6.5 POST-PROCESSING BEYOND THE REALM OF WEATHER AND INTO CLIMATE

David A. Unger
Innovim/NOAA/NCEP/CPC

Dan Collins
NOAA/NCEP/CPC

Arun Kumar
NOAA/NCEP/CPC

1. INTRODUCTION

Post-processing of numerical models for climate and extended range weather forecasting differs from that of shorter ranges. Post-processing of weather forecasts focuses on calibrating models to observations related to specific synoptic scale events. A climate forecast, on the other hand, focuses on the range and frequency of observations that are associated with a large ensemble of synoptic scale events that are individually unpredictable at long lead times. At intermediate ranges, referred to here as 'extended range weather forecasts', the skill in predicting individual synoptic scale events diminishes to the point that it becomes advantageous to predict conditions for multi-day periods. Thus, forecasts in these ranges transition in character from weather to climate forecasts.

The evolution of operational post-processing techniques used for climate and extended range weather forecasts will be examined in this paper. Extended ranges will be defined here from the standpoint of weather prediction skill in the 1970's when operational numerical weather prediction (NWP) models matured enough to predict weather at both the surface and upper levels beyond a few days. In 1970 the U.S. Weather Service extended its daily forecasts to 5-days, (O'Connor, 1980) leaving lead times of 6-days and beyond as extended ranges. In December, 1977 the Meteorological Operations Division of the U.S. Weather Service issued its first extended range forecast for leads beyond 6 days (Figure 1) (Andrews, 1977). The 6-10 day outlook has been issued continuously since that date and will be examined closely here.

.2. ORIGINS

In 1935, the U.S. Weather Bureau, together with the U.S. Department of Agriculture, funded research to predict weather beyond the 1 to 2 day lead times issued by forecasters at the time. Dr. Jerome Namias (Figure 2) led a team of scientists

Fig. 1. Jim Andrews and George Cressman announce the implementation of the 6-10 day forecast. December 14, 1977

at the Massachusetts Institute of Technology to investigate and develop methods of extended range forecasting (Namias, 1943).

After initial attempts, the team abandoned efforst to predict daily weather in favor of predicting mean conditions over multi-day periods. Five-days was chosen as an optimum period to filter out the unpredictable short-wave systems and focus on the more predictable long wave patterns. Extended range forecasts for mean temperatures and precipitation totals in the 2-6 day period were first issued in 1941.

In a 1953 publication, reflecting on those early efforts, Namias, when comparing short term weather forecasts to extended ranges, stated:

"In extended forecasting, on the other hand, one is not so much concerned with the individual features of the daily map, cyclones or even upper-level waves, as with the hemispheric ensemble of atmospheric circulation that has been evolving over a long period of time." (Namias, 1953, pg 2.)

This statement clearly expressed the philosophy that formed the basis of extended

Fig. 2. Jerome Namias in the late 1960's

range forecasting through at least the 1990s, and influenced the design of early model post-processing at extended ranges. Individual weather systems that were unpredictable by the technology at the time were filtered out by employing multiple day means. Emphasis was placed on upper level circulation; the 10,000 foot pressure levels in the 1930's, and later the 700-hpa heights, and its relation to surface weather.

In the absence of numerical weather prediction models Namias' team focused on graphical techniques to predict the relatively slowly evolving upper level waves and the associated surface weather. Surface conditions were predicted by statistical relationships between upper level information and surface weather elements.

The forecasting methods used for the 6-10 day forecasts issued in the 1970's were influenced largely from the experiences gained in over 35 years of 2-6 day forecasting. The forecast format itself was directly inherited from the early forecasts. Both the 2-6 day and 6-10 day products were for mean temperature and total precipitation in 5-day periods. Both express anomalies in terms of broad categories and were compared to climatological distributions.

It should come as no surprise that the experience gained from extended range forecasting for the 2-6 day forecast would be applied to days 6-10 when NWP extended the useful daily prediction ranges to 5-days.

3. EARLY POST-PROCESSING

Figure 3 shows a graphical technique developed for the 2-6 day forecast by the Extended Range Forecast Group of the Weather Bureau in the 1950's. This figure was reproduced from the review of the early 30-day forecasts (Namias, 1953). At that time the 30-day forecasts were based heavily on techniques developed for 5-day forecasting.

Mean temperature anomalies at forecast locations were linked to 5-day mean 700-hpa heights. Temperatures were related not only to conditions overhead, but also to anomalies at so called "centers of action" upstream. The figure on the right shows the upstream points at the head of the lines that associate with downstream anomalies for location their tails. Centers of action were based on correlations between upper level heights and mean temperatures at the forecast location. Correlations between 700-hpa heights and temperatures at Evansville, IN are shown on the leftmost panel in Figure 3. Contoured surface 5-day mean temperatures relating height anomalies overhead and at the remote location formed the basis for the prediction (Center frame).

These graphical techniques were emulated in early post-processing of NWP in the late 1950's in statistical relationships developed by William Klein (Klein et al., 1959). These "Klein Specifications" were regression equations based on concurrent

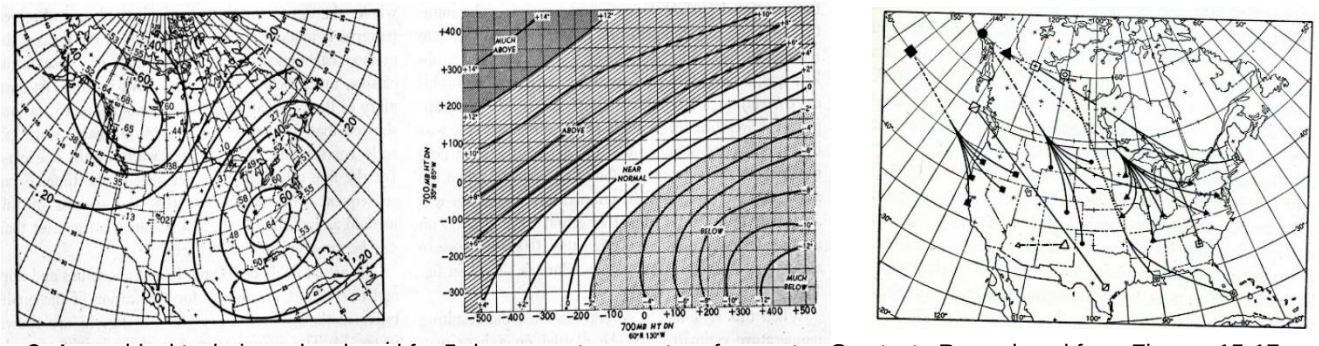

Fig. 3. A graphical technique developed for 5-day mean temperature forecasts. See text. Reproduced from Figures 15-17, Namias, 1953.

observations of heights and temperatures and were applied to forecasts from early models of the time and are among the first examples of routine numerical post-processing of NWP output. (Namias and Collaborators, 1958). Short term regression bias corrections were often applied to the model heights prior to application of Klein specifications. (Harnak, 1986)

Analog specification methods were used for both temperature and precipitation forecasts. Composite means or totals for each element were computed from the dates of the approximately 20 past cases that correlate highest with forecast 5-day mean 500-hpa anomalies. The correlation domain was restricted to points over North America and vicinity.

4. MODERN POST-PROCESSING

By the early 1990's general circulation models being used for weather prediction were able to realistically simulate the daily circulations well beyond the week or so that useful predictions could be made. Operational climate models used to support CPC's Monthly and Seasonal Outlooks have evolved through the years. The Coupled Model Project (Ji, et al. 1994) was used in the 1990's. This was replaced by the Climate Forecast System, CFS, in 2004 (Saha, et al. 2006) and upgraded to the CFS version 2 (CFSv2) in 2011 (Saha, et al. 2014). From their outset, the post-processing of climate model output often involved relatively simple calibration methods, at least for operational use. These post-processing methods while relatively simple, require large amount of historical data that is made available through retrospective runs on past data, known as hindcasts, for as long a period as possible.

4.1. Ensemble Mathematics

The relatively simple calibration methods commonly used on climate models are frequently quite close to an optimum linear least-squares solution to an ensemble forecast. Recall that at climate ranges, the observations can be thought of as probabilistic, in the sense that they result from an ensemble of unpredictable individual synoptic scale events that collectively produce a climatic distribution. Post-processing for climate ranges involves the preservation of the distributional aspects of the forecasts.

In its most fundamental interpretation, a forecast ensemble represents a sample drawn from a population of possible solutions. We will examine the ensemble here in terms of a non-parametric distribution.

The simplest calibration method is a bias correction, which may be written as follows:

$$Bias = (\bar{F} - \overline{Obs}) \ .$$

Where $\bar{F}$ and $\overline{OBS}$ are the respective model forecast and observation means on hindcast data. The terms can be rearranged as follows:

$$F_{cal} = F - Bias$$

$$F - \bar{F} = F_{cal} - \overline{Obs}$$

where F_{Cal} is the calibrated forecast. This can be rewritten as:

$$Anomaly(F) = Anomaly_{Cal}(Obs)$$

Which reads as: The forecast anomaly relative to the forecast climatology, Anomaly(F), is assumed equal to the calibrated forecast anomaly relative to the climatology of the observations ($Anomaly_{Cal}(Obs)$).

A continuous forecast distribution can be obtained by a kernel density estimation (KDE) technique (Silverman, 1986) that corrects for the known bias of the sample variance. An N-member ensemble might be regarded as a sample of the larger unknown population of solutions. The N-member ensemble spread is a biased estimate of the likely variance of the population of solutions. A Gaussian kernel applied around each member can be used to help estimate the forecast distribution. The Gaussian kernel width can be obtained by calculating the amount required to compensate for the bias in sample variance. In the following relationships, σ_p and σ_s are respectively the sample and population standard deviation, and N is the ensemble size.

$$\sigma_p = \sqrt{\frac{N}{N-1}}\sigma_S$$

$$\sigma_p^2 = \sigma_s^2 + \sigma_k^2$$

$$\sigma_k = \frac{\sigma_s}{\sqrt{N-1}} \qquad (1)$$

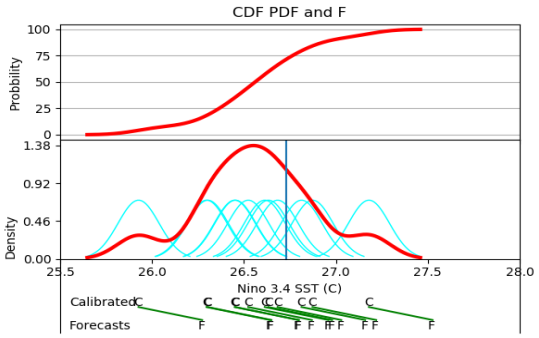

Fig. 4. Illustration of a distribution formed by a KDE estimate of a bias corrected ensemble forecast. See text for explanation.

A simple bias correction can be illustrated by the distribution in Figure 4. This illustrative case shows an ensemble forecast for the 3-month seasonal mean Nino 3.4 SSTs from the CFSv2 hindcasts. The forecast was initialized with data from November, 2005 and is a 3-month lead, valid for FMA 2006.

Original forecasts are represented by the letter F below the x-axis and the calibrated forecasts are shown by "C". The green lines connect the corresponding calibrated and uncalibrated forecasts. The cumulative probability distribution function (CDF) is shown on top and the probability density (PDF) is shown below it. The distributions are obtained by KDE with the kernels shown in light blue lines. Kernel amplitudes are not drawn to scale. Kernel widths are from Eq. 1. The observed SST for the period is indicated by the vertical blue bar.

Note that KDE is not commonly used in model post-processing and is shown here for later comparisons. The CDF is usually estimated by simple model counts or distributional assumptions applied to the ensemble.

Table 1 shows the most common post-processing methods applied to climate predictions. The error distributions for the non-regression based methods are obtained by KDE with kernel width set to correct for sample bias. σ_e is the sample standard deviations of the N=15 members, and is computed from the calibrated ensemble. This value is substituted for σ_s in Eq. 1. The term 'kernel' in the table signifies when the error distributions are applied to each ensemble member. 'Single' denotes a distribution around the calibrated ensemble mean.

For reasonably well behaved distributions, free from discontinuities and severe skews, percentile mapping is equivalent to the PDF correction. Percentile mapping/PDF correction is among the most common post-processing methods for climate prediction. It has, for example, been used heavily for post-processing of the North American Multimodel Ensemble (NMME) (Becker and van den Dool, 2014).

It seems reasonable that numerical modelers seek a linear relationship between forecasts and observations, so one might assume that linear regression is an appropriate calibration tool. Table 1 shows that regression is similar to PDF calibration except that the forecast is damped toward climatology by the correlation coefficient between the ensemble mean and the observations, R_M. Standard regression, however, filters out the case-dependent information from the forecast ensembles, and instead provides only a distribution around the ensemble mean based on model performance on the entire hindcast data.

A least squared solution based on linear regression exists for the ensemble set as a whole (Unger, 2009). This makes use of the *a priori* assumption that ensemble members represent an equally likely subset of possible solutions. With the assumption of equal weights, an expected value of

Table 1. Mathematical Formulations of Post-processing methods used for Extended Range forecasting.

Name	Formulation	Error Estimate
Bias Correction	$Anomaly(F) = Anomaly_{Cal}(Obs)$	$\frac{\sigma_e}{\sqrt{N-1}}$, Kernel
Percentile Mapping	$CDF(F) = CDF_{Cal}(Obs)$	$\frac{\sigma_e}{\sqrt{N-1}}$, Kernel
PDF Correction	$Z(F)$=Z(Obs)	$\frac{\sigma_e}{\sqrt{N-,1}}$, Kernel
Standard Regression	$R_M Z(F)$=Z_{Cal}(Obs)	$\sigma_{Obs}\sqrt{1-R_M^2}$, Single
Ensemble Regression	$R_{Best} Z(F)$=Z_{Cal}(Obs)	$\sigma_{Obs}\sqrt{1-R_{Best}^2}$, Kernel

the best-member (closest to the observation) correlation coefficient, R_{Best}, can be estimated statistically as:

$$R_{Best} = \frac{R_M^2}{R_i}$$

Where R_i is the correlation between the set of all individual ensemble members and the observation on hindcast data. The best member correlation damps the individual ensemble members always to a lesser extent than standard linear regression, because it accounts for the variance predicted by the ensemble spread. The regression error estimate is applied to each regression-calibrated member of the ensemble and takes the form of a KDE estimation with kernels determined statistically.

The ensemble mean is an indication of the signal while the ensemble spread represents the predicted component of the noise. The kernel distributions are the residual component of the noise, not already accounted for by the ensemble spread. The mean ensemble spread represents the model's estimate its own skill, and if it's realistic, R_{Best}, approaches 1 and the least squared solution approaches a PDF correction. If the overall forecast variance (of individual members) is close to that of the observation, a PDF correction becomes nearly equivalent to a bias correction.

Figure 5 shows a PDF correction for the forecast shown in Figure 4. Figure 6 illustrates a comparison of the distribution estimate from a PDF correction, standard regression and ensemble regression. It shows that a simple PDF correction in this case is quite close to the ensemble regression-based distribution. In this case both the PDF and ensemble regression method outperformed a standard regression (as measured by the forecast probability density near the observation) because they retained information on the clustering of solutions in the ensemble forecasts.

PDF corrections/Percentile Mapping and simple bias correction are frequently used for post processing of extended range forecasts. Major systems that rely upon bias correction alone include the NAEFS ((Candille, et al. 2010, Cui et al. 2012), and the CFSv2 for some elements (Saha, et al. 2014).

4.2. Impact on the 6-10 day forecasts

Post-processing methods used for the climate models began to be used for shorter range forecasts at CPC around 2014. Figure 7 shows the impact that these modern methods had on the skill of the official 6-10 day temperature forecasts. Prior to 2014, temperature forecasts were primarily based on subjectively modified Klein specifications. Recent forecasts are based on a blend of bias corrected models output, ensemble regression, Klein specification and analogs. The older tools are lightly weighted and appear because they are the last remaining tools that specifically account for observation-based large-scale teleconnections.

The skill metric used for Fig. 7 is the 3-category Heidke Skill Score (Jolliffe and Stephenson, 2012) and ranges between -.5 and 1, with a skill score of 1 for a perfect forecast and 0 for predictions that are no better than random

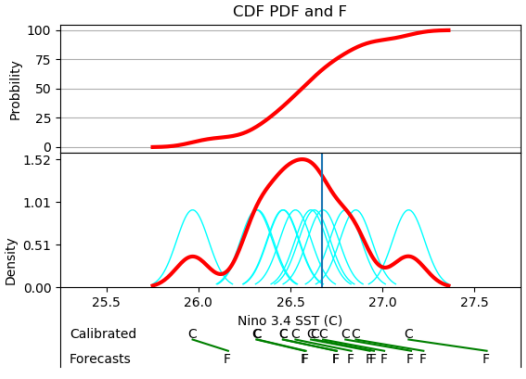

Fig 5. Same as Fig. 4 except for a PDF Correction.

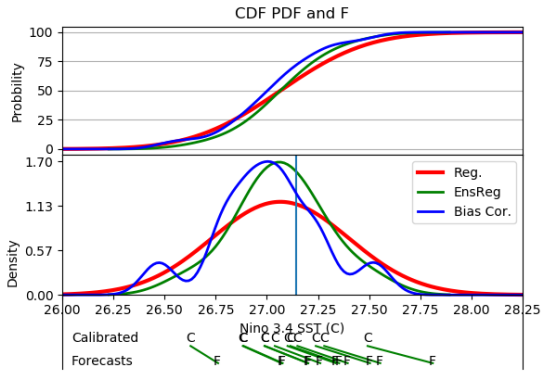

Fig. 6. Same as Fig. 4 except for ensemble regression, and PDF correction. The C values plotted are for ensemble regression.

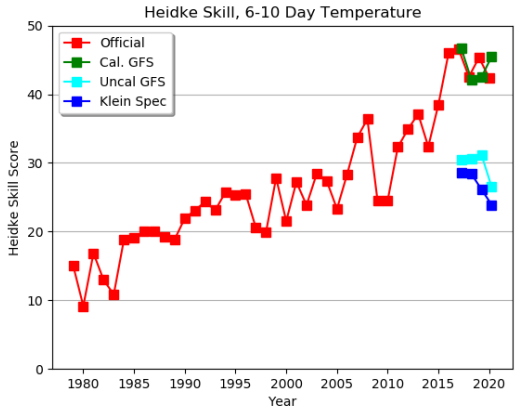

Fig. 7. Time series of CPC official 6-10 day temperature forecasts (Red) and major guidance tools.

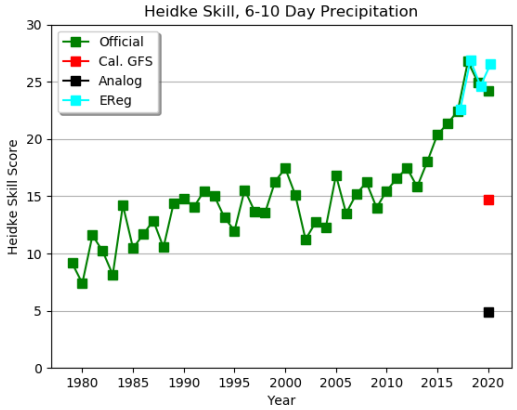

Fig. 8. Time series of CPC 6-10 day precipitation outlook (Green) together with major tools.

forecasts. A near discontinuous upward jump in skill in 2015 shows the improvement due to the newer post-processing methods based on calibrated model output. The skill of the Klein specification equations (blue) in recent years remained about the same as manual forecasts of the early 2000's when the specifications were the primary source of temperature guidance.

A simple bias correction of the NCEP Global Forecast System (GFS) ensembles (green) is just as skillful as both the official forecast, and the skill of the ensemble regression (ensemble regression skill is not shown because it is very similar to both the official and bias corrected GFS). Uncalibrated GFS (Cyan) is not competitive, being only slightly more skillful than the Klein specifications.

The improvement afforded by modern post-processing methods is even more evident in precipitation forecasts. Figure 8 shows the skill of the official precipitation forecasts (green). Scores plateaued following the introduction of ensemble model predictions in the 1990's remained at nearly the same levels until ensemble regression post-processing was introduced in 2015. The skill of the GFS model post-processed by ensemble regression is shown in cyan on this figure. In contrast to temperatures, simple bias correction of the GFS forecast (Cal GFS, red) is not competitive with precipitation forecasts from ensemble regression. The skill of analog forecasts (black) were never a competitive tool compared to the skill of the manual forecasts for precipitation. Before 2014 manual forecasts for precipitation were largely based on forecaster judgement using analogs and model forecast precipitation, so it is not surprising that the calibrated GFS forecast skill is about at the level of the official forecast prior to the introduction of ensemble regression.

5. CURRENT DEVOLOPMENTS

Current efforts to improve post-processing of extended range weather and climate forecasts in operational setting focus on three issues. First is the blending of information from a variety of models. Improvements in communications has enabled access to models run at a variety of weather forecasting centers worldwide. This led to considerable efforts in combining information from various numerical models. Most notable in these efforts is the efforts to provide integrated guidance from the climate models that form the National Multi-model Ensemble (NMME). (Becker et al., 2014).

Secondly, a focus on post processing methods that directly calibrate the probabilities from the ensemble forecasts to specific forecast elements. While the bias correction, percentile mapping, and ensemble regression frequently produce reliable probabilities, there is no guarantee of it. Probability Anomaly Correlations, PAC, (van den Dool et al., 2017) have successfully been used to improve the reliability extended range forecasts.

Finally, the methods discussed so far are point-by-point calibrations based on direct model output. These methods cannot correct biases and inaccuracies in the placement of mean synoptic scale features. Models can occasionally miss or distort major features, such as the teleconnection

patterns in upper level heights, or known patterns associated with ENSO, for example

Bayesian post-processing techniques (Wang et al. 2009) are being investigated to revise the output of climate models based on observations (such as observed or predicted ENSO states or the Madden-Julian Oscillation) or even to estimate weighting for different models in a multi-model ensemble. This method is used to incorporate information from other sources into climate and extended range forecasts. Bayesian methods have been successfully used to help revise climate model predictions (Strazzo, et al. 2019).

Figure 9 shows an example of the extensive use PAC calibration to produce a unified forecast info-rmation from a series of climate and statistical models used for seasonal predictions, The initial model calibration may be from percentile mapping, ensemble regression or Bayesian processing. These are then individually calibrated at each step in the consolidation process to improve reliability.

6. SUMMARY

Model post-processing for extended ranges has evolved from early efforts that were based on filtering daily weather systems to modern systems incorporate information from large forecast ensembles. Climate model post processing focus primarily on relatively simple calibration of model distributions. These are often close to a least-squared linear fit to the data. These techniques are now being used on the extended range weather forecasts. The shift from older methods to reliance on the direct calibrated model output resulted in notable increases in the skill of 6-10 day forecasts.

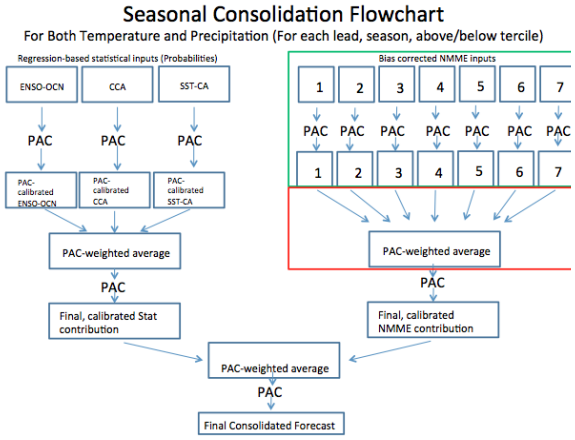

Fig. 9. Flow chart showing calibration and consolidation processing for seasonal outlook guidance.

Further improvements are provided by PAC calibration to improve reliability of the forecasts. Bayesian processing methods are also being investigated to revise model predicted distribution of forecast elements to better reflect information from observations such as teleconnection patterns associated with large scale climate anomalies.

7. REFERENCES

Andrews, James F., 1977: National Weather Service Inaugurates 6- to 10-Day Forecast Program. *Bull. Amer. Meteor. Soc,* **58**, 1304.

Becker, Emily and H. van den Dool, 2014: Probabilistic Seasonal Forecasts in the North American Multimodel Ensemble: A Baseline Skill Assessment. *J. Climate,* **29** 3015-3026

Candille, Guillem, S. Beauregard, and N. Gagnon, 2010: Bias Correction and Multiensemble in the NAEFS Context or How to Get a 'Free Calibration" through a Multiensemble Approach. *Mon. Wea. Rev.* **138**, 4268-4281.

Cui, Bo, Z. Toth, Y. Zhu, and D. Hou, 2012: Bias Correction for Global Ensemble Forecast, *Wea. Forecasting*, **27**, 396-410

Harnack, Robert P. 1986: Principles and Methods of Extended Period Forecasting in the United States. *Meteorological Monographs* National Weather Association, ISSN 0271-1044. 36pp.

Ji, Ming, A. Kumar and A. Leetmaa, 1994: A Multiseason Climate Forecast System at the National Meteorological Center. *Bull. Amer. Meteor. Soc*., **75**, 569-584.

Jolliffe, I. T., and D. B. Stephenson, 2012*: Forecast Verification: A Practitioners Guide in Atmospheric Science*. 2nd ed. Wiley, 292 pp.

Klein, W. H., B. M. Lewis, I. Enger, Objective Prediction of Five-Day Mean Temperatures during Winter. *J. Meteorology*, **16**, 672-682

Namias, Jerome 1943: Methods of Extended Forecasting. U.S. Department of Commerce, Weather Bureau, 65 pp.

Namias, Jerome, 1953: Thirty-Day Forecasting: a review of a ten-year experiment. Meteo-

rological Monographs, American Meteorological Society Vol. 2 No. 6.

Namias, J. and Collaborators, 1958: Application of Numerical Methods to Extended Forecasting Practices in the U.S. Weather Bureau. *Mon. Wea*. *Rev* **86,** 467-476.

O'Conner, James F.,1980: 30-Year Trend of the 5-Day Mean Temperature Forecast Skill. NMC Technical Attatchment no. 81 - 1.

Saha, S., S. Nadiga, C. Thiaw, and J. Wang, 2006: The NCEP Climate Forecast System, *J. Climate*, **19**, 3483-3517.

Saha. S., and Coauthors, 2014: The NCEP Climate Forecast System version 2*. J. Climate*, **27**, 2185-2208..

Silverman, B.W., 1986: *Density Estimation for Statistics and Data Analysis*, Chapman and Hall/CRC, 175 pp.

Strazzo, Sarah, D. Collins, A. Schepen, Q. J. Wang, E. Becker, and L. Jia, 2019: Application of a Hybrid Statistical Dynamical System to Seasonal Prediction of North American Temperature and Precipitation. *Mon. Wea. Rev*. **147**. 607-625.

Unger, D. A., H. van den Dool, E. O'Lennic, and D. Collins, 2009: Ensemble Regression. *Mon. Wea. Rev*., **13**7, 2365-2379.

van den Dool, Huug, E. Becker, L. Chen, and Q. Zhang: The Probability Anomaly Correlation and Calibration of Probabilistic Forecasts. *Wea. Forecasting*, **32**, 199-206.

Wagner, A. James, 1989: Medium- and Long-Range Forecasting. *Wea. Forecasting*. **4**. 413 - 426.

Wang, Q. J., D. E. Robertson, and F. H. S. Chiew, 2009: A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. *Water Resour. Res.,* **45.**