AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE

10.1002/2014WR016811

Key Points:
- Bayesian climate index weighting can postprocess an ensemble forecast
- The method can be applied without hindcast calibration
- Weighting adjusts to the strength of the relationship with the climate index

Supporting Information:
- Supporting information S1

Correspondence to:
A. A. Bradley,
allen-bradley@uiowa.edu

Citation:
Bradley, A. A., M. Habib, and S. S. Schwartz (2015), Climate index weighting of ensemble streamflow forecasts using a simple Bayesian approach, *Water Resour. Res.*, *51*, 7382–7400, doi:10.1002/2014WR016811.

Received 15 DEC 2014
Accepted 18 AUG 2015
Accepted article online 21 AUG 2015
Published online 12 SEP 2015

Climate index weighting of ensemble streamflow forecasts using a simple Bayesian approach

A. Allen Bradley[1], Mohamed Habib[1], and Stuart S. Schwartz[2]

[1]IIHR-Hydroscience & Engineering, The University of Iowa, Iowa City, Iowa, USA, [2]Center for Urban Environmental Research and Education, UMBC, Baltimore, MD, USA

Abstract Climate state can be an important predictor of future hydrologic conditions. In ensemble streamflow forecasting, where historical weather inputs or streamflow observations are used to generate the ensemble, climate index weighting is one way to represent the influence of climate state. Using a climate index, each forecast variable member of the ensemble is selectively weighted to reflect the climate state at the time of the forecast. A new approach to climate index weighting of ensemble forecasts is presented. The method is based on a sampling-resampling approach for Bayesian updating. The original hydrologic ensemble members define a sample drawn from the prior distribution; the relationship between the climate index and the ensemble member forecast variable is used to estimate a likelihood function. Given an observation of the climate index at the time of the forecast, the estimated likelihood function is then used to assign weights to each ensemble member. The weights define the probability of each ensemble member outcome given the observed climate index. The weighted ensemble forecast is then used to estimate the posterior distribution of the forecast variable conditioned on the climate index. The Bayesian climate index weighting approach is easy to apply to hydrologic ensemble forecasts; its parameters do not require calibration with hindcasts, and it adapts to the strength of the relation between climate and the forecast variable, defaulting to equal weighting of ensemble members when no relationship exists. A hydrologic forecasting application illustrates the approach and contrasts it with traditional climate index weighting approaches.

1. Introduction

Atmospheric circulation is affected by well-known global and regional climate oscillations. Seasonal to interdecadal variations of river flows at many locations around the world are related to these large-scale climate patterns [*Redmond and Koch*, 1991; *Kahya and Dracup*, 1993; *Piechota and Dracup*, 1996; *Eltahir*, 1996; *Piechota et al.*, 1997; *Chiew et al.*, 1998; *Chiew and McMahon*, 2002; *Tootle et al.*, 2005; *Tootle and Piechota*, 2006; *Tootle et al.*, 2008; *Aziz et al.*, 2010; *Lu et al.*, 2011, among others]. The state of a climate pattern is often represented by a climate index. One of the earliest examples is the Southern Oscillation Index (SOI), a measure of the El Niño Southern Oscillation (ENSO) pattern [*Walker and Bliss*, 1932; *Troup*, 1965; *Trenberth*, 1984]. Additional climate indexes now exist for ENSO and other major teleconnection patterns, such as the Pacific/North American (PNA) pattern, and the North American Oscillation (NAO). Given the linkages between streamflow and climate state, many approaches have explored the use of climate indexes for streamflow forecasting [*Piechota et al.*, 1998; *Piechota and Dracup*, 1999; *Souza Filho and Lall*, 2003; *Karamouz and Zahraie*, 2004; *Tootle and Piechota*, 2004; *Araghinejad et al.*, 2006; *Regonda et al.*, 2006a; *Block and Rajagopalan*, 2009; *Gobena and Gan*, 2009; *Golembesky et al.*, 2009; *Hay et al.*, 2009; *Kalra and Ahmad*, 2009; *Kennedy et al.*, 2009; *Timilsena et al.*, 2009; *Wei and Watkins*, 2011a, 2011b; *Kalra et al.*, 2013, among others].

For streamflow forecasting, ensemble forecasting techniques are growing in popularity [*Georgakakos and Krzysztofowicz*, 2001; *Schaake et al.*, 2007; *Cloke and Pappenberger*, 2009]. In most cases, an ensemble of streamflow variables is generated by some form of resampling of historical observations. One approach involves resampling of streamflow observations from the historical record [*Souza Filho and Lall*, 2003; *Grantz et al.*, 2005; *Regonda et al.*, 2006b]. Another, the ensemble streamflow prediction (ESP) approach [*Twedt et al.*, 1977; *Day*, 1985], uses a hydrologic forecast model with weather inputs resampled from the historical record. The simulations are initialized to represent the moisture conditions at the time of the forecast, and then the forecast model is run with alternate historical weather inputs to generate an ensemble of simulated streamflow time series.

Various approaches have been used to make hydrologic ensemble forecasts conditioned on climate information. One approach uses downscaled climate model forecasts to produce weather inputs for hydrologic prediction [*Leung et al*., 1999; *Wood et al*., 2002; *Clark and Hay*, 2004; *Thirel et al*., 2008; *Wood et al*., 2005; *Wood and Lettenmaier*, 2006; *Gobena and Gan*, 2010]. Another approach uses climate forecasts to weight ensemble members generated from resampled weather or streamflow observations [*Croley*, 1996, 1997, 2000; *Stedinger and Kim*, 2010]. The weighting adjusts the sample of observations from the climatological record to match forecast climate conditions.

A climate index may also be used to selectively weight ensemble members generated from resampled observations. With climate index weighting, the weight is based on the similarity of the climate index for the ensemble members to the climate index observed at the time the forecast. Examples of climate index weighting for hydrologic ensemble forecasts include *Smith et al*. [1992], *Hamlet and Lettenmaier* [1999], *Werner et al*. [2004], *Grantz et al*. [2005], *Regonda et al*. [2006b], *Wood and Lettenmaier* [2006], *Moradkhani and Meier* [2010], *Wang et al*. [2011], and *Najafi et al*. [2012].

In this paper, we introduce a new approach for climate index weighting of hydrologic ensemble forecasts. The approach is an application of Bayesian updating to the ensemble forecast probability distribution given the climate index at the time of the forecast and is based on the sampling-resampling approach of *Smith and Gelfand* [1992]. In the following sections, we describe the Bayesian climate index weighting approach and present an example to illustrate its use in ensemble forecasting.

2. Problem Statement

The following nonparametric framework for ensemble streamflow forecasting was introduced by *Smith et al*. [1992]. Let Y be a continuous random variable representing a forecast variable. Examples might include the discharge on a specified date, the cumulative discharge volume for some period, or the minimum discharge during the forecast interval. A probability distribution forecast of Y is

$$F(y)=P\{Y \leq y|\xi\}, \tag{1}$$

where $F(y)$ is the conditional distribution of Y given the state vector ξ. The state vector represents the state variables for the watershed at the time the forecast is made, such as initial conditions representing the soil moisture, snowpack, and river conditions. The state vector may also include variables representing the current climate state, such as a climate index.

Let the sample $y_i, i=1, \ldots, N$ represent an ensemble forecast of Y based on N ensemble members. Mathematically, $y_i, i=1, \ldots, N$ is just a sample drawn from the conditional distribution $F(y)$. The sample can be used to estimate the conditional distribution. Specifically, the sample estimated probability distribution forecast is

$$\hat{F}(y)=\sum_{i=1}^{N} w_i \cdot 1[y_i \leq y], \tag{2}$$

where w_i is a nonnegative weight for the ith ensemble member and $1[y_i \leq y]$ is the indicator function:

$$1[y_i \leq y]=\begin{cases} 1 & \text{if } y_i \leq y \\ 0 & \text{if } y_i > y \end{cases}. \tag{3}$$

Note that by definition, the sum of all the weights w_i must equal 1.

Often in forecasting, each member of the ensemble forecast is weighted equally, on the assumption that each realization in the sample is equally likely. In that case, the weights w_i for all ensemble members are

$$w_i=\frac{1}{N}. \tag{4}$$

The influence of the climate state can be represented by unequal weighting of ensemble members. Climate index weighting uses an appropriately chosen climate index available at the time of the forecast to selectively weight ensemble members. Let θ be the climate index at the time of the forecast. Let θ_i be the climate index representative of the ensemble member y_i. Climate index weighting is based on a measure of the similarity of θ_i to θ; ensemble members with similar climate indexes are weighted more than those with dissimilar ones.

One approach for climate index weighting is to assign each ensemble member y_i to a category based on θ_i, then make a block adjustment [*Stedinger and Kim*, 2010] that assigns the same weight to ensemble members in the same category. For example, *Hamlet and Lettenmaier* [1999] defined six climate categories based on two climate indexes; the climate category was determined for the forecast period based on current index values, and ensemble members from that category were given equal weight, whereas ensemble members for other categories were not used (zero weight). *Smith et al*. [1992] and *Croley* [1996, 1997, 2000] used a similar block adjustment approach based on seasonal climate outlook forecasts, weighting ensemble members based on categories defined by properties of their historical weather inputs.

In contrast, kernel climate index weighting assigns weights based on a kernel function $K(x)$ using

$$w_i = \frac{K(|\theta_i - \theta|)}{\sum_{j=1}^{N} K(|\theta_j - \theta|)}. \tag{5}$$

As $|\theta_i - \theta|$ represents the distance between index values, appropriate choices for a kernel function are those that produce larger values for smaller distances, such as an inverse distance or Gaussian function. As an example, *Werner et al*. [2004] used θ_i to find the *k*-nearest neighbor ensemble members to θ, and then assigned nonzero weights to these members using a kernel function. *Najafi et al*. [2012] compared *k*-nearest neighbor, formal and informal likelihood, and fuzzy clustering approaches for kernel climate index weighting with ensemble forecasts.

In the next section, we propose a new approach for climate index weighting. The approach is derived from Bayes' theorem and is equivalent to Bayesian updating of the probability distribution forecast given the current climate index observation.

3. Bayesian Climate Index Weighting

Bayes' theorem defines how prior estimates of probabilities can be updated given new information (e.g., a climate index). Let $F(y)$ be the prior cumulative distribution for a continuous random variable Y, and let $f(y)$ represent the prior density function. Given an observation θ, the updated (or posterior) density is given by Bayes' theorem to be

$$f(y|\theta) = \frac{f_\theta(\theta|y)f(y)}{f_\theta(\theta)}, \tag{6}$$

where $f_\theta(\theta)$ is the unconditional distribution of θ, and $f_\theta(\theta|y)$ is the likelihood function (or conditional distribution). In the context to be used, the prior density $f(y)$ describes a probability forecast. The posterior density $f(y|\theta)$ describes the updated forecast given the observed climate index θ. Analytical solutions to equation (6) are available when the prior density and the likelihood function are normally distributed (Gaussian). *Luo and Wood* [2008] illustrate such an example of Bayesian updating of precipitation and temperature priors using seasonal climate forecasts from multiple models. But in general, direct evaluation of equation (6) for non-Gaussian cases is very challenging.

A far simpler approach is to apply Bayes theorem with a data sample, rather than with density functions. In particular, let $y_i, i=1,\ldots,N$ represent a sample drawn from the prior density $f(y)$. *Smith and Gelfand* [1992] present a simple and elegant resampling approach to update this sample—using the likelihood function $f_\theta(\theta|y)$—to represent a sample drawn from the posterior distribution $f(y|\theta)$. Mathematically, this is done by a weighted resampling of the original sample, where the probability of selecting the discrete sample value y_i is given by a weight w_i. *Smith and Gelfand* [1992] show that this weight is simply

$$w_i = \frac{f_\theta(\theta|y_i)}{\sum_{j=1}^{N} f_\theta(\theta|y_j)}, \tag{7}$$

where by definition $\sum_{i=1}^{N} w_i = 1$. As a resampling method, the approach is a variant of the popular bootstrap resampling method [*Efron and Tibshirani*, 1993].

The proposed climate index weighting approach is an application of this same idea. For an ensemble forecast, the sample of forecast variables based on the ensemble members $y_i, i=1, \ldots, N$, is just a sample drawn from the prior probability distribution forecast $F(y)$ before conditioning on the climate state. The weights w_i defined in equation (7) are simply the Bayesian climate index weights given the climate index θ at the time of the forecast; the weights update the original (equally weighted) ensemble forecast to represent a posterior probability distribution forecast, as shown in equation (2). Conceptually, the weights are proportional to the likelihood of the current climate index observation θ given the ensemble member y_i. More simply, from the resampling perspective, one can think of the weight as the discrete probability of each ensemble member outcome in a sample given the current climate index.

In the next section, we show an example using Bayesian climate index weighting for ensemble forecasting. The example is for the Blue Nile, where river flow volumes are strongly correlated with the El Niño Southern Oscillation (ENSO) [*Eltahir*, 1996]. Ensemble streamflow predictions from a river forecast model are used as a sample of the prior distribution and reweighted using an ENSO index. This example is used to compare the Bayesian method with traditional climate index weighting methods and to illustrate how the likelihood function in equation (7) can be estimated using parametric and nonparametric approaches.

4. Nile Forecast System Example

This example examines ensemble forecasts for the Blue Nile at Diem, just over the Sudan border with Ethiopia. Figure 1 shows the daily flow climatology for 1992–2009. The annual flood in the summer is related to the timing of heavy Kiremt season rains in the Ethiopian highlands [*Seleshi and Demaree*, 1995; *Camberlin*, 1997; *Conway*, 2000; *Segele and Lamb*, 2005; *Block and Rajagopalan*, 2007]. Annual variations in the Kiremt rains are associated with ENSO, which affects the strength of the Indian Ocean monsoon. Ensemble forecasts will be generated for flood season flow volume, defined here as the cumulative flow volume from June through October.

Eltahir [1996] found that annual variability of the Nile flood is strongly related to ENSO. Larger floods are associated with La Niña conditions, whereas smaller annual floods are associated with El Niño. We examined relationships between several ENSO climate indexes and the Blue Nile flood volume. The correlation with May NINO4.0 index (-0.63) was the strongest among those that would be available for a forecast issued in June and will be used for climate index weighting.

4.1. Retrospective Forecasting With Climate Index Weighting

Retrospective forecasts (also known as hindcasts) were made for each flood season from 1992 to 2009, corresponding to the period with reliable flow observations for the Blue Nile at Diem. Ensemble streamflow forecasts were generated by the Nile Forecast System (NFS) [*Elshamy*, 2008]. The forecasting process is illustrated in Figure 2. For each forecast, the model is initialized using model state variables appropriate for basin moisture conditions on the date of the forecast; a flow data assimilation procedure within NFS accomplishes this task [*Nile Forecast Center*, 2007]. Then, the model is run to simulate streamflow time series (or traces) for alternate input weather sequences for June through October. These weather inputs are resampled from 50 historical years. The historical years are from 1952 to 1969, and 1977 to 2009, as historical weather inputs are not available from 1970 to 1976. The historical weather for the forecast year is not included. The model produces daily time series of streamflows for each weather sequence, which are used to compute the flood season flow volumes. Hence, the ensemble streamflow prediction process produces a 50-member ensemble of simulated flood volumes, each conditioned on the same initial conditions.

Climate index weighting was then used to represent the influence of climate state. For each flood volume y_i in the ensemble, a climate index θ_i is assigned; θ_i is the May NINO4.0 index for the same historical year as

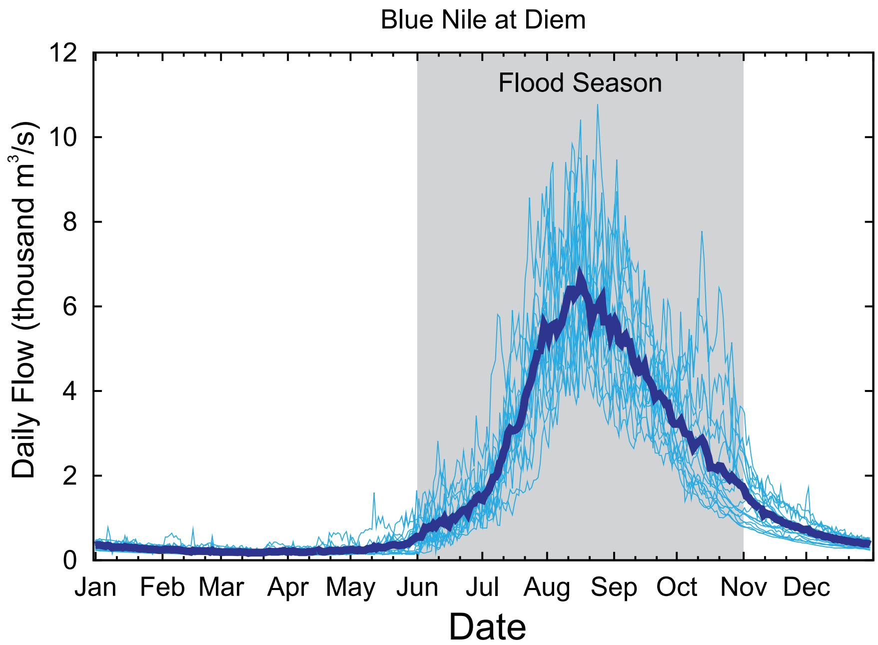

Figure 1. Annual cycle of daily flow (in m^3/s) for the Blue Nile at Diem. Daily flow time series for 1992–2010 are shown (thin lines). The average daily flow for the 1992–2010 period is also shown (thick line). The Nile flood season, defined by the months of June–October (shaded area), is related to seasonal rains in the Ethiopian highlands. Variability in the Nile flood from year to year is related to the El Niño Southern Oscillation.

the weather inputs. The climate state at the time of the forecast θ is represented by the May NINO4.0 index for the forecast year. Climate-weighted ensemble forecasts were then determined using the Bayesian method and two traditional climate index weighting methods.

4.2. Bayesian Climate Index Weighted Forecasts

With Bayesian climate index weighting, the original ensemble streamflow forecast from the Nile Forecast System, $y_i, i=1, \ldots, 50$, represents a sample drawn from the prior distribution. Hence, the prior represents a forecast conditioned on the initial basin moisture state in June of the forecast year. The next step is to define climate weights for each ensemble member using equation (7) to condition the forecast on the

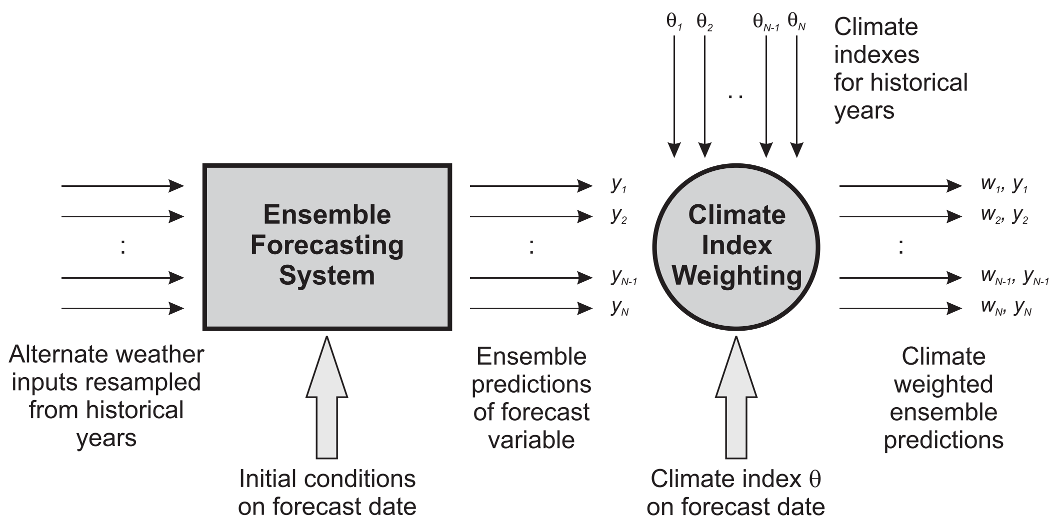

Figure 2. Ensemble streamflow prediction (ESP) forecasting process with climate index weighting postprocessing.

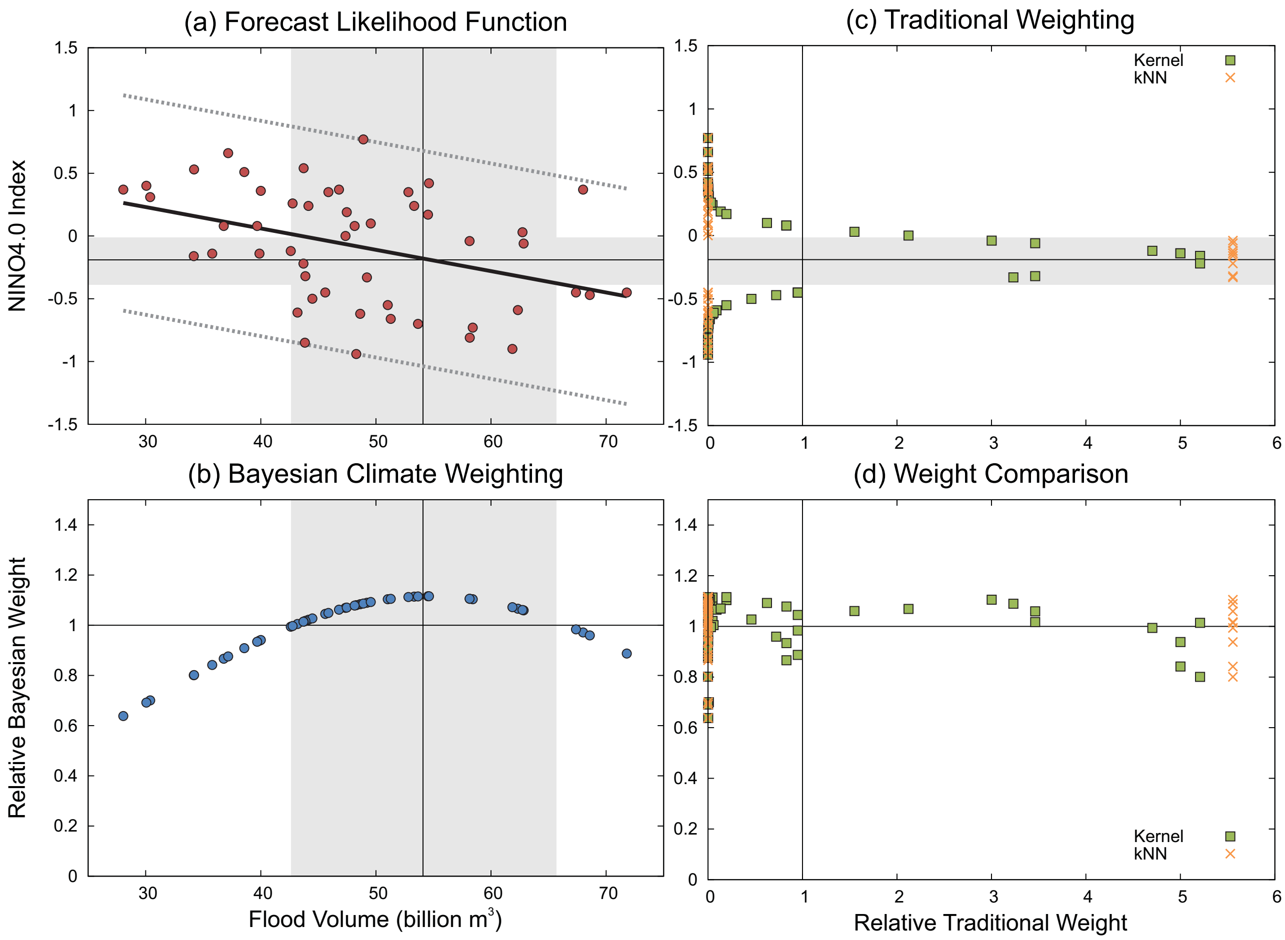

Figure 3. Ensemble forecast of 2001 flood volume for the Blue Nile at Diem. (a) The likelihood function for the relationship between the May NINO4.0 index and Nile flood volume for the ensemble members. The likelihood function is represented with a linear regression model. The expected value of the NINO4.0 index given the flood volume $\bar{\theta}(y)$ is shown by the solid line. The variability of the relationship is shown by the dashed lines at $\pm 2\sigma_\varepsilon$. The NINO4.0 index for May prior to the 2001 flood season (-0.19) is indicated by the horizontal line. A flood volume of 54.1 billion m^3, as shown by the vertical line, corresponds to an expected NINO4.0 index of -0.19. (b) The Bayesian climate index weights w_i for the 2001 forecasts; the weights relative to those for equal weighting are shown. (c) The kernel and kNN climate index weights w_i for the 2001 forecast; the weights relative to those for equal weighting are shown. (d) Compares relative Bayesian climate index weights versus the kernel and kNN climate index weights.

climate state. This requires the estimation of the likelihood function $f_\theta(\theta|y)$ for the forecast. An example is shown for the 2001 ensemble flood volume forecast in Figure 3.

Figure 3a shows a scatter plot of the May NINO4.0 index and the simulated flood volume for the 50 ensemble members. Also shown is the May NINO4.0 index value (-0.19) at the time of the June 2001 forecast. The likelihood function $f_\theta(\theta|y)$ is the distribution of NINO4.0 index θ conditioned on the ensemble flood volume y. This likelihood function can be estimated by a regression model [*Faber and Stedinger*, 2001]:

$$\theta = \bar{\theta}(y) + \varepsilon, \tag{8}$$

where $\bar{\theta}(y)$ is the expected value of the climate index θ given the observation y and ε is the residual model error. At this point, one must choose a mathematical form for the likelihood function model, which introduces some subjectivity to the method. Given the sample shown in Figure 3a, we hypothesize a simple linear model for $\bar{\theta}(y)$. The residual error ε is assumed to be normally distributed with constant variance σ_ε^2. For this hypothesized model, the likelihood $f_\theta(\theta|y)$ is a normal density function:

$$f_\theta(\theta|y) = \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} e^{-\frac{(\bar{\theta}(y)-\theta)^2}{2\sigma_\varepsilon^2}}. \tag{9}$$

The estimated model fitted by linear regression is shown in Figure 3a.

Using equation (9) and the NINO4.0 index of -0.19 for the 2001 forecast, Bayesian climate index weights are estimated using equation (7) for each ensemble member y_i. Figure 3b shows these weights w_i, relative to equal weighting of ensemble members; values greater than 1 apply more weight than equal weighting to the ensemble member, and values less than 1 apply less weight than equal weighting. Note that the Bayesian climate index weights vary smoothly with the ensemble member flood volume y_i. For a given volume y_i, if the expected value $\bar{\theta}(y_i)$ is close to the observed index value θ at the time of the forecast (-0.19), the Bayesian climate weight is large; in other words, the ensemble outcome is more likely given the climate state.

4.3. Traditional Climate Index Weighted Forecasts

We also produced forecasts using two traditional climate index weighting methods. The first method is kernel climate index weighting [*Smith et al*., 1992] using a Gaussian kernel function. The Gaussian kernel has the form

$$K(x)=\frac{1}{\sqrt{2\pi}h}e^{-\frac{x^2}{2h^2}}, \tag{10}$$

where h represents the kernel bandwidth parameter. The Gaussian kernel has a similar mathematical form to the likelihood function in equation (9). However, the values depend on a distance x, defined as $|\theta_i-\theta|$ using the climate index for the ensemble member (see equation (5)), as opposed to $|\bar{\theta}(y_i)-\theta|$ for the likelihood function. The second method is a k-nearest neighbor (kNN) approach [*Najafi et al*., 2012], which uses the distance $|\theta_i-\theta|$ to find the nearest neighbors to the climate index θ at the time of the forecast. The k nearest to θ are assigned equal weights of $1/k$, and all others are assigned a weight of zero. This method is analogous to a block adjustment approach [*Stedinger and Kim*, 2010], but the category is defined for each forecast based on the k-nearest neighbors.

To apply the Bayesian climate index weighting for the 2001 flood volume forecast, all that was needed was the ensemble forecast itself, the May NINO4.0 index values associated with each ensemble member, and the May 2001 NINO4.0 index. For the assumed likelihood function model, the required model parameters were estimated by linear regression. In contrast, for the two traditional climate index methods, we still need to select h for the kernel method and k for the kNN method. Rather than make an arbitrary choice, we will "stack the deck" in favor of the traditional methods by letting them use all the hindcasts and corresponding flood volume observations (including the one for 2001) to find optimal parameter values. In particular, using the hindcasts and observations, we found values of h and k that maximize the average probability forecast skill $\overline{SS}$, a weighted-average skill score that is analogous to the continuous ranked probability skill score [*Bradley and Schwartz*, 2011]. The optimal values are 0.14 for bandwidth h, and 9 for the nearest neighbors k. We will revisit the parameter selection issue in the next section.

Using these optimal parameters, Figure 3c shows the relative climate index weights for the kernel and kNN methods for the 2001 forecast. Note that the kernel climate index weights vary smoothly with the ensemble member climate index value θ_i. If the climate index θ_i is close to the index value θ at the time of the forecast (-0.19), the kernel climate weight is large; ensemble members are weighted more if their climate state is similar to that at the time of the forecast. The kNN method acts in a similar way, assigning large weights to the 9 ensemble members with climate index values θ_i closest to θ.

The arrangement of the plots in Figure 3 illustrates the orientation of weight assignment, which is different for Bayesian and traditional climate index weighting. For the Bayesian method, the weights are a function of the ensemble member outcome y_i, whereas with traditional methods the weights are a function of their climate index θ_i. This difference can lead to the assignment of very different weights to each ensemble member, as illustrated by the comparison in Figure 3d. Indeed, for the 2001 forecast, there is no apparent association between the Bayesian and traditional weights assigned. What is clear is that the traditional methods discriminate more for climate state; a fraction of the ensemble members receive very high weights, whereas the others receive virtually no weight.

4.4. Climate-Weighted Ensemble Probability Distribution Forecasts

Figure 4 compares the flood volume ensemble forecast probabilities for the 2001 season based on Bayesian and traditional climate index weighting (see equation (2)). Also shown are the probabilities for equal

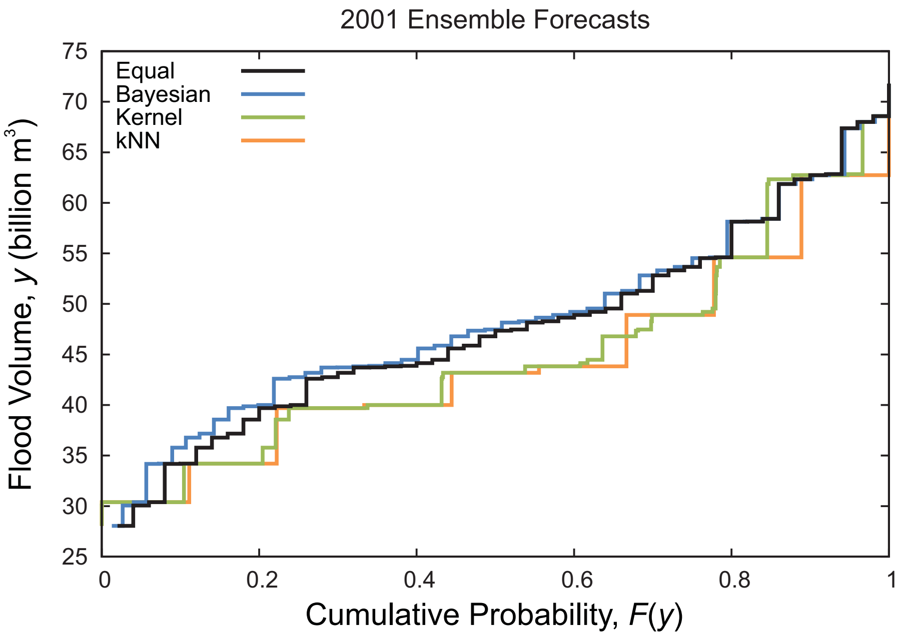

Figure 4. Ensemble forecasts of 2001 flood season flow volume for the Blue Nile at Diem. The estimated probability distribution is computed with the ensemble members using equation (2) and appears as a series of steps at the ensemble member values. Probability distribution forecasts are shown for equal weighting of ensemble members (black line), which represents the prior flood volume distribution without climate information, and with climate index weighting using the Bayesian method (blue line), the kernel method (green line), and the kNN method (orange line). Note that the ensemble forecasts weighted by the NINO4.0 climate index shift away from the equal-weighted prior distribution, based on how the different methods interpret the climate signal at the time of the forecast (the May 2001 NINO4.0 index).

weighting, which corresponds to the prior distribution before conditioning on the climate index (the original ensemble forecast). Obviously, the methods are interpreting the near-neutral ENSO state (NINO4.0 of -0.19) quite differently. Given the current climate index, the Bayesian forecast shifts slightly upward from the prior distribution, indicating that higher flood volumes are more likely given the climate state. In contrast, the kernel and kNN forecasts shift downward, indicating that much lower flood volumes are more likely. Also, because of the higher discrimination for the traditional climate index methods, the distribution exhibits a few long plateaus (corresponding to the fraction of ensemble members with large weights); in contrast, the plateaus are smaller for the Bayesian method and are a more consistent length (as weights vary smoothly with ensemble members y_i).

The climate index weighting methods were used to generate 18 retrospective ensemble forecasts, one for each flood season from 1992 to 2009. Figure 5 shows the ensemble forecasts using the Bayesian and kernel climate index weighting methods for 2001–2009. Also shown is the original ensemble forecast with equal weighting (the prior distribution before conditioning on the climate index); a shift away from the equal weighting forecast is the predicted influence of the climate index on forecast flood volumes. The observed flood volumes are indicated by the $\times$ symbol. In half the years, the two climate-weighted forecasts shift in a similar direction (2002, 2004, 2005, 2008, and 2009). In all those instances, the climate-weighted forecasts are accurate, shifting from the equal-weighted forecast in the same direction of the observed flood volume. However, in the other years, the two climate-weighted forecasts shift in opposite directions (as in the 2001 forecast illustrated in Figure 4). On some occasions, the Bayesian-weighted forecast is more accurate (2001 and 2003); in others, the kernel-weighted forecast is more accurate (2006 and 2007).

4.5. Forecast Verification Comparison

To assess the quality of the ensemble forecasts, Figure 6 shows the forecast skill for the equal-weighted forecasts and the three climate-weighted forecasts, based on all 18 forecast periods. For any given threshold y, a probability forecast that the flood volume will be below the threshold y is computed for all 18

10.1002/2014WR016811

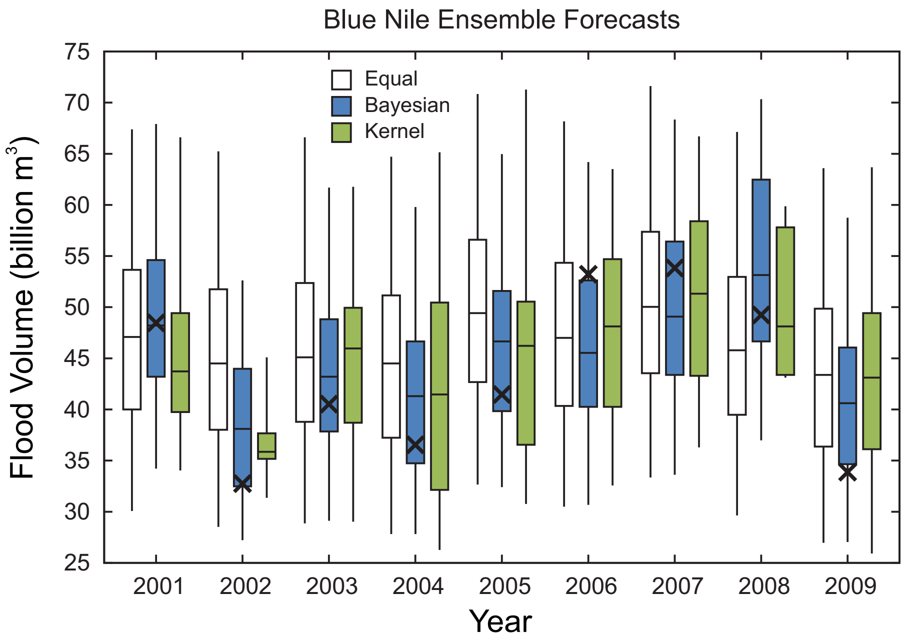

Figure 5. Ensemble forecasts of flood volume for the Blue Nile at Diem for 2001–2009. Three ensemble forecasts are shown for each year; one is the original forecast with on equal weighting (left box with no shading), the second is based on Bayesian climate index weighting (middle box with blue shading), and the third is based on a kernel climate index weighting (right box with green shading). The flood volume observed that year is indicated ($\times$ symbol). The ensemble forecasts are represented by box plots. The box shows the forecast 25% and 75% flow levels; the median (50%) level is indicated by the horizontal line within the box. Whiskers extending from box end at the forecast 5% and 95% flow levels.

ensemble forecast probability distributions (e.g., Figure 4). These are then compared to the actual outcome. A range of flood volume thresholds y were examined, chosen as the midpoint between the 18 observed flood volumes, to see how well the ensemble forecasts predict outcomes ranging from low to high flood

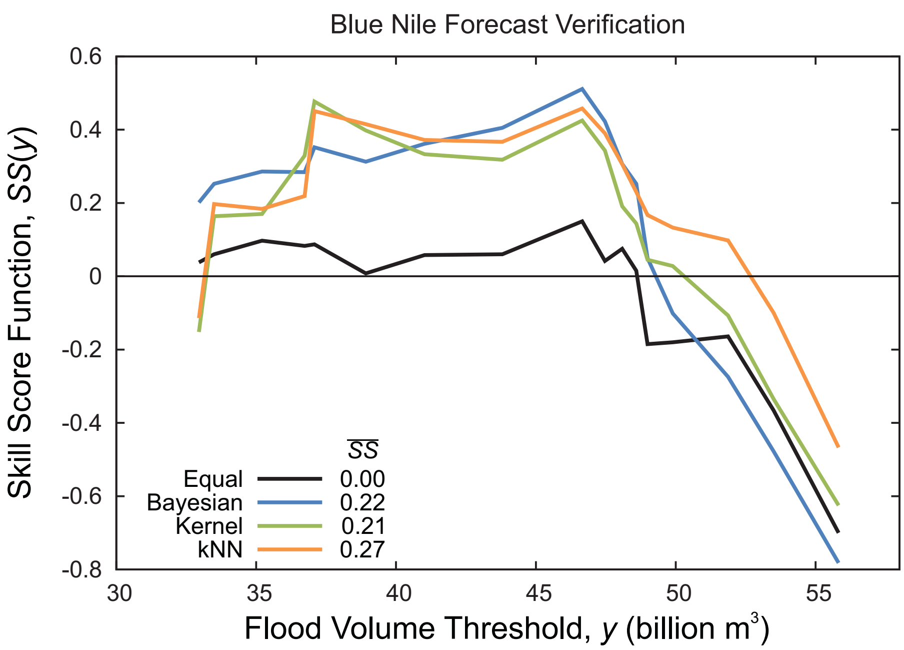

Figure 6. Verification of ensemble forecasts of flood volume for the Blue Nile at Diem. The MSE skill score SS is plotted as a function of the flood season flow volume threshold y for ensemble forecasts based on equal weighting of ensemble members (black line), which represents the prior distribution forecast without climate information, and with climate index weighting using the Bayesian method (blue line), the Gaussian kernel method (green line), and the kNN method (orange line). Verification statistics are based on 18 flood volume forecasts for 1992–2009.

volumes [*Bradley et al*., 2004]. Their relative accuracy was assessed using the mean squared error (MSE) skill score [*Wilks*, 2011], also commonly known as the Brier skill score [*Brier*, 1950]. A skill score greater than 0 indicates that the set of ensemble probability forecasts are skillful at predicting flood volume occurrences below (or above) the threshold; that is, they are more accurate than a reference (constant) climatology probability forecast for the threshold (i.e., the observed relative frequency of flood volumes below the threshold based on the historical record).

Without climate information (equal weighting), the forecast models' flood volume forecasts have modest skill only for flood volume thresholds less than about 49 billion m^3; the forecasts are less accurate than climatology forecasts for higher thresholds. This means that the basin state (at this time of year) has limited predictive ability. However, all three climate index weighting methods significantly improve forecast skill, except again at those high flood thresholds; the lowest threshold also has negative skill for the kernel and kNN methods. Note that seeing lower skill scores at the extremes is not uncommon; for a small verification sample and only a few flow volume observations more extreme than the threshold, the skill score is very sensitive to the performance in those cases. We have observed similar drops in skill for extreme events in other ensemble forecasting examples [*Bradley et al*., 2004; *Bradley and Schwartz*, 2011]. Still, the consistent negative skill scores for higher flood volumes does suggests that the forecast model is not predicting the highest flood volumes accurately.

Even though the Bayesian and traditional methods interpret climate information in a different way, and their climate-weighted forecasts are quite different (see Figures 4 and 5), the three climate index weighting methods have comparable probability forecast skill. Not surprisingly, the skill score functions for the two traditional climate index weighting methods are more consistent with each other. Overall, in terms of the average skill $\overline{SS}$, a weighted-average measure of the skill score functions shown, the kNN method has the highest average skill (0.27), which results from its superior performance for the high flood volume extremes. The Bayesian method (0.22) and the kernel method (0.21) have lower average skill. Given the way that the methods were applied, one could conclude that both the traditional and Bayesian climate index methods produce very different forecasts, but with comparable skill.

However, it is important to revisit how the traditional methods were applied. For these two methods, we used the 18 retrospective forecasts (hindcasts) and their corresponding flood volume observations to selected parameters that maximized $\overline{SS}$. That is, the parameters were optimized for the data set and verification metric used in our comparison; no other parameter choices would perform better. In contrast, the Bayesian method did not utilize hindcast information at all; its parameters are defined by the assumed likelihood function model and estimated only with data available at the time of the forecast (the ensemble members and their corresponding climate index values). To be on an equal footing, we would have had to make a (subjective) choice of parameters for the traditional methods without using hindcast information. To illustrate, consider the following choices. For the bandwidth h, one might choose to set $h = \sigma_\varepsilon$ from the likelihood function regression, since the kernel function has a similar mathematical form ($h = 0.43$ from the 2001 forecast example). For the nearest neighbors k, one might choose to use (say) one third of the ensemble members ($k = 17$). For these subjective choices, the average skill $\overline{SS}$ drops to 0.15 for the kernel method and 0.16 for the kNN method, both noticeably lower than for the Bayesian method (0.22). Hence, it would be fairer to conclude that the traditional methods can produce comparable skill, if there are hindcasts available to optimize their parameters. The fact that Bayesian climate index weighting performs as well, but does not require any hindcast information to apply it, is a significant advantage of the Bayesian approach.

4.6. Model Selection for Likelihood Function

Of course, there is a degree of subjectivity in Bayesian climate index weighting introduced by the choice of a likelihood function model. For the Blue Nile forecasts, we choose a linear regression model, with its corresponding likelihood function shown in equation (9). In this section, two alternate approaches for estimating the likelihood function are examined.

The first alternative uses a nonparametric locally weighted scatter plot smoothing (LOWESS) [*Cleveland*, 1979] to define the relationship between θ and y; a locally weighted estimate of the expected value $\bar{\theta}(y)$ and the standard deviation $\sigma_\varepsilon(y)$ are determined for each ensemble member y_i. The

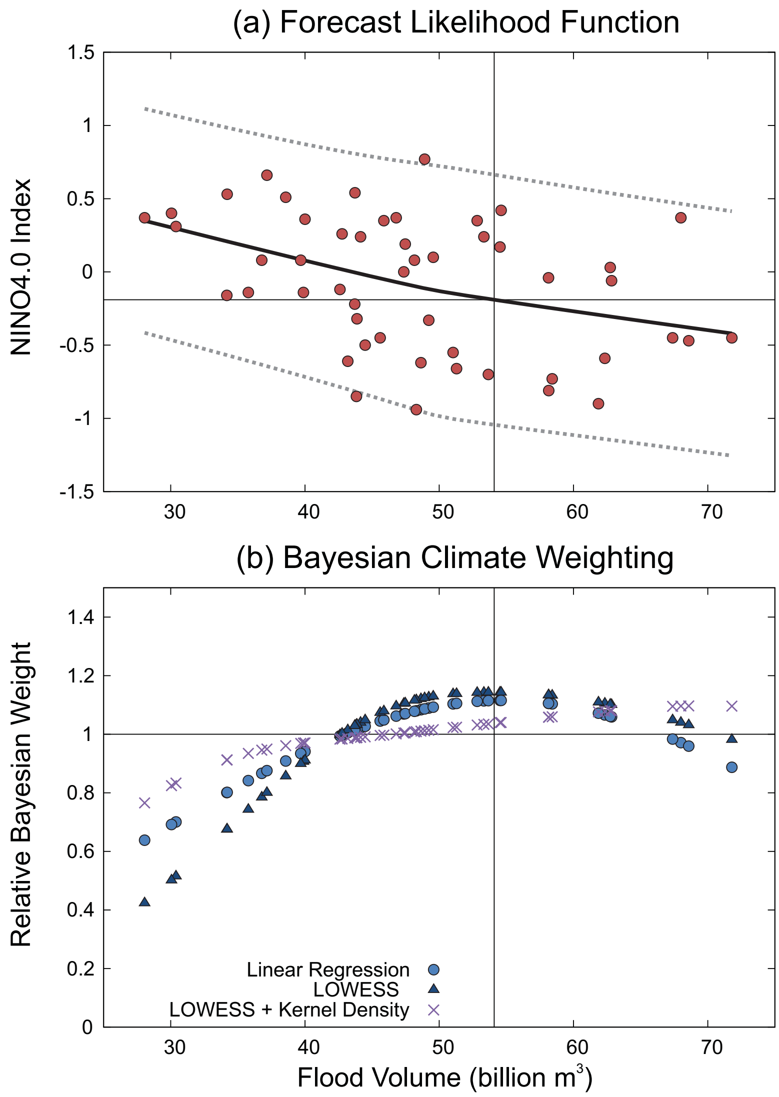

Figure 7. Ensemble forecast of 2001 flood volume for the Blue Nile at Diem. (a) The likelihood function for the relationship between the May NINO4.0 index and Nile flood volume for the ensemble members. The likelihood function is represented with a LOWESS regression model. The expected value of the NINO4.0 index given the flood volume $\bar{\theta}(y)$ is shown by the solid line. The variability of the relationship is shown by the dashed lines at $\pm 2\sigma_\varepsilon(y)$. The NINO4.0 index for May prior to the 2001 flood season (-0.19) is indicated by the horizontal line. A flood volume of 54.1 billion m^3, as shown by the vertical line, corresponds to an expected NINO4.0 index of -0.19. (b) The Bayesian climate index weights w_i for the 2001 forecasts; the weights relative to those for equal weighting are shown. The weights are based on three likelihood function models: a linear regression model (as shown in Figure 3a), a LOWESS regression model with varying standard deviation (as shown in Figure 7a), and LOWESS regression with a Gaussian kernel density estimation.

estimated relationship for the 2001 forecast is shown in Figure 7a. Using this hypothesized relationship, the likelihood function is

$$f_\theta(\theta|y) = \frac{1}{\sqrt{2\pi}\sigma_\varepsilon(y)} e^{-\frac{(\bar{\theta}(y)-\theta)^2}{2\sigma_\varepsilon(y)^2}}. \qquad (11)$$

That is, we assume that the residual model error is normally distributed, with zero mean and varying variance defined by $\sigma_\varepsilon^2(y)$.

The second alternative also uses the LOWESS estimate of $\bar{\theta}(y)$ but uses a nonparametric kernel density estimator for the likelihood function. Using $\bar{\theta}(y)$, the residual error from the regression is

$$z_i = \theta_i - \bar{\theta}(y_i). \qquad (12)$$

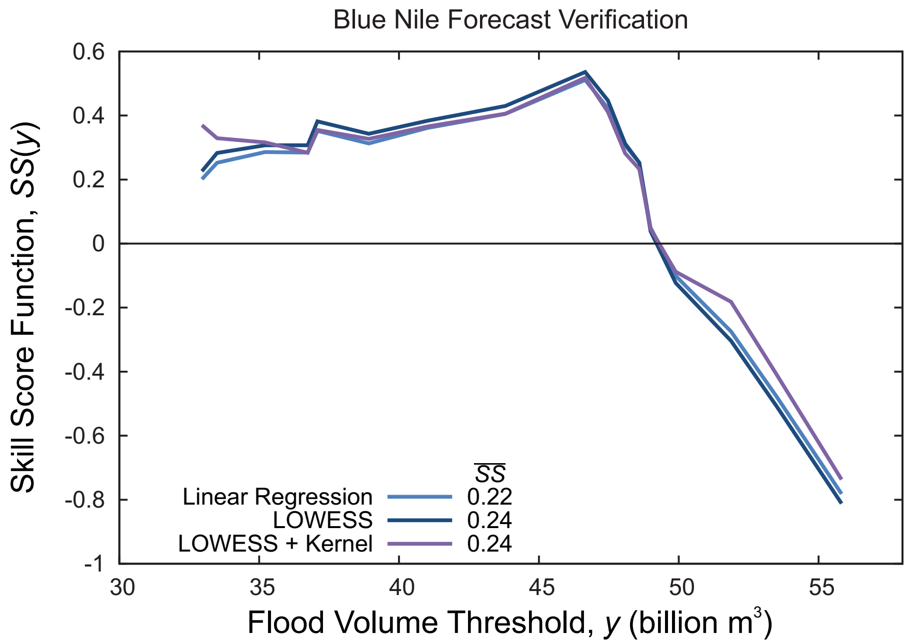

Figure 8. Verification of ensemble forecasts of flood volume for the Blue Nile at Diem. The MSE skill score *SS* is plotted as a function of the flood season flow volume threshold *y* for ensemble forecasts based on Bayesian climate index weighting with a linear regression model (light blue line), a LOWESS regression model with varying standard deviation (dark blue line), and a LOWESS regression model with Gaussian kernel density estimation (purple line). Also shown in the skill score for equal weighting of ensemble members (no climate information). Verification statistics are based of 18 flood volume forecasts for 1992–2009. Note that the probability forecasts are skillful ($SS > 0$) for most flow volume thresholds, except at low and high flow extremes. All Bayesian climate index weighted forecasts significantly increase probability forecast skill, except for high flood volumes, where all the forecasts are not skillful.

Using the sample of residuals $z_i, i=1,\ldots,N$, and the Gaussian kernel $K(x)$ shown in equation (10) with bandwidth h estimated from the sample of residuals using the "solve-the-equation" plug in method by *Sheather and Jones* [1991], the kernel density estimator of the likelihood function is

$$f_\theta(\theta|y_i)=\frac{1}{N}\sum_{i=1}^{N}K(\theta-\bar{\theta}(y_i)-z_i). \tag{13}$$

Note that this approach makes no distributional assumption for the likelihood function. However, it does assume that the model errors are homoscedastic (as is also assumed for the linear regression model).

Figure 7b shows Bayesian climate index weights w_i for the three approaches, relative to equal weighting of ensemble members. For the near-neutral ENSO conditions in June 2001, all three approaches weight higher simulated flood volumes more than lower flood volumes. As noted before, the weights vary smoothly assuming the parametric linear regression model for the likelihood function. Since the LOWESS regression closely approximates a linear relationship, and the estimated standard error $\sigma_\varepsilon(y)$ is nearly constant for all simulated flood volumes, the weights are similar to those by linear regression. In contrast, the weights for the kernel density estimator approach depart from the LOWESS weights, even though both share the same relation for $\bar{\theta}(y)$. The differences are due to the shape of the conditional density function $f_\theta(\theta|y)$.

Figure 8 shows the probability forecast skill for the three likelihood function alternatives based on their 18 ensemble forecasts for 1992–2009. Only slight differences are observed between the three alternatives. For instance, the additional flexibility of the LOWESS likelihood function model increases the probability forecast skill slightly over the linear regression model for most flood volume thresholds. The use of LOWESS and a kernel density estimator for the likelihood function model also tends to improve skill slightly; the improvement is most obvious for the lowest and highest flood thresholds. Yet overall, the forecasts and their quality are robust to the likelihood function model choice in this case, and all three effectively incorporate climate index information to improve ensemble forecasts. However, in other applications, the added flexibility of nonparametric approaches may be necessary. As this example shows, given a sufficient number of ensemble members, such approaches could be used in general without need to invoke distribution assumptions, even when simple parametric approaches are suitable.

5. Discussion

In this section, we explore issues related to the Bayesian climate index weighting method. In particular, we will discuss its relation to other Bayesian methods being used in forecasting, the advantages of the method over alternate approaches, some considerations for applying the Bayesian climate index weighting method, and its extension for use with multiple climate indexes.

5.1. Bayesian Methods in Hydrologic Forecasting

Although the Bayesian method presented is a new approach for climate index weighting, the concept is similar to other methods employed in ensemble forecasting. In particular, *Kelman et al*. [1990] use a Bayesian approach to assign probabilities (equivalent to weights) to a collection of streamflow scenarios (equivalent to ensemble traces) as part of their sampling stochastic dynamic programming (SSDP) method. Given an observation of streamflow volume up to some point in time, Bayes' theorem was used to assign probabilities for each streamflow scenario. As illustrated with the climate index weighting, this can be accomplished using a regression model relationship between historical observations and the forecast variable. *Stedinger and Kim* [2010] also present an approach for assigning probabilities to ensemble members based on a climate forecast; here the aim was to use Bayes' theorem to adjust the historical distribution of a climate variable (the prior) to match the climate forecast distribution.

Others have used Bayesian methods with climate indexes for streamflow forecasting [*Wang et al*., 2009; *Robertson and Wang*, 2013; *Bennett et al*., 2014; *Lima et al*., 2014] or nonstationary frequency analysis, a probability distribution forecast of a hydrologic variable conditioned on climate state [*El Adlouni et al*., 2007; *Kwon et al*., 2008; *Ouarda and El-Adlouni*, 2011; *Steinschneider and Brown*, 2012]. In these applications, a parametric model for the probability distribution was chosen for the hydrologic variable, and its model parameters are predicted by Bayesian inference; the Bayesian method begins with a prior distribution of the model parameters and estimates the posterior distribution of model parameters conditioned on climate state. A subjective prior is often needed. Bayesian climate index weighting differs from these approaches in that it predicts the forecast variable directly by Bayesian inference and uses an informative prior—the distribution of the forecast variable predicted by the ensemble forecast without climate information.

Finally, *Krzysztofowicz* [1999] has set the entire hydrologic forecast process within a generalized Bayesian framework. A Bayesian forecasting system has components to represent the uncertainties from the weather inputs and the model outputs, resulting in probabilistic statements that account for the total uncertainty of the forecasts. This approach has been extended to ensemble forecasting systems [*Herr and Krzysztofowicz*, 2015].

5.2. Advantages of Bayesian Climate Index Weighting

As enumerated in section 1, there are many ways to make hydrologic ensemble forecasts conditioned on climate information. However, Bayesian climate index weighting has some unique advantages. First, the approach has a strong theoretical basis; it assigns weights that accomplish Bayesian updating of an (unweighted) ensemble forecast to reflect the given climate state information. The weights are easy to interpret; they represent the relative likelihood of each ensemble member in the sample, given the observed climate index at the time of the forecast. The method is also simple and straight forward to apply; it can easily be added as a postprocessing step to an existing forecasting system. Compared to preprocessing methods, like creating weather inputs for a hydrologic forecast model by climate forecast downscaling, climate index weighting is far simpler and can even be more effective [*Werner et al*., 2004].

Given the ensemble forecast, and the climate index appropriate for each ensemble member, one can estimate the likelihood function by a variety of parametric and nonparametric methods. Because the data needed to develop a likelihood function model are contained within the ensemble forecast itself, the method is *self-calibrating* for individual forecasts; hindcasts are not needed to calibrate optimal model parameters. This self-calibrating feature means the method can be applied directly to any existing ensemble forecasting system, regardless of whether it is brand new or has been enhanced or changed in some way over time. Furthermore, the method is *self-adapting*, based on the strength of the relationship defined by

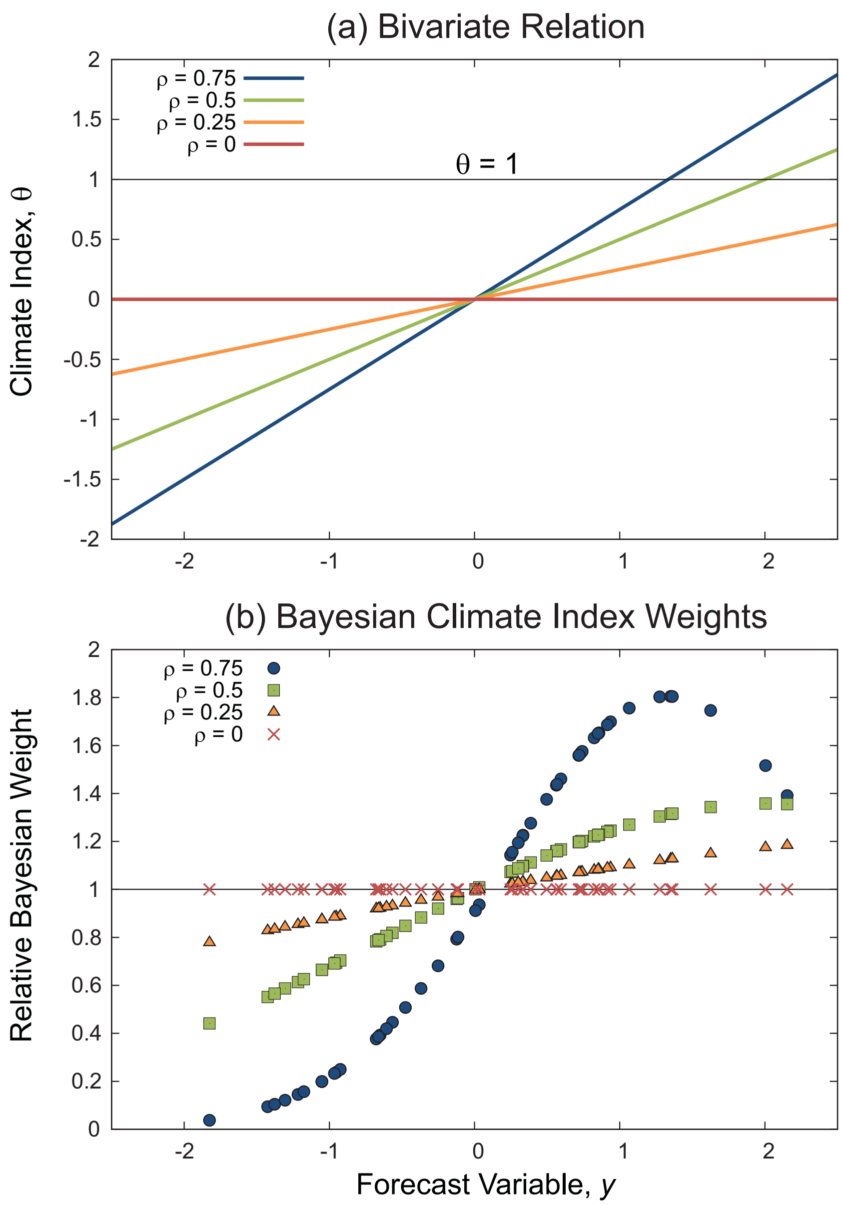

Figure 9. A hypothetical examples showing the self-adaptability of the Bayesian climate index weights to the strength of the relationship with the climate index. A bivariate normal relationship with zero means and unit variances is assumed between the climate index θ_i and the forecast variable y_i. The strength of the relationship depends on the correlation ρ. (a) The relation for different values of ρ. (b) The climate index weights assigned to a random sample assuming the climate index θ at the time of the forecast is 1.

the likelihood function. This fact is illustrated with a simple hypothetical bivariate likelihood model in Figure 9. Note that when the relationship between the ensemble members and the climate index is strong, the weights are able to discriminate ensemble members into more (or less) likely outcomes. But in the case, where there is no correlation of ensemble members to the climate index (the two are independent), the method defaults to equal weighting of ensemble members (the posterior distribution equals the prior distribution when there is no discernable climate signal).

5.3. Comparison With Traditional Climate Index Weighting

The Bayesian method approaches climate index weighting in a different way than traditional methods. Traditional climate index weighting assigns weights based on the similarity of the observed climate index θ (at the time of the forecast) with that for each ensemble member θ_i. However, the discrimination of the weights depends on the parameter values selected for the method. To assure optimal performance, the selection of the weights requires parameter calibration with hindcasts and observations. This can be a tedious process, even if hindcasts are already available. Consider a forecasting system that issues ensemble forecasts of multiple forecast variables on a monthly (or weekly) basis; the calibration process must be repeated for each forecast variable, and each forecast issuance date. Unfortunately, if the forecast system is enhanced or changed in some way, new hindcasts must be generated and the parameters must be calibrated again for the new setup. Yet even with parameter calibration in the Blue Nile example, the performance of the Bayesian method (which requires no hindcast calibration) was comparable. Clearly, a significant advantage of the Bayesian method is that it can be applied immediately to all forecast variables for each forecast issued; its self-calibrating and self-adapting features assure that any climate signal in forecast flows (if it exists) will be reflected properly in the climate-weighted forecasts.

The way that the Bayesian method assigns weights is based on the relative likelihood of each ensemble member outcome y_i in the sample, as defined by the likelihood function model. Since the weights depend on y_i, they can be assigned to ensemble members even when the climate index θ_i is unavailable. Consider an application where the climate index is available starting in 1950 (like the NINO4.0 index used in our example); however, the forecasting system generates some ensemble members using historical information available prior to 1950. Traditional climate index weighting cannot assign weights to the pre-1950 ensemble members that have no climate index. In contrast, once a likelihood function has been estimated using the subset of ensemble members where θ_i are available, the Bayesian method can use that model to assign weights using y_i for all the pre-1950 ensemble members. Hence, the Bayesian method is more versatile than traditional climate index weighting methods, as it can assign weights to all ensemble forecast members if there is a sufficient period of overlap with available climate index information.

5.4. Application Considerations

As the Blue Nile example shows, climate index weighting can be applied directly to ensemble forecast model-simulated flows. From the perspective of Bayes' theorem (equation (6)), the prior $f(y)$ represents the distribution of simulated flow conditioned on basin state; the updated posterior $f(y|\theta)$ represents the distribution of simulated flows conditioned on the climate state and the basin state. However, if a forecast model systematically underpredicts or overpredicts flows, the bias of the simulated flow needs to be accounted for to yield a reliable forecast of the future flow. Bias-correction methods are commonly used in ensemble streamflow forecasting to adjust the simulated flow for each ensemble member and produce an unbiased ensemble forecast [*Wood and Lettenmaier*, 2006; *Seo et al.*, 2006; *Hashino et al.*, 2007; *Bogner and Kalas*, 2008; *Brown and Seo*, 2010; *Zhao et al.*, 2011; *Brown and Seo*, 2013; *Pagano et al.*, 2013; *Pokhrel et al.*, 2013]. One could apply bias correction either before climate index weighting, and then use the bias-corrected flows to develop the likelihood function, or more simply, after climate index weighting. To carry out bias correction, the forecast model is typically run in a simulation mode to generate model-simulated flows for a historical period; the simulated flows are combined with observed flows to diagnose and correct biases. Therefore, some retrospective information is needed from the forecast model for bias correction (but not usually retrospective forecasts). Although bias correction was not applied in the Blue Nile example, it may be a necessary step in other applications.

Predictability in seasonal hydrologic forecasting can come from two sources—a climate signal and memory of the hydrologic system. If future flows depend on climate state, climate index weighting may be able to add skill to ensemble forecasts. Some preliminary analysis is required to select an appropriate climate index for index weighting. Many approaches have been used, including selection of the best climate index by correlation analysis, or by using methods like principal component analysis with multiple indexes to define an optimal climate predictor [*Moradkhani and Meier*, 2010; *Najafi et al.*, 2012]. The best climate index may vary by forecast calendar date, as the relationship can depend on when the forecast is issued. Hydrologic forecast models can exploit memory of the hydrologic system by properly representing the current basin moisture state as initial conditions at the time the forecast is made. In the Blue Nile example, the initial conditions only provided modest skill in predicting future flows (see Figure 6); knowledge of the climate state was a greater source of skill.

In situations where the basin moisture state offers no predictability, climate index weighting would be more easily applied without a forecast model. A second example, using historical flow observations as the ensemble forecast, is provided as supporting information to this paper. In this example, the historical record of observed flow is used as an ensemble forecast; that is, the prior distribution is the unconditional distribution (or climatology) of observed flows. The sample of observed flows is combined with historical climate index information to estimate the likelihood function, which is then used to find weights that represent the current climate state. Hence, the Bayesian climate index weighting method can be applied not only to hydrologic model-generated ensemble forecasts but also with ensemble forecasts generated by resampling observed flows from the historical record.

Furthermore, Bayesian climate index weighting need not be restricted to climate indexes. For example, if a climate forecast model (with available hindcasts) produces skillful forecasts of future conditions, one could construct indexes based on these forecasts. Likewise, if a measure of the basin moisture state (e.g., soil moisture or antecedent streamflow) is predictive of future flows, it could also be used as an index. Given these options, the ability to apply Bayesian climate index weighting with multiple indexes may be needed.

5.5. Extension to Multiple Climate Indexes

Although the Bayesian climate index weighting method was illustrated using a single climate index, the approach could easily be extended for use with multiple climate indexes. Consider the case where the streamflow variable Y depends on multiple climate indexes $\theta^1, \theta^2, \ldots, \theta^m$. For this case, Bayes' theorem can be written as

$$f(y|\theta^1, \theta^2, \ldots, \theta^m) = \frac{f_\theta(\theta^1, \theta^2, \ldots, \theta^m|y)f(y)}{f_\theta(\theta^1, \theta^2, \ldots, \theta^m)}, \tag{14}$$

and by extension, the weighting function can be written as

$$w_i = \frac{f_\theta(\theta^1, \theta^2, \ldots, \theta^m|y_i)}{\sum_{j=1}^{N} f_\theta(\theta^1, \theta^2, \ldots, \theta^m|y_j)}. \tag{15}$$

Therefore, to define the weights with multiple climate indexes, a multivariate likelihood function model must be developed. *Kelman et al.* [1990] illustrate such an example for streamflow forecasting. For the special case where the climate indexes are all independent, then $f_\theta(\theta^1, \theta^2, \ldots, \theta^m|y) = f_{\theta^1}(\theta^1|y) f_{\theta^2}(\theta^2|y)\cdots f_{\theta^m}(\theta^m|y)$, so the weighting function simplifies to

$$w_i = \frac{f_{\theta^1}(\theta^1|y_i) f_{\theta^2}(\theta^2|y_i)\cdots f_{\theta^m}(\theta^m|y_i)}{\sum_{j=1}^{N} f_{\theta^1}(\theta^1|y_j) f_{\theta^2}(\theta^2|y_j)\cdots f_{\theta^m}(\theta^m|y_j)}. \tag{16}$$

Clearly, the case of independent climate indexes shown in equation (16) is easier to manage, as it is simpler to estimate a likelihood function model between a single index and streamflow. But even when the multiple climate indexes are correlated (not independent), one could first apply a mathematical procedure like principle component analysis (PCA) [*Wilks*, 2011] to find uncorrelated variables. Replacing the original climate indexes with these uncorrelated variables would allow one to use equation (16) and its simpler likelihood function model. *Najafi et al.* [2012] illustrates the use of PCA with multiple indexes for climate index weighting.

6. Summary and Conclusions

A Bayesian method for climate index weighting of ensemble forecasts was presented. The method is based on a sampling-resampling approach for Bayesian updating. The original ensemble members define the prior distribution. The relationship between the ensemble members and a climate index is then used to define a likelihood function (or conditional distribution). Given an observation of the climate index at the time of the forecast, the likelihood function is used to assign weights to each ensemble member. The weights define the relative likelihood of each ensemble member given the observed climate index. The weighted ensemble forecast is then used to estimate the posterior distribution—the distribution of the forecast variable conditioned on the climate index.

The Bayesian climate index weighting method for ensemble forecasting was illustrated with an example for the Blue Nile, where flood season flow volumes are correlated with ENSO. Ensemble streamflow predictions from a river forecast model were used as a sample of the prior distribution of flood volumes. Bayesian climate index weighting was contrasted with more traditional climate index weighting methods. The two approaches assign very different weights to ensemble members, resulting in very different forecasts. Still, the overall forecast skills are comparable, if hindcasts and observations are used to calibrate parameters for the traditional weighting methods. However, the fact that the Bayesian method requires no hindcast information for calibration or use is a significant advantage of the approach. This example also compared parametric and nonparametric approaches for estimating the likelihood function. Although a simple parametric approach was suitable for this case, a more complex nonparametric approach, which may be necessary in other applications, performed as well or better.

After the selection of an appropriate climate index for conditioning, the Bayesian climate index weighting method is easy to apply, is self-calibrating using the data from the ensemble forecast, and is self-adjusting to the strength of the relationship between ensemble members (realizations of the forecast variable) and the climate index; if there is no relationship between the forecast variable and the climate index (the two are statistically independent), the method defaults to equal weighting of the ensemble members. As a result, the method can quickly be adapted to an existing ensemble forecasting system, or continue to be utilized after the forecasting system is updated or changed, as no hindcasts are required for parameter calibration. The method can also be applied directly to forecast model-simulated variables, which often are biased predictions. The resulting posterior ensemble model-simulated distribution can be adjusted with available bias-correction techniques to yield a reliable ensemble forecast of the forecast variable.

Acknowledgments
This work was supported in part by the National Oceanic and Atmospheric Administration (NOAA) Climate Program under grant NA09OAR4310196, as part of the Climate Prediction Program for the Americas (CPPA). We gratefully acknowledge this support. We would also like to thank the Associate Editor and the two anonymous reviewers for their thoughtful comments and suggestions. The ensemble streamflow forecasts and climate index records used in this study are available upon request from the first author (allen-bradley@uiowa.edu).

References

Araghinejad, S., D. H. Burn, and M. Karamouz (2006), Long-lead probabilistic forecasting of streamflow using ocean-atmospheric and hydrological predictors, *Water Resour. Res.*, *42*, W03431, doi:10.1029/2004WR003853.
Aziz, O. A., G. A. Tootle, S. T. Gray, and T. C. Piechota (2010), Identification of Pacific Ocean sea surface temperature influences of upper Colorado River basin snowpack, *Water Resour. Res.*, *46*, W07536, doi:10.1029/2009WR008053.
Bennett, J. C., Q. J. Wang, P. Pokhrel, and D. E. Robertson (2014), The challenge of forecasting high streamflows 1–3 months in advance with lagged climate indices in southeast Australia, *Nat. Hazards Earth Syst. Sci.*, *14*(2), 219–233, doi:10.5194/nhess-14-219-2014.
Block, P., and B. Rajagopalan (2007), Interannual variability and ensemble forecast of upper Blue Nile basin Kiremt season precipitation, *J. Hydrometeor.*, *8*(3), 327–343.
Block, P., and B. Rajagopalan (2009), Statistical-dynamical approach for streamflow modeling at Malakal, Sudan, on the White Nile River, *J. Hydrol. Eng.*, *14*(2), 185–196.
Bogner, K., and M. Kalas (2008), Error-correction methods and evaluation of an ensemble based hydrological forecasting system for the upper Danube catchment, *Atmos. Sci. Lett.*, *9*(2), 95–102, doi:10.1002/asl.180.
Bradley, A. A., and S. S. Schwartz (2011), Summary verification measures and their interpretation for ensemble forecasts, *Mon. Weather Rev.*, *139*(9), 3075–3089.
Bradley, A. A., S. S. Schwartz, and T. Hashino (2004), Distributions-oriented verification of ensemble streamflow predictions, *J. Hydrometeorol.*, *5*(3), 532–545.
Brier, G. W. (1950), Verification of forecasts expressed in terms of probability, *Mon. Weather Rev.*, *78*(1), 1–3.
Brown, J. D., and D. J. Seo (2010), A nonparametric postprocessor for bias correction of hydrometeorological and hydrologic ensemble forecasts, *J. Hydrometeorol.*, *11*(3), 642–665.
Brown, J. D., and D. J. Seo (2013), Evaluation of a nonparametric post-processor for bias correction and uncertainty estimation of hydrologic predictions, *Hydrol. Processes*, *27*(1), 83–105, doi:10.1002/hyp.9263.
Camberlin, P. (1997), Rainfall anomalies in the source region of the Nile and their connection with the Indian summer monsoon, *J. Clim.*, *10*(6), 1380–1392.
Chiew, F. H. S., and T. A. McMahon (2002), Global ENSO-streamflow teleconnection, streamflow forecasting and interannual variability, *Hydrol. Sci. J.*, *47*(3), 505–522.
Chiew, F. H. S., T. C. Piechota, J. A. Dracup, and T. A. McMahon (1998), El Niño Southern Oscillation and Australian rainfall, streamflow and drought: Links and potential for forecasting, *J. Hydrol.*, *204*(1–4), 138–149.
Clark, M. P., and L. E. Hay (2004), Use of medium-range numerical weather prediction model output to produce forecasts of streamflow, *J. Hydrometeorol.*, *5*(1), 15–32.
Cleveland, W. S. (1979), Robust locally weighted regression and smoothing scatterplots, *J. Am. Stat. Assoc.*, *74*(368), 829–836.
Cloke, H. L., and F. Pappenberger (2009), Ensemble flood forecasting: A review, *J. Hydrol.*, *375*(3–4), 613–626.
Conway, D. (2000), The climate and hydrology of the upper Blue Nile River, *Geogr. J.*, *166*, 49–62.
Croley, T. E., II (1996), Using NOAA's new climate outlooks in operational hydrology, *J. Hydrol. Eng.*, *1*(3), 93–102.
Croley, T. E., II (1997), Mixing probabilistic meteorology outlooks in operational hydrology, *J. Hydrol. Eng.*, *2*(4), 161–168.
Croley, T. E., II (2000), *Using Meteorology Probability Forecasts in Operational Hydrology*, Am. Soc. Civ. Eng. Press, Reston, Va.
Day, G. N. (1985), Extended streamflow forecasting using NWS-RFS, *J. Water Resour. Plann. Manage.*, *111*(2), 157–170.
Efron, B., and R. J. Tibshirani (1993), *An Introduction to the Bootstrap*, Chapman and Hall, N. Y.
El Adlouni, S., T. B. M. J. Ouarda, X. Zhang, R. Roy, and B. Bobée (2007), Generalized maximum likelihood estimators for the nonstationary generalized extreme value model, *Water Resour. Res.*, *43*, W03410, doi:10.1029/2005WR004545.
Elshamy, M. E. (2008), Assessing the hydrological performance of the Nile Forecast System in long term simulations, *Nile Basin Water Eng. Sci. Mag.*, *1*(3), 22–40.
Eltahir, E. A. B. (1996), El Niño and the natural variability in the flow of the Nile River, *Water Resour. Res.*, *32*(1), 131–137.

Faber, B. A., and J. R. Stedinger (2001), Reservoir optimization using sampling SDP with ensemble streamflow prediction (ESP) forecasts, *J. Hydrol.*, *249*(1–4), 113–133.

Georgakakos, K. P., and R. Krzysztofowicz (2001), Probabilistic and ensemble forecasting, *J. Hydrol.*, *249*(1–4), 1–1.

Gobena, A. K., and T. Y. Gan (2009), Statistical ensemble seasonal streamflow forecasting in the south Saskatchewan River basin by a modified nearest neighbors resampling, *J. Hydrol. Eng.*, *14*(6), 628–639.

Gobena, A. K., and T. Y. Gan (2010), Incorporation of seasonal climate forecasts in the ensemble streamflow prediction system, *J. Hydrol.*, *385*(1–4), 336–352.

Golembesky, K., A. Sankarasubramanian, and N. Devineni (2009), Improved drought management of Falls Lake Reservoir: Role of multimodel streamflow forecasts in setting up restrictions, *J. Water Resour. Plann. Manage.*, *135*(3), 188–197.

Grantz, K., B. Rajagopalan, M. Clark, and E. Zagona (2005), A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts, *Water Resour. Res.*, *41*, W10410, doi:10.1029/2004WR003467.

Hamlet, A. F., and D. P. Lettenmaier (1999), Columbia River streamflow forecasting based on ENSO and PDO climate signals, *J. Water Resour. Plann. Manage.*, *125*(6), 333–341.

Hashino, T., A. A. Bradley, and S. S. Schwartz (2007), Evaluation of bias-correction methods for ensemble streamflow volume forecasts, *Hydrol. Earth Syst. Sci.*, *11*(2), 939–950.

Hay, L. E., G. J. McCabe, M. P. Clark, and J. C. Risley (2009), Reducing streamflow forecast uncertainty: Application and qualitative assessment of the upper Klamath River basin, Oregon, *J. Am. Water Resour. Assoc.*, *45*(3), 580–596.

Herr, H. D., and R. Krzysztofowicz (2015), Ensemble Bayesian forecasting system. Part i: Theory and algorithms, *J. Hydrol.*, *524*(0), 789–802, doi:10.1016/j.jhydrol.2014.11.072.

Kahya, E., and J. A. Dracup (1993), United-States streamflow patterns in relation to the El-Niño Southern Oscillation, *Water Resour. Res.*, *29*(8), 2491–2503.

Kalra, A., and S. Ahmad (2009), Using oceanic-atmospheric oscillations for long lead time streamflow forecasting, *Water Resour. Res.*, *45*, W03413, doi:10.1029/2008WR006855.

Kalra, A., W. P. Miller, K. W. Lamb, S. Ahmad, and T. Piechota (2013), Using large-scale climatic patterns for improving long lead time streamflow forecasts for Gunnison and San Juan River basins, *Hydrol. Processes*, *27*(11), 1543–1559.

Karamouz, M., and B. Zahraie (2004), Seasonal streamflow forecasting using snow budget and El Niño-Southern Oscillation climate signals: Application to the Salt River basin in Arizona, *J. Hydrol. Eng.*, *9*(6), 523–533.

Kelman, J., J. R. Stedinger, L. A. Cooper, E. Hsu, and S. Q. Yuan (1990), Sampling stochastic dynamic-programming applied to reservoir operation, *Water Resour. Res.*, *26*(3), 447–454.

Kennedy, A. M., D. C. Garen, and R. W. Koch (2009), The association between climate teleconnection indices and upper Klamath seasonal streamflow: Trans-Nino index, *Hydrol. Processes*, *23*(7), 973–984.

Krzysztofowicz, R. (1999), Bayesian theory of probabilistic forecasting via deterministic hydrologic model, *Water Resour. Res.*, *35*(9), 2739–2750, doi:10.1029/1999WR900099.

Kwon, H. H., C. Brown, and U. Lall (2008), Climate informed flood frequency analysis and prediction in Montana using hierarchical Bayesian modeling, *Geophys. Res. Lett.*, *35*, L05404, doi:10.1029/2007GL032220.

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, *Bull. Am. Meteorol. Soc.*, *80*(11), 2313–2329.

Lima, L. M. M., E. Popova, and P. Damien (2014), Modeling and forecasting of Brazilian reservoir inflows via dynamic linear models, *Int. J. Forecasting*, *30*(3), 464–476, doi:10.1016/j.ijforecast.2013.12.009.

Lu, A., S. Jia, V. Zhu, H. Yan, S. Duan, and Z. Yao (2011), El Niño-Southern Oscillation and water resources in the headwaters region of the Yellow River: Links and potential for forecasting, *Hydrol. Earth Syst. Sci.*, *15*(4), 1273–1281.

Luo, L. F., and E. F. Wood (2008), Use of Bayesian merging techniques in a multimodel seasonal hydrologic ensemble prediction system for the eastern United States, *J. Hydrometeorol.*, *9*(5), 866–884.

Moradkhani, H., and M. Meier (2010), Long-lead water supply forecast using large-scale climate predictors and independent component analysis, *J. Hydrol. Eng.*, *15*(10), 744–762.

Najafi, M. R., H. Moradkhani, and T. C. Piechota (2012), Ensemble streamflow prediction: Climate signal weighting methods vs. climate forecast system reanalysis, *J. Hydrol.*, *442*, 105–116.

Nile Forecast Center (2007), The NFS Hydrologic Model Version 5.1 Manual, Ministry of Water Resources and Irrigation, Cairo, Egypt.

Ouarda, T., and S. El-Adlouni (2011), Bayesian nonstationary frequency analysis of hydrological variables, *J. Am. Water Resour. Assoc.*, *47*(3), 496–505, doi:10.1111/j.1752-1688.2011.00544.x.

Pagano, T. C., D. L. Shrestha, Q. J. Wang, D. Robertson, and P. Hapuarachchi (2013), Ensemble dressing for hydrological applications, *Hydrol. Processes*, *27*(1), 106–116.

Piechota, T. C., and J. A. Dracup (1996), Drought and regional hydrologic variation in the United States: Associations with the El Niño Southern Oscillation, *Water Resour. Res.*, *32*(5), 1359–1373.

Piechota, T. C., and J. A. Dracup (1999), Long-range streamflow forecasting using El Niño-Southern Oscillation indicators, *J. Hydrol. Eng.*, *4*(2), 144–151.

Piechota, T. C., J. A. Dracup, and R. G. Fovell (1997), Western US streamflow and atmospheric circulation patterns during El Niño Southern Oscillation, *J. Hydrol.*, *201*(1–4), 249–271.

Piechota, T. C., F. H. S. Chiew, J. A. Dracup, and T. A. McMahon (1998), Seasonal streamflow forecasting in eastern Australia and the El Niño Southern Oscillation, *Water Resour. Res.*, *34*(11), 3035–3044.

Pokhrel, P., D. E. Robertson, and Q. J. Wang (2013), A Bayesian joint probability post-processor for reducing errors and quantifying uncertainty in monthly streamflow predictions, *Hydrol. Earth Syst. Sci.*, *17*(2), 795–804, doi:10.5194/hess-17-795-2013.

Redmond, K. T., and R. W. Koch (1991), Surface climate and streamflow variability in the western United-States and their relationship to large-scale circulation indexes, *Water Resour. Res.*, *27*(9), 2381–2399.

Regonda, S. K., B. Rajagopalan, and M. Clark (2006a), A new method to produce categorical streamflow forecasts, *Water Resour. Res.*, *42*, W09501, doi:10.1029/2006WR004984.

Regonda, S. K., B. Rajagopalan, M. Clark, and E. Zagona (2006b), A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River basin, *Water Resour. Res.*, *42*, W09404, doi:10.1029/2005WR004653.

Robertson, D. E., and Q. J. Wang (2013), Seasonal forecasts of unregulated inflows into the Murray River, Australia, *Water Resour. Manage.*, *27*(8), 2747–2769, doi:10.1007/s11269-013-0313-4.

Schaake, J. C., T. M. Hamill, R. Buizza, and M. Clark (2007), The hydrological ensemble prediction experiment, *Bull. Am. Meteorol. Soc.*, *88*(10), 1541–1547.

Segele, Z. T., and P. J. Lamb (2005), Characterization and variability of Kiremt rainy season over Ethiopia, *Meteorol. Atmos. Phys*., *89*(1–4), 153–180.

Seleshi, Y., and G. R. Demaree (1995), Rainfall variability in the Ethiopian and Eritrean Highlands and its links with the Southern Oscillation Index, *J. Biogeogr*., *22*(4–5), 945–952.

Seo, D. J., H. D. Herr, and J. C. Schaake (2006), A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble stream prediction, *Hydrol. Earth Syst. Sci. Discuss*., *3*, 1987–2035.

Sheather, S. J., and M. C. Jones (1991), A reliable data-based bandwidth selection method for kernel density-estimation, *J. R. Stat. Soc., Ser. B*, *53*(3), 683–690.

Smith, A. F. M., and A. E. Gelfand (1992), Bayesian statistics without tears—A sampling resampling perspective, *Am. Stat*., *46*(2), 84–88.

Smith, J. A., G. N. Day, and M. D. Kane (1992), Nonparametric framework for long-range streamflow forecasting, *J. Water Resour. Plann. Manage*., *118*(1), 82–92.

Souza Filho, F. A., and U. Lall (2003), Seasonal to interannual ensemble streamflow forecasts for Ceara, Brazil: Applications of a multivariate, semiparametric algorithm, *Water Resour. Res*., *39*(11), 1307, doi:10.1029/2002WR001373.

Stedinger, J. R., and Y. O. Kim (2010), Probabilities for ensemble forecasts reflecting climate information, *J. Hydrol*., *391*(1–2), 11–25.

Steinschneider, S., and C. Brown (2012), Forecast-informed low-flow frequency analysis in a Bayesian framework for the northeastern United States, *Water Resour. Res*., *48*, W10545, doi:10.1029/2012WR011860.

Thirel, G., F. Rousset-Regimbeau, E. Martin, and F. Habets (2008), On the impact of short-range meteorological forecasts for ensemble streamflow predictions, *J. Hydrometeorol*., *9*(6), 1301–1317.

Timilsena, J., T. Piechota, G. Tootle, and A. Singh (2009), Associations of interdecadal/interannual climate variability and long-term Colorado River basin streamflow, *J. Hydrol*., *365*(3–4), 289–301.

Tootle, G. A., and T. C. Piechota (2004), Suwannee River long range streamflow forecasts based on seasonal climate predictors, *J. Am. Water Resour. Assoc*., *40*(2), 523–532.

Tootle, G. A., and T. C. Piechota (2006), Relationships between Pacific and Atlantic ocean sea surface temperatures and US streamflow variability, *Water Resour. Res*., *42*, W07411, doi:10.1029/2005WR004184.

Tootle, G. A., T. C. Piechota, and A. Singh (2005), Coupled oceanic-atmospheric variability and U.S. streamflow, *Water Resour. Res*., *41*, W12408, doi:10.1029/2005WR004381.

Tootle, G. A., T. C. Piechota, and F. Gutierrez (2008), The relationships between Pacific and Atlantic ocean sea surface temperatures and Colombian streamflow variability, *J. Hydrol*., *349*(3–4), 268–276.

Trenberth, K. E. (1984), Signal versus noise in the Southern Oscillation, *Mon. Weather Rev*., *112*(2), 326–332.

Troup, A. J. (1965), Southern Oscillation, *Q. J. R. Meteorol. Soc*., *91*(390), 490–506.

Twedt, T. M., J. C. Schaake, and E. L. Peck (1977), National Weather Service extended streamflow prediction, in *Proceedings, Western Snow Conference, 45th Annual Meeting, Albuquerque*, pp. 52–57.

Walker, G. T., and E. W. Bliss (1932), World weather V, *Mem. R. Meteorol. Soc*., *4*(36), 53–84.

Wang, E. L., Y. Q. Zhang, J. M. Luo, F. H. S. Chiew, and Q. J. Wang (2011), Monthly and seasonal streamflow forecasts using rainfall-runoff modeling and historical weather data, *Water Resour. Res*., *47*, W05516, doi:10.1029/2010WR009922.

Wang, Q. J., D. E. Robertson, and F. H. S. Chiew (2009), A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites, *Water Resour. Res*., *45*, W05407, doi:10.1029/2008WR007355.

Wei, W. E., and D. W. Watkins (2011a), Probabilistic streamflow forecasts based on hydrologic persistence and large-scale climate signals in central Texas, *J. Hydroinformatics*, *13*(4), 760–774.

Wei, W. W., and D. W. Watkins (2011b), Data mining methods for hydroclimatic forecasting, *Adv. Water Resour*., *34*(11), 1390–1400.

Werner, K., D. Brandon, M. Clark, and S. Gangopadhyay (2004), Climate index weighting schemes for NWS ESP-based seasonal volume forecasts, *J. Hydrometeorol*., *5*(6), 1076–1090.

Wilks, D. (2011), *Statistical Methods in the Atmospheric Sciences*, 3rd ed., 704 pp., Academic Press, Waltham, Mass.

Wood, A. W., and D. P. Lettenmaier (2006), A test bed for new seasonal hydrologic forecasting approaches in the western United States, *Bull. Am. Meteorol. Soc*., *87*(12), 1699–1712.

Wood, A. W., E. P. Maurer, A. Kumar, and D. P. Lettenmaier (2002), Long-range experimental hydrologic forecasting for the eastern United States, *J. Geophys. Res*., *107*(D20), 4429, doi:10.1029/2001JD000659.

Wood, A. W., A. Kumar, and D. P. Lettenmaier (2005), A retrospective assessment of National Centers for Environmental Prediction climate model-based ensemble hydrologic forecasting in the western United States, *J. Geophys. Res*., *110*, D04105, doi:10.1029/2004JD004508.

Zhao, L., Q. Duan, J. Schaake, A. Ye, and J. Xia (2011), A hydrologic post-processor for ensemble streamflow predictions, *Adv. Geosci*., *29*, 51–59, doi:10.5194/adgeo-29-51-2011.