Incorporating Medium-Range Numerical Weather Model Output into the Ensemble Streamflow Prediction System of the National Weather Service

KEVIN WERNER AND DAVID BRANDON

Colorado Basin River Forecast Center, Salt Lake City, Utah

MARTYN CLARK

Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado

SUBHRENDU GANGOPADHYAY

Cooperative Institute for Research in Environmental Sciences, and Department of Civil, Environmental and Architectural Engineering, University of Colorado, Boulder, Colorado

(Manuscript received 26 March 2004, in final form 5 August 2004)

ABSTRACT

This study introduces medium-range meteorological ensemble inputs of temperature and precipitation into the Ensemble Streamflow Prediction component of the National Weather Service River Forecast System (NWSRFS). The Climate Diagnostics Center (CDC) produced a reforecast archive of model forecast runs from a dynamically frozen version of the Medium-Range Forecast (MRF) model. This archive was used to derive statistical relationships between MRF variables and historical basin-average precipitation and temperatures. The latter are used to feed the Ensemble Streamflow Prediction (ESP) component of the NWSRFS. Two sets of ESP reforecasts were produced: A control run based on historically observed temperature and precipitation and an experimental run based on MRF-derived temperature and precipitation. This study found the MRF reforecasts to be generally superior to the control reforecasts, although there were situations when the downscaled MRF output actually degraded the forecast. Forecast improvements were most pronounced during the rising limb of the hydrograph—at this time accurate temperature forecasts improve predictions of the rate of snowmelt. Further improvements in streamflow forecasts at short forecast lead times may be possible by incorporating output from high-resolution regional atmospheric models into the NWSRFS.

1. Introduction and background

Ensemble-based river forecasts have been made by the Colorado Basin River Forecast Center (CBRFC) for many years with a component of the National Weather Service River Forecast System (NWSRFS) known as Ensemble Streamflow Prediction (ESP). The current operational practice at the CBRFC is to use "deterministic" meteorological forecasts of temperature for 1–10 days and precipitation for 1–3 days following the initialization time of the forecast period. After the meteorological forecast period, ensembles of historical temperature and precipitation data sequences are added to the end of the deterministic meteorological forecasts by blending to form ensemble model inputs extending out to several months. While this approach provides probabilistic forecasts on seasonal time scales, it does not provide probabilistic forecast information for days 1–10. Many applications (e.g., reservoir operations) can benefit from probabilistic forecast information on the shorter time scales as well. This study demonstrates the use of medium-term (lead times from 1 to 14 days) probabilistic forecasts of temperature and precipitation in the NWSRFS. Previous studies (e.g., Wood et al. 2002) have focused on improving streamflow forecasts using the ESP approach, but these studies are restricted to seasonal time scales.

This study incorporates medium-range (1–14 day) forecasted atmospheric model input into an ensemble hydrologic model. In doing so, the hydrologic model translates both the forecast and its associated uncertainty into a hydrologic forecast. A description of the methodology including the study area, the atmospheric model, the downscaling technique, and the river forecast model are presented. Results are presented for a headwater basin and for a larger main stem basin.

Corresponding author address: Kevin Werner, NWS/CBRFC, 2242 W. North Temple, Salt Lake City, UT 84116.
E-mail: Kevin.Werner@noaa.gov

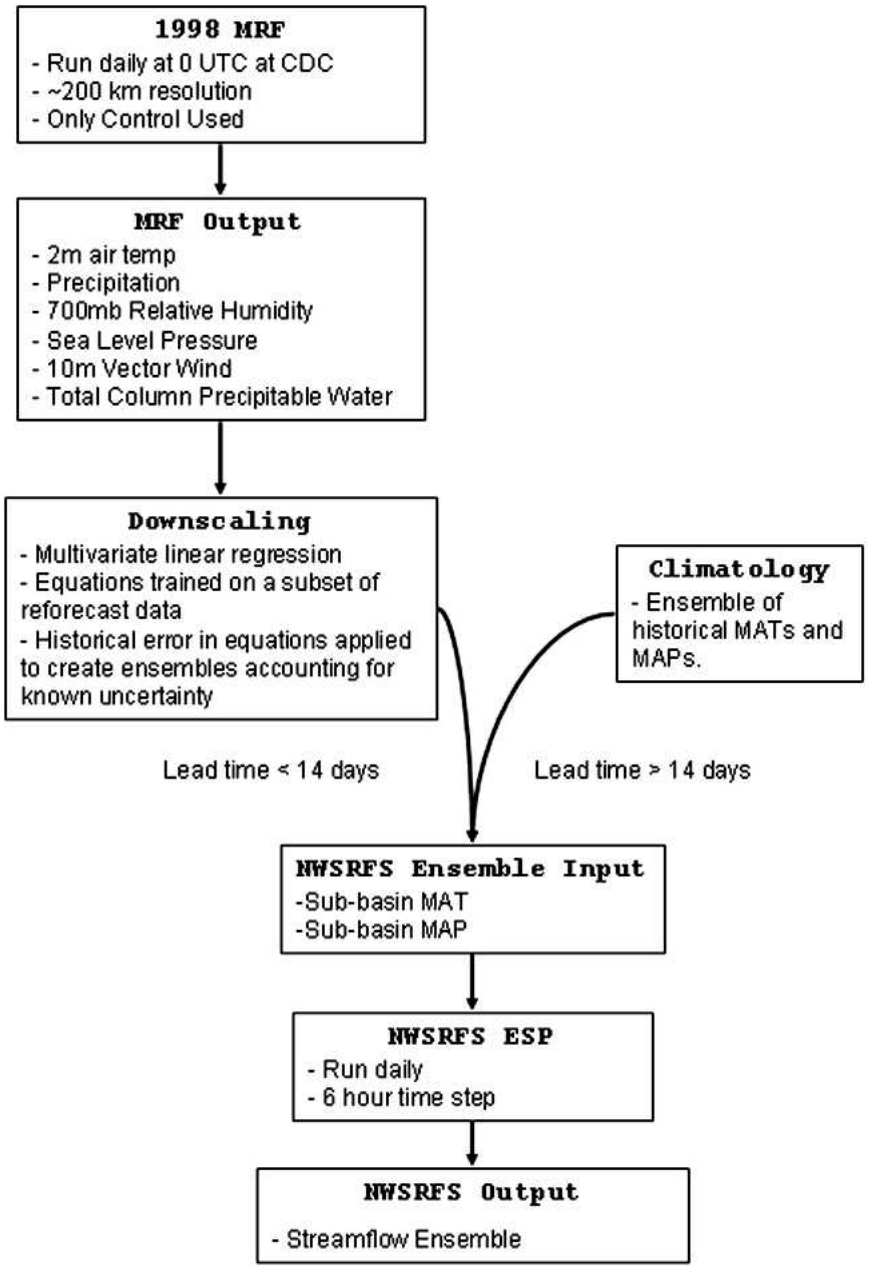

FIG. 1. Schematic of the daily process conducted at CBRFC to integrate meteorological information from the MRF model into river forecasts.

2. Method

Figure 1 shows the method by which meteorological forecasts are integrated into ESP river forecasts on an operational basis. Meteorological forecast information is derived from a Medium-Range Forecast (MRF) model. This information is downscaled to basin-scale temperature and precipitation in a manner that accounts for the forecast uncertainty. Finally, this information is fed into ESP to produce probabilistic river forecasts. This methodology is described in this section. In addition, the study area and the ranked probability skill score (RPSS) are described.

a. Study area

The upper Colorado River basin was chosen as the study area (Fig. 2). Average annual precipitation varies from about 6 in. in the lower valleys to 50 in. in the mountainous terrain on the Continental Divide (Daly et al. 2002). Snow water equivalent ranges from near zero in the valley bottoms to 25 in. in the mountains.

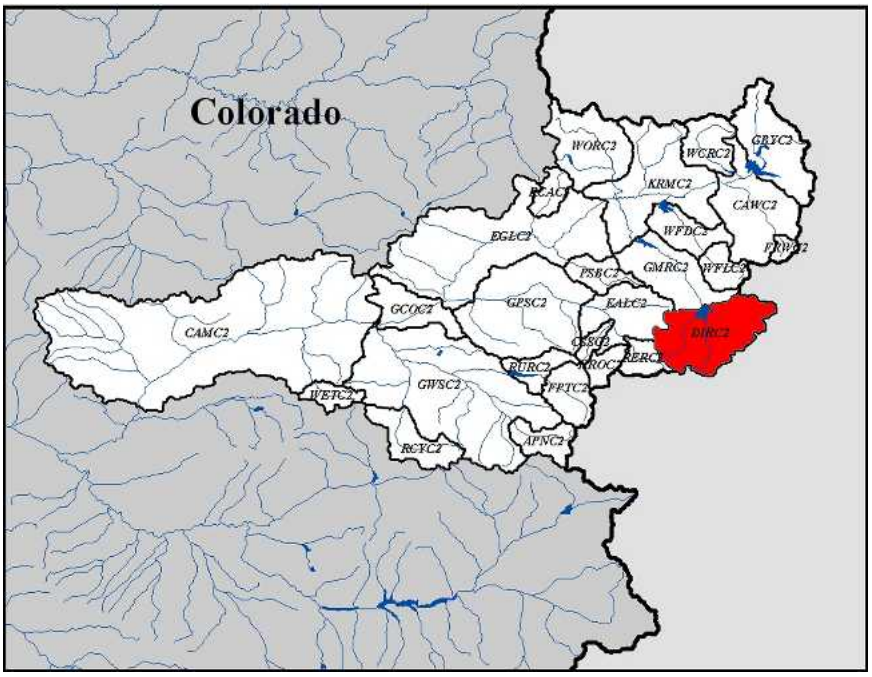

FIG. 2. CBRFC basins included in the study area. CAMC2 includes all white and red basins. DIRC2 is the red basin.

The hydrology of the upper Colorado River basin is dominated by snowmelt, for which temperature forecasts are important to determine the magnitude and timing of the melt. As with most meteorological forecasts, the MRF model forecasts of temperature are much better than precipitation, particularly at a lead time of more than 2 or 3 days (Clark and Hay 2004), so we expect larger improvements here than in rainfall-dominated river basins.

As modeled in the NWSRFS, the CBRFC has divided the upper Colorado into 27 smaller basins shown in Fig. 2. ESP is run separately over each of the smaller basins; downstream basins include water routed from the basin(s) above. Each of the small basins contains two or three subbasins divided by elevation. This allows for input of precipitation and temperature features that may be small in spatial extent or sensitive to elevation. ESP requires mean areal temperature (MAT) and mean areal precipitation (MAP) forecasted time series for each subbasin.

b. Ranked probability skill score

The RPSS is used to evaluate the probabilistic forecasts (Epstein 1969; Murphy 1969, 1971; Hersbach 2000). The continuous version of the ranked probability score (RPS) upon which the RPSS is based is given by

$$\mathrm{RPS} = \int_{-\infty}^{\infty} [P(x) - P_o(x)]^2 \, dx, \qquad (1)$$

where $P(x)$ is the forecasted exceedence probability of variable x (e.g., daily streamflow), and $P_o(x)$ is the observed exceedence probability of x. For a single value such as the observed volume on a particular day, the observed probability will be either zero or unity. The RPS is distance sensitive in that it increasingly penalizes forecasts that contain forecasted probability farther away from the observed quantity. The RPSS is based on the RPS and is given by

$$\mathrm{RPSS} = 1 - \frac{\mathrm{RPS_f}}{\mathrm{RPS_{ref}}}, \qquad (2)$$

where RPS_f and $\mathrm{RPS}_{\mathrm{ref}}$ are the forecast being evaluated and a reference forecast. The reference forecast is often taken to be climatology. RPSS values are less than or equal to unity. Positive RPSS values indicate percent improvement in forecast skill, while negative values indicate that the reference forecast is superior to the forecast being tested.

c. CDC reforecasting experiment

The atmospheric forecast model used in this study is the 1998 version of the National Centers for Environmental Prediction (NCEP) MRF model [this model was run by the National Oceanic and Atmospheric Administration (NOAA) Climate Diagnostics Center (CDC) as part of their reforecasting experiment; described here and in Hamill et al. (2004)]. For the CDC experiments the MRF is run with a grid spacing of about 200 km, whereas the CBRFC subbasin areas are irregularly

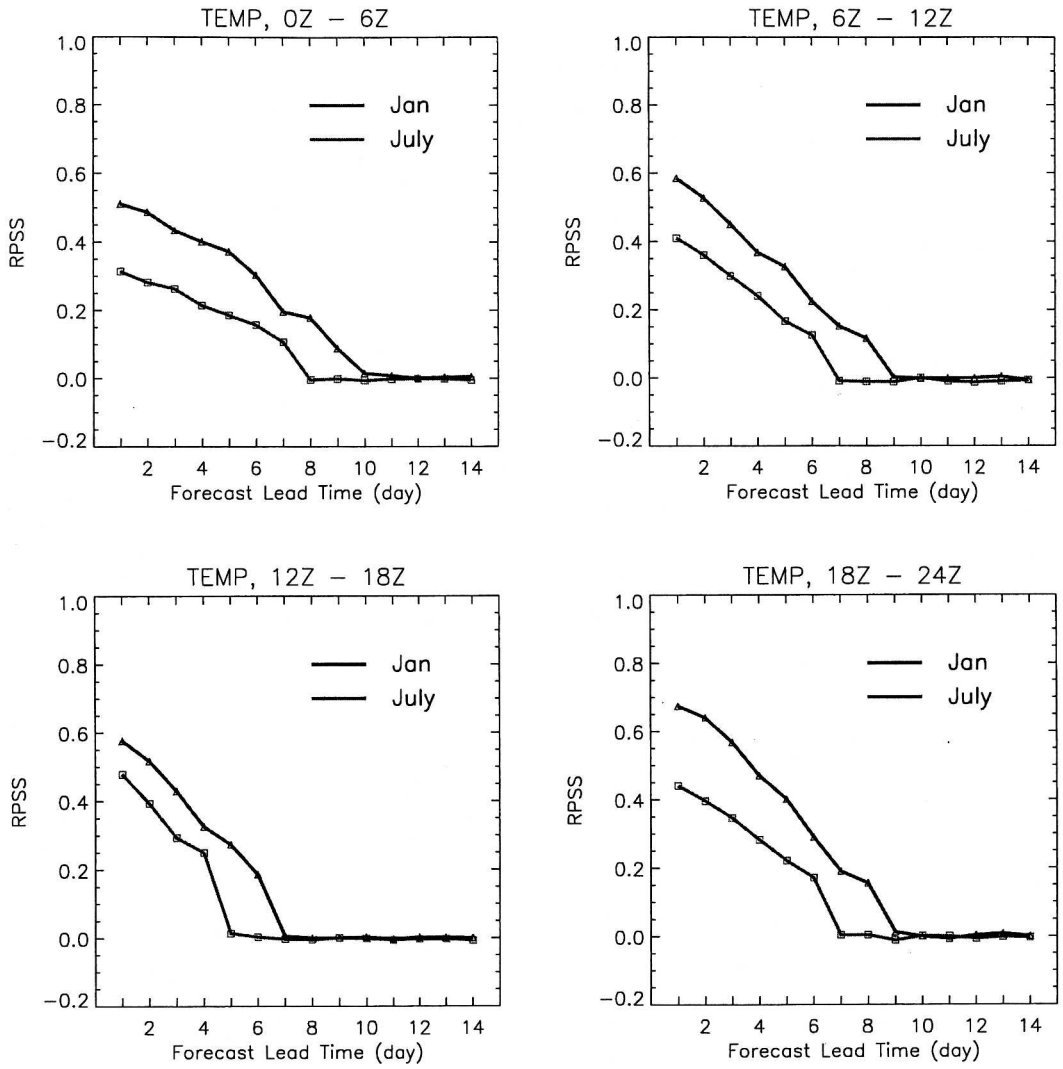

FIG. 3. RPSS values for MRF-derived mean areal temperatures for Jan and Jul as a function of lead time and forecast period.

shaped with highly variable spacing about of 20 km. Therefore a downscaling method [described later and in Clark and Hay (2004)] is used to relate the MRF output to the smaller basin scale. Although the atmospheric models currently run at NCEP for weather prediction have much finer resolution, this study takes advantage of the CDC reforecasts to integrate the observed model error into ESP.

A frozen version of the model is desirable rather than the most up-to-date version because this allows the estimation (and correction) of model biases and model errors. It also provides a suitably long archive that can be used to develop statistical relations between the (coarse resolution) MRF output and precipitation and temperature at local scales. Hamill et al. (2004) demonstrate that such postprocessing of model output can result in substantial improvements over operational forecasts.

Each MRF forecast initialization uses 15 different initial conditions from which meteorological forecast ensembles are derived. Only the control, or the best-guess initial condition, is used here because it was all that was available when this study began. However, use of the ensemble mean may result in future improvements. The CDC reforecast dates from 1979 to the present. Only the portion coinciding with the CBRFC MAT and MAP calibrations, 1979–98, was used in the study.

d. Downscaling

A statistical downscaling method is used to translate the coarse-resolution MRF output to the local scales important for hydrologic modeling (Clark and Hay 2004). Seven variables from the MRF model are interpolated to the center of each subbasin. These variables

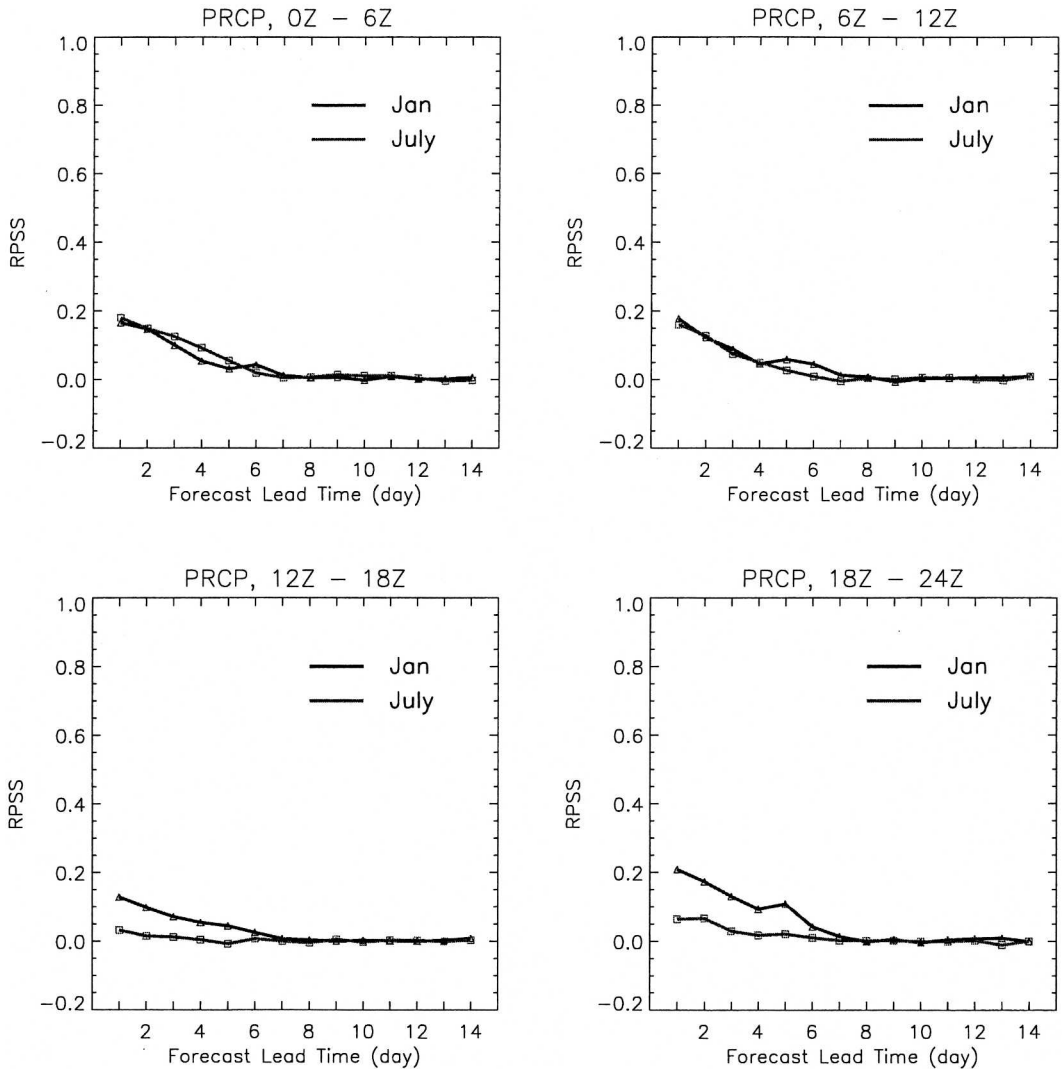

FIG. 4. RPSS values for MRF-derived mean areal precipitation for Jan and Jul as a function of lead time and forecast period.

have previously been found to be important for downscaling precipitation and temperature in the contiguous United States (Clark and Hay 2004). These variables are 2-m air temperature, precipitation, 700-mb RH, sea level pressure, 10-m u and v wind components, and total column precipitable water. The interpolation includes all MRF grid points within 500 km of the basin, but grid points are weighted by the inverse of their distance to the basin so that grid points nearer to a subbasin will be given more weight.

Unique multivariate linear regression equations were established for each subbasin, variable (i.e., MAP or MAT), and 6-h time step [see Clark and Hay (2004) for more details]. These equations are based on the seven MRF variables from the three nearest consecutive 12-h MRF time steps. For example, the MRF data valid on 1200 UTC 24 February, 0000 UTC 25 February, and 1200 UTC 25 February were used to predict the 0000–0600 UTC MAT on 25 February. This gives 21 predictors (seven variables at three time steps) for each MAT or MAP. Regression equations are developed using a forward-screening approach (Clark and Hay 2004). The coefficients are determined by training the regression equations on a subset of the data (i.e., 1980–85) and validated on the remainder of the data. During the validation process a residual term (i.e., the difference between observed and predicted) is computed. The standard deviation of the residuals is computed for each 6-h period, variable (i.e., MAT or MAP), and basin and saved for subsequent ensemble creation.

Ensemble traces are generated by stochastically modeling the error in the regression models:

$$y_{\text{iens}} = \hat{y} + \varepsilon, \tag{3}$$

where $\hat{y}$ is the value predicted from the multivariate

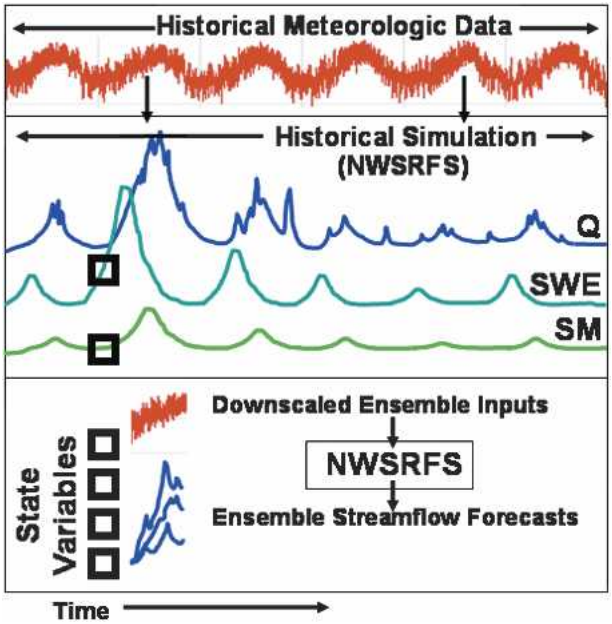

FIG. 5. Schematic representation ESP reforecast construction from hypothetical MATs and MAPs. Reforecasts here are made for 1 May between 1980 and 1984.

TABLE 1. NWSRFS hydrologic model states saved to initialize ESP reforecasts.

UZTWC	Upper-zone tension water
UZFWC	Upper-zone free water
LZTWC	Lower-zone tension water
LZFPC	Lower-zone primary free water
LZFSC	Lower-zone supplemental free
ADIMC	Additional impervious contents
SNOW	Snow model states

linear regression, and ε is a random number extracted from a normal distribution with mean of zero and standard deviation equal to the standard deviation of the regression residuals [$N(0, \sigma_e)$, computed above]. To create ensemble forecasts of MAT and MAP values for input into ESP, this equation is applied as many times as the number of ensemble traces are desired. For example, if 26 ensemble traces are desired, 26 random numbers are chosen yielding 26 predicted MATs (or MAPs). Note these predictions have no coherence between time steps, variables, or basins.

For MAPs a logical regression is done prior to determining the amount of precipitation to determine the occurrence of precipitation. If precipitation is chosen to occur, the amount is computed as the MAT value would be computed. If not, the amount is zero and the predicted MAP value is ignored.

Coherence is added through the "Schaake Shuffle" as described in Clark et al. (2004). The shuffle essentially integrates the coherence in the historical record into the forecast ensemble. The historical precipitation and temperature is sorted from lowest to highest. The precipitation and temperature forecast ensemble members are also sorted from lowest to highest. The precipitation and temperature series are sorted separately. 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 fifth highest precipitation and the 20th highest temperature, then the first ensemble member would be the ensemble with the fifth highest precipitation and the 20th highest temperature. This preserves the observed correlation between precipitation and temperature for the ensemble members (Clark et al. 2004). The downscaling process results in ensembles of MAT and MAP values for each subbasin,

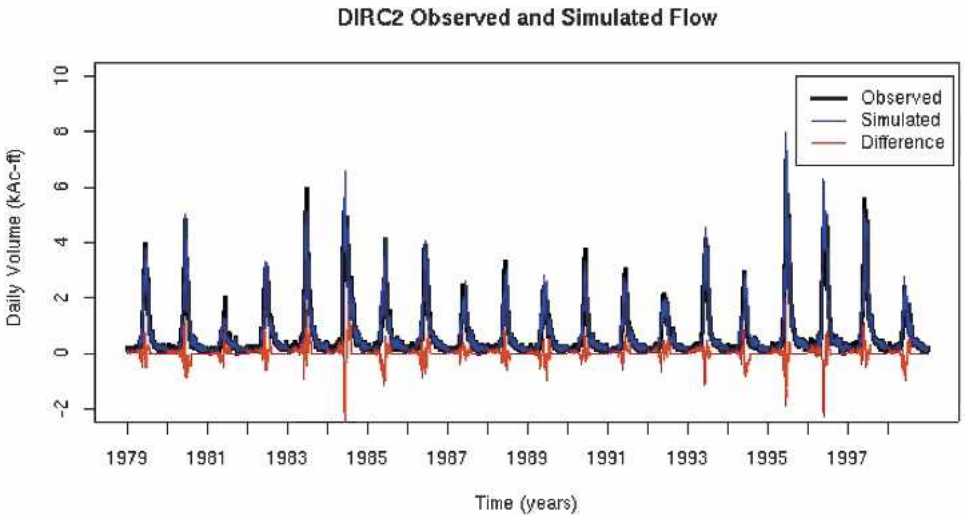

FIG. 6. Observed flows, simulated flows, and their difference for DIRC2.

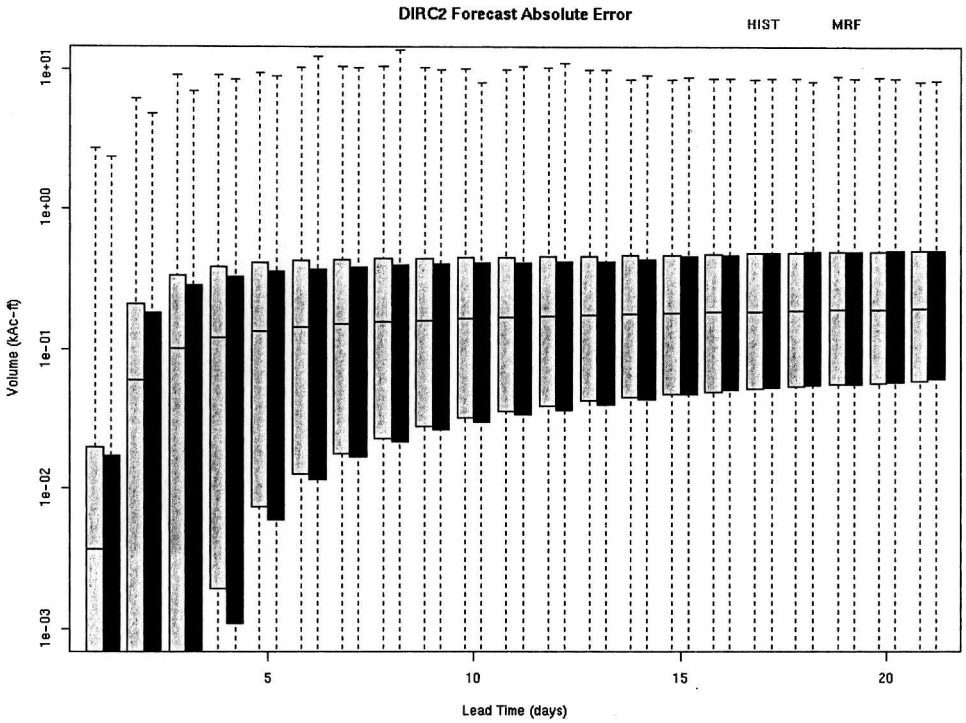

FIG. 7. Historical and MRF ESP forecast MAE for DIRC2. MAE is shown as function of forecast lead time for all forecasts initialized between Apr and Jul. Median, extreme, and quartile values are shown.

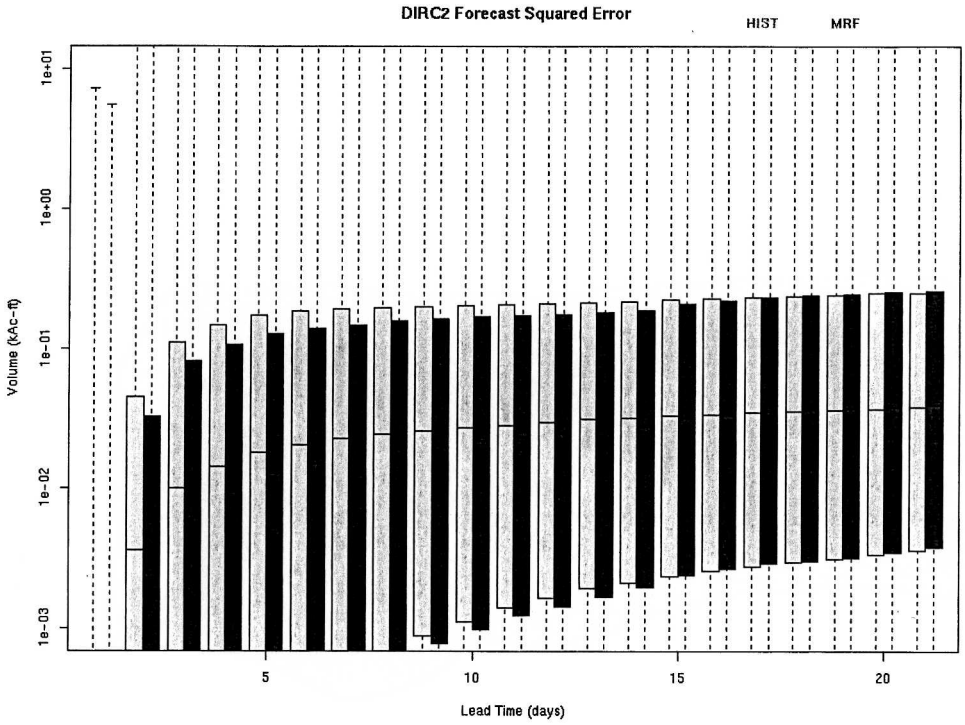

FIG. 8. Historical and MRF ESP forecast MSE for DIRC2. MSE is shown as function of forecast lead time for all forecasts initialized between Apr and Jul. Median, extreme, and quartile values are shown.

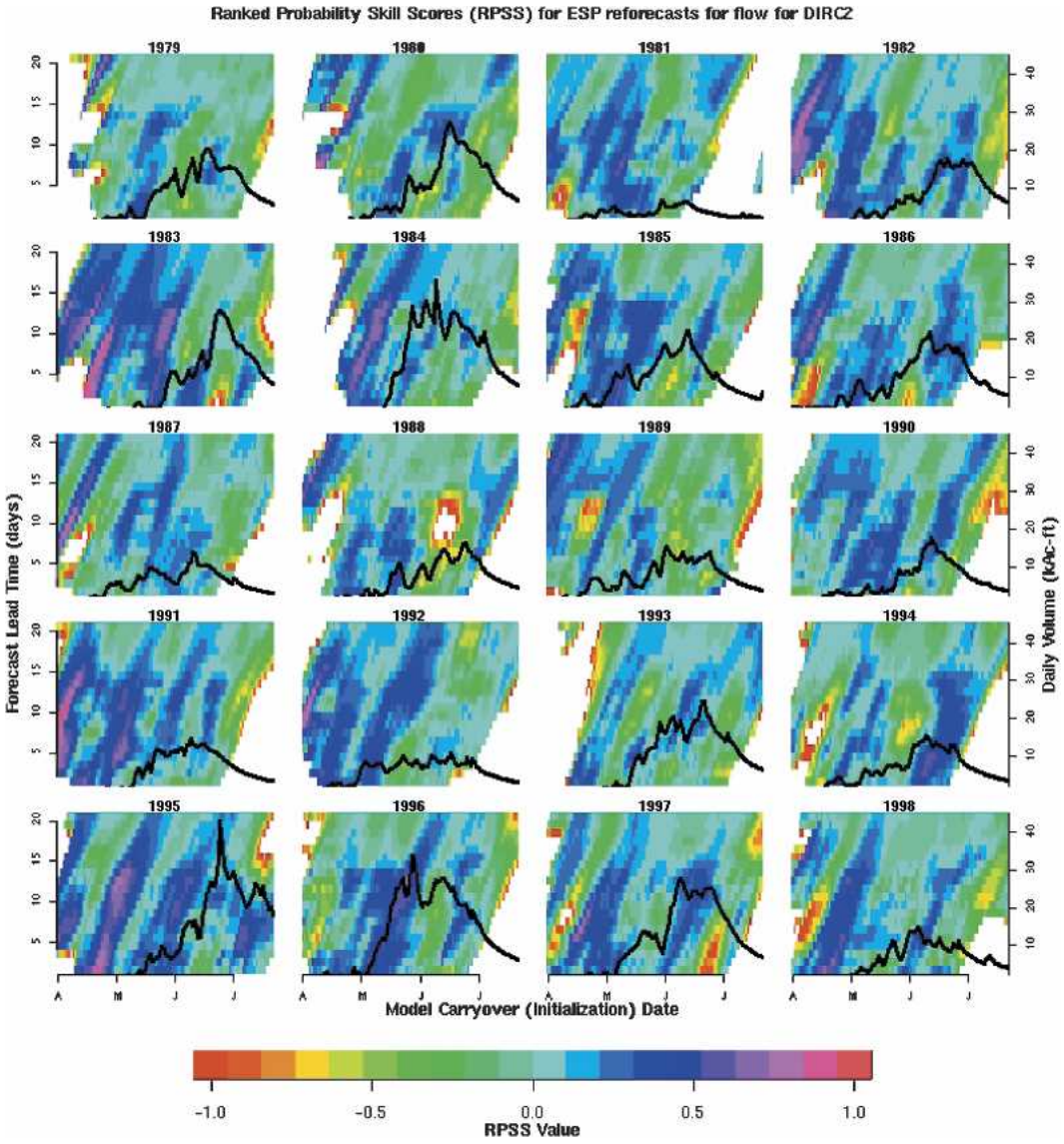

FIG. 9. DIRC2 MRF forecast RPSS by year. Forecast initialization time is on the *x* axis and forecast lead time is on the *y* axis. A 10-day running mean has been applied to the forecast intialization times. Superimposed black curves are the simulated flow for each year.

forecast initialization time between 1 January 1979 and 31 December 1998, and forecast 6-hourly lead time (up to 14 days). The variance of these ensembles reflects the observed forecast error in the downscaling process as well as the model error.

Figures 3 and 4 show the skill scores for the MATs and MAPs derived from this method as a function of forecast lead time. MAT skill scores show substantial improvement (up to 0.7) over climatological MATs during the first week. The second week shows only minimal improvement over climatology. RPSS values are greater for January than for July MATs for all lead times where positive RPSS values exist. As expected, the MATs show much higher skill than do the MAPs. MAP skill scores show only slight improvement (0.2) for the first 3–4 days after which no skill is added to the climatological MAPs.

e. ESP reforecasts

The ESP component of the NWSRFS was used to create reforecasts of the historical streamflow record (Day 1985). The reforecasting methodology is depicted graphically in Fig. 5. First, historical MATs and MAPs were used as input to the NWSRFS to create a simulated record of historical streamflow and model states

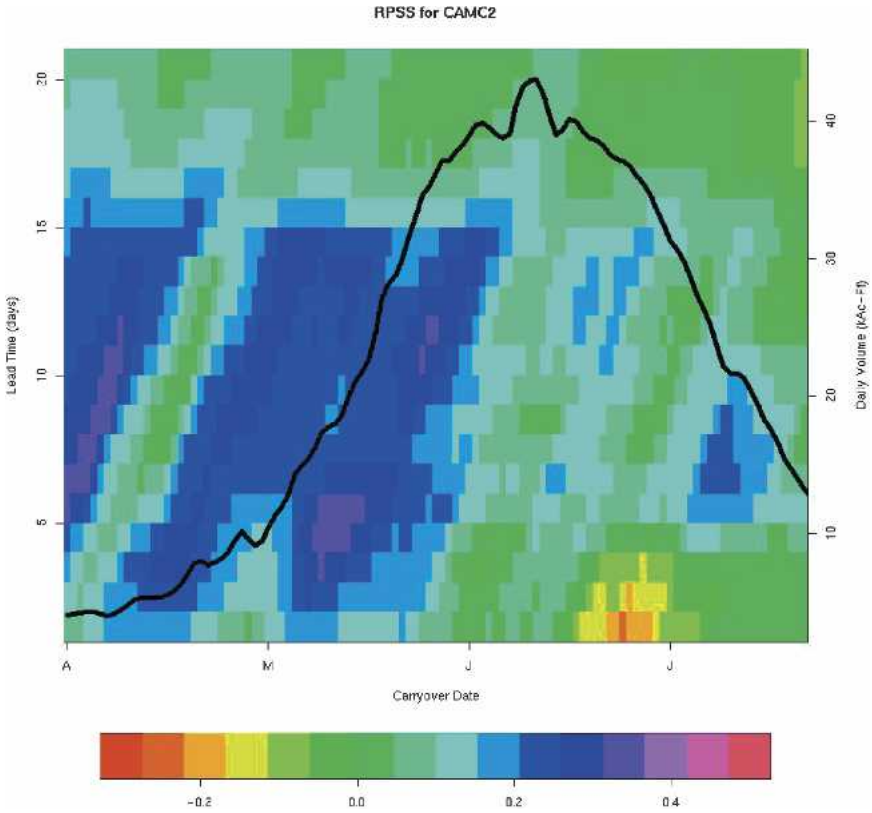

FIG. 10. DIRC2 MRF forecast mean RPSS over all reforecast years. Lead time (days) is displayed on the *y* axis and forecast intialization is displayed on the *x* axis. A 10-day running mean has been applied. Superimposed black curve is the climatological simulated flow.

for the period 1979–98 (Fig. 5). The model state is defined by the physical states shown in Table 1. Simulated hydrologic model states were saved for every day in the historical record. Next, for each reforecast in the historical record, the NWSRFS was initialized using the simulated model states, and the NWSRFS run forward in time using the downscaled MAP and MAT as input (Fig. 5). For comparison, the experiment depicted in Fig. 5 is repeated using historical MAP/MAT data in place of the downscaled ensemble inputs—for example, ensemble inputs to the NWSRFS from 1 to 15 May 1981 are composed of historical data from 1 to 15 May 1979, 1980, 1982, 1983, . . . , 1998 [this is the traditional implementation of the ESP approach for longer forecast lead times (Day 1985)]. To assess potential lagged effects of (potentially improved) inputs, the downscaled inputs are blended with the historical ensemble inputs on the 14th forecasted day, and all forecasts are run out to 25 days. In the blending period (day 14), the historical ensemble is linearly given more weight than the MRF ensemble.

Both sets of ESP reforecasts were made once each day between 1 January and 31 July for 1979 through 1999. Whereas the CDC MRF reforecasts were made every 12 h, ESP reforecasts may only be made once per day initializing at 0000 UTC. A 6-hr time step was used in ESP.

Traditional ESP forecasts weight the ESP ensemble member resulting from each year's historical observations equal to every other ESP ensemble member. In doing so, predictability is derived mostly by knowledge of the initial hydrologic states rather than from a forecast of weather or climate conditions. By using the traditional ESP forecast as a benchmark for comparison, we can assess the additional predictability that is obtained by including downscaled MRF forecast information.

3. Results

We assume a perfect hydrologic model for purposes of forecast verification. Here forecasts were verified against the simulated flow time series instead of the observed flow time series. This focuses attention on improvements due to the inclusion of meteorological forecast information instead of errors in the hydrologic model itself. Simulated and observed flows are plotted in Fig. 6 for DIRC2. Their difference is typically an order of magnitude less than the flow value, indicating the simulation is very near the observation in that head-

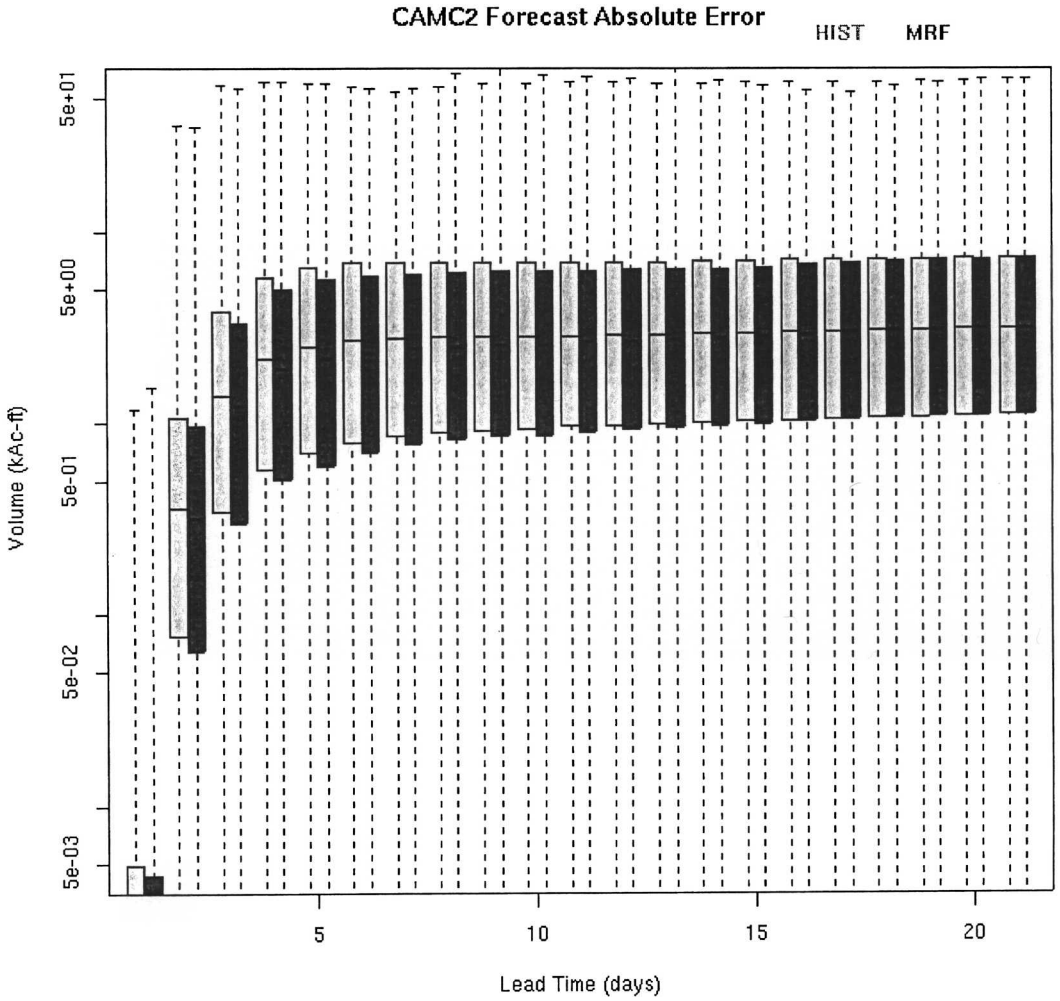

FIG. 11. Historical and MRF ESP forecast MAE for CAMC2. MAE is shown as function of forecast lead time for all forecasts initialized between Apr and Jul. Median, extreme, and quartile values are shown.

water basin. Outside of small headwater basins such as DIRC2, water regulation and diversions are widespread and sometimes large in magnitude, meaning that the perfect model assumption is necessary for forecast evaluation.

Traditional forecast verification statistics such as mean absolute error (MAE) and mean squared error (MSE) may be used for individual elements of a forecast ensemble. The mean and variance of these statistics will allow for rudimentary comparisons between different ensemble forecast systems. However, these statistics do not assess probabilistic forecast skill. The RPSS will be used to assess probabilistic forecast skill. Here, instead of using climatology as the reference forecast, we use the traditional ESP method (Day 1985).

Verification results are presented primarily for a headwater basin, inflow into Dillon reservoir (station identifier: DIRC2), and for the most downstream basin in the study area, the Colorado River at Cameo, CAMC2. DIRC2 is a headwater basin with minimal water diversions and/or regulations. Other headwater basins show results similar to DIRC2. By contrast, CAMC2 is the most downstream segment in the study area and includes water routed from all the other segments. CAMC2 also includes many water diversions and regulations. To compensate for this, simulated flow was used as the basis for forecast verification.

a. DIRC2

Figures 7 and 8 show the MAE and MSE averaged over all forecasts between April and July as a function of forecast lead time. Both MRF and historical reforecasts show increasing error with increasing lead time. For lead times up to 16 days the MRF reforecast has lower MSE and MAE values than for the historical reforecast. After 16 days the errors are essentially the

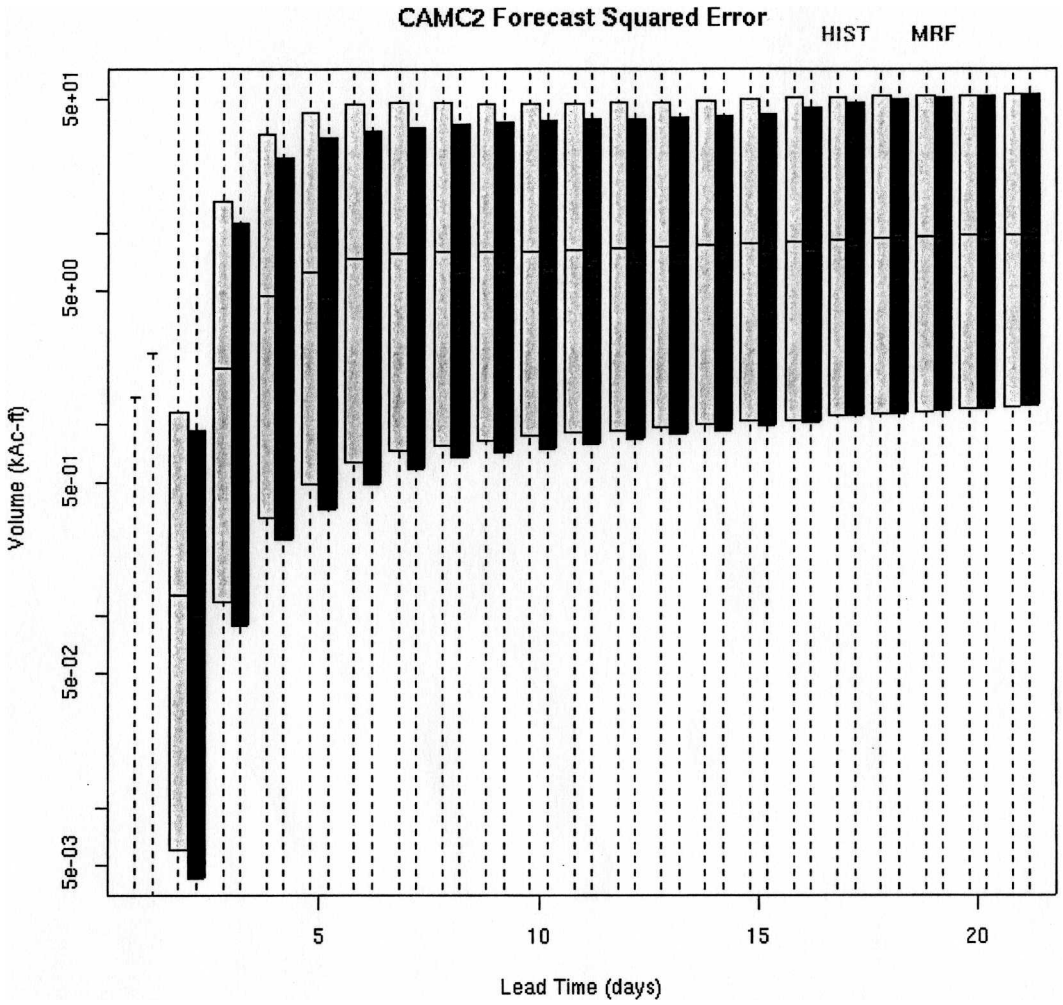

FIG. 12. Historical and MRF ESP forecast MSE for CAMC2. MSE is shown as function of forecast lead time for all forecasts initialized between Apr and Jul. Median, extreme, and quartile values are shown.

same. Forecast error reduction on the order of 10% is shown for forecast lead times of 1 to 9 days. These traditional measures give confidence that the MRF does in fact add value to the reforecasts. The value added will be further quantified through probabilistic forecast verification measures.

Figure 9 shows RPSS values (MRF against historical) as a function of forecast lead and initialization times for each reforecasted year. The hydrograph for each year is superimposed. In general, forecast skill improvements on the order of 40% are present during rises in the hydrograph. For example, 1996 shows a maximum RPSS value of about 60% in late May for forecast lead times of 10–15 days. This coincides with the final portion of the rising limb of the hydrograph, suggesting the MRF reforecast added significant information about precipitation, or more likely, temperature. Other years, notably 1988 and 1994, showed some negative RPSS values near the peak flow. In 1988, RPSS values less than 1.0 are present for lead times around 10 days near the mid-June peak flow, indicating a probable incorrect MRF forecast.

Figure 10 shows the climatological hydrograph and the mean RPSS values as a function of forecast lead and initialization times. Positive values are present throughout the entire domain with the exception of lead times greater than 8 days after about mid-July. RPSS values of 20%–40% are shown during the rising limb of the hydrograph for the duration of the MRF model run (14 days). This result indicates substantial and consistent improvements in forecast skill when the forecast is most important—before the peak flow occurs.

b. CAMC2

As was done for DIRC2, traditional forecast verification measures are applied to individual ensemble members. Figures 11 and 12 show the forecast MAE

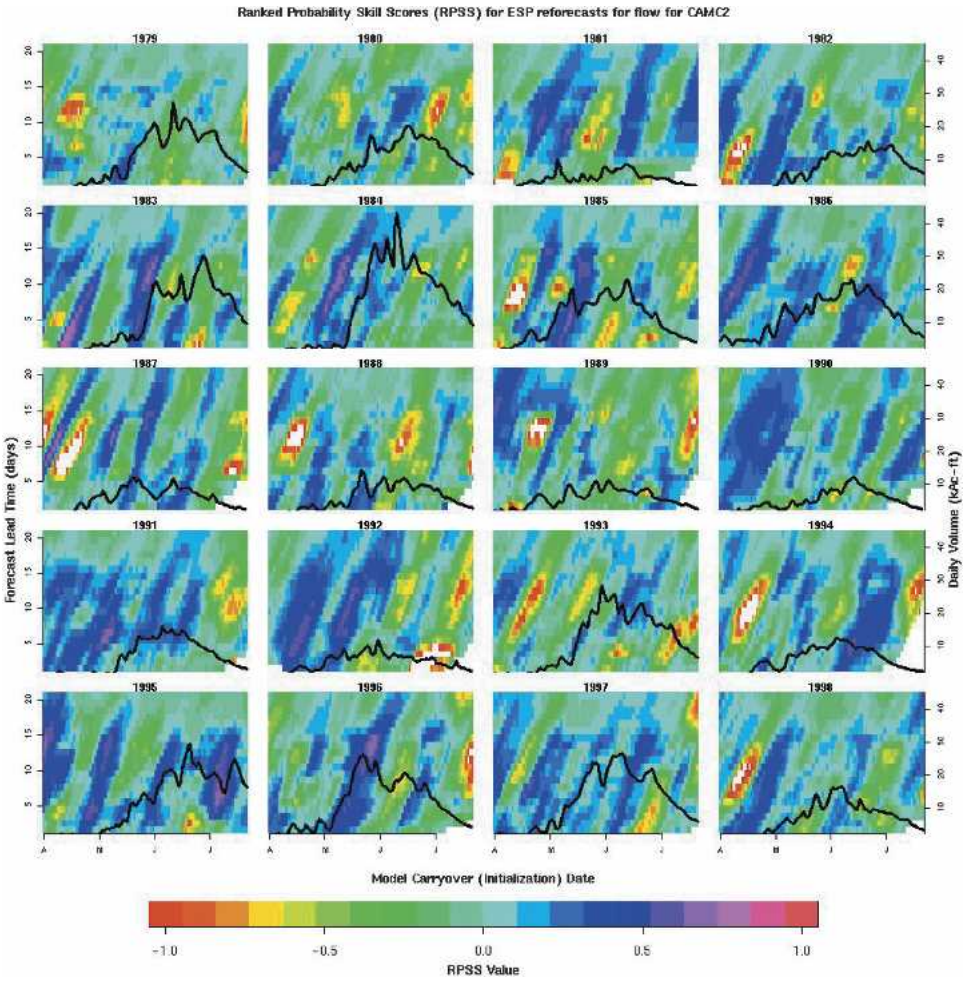

FIG. 13. CAMC2 MRF forecast RPSS by year. Forecast initialization time is on the *x* axis and forecast lead time is on the *y* axis. A 10-day running mean has been applied to the forecast intialization times. Superimposed black curves are the simulated flow for each year.

and MSE averaged over all forecasts between April and July as a function of lead time. As with DIRC2, the MRF reforecast error is less than the historical reforecast error up through lead times of 16 days. After that, the forecast errors are nearly equivalent. This gives confidence that the MRF does in fact add value to the reforecasts.

Figure 13 shows RPSS values for individual years. A 10-day running mean was applied to the forecast initialization date. Generally positive RPSS values are present during the rising limb of the hydrograph for all years and for all forecast lead times. Most years show a maximum RPSS value near 0.5, indicating a 50% improvement over the historical forecast. These maximum RPSS values generally occur near lead times of about one week as well as near the seasonal peak flow. Negative RPSS values exist for most years, particularly during the falling limb of the hydrograph, although for smaller areas than the positive values. Seasonal variations in RPSS may reflect cases when the forecast is most sensitive to temperature variations. Strong warming during the rising limb of the hydrograph will cause snowmelt, which will increase the rate of rise—the increased skill in the rising limb is likely because temperature forecasts have higher skill than precipitation forecasts (Figs. 3 and 4). Several years show substantial increases in forecast skill score out to three weeks.

Figure 14 shows the mean forecast RPSS for CAMC2 over all reforecast years as a function of forecast lead time and forecast initialization date. A 10-day running mean was applied to the forecast initialization times for all forecast lead times. The mean RPSS value during the April–July forecast period over forecast lead times up to 21 days is 12%. Strong forecast improvements are present during the entire rising limb of the hydrograph between mid-April and early June. Maximum RPSS

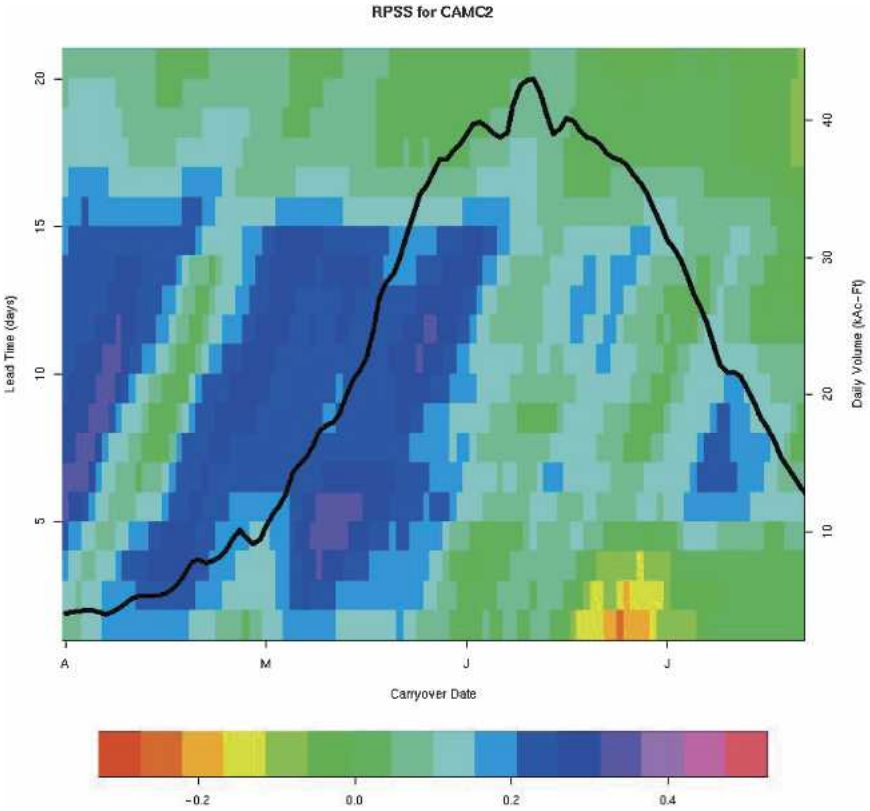

FIG. 14. CAMC2 MRF forecast mean RPSS over all reforecast years. Lead time (days) is displayed on the *y* axis and forecast intialization is displayed on the *x* axis. A 10-day running mean has been applied. Superimposed black curve is the climatological simulated flow.

values are about 0.4 and predominately occur at lead times of 3–6 days during the rising limb. Beyond 18 days, RPSS values are near zero, indicating, on average, that there is no forecast improvement beyond 18 days.

RPSS values were sorted according to the daily change in flow and plotted in Fig. 15. It is apparent from this figure that forecast skill is improved more strongly for situations where the daily flow is increasing from one day to the next. This again suggests the hydrologic states are sensitive to periods where temperatures are warmer than climatology in that the MRF is generally capable of forecasting with a 1–2-week lead time.

4. Summary and conclusions

The forecast skill improvements shown here through RPSS values are substantial. However, the RPS values were calculated against simulated flow rather than observed flow. Lower forecast skill should be expected when the RPSS is calculated against the observed flow and when the hydrologic model error is taken into account.

The current operational use of ESP does account for meteorological forecasts through the inclusion of deterministic temperature (days 1–10) and precipitation forecasts (days 1–3). It is hypothesized that the MRF ESP forecasts presented in this study may have a higher skill than the operational ESP forecasts because of the inclusion of probabilistic information. However, a historical archive of the deterministic station forecasts used in the operational ESP forecasts would be necessary to quantify the forecast skill of the operational ESP. Since the meteorological forecasts used in the operational ESP are based on the dynamic NCEP MRF and human-derived precipitation forecasts, such an archive of suitable length of record does not currently exist. A comparison of the operational ESP to the MRF ESP presented in this study will be addressed in a future study.

The increase in forecast skill due to the inclusion of the MRF forecast is encouraging. In 2004, CBRFC will have the capability to extend and use this method for any of its basins on a once per day operational basis. Similar methods may be applicable for other finer-resolution numerical weather models (i.e., the Meso Eta), which may lead to an increase in forecast skill for shorter lead times.

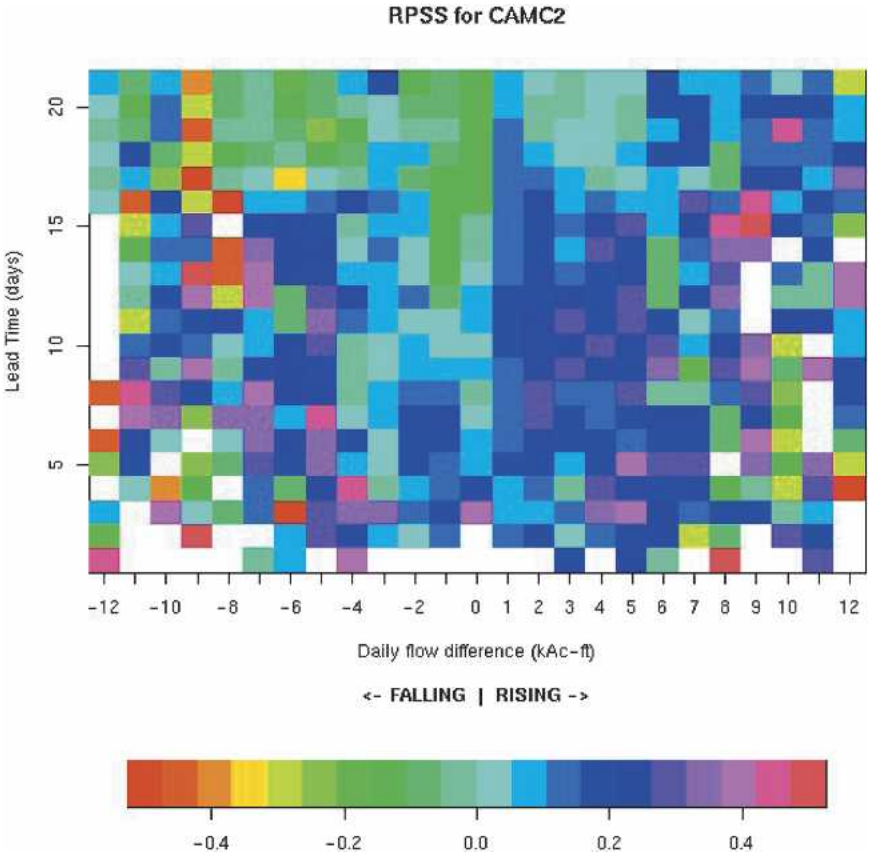

FIG. 15. RPSS values sorted and averaged according to the daily difference in simulated flow. Positive flow differences indicate a rising hydrograph, while negative differences indicate a receding hydrograph.

Acknowledgments. We are grateful to Jeffrey Whittaker and Tom Hamill at the Climate Diagnostics Center in Boulder, Colorado, for providing output from historical runs of the NCEP MRF model. This work was supported by the NOAA's NWS Advanced Hydrologic Prediction Services initiative. We thank Brent Bernard from CBRFRC for his GIS support. Clark and Gangopadhyay were supported by the NOAA GEWEX Americas Prediction Program (GAPP) and the NOAA Regional Integrated Science and Assessment (RISA) Program under award numbers NA16Gp2806 and NA17RJ1229, respectively.

REFERENCES

Clark, M. P., and L. E. Hay, 2004: Use of medium-range numerical weather prediction model output to produce forecasts of streamflow. *J. Hydrometeor.,* **5,** 15–32.

——, S. Gangopadhyay, L. E. Hay, B. Rajagopalan, and R. Wilby, 2004: The Schaake Shuffle: A method for reconstructing space–time variability in forecasted precipitation and temperature fields. *J. Hydrometeor.,* **5,** 243–262.

Daly, C., W. P. Gibson, G. H. Taylor, G. L. Johnson, and P. Pasteris, 2002: A knowledge-based approach to the statistical mapping of climate. *Climate Res.,* **22,** 99–113.

Day, G. N., 1985: Extended streamflow forecasting using NWSRFS. *J. Water Resour. Plann. Manage.,* **111,** 157–170.

Epstein, E. S., 1969: A scoring system for probability forecasts of ranked categories. *J. Appl. Meteor.,* **8,** 985–987.

Hamill, T. P., J. S. Whitaker, and X. Wei, 2004: Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts. *Mon. Wea. Rev.,* **132,** 1434–1447.

Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting,* **15,** 559–570.

Murphy, A. H., 1969: On the "ranked probability score." *J. Appl. Meteor.,* **8,** 988–989.

——, 1971: A note on the ranked probability score. *J. Appl. Meteor.,* **10,** 155–156.

Wood, A. W., E. P. Maurer, A. Kumar, and D. P. Lettenmaier, 2002: Long range experimental hydrologic forecasting for the eastern United States. *J. Geophys. Res.,* **107**, 4429, doi:10.1029/2001JD000659.