



Predictability of Western Himalayan river flow: melt seasonal inflow into Bhakra Reservoir in northern India

I. Pal^{1,2}, U. Lall³, A. W. Robertson², M. A. Cane⁴, and R. Bansal⁵

¹Department of Civil Engineering, University of Colorado, Denver, CO 80204, USA

²International Research Institute for Climate and Society, The Earth Institute at Columbia University, Palisades, NY 10964, USA

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

⁴Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY 10964, USA

⁵Bhakra Beas Management Board, Punjab, India

Correspondence to: I. Pal (indrani.pal@ucdenver.edu)

Received: 9 March 2012 – Published in Hydrol. Earth Syst. Sci. Discuss.: 4 July 2012

Revised: 18 March 2013 – Accepted: 5 May 2013 – Published: 5 June 2013

Abstract. Snowmelt-dominated streamflow of the Western Himalayan rivers is an important water resource during the dry pre-monsoon spring months to meet the irrigation and hydropower needs in northern India. Here we study the seasonal prediction of melt-dominated total inflow into the Bhakra Dam in northern India based on statistical relationships with meteorological variables during the preceding winter. Total inflow into the Bhakra Dam includes the Satluj River flow together with a flow diversion from its tributary, the Beas River. Both are tributaries of the Indus River that originate from the Western Himalayas, which is an understudied region. Average measured winter snow volume at the upper-elevation stations and corresponding lower-elevation rainfall and temperature of the Satluj River basin were considered as empirical predictors. Akaike information criteria (AIC) and Bayesian information criteria (BIC) were used to select the best subset of inputs from all the possible combinations of predictors for a multiple linear regression framework. To test for potential issues arising due to multicollinearity of the predictor variables, cross-validated prediction skills of the best subset were also compared with the prediction skills of principal component regression (PCR) and partial least squares regression (PLSR) techniques, which yielded broadly similar results. As a whole, the forecasts of the melt season at the end of winter and as the melt season commences were shown to have potential skill for guiding the development of stochastic optimization models to man-

age the trade-off between irrigation and hydropower releases versus flood control during the annual fill cycle of the Bhakra Reservoir, a major energy and irrigation source in the region.

1 Introduction

The Satluj River is one of the five main tributaries of the Indus river that traverse the Punjab region of northern India and Pakistan, whose name translates as “the land of five rivers”. The waters of the Satluj are allocated to India under the Indus Waters Treaty between India and Pakistan, and are mostly diverted to irrigation canals in India. There are several major hydroelectric plants on the Satluj, and the 1325 MW Bhakra Dam at the foothill of the Himalayas is the largest of them. The Bhakra Reservoir is the lifeline for water supply in three major states in northern India including India’s “bread basket” Punjab state. The current total inflow to the Bhakra Dam comprises the Satluj River flow (65–80 %) and the flow diverted from a tributary of the Satluj, the Beas River (via the Beas Satluj Link (BSL), 20–35 %) (location is shown in Fig. 1). Melting snow and ice provide the water supply to much of the Himalayan region during the dry months before the summer monsoon. Snowmelt is most important in the Western Himalayas where meltwater comprises up to about 70 % of the annual discharge of the Indus and its tributaries (Kattelmann, 1993; Archer, 2003). Both the Satluj and the

Beas originate from the Western Himalayas. More than 50 % of the annual flow of the Satluj River is contributed by snow and ice melt (Singh and Jain, 2002, 2003). Forecasting seasonal meltwater mean inflows into Bhakra has the potential to improve the operational efficiency of the hydroelectric and irrigation projects. The information about snow accumulation in winter provides a key to spring total inflow with lead times of a few months (Singh and Bengtsson, 2004; Schar et al., 2004; Stewart, 2009). However, because of limited data in this region of rugged topography and poor accessibility, there have been few long lead prediction studies of the Satluj and other Himalayan catchments. No published study exists for the seasonal prediction of the mean inflow of the Bhakra Reservoir.

Previous studies have developed regression relationships between the winter snow-covered area derived from remote sensing data and spring monthly accumulated Satluj runoff (Ramamoorthy and Haefner, 1991). But the limited period of data (1980–1990 with 3 yr data missing) used in that study does not provide much statistical confidence. With an improved data network, more recent studies have simulated the “daily” flow of the Satluj River based on daily precipitation, temperature and snow cover information from the satellite images (Singh and Quick, 1993; Jain et al., 1998, 2010; Singh and Jain, 2003). While these studies have reported better results with time, these conceptual rainfall-runoff models are not useful for longer-term (seasonal) forecasting since they are based on near-real-time daily weather data. In addition, the relationship between snow-covered areas versus streamflow can be complicated by variations of snow depth (Makhdoom and Solomon, 1986). However, snow water equivalent or snow measurement information has not been considered for Satluj Basin flow prediction because such data has not been previously available. Better skills could be achieved to predict spring seasonal (Mar/Apr/May/Jun (MAMJ)) total inflow using currently available winter snow information and other hydrometeorological data.

The size of the river basin also has an important impact: for instance, Li et al. (2009) found that the larger the basin, the stronger the influence of initial conditions. Meteorological forcings also contribute to the predictive skill of seasonal hydrological forecasts as total precipitation has a predominant effect on river flow (Li et al., 2009; Matera et al., 2010). Different climatic regions provide seasonal hydrological predictability based on different variables. Real-time climate data collected at nearby monitoring sites and/or large-scale climate indices are typically used for seasonal streamflow forecasting at different regions. For example, Wedgbrow et al. (2002) analyzed dependence of river flows on several climatic variables (as climatic indices and SST) and its usefulness to forecast the summer river flow for several rivers in England. Wilby et al. (2004) carried out a seasonal forecast of the River Thames flow in England using SSTs and other variables. In the Iberian Peninsula, SST fields of Atlantic Ocean were demonstrated to be useful in seasonal stream-

flow forecast (Gamiz-Fortis et al., 2008). In Australia, information based on ENSO streamflow teleconnection and serial correlation in streamflow was demonstrated to help irrigators to take more-informed risk-based management decisions (Chiew et al., 2003; Wang et al., 2010). The North Atlantic Oscillation (NAO) or the Southern Oscillation Index (SOI) was found to provide the magnitude of seasonal streamflow in Iran (Araghinejad et al., 2006). For the past years, quantitative methods for short-term and seasonal hydrological forecasting have been under development for several subcatchments of the central Asian rivers, which receive water through melting of snow accumulated in the previous winter (Schar et al., 2004; Barlow and Tippet, 2008). It has been indicated that winter climate information from regional-scale patterns is sufficient to capture a great deal of the variability of river flow in central Asia during subsequent warm seasons for the reason that river drainage basins act as a natural spatial integrator of regional climate (Schar et al., 2004; Barlow and Tippet, 2008).

Seasonal forecasts of streamflows were issued by a number of researchers using both dynamical and statistical approaches. A range of parametric/nonparametric statistical prediction-modeling techniques has been used globally, which exhibit various levels of skill in regional streamflow forecasting. Traditional parametric methods involve fitting a linear function, also known as linear regression that assumes a Gaussian distribution of data and errors, and a linear relationship between the predictors and the dependent variable. Schar et al. (2004) and Barlow and Tippet (2008) worked on the predictability of central Asian river flow using linear regression techniques. On the other hand several nonparametric data-driven approaches, which are claimed to overcome the limitations of linear regression, are also being used in practice, such as kernel-based ones, splines, K-nearest neighbor (K-NN) local polynomials (Owosina, 1992; Rajagopalan and Lall, 1999; Souza and Lall, 2003), and locally weighted polynomials (Loader, 1999). The K-NN local polynomials and the local weighted polynomial (LOC-FIT) approaches are very similar. Some of these methods were also used in conjunction with the multimodel ensemble forecasting framework that helps in determining the probability of exceedances of various thresholds useful for the water resource management (Regonda et al., 2006; Bracken et al., 2010). Existing studies also show that multimodel ensemble forecasts tend to perform much better than a single-model forecast, particularly in short-term and seasonal climate forecast (Krishnamurti et al., 1999, 2000; Rajagopalan et al., 2002; Hagedorn et al., 2005; Wood and Lettenmaier, 2006; Singla et al., 2012). Other alternative statistical techniques to multiple-regression models have also gained wider acceptance in many hydrologic applications, including artificial neural networks (ANN), genetic algorithms, multivariate adaptive regression splines, and partial least squares (Risley et al., 2001). Unlike multiple-regression models, which assume a linear relationship between variables, these

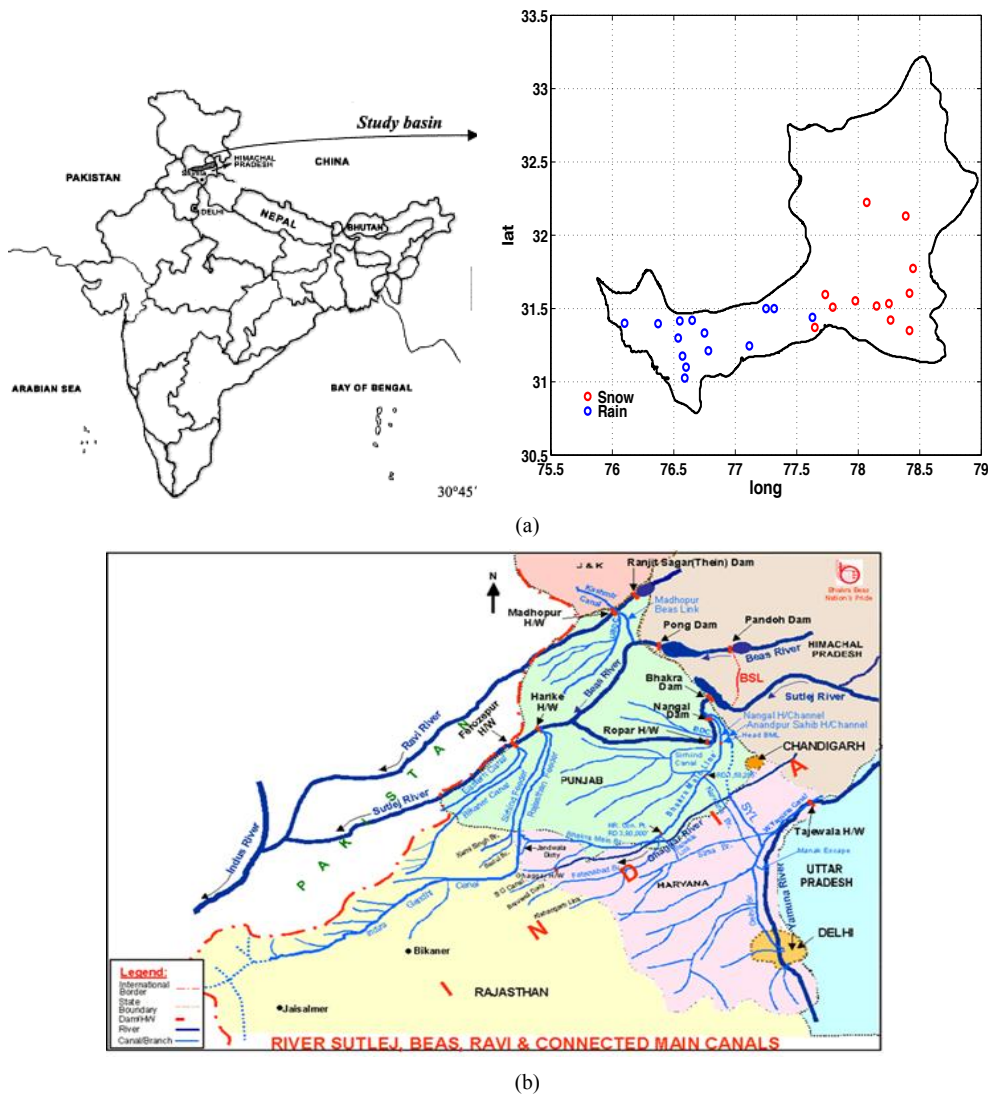


Fig. 1. (a) Indian part of Satluj River basin up to Bhakra Reservoir with location of hydrometeorological stations; (b) entire Satluj River network including location of Beas Satluj Link (BSL) channel.

methods are capable of efficiently modeling nonlinear processes that typically occur in natural systems. Canonical correlation analysis (CCA) and principal components regression (PCR) are also alternatives that are appropriate for situations where the independent variables are correlated with each other (Garen, 1992; Gamiz-Fortis et al., 2008; Barlow and Tippet, 2008).

In the current study we used basin scale winter climate information as predictors of spring seasonal river flow. Spring seasonal streamflow distribution of the Satluj River flow was found to be Gaussian; hence we proposed using a best-subset selection technique using AIC (Akaike information criteria) and BIC (Bayesian information criteria) from all variable combinations within a multiple linear regression framework and also compare this with PCR and partial least squares regression (PLSR) techniques, which will be discussed later.

2 Study region and data

The Satluj River originates from the Tibetan Plateau in the southern slopes of Mount Kailash at an elevation of more than 4500 m a.m.s.l. and flows generally west and southwest entering India in Himachal Pradesh. The entire basin area in the Tibetan Plateau experiences a cold desert winter climate and therefore the river has very low flow until it joins its major tributary, Spiti, near Namgia in India. The Spiti catchment (10 071 km²) experiences extensive snowfall due to westerly weather disturbances in the winter months that contribute to the Satluj flow in spring, or in MAMJ (Singh and Kumar, 1997), which is of our interest. In general the maximum snow cover area exists in March, by when most of the snowfall has occurred (Singh and Jain, 2003).

The Satluj River is the largest among the five rivers of Himachal Pradesh (Fig. 1). It leaves Himachal Pradesh to enter the plains of Punjab at Bhakra, where Asia's second highest gravity dam was constructed. The Bhakra Dam is the major point of water supply and electricity generation in northern India. The Satluj finally drains into the Indus in Pakistan. The river's total length is 1448 km and total drainage area up to Bhakra Reservoir is about 56 500 km². For the present study, the Indian part of the Satluj River basin up to Bhakra Reservoir (area: 22 275 km²; elevation range: 500–7000 m) was selected (Fig. 1).

The spring season (MAMJ) flow comes mainly from the runoff generated by snowmelt from the upper elevations in the greater Himalayas (Jain et al., 1998). About 65 % of the Satluj Basin area is covered with snow during winter, and about 12 % of the basin (2700 km²) is covered with permanent snowfields and glaciers (Singh and Bengtsson, 2004). The glacier melt runoff in the months of July, August and September occurs after the contribution of seasonal snowmelt, when glaciers become snow free (Singh and Quick, 1993). Peak values of total discharge in July and August are essentially due to combining this with monsoon rainfall in the lower elevations. Minimum flow is observed in winter when the base flow contribution sustains the river flow.

Spring seasonal inflow anomalies are found to be strongly correlated with large-scale precipitation and diurnal temperature range in the preceding winter over the Western Himalayas as well as adjoining northern and central Indian plains (Fig. 2), suggesting a potentially usable predictability for reservoir managers. Winter precipitation in the Western Himalayas and adjoining regions is mainly brought about by the midlatitude jet stream leading to the formation of low-pressure synoptic systems known as Western Disturbances (WD), and therefore the winter average precipitation over this entire region (Fig. 2) is related. Since total spring seasonal (MAMJ) inflow to Bhakra Dam is largely contributed by the winter snow melt, winter precipitation and temperature data available from the Indian side of Satluj Basin in addition to inflow itself were used as predictors.

Daily total inflow of Bhakra Reservoir (Satluj River flow + Beas Satluj Link (BSL) diversion) for 1963–2004, daily rainfall, snow and temperature data (T_{\max} , T_{\min}) were provided by the Bhakra Beas Management Board (BBMB) of India, the organization that is responsible for data collection and operation of the Bhakra Dam. The lower-elevation rainfall stations and upper-elevation snow stations are marked in Fig. 1. The daily snow measurements have a complete record from 1976–2006 for 12 stations (Table 1). Daily rainfall data had complete records for 1977–2006 for 15 stations (Table 1). Maximum temperature data were available only for 5 stations and a complete record was found for 1983–2005 (Table 1). Minimum-temperature data were available only for 2 stations in the entire basin (Table 1). The elevation information for various rainfall and snow stations indicates that snow

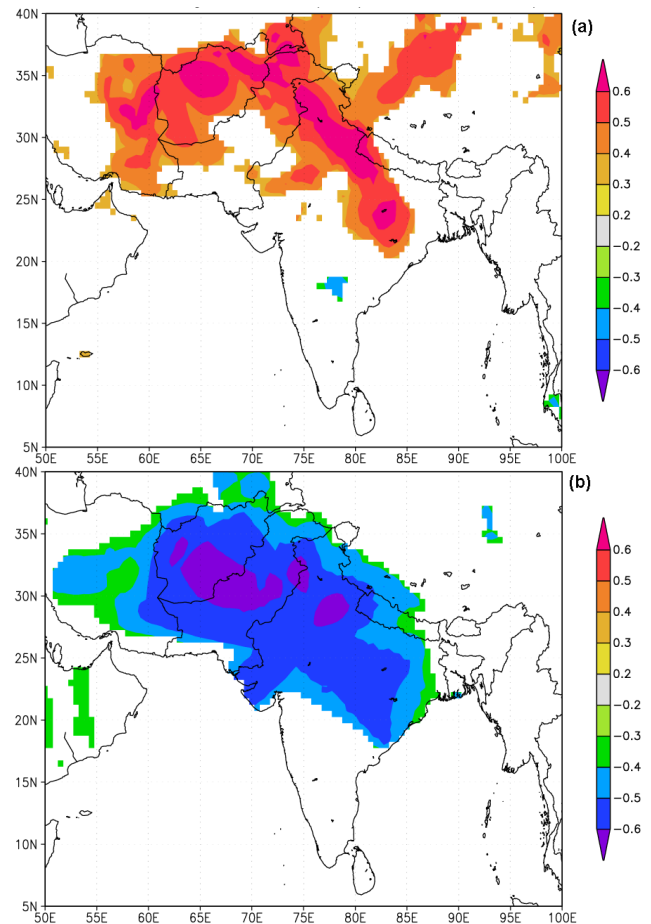


Fig. 2. Pearson's correlation fields of spring (MAMJ) seasonal total Bhakra inflow with preceding winter (DJF/DJFM) (a) precipitation and (b) daily temperature range over the Western Himalayas and adjoining northern and central Indian plains (also includes part of Pakistan and Afghanistan) for 1978–2004. Shading indicates local statistical significance level at 90 % confidence.

stations are at much higher elevation. We determined average temperature (T_{avg}) and diurnal temperature range (DTR) by averaging T_{\max} and T_{\min} , and subtracting T_{\min} from T_{\max} on a daily basis.

As mentioned earlier, total Bhakra inflow is a joint contribution from the Satluj River flow and BSL diversion that came into effect in November 1977. Consequently, the Bhakra total inflow data has a step jump since that time. To avoid this step change, we considered the flow data from 1978–2004. This study only considers the flow component in the spring season (MAMJ). Since station temperature data were available only since 1983 and also had poor spatial coverage (few stations in low elevations), we compared these data with the data available from other sources. The comparison of station temperature data with the Indian Meteorological Department (IMD) 1-degree gridded daily temperature data (Srivastava et al., 2008) for the Satluj Basin

Table 1. Information about different meteorological stations in Satluj River basin in India.

Station Name	Latitude (N)	Longitude (E)	Elevation (m a.m.s.l.)
Rainfall			
Berthin	31°25'11"	76°38'55"	668
Bhartgarh	31°6'0"	76°36'0"	284
Daslehra	31°24'56"	76°32'56"	562
Ganguwal	31°24'	76°6'	1220
Ghanauli	31°1'33"	76°35'22"	293
Kahu	31°12'43"	76°46'52"	526
Lohand	31°10'31"	76°34'14"	288
Naina Devi	31°17'56"	76°32'8"	985
Nangal	31°23'50"	76°22'21"	369
Rampur	31°26'24"	77°37'40"	987
Suni	31°14'43"	77°6'53"	701
Kasol	31°30'0"	77°19'0"	2614
Kotla	31°30'0"	77°15'0"	2824
Swarghat	31°20'	76°45'?	1220
Snow measurement			
Bahli	31°22'17"	77°38'48"	2285
Chitkul	31°20'59"	78°25'0"	3327
Giabong	31°46'24"	78°26'44"	2926
Jangi	31°36'15"	78°25'0"	2721
Kalpa	31°31'60"	78°15'0"	2662
Kaza	32°13'25"	78°4'11"	3618
Kilba	31°31'0"	78°9'0"	1988
Nichar	31°33'7"	77°58'34"	2225
Phancha	31°35'45"	77°43'54"	2348
Sangla	31°25'14"	78°15'44"	2780
Sarhan	31°30'34"	77°47'34"	2144
Tabo	32°7'51"	78°23'11"	4201
Maximum temperature			
Bhakra ¹	31°24'56"	76°26'5"	554
Kasol	31°30'0"	77°19'0"	2614
Nangal ¹	31°23'50"	76°22'21"	369
Rampur	31°26'24"	77°37'40"	987
Suni	31°14'43"	77°6'53"	701

¹ Also minimum temperature stations.

(30.5–33.5° N and 75.5–79.5° E) indicated that they have similar variability on average at monthly time scales for the study region. Therefore, given their longer period of record (1969–2004) and better spatial coverage, IMD gridded temperature data were used here for the Satluj Basin region. In addition, since IMD rainfall data (Rajeevan et al., 2005) also have good spatial and temporal (1-degree, daily) coverage for the basin region and pose a similar variability with the station data (Pal et al., 2013), we tested the prediction performance of both the rainfall datasets separately combined with station snow and IMD temperature data in this study.

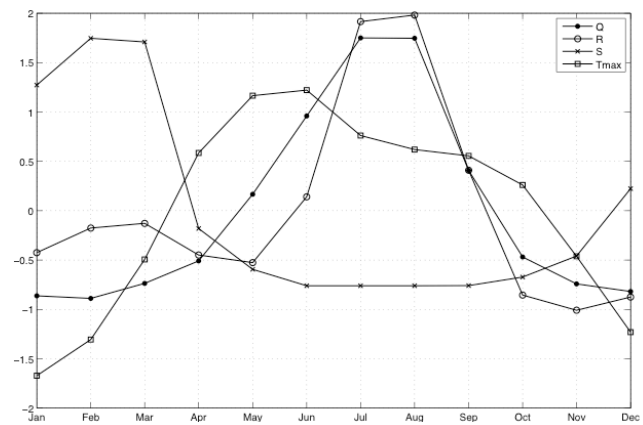


Fig. 3. Comparison of the seasonality of total inflow and basin climatology (Q = normalized inflow, R = normalized rainfall, S = normalized snow, T_{max} = normalized maximum temperature).

Table 2 summarizes the cross-correlations of average Dec/Jan/Feb/Mar (DJFM) climate data and spring seasonal flow from 1978–2004. The table indicates that average measured snow in winter in the high elevation is positively correlated with the lower-elevation rainfall, which was true both for the IMD and station rainfall (0.65 and 0.71, respectively). IMD rainfall is very highly correlated with the station rainfall (0.84). Winter precipitation (both rainfall and snow) is negatively and highly correlated with T_{max} and T_{avg} . Since T_{avg} is the simple average between T_{max} and T_{min} , the correlations between T_{avg} and T_{max} and T_{min} are very high (> 0.8).

Figure 3 shows the average seasonality of the inflow and Satluj River basin climatology (mean monthly over 1978–2004). The figure suggests that the basin snow and rainfall (IMD) patterns are masked by each other in different seasons. The figure also depicts that the flow starts increasing continuously from February relative to the increase in temperature. Therefore one might expect that warmer winters contribute positively to the melt in the months that followed. Temperature has a peak in June when the snowmelt-contributed flow usually reaches its peak. In July and August when the flow is highest, it is driven by a combination of monsoon rain at lower elevations and glacier melt (Singh and Quick, 1993). On the other hand, since T_{max} showed a strong and significant negative correlation with the spring inflow (Table 2), winter T_{max} might just be a reflection of concurrent precipitation amount.

Considering different redundancies in the data, we considered prediction of spring flow at different lead times, using two meteorological data combination settings:

- Setting 1: combining IMD rainfall and temperature data with station snow data.
- Setting 2: using station precipitation (snow and rainfall) data as the predictors and omitting temperature data. We considered this for three reasons: (i) station

Table 2. Cross-correlations for different hydrometeorological variables for 1978–2004.

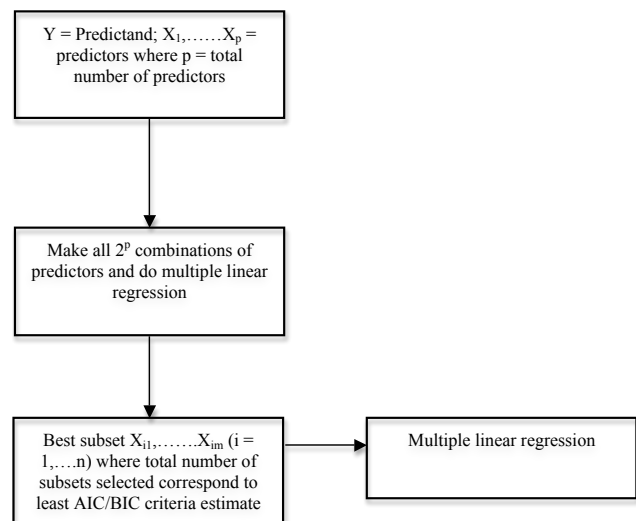
	Flow	Station snow	IMD rainfall	Station rainfall	IMD Tmax	IMD Tmin	IMD Tavg	IMD DTR
Flow	1.00	0.58	0.48	0.55	−0.50	−0.02	−0.33	−0.55
Station snow	0.58	1.00	0.65	0.71	−0.44	−0.18	−0.37	−0.34
IMD rainfall	0.48	0.65	1.00	0.84	−0.67	−0.49	−0.68	−0.32
Station rainfall	0.55	0.71	0.84	1.00	−0.78	−0.38	−0.69	−0.54
IMD Tmax	−0.50	−0.44	−0.67	−0.78	1.00	0.53	0.90	0.66
IMD Tmin	−0.02	−0.18	−0.49	−0.38	0.53	1.00	0.84	−0.30
IMD Tavg	−0.33	−0.37	−0.68	−0.69	0.90	0.84	1.00	0.27
IMD DTR	−0.55	−0.34	−0.32	−0.54	0.66	−0.30	0.27	1.00

Note: the meteorological data is for winter season (DJFM) and flow data is for spring (MAMJ). DTR = Diurnal Temperature Range = $T_{\max} - T_{\min}$.

temperature data has a shorter length (1983 onwards) and poor spatial coverage (2–5 stations); (ii) to avoid the multicollinearity issue, i.e., high correlations between precipitation and temperature, as in Table 2; and (iii) since there are difficulties associated with obtaining meteorological data from the IMD on a real-time basis (particularly the gridded datasets used in this study), we want to demonstrate the predictability of the spring inflows, based entirely on the BBMB network so that they can utilize the data in near-real time for making the best use of the prediction schemes developed here for any decisions they plan to initiate.

3 Predictor(s) selection

Selection of appropriate predictors is required for forecasting streamflows. Poorly designed predictor selection procedures can result in poor forecasts for independent events. Best predictors vary with location and forecast date. This section discusses the potential predictor(s) selection (most influential variables) for forecasting the spring seasonal total inflow into Bhakra Dam, determined using the “bestglm” technique at different lead times and for both settings stated above (McLeod and Xu, 2010). The bestglm tool is freely available to download for the publicly available statistical software R. This method provides a flexible framework to describe how a dependent variable can be explained by a set of predictors. Fitting a single model is not satisfactory in all circumstances (Rajagopalan et al., 2002, 2005). In particular, this assumes that the model used is true or at least optimal in some sense. Hence the resulting inference is conditional on this model. A special case of this general problem is variable selection in the multimodel and multiple regression framework (Miller, 2002). The usual method to select a model is stepwise – backward- or forward-selection approach based on significance tests. However, backward and forward approaches are not generally expected to converge to the same model (Venables and Ripley, 1997; Grafen and Hails, 2002). Several model selection techniques have been developed to avoid these pitfalls. One of these techniques is based on an

**Fig. 4.** Flowchart showing the bestglm method.

information criterion (IC) that compares all possible candidate models and ranks them based on their IC values. This method ensures that the “best” model (according to the IC) is identified, whereas stepwise explorations do not. Most importantly, this method allows one to assess model-selection uncertainty or to perform multimodel inference, rather than using one and only one best model (Buckland et al., 1997; Burnham and Anderson, 2002; Johnson and Omland, 2004).

The step-by-step procedure of the bestglm technique is shown in Fig. 4 as a flowchart. We usually consider the linear model of Y on p inputs, X_1, \dots, X_p ; but in many cases, Y can be more parsimoniously modeled and predicted using just a subset of $m < p$ inputs, X_{i_1}, \dots, X_{i_m} ($i = 1, \dots, n$). When $n > p$, the best-subset problem is to find out all the possible 2^p regressions (subsets) and the best fit (subset) according to some goodness-of-fit criterion, as in Fig. 4. When $p \leq 25$ or thereabouts (in our case maximum $p = 16$), this technique has proved to be very efficient, but problems with large enough p , such as $p > 100$, cannot be solved by this method (McLeod and Xu, 2010). Other high-dimensional

optimization algorithms such as genetic search and simulated annealing are recommended for that case (cross references from McLeod and Xu, 2010).

The bestglm model selection method includes a variety of IC. The information criteria include the usual AIC and BIC as well as two types of extended BIC. All the information criteria that might be considered in the bestglm technique are based on a penalized form of the deviance or minus the log-likelihood. In the multiple linear regression the deviance $D = 2\log L$, where L is the maximized log-likelihood, and $\log L = -(n/2) \log S/n$, where S is the residual sum of squares. Akaike (1974) showed that $AIC = D + 2k$, where k is the number of parameters, provides an estimate of the entropy. The model with the smallest AIC is preferred. While many other criteria (e.g., final prediction error criterion), which are essentially equivalent to the AIC, have also been suggested, in practice, with small n , all those criteria often select the same model. The BIC is derived using Bayesian methods. If a uniform prior is assumed for all possible models, the usual BIC may be written $BIC = D + k \log(n)$. The model with the smallest BIC corresponds to the model with maximum posterior probability. The difference between these criteria is in the penalty. When $n > 7$, the BIC penalty is always larger than for the AIC, and consequently the BIC will never select models with more parameters than the AIC. In practice, BIC often proved to select more parsimonious models than the AIC (McLeod and Xu, 2010). In large p problems ($p > 25$) a modified BIC called BICg is suggested that considers a prior that is uniform of models of fixed size instead of all possible models (as with the original BIC case). On the other hand, BICq is another modified version of BIC that is derived by assuming Bernoulli prior for the parameters. Each parameter has a priori probability of q of being included, where $q = [0, 1]$. When $q = 1/2$, BICq is equivalent to BIC, while $q = 0$ and $q = 1$ correspond to selecting the models with $k = p$ (full model) and $k = 0$ (no parameters), respectively. If q is within 0.01 and 0.51, it gives results equivalent to the AIC. Having a small dataset (only 27 data points), we used AIC and BIC in this paper and also compared them in terms of their robustness for the best models and the prediction skills, as mentioned in Fig. 4.

While bestglm technique has various advantages over many linear regression techniques, as discussed above, it is not clear how this technique handles the multicollinearity issue, which is a common redundancy with the hydrometeorological data and true here too, as Table 2 points out. Multicollinearity might create highly sensitive parameter estimators with inflated variances, and improper model selection. Regression techniques such as PCR or PLSR reduce the redundancy due to multicollinearity. Therefore, we also compared the prediction skills of best-subset multiple regression to that of PCR and PLSR, considering all variables.

We considered the prediction of mean MAMJ (total) inflow on 1 December, 1 January, 1 February and 1 March and of AMJ inflow on 1 April. All the possible predictors

at those lead times and corresponding Pearson's correlation coefficients with average MAMJ inflow are listed in Table 3. Predictors corresponding to the correlation coefficient within ± 0.10 were neglected. As stated earlier, March is a very important month since the maximum snow cover occurs during this time. Therefore, incorporating March data as a predictor is expected to show higher skills as also depicted in correlations in Table 3. In addition, $\sim 90\%$ of the volume of MAMJ inflow on average occurs in AMJ alone. Therefore, we also considered an 1 April prediction for AMJ inflow. As the subsets are unknown, various possible monthly combinations were considered in order to choose the best predictors at different lead times, as shown in Table 3. As Table 3 reveals, all the correlations for the first three lead times are very poor. This is because much of the snow accumulated in Nov/Dec/Jan usually melts away before March, usually starting in late January and early February (Pal et al., 2013).

The distribution of the predictand (MAMJ flow) was close to Gaussian based on Lillifor's test, and scatter plots between predictors and predictand revealed nearly linear structures. Therefore the link function used in this method is linear and Gaussian, and the models turned out to be multiple linear regressions. We also checked the linear relationship using LOCFIT, or local weighted polynomial approach, and generalized cross-validation (GCV) estimates for each fit between the predictors and predictand to determine the optimum window width used for the local regression that corresponds to the minimum GCV (Loader, 1999), as recommended by Rajagopalan et al. (2002, 2005). In all the cases, we got $\alpha = 0.9-1.0$, where α is the fraction of the total data length. Therefore, we used bestglm with much confidence.

4 Prediction results and skills

Variable selection, measurement and comparison of the model performance at different lead times were done using $K = 3$ leave-out cross-validation. This method leaves three randomly selected observations in turn, determines the best subset from the remaining training data points and the best multiple linear model fit. The dropped points form the testing set against which the goodness of fit of the estimated points are measured. This is repeated 100 times. PCR and PLSR were also applied in $K = 3$ leave-out cross-validation mode considering all the p variables at a time. Figure 5a–d show the bar charts of the total frequency of the variables selected in different lead times in 100 iterations corresponding to AIC and BIC. Different variables within the different settings are given the names V_1 to V_p for the convenience in plotting them, where $p =$ total number of variables used at certain lead time. Variables corresponding to the names V_1 to V_p in Fig. 5 are listed in Table 4. For example, in the first panel of Fig. 5a, a total of 7 variables were used for the 1 December prediction (Table 4), namely November flow (N_Q), November rainfall (N_R), November snow (N_S), November

Table 3. Pearson's correlation coefficients between average MAMJ inflow into Bhakra and spatially averaged Satluj River basin climatology (showing setting 1 datasets) and inflow for 1978–2004.

Prediction	1 December		1 January		1 February		
	N	D	ND	J	DJ	NDJ	
Flow (Q)	-0.05	0.17	0.04	0.07	0.12	0.05	
Snow (S)	0.24	0.13	0.22	-0.17	-0.05	0.03	
Rainfall (R)	0.30	0.43	0.48	-0.10	0.12	0.19	
T_{\max}	-0.06	-0.26	-0.19	0.18	-0.06	-0.07	
T_{\min}	-0.04	0.16	0.07	0.08	0.18	0.11	
T_{avg}	-0.06	-0.13	-0.11	0.16	0.06	0.01	
DTR	-0.02	-0.29	-0.21	0.10	-0.17	-0.14	
	1 March			1 April			
	F	JF	DJF	M	FM	JFM	
Flow (Q)	0.38	0.24	0.22	0.72	0.66	0.57	
Snow (S)	0.48	0.28	0.36	0.57	0.70	0.54	
Rainfall (R)	0.64	0.29	0.40	0.44	0.58	0.42	
T_{\max}	-0.42	-0.24	-0.31	-0.48	-0.53	-0.46	
T_{\min}	0.12	0.11	0.19	-0.27	-0.13	-0.07	
T_{avg}	-0.22	-0.08	-0.09	-0.41	-0.40	-0.33	
DTR	-0.64	-0.34	-0.41	-0.63	-0.71	-0.54	

T_{\max} ($N-T_{\max}$), November T_{\min} ($N-T_{\min}$), November T_{avg} ($N-T_{\text{avg}}$) and November DTR (N_DTR). Likewise, the variables corresponding to other lead times in other settings were also named V_1 to V_p . As mentioned above, BIC chooses the best models more parsimoniously than the AIC, which is also evident in the figure. However, the orders in which the variables are selected are the same for both of the criteria. Table 5 lists the final linear models those consider the best subsets chosen corresponding to the AIC. The variables in the linear models in Table 5 were consistently chosen ($> 80\%$ of the times) or chosen for the maximum number of times (e.g., 1 December). As noticed in Table 5, the variables selected for 1 December and 1 February predictions were the same for both the settings. For setting 1, temperature data were important for 1 January and 1 March predictions. Previous flow data (January, February and March) become important for the 1 April prediction of AMJ flow. Snow in February and March is important for MAMJ flow, which we have also noticed before and which is also reflected here in the linear models for 1 March and 1 April predictions.

To verify the cross-validated forecast model performance for three techniques considered (bestglm, PCR, PLSR), we divided both observed and predicted inflow time series into tercile categories (below normal/normal/above normal) and determined the joint distribution of them – characterized as a 3×3 contingency table (Wilks, 2006). We used 300 discrete random predictions ($K=3$, 100 iterations) and corresponding observations. Due to the fact that we have a small sample size and high-dimensional verification situation ($(3 \times 3) - 1 = 8$), we opted for traditional scalar approaches

such as (a) *accuracy*, or proportion correct (PC), and (b) *forecast skill score*, i.e., Heidke skill score (HSS) and Peirce skill score (PSS). Accuracy refers to the average correspondence between individual forecasts and the events they predict. Therefore, PC measure will summarize the overall quality of a set of forecasts in a single number determining the proportion of correct forecasts (Wilks, 2006). On the other hand, forecast skill scores will be interpreted as a percent improvement over the reference forecasts. While a large number of such scores are in use, one of the most frequently used skill scores is the HSS. A perfect forecast receives $HSS = 1$, forecasts equivalent to the reference forecasts receive zero scores, and forecasts worse than the reference forecasts receive negative scores. The reference accuracy measure in the HSS is the proportion correct that would be achieved by random forecasts that are statistically independent of the observations (Wilks, 2006). Another equivalent and popular skill score is the PSS, which can also be understood as the difference between the hit rate and the false alarm rate. Perfect forecast receives a score of 1, random forecast receives a score of zero, and forecasts inferior to the random forecasts receive negative scores (Wilks, 2006).

Table 6 lists different cross-validated prediction performance measures on the two data settings considered and for all of the three techniques (bestglm, PCR, PLSR). The table shows that PC usually improves from 1 December to 1 April gradually for all the cases. The best prediction skills corresponding to each lead time is marked with an asterisk (*). The table shows that for 1 January, 1 March and 1 April predictions corresponding to setting 1, bestglm performed better

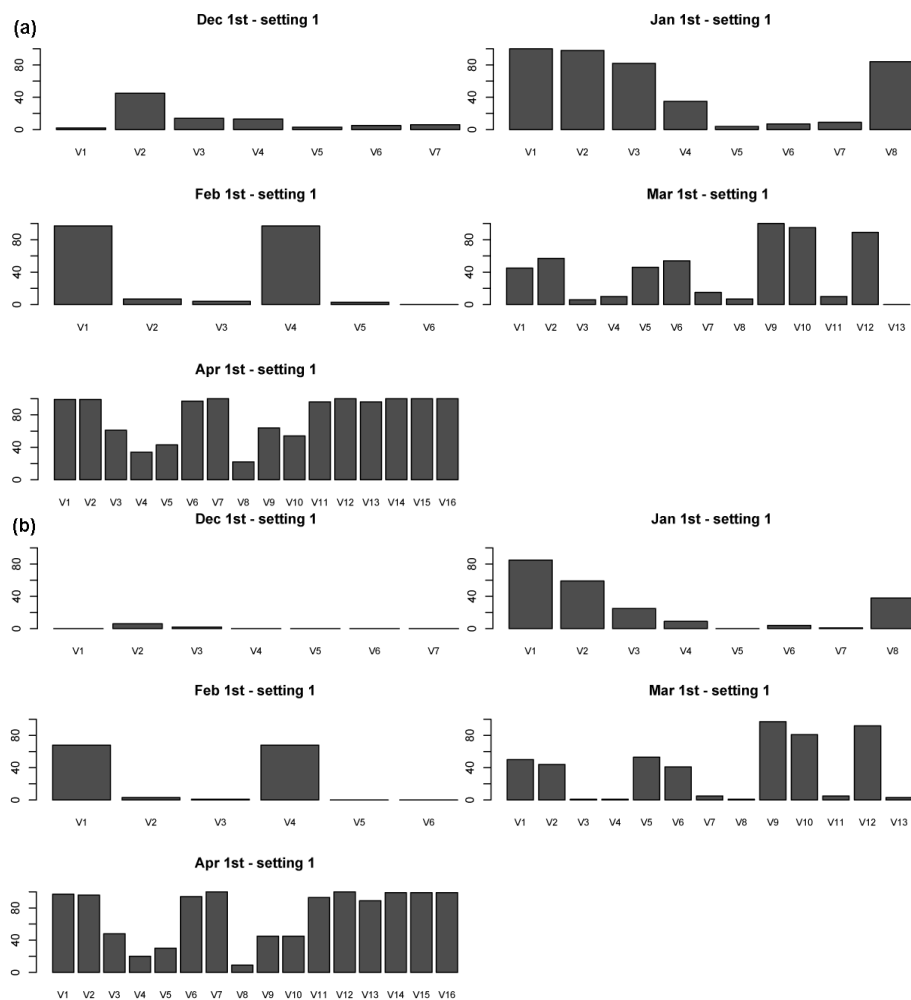


Fig. 5. (a) Number of times (in 100 iterations) the different variables were selected in setting 1 when AIC was used. V_1 to V_p names are listed in Table 4; (b) same as (a) but when BIC was used; (c) number of times (in 100 iterations) different variables were selected in setting 2. V_1 to V_p names are listed in Table 4; (d) same as (c) but when BIC was used.

than the other techniques, with AIC as well as BIC yielding more or less similar prediction skills. While 1 February and 1 December prediction skills were improved when PCR was used for setting 1, 1 December skills were still negative. For setting 2, bestglm showed the best performance for 1 January and 1 March; on the other hand, PCR showed best performance for 1 February and 1 April.

While the PC measurement corresponding to bestglm was better for setting 2 as compared to setting 1 for the 1 December, 1 January and 1 March forecasts, setting 1 prediction was improved for 1 April with 64 % of the forecasts falling into the correct category, while for setting 2 it was 48 %. When PCR was used in setting 2, PC jumped to 64 % but other skills did not improve from bestglm. Song and Kroll (2011) pointed out that for small sample size, PCR or PLSR decrease the variance of the watershed hydrologic parameters only when the true model is known (i.e., known subset) and do not improve model predictions compared to linear regres-

sion models. This might hold true in our case too since we just have 27 data points.

Furthermore, correlation coefficients in Table 6 were better for setting 1 for 1 January, 1 March and 1 April forecasts corresponding to bestglm (AIC). HSS and PSS estimates yielded the same estimates all the times, indicating that the forecasts exhibit little bias (Wilks, 2006). Like PC, HSS/PSS scores corresponding to bestglm were better for setting 2 for the 1 March forecast, although the 1 April forecast was much better when setting 1 was used. However, overall results imply that, even if the temperature information is not available, only the station precipitation information could be used to produce real-time forecasts of spring inflow with positive categorical skills. The availability of reliable station temperature information and inclusion in setting 2 might improve the categorical forecast skills for some lead times.

Figure 6 shows the range of cross-validated prediction square error and average absolute residual errors for each

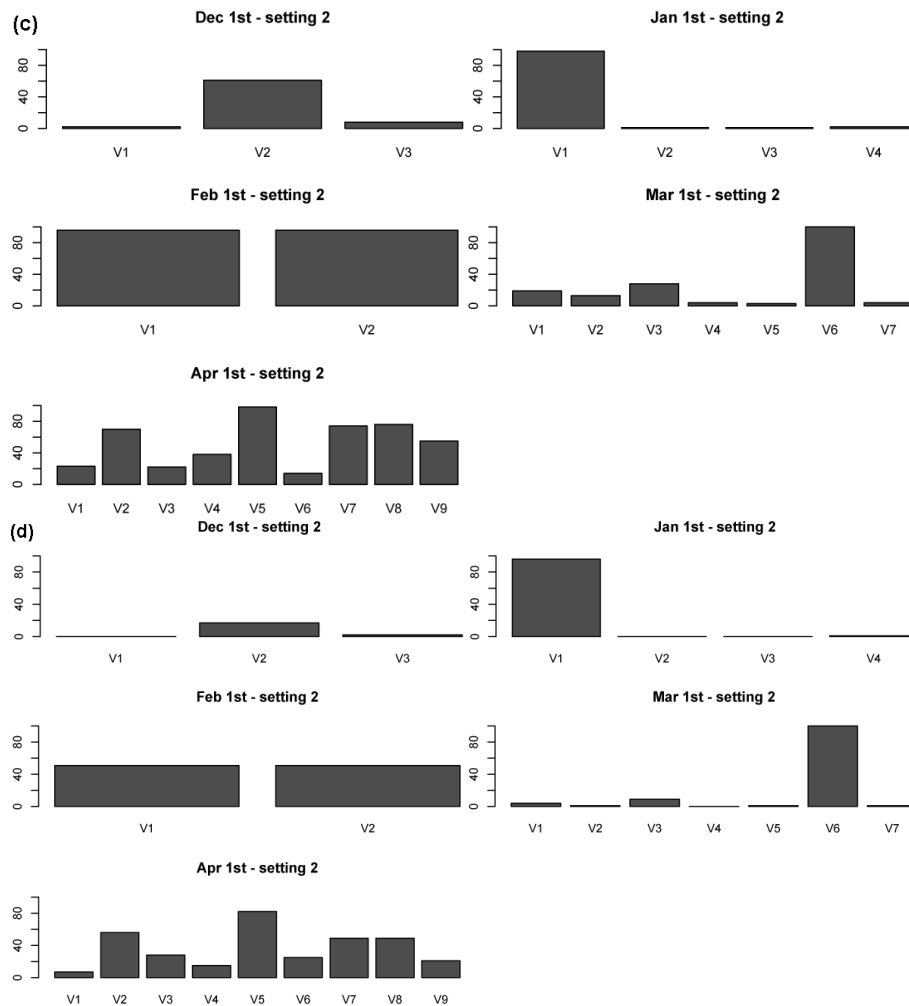


Fig. 5. Continued.

setting corresponding to bestglm technique. These were done after finding out the median forecast value of spring inflow for each year from the ensemble of (random) forecasts generated in 100 iterations. The figure depicts that the error is substantially reduced for 1 March predictions from the previous lead times for both settings, but the spread increases for 1 April again, which is true for both AIC as well as BIC best models. Figure 7 shows the (median) predictions made on 1 March and 1 April from the ensemble predictions after 100 iterations, with corresponding 5 and 95 % prediction intervals. Prediction intervals take into account uncertainty about mean prediction and uncertainty associated with the error term. 1 March nicely predicts all the years for setting 1 except a few in the early 1980s, 1996, 2003 and 2004, which are better predicted on 1 April. For setting 2, 1 April prediction is worse than 1 March, which is also reflected in the errors in Fig. 6 and forecast skill in Table 6. With high error (average absolute $\sim 18\%$), the 1 December prediction was the worst possible forecast for spring inflow while the

correlation coefficient was also negative (Table 5). All these results suggest that as far as preplanning is concerned, the 1 March prediction would give us the most accurate estimate of spring inflow before the season commences; however, the incorporation of the March data (i.e., forecast issued on 1 April) would be useful for AMJ inflow as well. Furthermore, while setting 1 dataset is more recommended for spring seasonal inflow prediction, setting 2 could be used interchangeably when IMD dataset is not available, albeit with less precision.

5 Summary and discussion

The research presented the potential predictability and the merits of different prediction models of melt-dominated inflow into the Bhakra Dam using the flow data and local winter climatology at the basin scale from 1978–2004. Several candidate methods for multiple linear regressions were explored in a situation where the short data length is a

Table 4. Names of different variables considered in best-subset selection in Fig. 5.

Setting 1				
1 December	1 January	1 February	1 March	1 April
V1 = N_Q	V1 = ND_R	V1 = NDJ_R	V1 = DJF_R	V1 = JFM_Q
V2 = N_R	V2 = ND_S	V2 = NDJ_T _{min}	V2 = DJF_S	V2 = JFM_R
V3 = N_S	V3 = ND_T _{max}	V3 = NDJ_DTR	V3 = DJF_T _{max}	V3 = JFM_S
V4 = N_T _{max}	V4 = ND_DTR	V4 = J_R	V4 = DJF_DTR	V4 = JFM_T _{max}
V5 = N_T _{min}	V5 = D_Q	V5 = J_T _{max}	V5 = JF_R	V5 = JFM_DTR
V6 = N_T _{avg}	V6 = D_R	V6 = J_T _{avg}	V6 = JF_S	V6 = FM_Q
V7 = N_DTR	V7 = D_T _{max}		V7 = JF_DTR	V7 = FM_R
	V8 = D_DTR		V8 = F_Q	V8 = FM_S
			V9 = F_R	V9 = FM_T _{max}
			V10 = F_S	V10 = FM_DTR
			V11 = F_T _{max}	V11 = M_Q
			V12 = F_T _{min}	V12 = M_R
			V13 = F_DTR	V13 = M_S
				V14 = M_T _{max}
				V15 = M_T _{avg}
				V16 = M_DTR

Setting 2				
1 December	1 January	1 February	1 March	1 April
V1 = N_Q	V1 = ND_R	V1 = NDJ_R	V1 = DJF_R	V1 = JFM_Q
V2 = N_R	V2 = ND_S	V2 = J_R	V2 = DJF_S	V2 = JFM_R
V3 = N_S	V3 = D_Q		V3 = JF_R	V3 = JFM_S
	V4 = D_R		V4 = JF_S	V4 = FM_Q
			V5 = F_Q	V5 = FM_R
			V6 = F_R	V6 = FM_S
			V7 = F_S	V7 = M_Q
				V8 = M_R
				V9 = M_S

Table 5. Variables selected using bestglm with AIC and the final models at different lead times for setting 1 (combining IMD rainfall and temperature data with BBMB station snow data) and setting 2 (using station rainfall and snow data only as the predictors and omitting temperature).

Setting 1	
1 December	$Q_{MAMJ} = 16\,974 + 44(N_R)$
1 January	$Q_{MAMJ} = -10\,622 + 213(ND_R) - 139(ND_S) + 2095(ND_T_{max}) - 1224(D_DTR)$
1 February	$Q_{MAMJ} = 16\,324 + 135(ND_R) + 80(J_R)$
1 March	$Q_{MAMJ} = 5546 + 47(F_R) + 18(F_S) + 1227(F_T_{min})$
1 April	$Q_{AMJ} = -2273 + 362(J_Q) - 136(J_R) - 82(F_Q) + 108(F_R) - 78(M_Q) + 35(M_R) + 25(M_S)$

Setting 2	
1 December	$Q_{MAMJ} = 17\,037 + 37(N_R)$
1 January	$Q_{MAMJ} = 16\,327 + 53(ND_R)$
1 February	$Q_{MAMJ} = 17\,189 + 79(ND_R) + 37(J_R)$
1 March	$Q_{MAMJ} = 13\,628 + 46(F_R)$
1 April	$Q_{AMJ} = 13\,995 - 71(J_R) + 64(F_R) + M_Q + 25(M_R)$

Note: flow (Q) is in cubic feet per sec (cfs), rainfall (R) and snow (S) are in millimeters (mm) and temperature (T) is in degrees centigrade ($^{\circ}C$)

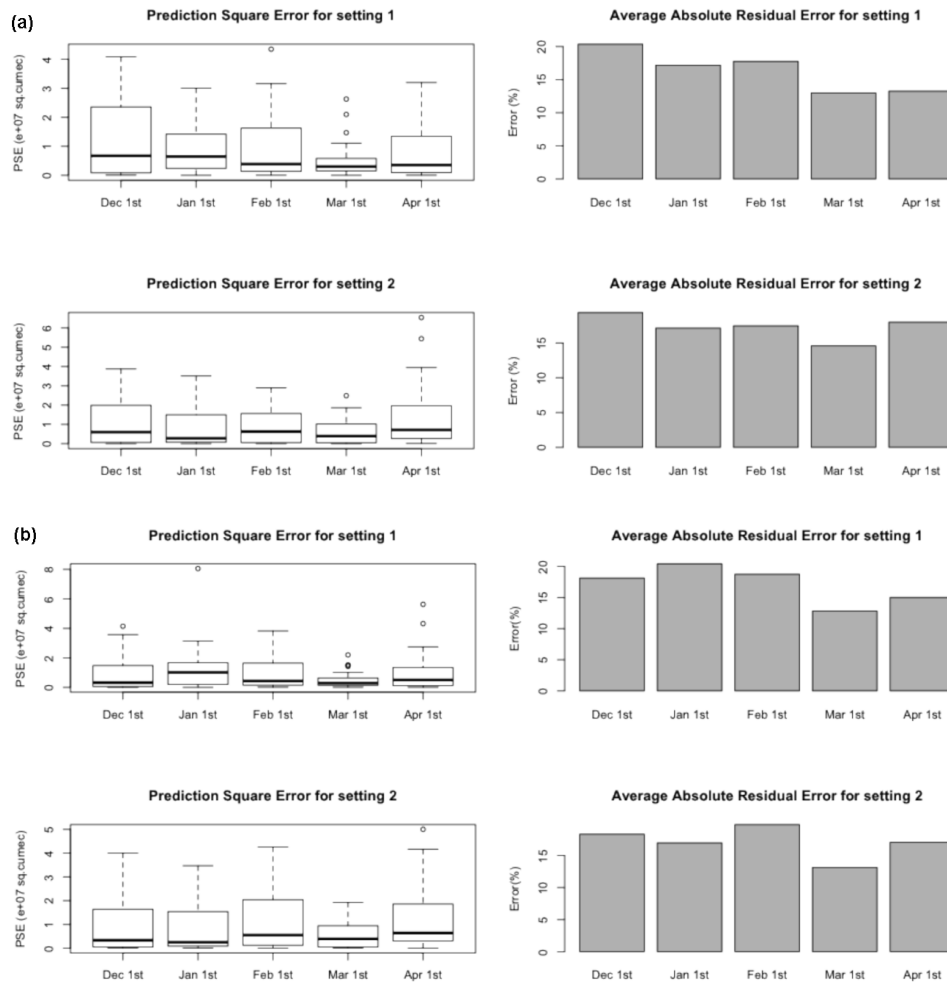


Fig. 6. (a) Cross-validated prediction square errors and average absolute errors corresponding to bestglm (AIC); (b) Cross-validated prediction square errors and average absolute errors corresponding to bestglm (BIC).

challenge. We used basin level winter climatology to find the predictability of spring seasonal streamflow. Given the direct physical link between winter precipitation and snowmelt-driven river flow, the spring inflow was found to be positively correlated with the high-elevation winter snow and corresponding rainfall in the lower elevations, and it is negatively correlated with the basin-averaged winter temperatures. The positive significant correlations of spring inflow with winter precipitation also suggest that the winter atmospheric circulation patterns driving these climate variables are also related to spring inflow (Pal et al., 2013), and are consistent with Schar et al. (2004), Tippet et al. (2003), and Barlow and Tippet (2008), who showed spring- and summertime river flow fluctuation due to winter climate patterns at the regional scale in central and southern Asia.

One apparent major assumption in the multiple linear regression prediction models is the stationary relationship between streamflow and snowmelt in time. Although precise consideration was given to where this was not true (e.g., step

change from the late 1970s), since step changes caused by the construction of a dam were obvious from the data as well as the dam manager's practical knowledge (Bhakra Beas Management Board), a closer look was also taken at whether there were more subtle changes such as nonstationarity of the hydro-climatological data due to climate variability/change that may also significantly change the relationship between lagged snowmelt and streamflow. In addition, this consideration could also potentially shed light on why the models performed better during certain periods as compared to others. To address this major assumption, first, a trend analysis was carried out for all the hydroclimatic variables considered in this paper, as predictors and predictand; it was found that the spatially averaged data from 1978–2004 (the study period) did not have any significant trend, and therefore they were not shown in the paper. This was partly consistent with the previous findings where some researchers noticed that the Western Himalaya responded to global warming differently as compared to the other regions (Fowler and Archer, 2006).

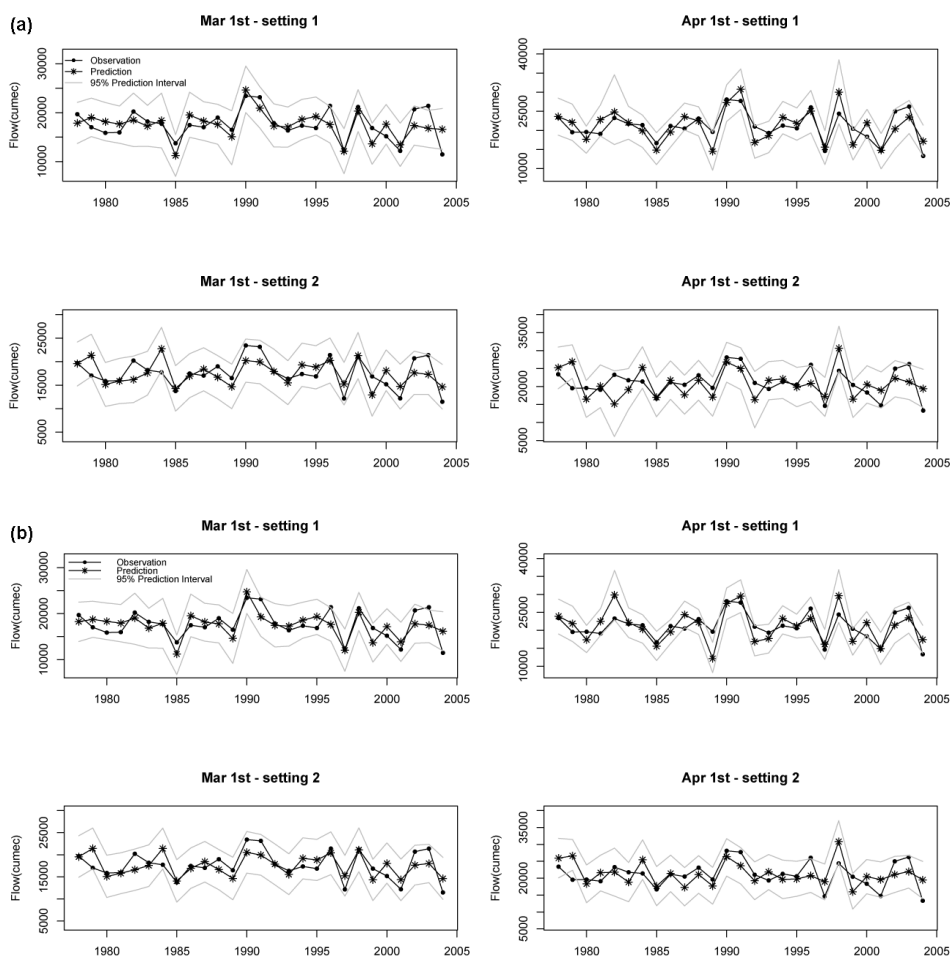


Fig. 7. (a) Cross-validated prediction for 1 March and 1 April (MAMJ and AMJ inflow) corresponding to different datasets used, i.e., settings 1 and 2 (AIC); (b) cross-validated prediction for 1 March and 1 April (MAMJ and AMJ inflow) corresponding to different datasets used, i.e., settings 1 and 2 (BIC).

In addition, to our surprise, we also did not find an increasing trend in the Satluj River basin average DTR as was reported by Yadav et al. (2004). However, the best hypothesis for this mismatch might be due to the fact that we aggregated all the station/gridded data to have a spatial basin average, which might have eventually smoothed out the spatially averaged data to the extent of not showing the trends. Individual station data might have had inconsistent precipitation trends, as published recently by Dimri and Dash (2012) while reporting the Western Himalayan climatic trend analysis, but this was not within the scope of our study. Second, the “linearity” of the relationship between the predictors and predictand was checked according to the LOCFIT method (Loader, 1999), a technique suggested by previous researchers (Rajagopalan et al., 2002, 2005); in addition, the predictand distribution was found to be Gaussian as per Lilifor’s test. The LOCFIT method confirmed that the predictors and predictand relationships were consistently found to be linear. Therefore, the utility of the bestglm technique in the multiple linear regression

scenario based on selective information criteria was skillfully used to forecast spring seasonal streamflow at the upstream of Bhakra Dam over the Satluj River at different lead times. We compared the forecasts to PCR and PLSR, which were more widely used for the central and southern Asian rivers (Schar et al., 2004; Tippet et al., 2004; Barlow and Tippet, 2008) and also in the US (Regonda et al., 2006). We find that overall, the proposed method is equally skillful to (and actually more skillful than) existing operational models while tending to better predict seasonal streamflow 1–4 months in advance.

The results as a whole suggest that winter climatology and inflow data allow a skillful forecast of the volume of MAMJ/AMJ flow in late winter or when the spring season commences. Inclusion of gridded rainfall data with station snow information enhanced the forecast skill scores possibly for the reason that rain gauge stations have poor coverage within the basin (in particular the mountains) and/or systematic rain gauge biases (wind losses during snowfall),

Table 6. Cross-validated prediction performance corresponding to different techniques.

Bestglm setting 1 (AIC)				
	CC	HSS	PSS	PC
1 December	-0.27	-0.18	-0.18	0.21
*1 January	0.37	-0.03	-0.03	0.29
1 February	0.18	-0.02	-0.02	0.32
*1 March	0.63	0.24	0.24	0.49
*1 April	0.72	0.47	0.47	0.64
Bestglm setting 1 (BIC)				
1 December	-0.26	-0.16	-0.16	0.22
1 January	0.06	-0.10	-0.10	0.27
1 February	0.02	-0.04	-0.04	0.31
1 March	0.61	0.23	0.23	0.49
1 April	0.70	0.39	0.39	0.59
PLSR setting 1				
1 December	-0.22	-0.14	-0.14	0.24
1 January	-0.12	-0.12	-0.12	0.32
1 February	0.30	-0.01	-0.01	0.33
1 March	0.14	0.14	0.14	0.42
1 April	0.51	0.27	0.27	0.52
PCR setting 1				
*1 December	-0.27	-0.1	-0.1	0.27
1 January	-0.19	-0.08	-0.08	0.28
*1 February	0.32	0.13	0.13	0.42
1 March	0.24	0.16	0.16	0.44
1 April	0.60	0.39	0.39	0.59
Bestglm setting 2 (AIC)				
1 December	-0.14	-0.14	-0.14	0.23
*1 January	0.21	0.12	0.12	0.41
1 February	0.19	-0.07	-0.07	0.28
1 March	0.53	0.37	0.37	0.58
1 April	0.52	0.21	0.21	0.48
Bestglm setting 2 (BIC)				
1 December	-0.28	-0.16	-0.16	0.23
1 January	0.22	-0.02	-0.02	0.32
1 February	-0.03	-0.13	-0.13	0.25
*1 March	0.61	0.47	0.47	0.64
1 April	0.48	0.16	0.16	0.44
PLSR setting 2				
*1 December	-0.17	-0.04	-0.04	0.30
1 January	-0.14	-0.05	-0.05	0.30
1 February	0.25	-0.02	-0.02	0.31
1 March	0.22	0.14	0.14	0.42
1 April	0.48	0.36	0.36	0.58
PCR setting 2				
1 December	-0.19	-0.09	-0.09	0.27
1 January	-0.16	-0.06	-0.06	0.29
*1 February	0.28	0.05	0.05	0.37
1 March	0.20	0.15	0.15	0.43
*1 April	0.58	0.46	0.46	0.64

Note: CC = Pearson's correlation coefficient; HSS = Heidke skill score; PSS = Peirce skill score; PC = proportion correct; * = best performance in a setting.

as speculated by Schar et al. (2004) while demonstrating the predictability of the central Asian river flow. Overall the above results indicate that the IMD rainfall and temperature data along with station snow information is able to credibly represent the variations of the spring seasonal Satluj River flow, while it probably misses smaller-scale features associated with the complex structure of the topography. An analysis was also undertaken to look at the ability of precipitation data from local meteorological stations to represent observed interannual variability of streamflow of the Satluj River, which was satisfactorily skillful. This is consistent with Schar et al. (2004) who also demonstrated that precipitation data in winter alone is sufficient to skillfully predict the next seasonal river flow in central Asia. Despite the low resolution of IMD precipitation data and poor coverage of station meteorological data, the results appeared highly promising. In most years the models provided excellent predictions on 1 March and 1 April, and only a few years (includes 2004) had a departure from the prediction interval (based on a $p = 95\%$ prediction interval, one would expect at least one miss in a 15 yr series, according to Schar et al. (2004)). The 1 March and 1 April models also enabled forecasts that represent an improvement in comparison with the climatological forecast.

Water managers depend on seasonal forecasts of reservoir inflow volume to support operations and planning in advance. Careful planning is necessary to meet water demands in the dry pre-monsoon season (MAMJ) under the stress of increased climate variability. This is particularly true for the Bhakra Reservoir because total volume of MAMJ inflow into Bhakra depends largely on the highly variable winter climatology of the Western Himalayas. The forecasting model developed here could significantly help water resource decision-making associated with Bhakra Reservoir operations.

For reservoir operation, a key question in April, May and June is how much space to leave for flood control and how full to keep the reservoir. Retaining water in the reservoir during this fill cycle allows more reliable irrigation and hydropower releases and a better buffering capacity for the failure of the monsoon. However, it increases exposure to flood risk. This trade-off is traditionally managed using a rule curve for storage allocation that is based on the historical climatology of the monsoon and the snowmelt period. The innovation presented here is now being used to foster the development of a stochastic optimization model for dynamic rule curve determination considering the probability distribution of fill during MAMJ/AMJ, and the flows and consumptive water demand in the subsequent monsoon season. Flood volume predictions are also needed to complete such an analysis, and to date very few efforts have been made at long lead flood volume forecasting. Nonlinear or nonparametric methods are often used to address such problems effectively (Kwon et al., 2012). However, given the short record available here with potential data quality issues, these methods

were contra-indicated based on our initial exploratory analyses. The list of abbreviations are included in the supplement.

Supplementary material related to this article is available online at: <http://www.hydrol-earth-syst-sci.net/17/2131/2013/hess-17-2131-2013-supplement.pdf>.

Acknowledgements. The use of gridded daily rainfall and temperature data from the Indian Meteorological Department is gratefully acknowledged. The Bhakra data were provided by the BBMB under an MOU with the Columbia Water Center. This work was supported by the NOAA through a cooperative agreement with the IRI, and by the Columbia Earth Institute.

Edited by: J. Freer

References

- Akaike, H.: A new look at the statistical model identification, *IEEE T. Automat. Contr.*, 19, 716–723, 1974.
- Araghinejad, S., Burn, D. H., and Karamouz, M.: Long-lead probabilistic forecasting of stream flow using ocean atmospheric and hydrological predictors, *Water Resour. Res.*, 42, W03431, doi:10.1029/2004WR003853, 2006.
- Archer, D.: Contrasting hydrological regimes in the upper Indus basin, *J. Hydrol.*, 274, 198–210, 2003.
- Barlow, M. A. and Tippett, M. K.: Variability and Predictability of Central Asia River Flows: Antecedent Winter Precipitation and Large-Scale Teleconnections, *J. Hydrometeorol.*, 9, 1334–1349, 2008.
- Bracken, C., Rajagopalan, B., and Prairie, J.: A multisite seasonal ensemble streamflow forecasting technique, *Water Resour. Res.*, 46, W03532, doi:10.1029/2009WR007965, 2010.
- Buckland, S. T., Burnham, K. P., and Augustin, N. H.: Model selection: An integral part of inference, *Biometrics*, 53, 603–618, 1997.
- Burnham, K. P. and Anderson, D. R.: *Model Selection and Multi-model Inference: A Practical Information Theoretic Approach*, New York: Springer-Verlag, 2002.
- Chiew, F. H. S., Zhou, S. L., and McMahon, T. A.: Use of seasonal streamflow forecasts in water resources management, *J. Hydrol.*, 270, 135–144, 2003.
- Dimri, A. P. and Dash, S. K.: Wintertime climatic trends in the western Himalayas, *Clim. Change*, 111, 775–800, 2012.
- Fowler, H. J. and Archer, D. R.: Conflicting Signals of Climatic Change in the Upper Indus Basin, *J. Climate*, 19, 4276–4293, 2006.
- Gamiz-Fortis, S., Pozo-Vazquez, D., Trigo, R. M., and Castro-Diez, Y.: Quantifying the predictability of winter river flow in Iberia. Part II: seasonal predictability, *J. Climate*, 21, 2503–2518, 2008.
- Garen, D. C.: Improved techniques in regression-based streamflow volume forecasting, *J. Water Resour. Plann. Manage.*, 118, 654–670, 1992.
- Grafen, A. and Hails, R.: *Modern Statistics for the Life Sciences*. Oxford: Oxford University Press, 351 pp., 2002.
- Hagedorn, R., Doblas-Reyes, F. J., and Palmer, T. N.: The rationale behind the success of multi-model ensembles in seasonal forecasting. Part I: Basic concept, *Tellus A*, 57, 219–233, 2005.
- Jain, S. K., Kumar, N., and Ahmed, T.: SLURP model and GIS for estimation of runoff in a part of Satluj catchment, India, *Hydrol. Sci.*, 43, 875–884, 1998.
- Jain, S. K., Goswami, A., and Saraf, A. K.: Assessment of snowmelt runoff using remote sensing and effect of climate change on runoff, *Water Resour. Manage.*, 24, 1763–1777, 2010.
- Johnson, J. B. and Omland, K. S.: Model selection in ecology and evolution, *Trends Ecol. Evol.*, 19, 101–108, 2004.
- Kattelmann, R.: Role of snowmelt in generating streamflow during spring in East Nepal, *Snow and Glacier Hydrology (Proceedings of the Kathmandu Symposium, November 2002)*, IAHS Publication No. 218, 103–111, 1993.
- Krishnamurti, T. N., Kishtawal, C. M., LaRow, T. E., Bachiochi, D. R., Zhang, Z., Williford, C. E., Gadgil, S., and Surendran, S.: Improved weather and seasonal climate forecasts from multi-model superensemble, *Science*, 285, 1548–1550, 1999.
- Krishnamurti, T. N., Kishtawal, C. M., Zhang, Z., LaRow, T. E., Bachiochi, D. R., Williford, C. E., Gadgil, S., and Surendran, S.: Multi-model ensemble forecasts for weather and seasonal climate, *J. Climate*, 13, 4196–4216, 2000.
- Kwon, H. H., Souza Filho, F. A., Sun, L., Block, P., Lall, U., and Dirceu, S.: Assessment of uncertainty of hydrologic model and climate forecast model in Northern Brazil, *Hydrol. Process.*, 26, 3875–3885, 2012.
- Li, H., Luo, L., Wood, E. F., and Schaake, J.: The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting, *J. Geophys. Res.*, 114, D04114, doi:10.1029/2008JD010969, 2009.
- Loader, C.: *Statistics and Computing: Local Regression and Likelihood*, Springer, New York, 1999.
- Makhdoom, M. T. A. and Solomon, S. I.: Attempting flow forecasts of the Indus river, Pakistan using remotely sensed snow cover data, *Nord. Hydrol.*, 17, 171–184, 1986.
- Materia, S., Dirmeyer, P. A., Guo, Z., Alessandri, A., and Navarra, A.: The sensitivity of simulated river discharge to land surface representation and meteorological forcings, *J. Hydrometeorol.*, 11, 334–351, 2010.
- McLeod, A. I. and Xu, C.: *bestglm: Best Subset GLM*, available at: <http://cran.r-project.org/web/packages/bestglm/vignettes/bestglm.pdf>, 2010.
- Miller, A. J.: *Subset Selection in Regression*. Chapman & Hall/CRC, Boca Raton, 2nd Edn, 2002.
- Owosina, A.: *Methods for assessing the space and time variability of groundwater data*, M.S. Thesis, Utah State University, Logan, Utah, 1992.
- Pal, I., Lall, U., Robertson, A., Cane, M., and Bansal, R.: Diagnostics of Western Himalayan Satluj River flow: Warm season (MAM/JJAS) inflow into Bhakra Dam in India, *J. Hydrol.*, 478, 132–147, 2013.
- Rajagopalan, B. and Lall, U.: A Nearest Neighbor Bootstrap Resampling Scheme for Resampling Daily Precipitation and other Weather Variables, *Water Resour. Res.*, 35, 3089–3101, 1999.
- Rajagopalan, B., Lall, U., and Zebiak, S.: Optimal categorical climate forecasts through multiple GCM ensemble combination and regularization, *Mon. Weather Rev.*, 130, 1792–1811, 2002.

- Rajagopalan, B., Grantz, K., Regonda, S., Clark, M., and Zagona, E.: Ensemble streamflow forecasting: Methods and Applications, in: *Advances in Water Science Methodologies*, edited by: Aswathanarayana, U., Taylor and Francis, Netherlands, 2005.
- Rajeevan, M., Bhatte, J., Kale, J. D., and Lal, B.: Development of a high resolution daily gridded rainfall data for the Indian region, India Meteorological Department, Met. Monograph Climatology No. 22/2005, 26 pp., 2005.
- Ramamoorthy, A. S. and Haefner, H.: Runoff modeling and forecasting of river basins, and Himalayan snowcover information system (HIMSIS). Hydrology for the Water Management of Large River Basins (Proceedings of the Vienna Symposium), IAHS Publ. no. 201, 1991.
- Regonda, S. K., Rajagopalan, B., Clark, M., and Zagona, E.: A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River Basin, *Water Resour. Res.*, 42, W09404, doi:10.1029/2005WR004653, 2006.
- Risley, J. C., Gannett, M. W., Lea, J. K., and Roehl Jr., E. A.: An analysis of statistical methods for seasonal flow forecasting in the Upper Klamath River Basin of Oregon and California: US Geological Survey Scientific Investigations Report 2005–5177, 44 pp., 2001.
- Schar, C., Vasilina, L., Pertziger, F., and Dirren, S.: Seasonal runoff forecasting using precipitation from meteorological data assimilation systems, *J. Hydrometeorol.*, 5, 959–973, 2004.
- Singh, P. and Bengtsson, L.: Hydrological sensitivity of a large Himalayan basin to climate change, *Hydrol. Process.*, 18, 2363–2385, 2004.
- Singh, P. and Jain, S. K.: Snow and glacier melt in the Satluj river at Bhakra dam in the western Himalayan region, *Hydrol. Sci.*, 47, 93–106, 2002.
- Singh, P. and Jain, S. K.: Modelling of streamflow and its components for a large Himalayan basin with predominant snowmelt yields, *Hydrol. Sci. J.*, 48, 257–276, 2003.
- Singh, P. and Kumar, N.: Impact assessment of climate change on the hydrological response of a snow and glacier melt runoff dominated Himalayan River, *J. Hydrol.*, 193, 316–350, 1997.
- Singh, P. and Quick, M. C.: Streamflow simulation of Satluj river in the Western Himalayas. Snow and Glacier Hydrology (Proceedings of the Kathmandu Symposium, November 2002), IAHS Publication No. 218, 261–271, 1993.
- Singla, S., Céron, J.-P., Martin, E., Regimbeau, F., Déqué, M., Habets, F., and Vidal, J.-P.: Predictability of soil moisture and river flows over France for the spring season, *Hydrol. Earth Syst. Sci.*, 16, 201–216, doi:10.5194/hess-16-201-2012, 2012.
- Song, P. and Kroll, C.: The impacts of multicollinearity on small sample hydrologic regional regression, *ASCE Conf. Proc.*, 3713–3722, doi:10.1061/41173(414)389, 2011.
- Souza, F. A. and Lall, U.: Seasonal to Interannual Ensemble Streamflow Forecasts for Ceara, Brazil: Applications of a Multivariate, Semi-Parametric Algorithm, *Water Resour. Res.*, 39, 1307–1320, 2003.
- Srivastava, A. K., Rajeevan, M., and Kshirsagar, S. R.: Development of a high resolution daily gridded temperature data set (1969–2005) for the Indian region. NCC Research Report 8, National Climate Centre, India Meteorological Department, Pune, India, 15 pp., 2008.
- Stewart, I. T.: Changes in snowpack and snowmelt runoff for key mountain regions, *Hydrol. Process.*, 23, 78–94, 2009.
- Tippett, M. K., Barlow, M., and Lyon, B.: Statistical correction of central Southwest Asia winter precipitation simulations, *Int. J. Climatol.*, 23, 1421–1433, 2003.
- Venables, W. N. and Ripley, B. D.: *Modern Applied Statistics with S-PLUS*, 2nd Edn, Springer, ISBN 0-387-98214-0, 1997.
- Wang, E., Zhang, Y., Luo, J., Chiew, F. H. S., and Wang, Q. J.: Monthly and seasonal streamflow forecasts using rainfall-runoff modeling and historical weather data, *Water Resour. Res.*, 47, W05516, doi:10.1029/2010WR009922, 2010.
- Wedgbrow, C. S., Wilby, R. L., Fox, H. R., and O'Hare, G.: Prospects for seasonal forecasting of summer drought and low river flow anomalies in England and Wales, *Int. J. Climatol.*, 22, 219–236, 2002.
- Wilby, R. L., Wedgbrow, C. S., and Fox, H. R.: Seasonal predictability of the summer hydrometeorology of the River Thames, UK, *J. Hydrol.*, 295, 1–16, 2004.
- Wilks, D. S.: *Statistical Methods in the Atmospheric Sciences*, Academic Press, USA, 2006.
- Wood, A. W. and Lettenmaier, D. P.: A test bed for new seasonal hydrologic forecasting approaches in the western United States, *B. Am. Meteorol. Soc.*, 8, 1699–1712, 2006.
- Yadav, R. R., Park, W. K., Singh, J., and Dubey, B.: Do the western Himalayas defy global warming?, *Geophys. Res. Lett.*, 31, L17201, doi:10.1029/2004GL020201, 2004.