Diagnostic Verification of 6-90 Day Ensemble Streamflow Predictions for AHPS

Principal Investigators:

A. Allen Bradley[1], Stuart S. Schwartz[2], Anton Kruger[1]

[1]IIHR–Hydroscience & Engineering
The University of Iowa

[2]Center for Urban Environmental Research and Education (CUERE)
University of Maryland-Baltimore County (UMBC)

Progress Report for 03/01/2005 - 08/31/2005

Project Objective

For water resource managers, ensemble streamflow predictions represent one of the most significant products of the National Weather Service's (NWS) Advanced Hydrologic Prediction Services (AHPS). This project seeks to advance a distributions-oriented (DO) framework for verification of probability distribution forecasts derived from ensemble streamflow predictions. DO forecast quality measures provide a consistent diagnostic framework to quantify the relative sources of forecast skill, which would allow water managers to match decision tools to forecast attributes, and enable forecasters to target research, resources, and development efforts to the most valuable improvements. Expected outcomes from this research include: (1) a consistent framework for verifying probability distribution forecasts, which will be demonstrated through the evaluation and comparison of forecast quality of 6-90 day NWS AHPS ensemble streamflow forecasts for the North-Central and Ohio River Forecast Centers, and (2) a set of diagnostic verification tools for elucidating relevant forecast quality attributes, for the management and targeted improvement of forecasts systems, and interpretation of forecasts for their operational use.

Progress Report

Our efforts during this period have focused on (1) the development and assessment of summary diagnostic measures for forecast intercomparisons, and (2) the generation of retrospective forecast verification data sets.

Diagnostic Summary Measurements for Intercomparison of Forecasts

The DO verification approach produces extremely detailed information on the quality of the ensemble forecasts for a single forecast date, a single forecast product, and at a single location. However, effective ways to summarize this information are needed to compare the quality of forecast issued on different dates, or for different forecasts products, or for different forecast locations.

We have extended DO verification concepts to develop summary measures to facilitate such comparisons [Bradley et al. 2006]. For instance, we have found (by derivation) that the ranked probability skill score (*RPSS*), a common verification metric used for multicategory probability forecasts, is equal to:

$$RPSS = \sum_i w_i SS(p_i) \tag{1}$$

where $SS(p_i)$ is the MSE (or Brier) skill score for a flow threshold with a nonexceedance probability of p_i, and w_i is a weighting factor (that depends only on p_i). Note that the detailed DO approach provides an estimate of the function $SS(p_i)$ for any and all values of p_i; the RPSS can therefore be interpreted as a weighted-average skill score $\overline{SS}$ over the range of flow thresholds. Furthermore, using a DO decomposition of the skill score $SS(p_i)$, we can use equation (1) to find weighted-average measures of the components, or:

$$RPSS = \overline{SS} = \overline{PS} - (\overline{SME} + \overline{SREL}) \tag{2}$$

where $\overline{PS}$ is the potential skill, $\overline{SME}$ is the standardized mean error (unconditional bias), and $\overline{SREL}$ is the slope reliability (a measure of conditional bias).

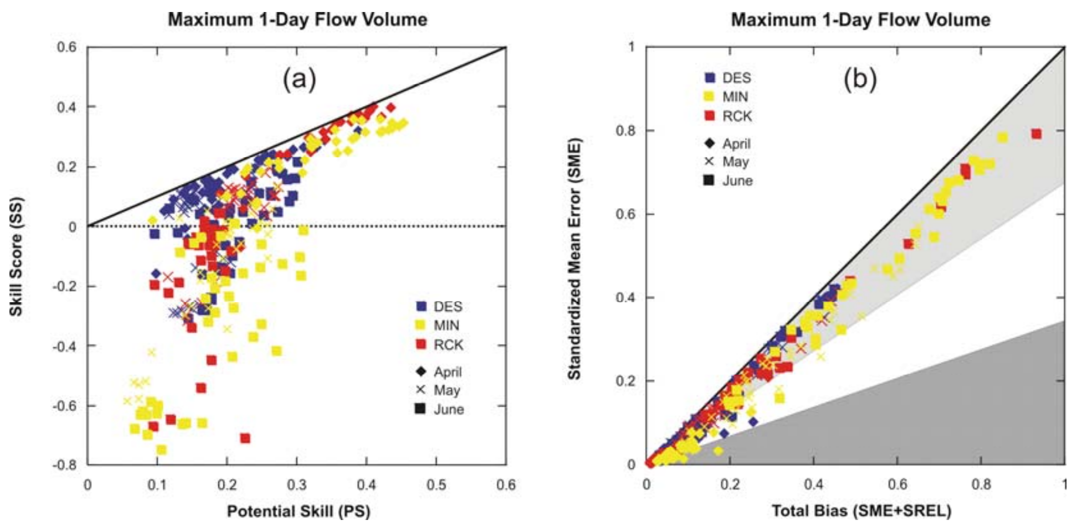

Figure 1: Intercomparison of summary forecast quality measures for maximum 1-day flow volume forecasts: (a) average skill score (*SS*) versus the average potential skill (*PS*), and (b) average standardized mean error (*SME*) versus the average total bias. Results are shown for mainstream forecast segments for the Des Moines (DES), Minnesota (MIN), and Rock River (RCK) basins, for forecast issued from April through June.

Figure 1 illustrates how these summary measures can be used for forecast intercomparison and diagnostic interpretation. Figure 1(a) compares the average skill $\overline{SS}$ with the potential skill $\overline{PS}$ of maximum 1-day volume forecast for mainstem sites in the Des Moines, Minnesota, and Rock River forecast groups. Overall, the best forecasts are those from the Minnesota and Rock River for April; the forecasts for the Des Moines tend to have lower average skill $\overline{SS}$, and the skill tends to decline later into the season. For the same set of forecasts, Figure 1(b) compares the unconditional biases ($\overline{SME}$) with the total biases ($\overline{SME} + \overline{SREL}$). Clearly, unconditional biases dominate. This has significant implications for operational forecasting. In our recent study [Hashino et al. 2006], we found that ensemble bias-correction procedures improve forecast skill by eliminating unconditional biases (*SME*); however, conditional biases (*SREL*) are largely unaffected, and would require a more extensive post hoc calibration to eliminate biases. Therefore, the inference based on Figure 1(b) is that ensemble bias correction would be sufficient to correct most of the biases observed in the maximum 1-day flow volume forecasts.

Generation of Retrospective Forecast Verification Data Sets

We have developed programs and scripts, which work within the architecture of the operational NWSRFS system, to generate archives of forecasts and verification products for assessing ensemble streamflow predictions. With the help of the North Central River Forecast Center (NCRFC), the NWSRFS system and model configuration for the NCRFC was installed on a Linux machine at IIHR. The system has since been used to generate retrospective forecasts, for 1950 through 1999, for the majority of forecast groups within the NCRFC.

The generation of the archive has three components. First, for a selected forecast group, a set of scripts control the creation the hypertext command language (HCL) for extended streamflow forecast (ESP) simulations for a forecast date (e.g., April 1st) for the historical period (1950-1999), and then run NWSRFS to produce ensemble traces. The process is repeated at weekly forecasts intervals, producing an archive of ESP traces in their native format for the 52 forecast dates per year. Next, the traces are processed to generate a suite of ensemble forecast products, including those routinely issued operationally (e.g., weekly flow volumes, maximum flows). To facilitate verification analyzes, we combine an ensemble forecast with its corresponding observation by accessing the observed flow data for the forecast segment contained within the *calb* subsystem of NWSRFS. Finally, we apply the DO verification methods described by Bradley et al. [2004] to generate archive information and data plots that characterize the skill and identify systematic biases for the ensemble streamflow forecast products. Because of the computational time required, the generation of archives of retrospective forecasts for the NCRFC is an ongoing activity

References

Bradley, A. A., S. S. Schwartz, and T. Hashino, Distributions-oriented verification of ensemble streamflow predictions, *Journal of Hydrometeorology*, 5(3), 532-545, 2004.

Hashino, T., A. A. Bradley, and S. S. Schwartz, Evaluation of bias-correction methods for ensemble streamflow volume forecasts, *Hydrology and Earth System Sciences Discussions*, 3, 561-594, 2006.

Bradley, A. A., S. S. Schwartz, and A. Kruger, Interpretation of summary verification measures for ensemble forecasts, *Eos Transactions AGU*, 87(36), Joint Assembly Supplement, Abstract A24A-05, 2006.