Seasonal and annual maximum streamflow forecasting using climate information: application to the Three Gorges Dam in the Yangtze River basin, China

HYUN-HAN KWON[1], CASEY BROWN[2], KAIQIN XU[3,4,5] & UPMANU LALL[3]

1 *Water Resources Research Division, Korea Institute of Construction Technology, Daehwa-Dong, Ilsan-Gu, Goyang-Si, Gyeonggi-Do 411-712, South Korea*

2 *Department of Civil and Environmental Engineering, University of Massachusetts, Amherst, Massachusetts, USA*
cbrown@ecs.umass.edu

3 *Department of Earth and Environmental Engineering, Columbia University, New York, New York 10027,USA*

4 *Asian Environmental Research Group, National Institute for Environmental Studies, Tsukuba 305-8506, Japan*

5 *State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan Unviversity, Wuhan 430072, China*

Abstract This paper explores the potential for seasonal prediction of hydrological variables that are potentially useful for reservoir operation of the Three Gorges Dam, China. The seasonal flow of the primary inflow season and the peak annual flow are investigated at Yichang hydrological station, a proxy for inflows to the Three Gorges Dam. Building on literature and diagnostic results, a prediction model is constructed using sea-surface temperatures and upland snow cover available one season ahead of the prediction period. A hierarchical Bayesian approach is used to estimate uncertainty in the parameters of the prediction model and to propagate these uncertainties to the predictand. The results show skill for both the seasonal flow and the peak annual flow. The peak annual flow model is then used to estimate a design flood (50-year flood or 2% exceedence probability) on a year-to-year basis. The results demonstrate the inter-annual variability in flood risk. The predictability of both the seasonal total inflow and the peak annual flow (or a design flood volume) offers potential for adaptive management of the Three Gorges Dam reservoir through modification of the operating policy in accordance with the year-to-year changes in these variables.

Key words Three Gorges Dam; Yangtze River (Changjiang); seasonal flow forecast; peak flow forecast; reservoir operations, hierarchical Bayesian model

Prévision d’écoulements saisonnier et maximum annuel à l’aide d’informations climatiques: application au Barrage des Trois Gorges dans le bassin du Fleuve Yangtze, Chine

Résumé Cet article explore le potentiel de prévision saisonnière de variables hydrologiques qui sont éventuellement utiles pour la gestion du Barrage des Trois Gorges en Chine. L’écoulement saisonnier de la principale saison d’alimentation et le débit maximum annuel sont étudiés à la station hydrologique de Yichang, vus comme proxy des apports au Barrage des Trois Gorges. Sur la base de la littérature et de résultats de diagnostic, un modèle de prévision est construit qui fait appel aux températures de surface de la mer et au couvert neigeux amont disponibles pour la saison précédant la période de prévision. Une approche Bayésienne hiérarchique est utilisée pour estimer l’incertitude dans les paramètres du modèle de prévision et pour propager ces incertitudes dans les prévisions. Les résultats sont pertinents pour l’écoulement saisonnier et pour le débit maximum annuel. Le modèle de débit maximum annuel est alors utilisé pour estimer une crue de projet (crue cinquantennale ou de probabilité de dépassement de 2%) sur une base année-après-année. Les résultats mettent en évidence la variabilité interannuelle dans le risque de crue. La prévisibilité de l’apport saisonnier total et du débit maximum annuel (ou du volume d’une crue de projet) rendent possible une gestion adaptative du Barrage des Trois Gorges via la modification des règles d’opération en accord avec les changements année-après-année de ces variables.

Mots clefs Barrage des Trois Gorges; Fleuve Yangtze (Changjiang); prévision d’écoulement saisonnier; prévision de débit maximum; gestion de barrage; modèle Bayésien hiérarchique

1 INTRODUCTION

Water resources planning and management focuses on supplying a steady water supply amid hydrological variability. Such variability occurs at many time scales, from hourly to daily and from seasonal to inter-annual and beyond. Engineering and management responses to hydrological

This paper is a revised and updated version of that presented at the Conference *Hydrological Sciences for Managing Water Resources in the Asian Developing World*, organized by the Center for Water Resources and Environment, Sun Yat-Sen University and Department of Geography and Resource Management, The Chinese University of Hong Kong, and held in Guangzhou, China, 8–10 June 2006.

Open for discussion until 1 December 2009

variability depend on the time frame of the variability. Infrastructure, such as dams, typically responds to longer temporal variations. For example, a dam is typically designed to provide water through a season (capturing wet season flows for supply during the dry seasons) or even through a year or more. In this way, the storage provided by the dam mitigates seasonal and inter-annual hydrological variability (drought). Widespread research has demonstrated that these hydrological events are often manifestations of climate fluctuations, the most familiar of which is the El Niño Southern Oscillation (ENSO; Ropelewski & Halpert, 1986, 1987; Halpert & Ropelewski, 1992; Piechota & Dracup, 1996). Due to the relatively slow evolution of climate processes (consider the persistence of sea-surface temperatures relative to a passing weather front), where hydrological extremes are physically connected with climate processes, there is hope for prediction of these events (Hamlet & Lettenmaier, 1999; Croley, 2003; Souza & Lall, 2003).

Seasonal forecasts of hydrological variables potentially improve water management (Brown & Rogers, 2006; Brown *et al.*, 2006). Reservoir operations, in particular, may benefit from a reduction in the uncertainty associated with future inflows (Kim & Palmer, 1997; Hamlet *et al*., 2002; Westphal *et al*., 2003; Souza & Brown, 2007). For example, with an expectation of enhanced inflows to the reservoir during the inflow period, extra water may be released to generate additional benefits (e.g. hydroelectricity) in lieu of overtopping of the reservoir and spilling water. Alternatively, with an expectation of diminished inflows to the reservoir, releases may be curtailed in order to reserve water for the dry period ahead. However, in many multipurpose reservoirs, a flood retention volume constrains the amount of water that can be retained. A volume of empty space is left in the reservoir in order to capture some design flood if it should occur. Since the design flood may not occur (and strictly speaking, regardless of whether it occurs or not), there is an opportunity cost for the lost potential of that reservoir volume to store water for the season ahead. This opportunity cost could be potentially reduced with a skilful forecast of the flood volume that is likely to occur.

Most studies of climate influence on hydrological variables have focused on 3-month rainfall totals or streamflow (Souza & Lall, 2003; Trigo *et al*., 2005; Kwon *et al*., 2006). However, there is increasing evidence that floods are also influenced by climate teleconnections (Jain & Lall, 2000, 2001; Franks & Kuczera, 2002; Milly *et al.*, 2002; Pizaro & Lall, 2002; Sankarasubramanian & Lall, 2003). As a result, their relative probability of occurrence may be predictable. In a previous paper, the authors described the climate teleconnections to seasonal flows of the Yangtze (Changjiang) River, at a location that serves as a proxy for inflows to the Three Gorges Dam (Xu *et al*., 2007). In the present paper, we investigate the climate teleconnections to the annual flood on the Yangtze at this site. A prediction model for the peak annual flow is developed using a Bayesian hierarchical model and climate indicators. In addition, a typical design flood value is estimated annually based on the state of the predictors. Given forecasts of the seasonal total reservoir inflow and the peak annual flow (or design flood volume) occurring in that season, the operating policies of the Three Gorges Dam could be operated in a dynamic way, responsive to the prevailing climate conditions instead of the typical decision-making processes based on historical (stationary) assumptions of inflow and flood risk.

The next section of the paper describes the data and methodology used for the analysis. Section 3 describes climate influences on Yangtze River streamflow. Section 4 describes the development of the prediction models and the estimation of the inter-annual variability in design flood estimation. The final section discusses the results of this analysis.

2 DATA AND METHODOLOGY

2.1 Data

The streamflow data for this study were recorded at the Yichang hydrological station (YHS, 111.28°E; 30.70°N, Fig. 1). The YHS, in western Hubei, is about 1837 km from the estuary in the upper-middle reach of the Yangtze River basin. The streamflow records date back to 1882.

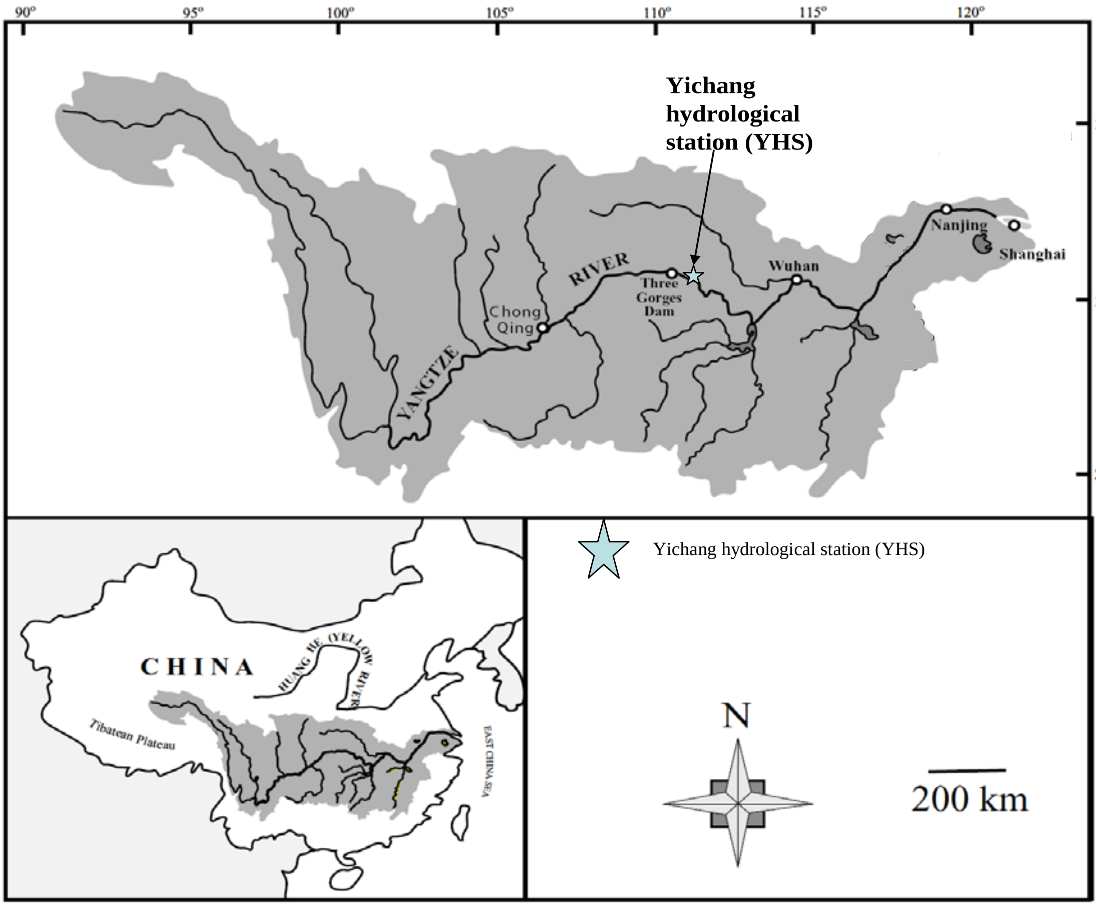

Fig. 1 Location of the Yichang hydrological station in the Yangtze River basin.

Yichang is known as the "Gateway to the Three Gorges". The well-known Three Gorges Dam (TGD) is about 40 km upstream. The drainage area above YHS is about 10^6 km^2 (Lu *et al.*, 2003; Xu *et al*., 2004, 2005). Xu *et al.* (2007) provide further details of the hydrological time series measured at YHS. While the construction and filling of the Three Gorges Dam will disrupt the natural flows of the Yangtze, the historical YHS data will continue to serve as a practical estimate of "historical inflows" into the Three Gorges Dam.

The annual mean flow, annual maximum flow, annual minimum flow and June-July-August (JJA) seasonal flow during 1882–2001 at YHS are shown in Fig. 2 (linear trend also shown). The mean streamflow time series shows a slightly, though not statistically significant, decreasing trend, consistent with the result reported by Xiong & Guo (2004). The annual hydrograph shows the seasonality of streamflow, with monsoon runoff (JJA) dominating (Fig. 3).

Because some of the climate time series data are only available from the 1970s, specifically the ocean precipitation data, which are available only from 1979, we divided the monthly time series into two parts, before and after 1979. Table 1 shows statistics for the split time series are similar. Streamflow values were summed for JJA during the period 1882–2001 to represent monsoon season streamflow. The series of annual maximum flood, the annual minimum flow and the annual mean flow during the period 1882–2001 (see Fig. 2) were extracted from the daily discharge records and used to represent the long-term hydrological characteristics at the Three Gorges Project site. Climate data was accessed from the International Research Institute for Climate and Society (IRI) Data Library (http://iridl.ldeo.columbia.edu/) and included global sea-surface temperature (SST) and snow cover. Data sets for SST are obtained from the anomaly grid product of Kaplan *et al.* (1998), and snow data was collected from the NOAA Climate Prediction

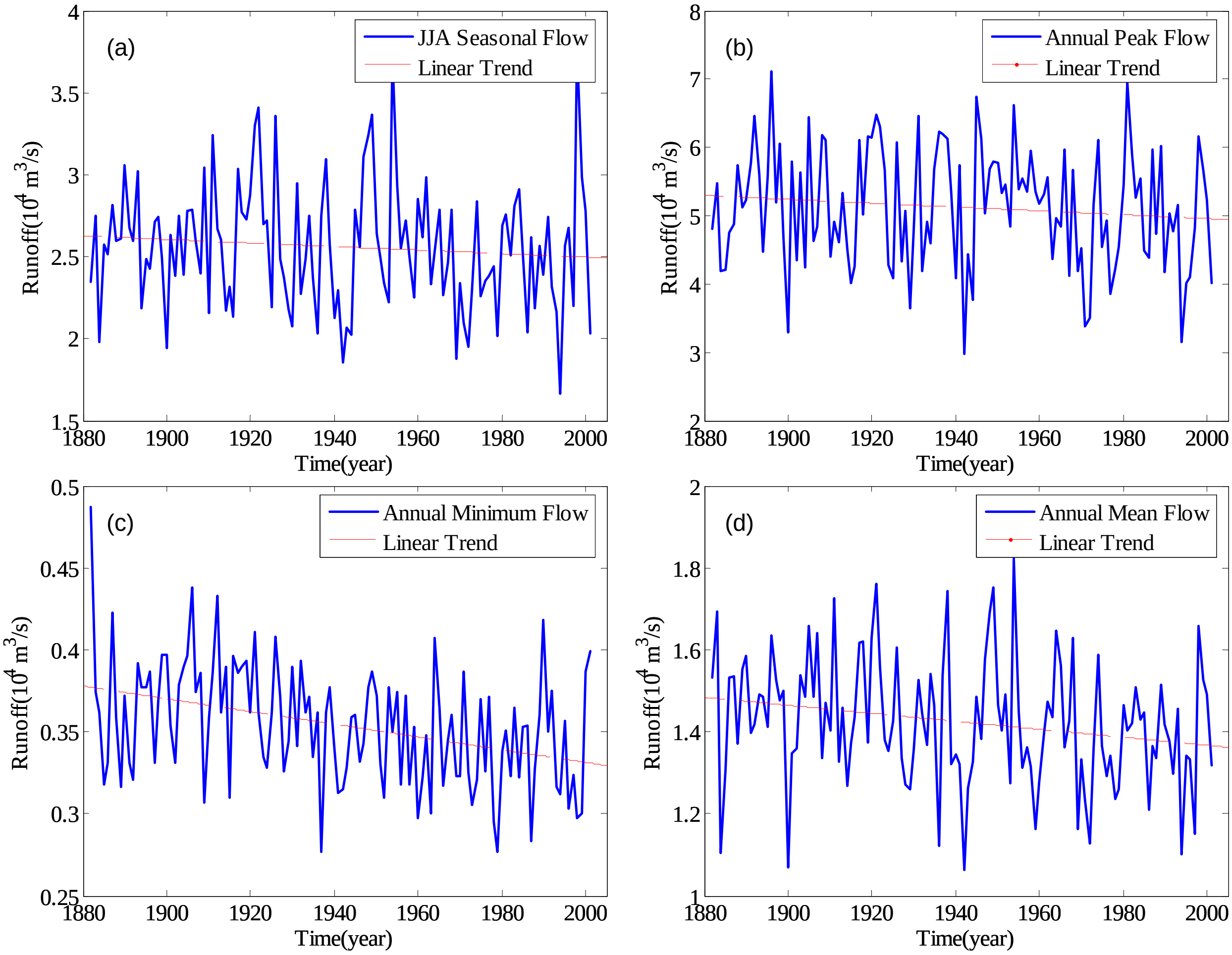

Fig. 2 (a) June-July-August (JJA) seasonal flow, (b) annual maximum flow, (c) annual minimum flow, and (d) annual mean flow at Yichang hydrological station (1882–2001).

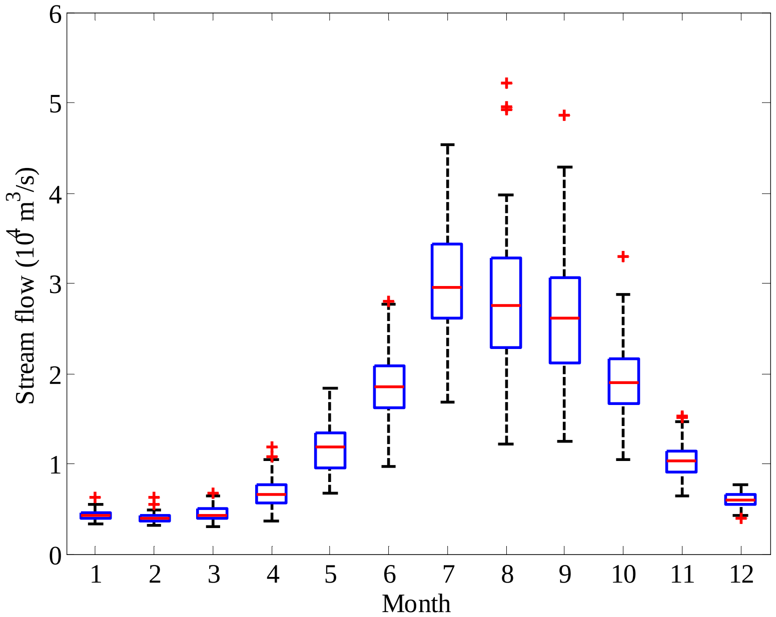

Fig. 3 Box plot of Yichang hydrological station monthly streamflow.

Table 1 Statistics of the monthly streamflow data at Yichang hydrological station.

Statistic	Total time series	1882–1978	1979–2001
Mean (10^4 m^3/s)	1.42	1.43	1.36
Standard deviation	1.02	1.03	1.00
Skewness	0.85	0.82	0.97
Minimum (10^4 m^3/s)	0.31	0.31	0.30
Maximum (10^4 m^3/s)	5.22	4.95	5.22
Range (10^4 m^3/s)	4.91	4.64	4.92
Coefficient of variation	0.72	0.72	0.74

Center website (Robinson *et al*., 1993). The available data for SST and snow cover are from years 1856 and 1970, respectively. The 3-month averages of March-April-May (MAM) values were used for prediction model development.

2.2 Forecasting model using climate information

A hierarchical Bayesian based prediction model was developed to incorporate parameter and model uncertainty in a statistical approach using time-dependent climate predictors. The objective of Bayesian inference is to compute the posterior distribution of the desired variables, in this case the parameters of the prediction models for seasonal mean flow and the annual maximum flood distribution. The posterior distribution $p(\theta|x)$ is given by the Bayes theorem, as follows:

$$p(\theta|x) = \frac{p(\theta)p(x\,|\,\theta)}{p(x)} = \frac{p(\theta)p(x\,|\,\theta)}{\int_{\Theta} p(\theta)p(x\,|\,\theta)d\theta} \propto p(\theta)p(x\,|\,\theta) \tag{1}$$

where θ is the vector of parameters of the distribution to be fitted, Θ is the space parameter, $p(x|\theta)$ is the likelihood function, x is the vector of observations and $p(\theta)$ is the prior distribution. Here, we present a method for incorporating climate information into updated estimates of the parameters for the distribution (e.g. Normal and Gumbel) used to represent the seasonal mean flow and annual maximum flood.

A hierarchical model allows the parameter values of the probability distribution (model) to be themselves functions of a (regression) model based on climate indices. Thus, the first model layer consists of the probability distribution, with two parameters and the second layer is a predictive regression model for these parameters. In this Bayesian hierarchical regression the parameters are hypothesized to be functions of climate indicators, such as SST and snow cover that are developed here, and others, such as ENSO and PDO (Pacific Decadal Oscillation), that are generally recognized climate phenomena. Analysis of the historical distribution of the mean seasonal flow for JJA shows that a normal distribution cannot be rejected, according to the chi-squared goodness-of-fit test (not shown). Using the normal distribution, the distribution of seasonal mean flood $Y_{\text{mean}}(t)$ can be modelled as follows:

$$Y_{\text{mean}}(t) \sim \text{Normal}(\mu(t), \sigma(t)) \tag{2}$$

where ~ signifies that the variable, $Y_{\text{mean}}(t)$ is distributed according to the distribution that follows, i.e. a normal distribution with mean $\mu(t)$ and standard deviation $\sigma(t)$. In the present application, the standard deviation, $\sigma(t)$, was not found to vary significantly as a function of time. As a result, equation (2) is changed to remove the time dependence of σ:

$$Y_{\text{mean}}(t) \sim \text{Normal}(\mu(t), \sigma) \tag{3}$$

Then, the mean of the distribution is modelled as a linear combination of the regression parameters and predictors:

$$\mu(t) = \beta_1 \boldsymbol{X}_{1i}(t) + ... + \beta_k \boldsymbol{X}_{ki}(t) \tag{4}$$

where $\boldsymbol{X}_i$, climate indices, are the ith rows of the known design matrices $\boldsymbol{X}$, and $\boldsymbol{\beta}$ is a vector of regression parameters which is normally distributed, $\beta_k \sim N(\eta_{\beta_k}, \sigma_k)$. The parameters are fitted such that the posterior distribution of the parameters is maximized for the given set of observations. In the same manner, the probability distribution of the annual peak flow $Y_{\text{peak}}(t)$ can be modelled by using the Gumbel type extreme distribution ($Y_{\text{peak}}(t) \sim \text{Gumbel}(\mu(t),\sigma)$). To formally account for the uncertainty in the regression parameters, they are assumed to be normally distributed with unknown parameters η_{β_k} and σ_k such that $\eta_\beta \sim N(0.0, \sigma_{\sigma_k})$. The hierarchical Bayesian structure also requires that we assume distributions for σ_k, and σ_{σ_k} and in both cases we assume they follow a half-Cauchy distribution, which is a standard approach for non-negative variables (Gelman, 2006). These distributions are then sampled using the Gibbs sampler method for parameter estimation, as described below.

We follow a Bayesian approach to estimation and inference, developing an efficient data augmentation algorithm for posterior computation. Specifically, the maximization of the posterior distribution expression is required to produce the posterior distribution of each parameter and hyper-parameter. Here we use the Gibbs sampler which is an effective Markov Chain Monte Carlo method for simulating the posterior probability distribution of the data field conditional on the current choice of parameters (Gelman *et al*., 2000, 2003; Godsill *et al*., 2001; Tsionas, 2001; Hue *et al*., 2002; Ridgeway & Madigan, 2003; Tucker & Liu, 2003; Chen *et al*., 2005). The use of the Gibbs sampler, as discussed in the context of hierarchical Bayesian models, enables a simple sampling-based solution to the problem of parameter estimation (Gilks *et al*., 1995). It is a special case of the Metropolis method, for which the acceptance parameter is set equal to one (a = 1). For more information on the use of the Gibbs sampler for Markov Chain Monte Carlo applications, see Gilks *et al*. (1995).

3 YANGTZE RIVER FLOW TELECONNECTION TO CLIMATE INFORMATION

Several studies of climate influences on precipitation in East Asia provide evidence that the streamflow of the Yangtze River may exhibit climate teleconnections. While the relationship between East Asia rainfall and ENSO is inconclusive, and similarly the relationship between streamflow and ENSO is not significant, there are other modes of ocean temperature variability that do have significant influence. Previous studies have found conditions in the western Pacific and Indian oceans and areas in the central and eastern Pacific that are closely linked with rainfall in the Yangtze River basin (Chang *et al*., 2000; Wang *et al*., 2000; Yang & Lau, 2004).

In a previous study, Xu *et al*. (2007) used the spring (MAM) averages of the sea-surface temperatures (SSTs), outgoing longwave radiation (OLR), precipitation, snow cover and sea level pressure (SLP) as the predictors of monsoon (JJA) seasonal mean flow measured at the YHS, the same gauge as used in this study. Strong correlations were found with SSTs, OLR, PREC (precipitation) and SLP in primarily the western Pacific, and in the case of SST, the eastern Indian Ocean. Rank correlations, which are more robust for non-normal data, yielded similar results.

Xu *et al*. (2007) constructed a prediction model of JJA seasonal flow at YHS using a quadratic model of the SSTs and snow cover variables. The results were consistent with previous studies showing SSTs in the eastern Indian and western Pacific influencing rainfall and streamflow in East Asia (Wu *et al*., 2003; Yang & Lau, 2004). In the current analysis, the objective was prediction of the wet season streamflow. In order to assess annual peak flood predictability, the preceding seasonal MAM SSTs and snow cover were re-examined and used in this study. As before, global SSTs and snow cover were evaluated as predictors. Linear correlation maps were constructed using the preceding seasonal climate variables (e.g. SST and snow cover) with streamflow at YHS. Data from these regions were extracted and transformed into climate indices. The indices were formed using a spatial average over the areas of interest of the gridded time series. The spatial pattern of correlations with the JJA mean flow and the annual peak flow series is displayed in

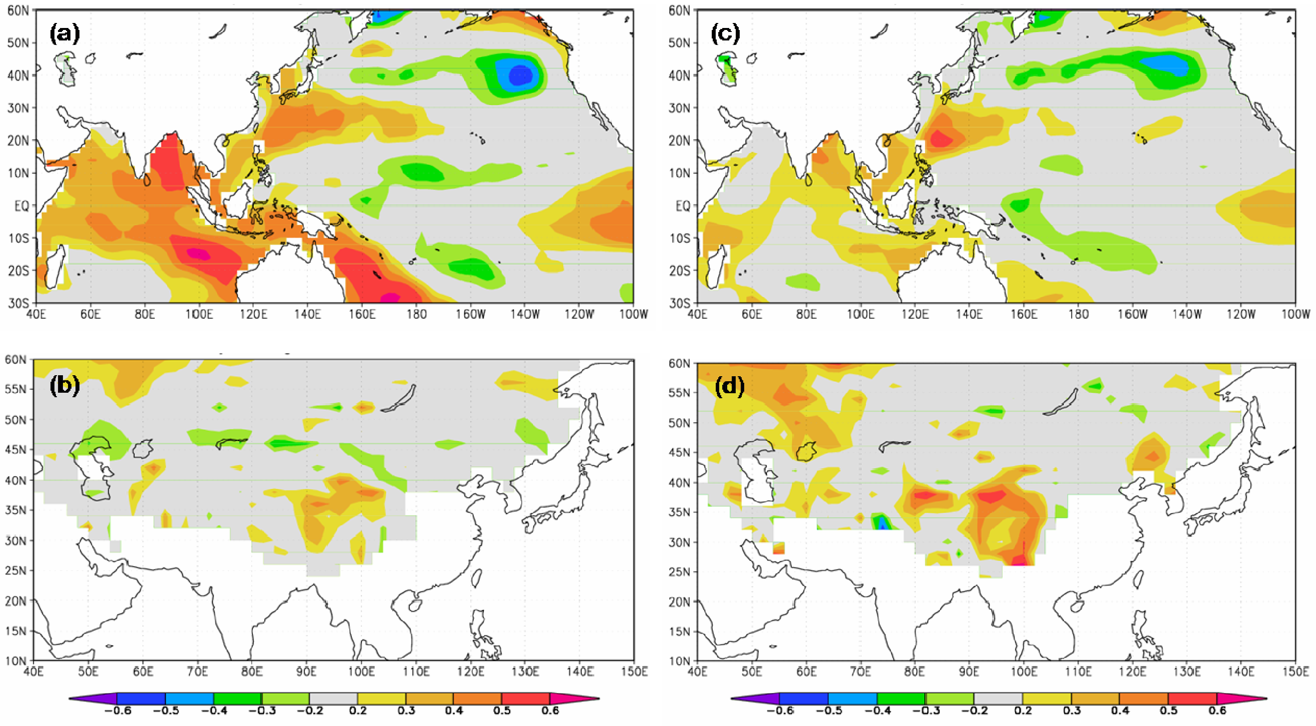

Fig. 4 Linear correlation maps between selected climate variables averaged over MAM and YHS streamflow for JJA for the years 1970–2001: (a) SST *vs* mean JJA flow (1970–2001); (b) snow *vs* mean JJA flow (1970–2001); (c) SST *vs* peak flow (1970–2001); and (d) snow *vs* peak flow (1970–2001).

Fig. 4(a) and (c), respectively. In both cases, anomalous conditions in SSTs are likely to strongly influence the flow at YHS. The area of negative correlation with SSTs located near the equator, which we designate SST1, is likely related to an ENSO signal. During ENSO warm events, East Asian monsoon circulation is weakened, causing less rainfall in northern China and leading to increased rainfall, streamflow and flooding in south central China as the subtropical high remains to the south. The area of positive SST correlation in the east Indian Ocean–western Pacific warm pool area, which we designate SST2, corresponds to increased rainfall (and thus streamflow) during non-ENSO years due to warm SSTs in that area. These correlations with the flood series are statistically significant. The area of positive SST correlation in the western part of the subtropical Pacific near the East China Sea, which we designate SST3, with the warm SST anomalies continuing east of Japan, relates to Tropical cyclone genesis which may play an important role in bringing moisture to that area. Camargo *et al*. (2007a,b) proposed a new probabilistic clustering technique, based on a regression mixture model, to describe tropical cyclone trajectories in the western North Pacific. These clusters are then analysed in terms of genesis location, trajectory, landfall, intensity, and seasonality. This study provides evidence that warm SST anomalies patterns in the western part of the subtropical Pacific, around 30°N, are strongly related to tropical cyclones clusters that make landfall in Southeast Asia and southern China.

In addition to summer monsoon rainfall, the snowmelt in the mountainous headwaters region in western China is a significant component of water flux in the Yangtze River. Although seemingly readily apparent, we have not seen snow cover used previously as a predictor of Yangtze flows. The mechanisms for peak values are heavy rains falling during the spring thaw of a large snow coverage area. Here we evaluate the influence of the preceding snow cover data near the streamflow station. The snow index is located in the high elevation headwaters of the Yangtze and represents potential snowmelt runoff. The relationships between snow cover and mean JJA flow and peak flow are displayed in Fig. 4(b) and (d), respectively.

From Fig. 4 potential predictors were identified from the SST and snow cover data sets according to regions of high correlation. Rectangular zones that encompassed these regions were specified as follows: SST (SST1: –10°N–+10°N to 150°E–180°E; SST2: –20°N–0° to 75°E–110°E; SST3: 10°N–30°N to 130°E–150°E), and snow cover (–10°N–0° 200°E–230°E). The MAM values for each variable were spatially averaged over the selected box. Table 2 lists the correlation coefficients of the spatially averaged MAM indices with mean JJA runoff and annual peak flow with 95% and 90% confidence.

Table 2 Predictors based on spatial average of selected zones from the global climate data sources and their correlation with each predictand.

Climate prediction	Zone selected:		JJA seasonal flow	Annual peak flow
SST1	–10°N–10°N	150°E–180°E	–0.27 †	–0.28 †
SST2	–20°N–0°	75°E–110°E	0.51 *	0.20 *
SST3	10°N–30°N	130°E–150°E	0.38 *	0.45 *
Snow	–10°N–0°	200°E–230°E	0.42 *	0.42 *

* Significant at 95% confidence.
† Significant at 90% confidence.

4 YANGTZE RIVER FLOW PREDICTION

In a previous study (Xu et al., 2007), a quadratic regression model was used for the prediction of seasonal (JJA) streamflow. Additional exploratory data analysis confirmed that the data shows mildly nonlinear relationships between the YHS streamflow and some of the predictors. Consequently, we explored the quadratic regression models. As described above, a hierarchical Bayesian prediction model is developed for seasonal mean flow and annual peak flow at YHS using selected climate variables, specifically SSTs and snow cover, using MAM values. The detailed prediction model for each case and for each year, *t*, can be formulated as follows:

Seasonal forecasting:

$$Y_{\text{mean}}(t) \sim \text{Normal}(\mu_{\text{mean}}(t), \sigma_{\text{mean}}) \tag{5}$$

$$\mu_{\text{mean}}(t) = \beta_{m1} + \beta_{m2} \cdot \text{SST1}(t) + \beta_{m3} \cdot \text{SST1}^2(t) + \beta_{m4} \cdot \text{SST2}(t) + \beta_{m5} \cdot \text{Snow}^2(t) \tag{6}$$

Peak flow forecasting:

$$Y_{\text{peak}}(t) \sim \text{Gumbel}(\mu_{\text{peak}}(t), \sigma_{\text{peak}}) \tag{7}$$

$$\mu_{\text{peak}}(t) = \beta_{p1} + \beta_{p2}\text{SST1}^2(t) + \beta_{p3}\text{SST3}(t) + \beta_{p4}\text{Snow}^2(t) \tag{8}$$

The hierarchical Bayesian regression models are solved simultaneously in a Bayesian framework. Non-informative priors are assumed for each of the parameters β_m and β_p, and their optimal values are selected through a maximization of the posterior likelihood associated with the quadratic regresion models. A Markov chain Monte Carlo (MCMC) procedure is used. In particular, the Gibbs sampling approach to MCMC (Gilks *et al*., 1995) has been used in this study. To get an idea for the Gibbs sampling, a simple example is provided. Suppose we want to sample s values of α from a joint distribution $f(\alpha,\beta)$. The Gibbs sampling begins with a value of β_0 and sample α by $a_i \sim p(\alpha|\beta = \beta_{i-1})$. Once that value of α is sampled, repeat by sampling for the next β by $\beta_i \sim p(\beta|\alpha = a_{i-1})$. Similarly, the β parameters in equations (6) and (8) are derived through the Gibbs sampler. We chose to run three chains simultaneously searching for optimal parameters. The evolution of each chain was monitored to check for convergence to a common value. Selection of the hyper-priors and the appropriateness of the prior distributions and the model structure were judged by the deviance information criterion (DIC) (Berg *et al*., 2004). The optimal significant predictors from a set of independent variables are selected by the stepwise regression method for each station.

Table 3 Posterior median estimates and credible intervals for selected parameters from hierarchical Bayesian seasonal flow prediction model at YHS.

Description	Node	Mean	Std dev.	2.50%	Median	97.50%
Seasonal flow model:						
Interceptor	β_1	2.273	0.074	2.129	2.273	2.420
SST1	β_2	–0.111	0.050	–0.209	–0.111	–0.011
SST1^2	β_3	0.130	0.048	0.035	0.130	0.224
SST2	β_4	0.276	0.051	0.176	0.276	0.377
Snow^2	β_5	0.083	0.025	0.034	0.083	0.132
Performance measure		R	CoE	IoA	Bias	RMSE
Seasonal (JJA)		0.802	0.643	0.886	0.001	0.231
Peak flow model:						
Interceptor	β_1	4.174	0.195	3.791	4.171	4.548
SST1	β_2	0.198	0.119	–0.055	0.203	0.423
SST1^2	β_3	0.699	0.148	0.410	0.706	0.986
SST2	β_4	–0.089	0.079	–0.264	–0.085	0.053
Snow^2	β_5	0.302	0.098	0.091	0.310	0.473
Performance measure		R	CoE	IoA	Bias	RMSE
Seasonal (JJA)		0.729	0.531	0.828	–0.001	0.602

Table 3 summarizes the results from the Bayesian model. Values are estimates and statistics of the regression coefficients for predictors. The one parameter is β_1 with the regression intercept, and others are the regression coefficients denoted with β_2, … β_{J+1}, where J is the number of predictors. The mean, standard deviation, and 95% credible interval are based on a hierarchical Bayesian specification of a nonlinear regression model where streamflow is governed by the predictors. It is found that the coefficients for the predictor variables are statistically significant.

The posterior distribution for each model parameter is presented in Fig. 5. The figures represent the uncertainty distribution as derived from the hierarchical Bayesian inference relating to the parameter's uncertainty for the seasonal and the annual peak prediction model. In the case of the seasonal flow prediction, the distribution of β_3 and β_4 show relatively tight uncertainty bounds compared to β_2 and β_5. An advantage of this approach is that the uncertainties of these parameters can be propagated to the model prediction, as can the uncertainty in the model structure itself.

There are several methods available to measure the goodness of fit or prediction skill for a particular hydrological forecasting model. Legates & McCabe (1999) have critically reviewed many of these principal statistics. For more details regarding goodness-of-fit measures, see Legates & McCabe (1999) and Willmott *et al.* (1985). Statistics of the seasonal flow model are shown in Table 3. The seasonal flow predicted by the model using SST1, SST2 and snow cover exhibits strong correlation with the observed streamflow, with R = 0.8 and a coefficient of efficiency of 0.64.

The best model for the peak flow as determined by the stepwise regression procedure consisted of a quadratic model of the variables SST1, SST3, and SNOW. Statistics of the model are shown in Table 3. The strongest predictors were SST3 and the snow cover variable. These were retained for the final model of peak annual flow at YHS. The time series of observed peak flow and the values predicted with the hierarchical Bayesian model with two predictors are shown in Fig. 6(b). The model shows reasonably good skill, with a correlation between observed and predicted peak annual flow of 0.73 and coefficient of efficiency of 0.53.

The peak annual flood model was then used to estimate the variability in a design flood estimation based on the state of the predictor variables. Estimates of the annual flood value with a 2% probability of occurring in a given year (50-year flood) were produced using the Bayesian model based on the predictors cited above. The time series are shown in Fig. 7 with the observed values of annual peak flood. The time series shows considerable variability in year-to-year flood risk. In other words, the estimated 50-year design flood value differs based on the state of the

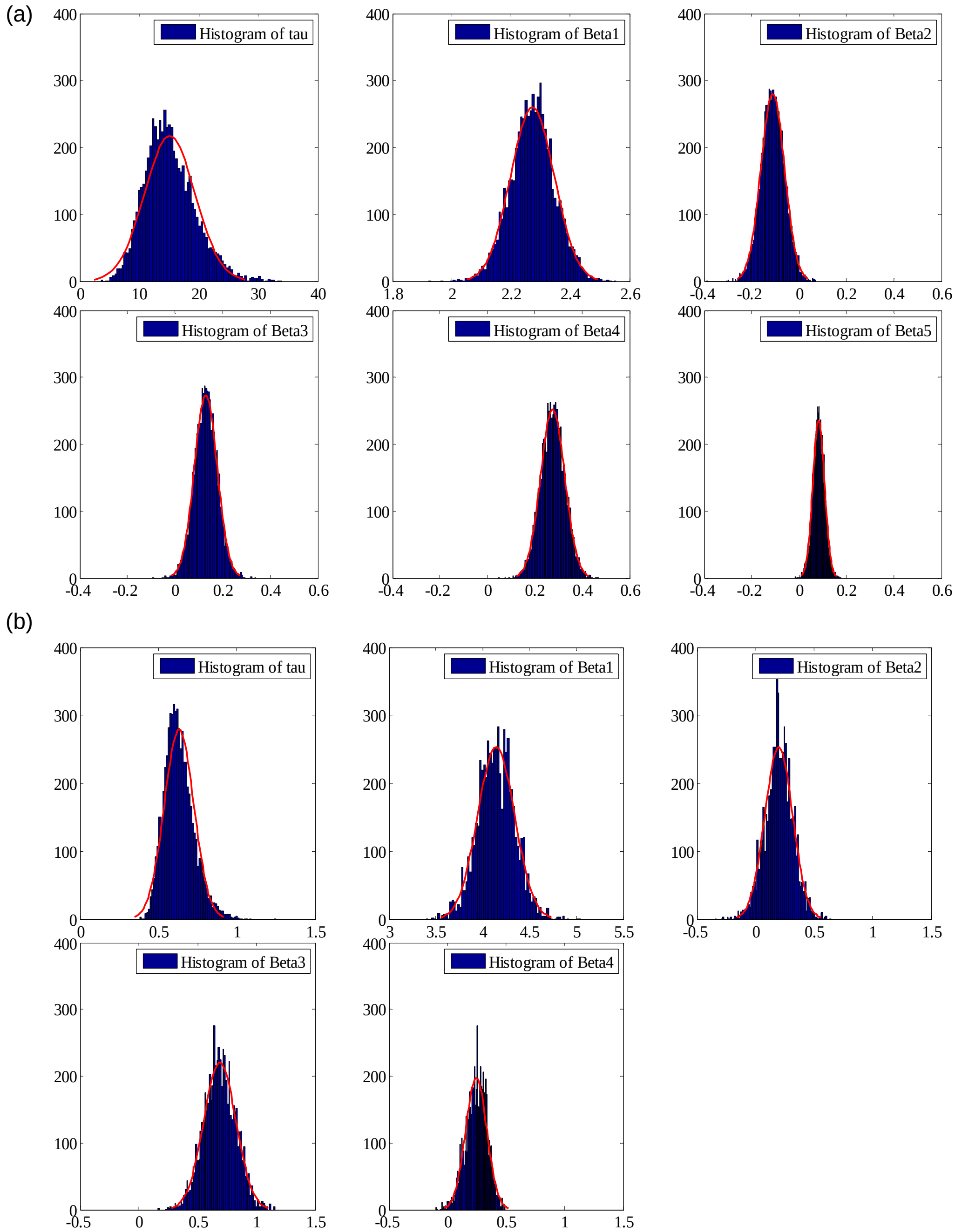

Fig. 5 Histogram of the parameters (regression coefficients) from hierarchical Bayesian regression model for: (a) seasonal flow model and (b) annual peak flow model.

predictor variables which vary from year to year, and in this case are a source of predictability. Each time series also includes the confidence interval based on the uncertainties associated with the parameter estimation of Table 4. The figure also portrays the skill of the climate predictors in indicating the pronounced increases and decreases in flood risk inter-annually.

Such a prediction of the changes in design flood could be valuable from a reservoir operations standpoint. Since, in some years, the design flood is smaller than others, potentially less volume need be reserved for flood retention and could be used for additional storage to provide more water for the dryer months ahead. Alternatively, given an indication of enhanced flood risk based on the state of the predictor variables, a greater volume could be reserved for flood retention, and

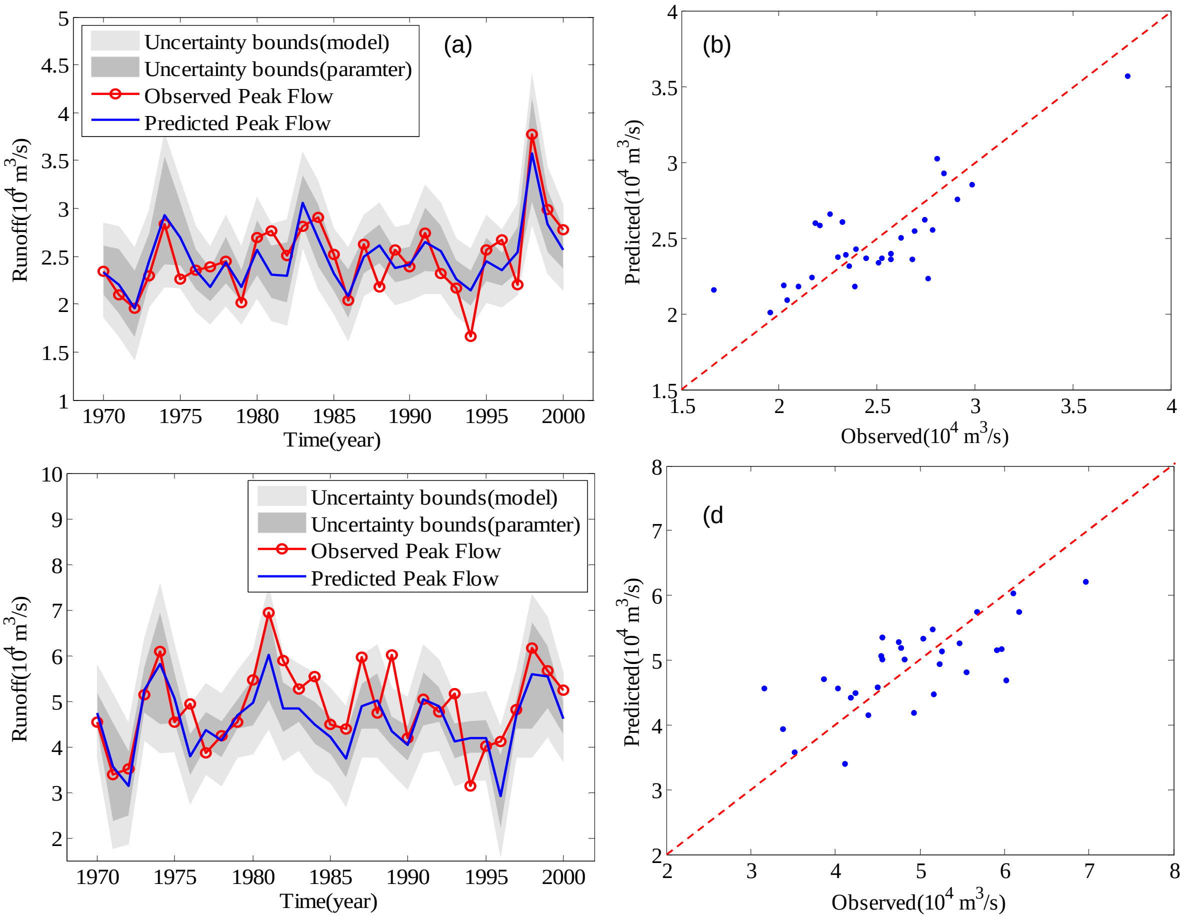

Fig. 6 Time series of: (a) seasonal flow prediction results (solid line) and observed JJA streamflow (open circle); (c) peak flow prediction results (solid line) and observed annual peak flow at Yichang Hydrological Station using hierarchical Bayesian regression with MAM values of SST1, SST2, SST3 and SNOW as predictors for 1970–2001. Scatter plot of (b) seasonal flow prediction model and (d) peak flow prediction model.

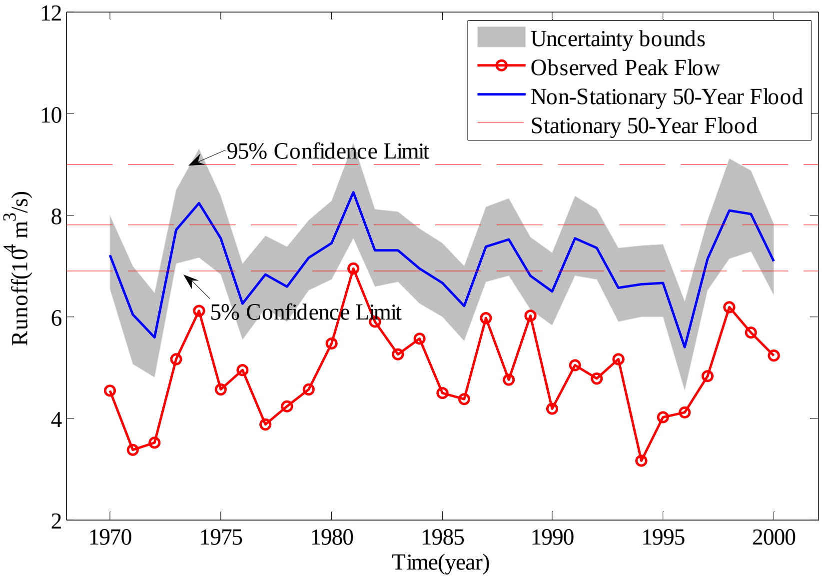

Fig. 7 Dynamic 50-year return period flood with SST and snow cover predictor. Solid line shows dynamic 50-year flood using different location parameters in time and dashed line indicates confidence interval derived by stationary flood frequency analysis.

Table 4 Gumbel distribution parameters and 50-year return period peak flow with stationary and non-stationary cases.

	Parameters		50-year return period peak flow		
	μ	σ	5%	Median	9.5%
Stationary	4.46	0.86	6.88	7.79	8.98
Modelled non-stationary	4.60	0.62	5.58	7.03	8.45

appropriate planning could be initiated to adjust to the reduced storage available. Such predictive skill may be especially useful for the Yangtze, where floods have continually wreaked havoc throughout the basin.

4 SUMMARY AND DISCUSSION

This study investigated the predictability of the peak annual flood on the Yangtze River at the site of the Three Gorges Dam. It also demonstrated the application of hierarchical Bayesian modelling for prediction of both peak annual and mean seasonal flow. The major advantage of the approach is that it allows an empirical estimation of the parameter uncertainty and higher confidence in the uncertainty bands accompanying the prediction. The model results show satisfactory predictive skill, with a correlation between predicted and observed mean seasonal flow (JJA) of 0.8 and between predicted and observed annual peak flow of 0.73. Each model utilizes predictors that are available one season ahead of the prediction period.

The skill in the prediction of the annual peak flow may be of particular interest for dam operations. Often, a volume of potential storage is left empty to capture some design flood. As the present results show, the expected flood volume changes significantly from year to year. As a result, the design flood capture volume is likely too large in many years, resulting in lost potential benefits, and too small in other years, leading to less than expected flood risk protection. Forecasts could be used to modify the flood storage volume to be reflective of the risk faced in each year, based on observations of MAM snow cover and SSTs. Although we would not advocate a modification of the operating policy of the Three Gorges Dam on these results alone, we do advocate further investigation of the possible use of seasonal forecasting to improve the performance of reservoir operations there.

REFERENCES

Berg, A., Meyer, R. & Yu, J. (2004) Deviance information criterion for comparing stochastic volatility models. *J. Business & Economic Statist.* **22**(1), 107–120.

Brown, C. & Rogers, P. (2006) Effect of forecast-based pricing on irrigated agriculture: a simulation. *J. Water Resour. Plan. Manage. ASCE* **122**(6), 403–413.

Brown, C., Rogers, P. & Lall, U. (2006) Demand management of groundwater with monsoon forecasting. *Agric. Systems* **90**, 293–311.

Camargo, S. J., Robertson, A. W., Gaffney, S. J., Smyth, P. & Ghil, M. (2007a) Cluster analysis of typhoon tracks. Part I: General properties. *J. Climate* **20**, 3635–3653.

Camargo, S. J., Robertson, A. W., Gaffney, S. J., Smyth, P. & Ghil, M. (2007b) Cluster analysis of typhoon tracks. Part II: Large-scale circulation and ENSO. *J. Climate* **20**, 3654–3676.

Chang, C. P., Zhang, Y. & Li, T. (2000) Interannual and interdecadal variations of the East Asian summer monsoon and tropical Pacific SSTs, Part I: Role of subtropic ridges. *J. Climate* **13**, 4310–4325.

Chen, H. J., Lee, K. C., Murphy-Chutorian, E. & Triesch, J. (2005) Toward a unified probabilistic framework for object recognition and segmentation. *Adv. Visual Computing* **3804**, 108–117.

Chiang, J. C. H., Kushnir, Y. & Zebiak, S. E. (2000) Interdecadal changes in eastern Pacific ITCZ variability and its influence on the Atlantic ITCZ. *Geophys. Res. Lett.* **27**, 3687–3690.

Croley, T. E. (2003) Weighted-climate parametric hydrologic forecasting. *J. Hydrol. Engng* **8**(4), 171–180.

Fasullo, J. & Webster, P. J. (2002) Hydrologic signatures relating the Asian summer monsoon and ENSO. *J. Climate* **15**, 3082–3095.

Franks, S. W. & Kuczera, G. (2002) Flood frequency analysis: Evidence and implications of secular climate variability, New South Wales. *Water Resour. Res*. **38**(5), 1062. doi:10.1029/2001WR000232.

Gelman, A. (2006) Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1**, 515–533.

Gelman, A. Carlin, J. B., Stern, H. S. & Rubin, D. B. (2003) *Bayesian Data Analysis* (second edn). Chapman & Hall/CRC, New York, USA.

Gilks, W. R., Best, N. G. & Tan, K. K. C. (1995) Adaptive rejection metropolis sampling within Gibbs sampling. *Appl. Stat.– J. Roy. Statist. Soc.* Ser. C **44**, 455–472.

Godsill, S., Doucet, A. & West, M. (2001) Maximum *a posteriori* sequence estimation using Monte Carlo particle filters. *Ann. Inst. Statist. Math*. **53**, 82–96.

Gong, D. Y. & Ho, C. H. (2002) Shift in the summer rainfall over the Yangtze River valley in the late 1970s. *Geophy. Res. Lett.* **29**(10), 78-1–78-4.

Halpert, M. S. & Ropelewski, C. F. (1992) Surface-temperature patterns associated with the Southern Oscillation. *J. Climate* **5**, 577–593.

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

Hamlet, A. F., Huppert, D. & Lettenmaier, D. P. (2002) Economic value of long-lead streamflow forecasts for Columbia River hydropower. *J. Water Resour. Plan. Manage.* **128**(2), 91–101.

Huang, R. H. & Wu, Y. F. (1989) The influence of ENSO on the summer climate change in China and its mechanism. *Adv. Atmos. Sci*. **6**, 21–33.

Hue, C. Le Cadre, J. P & Perez, P. (2002) Sequential Monte Carlo methods for multiple target tracking and data fusion. *IEEE Trans. Signal Processing* **50**, 309–325.

Jain, S. & Lall, U. (2000) Magnitude and timing of annual maximum floods: trends and large-scale climatic associations for the Blacksmith Fork River, Utah. *Water Resour. Res.* **36**, 3641–3651.

Jain, S. & Lall, U. (2001) Floods in a changing climate: does the past represent the future? *Water Resour. Res.* **37**, 3193–3205.

Kaplan, A., Cane, M. A., Kushnir, Y., Clement, A. C, Blumenthal, M. B. & Rajagopalan, B. (1998) Analyses of global sea surface temperature 1856–1991. *J. Geophy. Res*. **103**, 18 567–18 589.

Kim, Y. O. & Palmer, R. N. (1997) Value of seasonal flow forecasts in Bayesian stochastic programming. *J. Water Resour. Plan. Manage*. **123**(6), 327–335.

Kwon, H. H., Lall. U., Moon, Y.-I., Khalil, A. F. & Ahn, H. (2006) Episodic interannual climate oscillations and their influence on seasonal rainfall in the Everglades National Park. *Water Resour. Res*. **42**, W11404. doi:10.1029/2006WR005017.

Legates, D. R. & McCabe, G. J. (1999) Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* **35**(1), 233–241.

Lu, X. X., Ashmore, P. & Wang, J. F. (2003) Seasonal water discharge and sediment load changes in the Upper Yangtze, China. *Mountain Res. Develop.* **23**(1), 56–64.

Milly, P. C. D., Wetherald, R. T. , Dunne, K. A. & Delworth, T. L. (2002) Increasing risk of great floods in a changing climate. *Nature* **415**, 514–517.

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

Pizaro, G. & Lall, U. (2002) El Niño and floods in the US West: what can we expect? *EOS Trans. Am. Geophys. Union* **83**, 349–352.

Ridgeway, G. & Madigan, D. (2003) A sequential Monte Carlo method for Bayesian analysis of massive datasets. *Data Mining and Knowledge Discovery* **7**, 301–319.

Robinson, D., Dewey, K. & Heim. R. (1993) Global snow cover monitoring: an update. *Bull. Am. Met. Soc*. **74**(9), 1689–1696.

Ropelewski, C. F. & Halpert, M. S. (1986) North-American precipitation and temperature patterns associated with the El-Niño Southern Oscillation (ENSO). *Monthly Weather Review* **114**, 2352–2362.

Ropelewski, C. F. & Halpert, M. S. (1987) Global and regional scale precipitation patterns associated with the El-Niño Southern Oscillation. *Monthly Weather Review* **115**, 1606–1626.

Sankarasubramanian, A. & Lall, U. (2003) Flood quantiles in a changing climate: seasonal forecasts and causal relations. *Water Resour. Res*. **39**(5), 1134. doi:10.1029/2002WR001593.

Souza, F. A. & Lall, U. (2003) Seasonal to interannual ensemble streamflow forecasts for Ceara, Brazil: applications of a multivariate, semiparametric algorithm. *Water Resour. Res.* **39**(11), 1307.

Souza, F. A. & Brown, C. (2009) Performance of water policy reforms under scarcity conditions: a case study in Northeast Brazil *Water Policy* (in press).

Trigo, R. M, Zêzere, J. L., Rodrigues, M. L. & Trigo, I. F. (2005) The influence of the North Atlantic Oscillation on rainfall trig gering of landslides near Lisbon. *Natural Hazards* **36**, 331–354. doi:10.1007/s11069-005-1709-0.

Tucker, A. & Liu, X. H. (2003) Learning dynamic Bayesian networks from multivariate time series with changing dependencies. *Adv. Intell. Data Analysis* **V**, 2810, 100–110.

Tsionas, E. G. (2001) Bayesian multivariate Poisson regression. *Commun. in Statistics – Theory and Methods* **30**, 243–255.

Wang, B., Wu, R. & Fu, X. (2000) Pacific–East Asian teleconnection: how does ENSO affect East Asian climate? *J. Climate* **13**, 1517–1536.

Westphal, K. S., Vogel, R. M., Kirshen, P. & Chapra, S. C. (2003) Decision support system for adaptive water supply management. *J. Water Resour. Plan. Manage*. **129**(3), 165–177.

Willmott, C. J., Ackleson, S. G., Davis, R. E., Feddema, J. J., Klink, K. M., Legates, D. R. J., O'Donnell, J. & Rowe, C. M. (1985) Statistics for the evaluation and comparison of models. *J. Geophys. Res.* **90**(C5), 8995–9005.

Wu, R., Hu, Z. Z., Kirtman, B. P. (2003) Evolution of ENSO-related rainfall anomalies in East Asia. *J. Climate* **16**, 3741–3757.

Xiong, L. H. & Guo, S. L. (2004) Trend test and change-point detection for the annual discharge series of the Yangtze River at the Yichang hydrological station. *Hydrol. Sci. J.* **49**(1), 99–112

Xu, K. Q., Zhang, J. Q., Watanabe, M. & Sun, C. P.(2004) Estimating river discharge from very high-resolution satellite data: a case study in the Yangtze River, China. *Hydrol. Processes* **18**(10), 1927–1939. doi: 10.1002/hyp.1458.

Xu, K. Q., Chen, Z., Zhao, Y., Wang, Z., Zhang, J., Hayashi, S., Murakami, S. & Watanabe, M. (2005) Simulated sediment flux during 1998 big-flood of the Yangtze (Changjiang) River, China. *J. Hydrol.* 313, 221–233. doi:10.1016/j.jhydrol.

Xu, K. Q., Brown, C., Kwon, H., Lall, U., Zhang, J., Hayashi, S. & Chen, Z. (2007) Climate teleconnections to Yangtze River seasonal streamflow at the Three Gorges Dam, China. *Int. J. Climatol.* **27**(6), 771–780. doi:10.1002/joc.1437.

Yang, F. L. & Lau, K. M. (2004) Trend and variability of China precipitation in spring and summer: linkage to sea-surface temperatures. *Int. J. Climatol.* **24**, 1625–1644.

Received 15 January 2009; accepted 17 February 2009