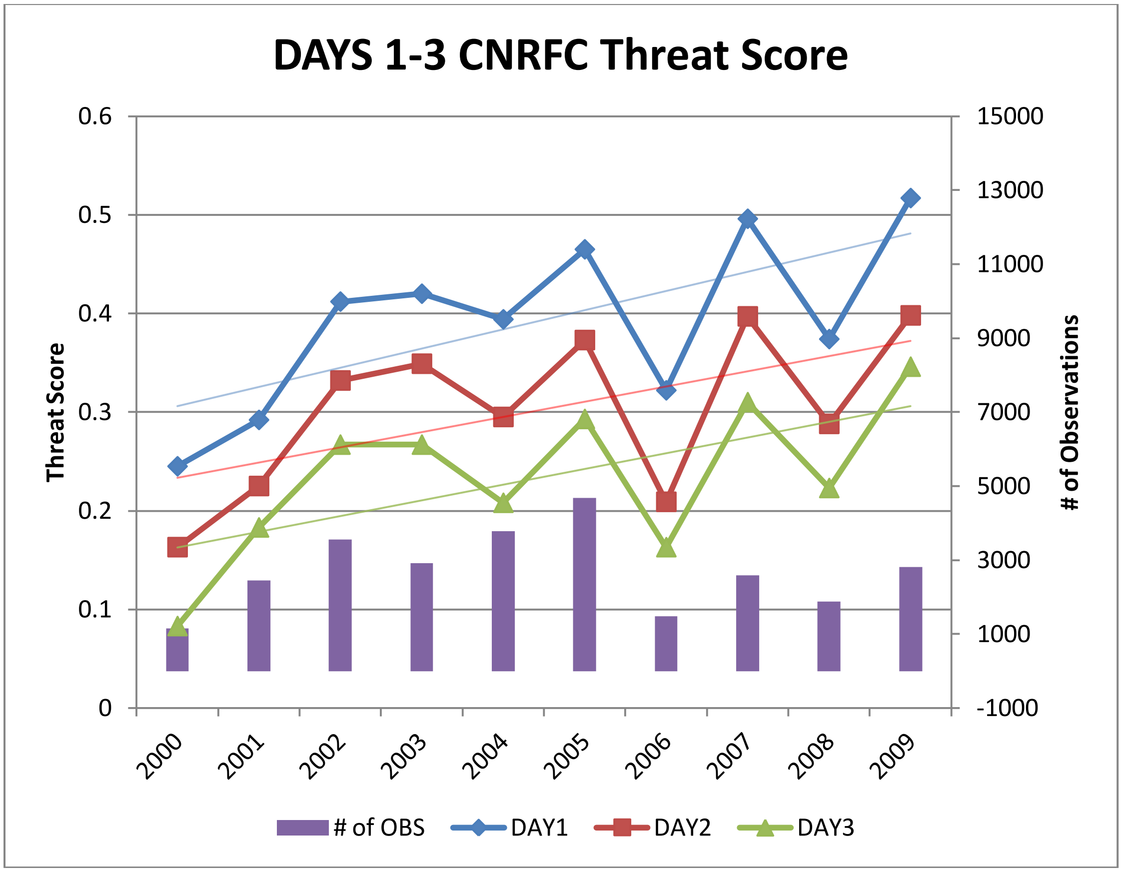

Fig 4. A comparison of CNRFC days 1-3 seasonal (Oct-Mar) Threat Score. The linear trends for each day are annotated and the number of observations is plotted in bar graph form for each season.

Summary and Conclusions

Upon examining seasonal statistics for the CNRFC, HPC and the GFS, as depicted in Threat Score for six-hour QPFs in excess of 0.50 inches for day 1, it is clear that the accuracy is trending higher with time and numerical model guidance is likely driving the upward trend with HPC and CNRFC forecasters successively adding value to these forecasts. The quantification of value added to the GFS QPF guidance by the CNRFC forecasters, which implicitly includes the value added by the HPC forecasters, translates to nearly two days of additional lead time or about a decade's worth of model development. Moreover, the CNRFC forecasts are gaining one day's worth of lead-time in accuracy about every five to six years. Thus we can say that in general, the CNRFC's day 2 forecasts today are about as accurate as the CNRFC's day 1 forecasts were six years ago.

This analysis indicates that the current paradigm in the production of QPF, which follows the "Snellman funnel" approach outlined in Snellman and Thaler (1993), is working. HPC modifies the QPF on a broad scale followed by more detailed scrutiny at the CNRFC where regional expertise is incorporated. Additionally, it appears that Snellman (1982) is vindicated in his strong conviction to keep meteorologists engaged in the forecast process in the face of increasing pressure to use straight model guidance, even after several decades of model development.

Furthermore, the foresight and investment in establishing the NPVU a decade ago seems to be bearing significant fruit. The feedback provided by the NPVU metrics is essential, both in helping forecasters to improve and in quantifying the value added by forecasters over the numerical model guidance.

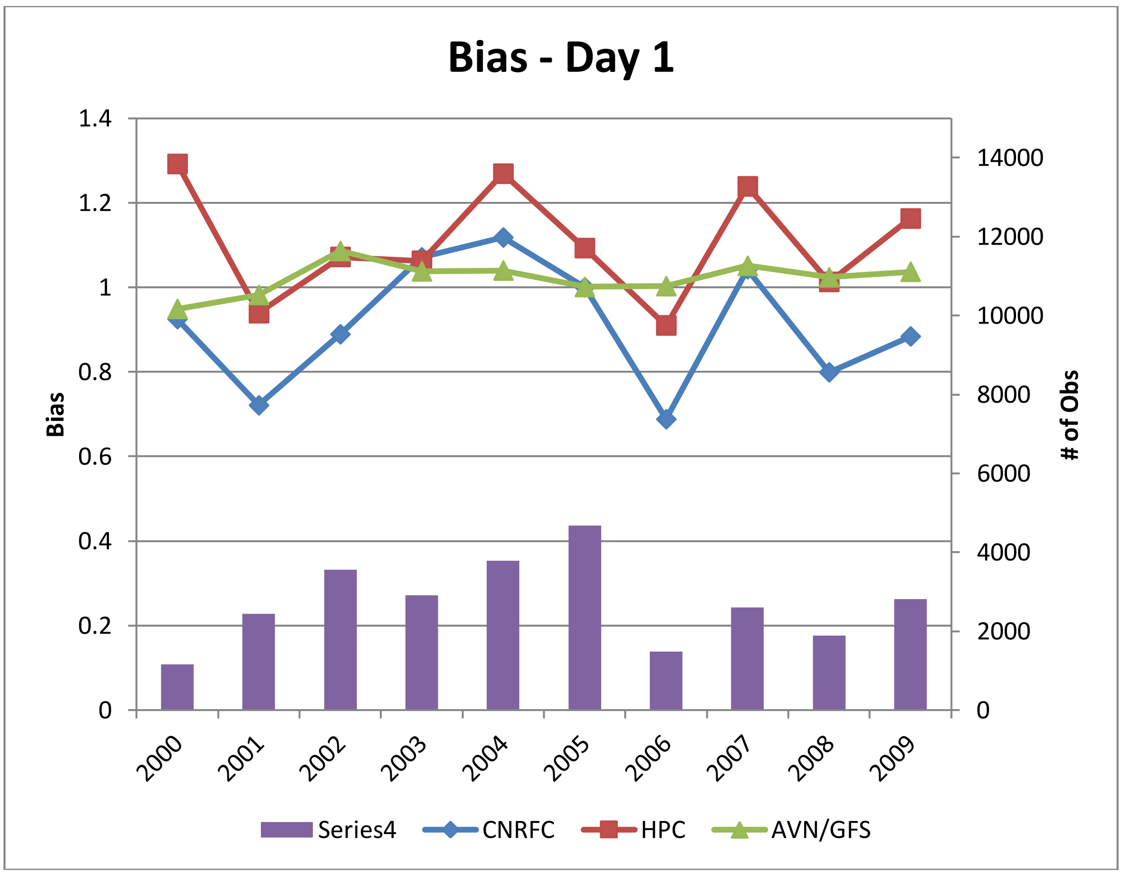

Fig 5. A comparison of seasonal (Oct-Mar) Biases for the CNRFC, HPC and the GFS QPFs for events greater than or equal to 0.50 inches. The number of observations is plotted in bar graph form for each season.

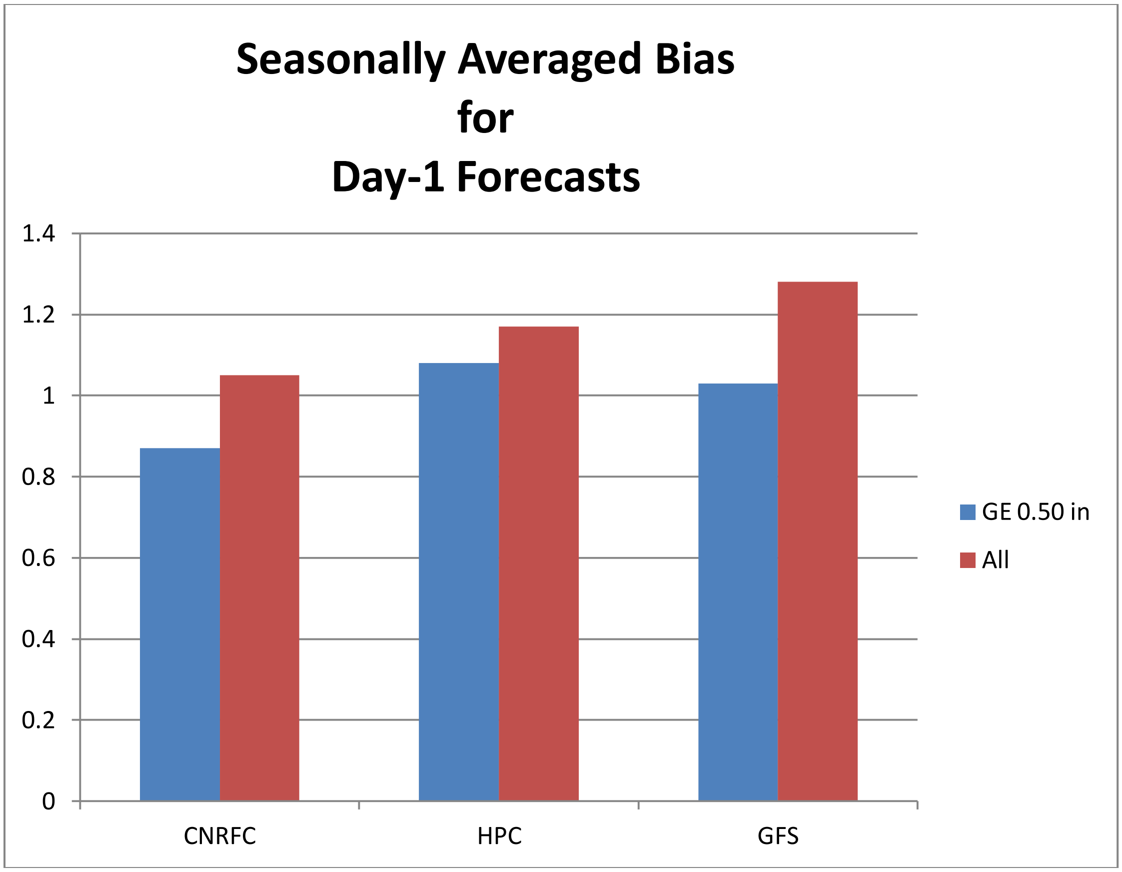

Fig 6. Seasonally (Oct-Mar) averaged Biases for the CNRFC, HPC and the GFS QPFs for day-1. Categorical events >= 0.50 inches are shown alongside all categories.

Acknowledgments

The authors are very grateful to David Novak for his review and provision of helpful comments toward this publication.

References

Doswell, C. A., III, and H. E. Brooks, 1998: Budget-cutting and the value of weather services. *Wea. Forecasting*, **13**, 206-212.

Novak, D., 2010: *Personal correspondence*.

Olson, D. A., Junker, W., and B. Korty, 1995: Evaluation of 33 Years of Quantitative Precipitation Forecasting at the NMC. *Wea. Forecasting*, **10**, 498-511.

Reynolds, D., 2003: Value-Added Quantitative Precipitation Forecasts: How Valuable is the Forecaster?, *Bull. Amer. Meteor. Soc.*, **84,** pp 876-878.

Snellman, L. W., 1982: Impact of AFOS on operational forecasting.
Preprints, *Ninth Conf. on Weather Forecasting and Analysis,* Seattle, WA, Amer. Meteor. Soc., 13–15.
——, and E. Thaler, 1993: Forecast process. COMET, CD-ROM version 1.0. [Available from COMET, UCA