Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences

Q. J. Wang[1] and D. E. Robertson[1]

Received 17 March 2010; revised 7 December 2010; accepted 29 December 2010; published 24 February 2011.

[1] Skillful and reliable forecasts of seasonal streamflows are highly valuable to water management. In a previous study, we developed a Bayesian joint probability (BJP) modeling approach for seasonal forecasting of streamflows at multiple sites. The approach has been adopted by the Australian Bureau of Meteorology for seasonal streamflow forecasting in Australia. This study extends the applicability of the BJP modeling approach to streams with zero flow occurrences. The aim is to produce forecasts for these streams in the form of probabilities of zero seasonal flows and probability distributions of above zero seasonal flows. We turn a difficult mathematical problem of mixed discrete-continuous multivariate probability distribution modeling into one of continuous multivariate probability distribution modeling by treating zero flow occurrences as censored data. This paper presents the mathematical formulation and implementation of the modeling approach, methods for forecast verification, and results of a test application to the Burdekin river catchment in northern Queensland, Australia.

Citation: Wang, Q. J. and D. E. Robertson (2011), Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences, *Water Resour. Res., 47*, W02546, doi:10.1029/2010WR009333.

1. Introduction

[2] Forecasts of future seasonal streamflows are potentially valuable to a range of water managers and users, including irrigators, urban and rural water supply authorities, environmental managers, and hydroelectricity generators. Such forecasts can inform planning and management decisions to maximize returns on investments and available water resources and to ensure security of supply [e.g., *Chiew et al.*, 2003; *Plummer et al.*, 2009]. For many river and storage systems, a joint forecast of streamflows at multiple sites that accounts for intersite correlations is needed for managing water resources at a system scale. *Regonda et al.* [2006], *Westra et al.* [2007], and *Bracken et al.* [2010] provide some of the methods for joint forecasting of streamflows at multiple sites.

[3] In a previous paper, we developed a Bayesian joint probability (BJP) modeling approach for seasonal forecasting of streamflows at multiple sites [*Wang et al.*, 2009]. Specifically, the BJP approach provides probabilistic forecasts of streamflow volume totals over a forecast period, say the next 3 months, at multiple sites in the form of ensembles. The approach uses a Yeo-Johnson transformed multivariate normal distribution to model the joint distribution of future streamflows and their predictors such as antecedent streamflows, El Niño–Southern Oscillation indices and other climate indicators. The model parameters and their uncertainties are inferred from historical data using a Bayesian method. The parameters are then used for producing joint probabilistic forecasts of streamflows at multiple sites for future events. The BJP modeling approach is completed with a method for selecting predictors from a large number of candidate predictors based on pseudo Bayes factors calculated from cross-validation predictive densities (D.E. Robertson and Q. J. Wang, A Bayesian approach to predictor selection for seasonal streamflow forecasting, submitted to *Journal of Hydrometeorology*, 2010), and with a suite of methods and tools for verification of probabilistic forecasts [*Wang et al.*, 2009].

[4] The BJP modeling approach has been adopted by the Australian Bureau of Meteorology for seasonal streamflow forecasting in Australia [*Plummer et al.*, 2009]. One of the limitations of the BJP modeling approach as presented by *Wang et al.* [2009] is that it does not deal with zero flows, which occur on many streams in Australia (and in arid and semiarid regions elsewhere in the world). Zero flows may occur in dry seasons occasionally on perennial streams and frequently on intermittent streams. Zero flows can be the dominant state of many ephemeral streams. This study extends the BJP modeling approach so that it is applicable to seasonal flow forecasting for streams with zero value occurrences. The aim is to produce forecasts that give probabilities of zero flows and probability distributions of above zero flows.

[5] Antecedent streamflows are often useful for representing the initial catchment conditions and thus for serving as predictors of future streamflows. Zero values may occur with either antecedent streamflows or future streamflows. In a multisite problem, zero flows may occur in any of the streams being considered. Thus, in the BJP modeling setting, zero flows may occur in any combination of predictors and predictands and of sites. Mathematically, this leads to a mixed discrete-continuous multivariate probability distribution, which is extremely difficult to formulate and manipulate.

[1]CSIRO Land and Water, Highett, Victoria, Australia.

0043-1397/11/2010WR009333

[6] Methods for stochastic generation and for downscaling of daily rainfall at multiple sites provide useful reference for modeling mixed discrete-continuous multivariate probability distributions. The most commonly used approach is to model rainfall occurrence and amounts separately [e.g., *Wilks*, 1998; *Charles et al.*, 1999; *Srikanthan and Pegram*, 2009]. *Bardossy and Plate* [1992] introduced an approach that directly transforms the discrete-continuous multivariate probability distributions to a continuous multivariate normal distribution. The transformation includes an allocation of the cumulative probabilities in the negative space of the transformed variables to the zero value points of the original rainfall variables. The approach does not require modeling rainfall occurrence and amounts separately and thus significantly simplifies the problem. A similar approach was used by *Frost et al.* [2007] in a context of stochastic generation of annual multisite hydrological data using a multivariate autoregressive lag-1 model. The approach uses a continuous multivariate distribution as the underlying distribution and then lumps the probability mass in the subspace below the zero thresholds of the variables to the zero value point of the variables. The lumping is done through a numerical integration.

[7] In this study we adopt essentially the same approach as *Bardossy and Plate* [1992] and *Frost et al.* [2007] as it is well suited to the extension of our original BJP modeling approach. However, we cast the problem to one of censored data. We treat the zero flows as censored data having unknown precise values but known to be below or equal to zero. In this way, the variables both before and after the Yeo-Johnson transform (see section 2) are considered to follow continuous multivariate distributions. The treatment allows us to greatly simplify the mathematical expressions required, and importantly, to deal with zero predictor values as well as zero predictand values. In the work of *Frost et al.* [2007], the zero threshold problem was considered only for variables at the current time step (equivalent to predictands), not for the variables at the lag-1 time step (equivalent to predictors) on which the variables at the current time step are conditioned.

[8] This paper presents the extended BJP modeling approach applicable to streams with zero value occurrences and demonstrates its use through a test application. In addition, improvements made to a number of detailed techniques in the work of *Wang et al.* [2009], including reparameterization, prior specification, and Markov chain Monte Carlo (MCMC) sampling, are also presented. A new skill score based on the root mean square error in probability (RMSEP) is introduced as an alternative to the linear error in probability space (LEPS). The paper is organized as follows. Model formulation is given in the next section. Model parameter inference in section 3 includes reparameterization, derivation of posterior distribution of parameters, specification of prior distribution of parameters, and MCMC sampling of the posterior distribution of parameters. Section 4 details the method for using the model to produce probabilistic forecasting, including the augmentation of censored predictor data. Section 5 includes a number of statistical measures and graphical methods for both overall and detailed verifications. Section 6 deals with model checking. A test application to forecasting streamflows at three river gauges in the Burdekin River catchment in northern Queensland, Australia, is given in section 7 to demonstrate the working of the extended BJP modeling approach. Section 8 completes the paper with a summary and conclusions. The paper also has three appendices to provide additional technical details.

2. Model Formulation

[9] Consider d variables that consist of both seasonal streamflows to be forecast at multiple sites and their predictors such as climate and catchment indicators

$$\mathbf{y}^{\mathrm{T}} = [\, y_1 \quad y_2 \quad \cdots \quad y_d \,], \tag{1}$$

where $\mathbf{y}$ is a column vector and $\mathbf{y}^{\mathrm{T}}$ is its transpose.

[10] As stated in section 1, streamflows may appear in $\mathbf{y}$ as predictors as well as predictands. Streamflows can sometimes have zero values, and this can lead to a mixed discrete-continuous joint probability distribution for $\mathbf{y}$ with various possible combinations of zero and nonzero values in terms of predictor and predictand streamflows and in terms of streamflows at different sites.

[11] Mathematically, multivariate mixed discrete-continuous distributions are extremely difficult to formulate and manipulate. We turn this problem into one of multivariate continuous distributions by treating the occurrences of zero flows as censored data. When an observed flow y_i is 0, we assign $y_i \leq 0$. In other words, we hypothetically allow the flow to be negative (and thus continuous), but its precise value is unknown. For future events, on the other hand, flows that are forecast to be negative are converted back to zero flows. This treatment allows the use of a continuous multivariate distribution for $\mathbf{y}$ and thus significantly simplifies the original problem.

[12] The vector $\mathbf{y}$ of predictor and predictand variables is normalized to $\mathbf{z}$ by applying the Yeo-Johnson transform, an extended form of the Box-Cox transform [*Yeo and Johnson*, 2000],

$$z_i = \mathrm{YJ}(y_i, \lambda_i)\,, \tag{2}$$

where $i = 1, 2, \ldots, d$ and λ is the Yeo-Johnson transform coefficient. The vector $\mathbf{z}$ is assumed to follow a multivariate normal distribution

$$\mathbf{z} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}\,\mathbf{R}\,\boldsymbol{\sigma}^{\mathrm{T}})\,, \tag{3}$$

where $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, and $\mathbf{R}$ are the mean vector, the standard deviation vector, and correlation coefficient matrix, respectively. Assuming a steady relationship among the variables, the model has a total of $3d + (d-1)d/2$ unknown parameters. For more details on the model formulation, readers are referred to *Wang et al.* [2009].

3. Model Parameter Inference

3.1. Reparameterization

[13] Before setting out for statistical inference of the unknown parameters, a number of parameters are reparameterized to ease the MCMC sampling [*Thyer et al.*, 2002; *Gelman et al.*, 2004]. Parameters $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are reparameterized to $\mathbf{m}$ and $\mathbf{s}$ by [*Wang et al.*, 2009]

$$\mu_i = \mathrm{YJ}(m_i, \lambda_i)\,, \tag{4}$$

$$\sigma_i = \begin{cases} (m_i+1)^{\lambda_i - 1} s_i & (m_i \geq 0) \\ (-m_i+1)^{1-\lambda_i} s_i & (m_i < 0) \end{cases}, \tag{5}$$

where $i = 1, 2, \ldots, d$, m is an approximation of the mean of a nontransformed variable y, and s is an approximation of the standard deviation of y [*Yeo and Johnson*, 2000]. In the work of *Wang et al.* [2009], the final parameters used were m and s^2. In this study s^2 is further parameterized to 2log s. The parameter 2log s is found to be less nonlinearly related to m than s^2, leading to more efficient MCMC sampling of the posterior distribution of the parameters using a multivariate normal proposal distribution (see section 3.4). The correlation coefficient matrix $\mathbf{R}$ is reparameterized to $\mathbf{\Phi}$ through an inverse hyperbolic tangent transform or Fisher Z-transform [*Wang et al.*, 2009].

3.2. Posterior Distribution of Parameters

[14] The $3d + (d-1)d/2$ unknown model parameters, denoted hereafter as $\boldsymbol{\theta}$, need to be inferred before the model can be used for forecasting. According to the Bayes theorem, given historically observed events $\mathbf{y}^t$ for year $t = 1,2,\ldots, n$, the posterior distribution of the model parameters is

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{y}^n, \mathbf{y}^{n-1}, \ldots, \mathbf{y}^1) &\propto p(\boldsymbol{\theta}) p(\mathbf{y}^n, \mathbf{y}^{n-1}, \ldots, \mathbf{y}^1|\boldsymbol{\theta}) \\ &= p(\boldsymbol{\theta}) \textstyle\prod_{t=1}^{n} p(\mathbf{y}^t|\boldsymbol{\theta})\,, \end{aligned} \tag{6}$$

where $p(\boldsymbol{\theta})$ is a prior distribution representing any information available about the parameters before the use of the historical data, $p(\mathbf{y}^n, \mathbf{y}^{n-1}, \ldots, \mathbf{y}^1|\boldsymbol{\theta})$ is the likelihood function defining the probability of observing the historical events $\mathbf{y}^t$, $t = 1,2,\ldots, n$, given the model and its parameter set.

[15] The formulation for the likelihood function $p(\mathbf{y}^t|\boldsymbol{\theta})$ for event t with complete data is straightforward and given by *Wang et al.* [2009]. The formulation for the likelihood function $p(\mathbf{y}^t|\boldsymbol{\theta})$ for an event with censored data is more complicated and requires a few steps. We drop the superscript t in the rest of this subsection for simplicity. We also allow general censoring thresholds $\mathbf{y}_\mathbf{c}$ rather than just zero for wider applications.

[16] First, we rearrange the $\mathbf{y}$ vector into two subvectors

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}(a) \\ \mathbf{y}(b) \end{bmatrix}, \tag{7}$$

where $\mathbf{y}(a)$ consists of variables whose values are above their respective censor thresholds $\mathbf{y}_\mathbf{c}(a)$ and precisely known, and $\mathbf{y}(b)$ consists of variables whose values are only known to be equal to or below their respective censor thresholds $\mathbf{y}_\mathbf{c}(b)$. Accordingly, the variable vector after the Yeo-Johnson transform is organized as

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}(a) \\ \mathbf{z}(b) \end{bmatrix}. \tag{8}$$

[17] The likelihood function $p(\mathbf{y}|\boldsymbol{\theta})$ is then given by

$$\begin{aligned} p(\mathbf{y}|\boldsymbol{\theta}) &= p(\mathbf{y}(a), \mathbf{y}(b) \leq \mathbf{y}_\mathbf{c}(b)|\boldsymbol{\theta}) \\ &= p(\mathbf{y}(a)|\boldsymbol{\theta}) \times p(\mathbf{y}(b) \leq \mathbf{y}_\mathbf{c}(b)|\mathbf{y}(a), \boldsymbol{\theta})\,, \end{aligned} \tag{9}$$

where

$$p(\mathbf{y}(a)|\boldsymbol{\theta}) = J_{\mathbf{z}(a)\to\mathbf{y}(a)} p(\mathbf{z}(a)|\boldsymbol{\theta}) = \textstyle\prod_{i=1}^{d_a} (dz_i/dy_i) p(\mathbf{z}(a)|\boldsymbol{\theta}) \tag{10}$$

$$\begin{aligned} p(\mathbf{y}(b) \leq \mathbf{y}_\mathbf{c}(b)|\mathbf{y}(a), \boldsymbol{\theta}) &= p(\mathbf{z}(b) \leq \mathbf{z}_\mathbf{c}(b)|\mathbf{z}(a), \boldsymbol{\theta}) \\ &= \int_{-\infty}^{\mathbf{z}_\mathbf{c}(b)} p(\mathbf{z}(b)|\mathbf{z}(a), \boldsymbol{\theta}) d\mathbf{z}(b). \end{aligned} \tag{11}$$

[18] In equation (10), $J_{\mathbf{z}(a)\to\mathbf{y}(a)}$ is the Jacobian determinant of the transform from $\mathbf{z}(a)$ to $\mathbf{y}(a)$, d_a is the dimension of $\mathbf{y}(a)$, and the derivative dz_i/dy_i can be found in the work of *Wang et al.* [2009]. In equation (11), the thresholds $\mathbf{z}_\mathbf{c}(b)$ correspond to the Yeo-Johnson transformed values of $\boldsymbol{y}_\mathbf{c}(b)$. The conditional probability distribution $p(\mathbf{z}(b)|\mathbf{z}(a), \boldsymbol{\theta})$ is multivariate normal, and its mean vector and covariance matrix can be found by applying equations (A4) and (A5) in Appendix A. The integration of the multivariate normal distribution requires numerical solution. The RANNRM [*Genz*, 1993] algorithm is used in this study.

3.3. Prior Distribution of Parameters

[19] The prior distribution for the various parameters in the model is specified as

$$p(\boldsymbol{\theta}) = \textstyle\prod_{i=1}^{d} p(\lambda_i) p(m_i,\, 2\log s_i) p(\mathbf{\Phi}). \tag{12}$$

[20] A uniform prior with a range of $[-0.5, 1.2]$ is used for each λ. A more elaborate derivation of prior for each pair of (m, 2log s) is made to deal with the effect of the Yeo-Johnson transform and reparameterization, giving

$$p(m,\, 2\log s) = J_{\mu,\sigma^2\to m,s^2} J_{s^2\to 2\log s} p(\mu, \sigma^2)\,, \tag{13}$$

where $J_{\mu,\sigma^2\to m,s^2}$ is the Jacobian determinant of the transform from (μ, σ^2) to (m, s^2) and can be found in the work of *Wang et al.* [2009]; $J_{s^2\to 2\log s}$ is the Jacobian determinant of the transform from s^2 to 2log s,

$$J_{s^2\to 2\log s} = d(s^2)/d(2\log s) = s^2; \tag{14}$$

and $p(\mu, \sigma^2)$ takes on the simplest form of priors commonly used for normal distribution mean and variance [*Gelman et al.*, 2004, p. 74],

$$p(\mu, \sigma^2) \propto 1/\sigma^2. \tag{15}$$

[21] The simple prior for $p(\mu, \sigma^2)$ as in equation (15) replaces the normal-inverse-chi-square prior previously used in the work of *Wang et al.* [2009, equations (23) and (24)]. The simple prior is parameter free. In contrast, the normal-inverse-chi-square prior consists of a number of parameters, which are difficult to empirically specify from samples with censored data. The prior for the reparameterized correlation coefficient matrix $\mathbf{\Phi}$ is as specified in the work of *Wang et al.* [2009].

3.4. Sampling of Parameters

[22] The technique of Metropolis MCMC sampling is used to draw a sample, say 1000 sets, of parameter values

that numerically represent the joint posterior distribution of the parameters. In this subsection, we describe the changes made to the Metropolis MCMC sampling procedure used in our previous study [*Wang et al.*, 2009].

[23] The Metropolis MCMC sampling requires initial parameter values to start off sampling. In our previous study, we used moment estimates of m and s^2 and 0.2 for λ as initial values for the MCMC sampling. However, when data are censored, moment estimates of m and 2 log s are not readily available. In this study, single series maximum likelihood values of m, 2log s, and λ are used as initial values for the MCMC sampling. In our previous study, we used the nearest correlation matrix that is positive semidefinite as the initial correlation matrix for the MCMC sampling. Our experience since then is that the nearest correlation matrix method generally works well but very occasionally strays into a parameter space (with one or more correlation coefficients close to 1) that is difficult for the MCMC sampling to jump out from. Although the shrinkage method [*Devlin et al.*, 1975; *Jobson*, 1992, p. 156] is an attractive alternative, we simply opt to use a unit matrix as the initial correlation matrix, which in our experience, is highly robust and results in little loss in computational efficiency.

[24] The Metropolis MCMC sampling also requires a proposal distribution for generating random jumps in the parameter space [*Gelman et al.*, 2004]. As in our previous study, a multivariate normal distribution is adopted as the initial proposal distribution. However, the posterior parameter distribution for problems with zero flows is typically more difficult to sample from, and therefore, greater care is needed in constructing the initial proposal distribution covariance matrix. In this study, the initial matrix is made of two parts of nonzero entries. The first part is the within series variances and covariances of m, 2log s, and λ, estimated from a sample of 30,000 sets of parameters numerically representing the posterior parameter distribution of the single series Yeo-Johnson transformed normal distribution. This sample of parameters is obtained by undertaking a short MCMC sampling initialized using the maximum likelihood parameters as the initial values and a crude diagonal covariance matrix estimate. The second part is the variances of the reparameterized correlation coefficients in $\mathbf{\Phi}$, estimated as $1/(n-3)$ from a stabilized variance approximation [*Fisher*, 1921; *Zhu and Hero*, 2007]. The proposal distribution covariance matrix is subsequently updated three times. Each update involves sampling 30,000 sets of parameters using MCMC sampling. The posterior covariance matrix of the sampled parameters is calculated and scaled using the method of *Gelman et al.* [2004] to achieve an acceptance rate close to the optimal value of 0.23.

4. Model Use for Probabilistic Forecasting

[25] Now separate the $\mathbf{y}$ vector into two subvectors

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}(1) \\ \mathbf{y}(2) \end{bmatrix}, \tag{16}$$

where $\mathbf{y}(1)$ consists of predictor variables and $\mathbf{y}(2)$ consists of the predictand variables. Given what is known about $\mathbf{y}(1)$, a probabilistic forecast of $\mathbf{y}(2)$ is to be made. *Wang et al.* [2009] described a procedure for generating probabilistic forecasts in the form of ensembles. The procedure is applicable to cases when all the predictor values $\mathbf{y}(1)$ are precisely known. In this section, we extend the procedure to cases where some of the predictor values are only known to be equal to or below the censor thresholds.

[26] Corresponding to equation (16), the variable vector after the Yeo-Johnson transform is expressed in two subvectors

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}(1) \\ \mathbf{z}(2) \end{bmatrix}. \tag{17}$$

[27] The subvector $\mathbf{z}(1)$ is further divided into

$$\mathbf{z}(1) = \begin{bmatrix} \mathbf{z}(1a) \\ \mathbf{z}(1b) \end{bmatrix}, \tag{18}$$

where $\mathbf{z}(1a)$ consists of predictors whose values are precisely known, and $\mathbf{z}(1b)$ consists of predictors whose values are only known to be equal to or below the censor thresholds.

[28] To generate a set of forecast values for $\mathbf{y}(2)$, we first augment the predictor data by generating a random set of $\mathbf{z}(1b)$ that follows $p(\mathbf{z}(1b)|\mathbf{z}(1a))$ but within the subspace of $\mathbf{z}(1b) \le \mathbf{z}_c(1b)$ using the Gibbs sampler [*Gelman et al.*, 2004, pp. 287–288]. The conditional probability distribution $p(\mathbf{z}(1b)|\mathbf{z}(1a))$ is multivariate normal and can be found by applying equations (A4) and (A5) in Appendix A. A description of the Gibbs sampler as implemented for this application is given in Appendix B.

[29] The generated $\mathbf{z}(1b)$ is then treated as known and together with $\mathbf{z}(1a)$, which is already known, gives $\mathbf{z}(1)$ to generate a set of forecast values for $\mathbf{z}(2)$ using the conditional probability distribution $p(\mathbf{z}(2)|\mathbf{z}(1), \boldsymbol{\theta})$. An inverse of the Yeo-Johnson transform is then applied to give a set of forecast values for $\mathbf{y}(2)$. Any negative flow forecast values are converted back to zero flows.

[30] The procedure is repeated for all the parameter sets in the representative parameter sample previously obtained through MCMC sampling as described in section 3.4. This means that different sets of randomly sampled values of $\mathbf{z}(1b)$ are used with different parameter sets. The total collection of all the forecast values provides a numerical representation of the probabilistic forecast of the streamflows.

[31] Special care needs to be taken when forecasting events with predictor values well beyond the range of the historical data, on which the model was established. This is especially true with initial catchment condition predictors such as antecedent streamflows and rainfall. Antecedent streamflows may not be good predictors of future streamflows when they are beyond certain high values. Antecedent streamflows are used to indicate catchment wetness, which is what gives rise to some of the predictability of future streamflows. In a wet period, however, streamflows may continue to increase even long after the catchment has become fully wet. The use of antecedent streamflows may in such cases lead to forecasts of future streamflows that are too high. The same problem also applies to using antecedent rainfall as a predictor.

[32] We take a pragmatic approach to dealing with this problem. For a forecast event, if the antecedent streamflow exceeds the highest historically observed value, y_H, that has

been used for model inference at a particular site, the antecedent streamflow is first adjusted to y_H for that site and then used to produce a forecast. The same approach also applies to antecedent rainfall if used as a predictor. In future work, we will investigate the use of water balance modeling to better represent both wet and dry initial catchment conditions. It should be noted that for the test application presented in section 7, this pragmatic approach makes little difference to the results, but in some of our other applications it leads to lower and more reasonable forecast distribution tails for a few large forecast events.

5. Forecast Verifications

[33] The quality of forecasts is assessed using a leave-one-out cross-validation procedure in this study. The cross-validation procedure is implemented by sampling the posterior distribution of the parameters using a likelihood function based on all available data except one event. The streamflows for the left-out event are then forecast and compared with the observed data. This cross-validation procedure produces, for each forecast variable y, a series of forecast cumulative probability distributions $y^t \sim F^t(y^t)$ for events $t = 1,2,\ldots, n$. Paired with these forecasts are the observed values y^t_{OBS}. There are many facets to the verification of probabilistic forecasts of a continuous variable [e.g., *Gneiting et al.*, 2007; *Laio and Tamea*, 2007]. In this study, a number of statistical measures and graphical methods are used for both overall and detailed verifications.

5.1. Overall Verifications

[34] Skill scores are used to provide some overall measures of forecast performance. Skill scores are defined as percentage reductions in error scores (ES) of forecasts being assessed from reference forecasts

$$\mathrm{SS} = \frac{\mathrm{ES}_{\mathrm{REF}} - \mathrm{ES}}{\mathrm{ES}_{\mathrm{REF}}}. \tag{19}$$

[35] The reference forecasts may be naïve forecasts based on observed historical (climatology) mean, median, or distribution, or forecasts based on established models that set the benchmarks for assessing other models. Perfect forecasts will have a score of 1. The error scores may be defined in a number of ways such as the mean square error (MSE), RMSEP introduced in Appendix C, and continuous ranked probability score (CRPS). For probabilistic forecasts, skill assessments may focus on just the forecast means or medians, or be on the full probability density distributions (pdf). The skill scores for the forecast means and medians are useful for comparison with deterministic forecasts and with probabilistic forecasts that may not be reliable in their probability distributions. The following six skill scores are used in this study.

[36] 1. $\mathrm{SS}_{\mathrm{MSE}}$ (mean): Skill score for forecast means based on mean square error using observed historical (climatology) mean as reference forecasts. It is commonly known in hydrology as the Nash-Sutcliffe efficiency [*Nash and Sutcliffe*, 1970]. Note that this skill score is used only for the forecast means. It is not used for assessing the full probability distributions of the forecasts, because the mean expected square error will be very sensitive to the tails of the distributions. Even for the forecast means, $\mathrm{SS}_{\mathrm{MSE}}$ can be oversensitive to just a few (usually very high flow) events with large forecast errors.

[37] 2. $\mathrm{SS}_{\mathrm{MSE}}$ (median): Skill score for forecast medians based on mean square error using observed historical (climatology) median as reference forecasts. Forecast medians are often of great interest to forecasters and forecast users and therefore assessed in addition to the forecast means.

[38] 3. $\mathrm{SS}_{\mathrm{RMSEP}}$ (median): Skill score for forecast medians based on RMSEP using observed historical (climatology) median as reference forecasts. $\mathrm{SS}_{\mathrm{RMSEP}}$ is a new skill score introduced in this paper. Its rationale and formulation can be found in Appendix C. The main advantage of this skill score is that all forecast events are given a similar opportunity to contribute to the overall assessment of the forecast skill.

[39] 4. $\mathrm{SS}_{\mathrm{RMSEP}}$ (mean): Skill score for forecast means based on RMSEP using observed historical (climatology) mean as reference forecasts.

[40] 5. $\mathrm{SS}_{\mathrm{RMSEP}}$ (pdf): Skill score for probabilistic forecasts based on RMSEP using observed historical (climatology) distribution as reference forecasts.

[41] 6. $\mathrm{SS}_{\mathrm{CRPS}}$: Skill score for probabilistic forecasts based on CRPS [*Wang et al.*, 2009]. Like $\mathrm{SS}_{\mathrm{MSE}}$, this skill score can also be oversensitive to just a few (usually very high flow) events with large forecast errors.

[42] Although we recommend the use of $\mathrm{SS}_{\mathrm{RMSEP}}$ (median), $\mathrm{SS}_{\mathrm{MSE}}$ (median), and $\mathrm{SS}_{\mathrm{CRPS}}$ as the primary skill score measures, results for all of the six skill scores are presented in the test application in section 7.

[43] We also assess the overall reliability of the forecast probability distributions by using the PIT (probability integral transform) uniform probability plot [*Wang et al.*, 2009]. Given a forecast in the form of nonexceedance probability distribution $y^t \sim F^t(y^t)$, the PIT of the observed value y^t_{OBS} is given by

$$\pi^t = F^t(y^t_{\mathrm{OBS}}). \tag{20}$$

[44] For a reliable forecast, π^t should be uniformly distributed. The uniformity can be checked by pooling together π^t values for all the forecast events $t = 1, 2, \ldots, n$ and displaying ranked π^t values in a uniform probability plot [*Wang et al.*, 2009]. Figure 1 shows how the PIT uniform probability plot (also termed predictive QQ plot by *Thyer et al.* [2009]) may be used for indicating whether the forecast probability distributions are predicting too high or too low or too wide or narrow [*Laio and Tamea*, 2007; *Thyer et al.*, 2009].

[45] The use of PIT uniform probability plot becomes problematic when the observed value is not precisely known and is only known to be equal to or below a certain threshold $y^t_{\mathrm{OBS}} \le y_c$. The precise value for π^t cannot be found from equation (20). We overcome this problem by randomly generating a pseudo π^t value from a uniform distribution with a range $[0, F^t(y_c)]$ and subsequently using it with other real and pseudo π^t values to construct the PIT uniform probability plot (Figure 3).

[46] In addition to the use of PIT uniform probability plots to assess the overall reliability of forecast probability distributions, we also use reliability diagrams to specifically assess the reliability of forecast probability of a zero flow event, of an event smaller than 25%, 50%, or 75% of

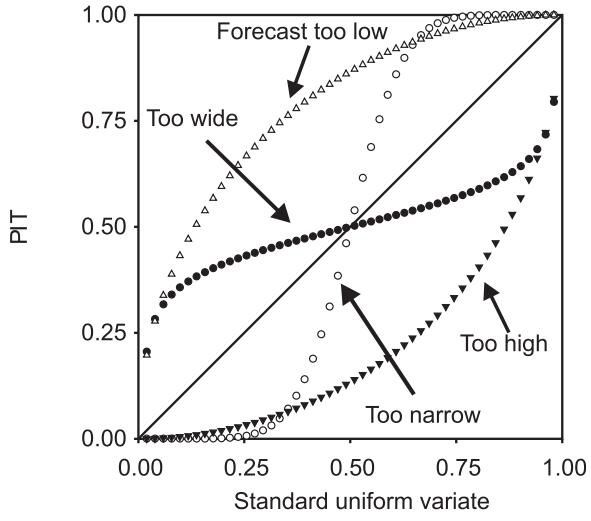

Figure 1. Possible outcomes on a PIT uniform probability plot [after *Laio and Tamea*, 2007].

at-site historical flows. A reliability diagram plots the observed frequency against the forecast probability and shows how well the predicted probability of an event corresponds to their observed frequency [*Wilks*, 1995].

5.2. Detailed Verifications

[47] For detailed verifications, we use a number of plots to examine the robustness of forecasts over time and event size [*Wang et al.*, 2009]. Forecast quantiles are plotted for individual events both chronologically and according to the forecast median and compared with observed values (Figures 5 and 6). Similarly, the PIT values for individual events are plotted both chronologically and according to the forecast median and compared with observed values (Figures 7 and 8). The four plots are visually examined for patterns and trends to identify if there are systematic errors in the forecasts.

6. Model Checking

[48] Forecast verification deals with model performance in a predictive mode. Model checking deals with the appropriateness of the model in a fitting mode. Model checking is particularly important here because of the additional complications involved in dealing with zero flows. The assumed Yeo-Johnson transformed multivariate normal probability model is checked for consistency with observations in terms of (1) the marginal distributions of individual predictor and predictand variables, (2) the marginal distributions of the principal components of the predictor and predictand variables, and (3) the marginal distribution of the sum of the predictand streamflows over multiple sites. The first check is to show how well the probability model describes the individual variables. The second and third checks are to show how well the probability model describes the variables jointly in different directions (as represented by the principal components and by the sum) in the multiple variable spaces. In addition, modeled cross-correlation coefficients of all predictor and predictand variables are compared with values directly calculated from observed data. This provides another check on how well the probability model describes the relationships among the variables. Detailed methods of model checking can be found in the work of *Wang et al.* [2009]. Because many of the variables are highly skewed, some of which contain a large number of zero values, and the relationships among the variables tend to be highly nonlinear, the Kendall's tau-b rank correlation coefficient is more appropriately used in this study instead of the more commonly used Pearson's correlation coefficient.

7. Application

7.1. Data

[49] To demonstrate the extended BJP modeling approach, it was applied to joint forecasting of streamflows at three gauging stations in the Burdekin River catchment in northeast Australia and catchment average rainfall (Figure 2). The three gauging stations measure the majority of the inflows into the Burdekin Falls Dam, the primary water storage in the Burdekin catchment. The Burdekin River catchment has a maximum allowable annual water use of 1975 GL, the majority of which is used by agricultural industries for irrigation [*Department of Environment and Resource Management*, 2009; *Beare et al.*, 2003]. Farm revenue from irrigated agricultural production, primarily of sugar cane and horticultural crops, exceeded $450 million

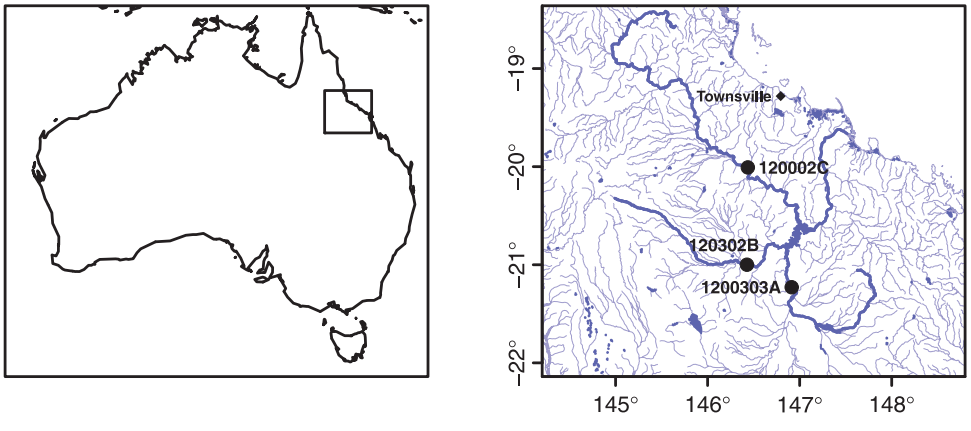

Figure 2. Locations of streamflow gauging stations.

in 2000 [*Beare et al.*, 2003]. Annual irrigation water allocations vary according to water resource availability. Progressive announcements are made throughout a season with allocations increasing according to the availability of water in storage. Irrigators wanting to plan their cropping strategies early in an irrigation season need reliable forecasts of likely irrigation allocations, which depend on reliable forecasts of storage inflows.

[50] The three gauging stations and their associated catchment areas and mean annual runoff as well as rainfall are given in Table 1. The location of the gauges is shown in Figure 2. A list of the predictors and predictands used are in Table 2. Streamflows and rainfall from June to August were forecast from predictors describing the initial catchment condition and oncoming climate. The total streamflows for April and May at the three sites were used to represent the initial catchment conditions, while the NINO4 index for May was used to represent the oncoming climate condition. NINO34 is the sea surface temperature anomaly over 150°W–160°E and 5°S–5°N. As the purpose of this application was to demonstrate the BJP modeling approach, other potentially useful predictors were not investigated here but are being followed up in a further study (Robertson and Wang, submitted manuscript, 2010).

[51] In the final joint probability model, the forecast rainfall variable is decoupled from the predictors representing initial catchment conditions by setting the correlations between them to zero. Thus, rainfall is forecast only from the NINO4 predictor, which represents the oncoming climate condition, but correlation between concurrent forecast rainfall and forecast streamflows are maintained. The provision of rainfall forecasts is a useful addition to streamflow forecasts, but in our experience the inclusion of rainfall as a copredictand also helps stabilize streamflow forecasts, especially in the upper tails of probability distributions of large streamflow forecast events. The most likely reason for this stabilization effect is that any extrapolated streamflow forecasts need to maintain a reasonable relationship with rainfall, which keeps streamflow forecasts in check.

[52] The data used covers a period of 42 years from 1967 to 2008 with some data missing (Table 2). In total there are 39 years (events) with complete records in which none of the predictors and predictands have data missing. Zero values of observed streamflows were treated as censored data. Two of the streamflow gauging stations had zero flows with approximately 20% of events for the predictors and 40–55% of events for the predictands. The rainfall data did not have zero values, and the NINO4 Index did not need censoring. Table 2 also shows the mean, median, minimum, and maximum values of data for each of the predictor and predictand variables. It is clear that the flow data are extremely skewed and require strong transformation to be normalized.

[53] Cross-validation was carried out by leaving out one of the events that have complete records of all the predictors and predictands. A sample of 1000 sets of parameter values were generated by MCMC sampling. A probabilistic forecast of the streamflows for June to August at the three gauging stations and June to August catchment rainfall was then made for the left-out event. The forecast was numerically represented by a sample of 1000 sets of values, one generated from each of the 1000 sets of parameter values. This procedure was repeated for all historical events with complete records.

7.2. Forecast Verification Results

7.2.1. MSE, RMSEP, and CRPS Skill Scores

[54] The skill scores for the forecast of June to August streamflows at the three gauges and catchment rainfall are given in Table 3. It shows that the streamflow forecasts are greater than 10% in all the skill scores. Forecasts for gauge 120002C are higher in most of the skill scores, reflecting the fact that it is generally easier to forecast for streams that have longer memory and fewer or no zero flow events. The RMSEP skills are higher than the MSE skill scores for gauges 120002C and 120302B because the forecasts are better in predicting the general direction of low or high flows than the exact flows, particularly for the 1990 high flow event (see Figure 5). The reverse is true for gauge 120303A, for which at least a few of the high flow forecasts, including for the 1990 event, are closer to the observed flows. In all cases, the MSE skill score for the forecast median is higher than for the forecast mean. This is because the forecast mean is sensitive to the forecast distribution, in particular the high flow tail which is the most difficult to get right.

[55] The skill scores for rainfall forecasts are mostly below 10%, demonstrating the challenge in achieving skillful seasonal rainfall forecasts [*Drosdowsky and Chambers*, 2001; *Fawcett*, 2008; *Fawcett et al.*, 2005]. The higher skill in streamflow forecasts than rainfall forecasts results from the use of information on initial catchment conditions (represented by antecedent streamflows here).

7.2.2. Pit Uniform Probability Plots

[56] The PIT uniform probability plots for the forecast of June to August streamflows at the three gauges and catchment rainfall are given in Figure 3. The PIT values are in general close to the diagonal line. Small departures from the diagonal line are observed, but they are well within the Kolmogorov 5% significance band [*Laio and Tamea*, 2007], suggesting that the departures are within acceptable sample variability. Therefore, our interpretation of the results is that the PIT values are distributed quite uniformly and the forecast probability distributions are overall reliable in that they are not predicting too high or too low or too wide or too narrow. In other words, the forecast

Table 1. Streamflow Gauging Stations and Associated Catchments

Gauge Number	Station Name	Catchment Area	Mean Annual Runoff or Rainfall
120002C	Burdekin River at Selheim	36,260 km^2	117 mm (4125 GL)
120303A	Suttor River at St. Anns	50,291 km^2	28 mm (1408 GL)
120302B	Cape River at Taemas	16,074 km^2	38 mm (611 GL)
Rainfall	Areal average	102,625 km^2	603 mm

Table 2. Predictors and Predictands[a]

	Predictor				Predictand			
Variable	NINO4	Flow	Flow	Flow	Flow	Flow	Flow	Rainfall
Gauge number		120002C	120303A	120302B	120002C	120303A	120302B	
Period	May	Apr–May	Apr–May	Apr–May	June–Aug	June–Aug	June–Aug	June–Aug
Missing data (years)	0	3	1	1	3	2	2	0
Number of zero values	Not applicable	0	8	9	0	17	23	0
Mean	0.06	8.8 mm (318 GL)	3.6 mm (183 GL)	2.9 mm (47 GL)	2.4 mm (89 GL)	0.7 mm (35 GL)	0.5 mm (9 GL)	61.1 mm
Median	0.15	4.6 mm (166 GL)	0.1 mm (3 GL)	0.1 mm (2 GL)	1.4 mm (49 GL)	0.0 mm (2 GL)	0.0 mm (0 GL)	47.9 mm
Minimum	–0.68	0.0 mm (1 GL0	0.0 mm (0 GL)	0.0 mm (0 GL)	0.1 mm (3 GL)	0.0 mm (0 GL)	0.0 mm (0 GL)	4.5 mm
Maximum	0.88	49.3 mm (1787 GL)	63.1mm (3175 GL)	43.5 mm (700 GL)	17.9 mm (649 GL)	5.2 mm (252 GL)	8.8 mm (142 GL)	194.3 mm

[a]Data from 1967 to 2008. Rainfall uses only NINO4 as its predictor.

Table 3. June to August Streamflow and Rainfall Forecast Skill Scores[a]

Gauge	120002C	120303A	120302B	Rainfall	Sum of flows
SS_{MSE} (mean)	15	22	13	11	27
SS_{MSE} (median)	30	30	16	17	36
SS_{RMSEP} (median)	38	19	29	9	24
SS_{RMSEP} (mean)	47	16	31	8	40
SS_{RMSEP} (pdf)	36	11	19	6	27
SS_{CRPS}	19	11	10	5	22

[a]Scores are in percentages.

probability distributions are overall unbiased and of appropriate spread.

7.2.3. Reliability Diagrams

[57] Forecast reliability diagrams of a zero flow event, of an event smaller than 25%, 50%, or 75% of at-site historical flows are presented in Figure 4. In producing these diagrams, forecasts for the three sites are pooled to increase the sample size and the range of forecast probability is divided into four bins (see inserts). The [0.05, 0.95] uncertainty interval of the observed relative frequency is calculated through bootstrap resampling of the forecasts and observed flows jointly at the three sites. For the majority of forecast probability ranges, the uncertainty interval of the observed relative frequency intersects the theoretical 1:1 line, indicating that the observed relative frequency is not significantly different from the forecast probability. Forecasts are therefore considered reliable with respect to all of the four forecast thresholds.

7.2.4. Forecast Quantile and Observed Value Comparison Plots

[58] Figures 5 and 6 provide a comparison of the forecast median and quantile ranges with an observed value for individual events. The events are displayed chronologically in Figure 5 and according to the forecast median in Figure 6. Figure 5 highlights the ability of the method to jointly handle concurrent forecasts of zero and nonzero flows. The forecast medians (and 0.75 quantiles) at site 120302B are equal to zero for the majority of events. For several of these events, concurrent nonzero forecasts are made at the other gauges. From Figure 5, there does not appear to be any obvious trend with time in the relationship between the forecasts and observed values. For example, the forecasts are not biased in any particular direction over time.

[59] The forecast medians and quantiles shown in Figure 6 are before the negative values were converted back to 0. This allows the many events with final forecast medians equal to 0 to be separated on the horizontal axis. From Figure 6, the forecast medians appear to be consistent with observed values. The forecast quantile ranges increase with the forecast median and also appear to be consistent with the observed values. There does not appear to be any trend with the forecast median in the relationship between the forecasts and the observed values. For example, the forecasts are not obviously biased in any particular direction over the forecast median. Figures 5 and 6 also provide a contrast of the BJP forecasts with climatology reference forecasts, showing the information gained from the BJP modeling over climatology only.

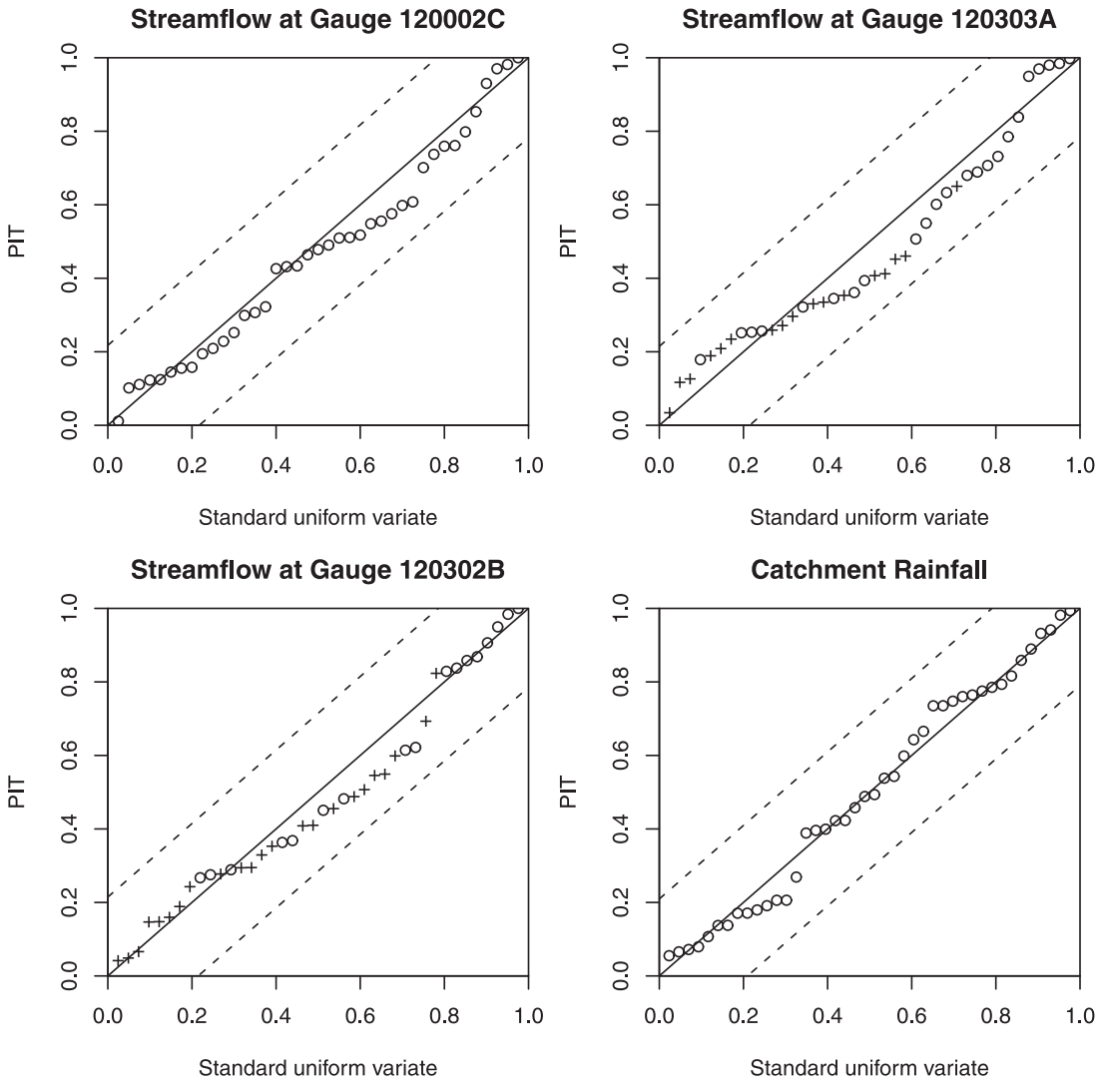

Figure 3. PIT uniform probability plot (1:1 solid line, theoretical uniform distribution; dashed lines, Kolmogorov 5% significance band; open circles, PIT value of observed streamflow or rainfall; crosses, pseudo PIT value).

7.2.5. PIT Plots

[60] Figures 7 and 8 present PIT values for individual events. The events are displayed chronologically in Figure 7 and according to the forecast median before it was converted to zero in Figure 8. Ideal forecasts should lead to PIT values that are uniformly scattered in [0, 1], and should not exhibit any obvious trends with time or with the forecast median. This appears to be the case both with time in Figure 7 and with the forecast median in Figure 8.

7.2.6. Verifications of Forecast Sum of Flows Over Multiple Sites

[61] The sum of flows over the three sites is not directly forecast, but its verifications are useful for showing whether the relationships between the forecasts at individual sites are sensible. Skill scores for the forecast sum of flows are shown in Table 3. They appear to be more consistent across the different measures than for individual sites. Because the average June to August flow at gauge 120002C is about twice of the average flows at the other two gauges put together (Table 2), the skill scores for the forecast sum of flows are generally closer to those for gauge 120002C. Other verification results are shown in Figures 9 and 10 and considered satisfactory. The forecasts show a slight bias toward high values, but the bias is within the 5% significance band (Figure 10a in reference to Figure 1).

7.3. Model Checking Results

[62] The modeled and observed marginal distributions of the predictors and predictands are compared in Figure 11. The observed data are reasonably well described by the modeled median marginal distributions. The majority of observed data points fall within the modeled [0.05, 0.95] uncertainty band, and the extent of modeled zero flows are consistent with the observed flows. A comparison of the modeled marginal distributions of the principal components of the predictors and predictands with those derived from observed data is given in Figure 12. Again, there is a good match between the modeled results and observed results,

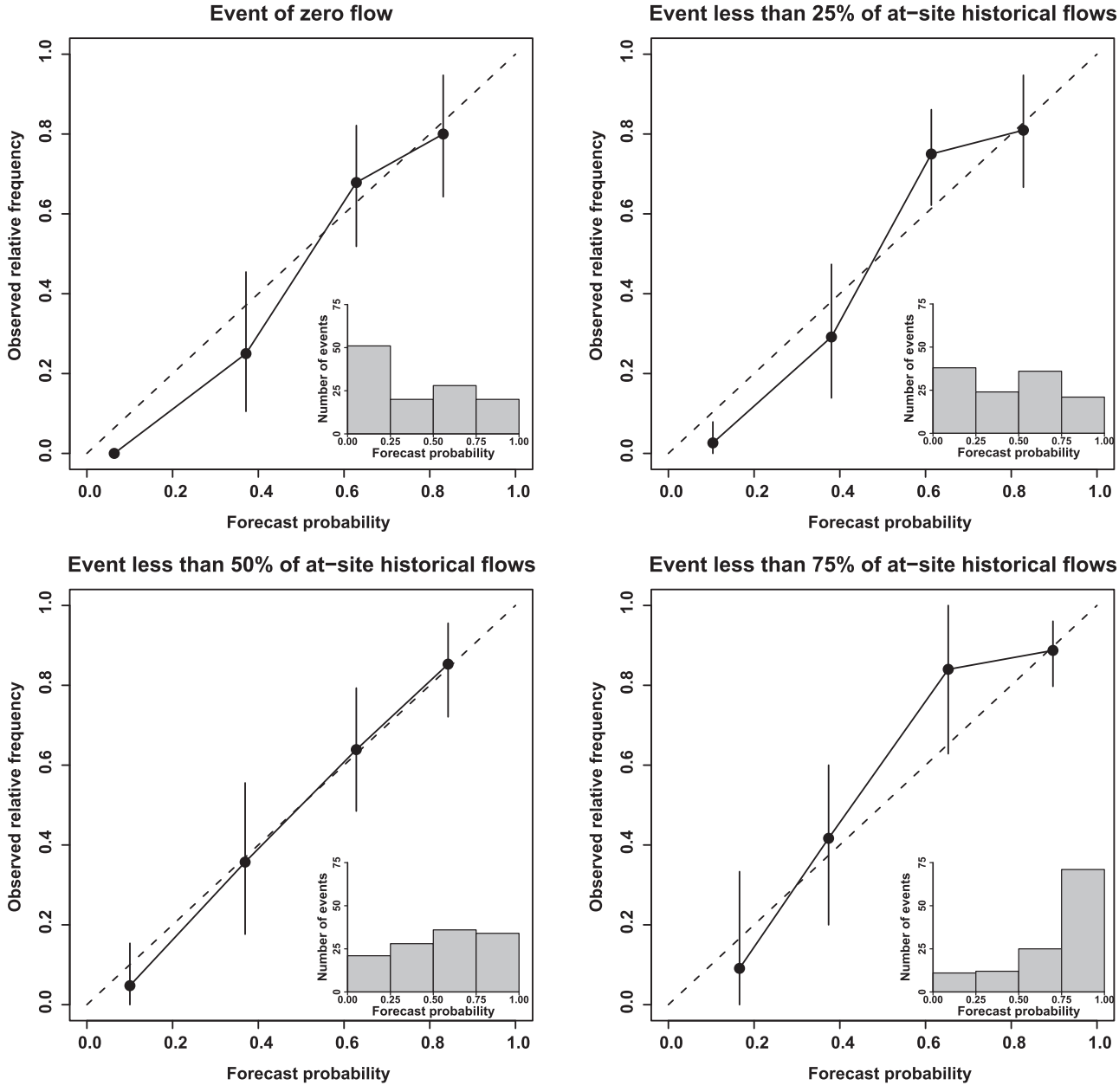

Figure 4. Forecast reliability diagrams of a zero flow event, of an event smaller than 25%, 50%, or 75% of at-site historical flows (1:1 dashed lines, perfectly reliable forecast; circles, observed relative frequency; vertical lines, [0.05, 0.95] uncertainty interval of observed relative frequency; inserts, number of events in different forecast probability ranges).

considering that the marginal distributions of the principal components were not explicitly modeled. Figure 13 shows the modeled marginal distribution of the sum of predictand streamflows at the three gauges. The modeled results match closely with the observed data. Figure 14 compares the modeled medians and [0.05, 0.95] uncertainty ranges of the (Kendall's tau-b rank) cross-correlation coefficients of all the predictor and predictand variables with values directly calculated from observed data. The modeled correlation coefficients agree well with the observed values. All the above results indicate that the model assumption and treatment of zero flows as censored data are reasonably consistent with observed data.

8. Summary and Conclusion

[63] Skillful and reliable forecasts of seasonal streamflows are highly valuable to water resources management. This paper extends the applicability of the previously developed BJP modeling approach for seasonal forecasting of streamflows at multiple sites to streams with the occurrences of zero flows. Forecasts are produced in the form of probabilities of zero seasonal flows and probability distributions of above zero seasonal flows. These forecasts are produced using a Yeo-Johnson transformed multivariate normal distribution to model the joint distribution of future streamflows and their predictors. In model parameter inference, the zero observed streamflows are treated as censored

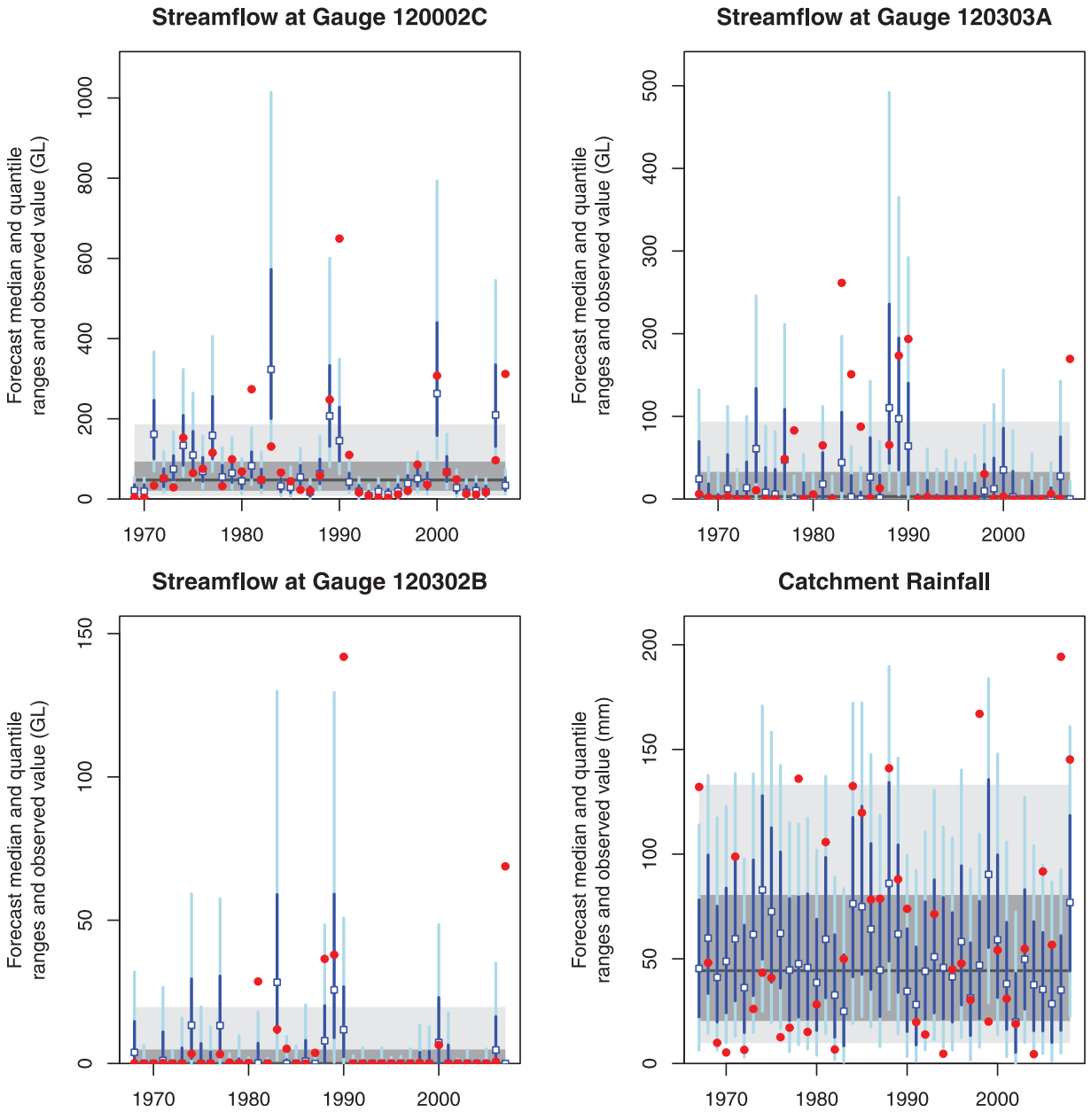

Figure 5. Forecast quantiles and observed values plotted chronologically (open squares, forecast median; dark blue vertical lines, forecast [0.25, 0.75] quantile range; light and dark blue vertical lines, forecast [0.10, 0.90] quantile range; dark gray horizontal lines, climatology median; mid gray shade, climatology [0.25, 0.75] quantile range; light and mid gray shade, climatology [0.10, 0.90] quantile range; red dot, observed streamflow or rainfall).

data, having unknown precise values but equal to or below zero. In forecasting, censored predictor values are augmented to "known" values, and negative values of streamflow (and rainfall) forecasts are converted to zero.

[64] In addition, improvements are made to a number of detailed techniques in the work of *Wang et al.* [2009], including reparameterization, prior specification, and MCMC sampling to both cater for censored data and gain computational efficiency. A new skill score based on the RMSEP is introduced as an alternative to LEPS. It retains the conceptually attractive characteristics of the LEPS skill score but is much easier to understand and consistent with the normalization used in many modern skill scores.

[65] The extended BJP modeling approach was applied to forecasting streamflows at three river gauges in the Burdekin River catchment in northern Queensland, Australia. June to August streamflows were forecast from the NINO4 index for the previous month and total streamflows for the previous 2 months. Catchment average rainfall from June to August was jointly forecast with streamflows for the same period but from the NINO4 index only. Cross-validation results show that the BJP probabilistic forecasts

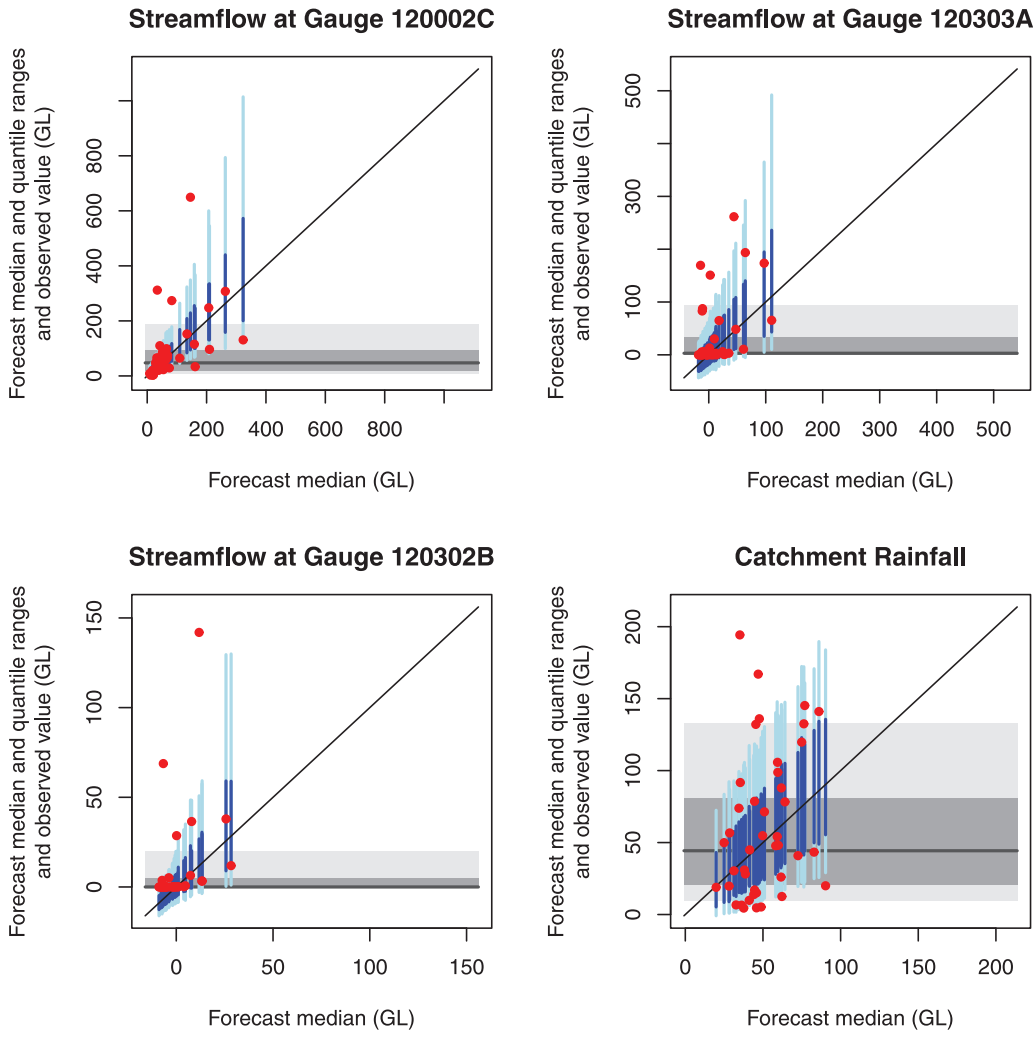

Figure 6. Forecast quantiles and observed value plotted according to the forecast median (1:1 line, forecast median; dark blue vertical lines, forecast [0.25, 0.75] quantile range; light and dark blue vertical lines, forecast [0.10, 0.90] quantile range; dark gray horizontal lines, climatology median; mid gray shade, climatology [0.25, 0.75] quantile range; light and mid gray shade, climatology [0.10, 0.90] quantile range; red dot, observed streamflow or rainfall).

of streamflows are of some skills, are free from obvious bias and other errors, and have appropriate uncertainty spread. The skill of rainfall forecasts is lower than streamflow forecasts, demonstrating the importance of using initial catchment condition for streamflow forecasts. Model checking results show that the model assumptions and the treatment of zero flows as censored data are consistent with the observed data.

Appendix A: Conditional Distributions of a Multivariate Normal Distribution

[66] Vector $\mathbf{z}$ follows a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. The vector can be partitioned into two subvectors

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}(1) \\ \mathbf{z}(2) \end{bmatrix}. \tag{A1}$$

[67] The corresponding partitions of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are given by

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}(1) \\ \boldsymbol{\mu}(2) \end{bmatrix}, \tag{A2}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}(1,1) & \boldsymbol{\Sigma}(1,2) \\ \boldsymbol{\Sigma}(2,1) & \boldsymbol{\Sigma}(2,2) \end{bmatrix}. \tag{A3}$$

[68] The conditional probability distribution of $\mathbf{z}(2)$ given $\mathbf{z}(1)$, $p(\mathbf{z}(2)|\mathbf{z}(1))$ is also normal with mean vector and covariance matrix [*Gelman et al.*, 2004, p. 579],

$$\boldsymbol{\mu}'(2) = \boldsymbol{\mu}(2) + \boldsymbol{\Sigma}(2,1)[\boldsymbol{\Sigma}(1,1)]^{-1}[\mathbf{z}(1) - \boldsymbol{\mu}(1)], \tag{A4}$$

$$\boldsymbol{\Sigma}'(2,2) = \boldsymbol{\Sigma}(2,2) - \boldsymbol{\Sigma}(2,1)[\boldsymbol{\Sigma}(1,1)]^{-1}\boldsymbol{\Sigma}(1,2). \tag{A5}$$

Appendix B: Sampling of a Multivariate Normal Distribution Below Thresholds Using the Gibbs Sampler

[69] The aim here is to produce a random sample from a known multivariate normal distribution $p(\mathbf{z})$ (corresponding

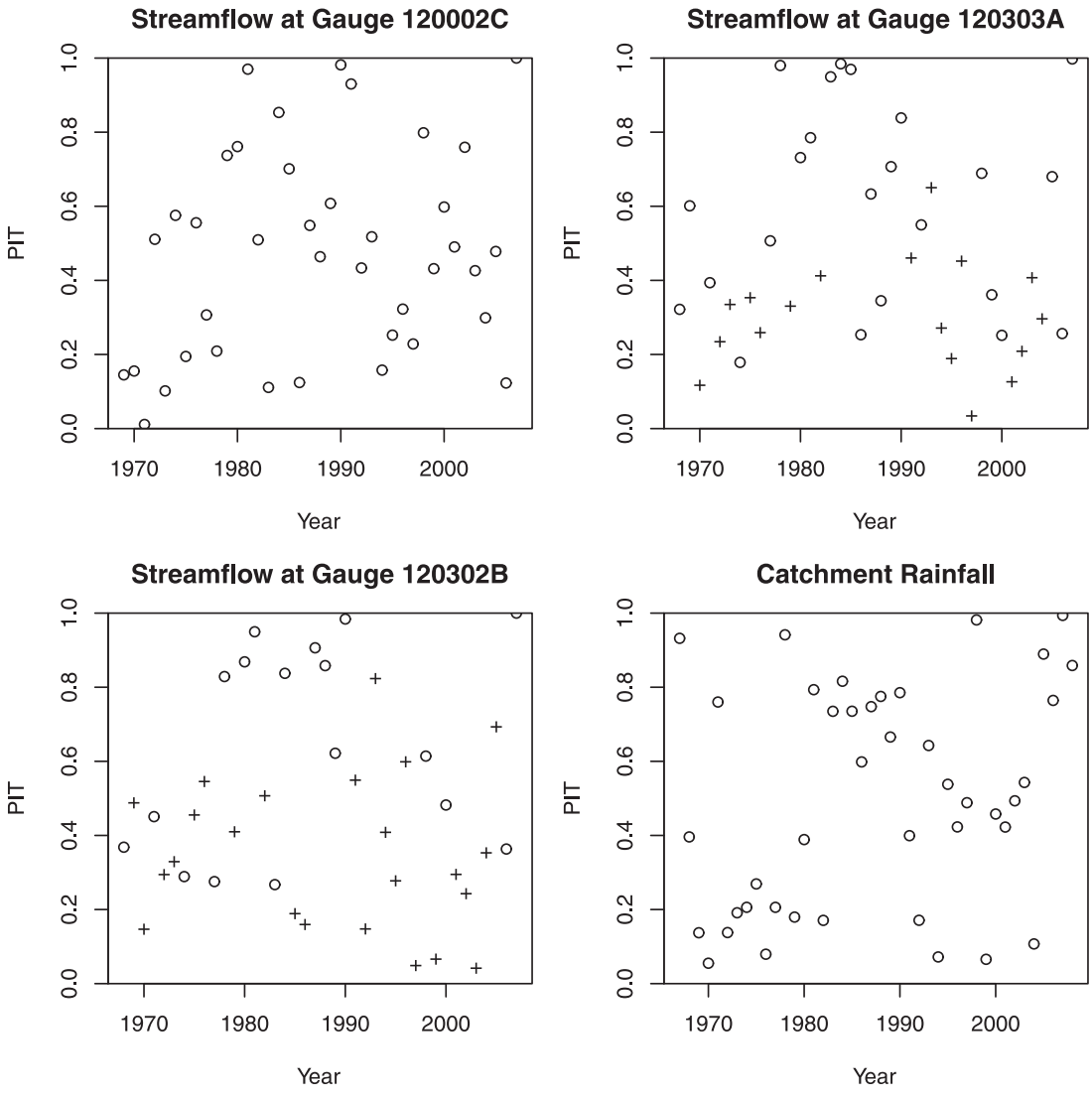

Figure 7. PIT values plotted chronologically (open circles, PIT values of observed streamflow or rainfall; crosses, pseudo PIT values).

to $p(\mathbf{z}(1b)|\mathbf{z}(1a))$ in section 4) in the variable space equal to or below specified thresholds $\mathbf{z}_c$ (corresponding to $\mathbf{z}(1b) \leq \mathbf{z}_c(1b)$ in section 4). The Gibbs sampler, which is a particular MCMC sampling algorithm, is highly suitable for this purpose. Assume that the vector $\mathbf{z}$ is d dimensional and its distribution has defined parameters. The Gibbs sampler takes the following steps:

[70] 1. Choose a starting point $\mathbf{z}^0$ in the variable space of interest.

[71] 2. Randomly assign the order of the variables in $\mathbf{z}$ to $\mathbf{z} = [z_1, z_2, \ldots, z_i, \ldots, z_d]$.

[72] 3. Randomly sample from the conditional probability distribution $p(z_1 | \mathbf{z}_{-1} = \mathbf{z}^0_{-1})$ in the space $z_1 \leq z_{c1}$, where $\mathbf{z}_{-1}$ represents all the variables in $\mathbf{z}$ except z_1.

[73] 4. Update z_1^0 to the newly sampled value from step 3.

[74] 5. Repeat steps 3 and 4 for the rest of the variables $i = 2, \ldots, d$, one variable at a time, to complete a cycle.

[75] 6. Repeat steps 2–5 for as many cycles as necessary so that the final sample is not influenced by the choice of the starting point in step 1.

[76] The conditional probability distribution $p(z_i|\mathbf{z}_{-i} = \mathbf{z}^0_{-i})$. in step 3 can be found by applying equations (A4) and (A5) in Appendix A. As the distribution is a univariate normal distribution, the cumulative distribution $F_{ci} = p(z_i \leq z_{ci} | \mathbf{z}_{-i} = \mathbf{z}^0_{-i})$ can be easily found [*Gelman et al.*, 2004]. A uniformly random number in the range of $[0, F_{ci}]$ can be generated to give a sampled value for z_i using the inverse of the normal distribution. For our application in this study, we used the threshold values as the starting point. We allowed 1000 cycles of sampling.

Appendix C: Skill Score Based on RMSEP

[77] In this appendix, we introduce a new forecast skill score based on RMSEP. It is built on the broad concept behind the LEPS skill score [*Ward and Folland*, 1991; *Potts et al.*, 1996] but presented in such a form that is easy to understand and consistent with the normalization used in many modern skill scores. The main advantages of the LEPS and RMSEP skill scores are that all events of forecasts are given a similar opportunity to contribute

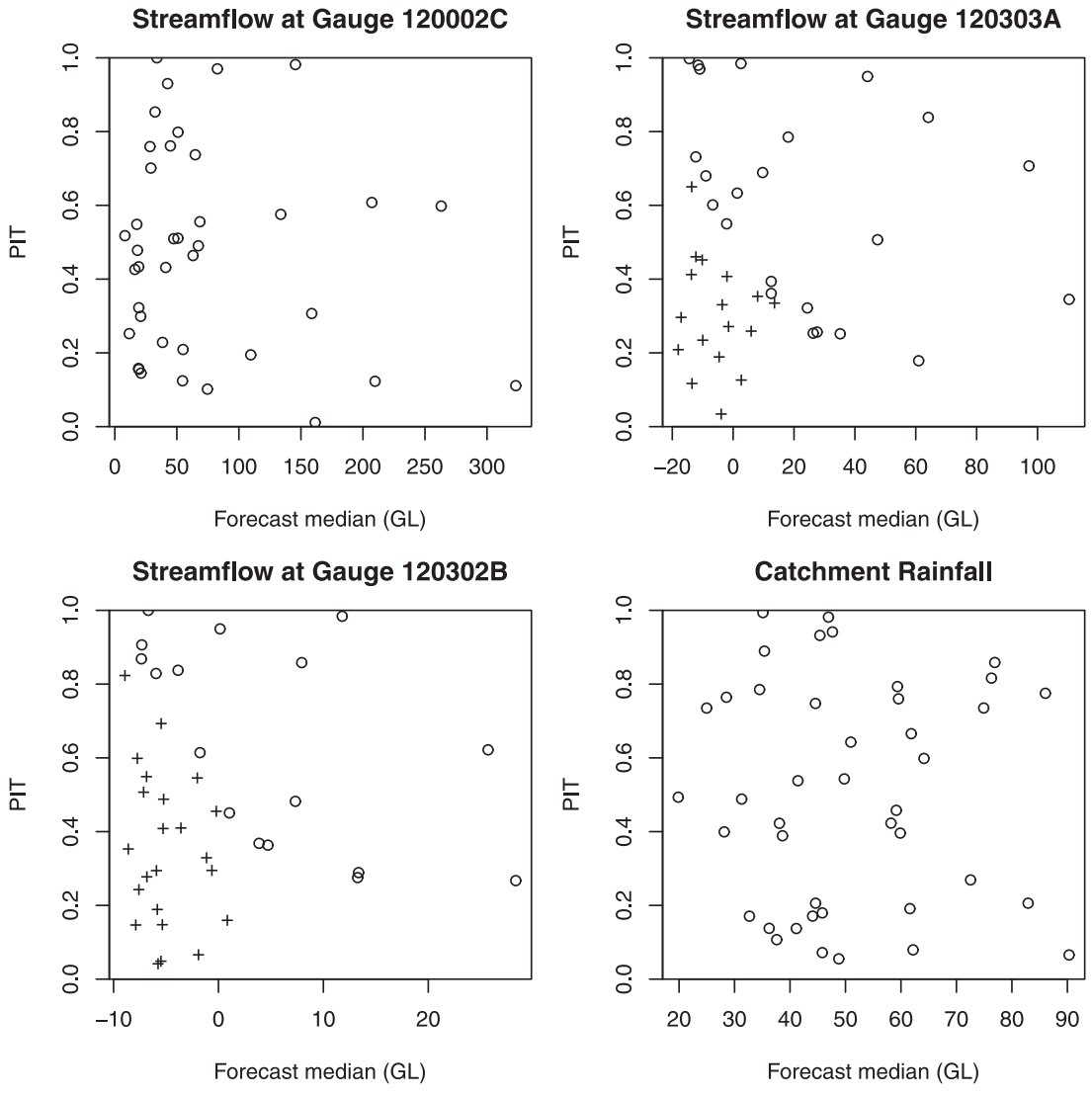

Figure 8. PIT values plotted according to the forecast median (open circles, PIT value of observed streamflow or rainfall; crosses, pseudo PIT value).

to the overall assessment of the forecast skill. In contrast, a number of available skill scores, such as the Nash-Sutcliff efficiency [*Nash and Sutcliffe*, 1970] and the continuous ranked probability score, are highly sensitive to just a few large errors, which tend to be associated with large events.

[78] First we introduce the RMSEP skill score for single value forecasts of a continuous variable y and denote the forecasts and observations, respectively, as y^t and y^t_{OBS}, $t = 1, 2, \ldots, n$ being the events under assessment. The most commonly used measure of forecast error for an event is on the original variable scale $(y^t - y^t_{\mathrm{OBS}})$. We use a measure of forecast error on a probability scale, which is the observed historical distribution (climatology) of y in the form of nonexceedance probability $F_{\mathrm{CLI}}(y)$. The forecast error in probability for an event is then $\left(F_{\mathrm{CLI}}(y^t) - F_{\mathrm{CLI}}(y^t_{\mathrm{OBS}})\right)$. Figure C1 is a schematic of forecast error in probability corresponding to error on the original variable scale. For a total of n events, the root mean square error in probability is

$$\mathrm{RMSEP} = \left[\frac{1}{n}\sum_{t=1}^{n}\left(F_{\mathrm{CLI}}(y^t) - F_{\mathrm{CLI}}(y^t_{\mathrm{OBS}})\right)^2\right]^{\frac{1}{2}}. \qquad \text{(C1)}$$

[79] The RMSEP of the forecasts can then be normalized to give a RMSEP skill score

$$\mathrm{SS}_{\mathrm{RMSEP}} = \frac{\mathrm{RMSEP}_{\mathrm{REF}} - \mathrm{RMSEP}}{\mathrm{RMSEP}_{\mathrm{REF}}}, \qquad \text{(C2)}$$

where $\mathrm{RMSEP}_{\mathrm{REF}}$ is the RMSEP of a set of reference forecasts. The most commonly used single value reference forecasts are climatology means or medians.

[80] The climatology distribution $F_{\mathrm{CLI}}(y)$ can be found by fitting a probability distribution function (such as a Yeo-Johnson transformed normal distribution) to the historically observed data. Alternatively, one may assign nonexceedance probabilities to the ranked historically observed data and then apply simple interpolation and extrapolation to estimate the nonexceedance probability for a given y value.

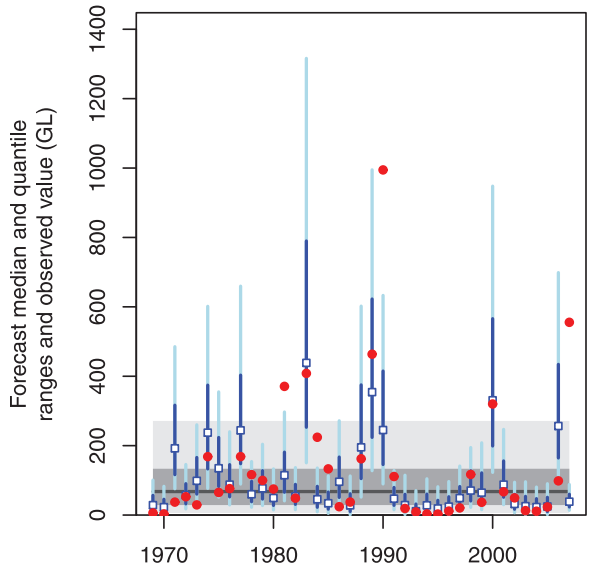

Figure 9. Verifications of forecast sum of flows over the three sites (the legends are the same as in Figure 5).

[81] Next we extend the RMSEP skill score to probabilistic forecasts. We denote the forecast density distribution for event t as $f^t(y^t)$. The RMSEP for a total of n events is then the root mean expected square error in probability

$$\mathrm{RMSEP} = \left[\frac{1}{n}\sum_{t=1}^{n}\int_{-\infty}^{\infty}\left(F_{\mathrm{CLI}}(y^t) - F_{\mathrm{CLI}}(y^t_{\mathrm{OBS}})\right)^2 f^t(y^t)dy^t\right]^{\frac{1}{2}}. \tag{C3}$$

[82] If the density distribution $f^t(y^t)$ has a probability mass at a point, for example, in zero flow problems, care needs to be taken to include the probability mass in the above integration.

[83] In practice, probabilistic forecasts are often presented in the form of ensemble forecasts. For example, the density distribution $f^t(y^t)$ is numerically represented by ensemble members $y^{t,k}$, $k = 1, 2, \ldots, m$. The integration in equation (C3) is then easily carried out numerically as

$$\mathrm{RMSEP} = \left[\frac{1}{n}\sum_{t=1}^{n}\frac{1}{m}\sum_{k=1}^{m}\left(F_{\mathrm{CLI}}(y^{t,k}) - F_{\mathrm{CLI}}(y^t_{\mathrm{OBS}})\right)^2\right]^{\frac{1}{2}}. \tag{C4}$$

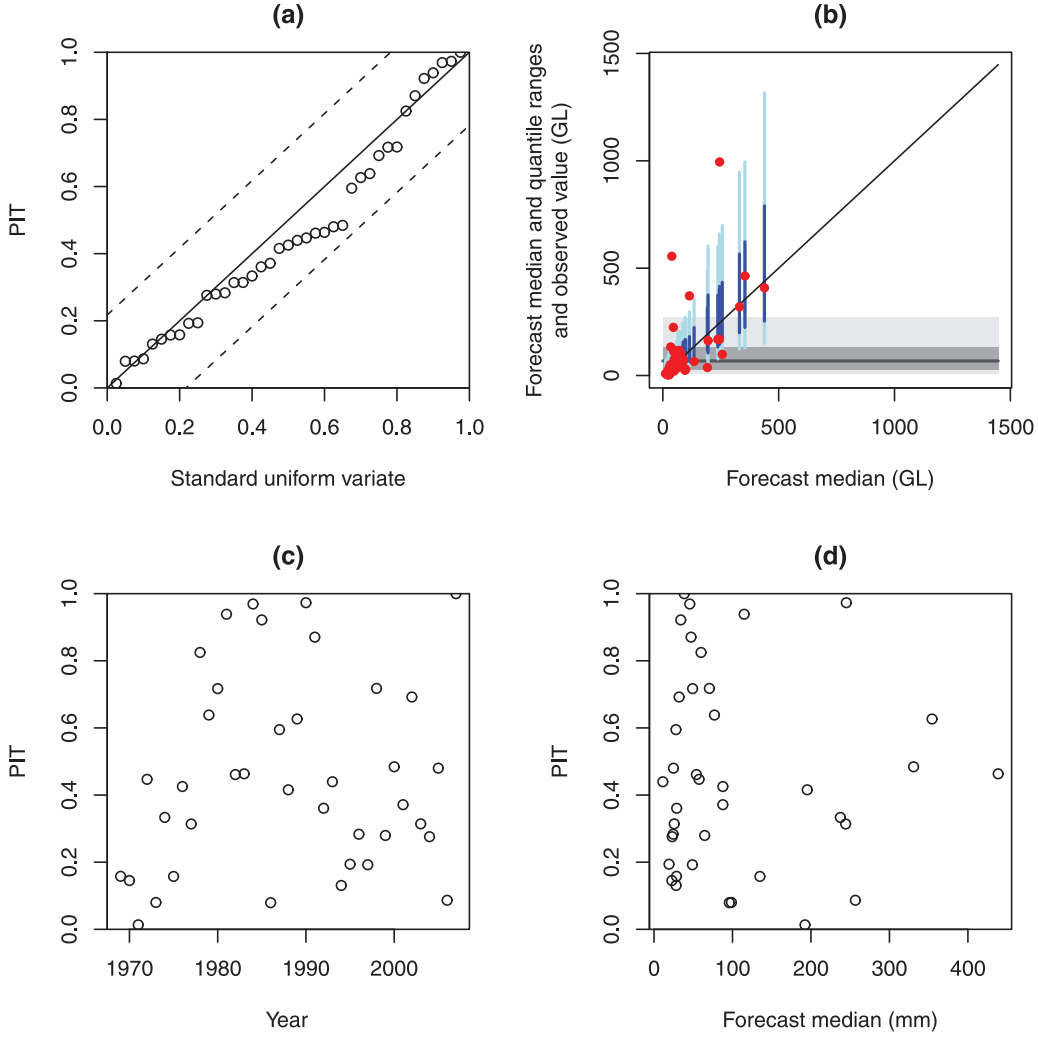

Figure 10. Verifications of forecast sum of flows over the three sites. ((a) the same as in Figure 3; (b) the same as in Figure 6; (c) the same as in Figure 7; and (d) the same as in Figure 8.)

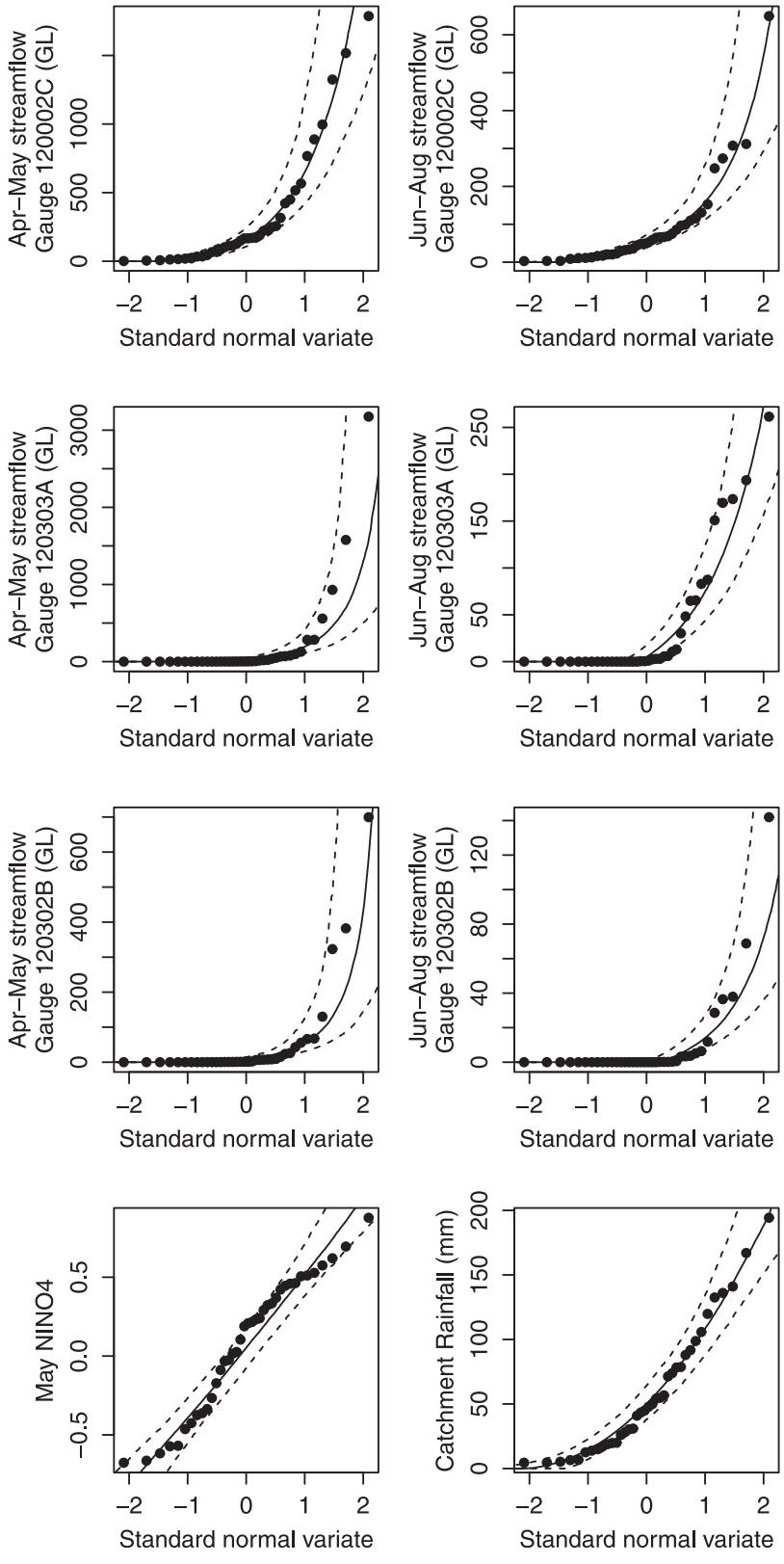

Figure 11. Marginal distribution of predictors and predictands (solid lines, modeled marginal distribution median; dashed lines, marginal distribution [0.05, 0.95] uncertainty band; dots, observed data).

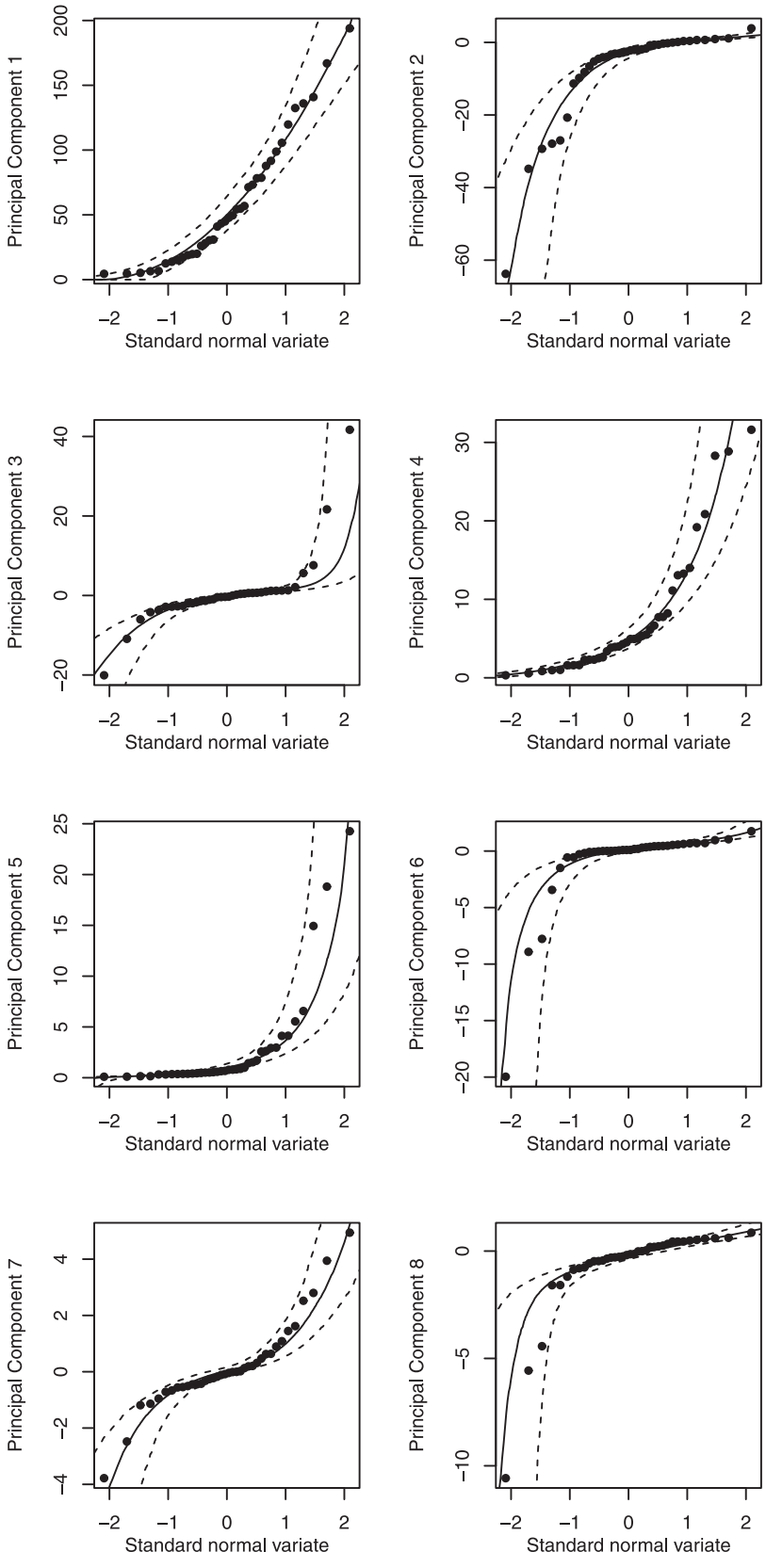

Figure 12. Marginal distribution of principal components of all predictors and predictands (solid lines, modeled marginal distribution median; dashed lines, marginal distribution [0.05, 0.95] uncertainty band; dots, value directly calculated from observed data).

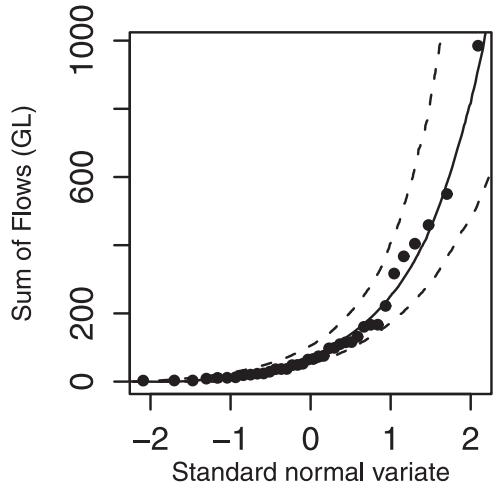

Figure 13. Marginal distribution of the sum of predictand streamflows at the three sites (solid lines, modeled marginal distribution median; dashed lines, marginal distribution [0.05, 0.95] uncertainty band; dots, observed data).

[84] The RMSEP skill score for probabilistic forecasts can be similarly calculated using equation (C2). The most commonly used reference forecasts for assessing probabilistic forecasts are the climatology distribution $F_{CLI}(y)$. When using equation (C4) to calculate $RMSEP_{REF}$, the historically observed data may simply be used as a set of ensemble members of $F_{CLI}(y)$, or if necessary a larger ensemble set can be generated from the fitted distribution $F_{CLI}(y)$.

[85] It should be noted that the RMSEP skill score for probabilistic forecasts tends to reward forecasts that are high in resolution (narrow in probability distribution spread). Therefore, the most meaningful interpretation of the RMSEP skill score for probabilistic forecasts is when the forecast probability distributions are found reliable (see section 5.1).

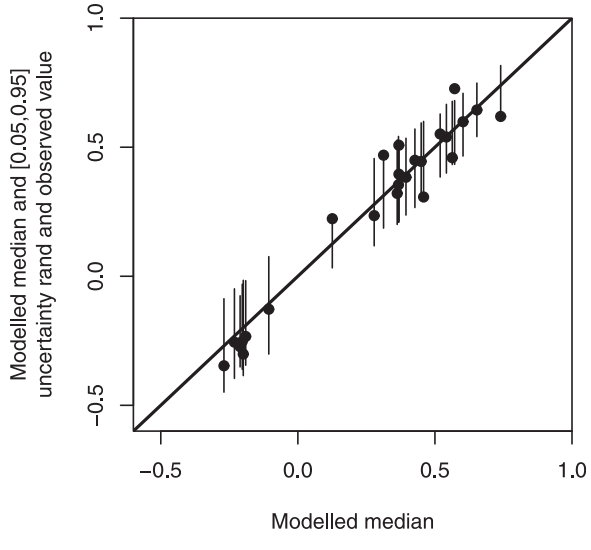

Figure 14. Cross-correlation coefficients of all the predictors and predictand variables (1:1 lines, modeled median; vertical lines, modeled [0.05, 0.95] uncertainty range; dots, value directly calculated from observed data).

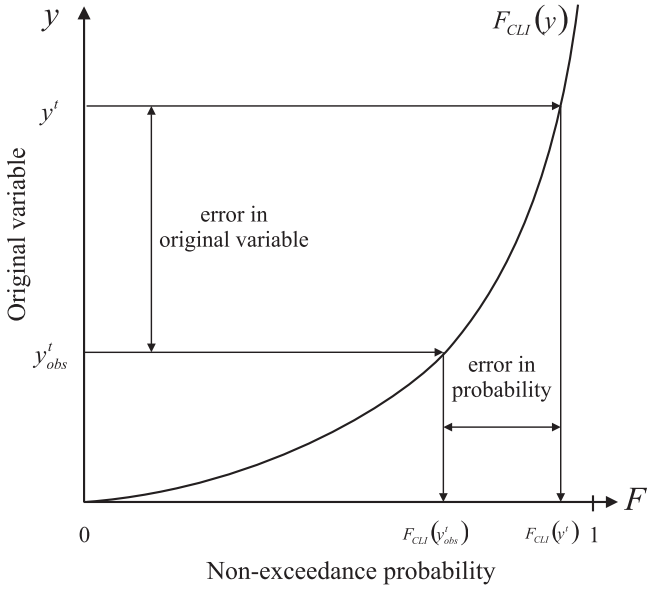

Figure C1. Schematic of forecast error in probability corresponding to error on the original variable scale.

[86] **Acknowledgments.** This research has been supported by the Water Information Research and Development Alliance between the Australian Bureau of Meteorology and CSIRO Water for a Healthy Country Flagship, the South Eastern Australian Climate Initiative, and the CSIRO OCE Science Leadership Scheme. We would like to thank Neil Plummer, Jeff Perkins, Senlin Zhou, Andrew Schepen, Trudy Peatey, Daehyok Shin, Andrew Frost, and Sri Srikanthan from the Australian Bureau of Meteorology for many valuable discussions and suggestions as well as providing the streamflow and rainfall data for this study. We are grateful to Elizabeth Ebert and Robert Fawcett also from the Australian Bureau of Meteorology for valuable discussions on the new root-mean-square error in probability (RMSEP) skill score and for sharing with us their knowledge on forecast verification. We would also like to acknowledge the thorough reviews by David Post from CSIRO and three anonymous reviewers.

References

Bardossy, A., and E. Plate (1992), Space-time model for daily rainfall using atmospheric circulation patterns, *Water Resour. Res.*, *28*, 1247–1260.

Beare, S. C., et al. (2003), Natural resource management in the Burdekin River Catchment: Integrated assessment of resource management at the catchment scale: A case study, 66 pp., ABARE, Canberra.

Bracken, C., B. Rajagopalan, and J. Prairie (2010), A multisite seasonal ensemble streamflow forecasting technique, *Water Resour. Res.*, *46*, W03532, doi:10.1029/2009WR007965.

Charles, S. P., B. C. Bates, and J. P. Hughes (1999), A spatiotemporal model for downscaling precipitation occurrence and amounts, *J. Geophys. Res.*, *104*, 31,657–31,669, doi:10.1029/1999JD900119.

Chiew, F. H. S., S. L. Zhou, and T. A. McMahon (2003), Use of seasonal streamflow forecasts in water resources management, *J. Hydrol.*, *270*, 135–144.

Department of Environment and Resource Management (2009), Burdekin Basin Resource Operations Plan, 130 pp., Dep. of Environ. and Resour. Manage., Brisbane Australia.

Devlin, S. J., R. Gnanadesikan, and J. R. Kettenring (1975), Robust estimation and outlier detection with correlation-coefficients, *Biometrika*, *62*, 531–545.

Drosdowsky, W., and L. E. Chambers (2001), Near global sea surface temperature anomalies as predictors of Australian seasonal rainfall, *J. Clim.*, *14*, 1677–1687.

Fawcett, R. J. B. (2008), Verification of the Bureau of Meteorology's seasonal forecasts: 2003–2005, *Aust. Meteorol. Mag.*, *57*, 273–278.

Fawcett, R. J. B., D. A. Jones, and G. S. Beard (2005), A verification of publicly issued seasonal forecasts issued by the Australian Bureau of Meteorology: 1998–2003, *Aust. Meteorol. Mag.*, *54*, 1–13.

Fisher, R. (1921), On the "probable error" of a coefficient of correlation deduced from a small sample, *Metron*, *1*, 32.

Frost, A. J., M. A. Thyer, R. Srikanthan, and G. Kuczera (2007), A general Bayesian framework for calibrating and evaluating stochastic models of annual multi-site hydrological data, *J. Hydrol.*, *340*, 129–148.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004), Bayesian data analysis, 2nd ed., in *Texts in Statistical Science Series*, 668 pp., Chapman and Hall, Boca Raton, Fla.

Genz, A. (1993), Comparison of methods for the computation of multivariate normal probabilities, in *Computing Science and Statistics*, vol. 25, *Statistical Applications of Expanding Computer Capabilities*, edited by M. E. Tarter and M. D. Lock, pp. 400–405, Interface Found. North Am., Fairfax, Va.

Gneiting, T., F. Balabdaoui, and A. E. Raftery (2007), Probabilistic forecasts, calibration and sharpness, *J. R. Stat. Soc., Ser. B*, *69*, 243–268.

Jobson, J. D. (1992), *Applied Multivariate Data Analysis*, vol. 2, *Categorical and Multivariate Methods*, 156 pp., Springer, New York.

Laio, F., and S. Tamea (2007), Verification tools for probabilistic forecasts of continuous hydrological variables, *Hydrol. Earth Syst. Sci.*, *11*, 1267–1277.

Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models, 1. A discussion of principles, *J. Hydrol.*, *10*, 282–290.

Plummer, N., N. K. Tuteja, Q. J. Wang, E. Wang, D. E. Robertson, S. Zhou, A. Schepen, O. Alves, B. Timbal, and K. Puri (2009), A seasonal water availability prediction service: Opportunities and Challenges, in *18th World IMACS/MODSIM Congress*, Modell. and Simul. Soc. of Aust. and N. Z. Cairns, Queensland, Australia.

Potts, J. M., C. K. Folland, I. T. Jolliffe, and D. Sexton (1996), Revised "LEPS" scores for assessing climate model simulations and long-range forecasts, *J. Clim.*, *9*, 34–53.

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

Srikanthan, R. and G. Pegram (2009), A nested multisite daily rainfall stochastic generation model, *J. Hydrol.*, *371*, 142–153.

Thyer, M., G. Kuczera, Q. J. Wang (2002), Quantifying parameter uncertainty in stochastic models using the Box-Cox transformation, *J. Hydrol.*, *265*, 246–257.

Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and S. Srikanthan (2009), Critical evaluation of parameter consistency and predictive uncertainty in hydrological modelling: A case study using Bayesian total error analysis, *Water Resour. Res.*, *45*, W00B14, doi:10.1029/2008WR006825.

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.

Ward, M. N., and C. K. Folland (1991), Prediction of seasonal rainfall in the Nordeste of Brazil using eigenvectors of sea-surface temperature, *J. Climatol.*, *11*, 711–743.

Westra, S., C. Brown, U. Lall, and A. Sharma (2007), Modeling multivariable hydrological series: Principal component analysis or independent component analysis?, *Water Resour. Res.*, *43*, W06429, doi:10.1029/2006WR005617.

Wilks, D. S. (1995), *Statistical Methods in the Atmospheric Sciences*, 648 pp., Elsevier, New York.

Wilks, D. S. (1998), Multisite generalization of a daily stochastic precipitation generation model, *J. Hydrol.*, *210*, 178–191.

Yeo, I., and R. A. Johnson (2000), A new family of power transformations to improve normality or symmetry, *Biometrika*, *87*, 954–959.

Zhu, D., and A. O. Hero (2007), Bayesian hierarchical model for large-scale covariance matrix estimation, *J. Comput. Biol.*, *14*, 1311–1326.

D. E. Robertson and Q. J. Wang, CSIRO Land and Water, PO Box 56, High-ett, Vic 3190, Australia. (qj.wang@csiro.au)