JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION

Vol. 45, No. 2 AMERICAN WATER RESOURCES ASSOCIATION April 2009

MID-RANGE STREAMFLOW FORECASTS BASED ON CLIMATE MODELING – STATISTICAL CORRECTION AND EVALUATION[1]

Jae H. Ryu, Richard N. Palmer, Matthew W. Wiley, and Sangman Jeong[2]

ABSTRACT: Mid-range streamflow predictions are extremely important for managing water resources. The ability to provide mid-range (three to six months) streamflow forecasts enables considerable improvements in water resources system operations. The skill and economic value of such forecasts are of great interest. In this research, output from a general circulation model (GCM) is used to generate hydrologic input for mid-range streamflow forecasts. Statistical procedures including: (1) transformation, (2) correction, (3) observation of ensemble average, (4) improvement of forecast, and (5) forecast skill test are conducted to minimize the error associated with different spatial resolution between the large-scale GCM and the finer-scale hydrologic model and to improve forecast skills. The accuracy of a streamflow forecast generated using a hydrologic model forced with GCM output for the basin was evaluated by forecast skill scores associated with the set of streamflow forecast values in a categorical forecast. Despite the generally low forecast skill score exhibited by the climate forecasting approach, precipitation forecast skill clearly improves when a conditional forecast is performed during the East Asia summer monsoon, June through August.

(KEY TERMS: streamflow forecast; climate variability; hydrologic models; monsoon.)

Ryu, Jae H., Richard N. Palmer, Matthew W. Wiley, and Sangman Jeong, 2009. Mid-Range Streamflow Forecasts Based on Climate Modeling – Statistical Correction and Evaluation. *Journal of the American Water Resources Association* (JAWRA) 45(2):355-368. DOI: 10.1111/j.1752-1688.2008.00292.x

INTRODUCTION

The water resources of Korea, China, and Japan depend largely on the summer monsoon season precipitation. Precipitation during this period is strongly influenced by the East Asian monsoon system (Ho and Kang, 1988; Kim *et al.*, 1998). Summer monsoons are caused by the physical and thermal interactions of the land and ocean. Since the land is more sensitive than the ocean to temperature during summer, it heats and cools more quickly. This results in the formation of low pressure at low levels, and is driven by the existence of lower atmospheric inflow from the ocean. This mechanism increases humidity and precipitation. Winter monsoons are characterized by the inverse of these conditions (Robinson, 1976; Lighthill *et al.*, 1981).

Heavy rains and humid weather occur during the summer monsoon. Winds arise from a southwesterly direction in Central and East Asia. More than 60% of the annual precipitation occurs during this period. Intense rain accompanying the tropical cyclones in

[1]Paper No. JAWRA-07-0103-P of the *Journal of the American Water Resources Association* (JAWRA). Received July 31, 2007; accepted July 28, 2008. **Discussions are open until October 1, 2009**.

[2]Respectively, Hydrologist, School of Natural Resources, University of Nebraska, Lincoln, Nebraska 68588-0988; Professor and Head, Department of Civil and Environmental Engineering, University of Massachusetts, Amherst, Massachusetts 01003-9293; Hydro Forecasting Engineer, 3TIER, 2001 Sixth Avenue, Suite 2100, Seattle, Washington 98121; and Professor, Department of Civil and Environmental Engineering, Kongju National University, Kongju, Republic of Korea (E-Mail/Ryu: jryu2@unl.edu).

the region contributes to flooding and can have significant impacts (e.g., dam failures, loss of human life, loss of livestock, and other economic damages). These characteristics of summer monsoons underline the importance of forecasts in mitigating damages associated with uncertain hydrologic events. Such complex climate characteristics, however, make it difficult to forecast the weather (precipitation and temperature) and streamflows several months into the future. This is also due to the weak relationship between the regional climate and the global physical hydrologic cycle, which is associated with the highly unstable climate variability in the northern hemisphere (latitude: N25-N45). Nonetheless, there has been strong interest in the quality of streamflow forecasts and their impacts on water resource operations.

From a hydrologic perspective, monsoon systems in East Asia impact the transition from periods of high to low seasonal flows depending on the magnitude of rainfall during the summer season. Accurate streamflow forecasts (particularly at lead times of three to six months), therefore, can provide valuable information to water managers. Such forecasts allow managers to provide extra flood control volumes when high flows are expected, to store more water at the beginning of a drawdown season when less than average flows are expected, and to encourage wise water use during unseasonably dry summer months. Mid-range forecasts, in particular, are valuable to water resource managers interested in supplying affordable water, sustaining the continuum of system reliability, and increasing annual water system revenues. The advantage of a forecast system is that responses can be initiated far in advance of the most significant impacts caused by hydrologic events, thereby minimizing the damages associated with them. Furthermore, these forecast systems can both inform decision makers, resource managers, and stakeholders and support decision making based on scientific facts and forecasts, rather than past operations or normal conditions.

Recently, general circulation models (GCMs) have been used for regional climate simulation and/or local scale forecasts, but substantial systematic precipitation and temperature biases preclude their direct use in hydrologic forecast modeling framework as forcing (Leung *et al.*, 1999; Roads *et al.*, 1999; Wood *et al.*, 2002). Because the spatial resolution of these products is relatively coarse, it is necessary to apply adjustment procedures to a specific regional watershed to minimize the error associated with a coarser grid. Leung *et al.* (1999) proposed a simple correction scheme that retains the basic statistics (e.g., mean and standard deviation) between the regional climate model (RCM) and the variable infiltration capacity (VIC) model, which is a macroscale hydrology model for water and energy balance (Liang *et al.*, 1994). This scheme uses an RCM-simulated monthly mean precipitation and surface temperature that is "corrected" with observed historical data. Hence, the new monthly mean is bounded by the range of the historic climate records. Alternatively, Wood *et al.* (2002) suggests a relatively simple approach to creating linkage between forecast outputs from GCMs and macroscale hydrologic models by using quantile-quantile mapping techniques on empirical distributions. First, the retrospective forecasts are compared with historic meteorological data and monthly cumulative distribution curves are generated to identify the bias. Next, the two cumulative distribution curves are used to transform GCM output into VIC input at given grid cells.

These two simple methods have been used to downscale a coarse climate model to a finer regional hydrologic model in both the Pacific Northwest and Eastern United States (U.S.) (Wood *et al.*, 2002). However, additional research is needed to improve the predictability of regional climate forecasts in terms of meteorological anomalies associated with precipitation and temperature forecast products. These products can be used, in turn, to forecast important hydrologic events.

This paper addresses the following issues: how to best develop seasonal streamflow forecasts using a climate model in the East Asian countries and then how to evaluate their quality in a hydrologic forecast framework. The outcome of the research and its statistical methodology can be used as a basis for similar forecasting efforts in diverse disciplines such as meteorology, agriculture, hydrology, and natural resources. This has potential to both maximize operational performance as well as provide economic benefits.

The paper is organized as follows: a brief description of the study area is first presented to justify the necessity of streamflow forecasts to better manage water resources in the region, followed by a detailed description of the models used in this study. The methodology is then described, including the statistical correction and evaluation techniques. Lastly, the quality of hydrologic forecasts associated with the climate forecasts is evaluated during the summer monsoon period, concluding with discussion of the results and future work.

MATERIALS AND METHODS

Study Area

The Geum River Basin is one of the largest watersheds in Korea. The Geum River drains some 9,810 km^2 and has a main stem length of 396 km

(Figure 1). The Geum River flows from south to north. Daechong Dam, which creates a reservoir of approximately 1,500 million m^3, provides water to several major cities with a combined population of approximately 3 million people. Upstream of Daechong Dam is Yongdam Dam (completed in 2001), which creates a reservoir of approximately 815 million m^3. Both dams play a key role in flood control, hydropower, and municipal and irrigation water in the region. The Geum River Basin, however, has experienced severe water conflicts that have led to considerable debate over what values should be established for environmental flows for both dams. In addition, uncertainty about future water demand in the face of rapid population growth in the region has fueled debate concerning water resources management.

Recent droughts, as well as flood concerns, highlight the necessity of streamflow forecasts to mitigate impacts caused by uncertain hydrologic events, and to provide useful insights for system managers in their decision-making processes. In this study, May was determined to be the appropriate month to predict hydrologic events (drought/flood) associated with the typical summer monsoon season (June-August). Earlier forecasts were found to be premature and forecasts later than this month were found to be too late to be effective.

Models. The Korea Meteorological Administration-Seoul National University (KMA-SNU) model and the Hydrologic Simulation Program-FORTRAN (HSPF) (EPA HSPF, 1970) are utilized as a climate model and a hydrologic model, respectively. The KMA-SNU has implemented a dynamic climate model based on global spectral models (Kim *et al.*, 1998). The KMA-SNU model includes a seasonal prediction system as part of the Seasonal Prediction Model Intercomparison Project 2 (SMIP2), which is organized by the Climate Variability and Predictability (CLIVAR) Working Group on Seasonal to Interannual Prediction (CLIVAR/WGSIP: http://www.clivar.org). SMIP2, the next generation of the SMIP forecasting system, extends the lead time of the forecast to two of the four seasons (winter, spring, summer, and autumn) and has analyzed recent climate retrospectively from 1980 to 2002. Each participating institute executes the model using two datasets (including an observed dataset) as well as its boundary conditions. This helps ensure that the monthly averages are preserved when monthly values are disaggregated to daily values. The model provides 10 ensemble members for seven months with observed initial conditions in February, May, August, and November for spring, summer, fall, and winter, respectively (Kang *et al.*, 2004).

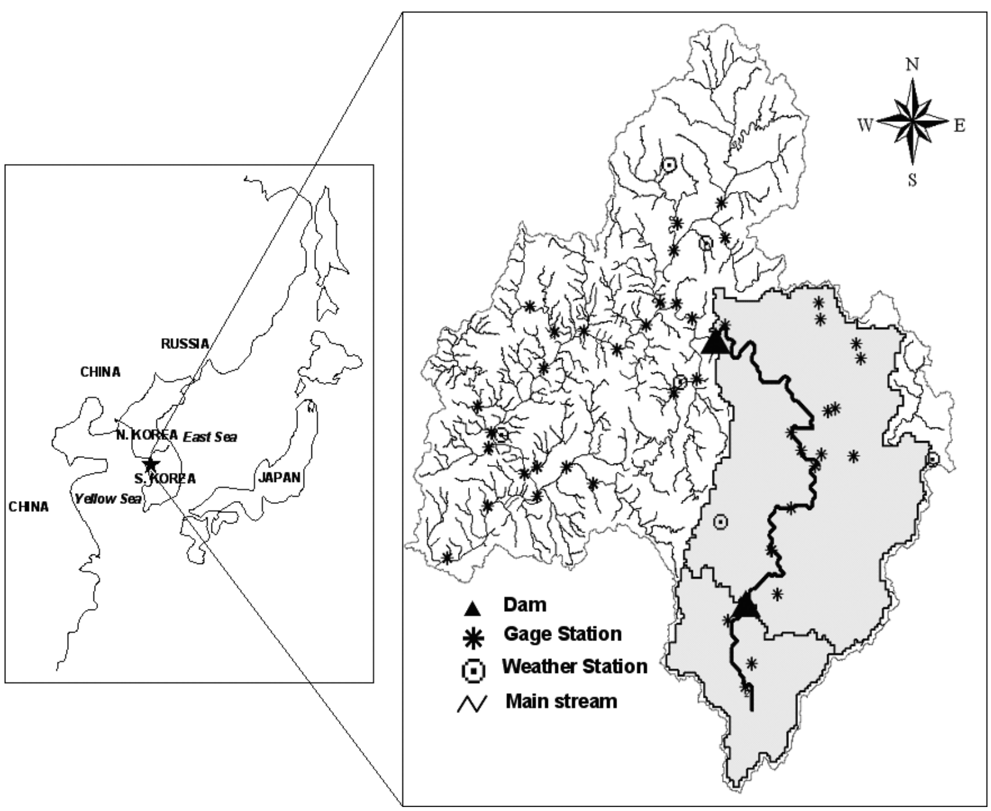

FIGURE 1. Map of Water System in the Geum River Basin.

As evidence of the successful performance of the KMA-SNU model, Kim *et al.* (1998) showed that the model is capable of simulating the climatology and interannual variations of global precipitation and circulation statistics reasonably well. The model's overall performance in simulating the climatological variations of summer monsoon statistics over the Asian-Western Pacific region is well-documented (Kang *et al.*, 2002). All models produced excessive rainfall in the Indian monsoon region. A second group of models, including KMA-SNU, simulated more precipitation in the subtropical western Pacific, but their rain band was somewhat shifted toward the north, which strongly affects the rainy season in East Asian countries (Kang *et al.*, 1999). Furthermore, Kang *et al.* (2002) emphasized that even if the northward climatologically intraseasonal oscillation components are simulated, their phases are shifted by about a month lead time so that the models indicate an early onset of the East Asian summer monsoon. This valuable information can be utilized to improve streamflow forecasts as well as system operation in this study.

The hydrologic model used in this research is a combination of the Better Assessment Science Integrating Point and Nonpoint Sources (BASINS) 3.0 (EPA BASINS, 2001) and the HSPF model. The main components of BASINS are: (1) a data management tool that allows the user to get nationally archived hydrologic and meteorological databases; (2) watershed delineation tools; (3) utilities for classifying the digital elevation maps, land use, soils, and water quality data; and (4) watershed characterization reports that allow the user to present the output of information on selected watersheds (http://www.epa.gov/waterscience/basins/basinsv3.htm).

The BASINS model provides a Digital Elevation Model (taken to be 6.5 km resolution in this study) as a primary dataset for HSPF. Each sub-basin is simulated using HSPF's pervious land segments (Forest, Agricultural, and Urban Built-up), impervious land segments (Urban Built-up), and streams or mixed reservoir segments (RCHRES). HSPF employs several storage zones to represent the storage processes that interact and occur simultaneously on the land surface and in the soil columns. Each sub-basin consists of at least two soil layers, including an upper-zone soil layer and a lower-zone soil layer. Raindrops move to existing storage and leverage the water level up for infiltration and runoff processes. Redundant soil moisture in the upper storage can infiltrate to the lower storage and groundwater storage, and may be routed as runoff. The upper-zone soil layer responds quickly to storm events, while the lower-zone soil layer controls interflow and ground base flow. The RCHRES simulates the flow of water in the tributary that drains each subwatershed. To create a forecast, meteorological data for the three years preceding the forecast is run through the model to capture the initial conditions.

Model calibration procedures are performed as follows: the annual water balance is maintained by adjusting evaporation data when necessary, then monthly water balances are evaluated by changing model parameters related to interflow and ground water. The generated hydrologic simulations, using observed station data (precipitation and maximum and minimum temperature) are used to determine forecast skill (described later). This process indicates the sensitivity of the forecast to the meteorological input rather than possible errors in the hydrologic model itself (Clark and Hay, 2004).

Data

The KMA-SNU model provides three-month predictions of temperature and precipitation. These forecasts are made with a model that uses a spatial domain of 2.5 degrees and a one-day time-step. The KMA-SNU also provides six different "hindcasts" of the GCM for each year from 1980 to 2002. For each year, six simulations are made with different initial conditions. For instance, the three-month forecast runs for 2004 are made from the December 2003 observed atmospheric initial states and are derived from six different initial conditions for every month. Hindcast runs are basically "retrospective" Atmospheric General Circulation Model (AGCM) simulations that are forced with the global observed sea surface temperatures (SSTs). These hindcasts represent the forecasts that would have been made had the forecasting system been in place.

The hindcasts can be used to compare the results of forecasts and historic weather conditions. Both the hindcasts and the forecasts are made for a three-month period from June to August (typical Asian summer monsoon). For application to a specific watershed, the KMA-SNU forecasts must be systematically corrected to minimize the error associated with a coarse grid. To be used in a hydrology model, the corrected forecasted data must be spatially distributed throughout the basin.

For weather station data, this study uses daily precipitation and maximum and minimum temperature data from a network of more than 57 KMA climate observing stations across South Korea. These data were downloaded from the KMA meteorological database. Records at most of these stations began in 1963 and are continuous to the present. Only stations that are well maintained and have long (and essentially continuous) records from 1963 to 2001 are used for evaluating the accuracy of the KMA-SNU forecasts. For hydrologic modeling purposes, only one station, Geumsan weather station (KMA 238), is used. The Geumsan station is located within the basin and has a long, continuous record (Figure 1).

Bias Correction

As a first step in evaluating streamflow forecasts associated with the KMA-SNU model, the systematic biases in the model's forecasts of precipitation and temperature are examined with five distinct procedures: (1) transformation, (2) statistical correction, (3) observation of ensemble average, (4) improvement of forecast, and (5) streamflow forecast skill test. Figure 2 illustrates the procedure for assessment of streamflow forecast measurements associated with the KMA-SNU climate forecast. A statistical correction procedure is employed to minimize the spatial and temporal differences between the two models (the climate model and the local hydrology model). This process is conducted in a probability framework using the normal density function. Then, an ensemble average forecast is produced. Potential forecast improvement is determined based on the correspondence between the observed conditional mean and the conditional forecast (Murphy, 1993). These measures classify the forecast/observation pairs into groups according to the value of the forecast variable and characterize the conditional distributions of the observations given in the forecasts. Finally, a forecast skill test is used to display the relative accuracy of a set of forecasts corresponding to an observed value, with respect to a set of reference forecasts (forecasts in random space in this case, e.g., randomly selected values from a normal density function in a given month). Forecast skill is usually measured using a

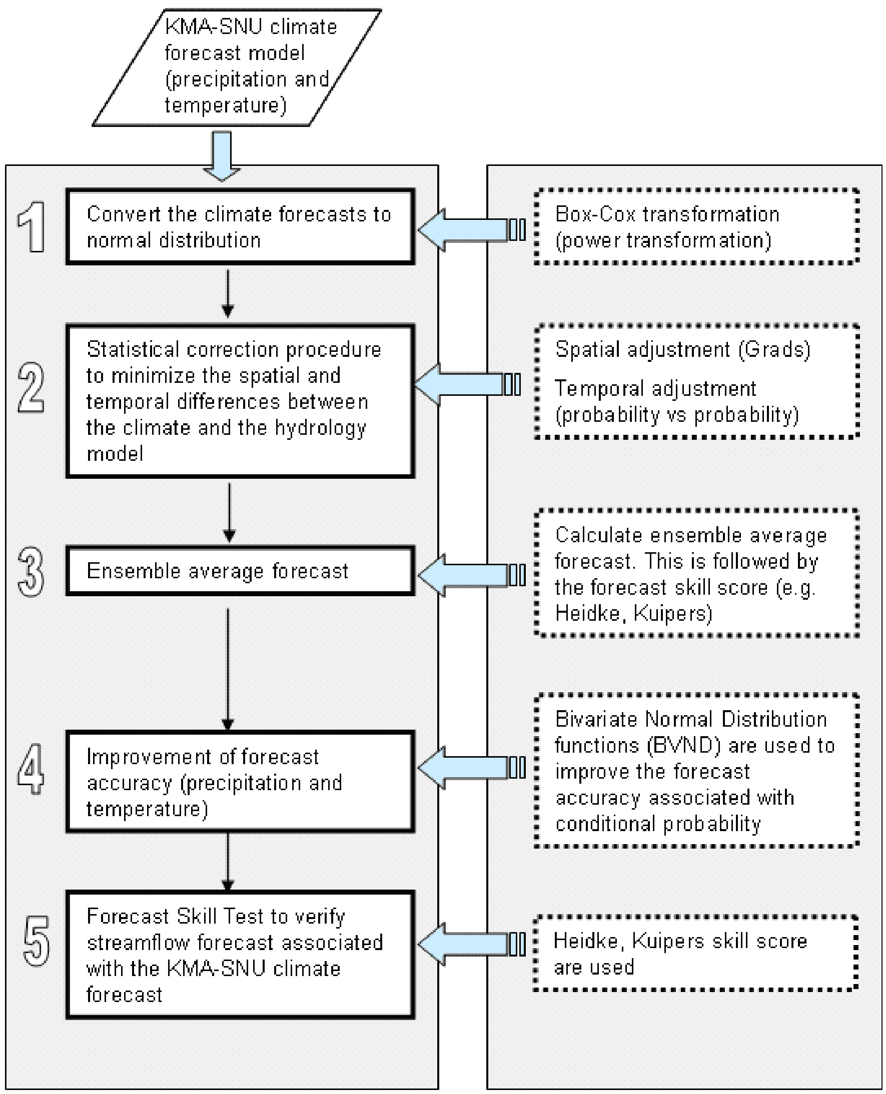

FIGURE 2. Methodology for Assessment of Streamflow Forecasts Associated with KMA-SNU Climate Forecast.

prescribed skill score method such as the Heidke Skill Score (HSS) or the Kuipers Skill Score (KSS) (Heidke, 1926; Hanssen and Kuipers, 1965). For this particular application, however, a root mean square error (RMSE) method is also used as an additional measure of the forecast performance.

Transformation. A common practice in hydrology is to transform the forecast and observed values into normally distributed random variables by taking logarithms or some other Box-Cox transformation (Box and Cox, 1964; Stedinger, 1980; Shapiro, 1990). One reason for the wide application of the normal distribution is that many climatic variables (such as the average monthly temperature) and hydrologic variables (such as the average monthly streamflow) are normally distributed or approximately normally distributed. Although the data are approximately normally distributed, they appear to preserve the normality through statistical goodness-of-fit tests such as histograms, probability plots, and quantile-quantile plots, as well as through classic normality tests (e.g., regression, chi-square, and moment tests) (Shapiro, 1990). Another reason for transforming data to a normal distribution is that other statistical tests can be derived from the normal distribution. The normal distribution can be easily related to many other theoretical distribution functions (Log-normal, Poisson, Binomial, and Gamma). Figure 3 presents the average January precipitation at the Geumsan weather station, and demonstrates the results of a typical transformation.

Statistical Correction. To apply the coarse gridded climate forecasts to local climatology, the regional biases and the spatial and temporal discrepancies between the climate model and the historical data must be addressed (Clark and Hay, 2004). Since the watersheds in this study are significantly smaller than the spatial grid of the KMA-SNU, these forecasts must be spatially adjusted. In addition, the KMA-SNU forecasts and the hydrology model operate at different time steps; therefore, the forecasts must also be temporally disaggregated. Note that for the temporal adjustment step of this research, the additive and multiplicative techniques are utilized for temperature and precipitation, respectively, as described by Wood *et al.* (2002). Many global and local statistical downscaling techniques such as regression (Enke and Spekat, 1997), canonical correlation (Karl, 1990), neural networks (Wilby and Wigley, 1997), analogues (Zorita and Storch, 1999), and clustering (Gutierrez *et al.*, 2004) have also been used as methods for incorporating global climate information into local meteorological and hydrological applications. However, the proposed method in this study is relatively simple and straightforward.

Statistical correction is performed for each weather station, treating each point individually, that is, the monthly climatologic distributions (precipitation and temperature) are independent of surrounding stations. For example, bias correcting a monthly retrospective forecast for June-August requires two individual normal distribution functions. The 144

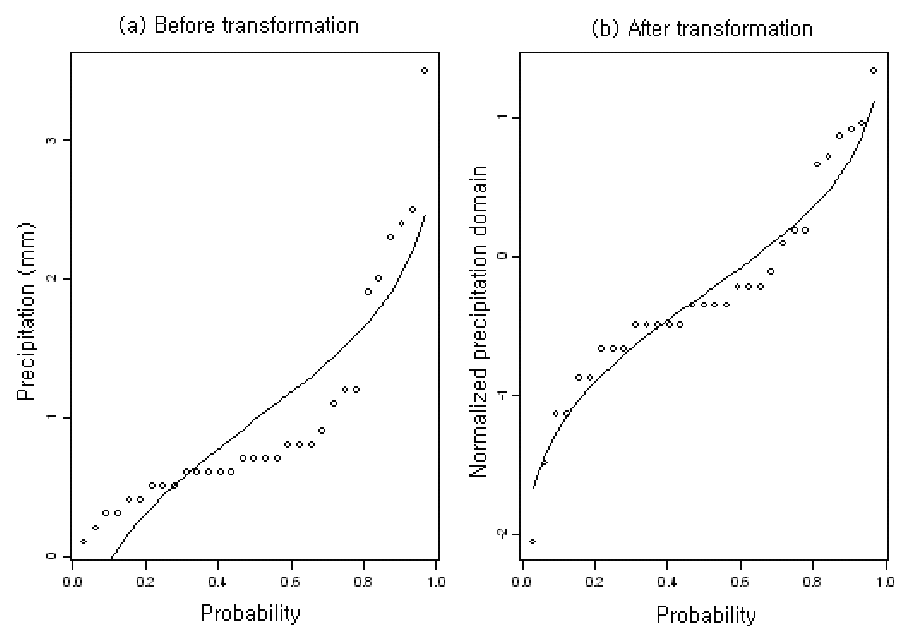

FIGURE 3. January Average Precipitation Data for Geumsan Weather Station in Geum River Basin, 1973-2003, (a) Before and (b) After Data Transformation. The solid line represents a normal curve fitted to the data.

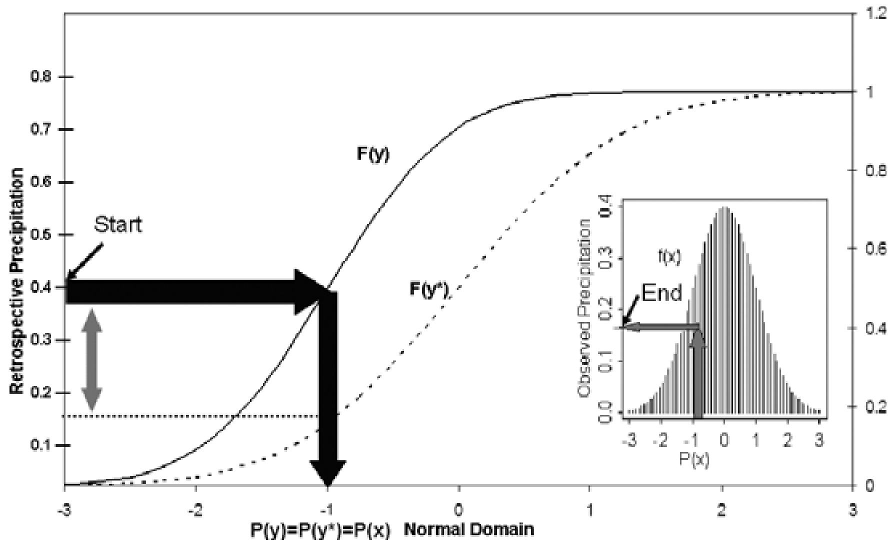

FIGURE 4. Statistical Correction Procedure of Precipitation Data (mm) for KMA-SNU Forecast (y) to Regional Observed Data (x).

values (24 years and 6 ensembles) of hindcasts are transformed to a normal distribution, and the 30 values of historic data are separately transformed to normal distribution. Each month has different means and standard deviations.

The first step of statistical correction compares the retrospective (hindcasts) to historic meteorological data. The two probability density functions (PDFs) – transformed normal density of hindcasts and historic data – are used to transform the KMA-SNU forecasts into appropriate values for the specific weather station. The transformations imply that probability mapping is the most appropriate way to relate the KMA-SNU forecast to the "statistically corrected forecast." Figure 4 represents the statistical correction procedure for monthly precipitation, which relates one KMA-SNU grid point with one weather station. The probability of the initial retrospective forecast of average precipitation for a given month is the probability $P(y)$ along the retrospective density function $f(y)$. This $P(y)$ relates to $P(x)$, which produces a "statistically corrected" forecast, $P(y^*)$. For instance, the hindcasted KMA-SNU indicates average monthly precipitation of 0.4 mm ("Start point" in Figure 4). This relates to about 16% probability in normal domain x, which produces an observed value (0.18 mm) having the same probability. Finally, the initial value of the hindcasted value (0.4 mm) derived from the climate forecast model is replaced by an observed local value (0.18 mm). This implies that the forecast value is statistically corrected by 55%. The double-sided direction of the arrow in Figure 4 indicates the value of the statistical correction. Figure 5 demonstrates a convenient means of summarizing the distribution of the hindcasts, observations, and statistically corrected hindcasts. For the Geumsan precipitation data, the box plots for the hindcasts in June indicate symmetrical distributions.

Obvious differences between the median of retrospective $f(y)$ and observed precipitation $f(x)$ suggest that the hindcasts may be biased. On the contrary, the central tendency and variability between statistically corrected $f(y^*)$ and observed data $f(x)$ appear to preserve the same statistical properties such as the mean and variance. In particular, the upper and lower whiskers are approximately equal in length and the means of the boxes are approximately the same size. Large changes in the median value between the "before" statistical correction, $f(y)$, and "after" statistical correction, $f(y^*)$, of the hindcast data suggest that the retrospective forecasts are

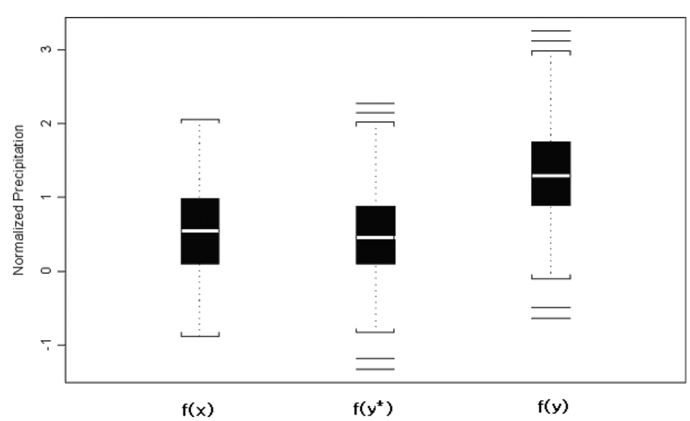

FIGURE 5. Box Plots of Marginal Distribution of June Precipitation Observations $f(x)$ and Forecasts $f(y^*)$ at Geumsan Weather Station, Respectively. Note: $f(y)$ is before statistical correction and $f(y^*)$ is after statistical correction.

biased. The observed precipitation distribution shows somewhat smaller variability than that exhibited by the retrospective precipitation forecasts, as indicated by the outliers of the boxes and the differences between the 0.90th and 0.10th quantile values.

Ensemble Average Forecast. Another valuable aspect of the forecast is improved accuracy, which is achieved by allowing a fuller range of estimates of the initial state of the atmosphere. Wilks (1995) noted that the ensemble average can be expressed as the atmosphere state corresponding to the center of the ensemble in phase space for some time in the future, and approximates the center of the stochastic-dynamic probability distribution at that future time. This makes the ensemble average forecast a good indication of the central tendency for future possible climate characteristics as well as in dynamic random space, making it another useful tool in determining the quality of the forecast (Wilks, 1995). The ensemble average forecast is obtained simply by averaging the ensemble members for the forecast period. Figure 6 presents the forecast distribution in each month. For example, the June forecast (Figure 6a) exhibits a tendency toward under forecasting precipitation during the period from the mid to late 1990s. Overall, average forecast is poor until 1999. On the other hand, during this period, the hindcasts still appear to be relatively biased over the entire range of forecast values.

Improvement of the KMA-SNU Forecast. Because the previous forecasts proved to be of poor quality despite statistical correction, other approaches were explored. A general framework for a forecast can be based on the joint (probability) distribution of forecasts and observations and on the conditional and marginal distributions (Murphy *et al.*, 1989). Since the normalized observed distribution $f(X)$ and forecast ensemble average distribution $f(Y^{**})$ are fairly symmetrical, it is reasonable to model their joint behavior as a bivariate normal distribution. A very useful property of the bivariate normal distribution is that the conditional distribution of one of the variables, given any particular value of the other, is normally distributed. If the average ensemble forecasts and observations are denoted by Y and X, respectively, then the joint distribution can be denoted by $f(Y, X)$. The joint distribution $f(Y, X)$ specifies the relative frequency of the occurrence of a particular combination of values of retrospective and observed data. Two normal distribution functions $f(Y)$ and $f(X)$ are already calculated using the series of statistical transformation and statistical correction procedures described previously. Based on the conditional distribution theorem, $f(Y, X) = f(Y|X)f(X)$ or $f(Y, X) = f(X|Y)f(Y)$, the joint distribution is the conditional distribution of Y at a given X multiplied by marginal distribution X. The value indicates a function proportional to a conditional distribution of X given a particular value of Y. The parameters for these conditional normal

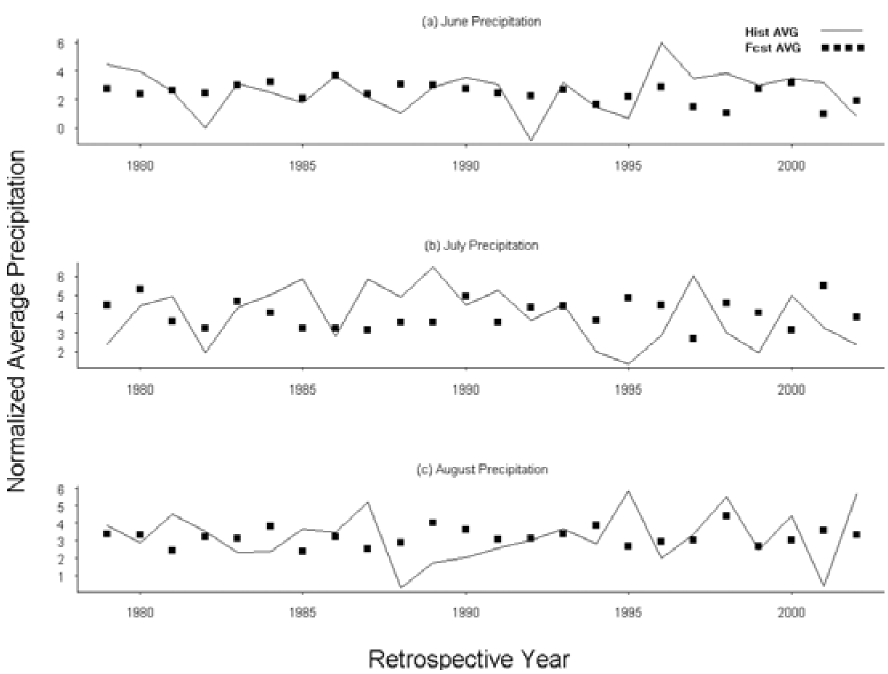

FIGURE 6. Comparison of the Precipitation Forecast for Three Different Months over Historic Records: (a) June, (b) July, and (c) August. Note: square dots indicate ensemble average of monthly precipitation.

distributions can be calculated from the five parameters of the bivariate normal distribution (see Appendix).

Forecast Skill Scores. The streamflow forecast is first analyzed for skill using HSS and KSS, which are widely used for forecast verification (Wilks, 1995). The Heidke score measures the accuracy of the forecast relative to the accuracy of random chance. This score calculates the fraction of correct forecasts after eliminating forecasts that would be correct because of purely random chance. This representation can also be interpreted as a percentage improvement over random sampling. A typical score ranges from −1.0 to 1.0. A value of 1 is considered a perfect forecast while a value of 0 indicates that the forecast has the same skill that would be expected by random sampling. A negative value implies that the forecast performs worse than random chance. The difference between HSS and KSS is that, unlike the HSS, the KSS formulation is based on an unbiased random sampling. Equations of HSS and KSS are available in the Appendix (see Equations A5 and A6).

To apply HSS and KSS to Geum River streamflow forecasts, the probability distributions created by the set of forecast values are converted into a categorical forecast. This conversion is performed by assuming a threshold that defines a region around historic average streamflow values; this is considered the average value region. The median forecast value is then placed into the multi-categorical format (e.g., above normal, normal, and below normal). The threshold is chosen in advance, using half of a standard deviation in the given month.

The probability space is thus divided into four regions. Figure 7 illustrates the threshold that would be chosen according to the mean of observed historic data. The upper and lower dotted lines represent the half of standard deviation above and below the mean of the given observed month, respectively. The hit rate in Equation (A5) is calculated as follows: if both forecast and observation fall in the same boundary, no matter what the boundary is (e.g., 1-1, 2-2, 3-3, and 4-4), a hit is considered to have occurred and is recorded; if both forecast and observation fall either above or below the normal (e.g., 1-2, 2-1, 3-4, and 4-3), a hit is recorded; and if both forecast and observation are adjacent to each other (e.g., 1-2, 2-1, 2-3, 3-2, 3-4, and 4-3), a hit is recorded. Based on this condition, Table 1 shows a 4-by-4 contingency table and the associated alphabetic letter corresponding to the hit criterion.

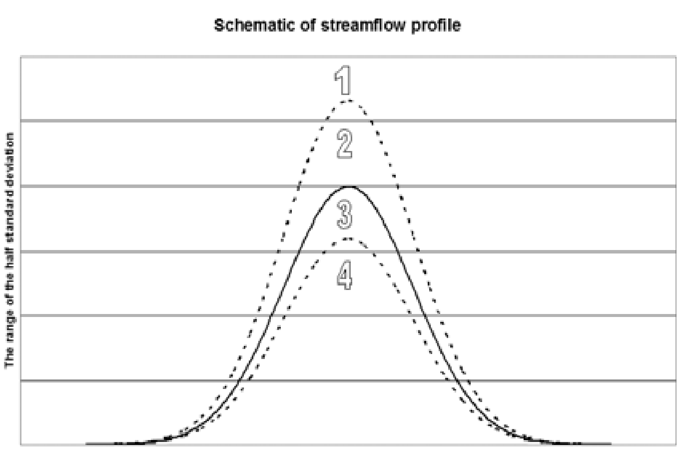

FIGURE 7. Schematic of Determination of Threshold.

RESULTS

Figure 8 illustrates a bivariate normal distribution between hindcasts of precipitation and the corresponding observed precipitation for Geumsan station for the period June 1980-August 2002. The correlation of actual July precipitation to the forecast appears to be greater than correlations for other months. The plots in Figure 9 present normalized data for the actual precipitation, the statistically corrected precipitation, and the conditional predicted precipitation, respectively. Overall, the result shows that the conditional predicted precipitation is more accurate during severe drought periods (1994-1995). It is also possible to express the increased accuracy of the forecast by evaluating each member of a collection of paired comparisons (observed *vs.* statistically corrected hindcasts and observed *vs.* conditional predicted forecasts after statistical correction). The RMSE between observed and forecasted precipitation is presented in Figure 10. The upper portion of the figure (above 45° linear line) represents a condition

TABLE 1. Categorical Streamflow Forecasts for Threshold Boundary.

	1	2	3	4
1	A*	B*	C	D
2	E*	F*	G*	H
3	I	J*	K*	L*
4	M	L	O*	P*

*Hits satisfying threshold boundary condition, alphabets indicate placeholders for the values in the cells when they are available. Thus, these letters will be replaced by the values in shaded areas in Table 3.

in which the hindcasts are more biased than the conditional prediction value. This result shows that the conditional forecast diminishes the absolute errors between hindcasts and observed data and improves forecast performance (Table 2), and implies that the conditional forecast technique is a better estimation technique than direct estimation after statistical correction. It also appears that there is little increase in skill for the June forecast, but that the forecast skill in July and August increased significantly (by up to 26%). However, this does not necessarily mean that the July forecast becomes more accurate than the other months, even if the forecast performance is significantly improved. This fact is discussed in the next section.

Figure 11 shows the distribution of hindcasts and modeled streamflow associated with climate forecasts during monsoon seasons (June-August) over historic periods. The solid line (diamonds) and dotted line (triangles) represent the total monthly streamflow as modeled and forecasted, respectively. The results show that streamflow forecasts in July do not make good predictions for the years of hydrologic events (e.g., 1987, 1989, and 1997) even if the precipitation forecast is much improved through statistical correction procedures (Figure 10). This implies that the July precipitation forecast in the climate model, in particular, is relatively more biased than the forecasts for other months. In other words, July is the most difficult month during the monsoon season to forecast both precipitation and streamflow.

Table 3 summarizes the four streamflow type outcomes on each forecasting occasion, respectively. Using the HSS equation, the Heidke score for the 4-by-4 contingency table in Table 3 is computed as follows. The hit rate (HR) is $\mathrm{HR} = \sum_{i=1}^{I} P(y_i, x_i) = (1/66) + (0/66) + (1/66) + (4/66) + (5/66) + (4/66) + (5/66) + (3/66) + (9/66) + (10/66) = 0.636$. The hit rate for the random reference forecasts (HRF) is $\mathrm{HRF} = \sum_{i=1}^{I} P(y_i)P(x_i) = (0.015)(0.227) + (0.227)(0.227) + (0.015)(0.212) + (0.227)(0.212) + (0.424)(0.212) + (0.227)(0.182) + (0.424)(0.182) + (0.333)(0.182) + (0.424)(0.379) + (0.333)(0.379) = 0.662$. Therefore, by the equation HSS, the Heidke Skill Score = −0.07. Similarly, KSS can be computed using the equation KSS that used an unbiased hit rate in random space denoted as $\sum_{i=1}^{I} P(x_i)^2 = (0.227)^2 + (0.212)^2 + (0.182)^2 + (0.379)^2 = 0.273$. The Kuiper Skill Score = (0.635-0.662)/(1-0.273) = −0.04. The small difference between HSS and KSS implies that the forecast is slightly biased in comparison with unbiased random space (Wilks, 1995).

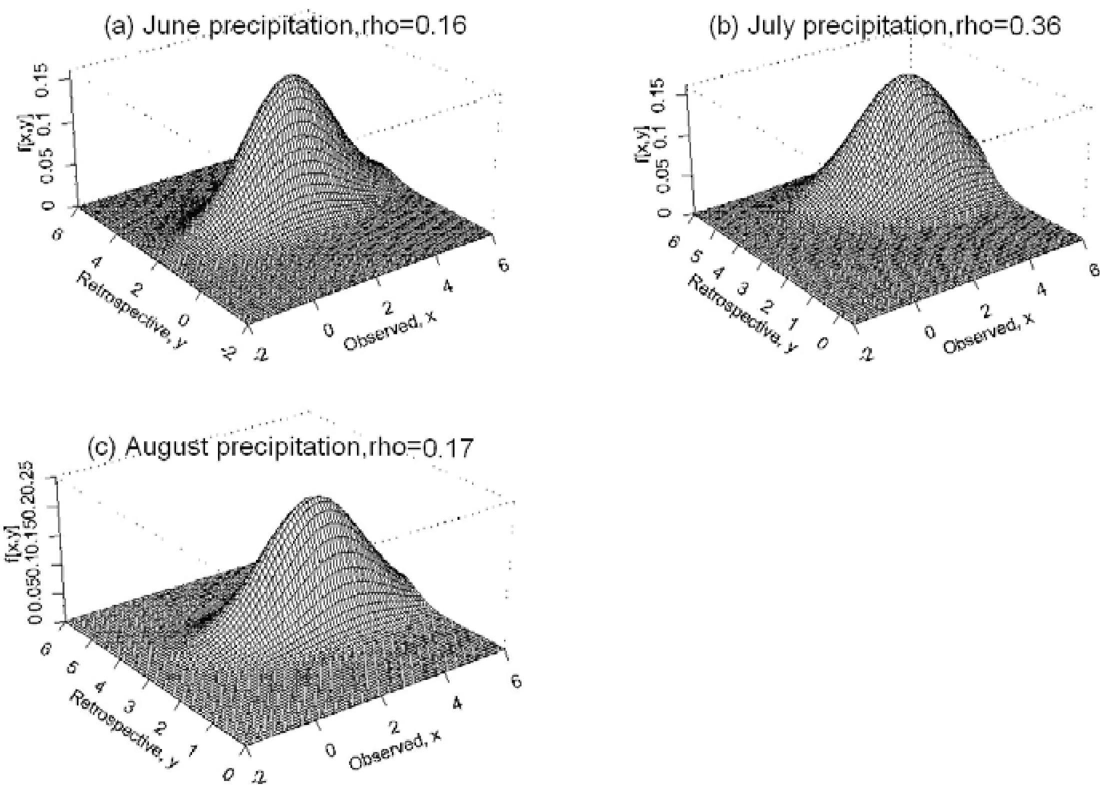

FIGURE 8. Perspective View of a Bivariate Normal Distribution for Geumsan Station in Different Months (June, July, and August).

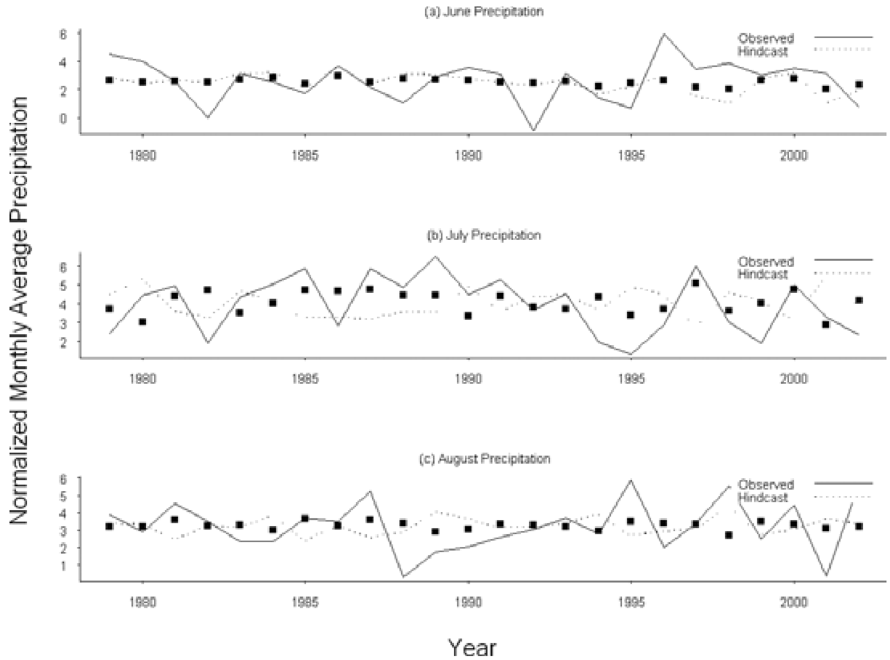

FIGURE 9. Comparisons of Normalized Monthly Average Precipitation with Observed, Hindcast, and Conditional Predicted Precipitation for (a) June, (b) July, and (c) August. Note: square dots indicate conditional predicted precipitation.

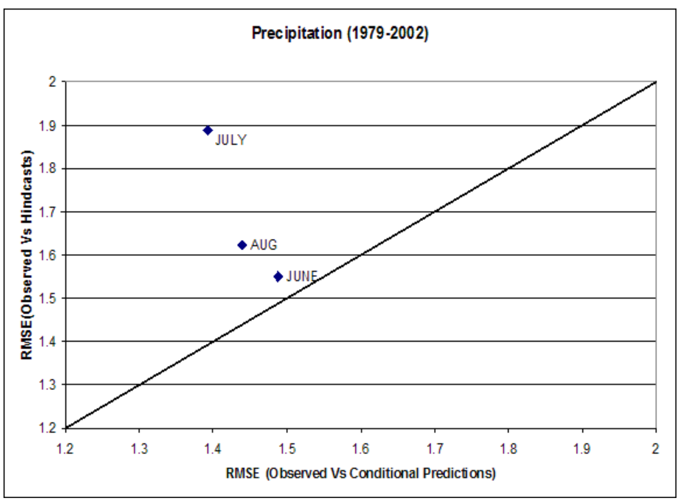

FIGURE 10. Root Mean Square Error Between Observed and Forecast, and Observed and Conditional Predicted Values over Historic Data.

TABLE 2. RMSE Between Observed and Forecasted, and Observed and Conditional Predicted Values Over Historic Data.

	RMSE	
Month	**OBS/Conditional Prediction**	**OBS/Hindcast**
June	1.49 (4%)	1.55
July	1.39 (26%)	1.89
August	1.44 (11%)	1.62

Note: The values in parentheses represent percent increase of RMSE between two cases.

CONCLUSION AND FUTURE WORK

This research evaluated the quality of the KMA-SNU forecast model and the systematic biases in the model's precipitation and temperature forecasts. Five distinct procedures were used to evaluate the forecasts, including (1) transformation, (2) statistical correction, (3) observation of ensemble average, (4) improvement of forecast, and (5) forecast skill test. The results show spatial adjustment, statistical correction, and temporal disaggregating techniques are needed to improve the forecast and contribute to modest improvement in the forecast skill. For precipitation, research results show a gain in forecast skill when a conditional forecast is performed. Recent personal communications with KMA-SNU model developers and forecasters indicate that these results are consistent with their experience, and that, unfortunately, the forecasts being generated are not sufficiently accurate to be used to support management decisions. Although precipitation forecasts for the month of July showed some improvement, the analysis also showed low skill (HSS and KSS) in the critical months (June through August). This could be related to the unstable regional climate and unexpected tropical cyclones (typhoons) created in the western Pacific Ocean.

Additionally, the results show that if the input forecast has low skills, an adjustment will rarely produce a credible hydrologic forecast for the targeted region.

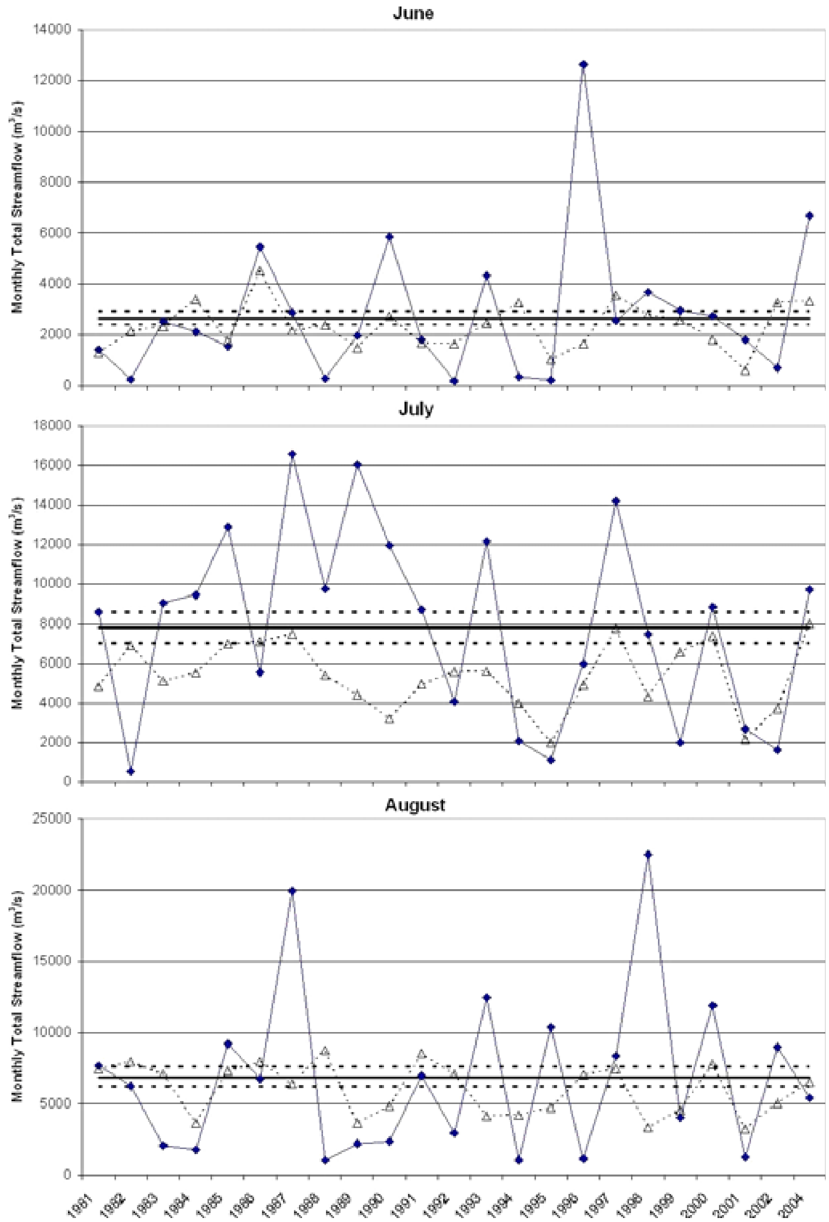

FIGURE 11. Accuracy of the Six-Month Streamflow Hindcast (1981-2000) and Forecast (2004) for Geum River Basin, January through June. Shown are monthly average streamflow (cms) between modeled and forecasted. Modeled streamflows are expressed with a solid line (diamonds) and the forecasted streamflows are expressed with a dotted line (triangles). Note that the solid and dotted constant lines are the monthly mean and ±0.5 SD, respectively.

A potential explanation for this could be that the study basin is too small compared with the spatial resolution of the seasonal forecast model. Therefore, working with basins of many different sizes, where KMA-SNU contains at least a dozen or more grid boxes, would help to determine if there is skill at the largest scale and how skill deteriorates as basins get smaller.

Many climate models have been established for more than a decade and are still evolving to enhance their accuracy and applicability. A point of this study, however, is to focus not only on skillful streamflow forecasts but also on feasible approaches to measuring the quality of streamflow forecasts associated with climate forecasts. Little work has been done to determine the most feasible approach to evaluating the quality of streamflow forecast using skill scores (e.g., HSS and KSS).

Finally, the authors anticipate that the proposed methodology will provide useful insights and directions for stakeholders to improve the skill of hydrologic forecasts when such climate models become available and operational with higher forecast skills.

TABLE 3. Categorical Geum River Streamflow Forecast for the Threshold Boundary.

	X_1	X_2	X_3	X_4	$P(Y_i)$
Y_1	1*	0*	0	0	0.015
Y_2	1*	4*	4*	6	0.227
Y_3	9	5*	5*	9*	0.424
Y_4	4	5	3*	10*	0.333
$P(X_i)$	0.227	0.212	0.182	0.379	n = 66

*Hits satisfying threshold boundary condition. Conditional on occurrence of some form of streamflow for the Geum River basin during monsoon season of June, July, and August 1981 through 2002. The verification data are presented as a 4-by-4 contingency table. The sample size n is 66. Here, for example, the marginal probability $P(Y^2) = (1 + 4 + 4 + 6)/66 = 0.227$.

Also, we anticipate that the proposed methodology could provide useful insights into improving streamflow forecast assessment via skill tests. This will improve forecasting accuracy for large watersheds (or other watersheds where an RCM is well established) by focusing on region-specific initial and boundary conditions that help minimize systematic errors inherited from significantly different spatial resolutions.

APPENDIX

Box-Cox Transformations

The Box-Cox power transformations (Box and Cox, 1964) are given by:

$$x(\lambda) = \frac{(x^{\lambda} - 1)}{\lambda}\ \lambda \neq 0$$
$$x(\lambda) = \ln(x)\ \lambda = 0 \qquad (A1)$$

Given a vector of data observations, $x = x_1, x_2, x_3, \ldots, x_n$, one way to select the power λ, which is an unknown parameter that allows for changing the shape of the data distribution by the exponentiation and effectively lowering or raising the data value, is to use the λ that maximizes the logarithm of the likelihood function.

$$f(x,\lambda) = -\frac{n}{2}\ln\left(\sum_{i=1}^{n}\frac{(x_i(\lambda) - x\_(\lambda))^2}{n}\right) + (\lambda - 1)\sum_{i=1}^{n}\ln(x_i), \qquad (A2)$$

where $\bar{x}(\lambda) = \frac{1}{n}\sum_{i=1}^{n} x_i(\lambda)$ is the arithmetic mean of the transformed data.

Bivariate Normal Distribution

The distributions of predicted X, given a particular forecast value of Y and of the conditional normal density function $f(X \mid Y = y)$, have parameters:

$$\mu_{X|Y} = \mu_X + \rho\sigma_X\frac{y - \mu_Y}{\sigma_X}$$
and
$$\sigma_{X|Y} = \sigma_X\sqrt{1 - \rho^2}, \qquad (A3)$$

where ρ is the correlation coefficient.

Basic Skill Score

A basic skill score equation given by:

$$SS_{\text{ref}} = \frac{A - A_{\text{ref}}}{A_{\text{perf}} - A_{\text{ref}}} \times 100\%, \qquad (A4)$$

where A_{perf} is the value given by a perfect forecast, and A_{ref} is the value given by the reference forecast drawn from random space. If the forecast being evaluated (A) is equal to the perfect forecast, A_{perf}, the skill score will be 100%. Similarly, if the forecast being evaluated is essentially random, the skill score will be 0%.

Heidke Skill Score

The general form of HSS is given by

$$\text{HSS} = \frac{\sum_{i=1}^{I} P(y_i, x_i) - \sum_{i=1}^{I} P(y_i)P(x_i)}{1 - \sum_{i=1}^{I} P(y_i)P(x_i)}, \qquad (A5)$$

where $P(y_i, x_i)$ is the joint distribution of forecasts and observation and $P(y_i)$ and $P(x_i)$ are the marginal distributions of the forecast and observation, respectively. The first term in the numerator, $\sum_{i=1}^{I} P(y_i, x_i)$, is the "hit rate" (HR), the proportion correct, or probability of detection. The second term in numerator $\sum_{i=1}^{I} P(y_i)P(x_i)$ is the HRF. This is the expected "yes" events that were correctly forecasted from random space.

Kuipers Skill Score

The KSS is defined as

$$KSS = \frac{\sum_{i=1}^{I} P(y_i, x_i) - \sum_{i=1}^{I} P(y_i)P(x_i)}{1 - \sum_{i=1}^{I} P(x_i)^2} \qquad (A6)$$

The only difference between HSS and KSS is that the KSS formulation is based on an unbiased random sampling by assuming that the marginal distribution of both observed and forecast values are treated equally in the random forecast domain (second term in the denominator in the equation above).

ACKNOWLEDGMENTS

The authors thank Dr. In-Sik Kang for providing valuable comments and data, and three anonymous reviewers for insightful comments and suggestions that helped improve the quality of the manuscript substantially.

LITERATURE CITED

Box, G.E.P. and D.R. Cox, 1964. An Analysis of Transformations. Journal of the Royal Statistical Society 26(2):211-252.

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(1):15-32.

Enke, W. and A. Spekat, 1997. Downscaling Climate Model Ouputs Into Local and Regional Weather Elements by Classification and Regression. Climate Research 8:195-207.

EPA BASINS, 2001. Better Assessment Science Integrating Point and Nonpoint Sources. BASINS V. 3.0 User's Manual, U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

EPA HSPF, 1970. Hydrologic Simulation Program-Fortran. HSPF's User Manual, U.S. Environmental Protection Agency, Office of Research and Development, Athens, Georgia.

Gutierrez, J.M., R. Cano, A.S. Cofino, and M.A. Rodriguez, 2004. Clustering Methods for Statistical Downscaling in Short-Range Weather Forecast. Monthly Weather Review 132(9):2169-2183.

Hanssen, A.W. and W.J.A. Kuipers, 1965. On the Relationship Between the Frequency of Rain and Various Meteorological Parameters. Mededeelingen en Verhandelingen 81:2-15.

Heidke, P., 1926. Berechnung des Erfolges und der Gute der Windstarkevorhersagen im Sturmwanungsdient. Geografiska Annaler 8:301-349.

Ho, C.-H. and I.-S. Kang, 1988. The Variability of Precipitation in Korea. Journal of Korean Meteorological Society 24:38-48.

Kang, I.-S., C.-H. Ho, Y.-K. Lim, and K.-M. Lau, 1999. Principal Modes of Climatological Seasonal and Intraseasonal Variations of the Asian Summer Monsoon. Monthly Weather Review 127(3):322-340.

Kang, I.-S., K. Jin, B. Wang, K.-M. Lau, J. Shukla, V. Krishnamurthy, S.D. Schubert, D.E. Wailser, W.F. Stern, A. Kitoh, G.A. Meehl, M. Kanamitsu, V.Y. Galin, V. Satyan, C.-K. Park, and Y. Liu, 2002. Intercomparison of the Climatological Variations of Asian Summer Monsoon Precipitation Simulated by 10 GCMs. Climate Dynamics 19:383-395.

Kang, I.-S., J.-Y. Lee, and C.-K. Park, 2004. Potential Predictability of Summer Mean Precipitation in a Dynamical Seasonal Prediction System With Systematic Error Correction. Journal of Climate 17(4):834-844.

Karl, T.R., 1990. A Method of Relating General Circulation Model Simulated Climate to the Observed Local Climate, Part I: Seasonal Statistics. Journal of Climate 3(10):1053-1079.

Kim, J.K., I.-S. Kang, and C.-H. Ho, 1998. East Asian Summer Monsoon Simulated by the Seoul National University GCM. Proceedings of International Conference on Monsoon and Hydrologic Cycle, 22-25 April 1998, Korean Meteorological Society, Kyongju, Korea.

Leung, L.R., A.F. Hamlet, D.P. Lettenmaier, and A. Kumar, 1999. Simulations of the ENSO Hydroclimate Signals in the Pacific Northwest Columbia River Basin. Bulletin of the American Meteorological Society 80(11):2313-2329.

Liang, X., D.P. Letternmaier, E.F. Wood, and S.J. Burges, 1994. A Simple Hydrologically Based Model of Land Surface Water and Energy Fluxes for General Circulation Models. Journal of Geophysical Research 99:14415-14428.

Lighthill, J., and R.P. Pearce, 1981. Monsoon Dynamics, Cambridge University Press, 735 pp.

Murphy, A.H., 1993. What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting. Weather and Forecasting 8(2):281-293.

Murphy, A.H., B.G. Brown, and Y.-S. Chen, 1989. Diagnostic Verification of Temperature Forecasts. American Meteorological Society 4(4):485-501.

Roads, J.O., S.C. Chen, M. Kanamitsu, and H. Juang, 1999. Surface Water Characteristics in NCEP Global Spectral Model and Reanalysis. Journal of Geophysical Research 104(D16):19307-19327.

Robinson, H. (1976). Monsoon Asia, Macdonald and Evans, Plymouth, Massachusetts.

Shapiro, S.S., 1990. How to Test Normality and Other Distributional Assumptions. American Society for Quality Control, Statistics Division 3:92.

Stedinger, J.R., 1980. Fitting Log Normal Distributions to Hydrologic Data. Water Resources Research 16(3):481-490.

Wilby, R.L. and T.M.L. Wigley, 1997. Downscaling General Circulation Model Output: A Review of Methods and Limitations. Progress in Physical Geography 21:530-548.

Wilks, D.S. (1995). Statistical Methods in the Atmospheric Sciences. Academic Press, Burlington, Massachusetts.

Wood, A.W., E.P. Maurer, A. Kumar, and D.P. Lettenmaier, 2002. Long-Range Experimental Hydrologic Forecasting for the Eastern United States. Journal of Geophysical Research 107(D20):ACL-6.

Zorita, E., and H.V. Storch, 1999. The Analog Method as a Simple Statistical Downscaling Technique: Comparison With More Complicated Methods. Journal of Climate 12:2474-2489.