J11.5 HYDROLOGIC APPLICATIONS OF SHORT AND MEDIUM RANGE ENSEMBLE FORECASTS IN THE NWS ADVANCED HYDROLOGIC PREDICTION SERVICES (AHPS)

Mary Mullusky*, Julie Demargne, Edwin Welles, Limin Wu and John Schaake
Hydrology Laboratory, Office of Hydrologic Development
National Weather Service, NOAA, Silver Spring, Maryland

1. INTRODUCTION

Ensemble precipitation and temperature forcing data are needed to drive the NWS Ensemble Streamflow Prediction (ESP) system to produce hydrologic forecasts for the Advanced Hydrologic Prediction Services (AHPS). These forcing data must be derived from existing atmospheric forecast information using pre-processing procedures that are being developed to do the required re-scaling and downscaling to produce forcing data at the space and time scales needed by ESP hydrologic forecast models. Experimental procedures are being tested in pilot projects at four NWS River Forecast Centers (RFCs).

This paper presents the strategy being followed to develop pre-processing procedures that use existing operational weather and climate forecasts. It presents some example verification statistics for short range precipitation ensemble forecasts. And it highlights some of the issues that must be resolved to use operational Numerical Weather Prediction (NWP) ensemble forecasts as input to the ESP procedures being used to support AHPS.

2. ESP HYDROLOGIC FORECAST MODEL FORCING REQUIREMENTS

NWS hydrologic forecast models represent hydrologic processes that occur within elemental river basin areas from about 100 km^2 to as large as a few thousand km^2. Most existing forecast models operate using a 6-hour time step but the forecast system can support time steps as short as one hour. More detailed spatial information is now available to permit future reduction in both the size of elemental river basin areas and the computational time step.

The ESP hydrologic forecast system allows forecast models to be integrated starting at prescribed initial conditions. This integration accounts for the movement of water through a network of river basin segments that comprise a large river basin. Each integration produces a time series of streamflow and river stage at prescribed forecast points in the river basin network. Each integration produces a single ensemble forecast member. Multiple integrations produce the complete ensemble.

Hydrologic ensemble forecasts provide probabilistic information about events at each forecast point. They also provide joint probability information about simultaneous and lagged events at multiple forecast points. This is important because some potential applications of hydrologic ensemble forecasts, e.g. certain reservoir operations, may require hydrologic events at different places in a river basin to be mutually consistent within each individual ensemble member.

To produce the individual ensemble hydrograph members for multiple locations in a river basin requires a space-time pattern of future precipitation and temperature forcing. Data values are required for each sub-basin and for each time step represented by the model for the entire river basin and for the full future lead time that may be a few days or several months. The ensemble forcing must preserve the space-time variability characteristics appropriate for the future events being predicted. This is essential because a river basin acts as a natural space-time integrator of the atmospheric forcing. The variability of the hydrologic forecasts is highly dependent on the space-time structure of the atmospheric inputs.

The most important atmospheric forecast variables are precipitation and air temperature. The models also require future values of potential evaporation. In the mountains of the west, the freezing height also is needed.

3. OPERATIONAL WEATHER AND CLIMATE FORECAST PRODUCTS

A wide range of operational atmospheric forecasts products are available as potential inputs to the ESP pre-processor systems. In time, optional features will likely be available to support hydrologic application of most of these products.

The primary short-term precipitation forecasts used by NWS RFCs are produced by NCEPs Hydrometeorological Prediction Center (HPC). These are single value gridded forecasts for precipitation at:

---

* Corresponding author address: John Schaake, NWS/NOAA, Office of Hydrologic Development, OHD12, 1325 East_West Highway, Silver Spring,MD 20910; e-mail: john.schaake@noaa.gov

(i) 6-hour intervals for day-1 and day-2, (ii) a daily value for day-3 and (iii) a 2-day total for days 4 and 5.

The primary short-term temperature forecasts (up to 7 days) are single-value daily maximum and minimum MOS temperature forecasts from the Global Forecast System (GFS).

CPC produces a 6-10 day probabilistic forecast product that can be used to prescribe the distribution of precipitation and average temperature for the period. Similar precipitation and temperature products are issued by CPC for a lead time of 8-14 days as well.

Long range probabilistic precipitation and temperature outlooks are issued by CPC at the middle of each month. This includes a probabilistic outlook for next month's precipitation and temperature. It includes seasonal precipitation and temperature outlooks for 3-month periods beginning at monthly intervals for the next year.

Ensemble weather forecasting began at NCEP more than a decade ago. Now, a suite of operational ensemble forecast products is being introduced at NCEP. This includes a new operational Short Range Ensemble Forecast (Herr et al, 1002) for the next 2 days. It includes operational ensemble forecasts from the GFS for the next 16 days. And it includes a new long range coupled ocean, land atmosphere system for ensemble climate prediction. The 1-16 day GFS ensemble system will soon include ensemble forecasts from a fixed version of the Medium Range Forecast (MRF) model as well as additional ensemble forecasts from a more recent version of the GFS.

The ensemble from a fixed MRF model is especially important for AHPS applications because a climatology of ensemble forecasts has been produced by NOAA's Climate Diagnostics Center (CDC). This is needed to calibrate the re-scaling and down-scaling procedures in the ESP pre-processor procedures.

4. STRATEGY TO PRODUCE ENSEMBLE FORCING DATA FOR ESP

The strategy to produce ensemble forcing data for ESP is to build on existing capabilities and operational experience in a way that meets basic ESP requirements and that can be implemented at many RFCs quickly. The strategy is to keep procedures as simple as possible and to introduce complexity gradually and only as much as absolutely necessary. The goal is not to produce the perfect system but to make continuing practical improvements that can be supported in an operational forecasting environment.

The most important requirement for improvement to ESP pre-processor capabilities is for the short to medium range period for lead times of 1 to 16 days. The approach is (i) to make improvements at the shortest lead times first and (ii) to develop ensemble pre-processor procedures to use existing single-value forecasts from HPC and MOS before developing procedures to use the ensemble atmospheric forecast products. Some example results for ensemble forecasts using single value HPC forecasts are presented below in Section 5.

Once appropriate operational procedures to use HPC and MOS products are in place, procedures to use atmospheric ensemble forecasts will be introduced. Procedures for longer lead times (beyond 3-5 days) will be done first. Temperature forecast procedures will likely become operational before precipitation forecasts. At present, procedures to use atmospheric ensemble forecasts are very much in the research stage and are discussed briefly in Section 6 below. The existing ESP forecast system includes procedures to use all of the CPC probabilistic forecast products.

5. EXAMPLE VERIFICATION STATISTICS FOR ESP SHORT TERM ENSEMBLE FORCING

Forecast verification requires a large enough sample size so that verification statistics are reliable and not subject to large sampling variations. In the case of ensemble verification this means that real-time verification of ensemble procedures can only be done over large areas. Verification of ensemble procedures for specific locations can only be done over a period of time long enough to produce a large sample of the event to be verified.

This paper presents some verification statistics for ensemble precipitation forecasts for one of the short-term ensemble pilot project areas, the California Nevada River Forecast Center (CNRFC). The most important precipitation forecasts at CNRFC are for the winter season from mid October to mid April. Ensemble procedures for precipitation and temperature forecasts for lead times of 1 to 5 days have been developed and are ready for testing at the CNRFC.

These procedures use single value precipitation forecast from HPC. The procedures are calibrated using archived single value forecasts and corresponding observations. The forecasts and observations are used to define the joint distribution of forecasts and observations. The ensemble procedure is applied to a given future event by using the single value forecast for the event together with the estimated joint distribution of forecasts and observations to estimate the conditional distribution for precipitation associated with the single value forecast. (Refs)

In this section, example verification statistics are presented to illustrate that the forecast distributions of precipitation agree well with the observations for the events that actually occurred. All of the verification statistics presented are based on retrospective forecasts for 21 representative locations throughout the state of California. Results are presented for several verification statistics. Separate verification statistics are presented for sets of events that occurred in the 2001 and 2002 water years. One question being investigated is how much data are needed to calibrate the parameters. To test the effect of the amount of historical data used to calibrate the procedures, ensemble forecasts for each of these sets were made using parameter sets from 3 different periods of historical data. The calibration periods have increasing durations, all beginning in 1997 at the start of the CNRFC forecast data archive. The three calibration periods end in 2000, 2001 and 2002 respectively. Note that the verification statistics for events predicted using the 2000 parameter sets are a true verification of the procedures; the events being verified were not used to estimate the parameters.

The first verification statistic is the Nash-Sutcliffe efficiency that measures how well the ensemble mean agrees with the observed values. Values of the Nash-Sutcliffe efficiency for 2001 and 2002 events are shown in Figures 1a and 1b, respectively. This statistic is a skill score that equals 1.0 for perfect forecasts, that is positive for forecasts that are better than climatology and negative, otherwise. The efficiency decreases substantially during the 5-day forecast period but there is some skill remaining for 24 hr precipitation forecasts on day-5. There is some sensitivity to the values of the parameters used. But it appears there is an adequate amount of data, even for the 2000 parameter values that are based on 3-4 years of data. Results for 2002 in Figure 1b are similar to those for 2001 in Figure 1a.

Each ensemble forecast includes an estimate of the probability of precipitation (POP) and an estimate of the distribution of precipitation amount if precipitation occurs. The mean value of this conditional distribution was compared to the corresponding observed value of precipitation for events when precipitation actually occurred. The Nash-Sutcliffe efficiency statistics for the conditional mean precipitation are given for 2001 and 2002 events in Figures 2a and 2b, respectively. The skill in the ensemble forecasts of precipitation amount, given that precipitation occurs is much less than the skill in the unconditional ensemble mean given in Figure 1. Beyond the first 2 days the ensemble forecasts of the conditional distribution of precipitation are only slightly less skillful than climatology.

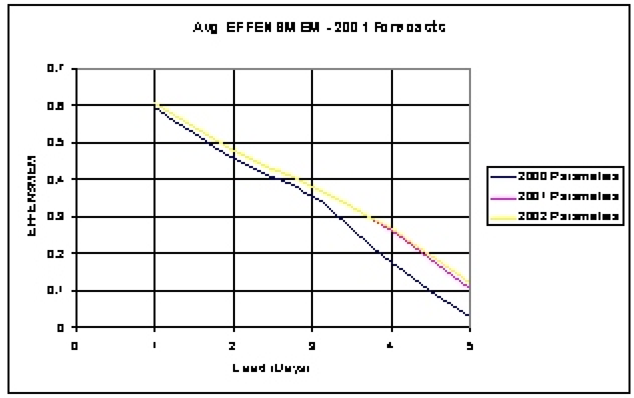

Figure 1a - Nash-Sutcliffe Efficiency for 2001 Forecasts

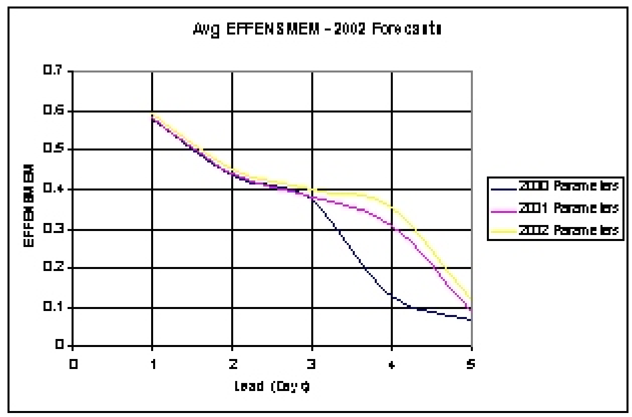

Figure 1b - Nash-Sutcliffe Efficiency for 2002 Forecasts

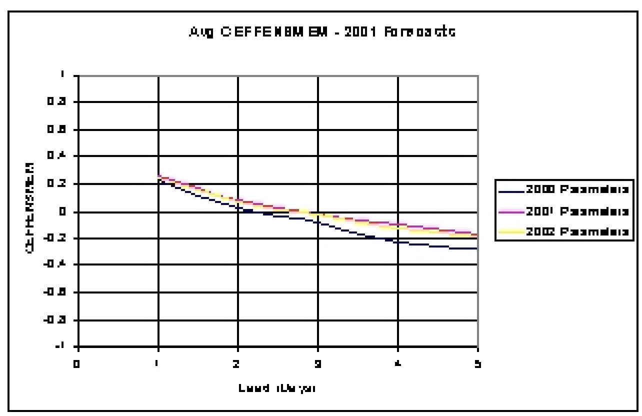

Figure 2a - Nash-Sutcliffe Efficiency for Conditional Mean of 2001 Forecasts

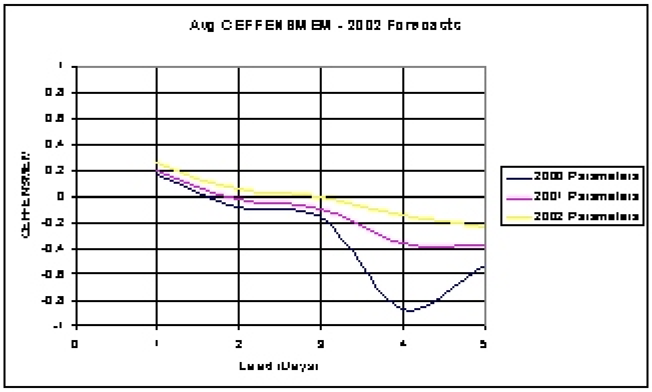

Figure 2b - Nash-Sutcliffe Efficiency for Conditional Mean of 2002 Forecasts

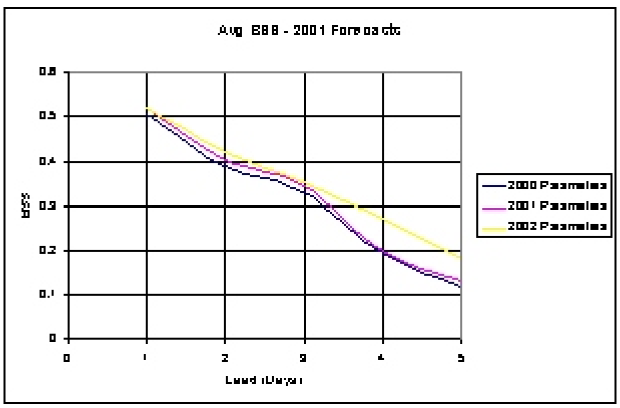

Figure 3a - Brier Skill Score for Probability of Precipitation 2001 Forecasts

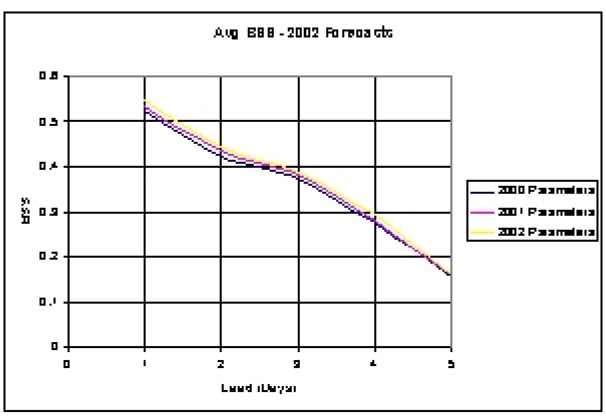

Figure 3b - Brier Skill Score for Probability of Precipitation 2002 Forecast

The skill in ensemble forecasts of the probability of precipitation can be measured using the Brier Skill Score (BSS). Values of the BSS for 2001 and 2002 events are presented in Figures 3a and 3b

The reliability of the probability of precipitation forecasts can be measured using a root mean square error statistic. This statistic is estimated by partitioning all of the forecasts into 4 quartiles of POP forecast The fraction of forecasts for times when precipitation occurs is compared to the average POP forecast for each quartile and is used to compute the rms error. The results are shown in Figures 4a and 4b for 2001 and 2002 events respectively. The results suggest that the POP forecasts have rms errors less than about 15 percent and that the results are not sensitive to the amount of data used to calibrate the procedures.

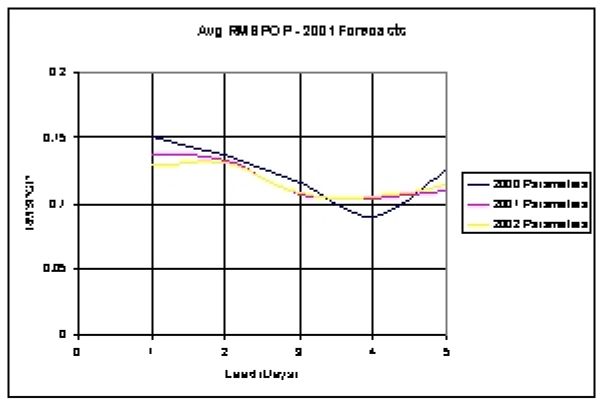

Figure 4a - RMS Error of Probability of Precipitation 2001 Forecasts

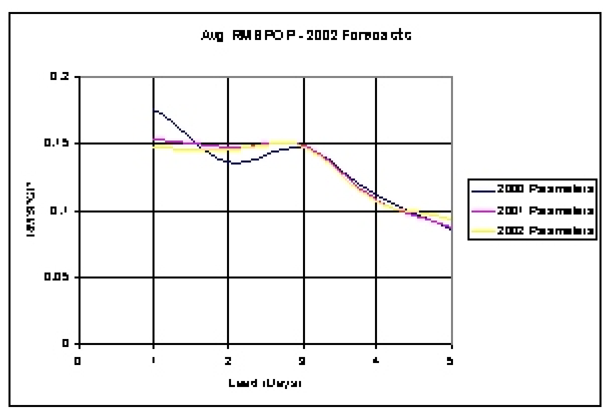

Figure 4b - RMS Error of Probability of Precipitation 2002 Forecasts

The reliability of the probability forecasts of precipitation amount given by the ensemble conditional distribution can be measured by analyzing the forecast exceedence probability associated with each observed precipitation event. If the forecast probabilities are reliable, the distribution of exceedence probabilities will have a uniform distribution. The rms difference between the distribution of exceedence probabilities and the uniform distribution is a measure of reliability of the conditional probability forecasts of precipitation amount. The results are shown in Figures 5a and 5b for 2001 and 2002 events respectively. These figures show that the conditional probability forecasts are reliable and not sensitive to the amount of data used to calibrate the procedures

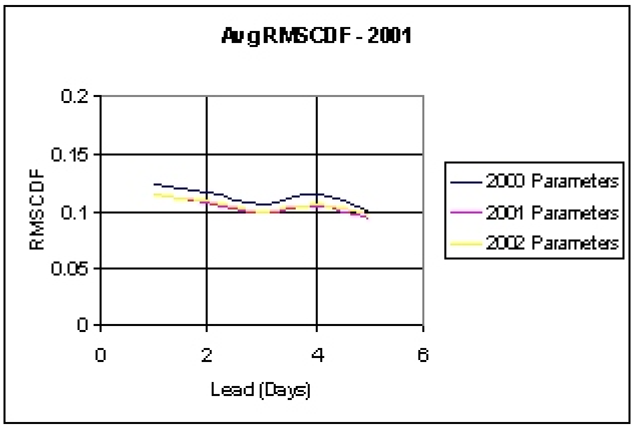

Figure 5a - RMS Error of 2001 Conditional Probability of Precipitation Amount Forecasts

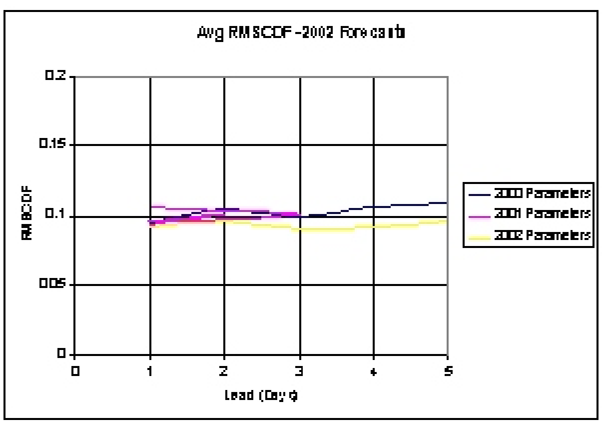

Figure 5b - RMS Error of 2002 Conditional Probability of Precipitation Amount Forecasts

6. FUTURE APPLICATIONS OF ATMOSPHERIC ENSEMBLE FORECASTS

Atmospheric ensemble forecasts hopefully can be used in the future to provide forcing data for AHPS ESP forecasts. But this requires development of reliable procedures to remove local biases in the atmospheric forecasts. Additional procedures to compensate for the tendency of atmospheric models to underestimate the spread (i.e. uncertainty) in the distribution of events that might occur are needed as well. The main problem to remove bias is to have enough historical forecasts and observations to define the climatology of the forecast system and to estimate the local parameters of bias removal techniques.

The most difficult issue in using the atmospheric ensemble forecasts is to compensate for the error in the ensemble spread. This error occurs because atmospheric ensemble forecast models do not account for all of the important sources of uncertainty. To illustrate this issue, Figure 6 presents results for 4 different approaches to making adjustments to ensemble forecasts from the NWS GFS. Three years of ensemble precipitation forecasts were analyzed at several locations throughout the U.S. Analysis results for July for a grid element in the southeast U.S. were used for Figure 6. The forecast exceedence probabilities for the observed precipitation amounts were partitioned into quartiles to produce the Talagrand diagram shown in Figure 6. The first adjustment approach was to use the raw ensemble data values without adjustment. The second was to multiply each ensemble value by a constant so that the average ensemble mean was the same as the average observation. The third approach modified the cumulative distribution of ensemble members so that it was the same as the cumulative distribution of observations. The fourth approach was to apply the short-term ensemble procedures being used to in the AHPS ESP pilot projects to the joint distribution of observations and ensemble mean values from the GFS ensemble forecasts. The ensemble mean value from the GFS is used as a surrogate for the HPC single-value forecast.

If probability forecasts from the ensemble were reliable, each bar in Figure 6 would have a value of 0.25. The raw ensemble forecasts tended to occur in just one of the 4 quartiles, a result of bias in the ensemble members. All of the other approaches remove this bias. But neither the second nor third approach that remove bias can correct the problem associated with the ensemble spread.

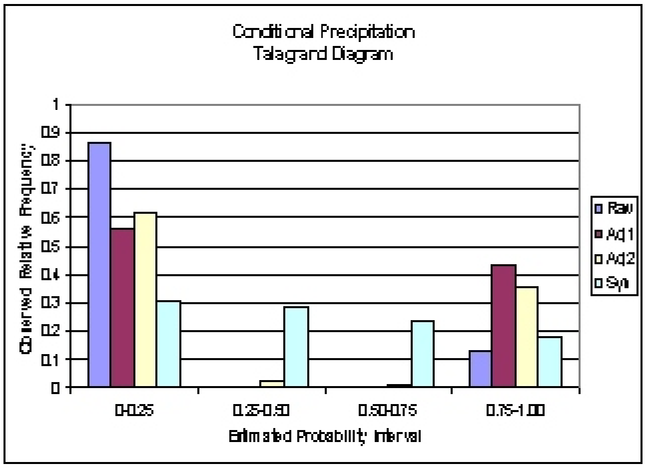

Figure 6 - Talagrand Diagram for 4 Different Approaches to Adjusting NWS GFS Ensemble Forecasts

Because the spread is underestimated observed precipitation values do not lie within the main part of the forecast probability distribution. As a result they appear in the first and fourth quartiles for adjustment approaches 2 and 3. The fourth adjustment approach is based on the short-term ensemble procedure initially being used in AHPS. This approach did a much better job of compensating for the GFS underestimate of the ensemble spread. Almost equal fractions of the observations occurred in each of the quartiles.

7. REFERENCES

Herr, H., E. Welles, M. Mullusky, L. Wu, J. Schaake, 2002, "Simplified Short Term Precipitation Ensemble Forecasts: Theory," Preprints of the Symposium on Observations, Data Assimilation and Probabilistic Prediction, 82nd AMS Annual Meeting, Orlando, Florida.

Tracton, M. S., J. Du, Z. Toth, and H.-M. Juang, 1998: Short Range Ensemble Forecasting (SREF) at NCEP/EMC. *Preprint 12th Conference on Numerical Weather Prediction* , 78th AMS Annual Meeting, January 11-16, 1998, Phoenix, Arizona, 269-272.