A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites

Q. J. Wang,[1] D. E. Robertson,[1] and F. H. S. Chiew[2]

Received 12 August 2008; revised 27 January 2009; accepted 2 March 2009; published 9 May 2009.

[1] Seasonal forecasting of streamflows can be highly valuable for water resources management. In this paper, a Bayesian joint probability (BJP) modeling approach for seasonal forecasting of streamflows at multiple sites is presented. A Box-Cox transformed multivariate normal distribution is proposed to model the joint distribution of future streamflows and their predictors such as antecedent streamflows and El Niño-Southern Oscillation indices and other climate indicators. Bayesian inference of model parameters and uncertainties is implemented using Markov chain Monte Carlo sampling, leading to joint probabilistic forecasts of streamflows at multiple sites. The model provides a parametric structure for quantifying relationships between variables, including intersite correlations. The Box-Cox transformed multivariate normal distribution has considerable flexibility for modeling a wide range of predictors and predictands. The Bayesian inference formulated allows the use of data that contain nonconcurrent and missing records. The model flexibility and data-handling ability means that the BJP modeling approach is potentially of wide practical application. The paper also presents a number of statistical measures and graphical methods for verification of probabilistic forecasts of continuous variables. Results for streamflows at three river gauges in the Murrumbidgee River catchment in southeast Australia show that the BJP modeling approach has good forecast quality and that the fitted model is consistent with observed data.

Citation: 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.

1. Introduction

[2] Reliable forecasts of streamflows several months or seasons ahead are highly valuable to water resources managers and water users [*Chiew et al.*, 2003]. One approach for seasonal forecasting of streamflows is to run dynamic climate models to produce forecasts of rainfall and other weather variables, which are then fed into hydrological models to produce forecasts of streamflows. However, the skill of rainfall forecasts by dynamic climate models is still low for lead time longer than 1 month. Another approach is to use statistical relationships derived directly from observed data. Many climate indicators based on atmospheric pressure and sea temperature have been linked to future seasonal rainfalls. For example, sea surface temperature anomalies in the Pacific and Indian oceans have been found to correlate with rainfall in Australia [*Ashok et al.*, 2003; *Drosdowsky and Chambers*, 2001; *Hendon et al.*, 2007; *Murphy and Timbal*, 2008; *Ruiz et al.*, 2006; *Saji et al.*, 1999; *Verdon and Franks*, 2005]. Such correlations have also been found with streamflows [e.g., *Chiew and McMahon*, 2002; *Ruiz et al.*, 2007]. These observed correlations together with observed persistence in streamflows have led to the development of a number of statistical approaches for seasonal forecasting of streamflows. *Piechota et al.* [2001], *Sharma* [2000], and F. H. S. Chiew and L. W. Siridardena (Probabilistic seasonal streamflow forecasting methods, paper presented at 29th Hydrology and Water Resources Symposium, Institution of Engineers, Australia, Canberra, Australia, 2005) introduced approaches for single-site seasonal streamflow forecasting. A. Sharma et al. (Multiple reservoir probabilistic inflow forecasting system: Application to the Hydro-Tasmania reservoir network, paper presented at 30th Hydrology and Water Resources Symposium, Institution of Engineers, Launceston, Tasmania, Australia, 4–6 December 2006) extended a nonparametric single-site approach to forecasting streamflows at multiple sites, while *Westra et al.* [2008] proposed the use of independent component analysis for dealing with spatially correlated streamflows. None of the available approaches allow the use of data that contain nonconcurrent and missing records.

[3] This paper introduces a Bayesian joint probability (BJP) modeling approach for seasonal forecasting of streamflows at multiple sites. A Box-Cox transformed [*Box and Cox*, 1964; *Yeo and Johnson*, 2000] multivariate normal distribution is used to model the joint distribution of future streamflows and their predictors, such as antecedent streamflows and El Niño-Southern Oscillation indices and other climate indicators. A Bayesian inference of model parameters and their uncertainties is implemented using Markov chain Monte Carlo (MCMC) sampling. Joint probabilistic forecasts of streamflows at multiple sites are generated.

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

0043-1397/09/2008WR007355

[4] The motivations for developing the BJP modeling approach are as follows: (1) The use of an explicit model structure (through its parametric form) facilitates model testing and further improvement, and assists transferring of learning acquired through applications within a region and across regions. (2) The Box-Cox transformed multivariate normal distribution provides a simple structure to model intersite correlations while possessing considerable flexibility to model a wide range of potentially useful predictors and predictands. (3) The statistical inference method allows data to contain nonconcurrent and missing records, making full use of all available information without the need to truncate the data to concurrent series or infill missing data points. In practice, many model variables (e.g., streamflows at multiple sites) often do not have fully contiguous records, and therefore the model must be able to handle such data. The approach can be particularly useful for including those predictors that have much shorter records than other predictors. For example, modeled quantities from dynamic climate models can be good candidate predictors for use alongside directly observable predictors. However, available hindcast data from many dynamic climate models often do not extend to the same record lengths as observed predictors.

[5] This paper gives a detailed description of model formulation, parameter inference, model use for forecasting, forecast verification, and model checking. The paper includes an application of the BJP modeling approach to forecasting streamflows at three river gauges in the Murrumbidgee River catchment in southeast Australia.

2. Model Formulation

[6] Consider d variables that consist of both seasonal totals of streamflows to be forecast and their predictors such as climate and catchment indicators

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

where $\mathbf{y}$ is a column vector and $\mathbf{y}^{\mathrm{T}}$ is its transpose. Each of the variables, y_i, $i = 1,2, \ldots, d$, is normalized through the extended form of the Box-Cox transform introduced by *Yeo and Johnson* [2000]

$$z = \begin{cases} \left\{(y+1)^{\lambda} - 1\right\}/\lambda & (y \geq 0, \lambda \neq 0) \\ \log(y+1) & (y \geq 0, \lambda = 0) \\ -\left\{(-y+1)^{2-\lambda} - 1\right\}/(2-\lambda) & (y < 0, \lambda \neq 2) \\ \log(-y+1) & (y < 0, \lambda = 2) \end{cases} \tag{2}$$

For simplicity, the subscript i has been omitted from the above equation. (Various subscripts are also omitted from a number of equations throughout the paper for the same reason. It should be obvious where such omissions are made.) The advantage of this extended form of Box-Cox transform is that y does not need to be restricted to a positive value, avoiding the need to introduce a shift parameter otherwise required by the original Box-Cox transform [*Box and Cox*, 1964] for situations where y goes negative. The transform coefficient in equation (2) is allowed to vary for different variables

$$\boldsymbol{\lambda}^{\mathrm{T}} = [\lambda_1\ \lambda_2\ \cdots\ \lambda_d] \tag{3}$$

giving

$$\mathbf{z}^{\mathrm{T}} = [z_1\ z_2\ \cdots\ z_d] \tag{4}$$

The vector $\mathbf{z}$ is assumed to follow a multivariate normal distribution

$$\mathbf{z} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{5}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean vector and covariance matrix, respectively:

$$\boldsymbol{\mu}^{\mathrm{T}} = [\mu_1\ \mu_2\ \cdots\ \mu_d] \tag{6}$$

$$\boldsymbol{\Sigma} = \boldsymbol{\sigma}\mathbf{R}\boldsymbol{\sigma}^{\mathrm{T}} \tag{7}$$

and $\boldsymbol{\sigma}$ and $\mathbf{R}$ are the standard deviation vector and correlation coefficient matrix, respectively:

$$\boldsymbol{\sigma}^{\mathrm{T}} = [\sigma_1\ \sigma_2\ \cdots\ \sigma_d] \tag{8}$$

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1d} \\ r_{21} & 1 & \cdots & r_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ r_{d1} & r_{d2} & \cdots & 1 \end{bmatrix} \tag{9}$$

noting that $r_{ij} = r_{ji}$ for variables indexed i and j. Assuming a steady relationship among the variables, the model has a total of $3d + (d - 1)\ d/2$ unknown parameters.

[7] The assumption of a multivariate normal distribution ignores the theoretical bounds for z introduced by the extended Box-Cox transform. When $\lambda < 0$

$$z \to \begin{cases} -1/\lambda & (\text{as } y \to +\infty) \\ -\infty & (\text{as } y \to -\infty) \end{cases} \tag{10a}$$

and when $\lambda > 2$

$$z \to \begin{cases} +\infty & (\text{as } y \to +\infty) \\ 1/(2-\lambda) & (\text{as } y \to -\infty) \end{cases} \tag{10b}$$

This assumption is rarely unacceptable as z values of observed data, and thus the range of z of interest, is usually well within the above bounds in practical applications. It is

a standard practice in statistical modeling to use theoretical distributions that are strictly incorrect but are considered appropriate for the variable ranges of interest. For example, one may use the normal distribution to model the probability of outdoor temperature. The outdoor temperature has a physical range, but the normal variable has an unlimited range $(-\infty, \infty)$. However, the normal model may still be considered appropriate as long as it is not overly extrapolated to ranges much beyond observed values. Problems can arise if models are over extrapolated. This will be discussed in section 4 in the context of generating forecast probability distributions.

[8] A more serious bound-related issue can arise when y is subject to a physical bound (such as streamflow is greater or equal to zero) and has been observed to fall on or near that bound. In such a case, the variable bound must be appropriately taken into account. The model formulation in this paper is for nonephemeral streams, for which the zero flow bound does not come into effect. Seasonal forecasting of flows in ephemeral streams will be dealt with in a future paper.

3. Model Parameter Inference

3.1. Reparameterization

[9] Before setting out for statistical inference of the unknown parameters, a number of parameters are reparameterized to ease the statistical inference. Parameters μ and σ are reparameterized to m and s by

$$\mu = \begin{cases} \left\{(m+1)^{\lambda} - 1\right\}/\lambda & (m \geq 0, \lambda \neq 0) \\ \log(m+1) & (m \geq 0, \lambda = 0) \\ -\left\{(-m+1)^{2-\lambda} - 1\right\}/(2-\lambda) & (m < 0, \lambda \neq 2) \\ \log(-m+1) & (m < 0, \lambda = 2) \end{cases} \tag{11}$$

$$\sigma = \begin{cases} (m+1)^{\lambda-1} s & (m \geq 0) \\ (-m+1)^{1-\lambda} s & (m < 0) \end{cases} \tag{12}$$

to give

$$\mathbf{m}^{\mathrm{T}} = [m_1 \; m_2 \; \cdots \; m_d] \tag{13}$$

$$\mathbf{s}^{\mathrm{T}} = [s_1 \; s_2 \; \cdots \; s_d] \tag{14}$$

For ease of prior specification to be described later in section 3.3, s^2 is to be used as a parameter instead of s. The above reparameterization follows the treatment by *Thyer et al.* [2002] of the original Box-Cox transform. Parameter m is an approximation of the mean of a nontransformed variable y. Parameter s^2 is an approximation of the variance of y [*Yeo and Johnson*, 2000]. This reparameterization is to resolve the problem that parameters μ and σ^2 after the Box-Cox transform are highly correlated with λ and difficult for statistical inference [*Thyer et al.*, 2002].

[10] The correlation coefficient matrix $\mathbf{R}$ of equation (9) is reparameterized through an inverse hyperbolic tangent transform (or Fisher Z transform) to give

$$\mathbf{\Phi} = \begin{bmatrix} \infty & \varphi_{12} & \cdots & \varphi_{1d} \\ \varphi_{21} & \infty & \cdots & \varphi_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{d1} & \varphi_{d2} & \cdots & \infty \end{bmatrix} \tag{15}$$

where

$$\varphi_{ij} = \tanh^{-1}\left(r_{ij}\right) \tag{16}$$

The reparameterized correlation coefficients φ_{ij} are asymptotically normal [*Zhu and Hero*, 2007] and thus more convenient for parameter and uncertainty inference than the original correlation coefficients r_{ij} which range between $[-1,1]$. This will be further elaborated in section 3.4.

3.2. Posterior Distribution of Parameters

[11] 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. A Bayesian statistical inference is used here. According to the Bayes theorem, given historically observed cases $\mathbf{y}^t$ for year $t = 1,2, \ldots, n$, the posterior distribution of the model parameters is

$$\begin{aligned} p\left(\boldsymbol{\theta}|\mathbf{y}^n, \mathbf{y}^{n-1}, \ldots, \mathbf{y}^1\right) &\propto p(\boldsymbol{\theta}) p\left(\mathbf{y}^n, \mathbf{y}^{n-1}, \ldots, \mathbf{y}^1|\boldsymbol{\theta}\right) \\ &= p(\boldsymbol{\theta}) \textstyle\prod_{t=1}^{n} p(\mathbf{y}^t|\boldsymbol{\theta}) \end{aligned} \tag{17}$$

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 cases $\mathbf{y}^t$, $t = 1,2, \ldots, n$, given the model and its parameter set. The multiplication of the likelihood function $p(\mathbf{y}^t|\boldsymbol{\theta})$ for all the cases $t = 1,2, \ldots, n$ implies the assumption of independent cases. For forecasting streamflow totals over the next 3–6 months starting from a fixed month, the assumption of independent cases is considered reasonable. Future work will extend this approach to joint forecasting of streamflows for multiple seasons.

[12] The likelihood function for each case $\mathbf{y}^t$ can be expressed as

$$p(\mathbf{y}^t|\boldsymbol{\theta}) = J_{\mathbf{z}\to\mathbf{y}}|^t p(\mathbf{z}^t|\boldsymbol{\theta}) = \textstyle\prod_{i=1}^{d} dz_i/dy_i|^t p(\mathbf{z}^t|\boldsymbol{\theta}) \tag{18}$$

where $J_{\mathbf{z}\to\mathbf{y}}$ is the Jacobian determinant for the transform from $\mathbf{z}$ to $\mathbf{y}$,

$$dz_i \Big/ dy_i = \begin{cases} (y_i+1)^{\lambda_i - 1} & (y_i \geq 0) \\ (-y_i+1)^{1-\lambda_i} & (y_i < 0) \end{cases} \tag{19}$$

and d is the number of variables as defined in section 2.

[13] When the values for some of the elements in the vector $\mathbf{y}^t$ are missing, a subset of equations (18) and (5)

corresponding to the nonmissing elements are used in calculating $p(\mathbf{y}^t|\boldsymbol{\theta})$. This is based on the fact that if $\mathbf{z}$ follows a multivariate normal distribution, a subset of $\mathbf{z}$ is also normally distributed with its mean vector and covariance matrix equal to their respective subsets of the mean vector and covariance matrix for $\mathbf{z}$. When the values for all the elements in the vector $\mathbf{y}^t$ are missing, the time step t is omitted from equation (17). Thus, missing and nonconcurrent data are easily accommodated in the formulation of the posterior distribution of the model parameters.

3.3. Prior Distribution of Parameters

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

$$p(\boldsymbol{\theta}) = \prod_{i=1}^{d} p(\lambda_i) p\left(m_i, s_i^2\right) p(\boldsymbol{\Phi}) \tag{20}$$

A uniform prior with a range of $[-2, 2]$ is used for each λ. A more elaborate prior is derived for each pair of (m,s^2) to deal with the effect of the Box-Cox transform (equation (2)) and reparameterization (equations (11) and (12))

$$p\left(m, s^2\right) = J_{\mu,\sigma^2\to m,s^2} p\left(\mu, \sigma^2\right) = J_{\mu,\sigma^2\to m,s^2} p\left(\mu|\sigma^2\right) p\left(\sigma^2\right) \tag{21}$$

where $J_{\mu,\sigma^2\to m,s^2}$ is the Jacobian determinant of the transform from (μ, σ^2) to (m, s^2)

$$J_{\mu,\sigma^2\to m,s^2} = \begin{vmatrix} \partial\mu/\partial m & \partial\mu/\partial s^2 \\ \partial\sigma^2/\partial m & \partial\sigma^2/\partial s^2 \end{vmatrix} = \begin{cases} (m+1)^{3\lambda-3} & (m \geq 0) \\ (-m+1)^{3-3\lambda} & (m < 0) \end{cases} \tag{22}$$

and

$$p\left(\mu|\sigma^2\right) = N\left(\mu_0,\ \sigma^2/\kappa_0\right) \qquad \left(\kappa_0 = 1,\ \mu_0 = \mu(\hat{m})\right) \tag{23}$$

$$p\left(\sigma^2\right) = \text{Inverse} - \chi^2\left(\nu_0, \sigma_0^2\right) \qquad \left(\nu_0 = 2,\ \sigma_0^2 = \sigma^2\left(\hat{m}, \hat{s}^2\right)\right) \tag{24}$$

$\hat{m}$ and $\hat{s}^2$ being the sample mean and variance of y, respectively, and functions $\mu(\cdot)$ and $\sigma^2(\cdot,\cdot)$ being based on equations (11) and (12), respectively. The rationale for this prior is that the classical prior of equations (23) and (24) [*Gelman et al.*, 1995] derived for the mean and variance of a normal distribution is appropriate here only for (μ, σ^2) of the normalized variable z, and this prior needs to be projected to the reparameterized space to give an appropriate prior for (m, s^2). Even though the prior for (μ, σ^2) is quite diffuse because of the choice of values for κ_0 and ν_0 as in equations (23) and (24), the projected prior for (m, s^2) is more defined depending on the values of m and λ. In the course of this study, it was found that using a highly diffused prior directly on (m, s^2) resulted in significantly biased model inference when data exhibited strong nonnormality. This led to the derivation of the prior as shown in equations (21)–(24).

[15] The prior for the reparameterized correlation coefficient matrix $\boldsymbol{\Phi}$ of equation (15) is related to the prior for the original correlation coefficient matrix $\mathbf{R}$ of equation (9) by

$$p(\boldsymbol{\Phi}) = J_{\mathbf{R}\to\boldsymbol{\Phi}} p(\mathbf{R}) = \prod_{i=1}^{d-1} \prod_{j=i+1}^{d} dr_{ij}/d\varphi_{ij} p(\mathbf{R}) \tag{25}$$

where $J_{\mathrm{R}\to\Phi}$ is the Jacobian determinant for the transform from $\mathbf{R}$ to $\boldsymbol{\Phi}$, and

$$dr_{ij}\Big/d\varphi_{ij} = \left[\cosh\left(\varphi_{ij}\right)\right]^{-2} \tag{26}$$

[16] One of the candidate priors considered for the correlation matrix $\mathbf{R}$ is a jointly uniform prior:

$$p(\mathbf{R}) \propto 1 \tag{27}$$

Because a correlation matrix is valid only if it is positive semidefinite, this jointly uniform prior leads to marginal priors for individual correlations, r_{ij}, that are not uniform. In fact, the marginal priors can be highly defined, especially for high-dimension d, favoring values close to zero over values close to ±1 [*Barnard et al.*, 2000]. For this reason, this jointly uniform prior is not adopted. Instead, the marginally uniform prior for the correlation matrix [*Barnard et al.*, 2000] is used as

$$p(\mathbf{R}) \propto |\mathbf{R}|^{d(d-1)/2-1} \left(\prod_{i=1}^{d} |\mathbf{R}_{ii}|\right)^{-(d+1)/2} \tag{28}$$

where $\mathbf{R}_{ii}$ is the ith principal submatrix of $\mathbf{R}$.

3.4. Sampling of Parameters

[17] The posterior distribution of the parameters, as expressed by equation (17), is a nonstandard distribution and does not allow analytical integration. Therefore, numerical solution is necessary for probabilistic forecasting. 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 [e.g., *Frost et al.*, 2007; *Gelman et al.*, 1995; *Thyer et al.*, 2002; *Wang*, 2008]. A visual inspection of the sampled parameter chains is made to monitor the convergence of the sample [*Cowles and Carlin*, 1996].

[18] The technique of Metropolis MCMC sampling is well developed, but its implementation allows considerable variation in specifics. Two aspects of the Metropolis MCMC sampling implemented in this study are noted here. MCMC sampling requires an initialization of parameter values to start off sampling, By carrying out sufficiently long warming-up runs as practiced in standard MCMC sampling, the actual initial values used should have little impact on the final sampling results. Nonetheless, a feasible point is still needed to get MCMC sampling started, and ideally the point is of a reasonable probability density to speed up the warming up. In this study, an initial value of 0.2 is assigned to all the Box-Cox transform coefficients λ. Moment estimates of the means (m), variances (s^2) and correlation coefficients (r_{ij}) from nontransformed data series are used. Where missing (including nonconcurrent) data are involved, the correlation coefficient of any two data series is

estimated from only pairs of nonmissing data. However, the resulting correlation matrix **R** can sometimes be invalid because it is not positive semidefinite. This problem is more likely to occur when some of the data series are highly correlated with each other, missing and nonconcurrent data are prevalent, or the dimension d is high. A solution to this problem is to find the nearest correlation matrix that is positive semidefinite and use it as the initial correlation matrix for the MCMC sampling. A modified alternating projection method is used in this study [*Higham*, 2002; *Monserrat et al.*, 2007; *Xiong et al.*, 2007]. As reparameterization has been applied to the correlation coefficients, the initial values for r_{ij} are used to set the initial values for φ_{ij} by applying equation (16).

[19] MCMC sampling also requires the use of a proposal distribution for generating random jumps in the parameter space [*Gelman et al.*, 1995]. The multivariate normal distribution is the most commonly used proposal distribution and is adopted in this study. It has been pointed out in section 3.1 that the reparameterized correlation coefficients φ_{ij} are asymptotically normal. They are thus consistent with the use of the multivariate normal proposal distribution, unlike the original correlation coefficients r_{ij}. This reparameterization leads to much more efficient MCMC sampling as confirmed in the course of this study.

4. Model Use for Probabilistic Forecasting

[20] Now separate the **y** vector into two subvectors

$$\mathbf{y}^{\mathrm{T}} = \left[\mathbf{y}^{\mathrm{T}}(1) \quad \mathbf{y}^{\mathrm{T}}(2)\right] \tag{29}$$

where **y**(1) consists of the predictors

$$\mathbf{y}^{\mathrm{T}}(1) = [y_1\ y_2\ \cdots\ y_{d_1}] \tag{30}$$

and **y**(2) consists of the predictands

$$\mathbf{y}^{\mathrm{T}}(2) = [y_{d_1+1}\ y_{d_1+2}\ \cdots\ y_d] \tag{31}$$

Accordingly the variable vector after the Box-Cox transform, **z**, is organized as

$$\mathbf{z}^{\mathrm{T}} = \left[\mathbf{z}^{\mathrm{T}}(1)\ \mathbf{z}^{\mathrm{T}}(2)\right] \tag{32}$$

and the mean vector and covariance matrix of **z** as in equations (6) and (7) are expressed respectively as

$$\boldsymbol{\mu}^{\mathrm{T}} = \left[\boldsymbol{\mu}^{\mathrm{T}}(1)\ \boldsymbol{\mu}^{\mathrm{T}}(2)\right] \tag{33}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}(1,1) & \boldsymbol{\Sigma}(1,2) \\ \boldsymbol{\Sigma}(2,1) & \boldsymbol{\Sigma}(2,2) \end{bmatrix} \tag{34}$$

Given model parameters $\boldsymbol{\theta}$, a forecasting problem is stated as finding the predictands **y**(2) given the predictors **y**(1). Given the multivariate normal distribution of equation (5), it can be shown that

$$p(\mathbf{z}(2)|\mathbf{z}(1), \boldsymbol{\theta}) = \mathrm{N}(\boldsymbol{\mu}'(2), \boldsymbol{\Sigma}'(2,2)) \tag{35}$$

where

$$\boldsymbol{\mu}'(2) = \boldsymbol{\mu}(2) + \boldsymbol{\Sigma}(2,1)[\boldsymbol{\Sigma}(1,1)]^{-1}[\mathbf{z}(1) - \boldsymbol{\mu}(1)] \tag{36}$$

$$\boldsymbol{\Sigma}'(2,2) = \boldsymbol{\Sigma}(2,2) - \boldsymbol{\Sigma}(2,1)[\boldsymbol{\Sigma}(1,1)]^{-1}\boldsymbol{\Sigma}(1,2) \tag{37}$$

To obtain probabilistic forecasts of streamflows, the multivariate normal distribution of equation (35) can be used to generate a sample, say one set, of **z**(2) for a given set of parameters $\boldsymbol{\theta}$. An inverse of the extended Box-Cox transform (equation (2)) is then applied to give the forecast values

$$y = \begin{cases} (\lambda z + 1)^{1/\lambda} - 1 & (z \geq 0, \lambda \neq 0) \\ \exp(z) - 1 & (z \geq 0, \lambda = 0) \\ -\{-(2-\lambda)z + 1\}^{1/(2-\lambda)} + 1 & (z < 0, \lambda \neq 2) \\ -\exp(-z) + 1 & (z < 0, \lambda = 2) \end{cases} \tag{38}$$

This procedure is repeated for all the parameter sets in the representative parameter sample previously obtained through MCMC sampling as described in section 3.4. The total collection of all the forecast values provides a numerical representation of the probabilistic forecast of the streamflows.

[21] Note that when data for some of the predictors in **y**(1) are not available, a subset of equations (32)–(38) excluding the elements that are related to the missing predictors can be used to forecast **y**(2). This is again based on the fact that if **z** follows a multivariate normal distribution, a subset of **z** is also normally distributed with its mean vector and covariance matrix equal to their respective subsets of the mean vector and covariance matrix for **z**. Thus, forecasts can be issued uninterrupted even when data for some of the predictors are not available. The forecasts are however expected to show higher uncertainty than if data for all the predictors are available.

[22] On rare occasions, the generated z values can approach or go beyond the bounds of equation (10), resulting in unrealistically large, low or even undefined y streamflows according to equation (38). This problem is caused by over extrapolation of the model to distribution tails when very large samples of z are generated. The problem is dealt with here by specifying a feasible range for y. When y values corresponding to the generated z are outside the specified range, they are recorded only as having values equal or above the upper range or equal or below the lower range. The exact y values are ignored. To use the same outdoor temperature analogy as in section 2, values generated from the normal distribution that describes the outdoor temperature may on occasions be unrealistically low or high because of over extrapolation of the normal model. These unrealistic values may be considered only as very cold or very hot events. Their exact temperature values are ignored. Note that the specification of a variable range for data generation here does not affect model inference (described in section 3.2) as long as the observed data values are within the range. In the application to be described in section 7, less than 1 in 1000 generated values

was considered unrealistically high or low. If for some reasons, there is a large proportion of unrealistic values generated, one needs to consider the use of models suitable for truncated variables [*Frost et al.*, 2007]. Seasonal forecasting of flows in ephemeral streams, which requires such a treatment, will be dealt with in a future paper.

5. Forecast Verification

[23] 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 case. The streamflows for the left-out case 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 cases $t = 1,2, \ldots, n$. Paired with these forecasts are the observed values y^t_{OBS}.

[24] 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 adopted, adapted or devised to provide both overall and detailed verifications. For overall verifications, two skill scores are used, one based on the linear error in probability space score (LEPS) and another based on the continuous ranked probability score (CRPS). In addition, a probability plot is used to show the overall uniformity of the forecast probability integral transform (PIT) of the observed streamflows.

5.1. Skill Score Based on LEPS

[25] The linear error in probability space (LEPS) skill score is one of the verification measures that have been used for weather forecasts [*Ward and Folland*, 1991; *Potts et al.*, 1996]. For a continuous forecast variable y, the LEPS score was originally formulated for single value (or point) forecasts y^t, $t = 1,2, \ldots, n$, rather than for probabilistic forecasts. It measures the errors between the forecast and observed values in the space of a reference (usually climatology based) cumulative probability distribution $F_{REF}(y)$. In other words, the errors are in terms of the reference probability integral transform (PIT) of the forecast and observed values, $F_{REF}(y^t)$ and $F_{REF}(y^t_{OBS})$. The (revised) LEPS score introduced by *Potts et al.* [1996] is given by

$$LEPS^t(y^t) = 3\Big\{1 - \big|F_{REF}(y^t) - F_{REF}\big(y^t_{OBS}\big)\big| + [F_{REF}(y^t)]^2 - F_{REF}(y^t) + \big[F_{REF}\big(y^t_{OBS}\big)\big]^2 - F_{REF}\big(y^t_{OBS}\big)\Big\} - 1 \tag{39}$$

This LEPS score is adapted here for a probabilistic forecast $y^t \sim F^t(y^t)$ (or in probability density $y^t \sim f^t(y^t)$) by taking expectation

$$LEPS^t = \int_{-\infty}^{\infty} LEPS^t(y^t) f^t(y^t) dy^t = \int_0^1 LEPS^t(y^t) dF^t(y^t) \tag{40}$$

Note that equation (39) already includes a normalization, giving a zero LEPS score for a climatology-based reference probabilistic forecast.

[26] A skill score based on the LEPS score is given by [*Ward and Folland*, 1991; *Potts et al.*, 1996]

$$SS_{LEPS} = \overline{LEPS^t} \Big/ \overline{LEPS^t_m} \tag{41}$$

The overbar denotes averaging over the forecast cases. When $\overline{LEPS^t} \geq 0$, $LEPS^t_m$ is the LEPS score for a perfect single value forecast $y^t \equiv y^t_{OBS}$. When $\overline{LEPS^t} < 0$, $LEPS^t_m$ is the worst possible LEPS score given by either $F_{REF}(y^t) \equiv 0$ or $F_{REF}(y^t) \equiv 1$. Thus, this skill score is scaled to -100% to 100%.

5.2. Skill Score Based on CRPS

[27] The CRPS was formulated for verification of probabilistic forecasts of continuous variables [e.g., *Brown*, 1974; *Matheson and Winkler*, 1976; *Bouttier*, 1994; *Hersbach*, 2000; *Gneiting et al.*, 2007; *Laio and Tamea*, 2007; D. A. Unger, A method to estimate the continuous ranked probability score, paper presented at Ninth Conference on Probability and Statistics in Atmospheric Sciences, American Meteorological Society, Virginia Beach, Virginia, 1985]. It expresses some kind of distance between a probabilistic forecast $y^t \sim F^t(y^t)$ and the observed value y^t_{OBS} and is defined as

$$CRPS^t = \int_{-\infty}^{\infty} \big[F^t(y^t) - H^t\big(y^t - y^t_{OBS}\big)\big]^2 dy^t \tag{42}$$

where

$$H\big(y^t - y^t_{OBS}\big) = \begin{cases} 0 & \big(y^t < y^t_{OBS}\big) \\ 1 & \big(y^t \geq y^t_{OBS}\big) \end{cases} \tag{43}$$

is the Heaviside function, representing here the observed cumulative probability distribution. The CRPS is a generalization of the ranked probability score for probabilistic forecasts of ordinal multicategory variables [e.g., *Epstein*, 1969; *Murphy*, 1969, 1971]. The ranked probability score is in turn a generalization of the well-known Brier score for probabilistic forecasts of two-category variables. For a deterministic (single value or point) forecast, the CRPS is equal to the absolute error. The minimum CRPS value of zero is achieved only in the case of a perfect single value forecast.

[28] A skill score based on the CRPS is defined as

$$SS_{CRPS} = \Big(\overline{CRPS^t} - \overline{CRPS^t_{REF}}\Big) \Big/ \Big(0 - \overline{CRPS^t_{REF}}\Big) \tag{44}$$

where the overbar again denotes averaging over the forecast cases, $CRPS^t_{REF}$ is the CRPS when a reference (usually climatology based) probabilistic forecast $-y^t \sim F_{REF}(y^t)$ is used, and 0 is the perfect forecast CRPS. It is obvious that the skill score for reference probabilistic forecasts is zero.

5.3. PIT Uniform Probability Plot

[29] Given a probabilistic forecast $y^t \sim F^t(y^t)$, a probability integral transform can be applied to the observed value y^t_{OBS} to give

$$\pi^t = F^t\big(y^t_{OBS}\big) \tag{45}$$

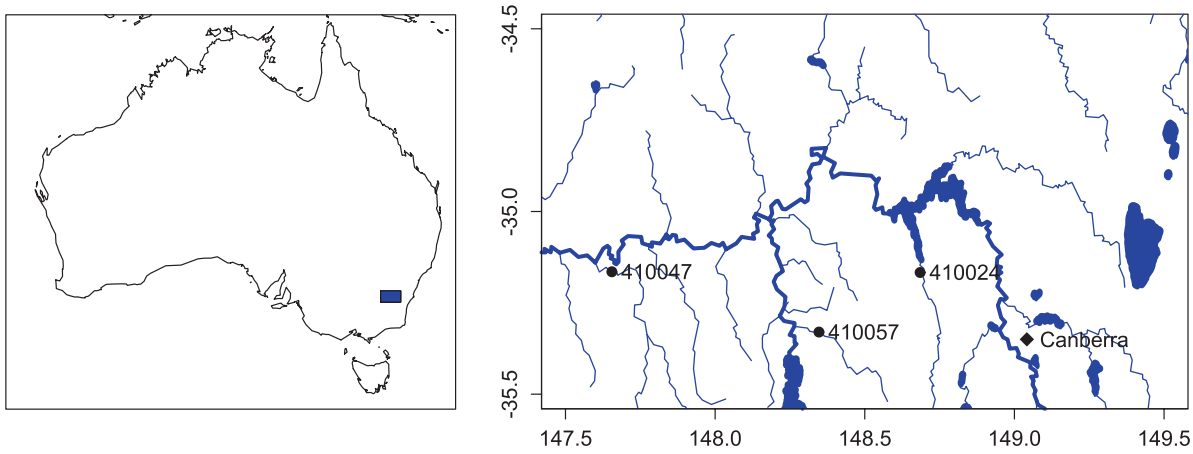

Figure 1. Location of streamflow gauging stations.

For an ideal forecast, π^t should be uniformly distributed. The uniformity can be checked by pooling together π^t values for all the forecast cases $t = 1,2, \ldots, n$ and displaying ranked π^t values in a uniform probability plot (Figure 2 later in the paper). This PIT uniform probability plot (termed predictive QQ plot by *Thyer et al.* [2009]) is useful 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]. The Kolmogorov confidence bands can be presented on the same graph to provide a more formal test of uniformity [*Laio and Tamea*, 2007]. It is noted that the well known rank histogram (Talagrand diagram) is an alternative method for displaying PIT uniformity but more suited for large samples. It is also noted that the PIT uniform probability plot for probabilistic forecasts of continuous variables is similar to the reliability diagram for probabilistic forecasts of ordinal multicategory variables [*Hamill*, 1997]. The latter is a generalization of the well known reliability diagram for probabilistic forecasts of two-category variables [*Wilks*, 1995].

[30] The LEPS and CRPS skill scores and the PIT uniform probability plot provide only some overall verifications. They do not tell whether the quality of the forecasts have changed with time or varied with different classes of cases. For example, the PIT uniform probability plot may show that the forecasts for all the cases mixed together are overall unbiased even when the individual forecasts are systematically biased, some positively and some negatively. Therefore, additional more detailed verifications are needed. A number of plots are devised here for this purpose.

5.4. Forecast Quantile and Observed Value Comparison Plots

[31] For each case, the observed value is compared with the forecast median and forecast [0.10, 0.90] quantile range. Comparisons for all the cases are visually shown together by plotting the cases both chronologically and according to the forecast median (Figures 3 and 4 later in the paper). The plots are then visually examined for patterns and trends to identify if there are systematic errors in the forecasts.

5.5. PIT Plots

[32] For each case, the forecast PIT of the observed value is calculated. The PIT values for all the cases are visually shown together by plotting the cases both chronologically and according to the forecast median (Figures 5 and 6 later in the paper). Again, the plots are visually examined for patterns and trends to identify if there are systematic errors in the forecasts.

6. Model Checking

[33] Forecast verification deals with model performance in a predictive mode. Model checking here deals with model appropriateness in a fitting mode. The assumed Box-Cox transformed multivariate normal probability model is checked for consistency with observations including (1) consistency in the marginal distributions of the predictors and predictands and (2) consistency in the joint distribution of all the predictors and predictands.

6.1. Marginal Distributions

[34] The marginal distribution of a predictor or predictand is the probability distribution when all other predictors and predictands are integrated out of the joint probability distribution of the predictors and predictands. It can also be understood as the probability distribution of a predictor or predictand when all other predictors and predictands are unknown. In this study, the marginal distributions of the predictors and predictands are derived numerically as follows. For a given set of parameters $\boldsymbol{\theta}$, a large sample of size, say 10,000, of vector $\mathbf{z}$ is drawn from the multivariate normal distribution of equation (5). The generated z sample is converted to y by using equation (38). For each of the variables y_i, $i = 1,2, \ldots, d$, the generated 10,000 values of y_i are used to form a cumulative marginal distribution of that variable. This procedure is repeated for the different sets of parameters in the representative sample of parameters obtained through MCMC simulations as described in section 3.4, forming a sample of the cumulative marginal distribution. The median and the [0.05, 0.95] uncertainty band of the cumulative marginal distribution are extracted from the sample and compared with observed data in a probability plot (Figure 7 later in the paper).

6.2. Joint Distribution

[35] The consistency check of the marginal distributions of the predictors and predictands examines each of the

Table 1. Streamflow Gauging Stations and Associated Catchments

Gauge Number	Station Name	Catchment Area (km^2)	Mean Annual Runoff (mm)
410024	Goodradigbee River at Wee Jasper	990	275
410057	Goobarragandra River at Lacmalac	663	399
410047	Tarcutta Creak at Sunnyside	1645	100

Table 3. September to November Streamflow Forecast Skill Scores[a]

	Gauge			
	410024	410057	410047	Sum of Flows
SS_{LEPS}	32	30	35	33
SS_{CRPS}	24	16	22	21

[a]Values are given in percentage.

predictors and predictands individually. To assess the consistency of the modeled joint distribution of all the predictors and predictands with observations, the marginal distributions of different combinations of the variables are examined. Different linear combinations represent different directions in the multiple variable space and can be arbitrarily chosen. A convenient set of linear combinations to use is the principal components derived from observed **y** data (for only cases with complete records). The modeled **y** data previously generated for deriving the modeled marginal distributions of the predictors and predictands as in section 6.1 are mapped to these principal components and then used to extract the median and the [0.05, 0.95] uncertainty band of the cumulative marginal distribution of each of the principal components. Again, the modeled cumulative marginal distribution is compared with values directly calculated from observed data in a probability plot (Figures 9a–9g later in the paper).

[36] The modeled marginal distribution of the sum of forecast streamflows at the multiple sites are also derived and compared with observed data (Figure 9h later in the paper). As the behavior of the sum of streamflows at the multiple sites is influenced by the correlations of streamflows across sites, this analysis provides another check on the modeled joint distribution. In addition, cross-correlation coefficients of streamflows across sites are calculated from modeled data as in section 6.1 and compared with values directly calculated from observed data (Figure 10 later in the paper).

7. Application

7.1. Data

[37] To demonstrate the BJP modeling approach, it was applied to forecasting streamflows at three gauging stations in the Murrumbidgee River catchment in southeast Australia (Figure 1). The Murrumbidgee River catchment supports a substantial irrigation industry with a total of irrigation water entitlement over 2400 GL and producing over $600 million in farm revenue in 2000/01. More than half of the farm revenue is sourced from annual grain and vegetable crops, while the remaining is sourced from permanent plantings of fruit, nuts and grapes [*Meyer*, 2005]. Annual irrigation water allocations are variable. Progressive announcements are made throughout a season with allocations increasing according to the availability of water in storages. Irrigators wanting to plan their cropping and water trading activities early in an irrigation season need reliable forecasts of likely irrigation allocations, which in turn depend on reliable forecasts of storage inflows.

[38] The three gauging stations and their associated catchment areas and mean annual runoff are given in Table 1. The location of the gauges is shown in Figure 1. Streamflows for September to November were forecast from antecedent climate and catchment conditions. The Southern Oscillation Index (SOI) for August was used to represent the antecedent climate condition, while the streamflows in August at the three sites were used to represent the antecedent catchment conditions. 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. Table 2 gives a list of the predictors and predictands used. The data used covers a period of 90 years from 1916 to 2005 with some data missing. In total there are 40 years (cases) with complete records, in which none of the predictors and predictands have data missing.

[39] Cross validation was carried out by leaving out one of the cases that have complete records of all the predictors and predictands. A sample of 1000 sets of parameter values were generated from MCMC simulations. A probabilistic forecast of the streamflows for September to November at the three gauging stations was then made for the left-out case. 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 the cases that have complete records.

[40] Model checking on the marginal distributions and the joint distribution was undertaken using all the available data. The principal components of the observed data could

Table 2. Predictors and Predictands[a]

	Predictor				Predictand		
	SOI	410024 Flow	410057 Flow	410047 Flow	410024 Flow	410057 Flow	410047 Flow
Period	Aug	Aug	Aug	Aug	Sep–Nov	Sep–Nov	Sep–Nov
Missing data (years)	0	9	32	36	6	32	40

[a]Data are from 1916 to 2005.

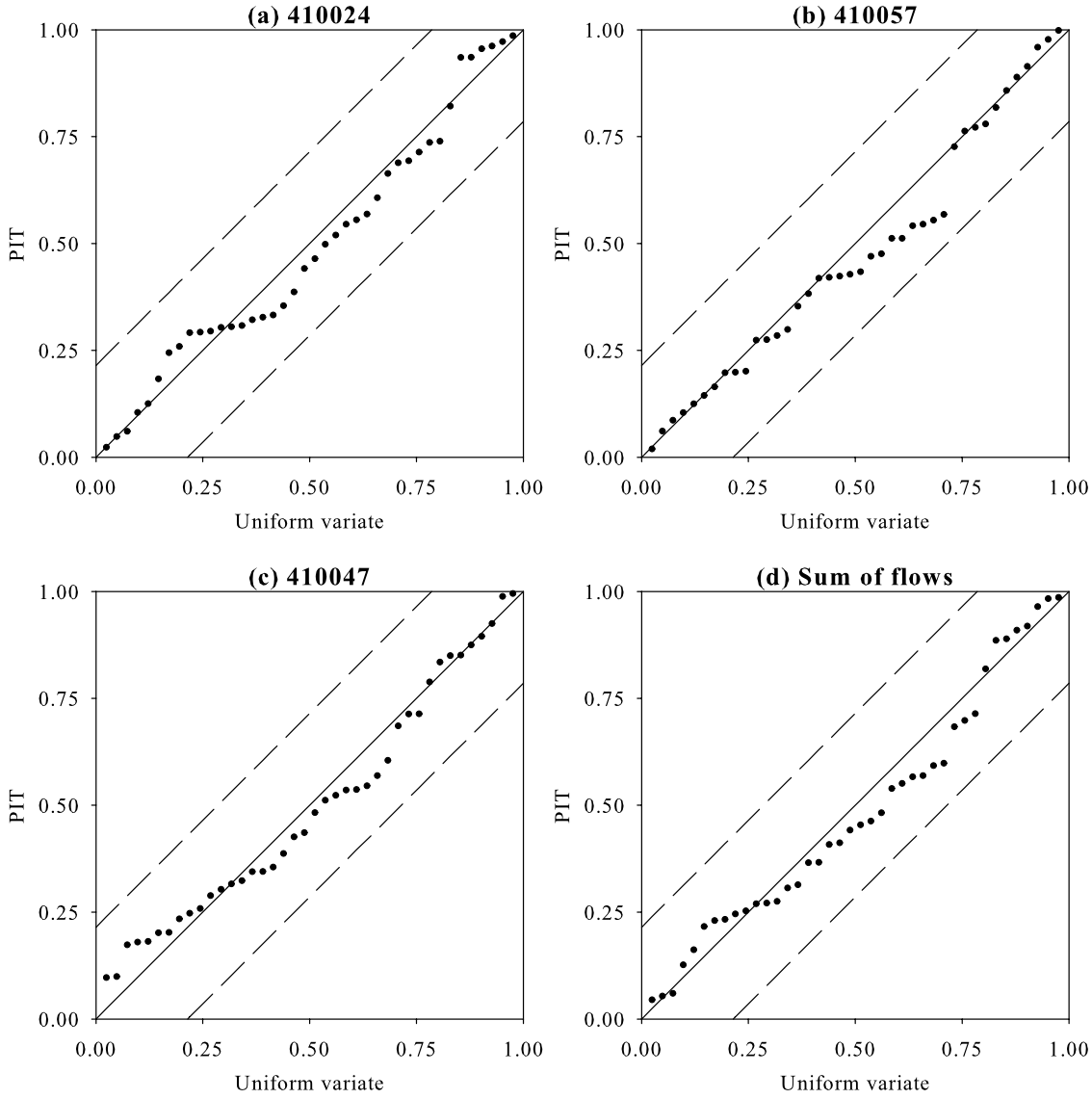

Figure 2. PIT uniform probability plots (1:1 solid line, theoretical uniform distribution; dashed lines, Kolmogorov 5% significance band; dots, PIT value of observed streamflow).

only be derived for the cases that have complete records of all predictors and predictands. The sum of observed streamflows for September to November could only be calculated for cases that have observed September to November streamflows at all three gauging stations.

7.2. Forecast Verification Results

7.2.1. LEPS and CRPS Skill Scores

[41] The skill scores for the forecast September to November streamflows at the three gauges are given in Table 3. The LEPS skill scores (SS_{LEPS}) for the forecast streamflows at the three gauges are 30% or higher. *Piechota et al.* [2001] suggested that forecasts with a LEPS skill score of 10% or more is of good skill. Table 3 also shows that the LEPS skill score for the sum of the forecast streamflows over the three gauges is 33%, which is comparable with the LEPS skill scores for the individual gauges. In fact, the level of skill is also comparable with a stand-alone model for forecasting only the sum of the streamflows (results not shown). Thus, the forecast streamflows at the three gauges are sensibly correlated with each other. The CRPS skill scores (SS_{CRPS}) shown in Table 3 are lower in numerical values than the LEPS skill scores, but they are in general consistent with the LEPS skill scores except for gauge 410057, where the CRPS skill score is disproportionally lower than the LEPS skill score compared with the other two gauges. The cause for this will be discussed later.

7.2.2. PIT Uniform Probability Plots

[42] The plots for the forecast September to November streamflows at the three gauges and their sum are given in Figure 2. The PIT values align well with the diagonal line and lie well within the Kolmogorov 5% significance bands [*Laio and Tamea*, 2007]. This suggests that the PIT values are distributed quite uniformly and thus the forecast probability distributions are overall sensible in that they are not predicting too high or too low or too wide or too narrow [*Laio and Tamea*, 2007; *Thyer et al.*, 2009]. In other words, the forecast probability distributions are overall unbiased and of appropriate spread.

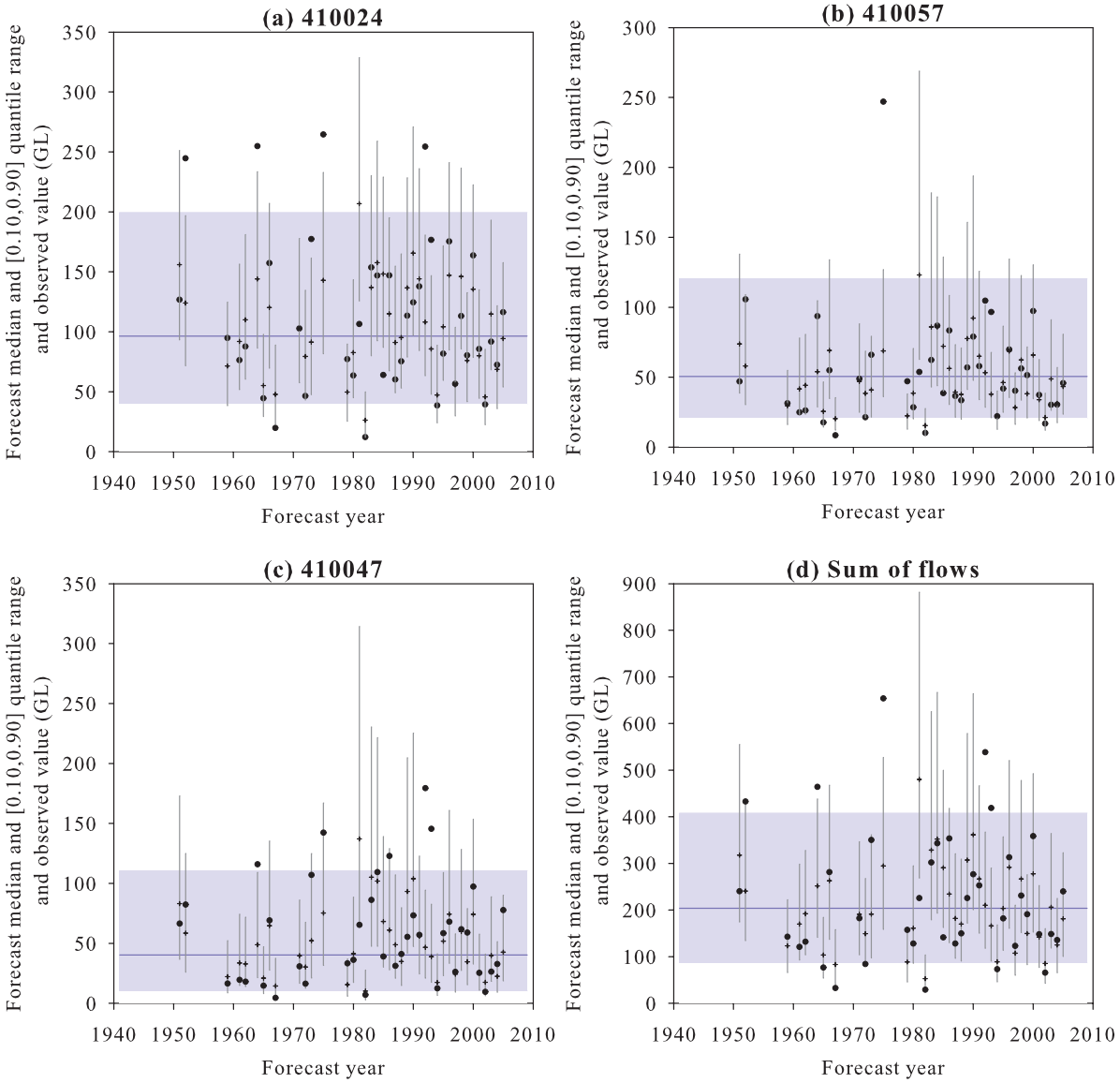

Figure 3. Forecast quantiles and observed value plotted chronologically (short horizontal lines, forecast median; vertical lines, forecast [0.10, 0.90] quantile range; blue horizontal line, climatology median; shading, climatology [0.10, 0.90] quantile range; dots, observed streamflow).

7.2.3. Forecast Quantile and Observed Value Comparison Plots

[43] Figures 3 and 4 provide a comparison of forecast median and [0.10, 0.90] quantile range with observed value for individual cases. The cases are displayed chronologically in Figure 3 and according to forecast median in Figure 4. From Figure 3, 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. From Figure 4, the forecast median appears to be consistent with observed values. The forecast [0.10, 0.90] quantile range generally increases with forecast median and also appears to be consistent with observed values. There does not appear to be any trend with forecast median in the relationship between the forecasts and observed values. For example, the forecasts are not obviously biased in any particular direction over forecast median. Figures 3 and 4 also provide a contrast of the BJP forecasts with climatology reference forecasts, showing the information gained from the BJP modeling over climatology only. The plots in Figure 4 are particularly useful for giving a "feel" for what the forecasts may look like when they have achieved certain LEPS and CRPS skill scores (Table 3). It was noted earlier that the CRPS skill score for gauge 410057 is disproportionally lower than the LEPS skill score compared with the other two gauges. This is caused by one very large flow event (observed in 1975) that was poorly forecast (Figures 3 and 4). The CRPS skill score is more sensitive to large events which are prone to large forecast errors in absolute term, while the LEPS skill score treats large and small events equally as it is based on errors in the probability distribution space only.

7.2.4. PIT Plots

[44] Figures 5 and 6 present PIT values for individual cases. The cases are displayed chronologically in Figure 5

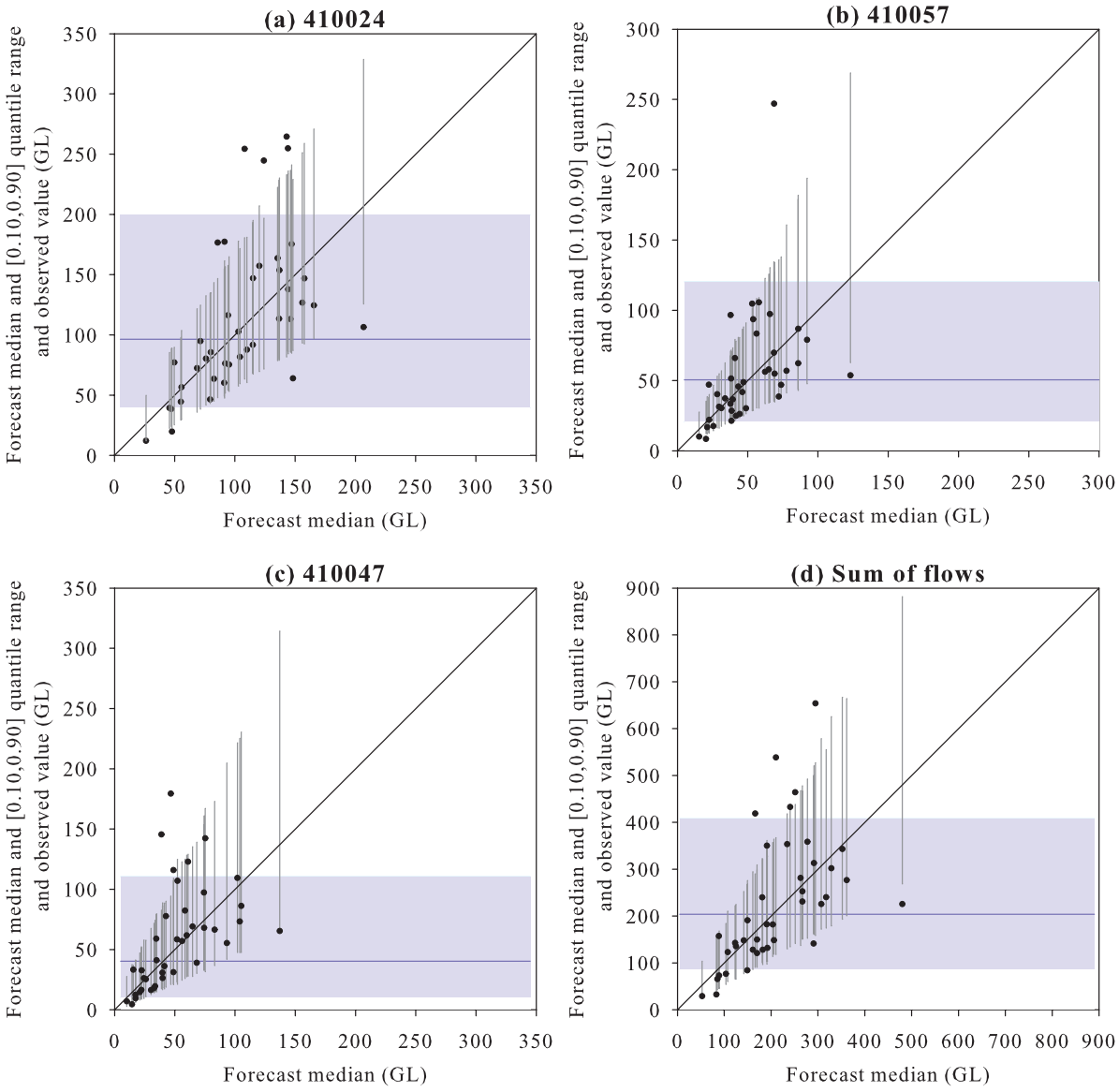

Figure 4. Forecast quantiles and observed value plotted according to forecast median (1:1 line, forecast median; vertical line, forecast [0.10, 0.90] quantile range; blue horizontal line, climatology median; shading, climatology [0.10, 0.90] quantile range; dots, observed streamflow).

and according to forecast median in Figure 6. 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 forecast median. This appears to be the case with forecast median in Figure 6 and appears to be case with time in Figure 5 before the mid-1990s. After the mid-1990s, the PIT values are more clustered toward 0.50, indicating that the forecasts are still unbiased but the forecast distributions could perhaps have been sharper (or narrower). Whether this represents a real change in the relationship between predictands and predictors is unclear from only about 10 data points since the mid-1990s. Additionally, any interdecadal modulation of the impact of ENSO on climate and hydrology that may exist [*Power et al.*, 1999] is difficult to detect from Figure 5 and previously Figure 3, which cover only a relatively a short period of record.

7.3. Model Checking Results

[45] The modeled and observed marginal distributions of the predictors and predictands are compared in Figure 7. Except for gauge 410057 (Figure 7d), the observed data are reasonably well described by the modeled median marginal distributions and the majority of observed data points fall within the modeled [0.05, 0.95] uncertainty band. The mismatch for gauge 410057 is as a result of the transfer of information from other gauges, in particular gauge 410024, which has many more data points. In the 90 years of record used for model inference, the September–November streamflow at gauge 410057 has 32 missing data points, while the September–November streamflow at gauge 410024 has only 6 missing data points (Table 2). The joint model inference using all the available data points leads to the transfer of information between two correlated data series [*Wang*, 2001, 2008], in this case mostly from gauge 410024 to gauge 410057. Thus, the marginal distribution of

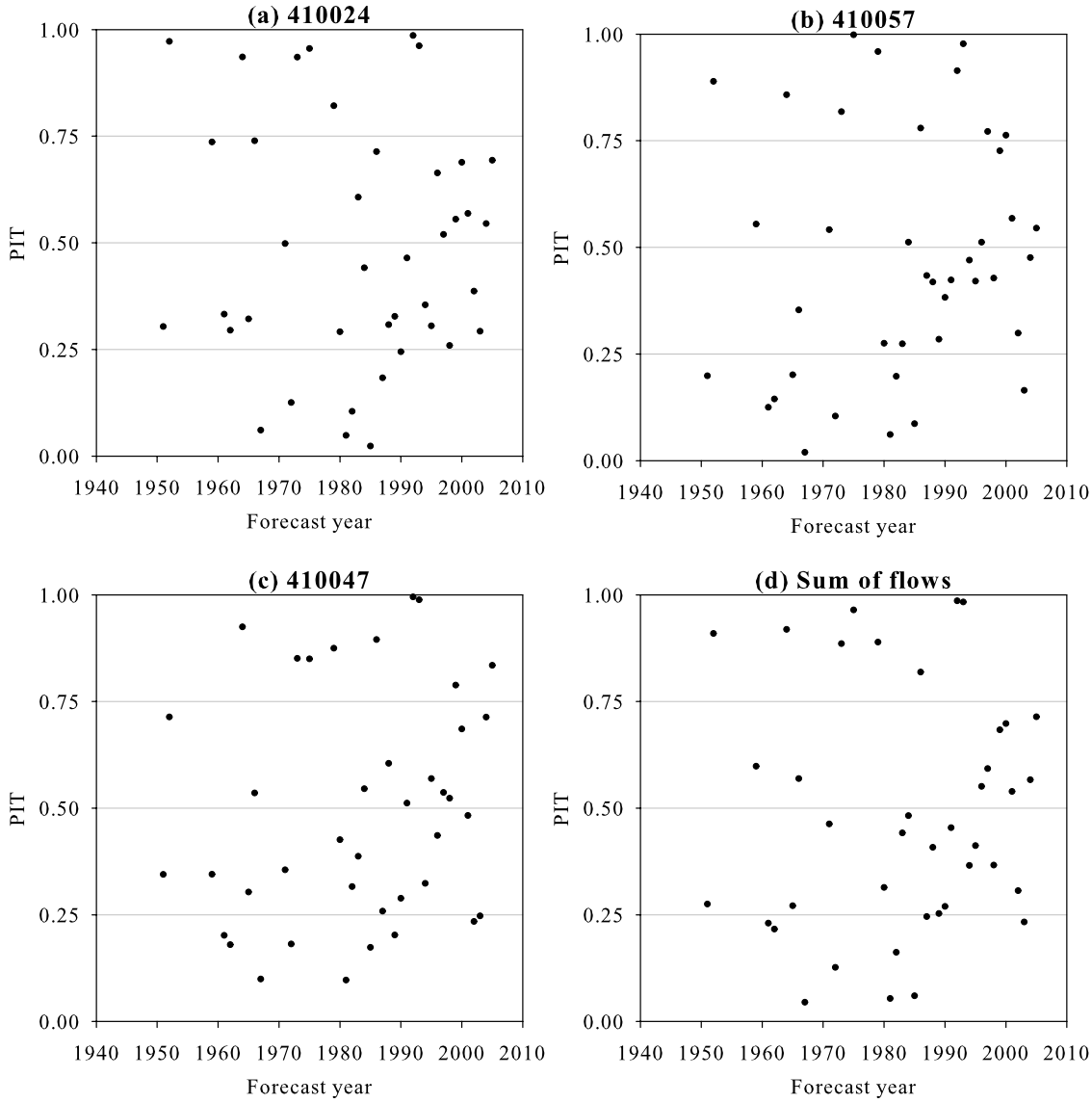

Figure 5. PIT values plotted chronologically.

the September–November streamflow at gauge 410057 derived from the joint model inference (Figure 7d) is influenced not only by data from that gauge but also by data from gauge 410024 (and other gauges as well as model assumptions). In Figure 8a, missing data are plotted as zeros, and therefore points on the x axis represent data missing at gauge 410057. The streamflows at the two gauges are highly correlated as shown by the nonmissing pairs of data. The September–November streamflows at gauge 410024 that correspond to the missing data at gauge 410057 are shown to be mostly in the lower half of its flow range. Therefore the missing streamflows at gauge 410057, had they been observed, are also likely to be in the lower half of its flow range. This additional information is taken into account in the joint model inference and leads to lower marginal distribution quantiles than otherwise derived from using the at-gauge data only. This is clearly shown in Figure 7d.

[46] The September–November streamflow at gauge 410047 also has a large number of missing data (Table 2 and Figure 8b). However, the missing data are likely to have spread throughout the flow range at that gauge as shown by how the points on the x axis spread. For this reason, the information transfer from gauge 410024 to gauge 410047 is not expected to noticeably shift the derived marginal distribution of streamflow at gauge 410047. Another point to be made is that the information transfer between highly correlated data series with unequal number of data points also has the effect of reducing the uncertainty of the marginal distribution estimation [*Wang*, 2001, 2008]. This may explain Figure 7 in that there are perhaps a slightly larger than expected number of points close to or outside the [0.05, 0.95] uncertainty bands, which would have been wider if derived from using at-gauge data only.

[47] 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 Figures 9a–9g. Overall there is a good match between the modeled results and observed results, considering that the marginal distributions of the principal components were not

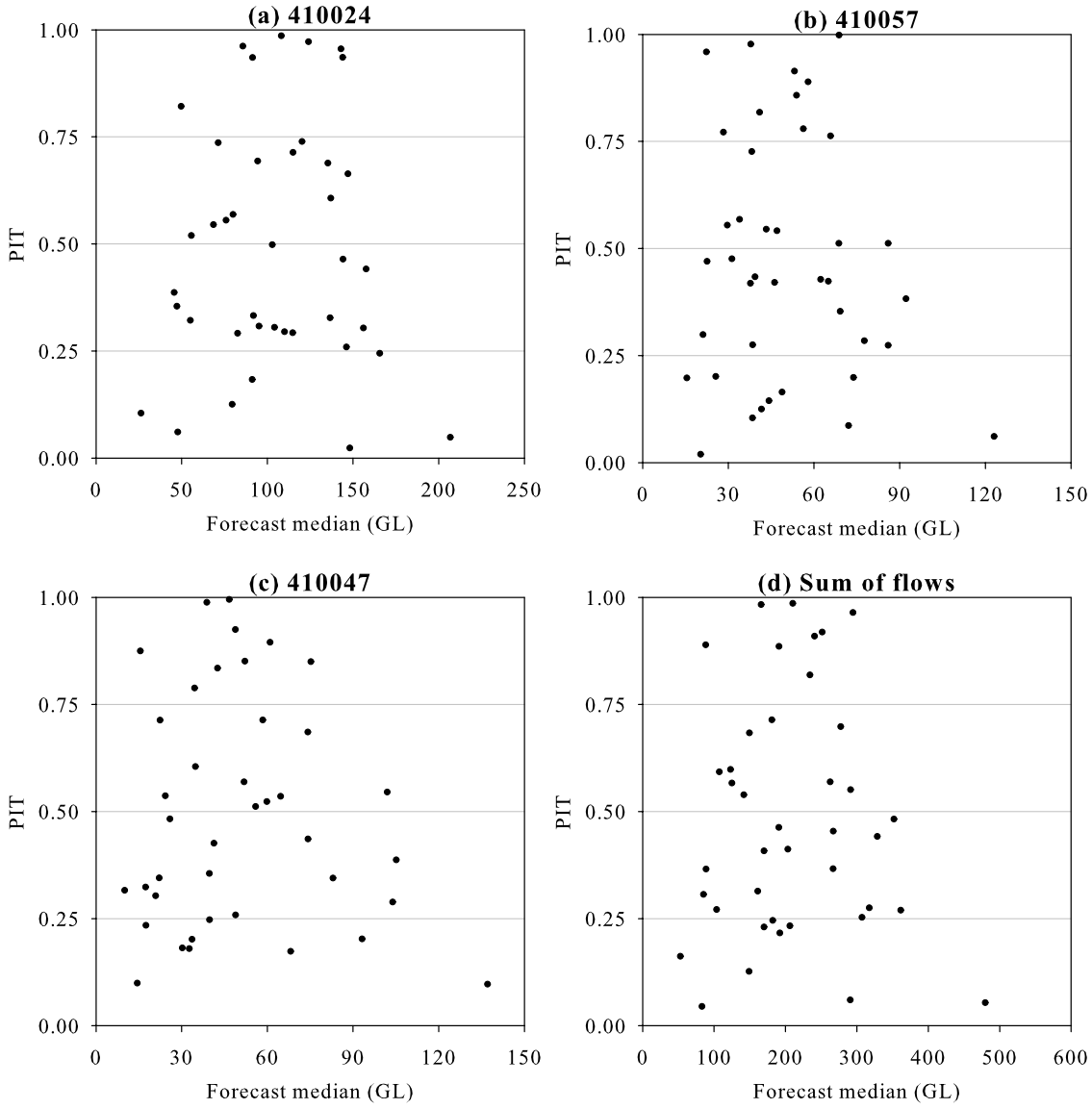

Figure 6. PIT values plotted according to forecast median.

explicitly modeled. Figure 9h shows the modeled marginal distribution of the sum of streamflows at the three gauges. The modeled results match closely with the observed data. Figure 10 compares the modeled medians and [0.05, 0.95] uncertainty ranges of the 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 in the high range, are slightly positively biased in the middle range and negatively biased in the lower range. Overall, the modeled correlation coefficients are considered consistent with the observed values. All the above results indicate that the Box-Cox transformed multivariate normal formulation of the joint probability distribution model is reasonably consistent with observed data.

8. Discussion

[48] The BJP modeling approach presented in this paper assumes that predictors are already chosen for inclusion in the model. In practice, there is often a large number of candidate predictors to choose from. A stepwise selection procedure, similar to stepwise regression, is likely to be the most feasible way to decide what predictors to use. The procedure will require a selection criterion. To assess the true quality of forecasts, a double cross validation will be needed to account for any effect of predictor optimization. This can be computationally highly demanding. A stepwise selection procedure is currently being developed and will be presented in the future.

[49] The number of parameters in the Box-Cox multivariate normal distribution increases quickly as the number of joint forecast sites increases. Model inference by MCMC sampling is likely to be able to cope with only a relatively small number of joint sites. As many of the parameters are related to intersite correlation coefficients, it is desirable to incorporate a (hierarchical) spatial model of intersite correlations to reduce the effective number of parameters. Unfortunately, suitable spatial models of intersite correlations of streamflows are unavailable. A more feasible approach is to first apply the BJP modeling to a small

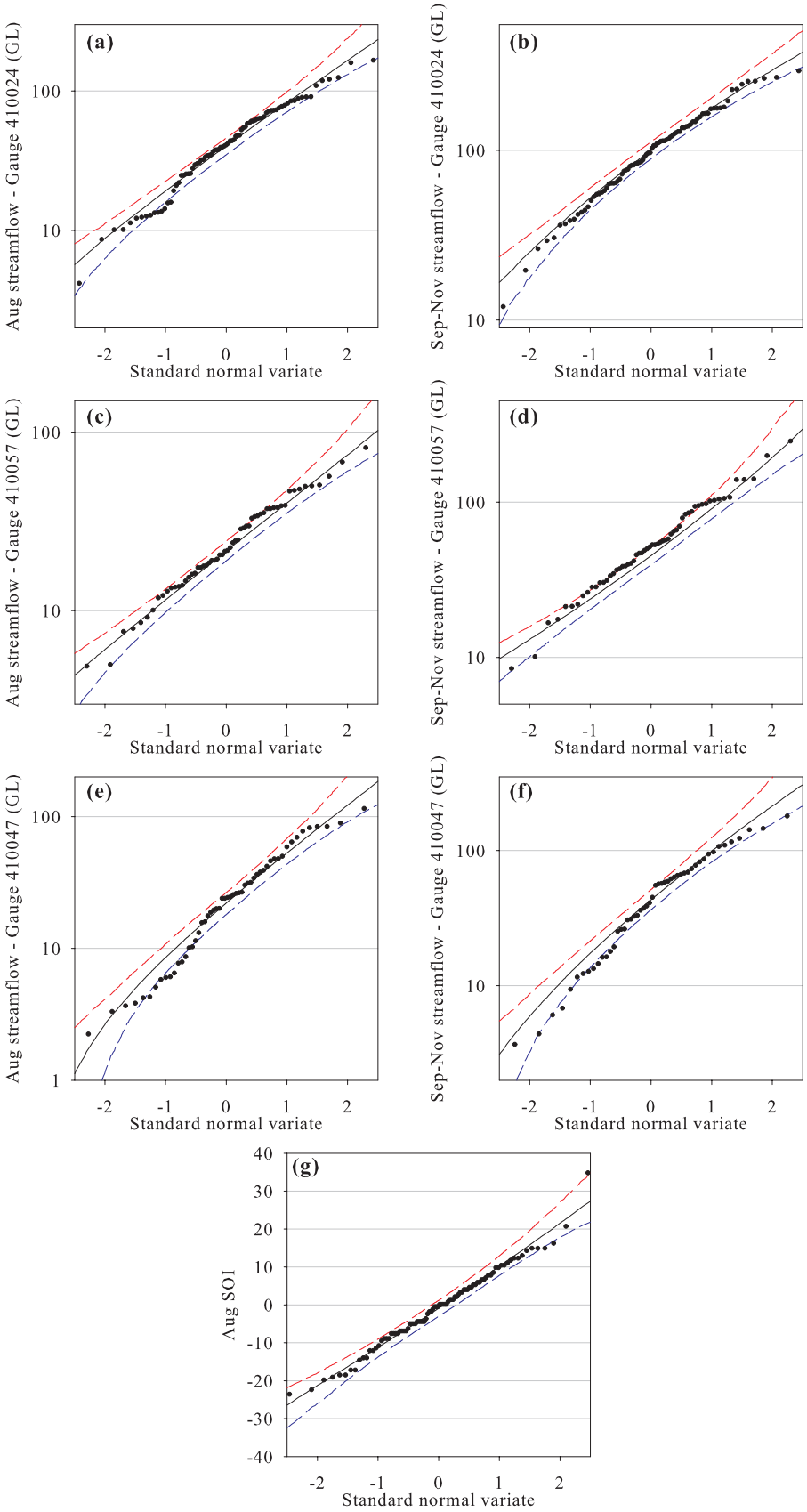

Figure 7. Marginal distribution of predictors and predictands (solid line, modeled marginal distribution median; dashed lines, marginal distribution [0.05, 0.95] uncertainty band; dots, observed data).

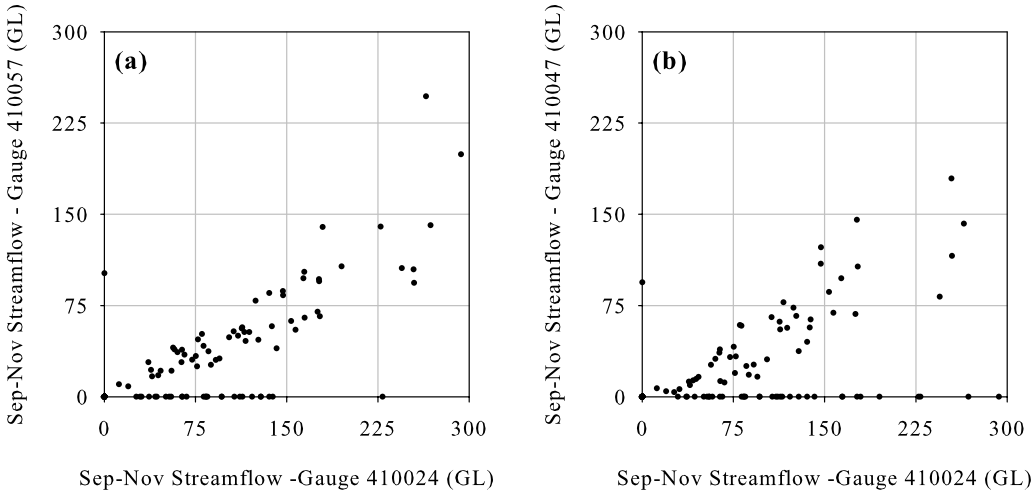

Figure 8. Streamflows at (a) gauge 410057 and (b) 410047 plotted against streamflows at gauge 410024. Missing data are plotted as having zero values.

number of key sites and then relate other sites to the key sites in some way. This will be investigated in future work.

[50] The Box-Cox multivariate normal distribution model assumes a stationary relationship between predictors and predictands. Some of the detailed forecast verification methods presented in this paper (such as those represented by plots in Figures 3 and 5) are helpful for detecting if the stationarity breaks down. Nonstationarity may be caused by changes in land use and cover and in water use practice or as a result of shifts in climate systems. It is possible to incorporate nonstationary relationships into statistical models, provided that such relationships are identified. Note that the discussion here concerns nonstationarity in relationships rather than nonstationarity in data series.

[51] The Box-Cox multivariate normal distribution model also assumes that the predictors are related to the predictands in particular ways. In the transformed variable space, the influences of the different predictors on the predictands are additive. Other types of relationships are possible. For example, initial catchment conditions may dampen or amplify future climate influences on streamflows depending on whether the catchment is dry or wet. This will be investigated in future work.

[52] Other aspects of the BJP modeling approach have been discussed throughout the paper. Some of the issues are specifically related to the BJP modeling approach, but many are pertinent to statistical modeling in general. A dynamic climate and hydrological modeling offers the potential to bypass some of these issues. However, much work is still needed to improve dynamic climate modeling to produce sufficiently skilful forecasts of climate variables to drive dynamic hydrological models. Even then, some kind of statistical postprocessing is likely to be needed to deal with modeling bias and uncertainty. Parallel work is being carried out on a dynamic climate and hydrological modeling approach.

9. Conclusions

[53] In this paper, a Bayesian joint probability (BJP) modeling approach for seasonal forecasting of streamflows at multiple sites is presented. A Box-Cox transformed multivariate normal distribution is used to model the joint distribution of future streamflows and their predictors such as antecedent streamflows, El Niño-Southern Oscillation indices and other climate indicators. Bayesian inference of model parameters and their uncertainties is implemented using MCMC simulations. Joint probabilistic forecasts of streamflows at multiple sites are generated. The BJP modeling approach is potentially of wide applications because (1) the Box-Cox transformed multivariate normal distribution is highly flexible for modeling a wide range of candidate predictors as well as predictands and (2) the statistical inference method allows the use of data that contains missing and nonconcurrent records.

[54] The paper also presents a number of statistical measures and graphical methods adopted, adapted or devised for verification of probabilistic forecasts of continuous variables. These include LEPS and CRPS skill scores and PIT uniform probability plot for overall verifications and forecast quantile and observed value comparison plots and PIT plots for detailed verifications.

[55] The BJP modeling approach was applied to forecasting streamflows at three river gauges in the Murrumbidgee River catchment in southeast Australia. September to November streamflows were forecast from Southern Oscillation Index for the previous month and streamflows over the previous month. Cross validation results show that the BJP forecasts of streamflows are of good quality in that the probabilistic forecasts achieved reasonably high forecast skill scores, are in general free from obvious bias and other errors, and have appropriate uncertainty spread. The formulation of the model, the Box-Cox transformed multivariate normal distribution, are shown to be consistent with observed data. Specifically, the modeled marginal distributions of the predictors and predictands, marginal distributions of the principal components of the predictors and predictands, marginal distribution of the sum of forecast streamflows at multiple sites, and cross-correlation coefficients of the predictors and predictands are all shown to be consistent with the observed data.

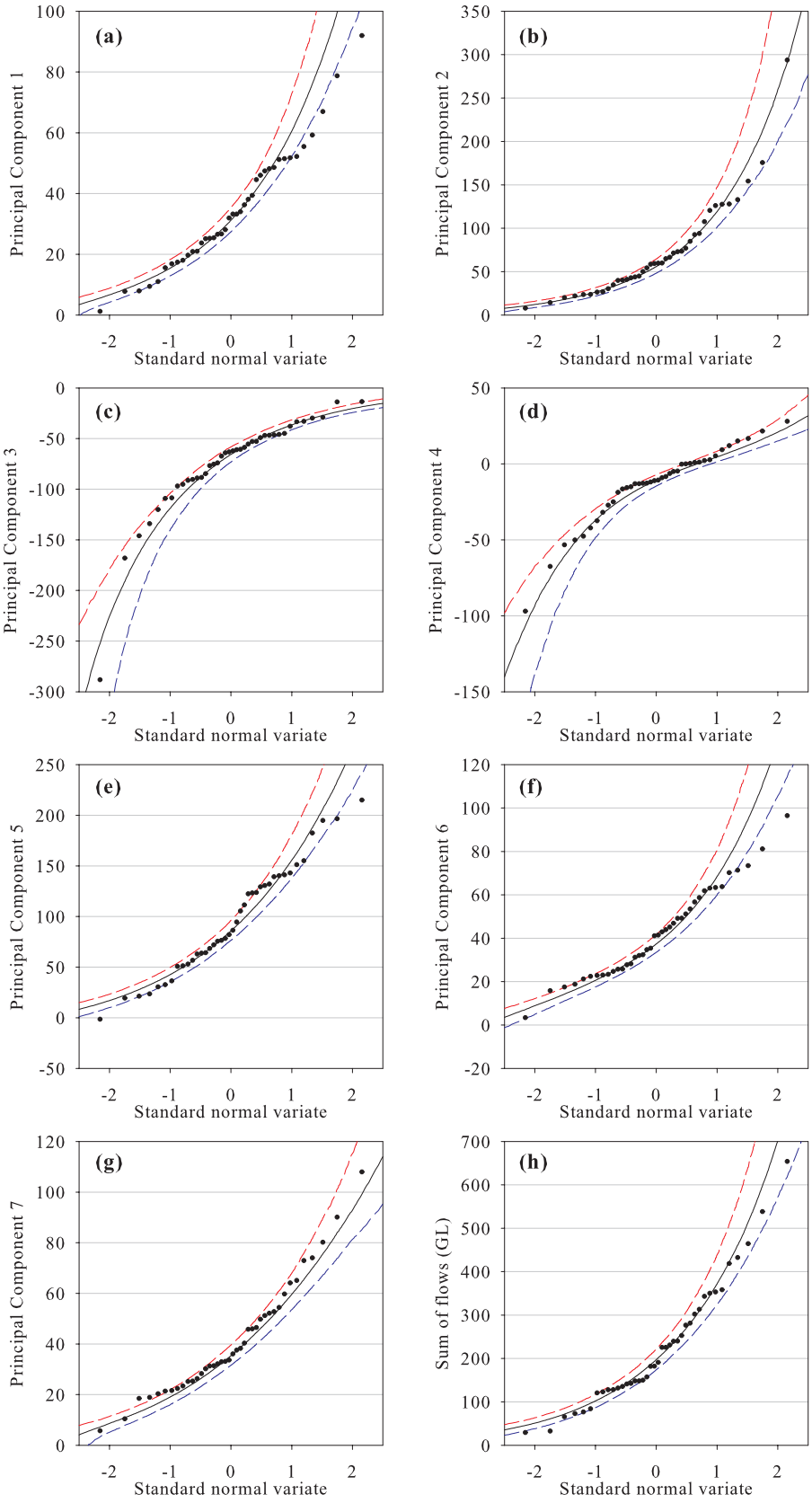

Figure 9. (a–f) Marginal distribution of principal components of all predictors and predictands and (h) marginal distribution of the sum of streamflows at the three gauges (solid line, modeled marginal distribution median; dashed lines, marginal distribution [0.05, 0.95] uncertainty band; dots, value directly calculated from observed data).

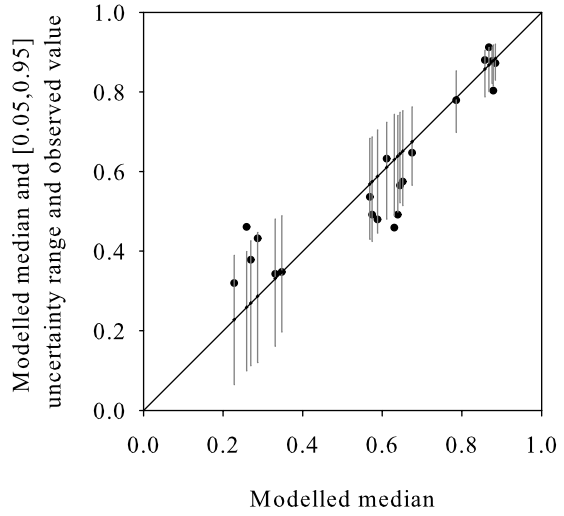

Figure 10. Cross-correlation coefficients of all the predictors and predictand variables (1:1 line, modeled median; vertical lines, modeled [0.05, 0.95] uncertainty range; dots, value directly calculated from observed data).

[56] Further work is being undertaken on predictor selection, potential incorporation of outputs from dynamic climate models as predictors, joint forecasting for multiple seasons ahead, handling zero flows for ephemeral streams, dealing with a large number of streamflow sites, investigating effects of predictor interactions, and applications to major water resources systems across Australia.

[57] A copy of FORTRAN subroutines for finding the nearest correlation matrix, for evaluating the marginally uniform prior distribution for the correlation matrix, and for calculating the LEPS and CRPS skill scores may be made available on request to the corresponding author.

[58] **Acknowledgments.** This research has been supported by the CSIRO CEO Science Leadership Scheme, the South Eastern Australian Climate Initiative, and the Water Information Research and Development Alliance between the Australian Bureau of Meteorology and CSIRO Water for a Healthy Country Flagship. Andrew Frost and Peter Hairsine made valuable suggestions on an early draft of the paper. Constructive comments from three anonymous reviewers led to substantial strengthening of the paper, in particular on forecast verification. Discussions with a number of colleagues from the Australian Bureau of Meteorology, CSIRO, and other organizations helped clarify a number of ideas.

References

Ashok, K., Z. Guan, and T. Yamagata (2003), Influence of the Indian Ocean Dipole on the Australian winter rainfall, *Geophys. Res. Lett.*, *30*(15), 1821, doi:10.1029/2003GL017926.

Barnard, J., R. McCulloch, and X.-L. Meng (2000), Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage, *Stat. Sin.*, *10*(4), 1281–1311.

Bouttier, F. (1994), Sur la prévision de la qualité des prévisions météorologiques, Ph.D. thesis, 240 pp., Univ. Paul Sabatier, Toulouse, France.

Box, G. E. P., and D. R. Cox (1964), An analysis of transformations, *J. R. Stat. Soc., Ser. B*, *26*, 296–311.

Brown, T. A. (1974), Admissible scoring systems for continuous distributions, *Manuscr. P-5235*, 22 pp., RAND Corp., Santa Monica, Calif.

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., S. L. Zhou, and T. A. McMahon (2003), Use of seasonal streamflow forecasts in water resources management, *J. Hydrol.*, *270*(1–2), 135–144, doi:10.1016/S0022-1694(02)00292-5.

Cowles, M., and B. P. Carlin (1996), Markov chain Monte Carlo convergence diagnostics: A comparative review, *J. Am. Stat. Assoc.*, *91*, 883–904, doi:10.2307/2291683.

Drosdowsky, W., and L. E. Chambers (2001), Near global sea surface temperature anomalies as predictors of Australian seasonal rainfall, *J. Clim.*, *14*, 1677–1687, doi:10.1175/1520-0442(2001)014<1677:NACNGS>2.0.CO;2.

Epstein, E. S. (1969), A scoring system for probability forecasts of ranked categories, *J. Appl. Meteorol.*, *8*, 985–987, doi:10.1175/1520-0450(1969)008<0985:ASSFPF>2.0.CO;2.

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*(3–4), 129–148, doi:10.1016/j.jhydrol.2007.03.023.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (1995), *Bayesian Data Analysis*, 526 pp., Chapman and Hall, London.

Gneiting, T., F. Balabdaoui, and A. E. Raftery (2007), Probabilistic forecasts, calibration and sharpness, *J. R. Stat. Soc., Ser. B*, *69*, 243–268, doi:10.1111/j.1467-9868.2007.00587.x.

Hamill, T. M. (1997), Reliability diagrams for multicategory probabilistic forecasts, *Weather Forecast.*, *12*, 736–741, doi:10.1175/1520-0434(1997)012<0736:RDFMPF>2.0.CO;2.

Hendon, H. H., D. W. J. Thompson, and M. C. Wheeler (2007), Australian rainfall and surface temperature variations associated with the Southern Annular Mode, *J. Clim.*, *20*, 2452–2467, doi:10.1175/JCLI4134.1.

Hersbach, H. (2000), Decomposition of the continuous ranked probability score for ensemble prediction systems, *Weather Forecast.*, *15*(5), 559–570, doi:10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

Higham, N. J. (2002), Computing the nearest correlation matrix—A problem from finance, *IMA J. Numer. Anal.*, *22*(3), 329–343, doi:10.1093/imanum/22.3.329.

Laio, F., and S. Tamea (2007), Verification tools for probabilistic forecasts of continuous hydrological variables, *Hydrol. Earth Syst. Sci.*, *11*(4), 1267–1277.

Matheson, J. E., and R. L. Winkler (1976), Scoring rules for continuous probability distributions, *Manage. Sci.*, *22*, 1087–1095, doi:10.1287/mnsc.22.10.1087.

Meyer, W. S. (2005), The irrigation industry in the Murray and Murrumbidgee basins, *Tech. Rep. 03/05*, CRC for Irrig. Futures, Adelaide, South Aust., Australia.

Monserrat, J. F., R. Fraile, and L. Rubio (2007), Application of alternating projection method to ensure feasibility of shadowing cross-correlation models, *Electron. Lett.*, *43*(13), 724–725, doi:10.1049/el:20070236.

Murphy, A. H. (1969), On the "ranked probability score", *J. Appl. Meteorol.*, *8*, 988–989, doi:10.1175/1520-0450(1969)008<0988:OTPS>2.0.CO;2.

Murphy, A. H. (1971), A note on the ranked probability score, *J. Appl. Meteorol.*, *10*, 155–156, doi:10.1175/1520-0450(1971)010<0155:ANOTRP>2.0.CO;2.

Murphy, B. F., and B. Timbal (2008), A review of recent climate variability and climate change in south eastern Australia, *Int. J. Climatol.*, *28*(7), 859–879, doi:10.1002/joc.1627.

Piechota, T. C., F. H. S. Chiew, J. A. Dracup, and T. A. McMahon (2001), Development of exceedance probability streamflow forecast, *J. Hydrol. Eng.*, *6*(1), 20–28, doi:10.1061/(ASCE)1084-0699(2001)6:1(20).

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*(1), 34–53, doi:10.1175/1520-0442(1996)009<0034:RSFACM>2.0.CO;2.

Power, S., T. Casey, C. K. Folland, A. Colman, and V. Mehta (1999), Inter-decadal modulation of the impact of ENSO on Australia, *Clim. Dyn.*, *15*, 319–324, doi:10.1007/s003820050284.

Ruiz, J. E., I. Cordery, and A. Sharma (2006), Impact of mid-Pacific Ocean thermocline on the prediction of Australian rainfall, *J. Hydrol.*, *317*(1–2), 104–122, doi:10.1016/j.jhydrol.2005.05.012.

Ruiz, J. E., I. Cordery, and A. Sharma (2007), Forecasting streamflows in Australia using the tropical Indo-Pacific thermocline as predictor, *J. Hydrol.*, *341*(3–4), 156–164, doi:10.1016/j.jhydrol.2007.04.021.

Saji, N., B. Goswami, P. Vinaychandran, and T. Yamagata (1999), A dipole mode in the tropical Indian Ocean, *Nature*, *401*, 360–363.

Sharma, A. (2000), Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 3—A nonparametric probabilistic forecast model, *J. Hydrol.*, *239*(1–4), 249–258, doi:10.1016/S0022-1694(00)00348-6.

Thyer, M., G. Kuczera, and Q. J. Wang (2002), Quantifying parameter uncertainty in stochastic models using the Box-Cox transformation, *J. Hydrol.*, *265*(1–4), 246–257, doi:10.1016/S0022-1694(02)00113-0.

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 modeling: A case study using Bayesian total error analysis, *Water Resour. Res.*, *45*, W00B14, doi:10.1029/2008WR006825.

Verdon, D. C., and S. W. Franks (2005), Indian Ocean sea surface temperature variability and winter rainfall: Eastern Australia, *Water Resour. Res.*, *41*, W09413, doi:10.1029/2004WR003845.

Xiong, J., V. Zolotov, and H. Lei (2007), Robust extraction of spatial correlation, *IEEE Trans. Comput. Aided Design Integrated Circuits Syst.*, *26*(4), 619–631, doi:10.1109/TCAD.2006.884403.

Wang, Q. J. (2001), A Bayesian joint probability approach for flood record augmentation, *Water Resour. Res.*, *37*(6), 1707–1712, doi:10.1029/2000WR900401.

Wang, Q. J. (2008), A Bayesian method for multi-site stochastic data generation: Dealing with non-concurrent and missing data, variable transformation and parameter uncertainty, *Environ. Modell. Software*, *23*(4), 412–421, doi:10.1016/j.envsoft.2007.04.013.

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

Westra, S., A. Sharma, C. Brown, and U. Lall (2008), Multivariate streamflow forecasting using independent component analysis, *Water Resour. Res.*, *44*, W02437, doi:10.1029/2007WR006104.

Wilks, D. S. (1995), *Statistical Methods in the Atmospheric Sciences*, 467 pp., Academic, San Diego, Calif.

Yeo, I., and R. A. Johnson (2000), A new family of power transformations to improve normality or symmetry, *Biometrika*, *87*(4), 954–959, doi:10.1093/biomet/87.4.954.

Zhu, D., and A. O. Hero (2007), Bayesian hierarchical model for large-scale covariance matrix estimation, *J. Comput. Biol.*, *14*, 1311–1326, doi:10.1089/cmb.2006.0151.

F. H. S. Chiew, CSIRO Land and Water, GPO Box 1666, Canberra, ACT 2601, Australia.

D. E. Robertson and Q. J. Wang, CSIRO Land and Water, P.O. Box 56, Highett, Vic 3190, Australia. (qj.wang@csiro.au)