JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION

Vol. 48, No. 4 AMERICAN WATER RESOURCES ASSOCIATION August 2012

GENERATION OF ENSEMBLE STREAMFLOW FORECASTS USING AN ENHANCED VERSION OF THE SNOWMELT RUNOFF MODEL[1]

Brian J. Harshburger, Von P. Walden, Karen S. Humes, Brandon C. Moore, Troy R. Blandford, and Albert Rango[2]

ABSTRACT: As water demand increases in the western United States, so does the need for accurate streamflow forecasts. We describe a method for generating ensemble streamflow forecasts (1-15 days) using an enhanced version of the snowmelt runoff model (SRM). Forecasts are produced for three snowmelt-dominated basins in Idaho. Model inputs are derived from meteorological forecasts, snow cover imagery, and surface observations from Snowpack Telemetry stations. The model performed well at lead times up to 7 days, but has significant predictability out to 15 days. The timing of peak flow and the streamflow volume are captured well by the model, but the peak-flow value is typically low. The model performance was assessed by computing the coefficient of determination (R^2), percentage of volume difference (Dv%), and a skill score that quantifies the usefulness of the forecasts relative to climatology. The average R^2 value for the mean ensemble is >0.8 for all three basins for lead times up to seven days. The Dv% is fairly unbiased (within ±10%) out to seven days in two of the basins, but the model underpredicts Dv% in the third. The average skill scores for all basins are >0.6 for lead times up to seven days, indicating that the ensemble model outperforms climatology. These results validate the usefulness of the ensemble forecasting approach for basins of this type, suggesting that the ensemble version of SRM might be applied successfully to other basins in the Intermountain West.

(KEY TERMS: snow hydrology; water supply; surface water hydrology; quantitative modeling; ensemble streamflow forecasting.)

Harshburger, Brian J., Von P. Walden, Karen S. Humes, Brandon C. Moore, Troy R. Blandford, and Albert Rango, 2012. Generation of Ensemble Streamflow Forecasts Using an Enhanced Version of the Snowmelt Runoff Model. *Journal of the American Water Resources Association* (JAWRA) 48(4): 643-655. DOI: 10.1111/j.1752-1688.2012.00642.x

INTRODUCTION

Realistic and accurate streamflow forecasts are essential for water resources planning and management. In many regions, agricultural, municipal, and environmental water uses increase demands on limited freshwater resources. In the western United States (U.S.), the planning and management of water resources is particularly challenging because water

[1]Paper No. JAWRA-10-0191-P of the *Journal of the American Water Resources Association* (JAWRA). Received November 2, 2010; accepted December 19, 2011. **Discussions are open until six months from print publication**.

[2]Respectively, Researcher (Harshburger), Aniuk Consulting, LLC, 7105 61st Avenue, Kenosha, Wisconsin 53142; Associate Professor (Walden), Professor (Humes), and Researcher (Moore), Department of Geography, University of Idaho, Moscow, Idaho; Researcher (Blandford), Department of Water Resources, State of Montana, Helena, Montana; and Research Hydrologist (Rango), USDA, Agricultural Research Service, Jornada Experimental Range, Las Cruces, New Mexico (E-Mail/Harshburger: brian.harshburger@aniukconsulting.com).

supplies are often derived from runoff due to snowmelt. As a result, knowledge of the timing, rate, and volume of snowmelt is crucial for the management of water resource systems, which are designed for purposes such as irrigation, wildlife management, recreation, flood control, and hydroelectric power generation.

Probabilistic streamflow forecasting provides users (e.g., reservoir operators) with information regarding the uncertainty or range of forecasts over a given time period. The goal is to provide the best, most accurate forecast possible, while minimizing the uncertainty commonly associated with input data and streamflow forecast models (Singh and Singh, 2001). In recent decades, considerable efforts have been devoted to improving probabilistic streamflow forecasts for lead times of days, weeks, and months. This has led to the development of ensemble streamflow prediction systems, which generate a set of forecasts that are intended to represent the range of possible streamflow values. In an early attempt, Day (1985) used weather inputs from historical climate records to generate ensemble streamflow predictions. More recently, ensemble inputs obtained from numerical weather prediction models, which have been found to be useful in the short-to-medium range (1 to 15 days) (Buizza *et al.*, 1999; Mullen and Buizza, 2001; Clark *et al.*, 2004; Hamill *et al.*, 2004), have been used to generate such forecasts (e.g., Wigmosta *et al.*, 1994; Westrick *et al.*, 2002; Clark and Hay, 2004). Weather forecast ensembles are input into a hydrologic model and a number of possible predictions are generated, each with an assigned probability of occurrence. This approach is advantageous when compared with traditional forecasting methods (single forecast only), which provide no indication of how likely a forecast is to be correct.

The scope of this study is to evaluate an ensemble prediction system for short-to-medium range streamflow forecasts (1 to 15 days in advance) for three basins in the state of Idaho in the U.S. Idaho is located in a region called the Intermountain West, which is roughly defined here as being the region between the Cascade and Sierra Nevada ranges and the Rocky Mountains and, thus, includes portions of Idaho, Utah, Nevada, Arizona, and Colorado. The hydrologic model used in this study is the snowmelt runoff model (SRM), which was designed to simulate and forecast streamflow in mountainous areas where snowmelt is the major contributing factor to runoff (Martinec *et al.*, 1994; Mitchell and DeWalle, 1998). SRM was used in this study because it has been successfully tested in numerous mountainous watersheds around the world (e.g., Martinec, 1975; Shafer *et al.*, 1982; Dey *et al.*, 1989; Rango and Katwijk, 1990; Mitchell and DeWalle, 1998; Nagler *et al.*, 2000; Gomez-Landesa and Rango, 2002; Hong and Guodong, 2003) and is a widely accessible model with minimal data requirements. Two enhancements were made to the model by Harshburger *et al.* (2010): (1) the use of an antecedent temperature index method to track snowpack cold content and determine when the snowpack is ripe (snowpack cannot retain any more liquid water), and (2) the use of both maximum and minimum critical temperatures to partition precipitation into rain, snow, or a mixture of rain and snow. Harshburger *et al.* (2010) tested the enhancements in three mountainous basins located in Idaho and found that they improved upon the original version of the model.

SRM has been used previously to produce single-deterministic streamflow forecasts (e.g., Rango and Martinec, 1994; Nagler *et al.*, 2000; Harshburger *et al.*, 2010). Nagler *et al.* (2008) used SRM to generate ensemble streamflow forecasts for the Ötztal drainage basin located in the Austrian Alps. Forecasts were generated for two snowmelt seasons at lead times of one to six days. The forecast results revealed that there was a good agreement between the ensemble streamflow predictions and the observed streamflow. However, quantitative statistics were not used to assess the usefulness of the ensemble predictions.

The objective of the work presented here is to produce and evaluate short-to-medium range ensemble streamflow forecasts (1 to 15 days) using an enhanced version of SRM. Streamflow forecasts are evaluated for three mountainous basins located in Idaho for four snowmelt seasons (2003-2006), the same basins used by Harshburger *et al.* (2010). The three basins modeled in this research are typical of many mid-elevation basins throughout the Intermountain West, in terms of physiographic characteristics (area, elevation range), as well as the availability and density of surface data used in the analysis. Therefore, this modeling framework may be applicable to other basins within the Intermountain West.

ENSEMBLE PREDICTION SYSTEM

Figure 1 shows a flow diagram for the ensemble forecasting system using the enhanced version of SRM. Below is a brief description of SRM, followed by descriptions of the input data used in the ensemble forecasting system, including the downscaled temperature and precipitation forecasts, as well as the satellite imagery used to obtain estimates of snow-covered area (SCA). SRM can be obtained at

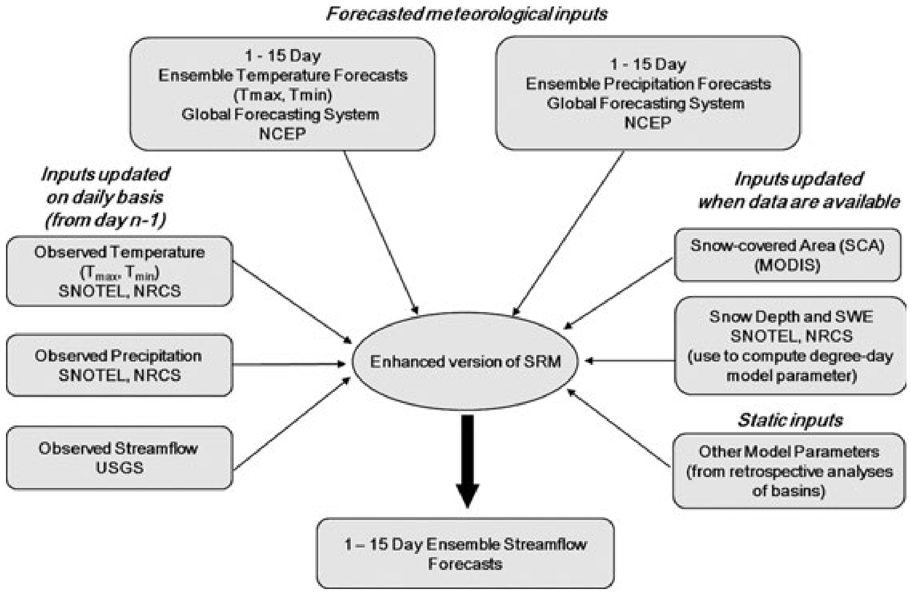

FIGURE 1. Flow Diagram for the Ensemble Forecasting System Using the Enhanced Version of the Snowmelt Runoff Model (SRM). The acronyms in this figure are (starting from the box in the upper left and proceeding clockwise): T_{max}, maximum daily temperature; T_{min}, minimum daily temperature; NCEP, National Center for Environmental Prediction; SCA, snow-covered area; MODIS, Moderate Resolution Imaging Spectroradiometer; SWE, snow water equivalent; SNOTEL, Snowpack Telemetry; NRCS, National Resources Conservation Service; USGS, United States Geological Survey.

http://hydrolab.arsusda.gov/cgi-bin/srmhome, or by contacting Al Rango (alrango@nmsu.edu) or Ralph Roberts (ralph.roberts@ars.usda.gov) directly. The enhanced version of SRM can be obtained by contacting Brian Harshburger (brian.harshburger@aniukconsulting.com).

Model Description

The SRM is a conceptually based, degree-day (temperature index) model (Martinec *et al.*, 1994). Model input variables include daily average air temperature, daily total precipitation, and SCA. As SRM is a semidistributed model, these variables are distributed among elevation zones (each with approximately 500 m of relative relief). The following equation is used in SRM to simulate the daily streamflow discharge Q (m^3/s):

$$Q_{n+1} = Q_n k_{n+1} + (1 - k_{n+1}) f \sum_i (c_{Si,n} \times a_{i,n} (T_{i,n} + \Delta T_{i,n}) S_{i,n} + c_{Ri,n} \times P_{i,n}) A_i, \tag{1}$$

where n is the day number, i is the index for each elevation zone, and f is a conversion factor (cm km^2/day to m^3/s). The recession coefficient, k, is the proportion of daily melt water production that immediately appears as runoff (Martinec *et al.*, 1994), and corresponds to the ratio of runoff on consecutive days without snowmelt and rainfall. Snowmelt and rainfall contributions are calculated separately for each elevation zone (area of A), and require the following input variables and parameters: T (°C/day), the number of degree-days, the temperature-lapse-rate adjustment ΔT (°C), the precipitation P contributing to runoff (cm), the fraction of SCA S, the degree-day factor a (cm/°C/day), and the runoff coefficients for snow and rain (c_{S} and c_{R}), which represent the difference between the available water volume and the outflow from the basin. A description of the model enhancements, along with results from model simulations used to test the enhancements can be found in Harshburger *et al.* (2010). The parameters used in the model are also described, along with the methods that are used to compute them.

Model Inputs

Temperature and Precipitation. Ensemble forecasts of temperature and precipitation are obtained from the Global Forecasting System (GFS) model (2.5 degree grid cells) produced by the National Center for Environmental Prediction. Due to the coarse spatial resolution of the GFS forecast data, the forecasted values are downscaled to the locations of Snowpack Telemetry (SNOTEL) stations, located within or surrounding each basin. Seven SNOTEL stations were used for the Big Wood Basin and five stations each for the South Fork of the Boise and North Fork of the Clearwater Basins. SNOTEL data were provided by the Natural Resources Conservation Service (NRCS).

The downscaling method used here closely follows the work of Clark and Hay (2004) and Clark *et al.* (2004). Here, we briefly describe the aspects of the method that are important for this study. The downscaling process uses historical forecast data from the GFS to assess the statistical relationship between weather forecast variables and observed temperature and precipitation values from the SNOTEL stations (Clark *et al.*, 2004; Moore, 2005). GFS forecasts are generated every 12 h and extend out 1 to 15 days in advance (Hamill *et al.*, 2004). Each forecast initialization uses 15 different initial conditions from which 15 meteorological forecast ensemble members are derived; however, only the control forecast (the average of the 15 ensemble members) is used to determine the regression coefficients used in the downscaling process. Seven forecast variables are used, including, 2-m air temperature, precipitation, 700-millibar relative humidity, sea-level pressure, 10-m meridional and zonal wind components, and total column precipitable water. These variables have been previously found to be important predictor variables for downscaling temperature and precipitation in the contiguous U.S. (Clark and Hay, 2004) and verified for Idaho by Moore (2005).

Using the technique outlined by Clark and Hay (2004), multiple-linear regression with forward selection is used to downscale the temperature and precipitation forecasts to the location of SNOTEL sites located within or surrounding each basin. Unique regression equations are generated for each SNOTEL site, variable (temperature and precipitation), month (March-July), and forecast lead time (30 lead times extending out 15 days). These equations are based on the seven GFS variables from the three nearest consecutive 12-h time steps. This results in 21 predictors (7 variables at 3 time steps) each for both temperature and precipitation. In addition to the multiple linear regression method described above, logistic regression with forward selection is used to estimate the probability of precipitation occurrence (Clark and Hay, 2004). The coefficients are determined by training the regression equations (multiple linear regression and logistic regression) on a subset of the data (i.e., 1995-2001); the remainder of the data is then used for validation.

Once the regression coefficients have been determined, they can be applied to real-time GFS ensemble forecasts to obtain ensemble forecasts of temperature and precipitation. The coefficients are applied to each of the 15 forecast ensembles and all 30 forecast lead times. This results in 15 ensemble forecasts of temperature and precipitation amount that extend out 15 days (30 lead times, each with a length of 12 h). Once the precipitation forecasts have been computed, the logistic regression coefficients obtained during the downscaling process are applied to estimate the probability of precipitation occurrence. Using the method described by Clark *et al.* (2004), random numbers are then generated from a uniform distribution (between 0 and 1) for each ensemble and forecast lead time. If the probability of precipitation occurrence is less than the random number, we assume that there is no precipitation. However, if the probability of precipitation occurrence is greater than the random number, we assume that the precipitation will occur and the amount determined from the multiple regression is used in the forecast.

Because SRM runs on a daily time step, the 15 temperature and precipitation ensemble forecasts (30 lead times, extending out 15 days) are converted to daily forecast values. This is accomplished by temporally matching the forecasts with the SNOTEL observations. GFS forecasts are generated at both 1200 UTC (05:00 h Mountain Standard Time; MST) and 0000 UTC (17:00 h MST) and are valid for the previous 12-h period. The analysis of hourly SNOTEL data records indicates that the minimum daily temperature ($T_{\min}$) typically occurs between 17:00 h and 05:00 h, and the maximum temperature ($T_{\max}$) typically occurs between 05:00 h and 17:00 h. Therefore, the downscaled $T_{\min}$ ensemble forecasts are calculated using forecast lead times 1, 3, 5, … , 29, and the downscaled $T_{\max}$ ensemble forecasts are derived from forecast lead times 2, 4, 6, …, 30. As a result, daily $T_{\min}$ and $T_{\max}$ ensemble forecasts are computed out to 15 days. The forecasted ensemble values $T_{\max}$ and $T_{\min}$ are then converted to forecast values of daily average temperature (15 ensembles) by simply averaging them. The ensemble precipitation forecasts are converted to daily values by taking the sum of the precipitation forecasts obtained for each of the two time steps for a given day. For example, the precipitation forecast for Time step 1 is added to the forecast for Time step 2.

Coherence between the temperature and precipitation forecasts is achieved through the use of the "Schaake shuffle" as described by Clark *et al.* (2004). The Schaake shuffle essentially integrates the coherence in the historical record into the forecast ensembles. Time-series historical daily-average temperature values (extending out 15 days) from the SNOTEL sites used in the downscaling process are collected so as to lie within seven days before and seven days after the forecast date; dates can be selected from all years in the historical record except from the year that is being forecasted. This process is completed separately for each of the 15 forecast ensembles; however, the same historical dates are used for each station. The historical daily-average temperature values are then sorted from lowest to highest. In addition, the daily-average temperature forecast ensemble members are also sorted from the lowest to highest. The sorted historical data are replaced with the sorted ensemble forecasts, and then resorted by (historical) year. For example, if the first year in the historical time series (say 1979) had the 20th highest temperature, then the first temperature ensemble member would be the ensemble with the 20th highest temperature. The corresponding precipitation forecast ensemble would then be used for Ensemble #1. This preserves the observed correlation between temperature and precipitation for the ensemble members (Clark *et al.*, 2004). The downscaling process results in 15 forecast ensembles of daily-average temperature and precipitation for each SNOTEL station and each of the 15 forecast lead times (days). A 16th ensemble member (for temperature and precipitation) is then created by taking the average of the 15 ensemble members described above. This ensemble member is now referred to as the "mean" ensemble.

Finally, the daily-average temperature ensemble forecasts for each station and lead time are averaged to create a synthetic station and are extrapolated to the hypsometric mean elevation of each elevation zone using monthly mean lapse rates from Blandford *et al.* (2008). This process is completed separately for each ensemble member. For example, Temperature Ensemble #1 from Station #1 is averaged with Temperature Ensemble #1 from the remaining stations. The elevation of the synthetic station is the mean elevation of all of the SNOTEL stations used to model each basin (Richard and Gratton, 2001). The precipitation ensemble forecasts are also averaged to create a synthetic station; however, the average values are applied across the entire basin. No adjustment is made to account for changes in precipitation with elevation because the measurements from the SNOTEL sites in each of the study basins indicated a weak dependence of precipitation on elevation.

Snow-Covered Area. SCA data are obtained from the MODIS eight-day composite snow cover data product (MOD10A2). This data product was obtained from the National Snow and Ice Data Center. Eight-day composite data are used to maximize the amount of useable SCA images and to minimize the effect of cloud cover. Harshburger *et al.* (2010) gives a full description of how the eight-day composite SCA data are used to create snow depletion curves, and how the curves are then used to produce the streamflow forecasts.

Model Updating. The enhanced version of SRM is updated on a daily basis with observed temperature, precipitation, and streamflow values from the previous day. These observational data are provided by the NRCS and the U.S. Geological Survey (USGS) (http://water.usgs.gov/data). Model updating is completed to avoid the propagation of errors in the streamflow forecasts. Harshburger *et al.* (2010) contains a full description of the model-updating process, along with a description of how the data are used in the model.

APPLICATION OF THE SNOWMELT RUNOFF MODEL ENSEMBLE APPROACH TO THREE STUDY BASINS

Ensemble streamflow forecasts were generated for the Big Wood, South Fork of the Boise, and North Fork of the Clearwater Basins in the state of Idaho (Figure 2) for the 2003, 2004, 2005, and 2006 snowmelt seasons. The streamflow was modeled at a single stream gauge located in each of the three basins. These gauges were selected based upon their hydrological significance, data availability, and the fact that they are located upstream of all significant dams

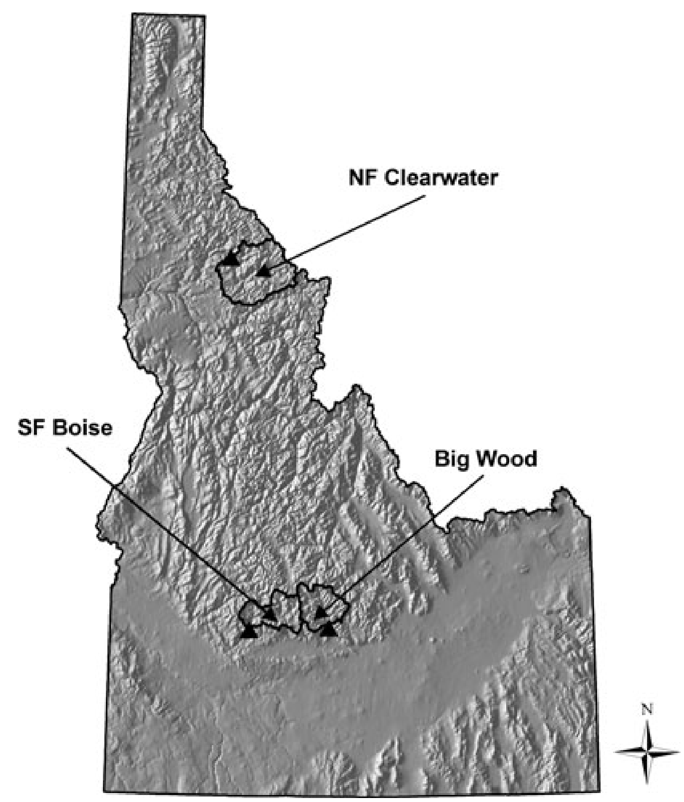

FIGURE 2. Map of Idaho Showing the Location of the Basins Modeled in This Study. The triangles represent the locations of the stream gauges.

and diversions and, therefore, represent natural flow (Harshburger *et al.*, 2010). Characteristics of each of the three test basins, along with the number of elevation zones that are used to model each basin, are found in Table 1.

RESULTS AND DISCUSSION

To assess the performance of the ensemble streamflow predictions relative to actual streamflow measurements, coefficients of determination (R^2), percentages of volume difference (Dv%), and skill score statistics were computed for all forecast cases. The results based on the mean ensemble, averaged over all four of the study years, are summarized in Table 2 and are discussed below. In addition, forecast hydrographs were generated to visually compare the actual (measured) stream discharge with the forecasted values at lead times of 1, 3, 7, and 10 days, to make qualitative observations related to the timing of peak flows and other forecast characteristics. As four years (2003-2006) of ensemble streamflow forecasts were generated for each of the three basins at every forecast lead time, numerous plots were generated; only two sets of forecast plots are shown here, one for an average flow year and the other for a low flow year (Figures 3 and 4). In addition, the streamflow forecast ensembles were converted to exceedance probabilities to limit the number of lines shown on the plots. The 10, 25, 50 (median), 75, and 90% exceedance probability lines are shown on the ensemble forecast plots. Forecast results are shown for lead times of 10 days or less because the usefulness of the forecasts decreases at lead times longer than that.

Figure 3 shows the forecast results for the Big Wood River Basin during the 2003 snowmelt season (April 1-July 31). This year was noteworthy because it was an average flow year (Harshburger *et al.*, 2010) and illustrates how the ensemble forecasts perform under normal conditions. As can be seen, the ensemble forecasts follow the measured flow fairly well, although there are small differences in the timing. The quality of the streamflow forecasts also degrades with increasing lead time. This can be attributed to uncertainties in the model estimates and the fact that the accuracy of the temperature and precipitation forecasts generally decrease as lead time increases. Even though there are small timing issues in the streamflow forecasts, the model correctly forecasted the timing of the peak flow seven days in advance. The streamflow forecast ensembles capture the measured streamflow very well; however, there are times when the measured streamflow lies above the 10% exceedance probability line and below the 90% exceedance probability line. In addition, the magnitude of the peak discharge is between the 10% and 25% exceedance probability lines for all of the

TABLE 1. Characteristics for the Three Test Basins Modeled in This Study.

Basin	Contributing Area (km^2)	Elevation Range (m)	No. Elevation Zones	Mean Average Annual Discharge (cm)
Big Wood	1,625	1,618-3,630	5	12.96
South Fork of the Boise	1,639	1,316-3,159	4	19.71
North Fork of the Clearwater	3,520	504-2,407	4	92.44

TABLE 2. The Average Coefficient of Determination (R^2), Percentage of Volume Difference (Dv%), and Skill Score Values Calculated for the Mean Ensemble at Various Forecast Lead Times (in days).

Lead Time (days)	R^2			Dv%			Skill Score		
	Bwood	SF Boise	NF Clwater	Bwood	SF Boise	NF Clwater	Bwood	SF Boise	NF Clwater
1	0.96	0.96	0.94	−0.7	0.5	3.1	0.82	0.83	0.80
2	0.93	0.93	0.91	0.5	2.7	6.3	0.80	0.80	0.76
3	0.91	0.91	0.89	0.1	3.1	8.2	0.78	0.79	0.73
4	0.89	0.88	0.87	0.0	3.6	10.2	0.77	0.77	0.69
5	0.86	0.86	0.83	0.1	4.2	11.7	0.74	0.75	0.65
7	0.79	0.80	0.78	1.2	6.4	14.6	0.69	0.70	0.57
10	0.75	0.74	0.72	3.6	10.6	19.2	0.63	0.62	0.45
15	0.59	0.58	0.63	8.7	15.0	24.2	0.50	0.47	0.28

Notes: The averaging is performed over the four study years for each basin. More detailed explanations of the three variables are described in the text. "Bwood" refers to the Big Wood Basin, "SF Boise" to the South Fork of the Boise River Basin, and "NF Clwater" to the North Fork of the Clearwater River Basin.

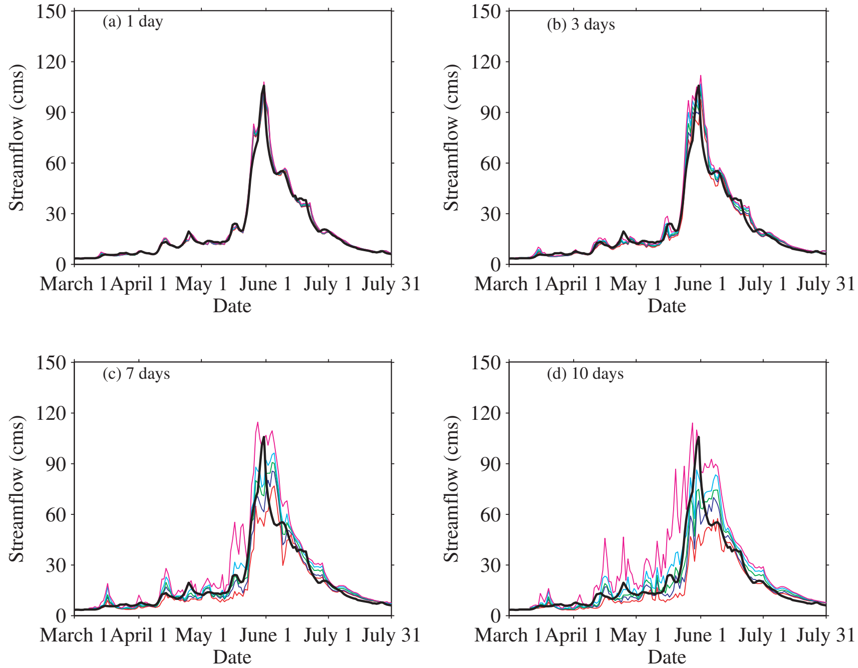

FIGURE 3. Hydrographs of the Daily Streamflow Amount Generated by the Ensemble Model for the Big Wood River Basin in 2003 at Lead Times of (a) 1 Day, (b) 3 Days, (c) 7 Days, and (d) 10 Days. The plots show the measured daily streamflow (black), the 10% exceedance probability (magenta), the 25% exceedance probability (cyan), the 50% exceedance probability (green), the 75% exceedance probability (blue), and the 90% exceedance probability (red). The instantaneous measured streamflow values were averaged to daily streamflow values to be compatible with the ensemble model simulations.

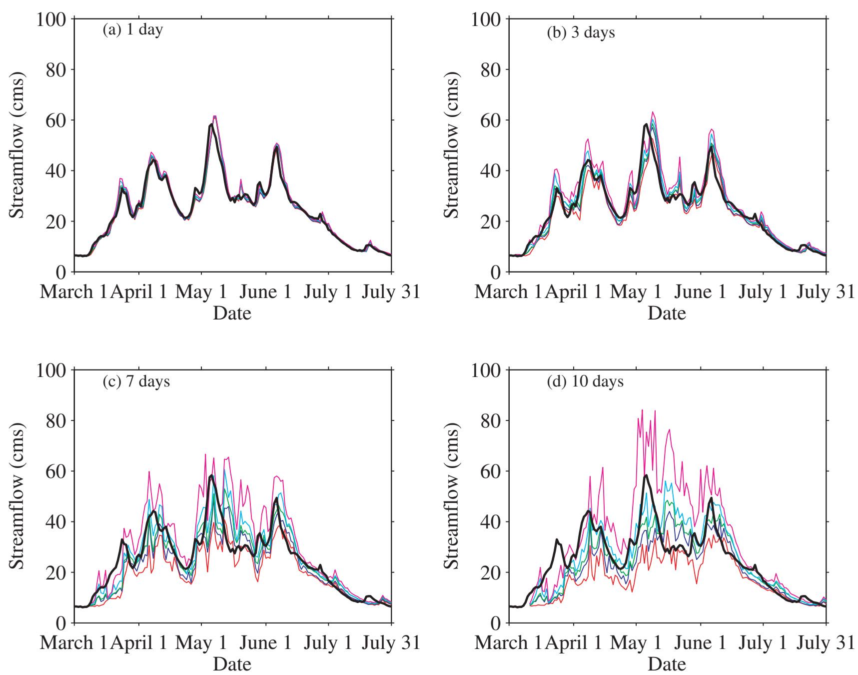

FIGURE 4. Hydrographs of the Daily Streamflow Amount Generated by the Ensemble Model for the South Fork of the Boise River Basin in 2004 at Lead Times of (a) 1 Day, (b) 3 Days, (c) 7 Days, and (d) 10 Days. The plots show the measured daily streamflow (black), the 10% exceedance probability (magenta), the 25% exceedance probability (cyan), the 50% exceedance probability (green), the 75% exceedance probability (blue), and the 90% exceedance probability (red). The instantaneous measured streamflow values were averaged to daily streamflow values to be compatible with the ensemble model simulations.

four forecast lead times illustrated here (Figure 3). This can be attributed to the fact that the peak flow typically occurs as a result of significant increases in temperature or large rainfall events, and the meteorological forecasts do not always capture these extreme events. The spread, or distance between the exceedance probability lines, is increasing with increasing lead time, which indicates more uncertainty in the forecasts at longer lead times.

Figure 4 shows the ensemble forecast results for the South Fork of the Boise River Basin during the 2004 snowmelt season, which was a low-flow year (Harshburger *et al.*, 2010). The ensemble forecast results are very similar to those presented in Figure 3 and degrade with increasing lead time.

In addition to the forecast hydrographs, the coefficient of determination (R^2), percentage of volume difference (Dv%), and skill score statistic are used to assess the accuracy of the ensemble streamflow forecasts. The R^2 value is used to evaluate the goodness of fit for the mean streamflow ensemble and is a direct measure of the proportion of the variance of the measured flow that is explained by the forecasted flow (Kite, 1975). The Dv% statistic is used to estimate the bias in the model simulations. Negative Dv% values indicate that the mean streamflow ensemble is overpredicting the observed flow, and positive values indicate that the ensemble forecast is underpredicting. Finally, a quantitative skill metric (skill score) is used to assess the performance of the ensemble forecasts relative to the climatology. Mean-squared error values are calculated for each of the 16 streamflow forecast ensemble members, relative to the observed streamflow, and the average of these squared errors over all ensemble members is calculated. This process is also completed using climatological (historical) streamflow values, in a similar way. The skill score statistic is calculated using the following equation:

$$\text{Skill score} = 1 - \frac{\sum_{i=1}^{N} \frac{(\text{forecast}_i - \text{observed}_i)^2}{N}}{\sum_{j=1}^{M} \frac{(\text{climatology}_j - \text{observed}_j)^2}{M}}, \quad (2)$$

where N is the number of forecast ensemble members and M is the number of climatological observations. A skill score value of 1.0 indicates a perfect forecast, a skill of 0.0 is equivalent to climatology, and a negative value means that the climatology outperforms the ensemble forecast. The skill score statistic rewards accuracy, but punishes forecast spread.

Figure 5 is used to illustrate how the R^2 values for the mean ensemble member vary with forecast lead time. Separate plots were generated for each of the three basins and each line represents a different forecast year. As can be seen, the R^2 values generally range from about 0.9 to 0.95 out 1 day to anywhere from 0.6 to 0.8 out 15 days. At lead times of seven days, the R^2 values are typically greater than about 0.70-0.75. There are, however, two exceptions. The R^2 values for the 2004 snowmelt season decrease at a greater rate with forecast lead time than those for the other forecast years in both the Big Wood and South Fork of the Boise River Basins (Figures 5a and 5b). This result is significant as both of these basins experienced abnormally low flows during the 2004 snowmelt season (Harshburger *et al.*, 2010). However, the same decrease is not seen in the R^2 values for the North Fork of the Clearwater Basin for the 2005 snowmelt season, which also had abnormally low flows.

In addition to the plots in Figure 5, the average R^2 values using the mean ensemble member for each of the three basins are listed in Table 2. Values are provided for forecast lead times 1-5, 7, 10, and 15 days. Average R^2 values for the four forecast years were used to limit the amount of data presented in the table. The average R^2 values are very comparable for all three stations and range from 0.96 at a lead time of 1 day to 0.58 at a lead time of 15 days. As seen in Figure 5, the R^2 values decrease with increasing lead time.

Figure 6 illustrates how the Dv% values for the mean ensemble member vary with forecast lead time. Separate plots were generated for each of the three basins and each line represents a different forecast year. The Dv% values for the mean ensemble member generally increase steadily as the forecast lead time increases. This indicates that the mean ensemble is underpredicting the actual flow more as the lead time increases. As the Dv% values at early forecast lead times (i.e., 1-day) are indicative of the level of error in discharge due to model errors, the steady increase with lead time is attributable to meteorological forecast errors, especially the inability of the mean ensemble member to forecast events far in advance that trigger large flow volumes, such as significant increases in temperature or large rainfall events. Predicted flows for the Big Wood and South Fork of the Boise River Basins during all years were nearly unbiased (within about ±10%) out to a lead time of seven days.

Table 2 also contains average Dv% values for each of the three basins. The results shown in the table indicate that the average Dv% values for the Big Wood and South Fork of the Boise River Basins are at or below 10% at a lead time of 10 days. However, the Dv% values are much higher for the North Fork of the Clearwater Basin. To assess the reasons for the inflated Dv% for the North Fork of the Clearwater, Dv% values were plotted against the number of rain-on-snow days (Figure 7). This was completed for each basin and year. Rain-on-snow was deemed to

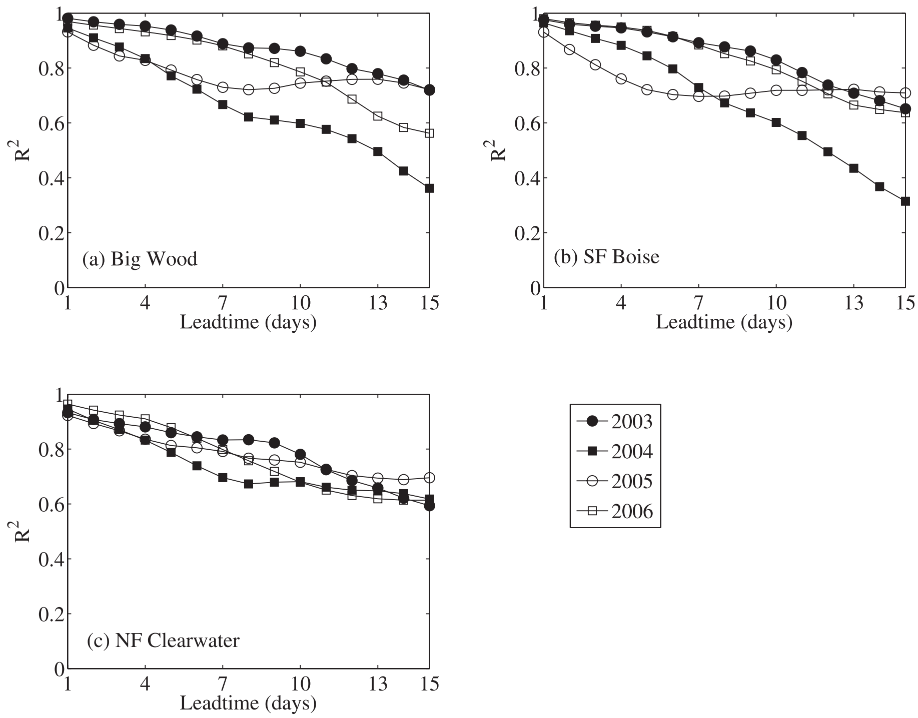

FIGURE 5. The Coefficient of Determination (R^2) for the Mean Ensemble as a Function of Forecast Lead Time for Years 2003, 2004, 2005, and 2006 for the (a) Big Wood, (b) South Fork of the Boise, and (c) North Fork of the Clearwater Basins.

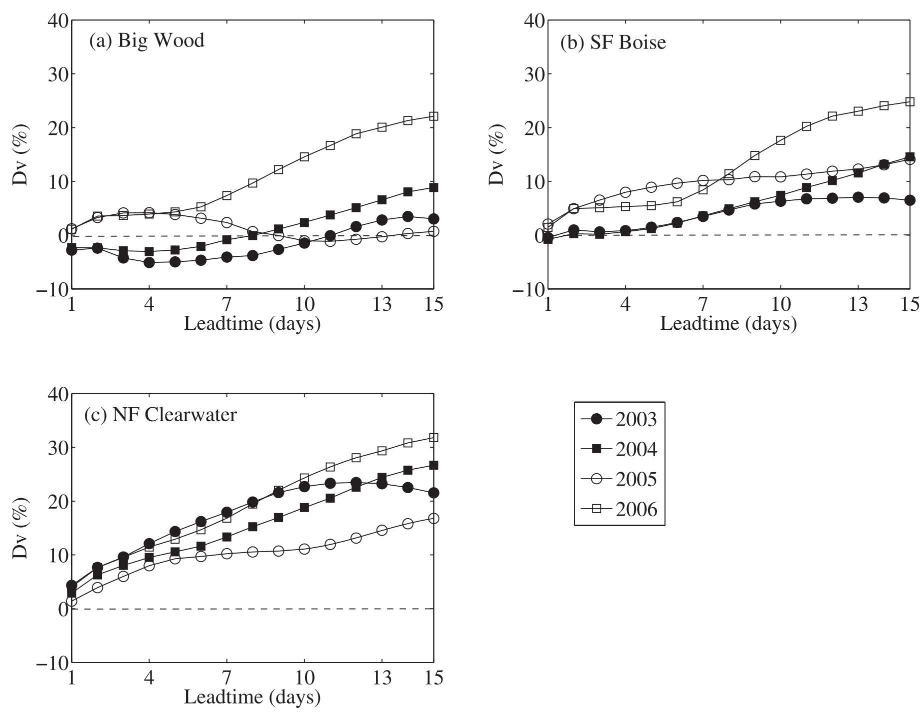

FIGURE 6. The Percentage of Volume Difference (Dv%) for the Mean Ensemble as a Function of Forecast Lead Time for Years 2003, 2004, 2005, and 2006 for the (a) Big Wood, (b) South Fork of the Boise, and (c) North Fork of the Clearwater Basins. Negative values indicate that the mean streamflow ensemble is overpredicting the observed flow, whereas positive values indicate that the ensemble forecast is underpredicting the observed flow.

have occurred if over half of the basin was experiencing rain falling on the snowpack (determined from SCA data) and if the rainfall amount exceeded 0.25 cm during a 24-h period. The results shown in Figure 7 indicate that the North Fork of the Clearwater receives rain-on-snow more regularly than the other two basins, which might explain why it generally has higher Dv% values.

Figure 8 is used to illustrate how the skill score statistics for the ensemble streamflow forecasts vary with forecast lead time. Separate plots were generated for each of the three basins and each line represents a different forecast year. The skill score values generally range from about 0.8 at a lead time of 1 day to anywhere from 0.8 to 0.2 at 15 days, although one case (North Fork of the Clearwater in 2006) is negative. This indicates that the ensemble forecasts generally outperform the climatology even at forecast lead times of up to 15 days. Skill score values generally degrade with increasing lead time, which can be explained by the fact that the forecasts generally degrade at these longer lead times (Figures 3 and 4). The skill score values for the North Fork of the Clearwater Basin for the 2006 snowmelt season go below zero at a lead time of 13 days (Figure 8c), which illustrates that the climatology is outperforming the ensembles at these lead times for this particular case. This is perplexing, however, because the skill score values were much higher for the other three years, and the North Fork of the Clearwater did not experience an abnormally large amount of rain-on-snow events during the 2006 snowmelt season (Figure 7). The skill score values are also generally higher for the low-flow years (2004 for the Big Wood and South Fork of the Boise and 2005 for the North Fork of the Clearwater) than for the high-flow years.

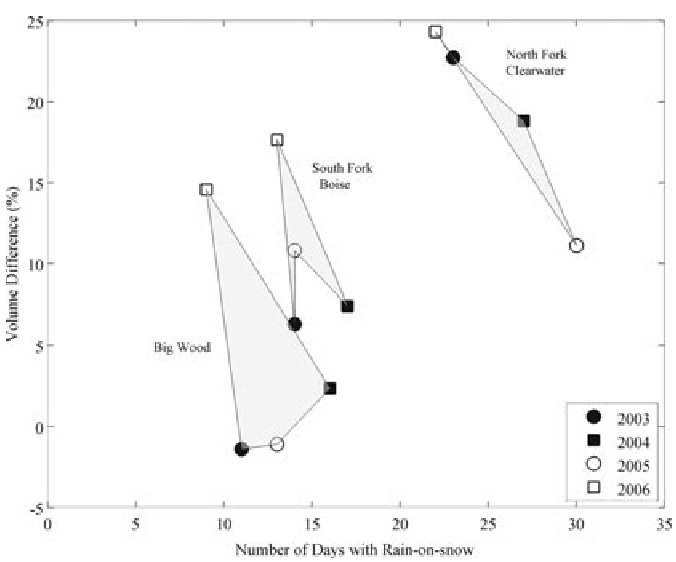

FIGURE 7. The Percentage of Volume Difference (Dv%) Plotted as a Function of the Number of Days with Rain-on-Snow for the 2003, 2004, 2005, and 2006 Snowmelt Seasons. The data for the Big Wood Basin, the South Fork of the Boise, and the North Fork of the Clearwater are shown as polygons to more easily visualize the parameter space of each basin.

Table 2 contains average skill score values for each of the three basins. The average skill scores are similar for each of the three basins; however, they are slightly lower for the North Fork of the Clearwater. The average values for the basins range from 0.83 to 0.57 for lead times of up to 7 days, but still have positive values of between 0.63 and 0.28 for lead times between 10 and 15 days. This indicates that, on average, the ensemble prediction system outperforms the assumption of climatology for all lead times.

Figures 9 and 10 are used to compare average R^2 and Dv% values for the mean ensemble member with values obtained for single deterministic forecasts for the same basins and years (Harshburger *et al.*, 2010). The R^2 results indicate that there is significantly less bias in the ensemble predictions. The Dv% values for the mean ensemble are typically slightly less than those for the deterministic forecasts, but are comparable for both approaches. However, in addition to improved performance, the ensemble forecasts provide the user with information regarding how likely a forecast is to be correct.

CONCLUSIONS

A new methodology was developed and tested for the generation of short-to-medium range ensemble streamflow forecasts (1 to 15 days) using an enhanced version of the SRM. Although SRM has been tested in numerous basins around the world, it has not been widely used to generate ensemble streamflow forecasts. Enhancements were made to the model to optimize model efficiency and aid in its operational implementation. The ensemble approach described here generates 15 different realizations of the potential streamflow in three basins in Idaho for the years 2003, 2004, 2005, and 2006. The results from the ensemble streamflow forecasts indicate that the model performed very well at lead times up to 7 days; however, there was still some predictability at the longer lead times up to 15 days. The model performed well in years with average to above-average flows, as well as low-flow years. Hydrographs of the forecasted streamflow show that the timing of peak flow and the overall measured streamflow are captured well by the ensemble modeling approach. However, the value of the predicted peak flow is typically low relative to the

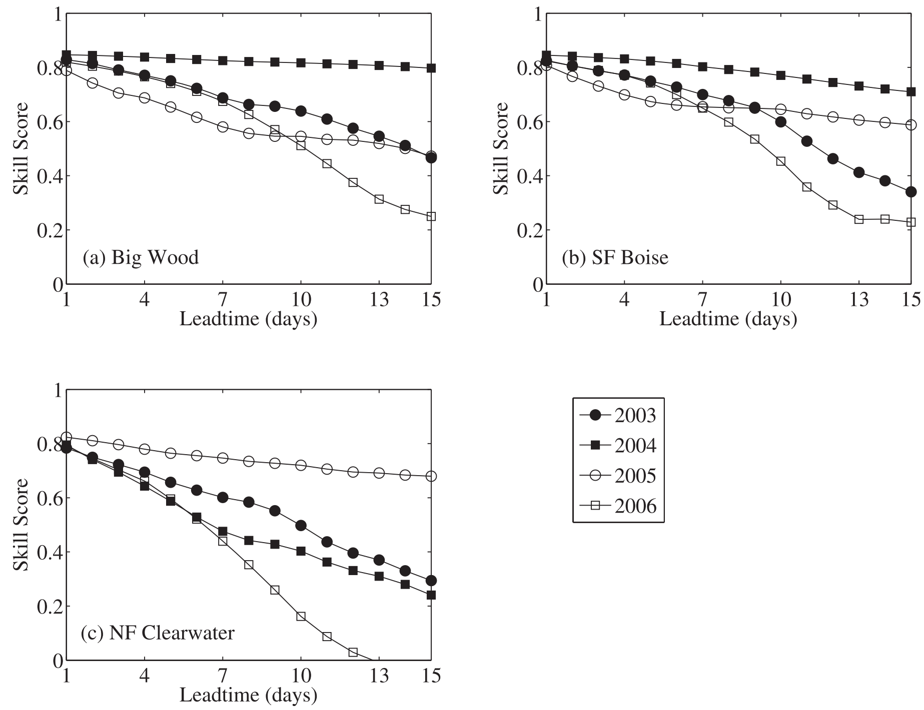

FIGURE 8. The Skill Score Values for the Mean Ensemble as a Function of Forecast Lead Time for Years 2003, 2004, 2005, and 2006 for the (a) Big Wood, (b) South Fork of the Boise, and (c) North Fork of the Clearwater Basins. A positive skill score indicates that the ensemble forecasts outperform the assumption of climatology.

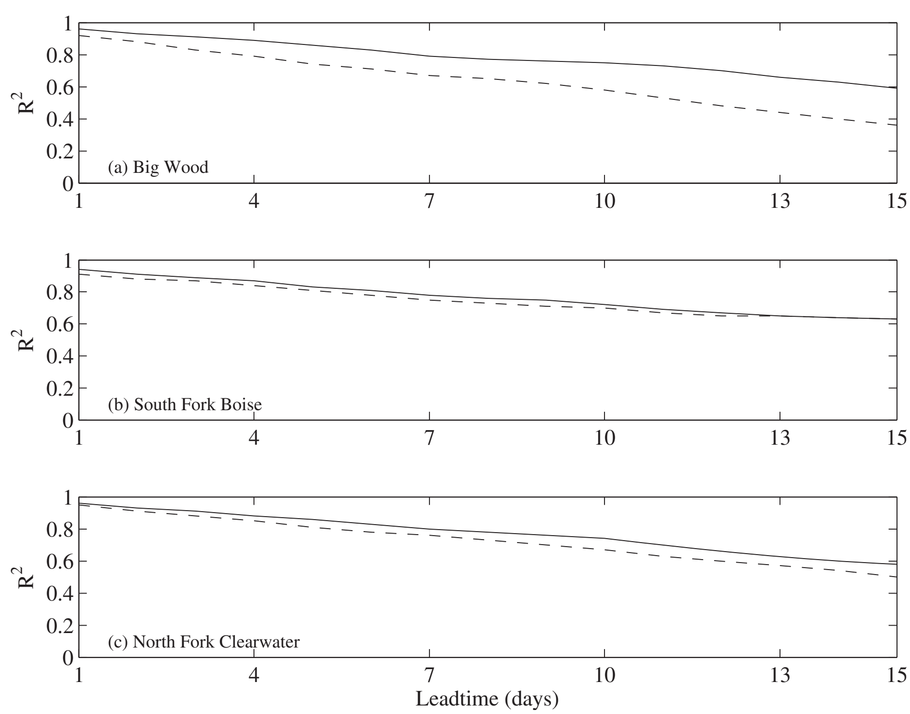

FIGURE 9. Comparison of the Average Coefficient of Determination (R^2) Values for the Mean Ensemble Member (solid lines) and Results for Single, Deterministic Forecasts (dashed lines) from Harshburger *et al.* (2010) for Various Forecast Lead Times. Forecast results are shown for the (a) Big Wood, (b) the South Fork of the Boise, and (c) the North Fork of the Clearwater Basins.

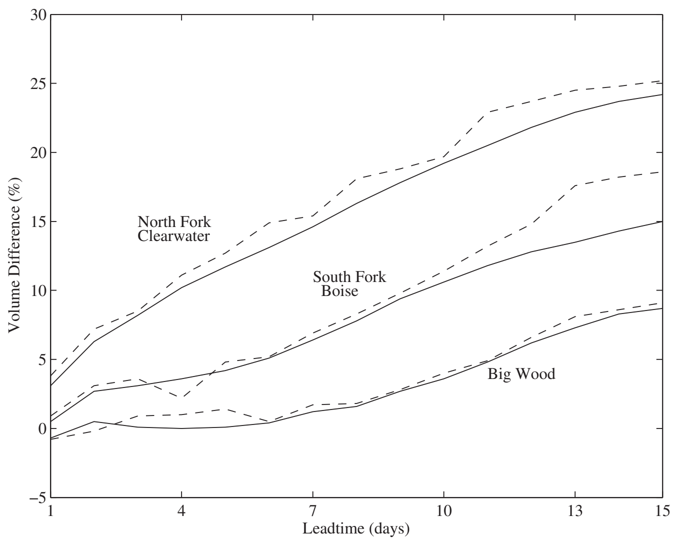

FIGURE 10. Comparison of the Average Percentage of Volume Difference (Dv%) for the Mean Ensemble Member (solid lines) and Results for Single, Deterministic Forecasts (dashed lines) Obtained from Harshburger *et al.* (2010) as a Function of Forecast Lead Time. The results are shown for the Big Wood, the South Fork of the Boise, and the North Fork of the Clearwater Basins.

measured peak flow. The performance of the ensemble approach was assessed by computing the coefficient of determination (R^2), the percentage of volume difference (Dv%), and a skill score that quantifies the usefulness of the forecasts relative to climatology. The average R^2 value for the mean ensemble is greater than about 0.8 for all three basins for lead times up to seven days. These results are better than those found by Harshburger *et al.* (2010) for single, deterministic forecasts (average R^2 value of 0.73 at a lead time of seven days). Two of the basins (Big Wood and South Fork of the Boise) are fairly unbiased (within ±10%) in terms of the streamflow volume at lead times of up to seven days. The model underpredicts the streamflow volume for the North Fork of the Clearwater, which may be due to the large number of rain-on-snow events that occur in that basin. The average skill scores for the basins are greater than about 0.6 for lead times up to seven days, which indicates that the ensemble forecasting system significantly outperforms the assumption of climatology. As the ensemble streamflow forecasting system was successfully tested in three basins for four separate years, there is potential for these methods to be applied in other snowmelt-dominated basins. In addition, the simplicity of the model should make it relatively easy for water managers and other operational forecasters (i.e., federal agencies) to generate accurate ensemble streamflow forecasts in a timely manner.

ACKNOWLEDGMENTS

The authors thank the Natural Resources Conservation Service, United States Geological Survey, and the National Snow and Ice Data Center for providing the data required to complete this study. This research was supported by the Pacific Northwest Regional Collaboratory as part of Raytheon Corporation's Synergy project, funded by NASA through NAS5-03098, Task No. 110. This grant also funded the graduate research for Brian Harshburger (Ph.D.), Brandon Moore (M.S.), and Troy Blandford (M.S.).

LITERATURE CITED

Blandford, T.R., K.S. Humes, B.J. Harshburger, B.C. Moore, V.P. Walden, and H. Ye, 2008. Seasonal and Synoptic Variations in Near-Surface Air Temperature Lapse Rates in a Mountainous Basin. Journal of Applied Meteorology and Climatology 47:249-261.

Buizza, R., A. Hollingsworth, F. Lalaurette, and A. Ghelli, 1999. Probabilistic Predictions of Precipitation Using ECMWF Ensemble Prediction System. Weather and Forecasting 14:168-189.

Clark, M.P., S. Gangopadhyay, L. Hay, B. Rajagopalan, and R. Wilby, 2004. The Schaake Shuffle: A Method for Reconstructing Space-Time Variability in Forecasted Precipitation and Temperature Fields. Journal of Hydrometeorology 5:243-562.

Clark, M.P. and L.E. Hay, 2004. Use of Medium-Range Numerical Weather Prediction Model Output to Produce Forecasts of Streamflow. Journal of Hydrometeorology 5:15-32.

Day, G.N., 1985. Extended Streamflow Forecasting Using NWSRFS. Journal of Water Resources Planning and Management 111:157-170.

Dey, B., V.K. Sharma, and A. Rango, 1989. A Test of Snowmelt-Runoff Model for a Major River Basin in Western Himalayas. Nordic Hydrology 20:167-178.

Gomez-Landesa, E. and A. Rango, 2002. Operational Snowmelt Runoff Forecasting in the Spanish Pyrenees Using the Snowmelt Runoff Model. Hydrological Processes 16:1583-1591.

Hamill, T., J.S. Whitaker, and X. Wei, 2004. Ensemble Reforecasting: Improving Medium-Range Forecast Skill Using Retrospective Forecasts. Monthly Weather Review 132:1434-1447.

Harshburger, B.J., V.P. Walden, K.S. Humes, B.C. Moore, T.R. Blandford, and A. Rango, 2010. Evaluation of Short-to-Medium Range Streamflow Forecasts Obtained Using an Enhanced Version of SRM. Journal of the American Water Resources Association 46:603-617.

Hong, M.A. and C. Guodong, 2003. A Test of Snowmelt Runoff Model (SRM) for the Gongnaisi River Basin in the Western Tianshan Mountains, China. Chinese Science Bulletin 48:2253-2259.

Kite, G.W., 1975: Performance of Two Deterministic Hydrologic Models. *In*: Proceedings of the Symposium on Applications of Mathematical Models in Hydrology and Water Resource Systems, Bratislava. IAHS Publ. No. 115, pp. 136-142.

Martinec, J., 1975. Snowmelt Runoff Model for River Flow Forecasts. Nordic Hydrology 6:145-154.

Martinec, J., A. Rango, and R. Roberts, 1994. The Snowmelt Runoff Model (SRM) User's Manual. Geographica Bernensia. Department of Geography, University of Berne, Berne, Switzerland.

Mitchell, K.M. and D.R. DeWalle, 1998. Application of the Snowmelt Runoff Model Using Multiple-Parameter Landscape Zones on the Towanda Creek Basin, Pennsylvania. Journal of the American Water Resources Association 34:335-346.

Moore, B.C., 2005. Statistical Downscaling of Medium-Range Weather Forecasts Within Snowmelt Dominated Basins in the

Intermountain West. M.S. Thesis, University of Idaho, Moscow, Idaho, 70 pp.
Mullen, S.L. and R. Buizza, 2001. Quantitative Precipitation Forecasts Over the United States by the ECMWF Ensemble Prediction System. Monthly Weather Review 129:638-663.
Nagler, T., S. Quegan, and H. Rott, 2000. Real Time Snowmelt Runoff Forecasting Using ERS SAR PRI Data. Proceedings of ERS-ENVISAT Symposium, Goteborg, Sweden.
Nagler, T., H. Rott, P. Malcher, and F. Müller, 2008. Assimilation of Meteorological and Remote Sensing Data for Snowmelt Runoff Forecasting. Remote Sensing of Environment 112:1408-1420.
Rango, A. and V.V. Katwijk, 1990. Development and Testing of a Snowmelt-Runoff Forecasting Technique. Water Resources Bulletin 26:135-144.
Rango, A. and J. Martinec, 1994. Model Accuracy in Snowmelt-Runoff Forecasts Extending From 1 to 20 Days. Water Resources Bulletin 30:463-470.
Richard, C. and D.J. Gratton, 2001. The Importance of the Air Temperature Variable for the Snowmelt Runoff Modeling Using the SRM. Hydrological Processes 15:3357-3370.
Shafer, B.A., E.B. Jones, and D.M. Frick, 1982: Snowmelt Runoff Simulation Using the Martinec-Rango Model on the South Fork Rio Grande and Conejos River in Colorado. AgRISTARS Report CP-G1-04072, Goddard Space Flight Center, Greenbelt, Maryland, 88 pp.
Singh, P. and V.P. Singh, 2001. Snow and Glacier Hydrology. Kluwer Academic Publishers, The Netherlands, 742 pp.
Westrick, K.J., P. Storck, and C.F. Mass, 2002. Description and Evaluation of a Hydrometeorological Forecast System for Mountainous Watersheds. Weather and Forecasting 17:250-262.
Wigmosta, M.S., L.W. Vail, and D.P. Lettenmaier, 1994. A Distributed Hydrology-Vegetation Model for Complex Terrain. Water Resources Research 30:1665-1679.