A STREAMFLOW FORECASTING FRAMEWORK USING MULTIPLE CLIMATE AND HYDROLOGICAL MODELS

Paul J. Block, Francisco Assis Souza Filho, Liqiang Sun, and Hyun-Han Kwon

ABSTRACT: Water resources planning and management efficacy is subject to capturing inherent uncertainties stemming from climatic and hydrological inputs and models. Streamflow forecasts, critical in reservoir operation and water allocation decision making, fundamentally contain uncertainties arising from assumed initial conditions, model structure, and modeled processes. Accounting for these propagating uncertainties remains a formidable challenge. Recent enhancements in climate forecasting skill and hydrological modeling serve as an impetus for further pursuing models and model combinations capable of delivering improved streamflow forecasts. However, little consideration has been given to methodologies that include coupling both multiple climate and multiple hydrological models, increasing the pool of streamflow forecast ensemble members and accounting for cumulative sources of uncertainty. The framework presented here proposes integration and offline coupling of global climate models (GCMs), multiple regional climate models, and numerous water balance models to improve streamflow forecasting through generation of ensemble forecasts. For demonstration purposes, the framework is imposed on the Jaguaribe basin in northeastern Brazil for a hindcast of 1974-1996 monthly streamflow. The ECHAM 4.5 and the NCEP/MRF9 GCMs and regional models, including dynamical and statistical models, are integrated with the ABCD and Soil Moisture Accounting Procedure water balance models. Precipitation hindcasts from the GCMs are downscaled via the regional models and fed into the water balance models, producing streamflow hindcasts. Multi-model ensemble combination techniques include pooling, linear regression weighting, and a kernel density estimator to evaluate streamflow hindcasts; the latter technique exhibits superior skill compared with any single coupled model ensemble hindcast.

(KEY TERMS: surface water hydrology; precipitation; computational methods; streamflow; forecast; uncertainty; multi-model; ensemble; Brazil.)


INTRODUCTION

Water resources planning and management efficacy is subject to capturing inherent uncertainties stemming from climatic and hydrological inputs and models. Streamflow forecasts, critical in reservoir operation and water allocation decision making, fundamentally contain uncertainties arising from assumed initial conditions, model structure, and...
modeled processes (Georgakakos et al., 2004; Doblas-Reyes et al., 2005). Accounting for these propagating uncertainties remains a formidable challenge.

Approaches to streamflow forecasting predominantly fall into two categories: statistical or dynamical (climatic-hydrological model integration). The former frequently utilizes predictors of sea-surface temperature or a related index to directly estimate streamflow through statistical techniques (Souza Filho and Lall, 2003). The second approach seeks to couple climate and hydrological processes by passing downscaled information in an iterative (online) or static (offline) fashion. Although the mechanics differ, accounting for forecast uncertainty is achievable in either approach by applying Monte Carlo or stochastic methodologies to generate forecast ensembles. Recent enhancements in climate forecasting skill and hydrological modeling serve as an impetus for further pursuing models capable of delivering improved streamflow forecasts (Cane et al., 1986; Barnston et al., 1999a,b; Mason et al., 1999; Landman et al., 2001; Rajagopalan et al., 2002).

Significant attention has been given lately to incorporation of multiple climate models or multiple hydrological models for forecasting climatic or hydrological variables (Krishnamurti et al., 1999, 2000; Doblas-Reyes et al., 2000; Hagedorn, 2001; Kumar et al., 2001; Robertson et al., 2004; Hagedorn et al., 2005; Tebaldi et al., 2005; Block and Rajagopalan, 2008; Chowdhury and Sharma, 2008; Devineni et al., 2008; Mujumdar and Ghosh, 2008). Previous research supports the notion that combinations of model forecasts may produce more robust forecast skill and reliability than single model forecasts, attributable to inclusion of varying initial conditions and processes (Beven and Freer, 2001; Georgakakos et al., 2004; Doblas-Reyes et al., 2005; Regonda et al., 2006). Numerous multi-model ensemble combination techniques have been developed, demonstrating improved streamflow forecast skill (Rajagopalan et al., 2002; Doblas-Reyes et al., 2005; Yun et al., 2005; Ajami et al., 2006; Duan et al., 2006). Little consideration, however, has been given to methodologies that include dynamical coupling of both multiple climate and multiple hydrological models, further increasing the pool of streamflow forecast ensemble members and accounting for cumulative sources of uncertainty. This approach, presented here, creates a robust, inclusive system, demonstrating various options available for improved streamflow forecasting.

In this paper, a general streamflow forecasting framework incorporating multiple climate and hydrologic models from dynamical and statistical approaches is proposed. The Problem Setting, including descriptions of existing climate, hydrologic and coupled model systems, and processes is initially outlined. Following is a Description of the Application Site Chosen and Associated Datasets employed for demonstration. The proposed forecasting framework is subsequently outlined in detail, followed by Results of imposing this framework through a hindcast on the application site. The paper concludes with a Summary and Discussion.

PROBLEM SETTING

Perfect forecast models in climate and hydrology are nonexistent. For improved skill, risk assessment, and probabilistic interpretation, forecasts must be placed in a context based on inherent and cumulative uncertainty created throughout the modeling process. Climate models suffer from assumptions of initial and boundary conditions; dynamical hydrological (i.e., rainfall-runoff) models lack in process description and resolution, parameter estimation, and model structure (Stern and Miyakoda, 1995; Goddard et al., 2001; Georgakakos et al., 2004; Doblas-Reyes et al., 2005; Kang and Yoo, 2006).

Global climate models (GCMs) are based on the general principles of fluid dynamics and thermodynamics. Other processes, such as convection, occurring on scales too small to be resolved directly, require parameterization. GCMs are typically run at relatively coarse spatial resolutions, generally greater than 2.0° latitudinally and longitudinally. The direct result of the poor spatial resolution produces a serious spatial scale mismatch between the available climate forecasts and the scale of interest to most climate forecast users. To overcome this, statistical or dynamical regional climate models (RCMs), with a higher spatial resolution, are constructed for limited areas. Relatively high-resolution RCMs, driven by low resolution GCMs, can provide meaningful small-scale features over a limited region at affordable computational cost compared with high-resolution GCMs.

Dynamical hydrological rainfall-runoff models are generally classified in two broad categories: lumped and distributed. The two differ primarily in spatial variation. Lumped models consider the integrated response of the basin as a whole, spatially averaging climatic and physical variables over the entire area. Distributed models, however, resolve these variables to a finer scale, disaggregating the basin into grid boxes, accounting for some basin heterogeneity (Russo et al., 1994; Vieux, 2001; Smith et al., 2004). Additionally, distributed models typically perform
some routing of streamflow between grid boxes (Ajami et al., 2006). Although distributed models are considered more complex than lumped models, they have not definitively shown their superiority in terms of streamflow forecast skill (Reed et al., 2004) and continue to suffer from inherent problems (Beven, 2001).

Coupling climate and dynamical hydrologic models to function in combination may occur through either an online or offline process. Online processes allow feedback between models in an iterative approach, and typically require the hydrological models to be distributed such that the grid box size is equivalent to the climate model grid box size. Offline coupled models, however, operate in a consecutive manner; the climate model completes its simulation before passing climatic values to the hydrological model. Hydrological models may therefore be of either a lumped or distributed nature. The framework outlined in this work adopts an offline approach.

As previously stated, ensemble forecasting is becoming increasingly popular (Tracton and Kalnay, 1993; Harrison et al., 1995; Doblas-Reyes et al., 2000, 2005; Fritsch et al., 2000; Palmer et al., 2000, 2004). Combining forecasts (from single or multiple models) initialized with differing conditions or assumptions often better predicts the forecast probability density function and has been shown to generally provide superior overall skill (Doblas-Reyes et al., 2005). Additionally, inclusion of forecasts from multiple models may better account for uncertainties in physical processes and model structure (Georgakakos et al., 2004). The driving motivation for this research is to create a framework in which the advantages of including multiple models may be exploited at several stages of the process.

Various techniques exist for combining forecasts from multiple predictions, however applications in a hydrologic context have only been applied in the relatively recent past (McLeod et al., 1987; Shamseldin and O’Connor, 1999; Ajami et al., 2006). The simplest approach is to equally weight all forecasts and pool them together (Barnston et al., 2003; Robertson et al., 2004; Hagedorn et al., 2005). Optimally weighted models based on historical performance advances this methodology. More sophisticated combination techniques include regression and Bayesian approaches (Raftery et al., 1997; Rajagopalan et al., 2002; Doblas-Reyes et al., 2005; Marshall et al., 2005; Ajami et al., 2006; DelSole, 2006). Ensemble means do not necessarily superior forecasts in comparison with single model predictions; however, the clear advantage of ensemble forecasting is apparent in a probabilistic framework (Palmer et al., 2004; Doblas-Reyes et al., 2005).

DESCRIPTION OF APPLICATION SITE AND DATA

**Iguatu Basin in the Jaguaribe River Basin, Brazil**

The streamflow forecasting framework proposed, and outlined in the subsequent section, is applied to the Iguatu basin (19,100 km²), which lies within the larger Jaguaribe basin, a 72,000 km² semi-arid area in northeast Brazil. The city of Iguatu lies at the outlet of this basin on the Jaguaribe River. The basin experiences one rainy season annually, from January to May. During this time, the Atlantic Intertropical Convergence Zone (ITCZ) reaches its southernmost position, lying very near to or over the region, enhancing atmospheric instability and producing precipitous systems. Abnormal latitudinal migrations of the ITCZ are associated with excess (southward) or deficit (northward) rainfall (Hastenrath and Heller, 1977). Previous investigations have firmly established that sea-surface temperature (SST) anomaly forcing is the primary factor responsible for the interannual variability of rainfall in northeast Brazil (Moura and Shukla, 1981; Ward and Folland, 1991; Sun et al., 2005). Positive (negative) rainfall anomalies are frequently observed when the Atlantic SSTs are colder (warmer) than normal north of the Equator and warmer (colder) than normal south of the Equator. Droughts also tend to coincide with the El Nin˜o-Southern Oscillation ENSO episodes. Slowly evolving SST anomalies, particularly in the tropical oceans, can be predicted with some degree of skill at lead times of several months (Zebiak and Cane, 1987). Seasonal rainfall forecasts (February to May) issued in January are skillful over northeast Brazil (Sun et al., 2005). Since 2001, an operational regional climate forecast has been maintained by the climate and water foundation for the State of Ceará (FUNCEME) and the International Research Institute for Climate and Society (IRI). Offline coupling of an RCM and the Soil Moisture Accounting Procedure (SMAP) hydrologic model was developed by Souza Filho and Porto (2003).

**Application Site Data**

Observed precipitation and streamflow time-series exist for the 1912-1996 period, although the monthly streamflow record at Iguatu is incomplete. Figure 1 illustrates January to June streamflow at Iguatu (essentially annual) for years with no missing months. Average daily precipitation over the basin is calculated using the Thiessen polygon method. Monthly potential evaporation is based on climatological values (basin data acquired from the Planning Study of the Jaguaribe River Basin) (COGERH, 1998).
ECHAM4.5 and NCEP/MRF9 GCM precipitation data is obtained from the IRI Data Library and based on observed SSTs and 10 ensemble members each (Kumar et al., 1996; Livezez et al., 1996; Roeckner et al., 1996; Saha et al., 2006).

FORECASTING APPROACH AND MODEL DESCRIPTIONS

This approach proposes the integration of multiple GCMs, RCMs, hydrologic models, and multi-model combination techniques in successive fashion for ensemble streamflow forecasting. The overall framework proposed is presented pictorially in Figure 2. Generally, persisted or forecasted sea-surface temperatures drive GCMs, producing low resolution precipitation that may be downscaled with statistical or dynamical RCMs. Dynamical approaches often require bias correction based on hindcasts and historical observations. If desired, downscaled precipitation (and other climatic variables) may be run through a weather generator to produce an ensemble of plausible scenarios, further increasing the forecast pool. Downscaled precipitation is fed into hydrological models to generate streamflow forecasts, which are subsequently weighted and combined, as a final step, to create a multi-model ensemble for probabilistic evaluation. To demonstrate the framework on the Iguatu basin, a streamflow hindcast is performed over the 1974-1996 period. The following sections describe specific models chosen for analysis, and by no means constitute the full array of possibilities available in the proposed framework.

Climate Models

Global Climate Models. The atmospheric GCMs chosen for framework demonstration are the ECHAM4.5, as developed at the Max Plank Institute for Meteorology in Germany (Roeckner et al., 1996), and the NCEP/MRF9, developed by
the U.S. National Weather Service’s National Centers for Environmental Prediction (Kumar et al., 1996; Livezey et al., 1996; Saha et al., 2006). For this hindcast demonstration, each model is run in simulation mode, driven by concurrent SSTs. Alternatively, a retrospective format could be evaluated, such that forecasted or persisted SSTs replace observed SSTs.

Regional Climate Model. Dynamical:Regional Spectral Model. The NCEP RSM (Juang and Kanamitsu, 1994; Juang et al., 1997) is selected for dynamical downscaling. An ensemble of 10 runs with the nested NCEP RSM – ECHAM4.5 AGCM system, using observed SSTs, is created for the period of January to June 1971-2000. For further model details, including coupling with the ECHAM4.5 GCM, please contact the authors. Dynamical downscaling models utilizing the NCEP/MRF9 GCM are not yet available; additional RCMs over northeast Brazil are currently under development.

Statistical:Linear Regression With Principal Components. Statistical downscaling of climatic variables from the GCMs offers an alternative approach to dynamical downscaling, and is traditionally significantly simpler in nature (Wilby et al., 1998; Murphy, 1999; Landman et al., 2001). The methodology adopted for downscaling of precipitation utilizes a cross-validated linear regression model with principal components (PCs) of the GCM forecasted precipitation ensemble mean acting as predictors. The spatial domain included is identical to that employed by the Regional Spectral Model (RSM) Principal component analysis, widely used in climate research, decomposes a space-time random field into orthogonal space [Eigenvector (E)] and time (PC) patterns using Eigen decomposition (Von Storch and Zwiers, 1999). The patterns are ordered according to the percentage of variance captured. This analysis technique additionally eliminates multicollinearities and limits unstable or unreasonable estimates of weights. The precipitation downscaling methodology, outlined in the following algorithm, is repeated for each month (six) from each GCM for January to June, 1971-1996:

1. Model PCs are based on the 1950-1996 GCM monthly precipitation ensemble mean, $X$, and constructed with the following relationship:

$$PC^T = E^T X^T$$  \hspace{1cm} (1)

The first $n$ PCs, explaining the majority of the variance, are retained and form the suite of model predictors.

2. Regression coefficients, $\beta_0, \ldots, \beta_n$, based on the ensemble mean, are identified through a least squares linear regression model. The following equation creates a best fit estimate by optimally weighting model PCs to minimize errors.

$$OBS_t = \beta_0 + \beta_1 PC_{1,t} + \beta_2 PC_{2,t} + \ldots + \beta_n PC_{n,t} + \varepsilon_t \hspace{1cm} t = 1 \text{to} t^*, \hspace{1cm} (2)$$

where OBS is the observed precipitation value at time $t$, $t^*$ is the number of time periods (years), and $\varepsilon$ corresponds to model residuals.

3. PCs for the prediction month, PCP, (individual ensemble member) are acquired using the eigenvectors developed in step (1) and applying Equation (1) with $X$ representing the GCM prediction month hindcast.

4. Retaining the regression coefficients from Equation (2), the prediction month PCs are applied to the following equation to determine the monthly precipitation hindcast, $H$.

$$H_t = \beta_0 + \beta_1 PC_{1,t} + \beta_2 PC_{2,t} + \ldots + \beta_n PC_{n,t} \hspace{1cm} t = 1 \text{to} t' \hspace{1cm} (3)$$

Equation (3) is utilized for each of the 10 ensemble members corresponding to the specified month and GCM.

Bias Correction: Probability Mapping. Dynamical or physically based model outputs typically contain some systematic bias and require correction, unlike statistical models which inherently account for biases by their empirical nature. Probability mapping is selected here for bias correction of the monthly RSM precipitation data (averaged over the 12 RSM grid boxes), and is based on two cumulative distribution functions (CDF): (1) the historical observed data, and (2) all RSM ensemble data pooled by months (akin to Ines and Hansen, 2006). The latter CDF is created by combining all hindcasts for a given month using all ensemble members. Each CDF is fit with a gamma distribution, saving the shape and scale parameters. A given monthly RSM precipitation value from a hindcast ensemble member is then bias corrected by mapping it from the corresponding month’s pooled RSM CDF to the cross-validated observed CDF, as demonstrated in Figure 3.
Hydrological Models

A plethora of lumped and distributed hydrological models have been developed over the past decades, providing a wide array of choices. Two rainfall-runoff models are selected for inclusion in this work: the ABCD model, well recognized and accepted in the hydrology community (Thomas, 1981; Thomas et al., 1983) and the SMAP applied to various basins within Brazil (Lopes et al., 1982; Lopes and Porto, 1993). The models are both run at monthly time-steps, effectively relegating them to water balance models. Models requiring higher temporal resolution may alternatively be selected, dependent on forecasting needs.

SMAP is a conceptual, lumped model containing two reservoirs (subsurface and ground water) and four parameters: soil saturation capacity, surface flow, a recharge coefficient, and a base flow recession coefficient. The rainfall-runoff component is founded on the Soil Conservation Service equation and utilizes basin average precipitation and evapotranspiration.

ABCD is a nonlinear watershed model, which represents soil moisture storage, ground water storage, direct runoff, ground water outflow to the stream channel, and actual evapotranspiration. Inputs include precipitation and potential evapotranspiration. Its performance in comparison with other monthly water balance models has lead to its recommended use (Alley, 1984, 1985; Vandewiele et al., 1992).

Multi-Model Hindcast Ensemble Combinations and Skill Scores

Three techniques for combining hindcast ensembles from multiple models are selected to demonstrate various levels of sophistication. Monthly streamflow hindcasts (January to June) from the hydrological models are aggregated to seasonal totals prior to combination. Aggregating temporally tends to smooth the data and increase skill by reducing the noisy month-to-month variability. Seasonal totals are also of particular interest for the application site in regards to agriculture and water resources infrastructure planning.

The most straightforward technique for combining ensembles is pooling, in which all ensemble members are given equal weight and joined into a single multimodel ensemble (Barnston et al., 2003; Robertson et al., 2004; Hagedorn et al., 2005). Basic statistics (median, standard deviation, etc.) are easily computed. It has been shown that even this simple approach typically proves superior to a single forecast due to its higher reliability (Doblas-Reyes et al., 2005; Hagedorn et al., 2005).

The second methodology utilizes a least squares linear regression technique, of which a form has been used for some time by NCEP for seasonal climate forecasts (Van den Dool, March 27, 2008, IRI Seminar Series). The particular version adopted here assigns a weight to model ensembles based on the regression coefficient created by fitting individual ensemble means to observed conditions. Regression coefficients closer to a value of one (similar to observed conditions) are assigned greater proportional weight than coefficients deviating further from one.

The third approach applies a normal kernel density estimator (Bowman and Azzalini, 1997; Bishop, 2006) to calculate the probability density, $P_{mt}$, at each hindcast observation from each model using the following equation:

$$P_{mt} = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{\sqrt{2\pi h^2}} \exp \left( -\frac{(x_{0,t} - x_{m,tn})^2}{2h^2} \right),$$

where $N$ is the total number of ensemble members, $h$ is the kernel bandwidth, $x_{m,tn}$ is the hindcast value from model $m$, in period $t$, and ensemble member $n$, and $x_{0,t}$ is the observation in period $t$ (where the probability density is desired). Bandwidth is set as a function of ensemble variance. Optimal weight for each model ($w_m$), constrained between zero and one, is obtained by maximizing the total likelihood ($L$) over all time periods from all models (Rajagopalan et al., 2002). For “$z$” models:

$$L = \prod_{t=1}^{T} \left( \sum_{m=1}^{z} w_m P_{m,t} \right)$$

FIGURE 3. Probability Mapping Demonstration for Bias Correction.
Skill scores for ensemble hindcast evaluation selected here include median and total root mean square error, Pearson’s correlation coefficients, and rank probability skill score (RPSS) (Wilks, 1995; Sun et al., 2006; Block and Rajagopalan, 2008). Pearson’s correlation coefficients are calculated using ensemble medians. Relevant equations and descriptions are provided in Appendix I.

RESULTS AND ANALYSIS

Following the framework proposed and models selected, streamflow hindcasts for 1974-1996 at Iguaçu are generated and compared with the observed record. Results at each major stage of the framework are presented.

Climate Model Downscaling

The ability of the nested ECHAM4.5 GCM-RSM model to simulate interannual variability of precipitation is illustrated in Figure 4. Temporal anomaly correlation coefficients between observed and simulated precipitation ensemble means during the rainy season (Figure 4A) exceed 0.3 (the 90% confidence level) over most of northeast Brazil, with a correlation coefficient of 0.72 for the Jaguaribe basin. The strong performance by the model suggests that much of the interannual variability of basin rainfall is associated with global SST variations.

Probabilistic rainfall hindcasts by the nested model provide an estimate of the predictability that may be achievable given a perfect SST forecast. The RPSS for the RSM precipitation ensemble mean averaged over the 30-year hindcast is shown in Figure 4B, and indicates moderate skill (<10%) south of latitude 9°S, but stronger skill (>10%) in northeast Brazil, particularly over the basin.

Figure 5 illustrates the raw RSM and observed CDFs required for bias correction of monthly precipitation, and clearly indicates a larger bias in the drier January, May, and June months.

The bias correction methodology adopted here is fairly effective for the Jaguaribe basin, yet some years still indicate insufficient or excess precipitation in comparison with the observed record. Figure 6 (top) depicts box plots of the January to June total precipitation for each corrected ensemble member and the observed record for 1971-1996. While for most years the ensemble members envelop the observed value, as desired, a few years indicate under- or over-predictions by all ensemble members (e.g., 1974 and 1976) even after bias correction, partially attributable to scaling with ensemble means. However, this is expected with ensemble forecasting, and the general wet or dry bias of the RSM model appears to be effectively removed. The correlation coefficient between the dynamically downscaled ensemble mean and observed precipitation over the entire season is 0.82 for the validation period (see Table 1); the median RPSS for the ensemble years equals 0.35, indicating a notable improvement over climatology.
FIGURE 5. Cumulative Distribution Functions for 1971-1996 Monthly Precipitation. Solid line represents Iguatu basin observed; dashed lines represent 10 uncorrected RSM ensemble members. Note different scales.

Statistical downscaling of precipitation from the ECHAM4.5 and NCEP/MRF9 GCMs is illustrated similarly in Figure 6 (middle and bottom). Ten PCs are retained for use as predictors in both models, chosen according to scree plot interpretations. These 10 explain approximately 79 and 65% of the variance, respectively (percentages are averages over the 26-year period, or equivalently 26 unique models, as required by the cross-validation methodology). Associated regression coefficients (excluding $b_0$ – the $y$-intercept) vary between $-2$ to $+2$. General performance is comparable with dynamical downscaling, with high correlation coefficients (0.88 and 0.85), and robust median RPSS (0.27 and 0.55) for the two models, respectively. The nature of the statistical model allows for the possibility of negative precipitation hindcasts, however the rate of occurrence is very small ($\sim1\%$); negative values are reset to zero.

The overall high skill scores between the dynamical and statistical downscaling models are not surprising given ensemble forecasts and temporal aggregation to seasonal values. More interestingly, perhaps, is the performance of the two statistical downscaling methods, being on par or bettering the complex and computationally demanding dynamical method. Retaining both methods, however, remains prudent for capturing structural uncertainty.

**Hydrological Model Calibration and Validation**

The SMAP and ABCD water balance models are calibrated against 1912-1969 monthly streamflow at Iguatu, as indicated in Table 1, by optimizing model parameters. The objective function for optimization in both cases centers on minimizing the sum of the squared errors. Required evapotranspiration values for the calibration period are based on observations. Parameter values for SMAP fell within a reasonable and expected range, with soil saturation capacity, surface flow, the recharge coefficient, and base flow coefficient equating to 624, 2.7, 0.01, and 0, respectively. ABCD model parameters over the calibration period are 0.98, 284, 0.49, and 0, respectively. Parameter value $b$ is high, given its physical interpretation as the upper limit of the sum of evapotranspiration and soil moisture storage, but this phenomenon has been noted by others as well (Thomas et al., 1983; Alley, 1984).

The seasonal time-series (aggregation of January to June) of observed and model produced streamflow during the calibration period are illustrated in Figure 7. Correlation coefficients for both models with observed values are nearly identical at 0.90 and 0.89 for SMAP and ABCD, respectively. The models mimic observed streamflow quite well, with the exception of underestimation in 1924, the wettest year in the calibration period. Monthly time-series (not shown here) also indicate strong correlation, with coefficients of approximately 0.87.

Validation of both models is performed over the 1974-1996 period. Evapotranspiration inputs to the model are based on climatological values of the period (to reflect similarity with the coupled model hindcast case). Seasonal correlation coefficients for the SMAP and ABCD models, respectively, are 0.95 and 0.97, indicating very strong model performance according to this metric. In slight contrast to the calibration period performance, both models do remarkably well in estimating high flow seasonal quantities, but are less proficient for low flow seasons.

**Coupled Climate and Hydrologic Model Hindcast**

Coupling the climate and hydrological models, a streamflow hindcast is performed on the 1974-1996 period. Six ensembles of hindcasts (10 members each) are produced by combining the three downscaling techniques with the two hydrological models. Figure 8 portrays results of driving the calibrated SMAP and

<table>
<thead>
<tr>
<th>Model</th>
<th>Calibration Period</th>
<th>Correlation Coefficient</th>
<th>Validation Period</th>
<th>Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downscaling</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dynamical – bias correction</td>
<td>1971-1996</td>
<td>0.82</td>
<td>1971-1996</td>
<td>0.82</td>
</tr>
<tr>
<td>Statistical – ECHAM</td>
<td>1971-1996</td>
<td>0.88</td>
<td>1971-1996</td>
<td>0.88</td>
</tr>
<tr>
<td>Statistical – NCEP</td>
<td>1971-1996</td>
<td>0.85</td>
<td>1971-1996</td>
<td>0.85</td>
</tr>
<tr>
<td>Hydrological models</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABCD</td>
<td>1912-1969</td>
<td>0.89</td>
<td>1974-1996</td>
<td>0.97</td>
</tr>
<tr>
<td>SMAP</td>
<td>1912-1969</td>
<td>0.90</td>
<td>1974-1996</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Notes: Pearson correlation coefficients use ensemble medians; concurrent calibration and validation periods imply a leave-one-out cross-validated methodology.
ABCD models with the ECHAM4.5 GCM precipitation ensemble dynamically downscaled and bias corrected. Figures 9 and 10 are similar excepting that statistical downscaling is employed for the ECHAM4.5 and NCEP/MRF9 GCMs. Streamflow ensembles are displayed as box plots; observed values are indicated by filled circles. Seasonal time-series of ensemble medians and observed streamflow are illustrated by solid and dashed lines, respectively. Notable features include better approximation of 1974 (wettest year in hindcast period) by the ECHAM4.5-statistical downscaling technique in both hydrological models (Figure 9), less variance in the NCEP/MRF9-statistical downscaling method under both hydrological models (Figure 10), compared with the other two methods, and solid capture of the dry years by all coupled methods. Coupled model skill scores, displayed in Table 2, are reasonably strong, and suggest that the models and precipitation downscaling techniques generally capture the features of the basin. The SMAP coupled models appear slightly superior to the ABCD models, however larger differences are evident between downscaling techniques. The ECHAM4.5 GCM-statistical downscaling-SMAP coupled model demonstrates the highest skill scores for correlation and root mean squared error, but performs inferiorly to climatology for the RPSS categorical skill score. Evaluation of rank histograms (Hamill, 2001) indicates suitable capture of streamflow variance within ensemble hindcasts, however interpretation is limited by the short time-series.

**Multi-Model Ensemble Combinations**

Ensemble combination of the coupled models constitutes the final step in streamflow forecasting laid...
Skill scores for the three combination techniques proposed (pooled, linear regression weighting, and kernel density estimator) are displayed in Table 3, with the kernel density estimator approach emerging as most skillful. Correlation coefficients are superior to single model ensemble results (Table 2), errors are less than or on par with single models, and categorical skill score (RPSS) is
again as good or better. Total root mean squared error for pooling was omitted, as it comprises 60 ensemble members, while the other techniques only 10. Optimal least squares linear regression model weights were quite similar across models, with the ECHAM4.5-statistical downscaling coupled models receiving the highest proportion. The kernel density estimator technique, however, placed most weight on the three SMAP coupled models, with the two statistical downscaling methods being highest (dynamical contributes approximately 28%, the ECHAM statistical 36%, and the NCEP statistical 35%). Very little weight is reserved for the ABCD coupled models under this methodology (approximately 1%), adding minimal additional information for this region. The kernel density estimator slightly outperforms the ECHAM4.5-statistical downscaling-SMAP single model when evaluating across all skill score metrics, generally reflecting an improvement in streamflow estimation for most years. Greater separation between these two (e.g., even stronger performance of the kernel model over the single model) is observed when 1974 (wettest year in hindcast) is omitted, and would also be expected as the time-series length is increased (more than 23 years). As the skill scores are already remarkably high for this region, it is exceptionally difficult to improve substantially over a single model hindcast. More distinct differences between single and multi-model approaches may be expected for regions of moderate to low skill. Figure 11 portrays median seasonal streamflow from the six hydrological model ensembles and the kernel density estimator multi-model combination, and the observations. Clearly the kernel combination does not produce the best estimate for any given year, but outperforms any single hydrological model approach over the hindcast period.

Streamflow forecast ensembles may also be used to construct probability density functions (PDFs), beneficial for risk-based decision making within the basin. As a demonstration, Figure 12 illustrates PDFs of seasonal streamflow from the pooling combination technique and from the observed record (climatology) for 1991, a dry year. Actual 1991 streamflow is included as a dotted vertical line. The PDFs are estimated using a nonparametric kernel density estimator (Bowman and Azzalini, 1997). The pooled ensemble hindcast PDF is clearly shifted left of the climatological PDF, indicative of relatively less streamflow, more closely matching observed conditions. This type of information is useful for risk assessment and improved decision making concerning reservoir operations, water allocation, crop irrigation, etc.

### SUMMARY AND DISCUSSION

Knowledge of forecast uncertainty stemming from climatic and hydrological models allows for improved forecasting and probabilistic evaluation within water resources decision making. Properly accounting for this uncertainty, however, remains a formidable challenge. Little consideration has been given to streamflow modeling methodologies that include coupling of both multiple climate and multiple hydrological models, further increasing the pool of streamflow forecast ensemble members and accounting for cumulative sources of uncertainty. The framework presented here proposes integration and offline coupling of GCMs, multiple RCMs, and numerous hydrologic models to improve streamflow forecasting through generation of ensemble forecasts. For demonstration purposes, the framework is imposed on the Jaguari basin in northeastern Brazil for a hindcast of 1974-1996 monthly streamflow. The ECHAM4.5 and NCEP/MRF9 GCMs are integrated with regional models, including dynamical and statistical models, and water balance models, specifically the SMAP and NCEP.
Precipitation hindcasts from the GCMs are downscaled via the regional models and fed into the hydrological models, producing streamflow estimates. Multi-model ensemble combination techniques include pooling, linear regression weighting, and a kernel density estimator.

The kernel density estimator multi-model ensemble demonstrates superior skill considering numerous metrics in comparison with any single hydrological model approach (formulated from multi-climate models). This improvement would only be expected to increase as the hindcast or forecast is lengthened (more than 23 years). It is anticipated that the coupled ABCD models, currently contributing little information, would also potentially receive greater weights under longer time-series. Their lack of weight for this case may be partially attributable to possible stability issues with kernel estimators at high dimensions, and requires further investigation.

The respectable performance of the simple pooling combination approach is also noteworthy, but not unexpected, given the relatively short hindcast period. The Jaguaribe basin clearly demonstrates high predictive skill, dampening potential improvements of multi-model approaches over single model forecasts. More distinct differences between single and multi-model approaches may be expected for regions of moderate to low skill.

The overall performance of the framework on the Jaguaribe basin appears robust in comparison with similar basin studies. It demonstrates an improvement over the Souza Filho and Porto (2003) ECHAM4.5-RSM-SMAP coupled model in terms of skill and uncertainty accounting. Souza Filho and Lall (2003) construct a semiparametric streamflow forecast model conditioned on an ENSO index and sea-surface temperatures, issuing forecasts for the January to May rainy season the prior July. The proposed framework produces similar skill, but differs by conducting a hindcast with observed SSTs, issuing concurrent monthly forecasts. Although the framework proposed here is considerably more complex and multifaceted than the
semiparametric model, it does boast sufficient flexibility for improvements at many stages, model substitutions, or other potential lines of analysis.

While the bulk of uncertainty is accounted for in the proposed framework, including initial conditions, model structure, and modeled processes, it is not entirely inclusive. To truly capture end-to-end uncertainty, two additional aspects must be targeted: model parameter and objective function uncertainty. The current framework can adequately capture model process uncertainty, but not explicitly model parameter uncertainty, specifically for hydrological models, although this may be simply addressed by calibrating the models over various time periods and evaluating parameter ranges. Considering objective functions, model optimization in the current framework utilizes minimization of model errors; alternative objective functions – independently or in combination – may better reflect existing uncertainty. These aspects are currently being developed for inclusion in the framework. On-going work utilizes the framework for apportioning uncertainty, comparing variances between single and multi-model approaches at each stage, to determine which processes or stages are contributing most significantly to the overall uncertainty in the system.

Other aspects also warrant attention for future study, specifically multi-model combination techniques. Categorical forecasts (Rajagopalan et al., 2002) or providing prior information through Bayesian weights (Marshall et al., 2005; DelSole, 2006; Duan et al., 2006) provide attractive alternatives, and may assist in improving streamflow forecast skill.

APPENDIX I – VERIFICATION MEASURES

Total root mean square error:

\[
\text{TRMSE}_m = \sum_{n=1}^{N} \sqrt{\frac{\sum_{t=1}^{T} (x_{o,t} - \bar{x}_m)^2}{t^*}} \tag{A1}
\]

Median root mean square error is identical, excepting the summation and using ensemble medians. Pearson’s correlation coefficient:

\[
R_m = \frac{\sum_{t=1}^{T} (\bar{x}_{m,t} - \bar{x}_m)(x_{o,t} - \bar{x}_o)}{(t^* - 1)\sigma_m\sigma_o}, \tag{A2}
\]

where \( m \) is the model number, \( n \) is the ensemble members, \( x_{o,t} \) is the observed streamflow value at time \( t \), \( \bar{x}_{m,t} \) is the hindcasted streamflow value, \( \bar{x}_o \) and \( \bar{x}_m \) represent the observed mean and model ensemble median mean, respectively, and \( \sigma_o \) and \( \sigma_m \) are the observed and model ensemble median standard deviations, respectively.

Rank Probability Skill Score

The rank probability skill score (RPSS) (Wilks, 1995; Sun et al., 2006; Block and Rajagopalan, 2008) is a measure of the skill of ensemble forecasts in comparison with predictions by climatology forecasts. The general rank probability score (RPS) equation for any year takes the form

\[
\text{RPS} = \sum_{m=1}^{R} \left( \text{CP}_{F,m} - \text{CP}_{O,m} \right)^2, \tag{A3}
\]

where \( R \) is the number of categories (three in this study, i.e., above-, near-, and below-normal), and \( \text{CP}_{F,m} \) and \( \text{CP}_{O,m} \) are the cumulative predicted and observed probabilities, respectively, through category \( m \). For three categories of equal size, the climatological probability of being in each is 33%; for the category that is observed the probability is 100% and zero in the other two. Perfect forecasts result in RPS equal to zero. The RPSS is subsequently defined as

\[
\text{RPSS} = 1 - \frac{\text{RPS}_{\text{FORECAST}}}{\text{RPS}_{\text{CLIMATOLOGY}}} \tag{A4}
\]

Rank probability skill score values range from \(-\infty\) to +1. A value of +1 represents perfect skill (i.e., a perfect forecast), while negative values symbolize poor skill; any value greater than zero corresponds to an improved forecast over climatology. The RPSS is calculated for each year.

ACKNOWLEDGMENT

This paper is funded by a grant/cooperative agreement from the National Oceanic and Atmospheric Administration, NA00AR4311004. The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its subagencies. The authors graciously thank Michael Tippett of the IRI and the three anonymous reviewers for their careful review and insightful comments and suggestions, certainly improving the quality of this paper.
LITERATURE CITED


A Streamflow Forecasting Framework Using Multiple Climate and Hydrological Models


