A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River Basin

Satish Kumar Regonda,[1,2] Balaji Rajagopalan,[1,2] Martyn Clark,[3] and Edith Zagona[4]

[1] We propose a multimodel ensemble forecast framework for streamflow forecasts at multiple locations that incorporates large-scale climate information. It has four broad steps: (1) Principal component analysis is performed on the spatial streamflows to identify the dominant modes of variability. (2) Potential predictors of the dominant streamflow modes are identified from several large-scale climate features and snow water equivalent information. (3) Objective criterion is used to select a suite of candidate nonlinear regression models each with different predictors. (4) Ensemble forecasts of the dominant streamflow modes are generated from the candidate models and are combined objectively to produce a multimodel ensemble, which are then back transformed to produce spatially coherent streamflow forecasts at all the locations. The utility of the framework is demonstrated in the skillful forecast of spring seasonal streamflows at six locations in the Gunnison River Basin at several lead times. The generated ensemble streamflow forecast provides valuable and useful information for optimal management and planning of water resources in the basin.

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

1. Introduction

[2] Fresh water resources in the western United States are under great stress in the wake of increased development, population, climate variability and climate change [e.g., *Hamlet et al.*, 2002; *Piechota et al.*, 2001]. The recent 5-year-long dry spell is raising many questions about system reliability, basin planning, water compacts, etc. In light of these concerns, careful planning and organized development is necessary to manage competing water demands of the system (e.g., municipal and industrial supplies, releases required for ecological and environmental purposes), so as to make the water resources sustainable.

[3] Larger rivers in the western United States are heavily controlled by water storage structures, which then are managed and operated to provide a reliable water source for a variety of needs. Furthermore, this region has extensive water laws, decrees, and compacts that impact the water resources management. The region is largely semiarid and prone to frequent dry spells; hence the storage structures are key to reliable water supply. Thus it can be seen that two broad aspects emerge as important for efficient water management: (1) skillful forecast of streamflows at

long lead times and (2) a decision support system that can evaluate various strategies incorporating the forecasts to guide the decision makers.

[4] On the Gunnison River, the study area in this paper, the major water development feature is the Aspinall unit of the Colorado River Storage Project. It consists of Blue Mesa, Morrow Point, and Crystal dams, reservoirs, power plants, and the Gunnison tunnel, which diverts water from the Gunnison River to the Uncompahgre Valley. These three reservoirs have combined water storage of approximately 1357 million m^3 and generation capacity of 287 MW. The Aspinall unit supplies water to prioritized water rights holders within Colorado, meets agriculture releases, downstream interstate compact deliveries, municipal and industrial demands, produces hydropower, and operates for other benefits such as minimum fish flows, recreation, flood control, federal water rights, Black Canyon National Park rights, and stream and lake protection.

[5] Most of the inflows (70% of the annual flow) to the Gunnison River comes as spring runoff of snowmelt from April to July and hence is a major component in the water management of the basin that needs to be forecasted. One of the major challenges for water managers is finding the optimal release schedule of spring flows from the Aspinall unit that satisfies the needs of endangered habitat while maximizing hydropower generation and meeting federal water rights. To operate the system efficiently under competing demands requires skillful forecast of spring season streamflows. Also several key decisions are made in January and February, and so long-lead spring streamflow forecasts are crucial. Currently, water managers make operational decisions based on a "24-month study" operation plan. In this, at the start of each month, streamflow fore-

[1]Department of Civil, Environmental and Architectural Engineering, University of Colorado, Boulder, Colorado, USA.

[2]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA.

[3]National Institute for Water and Atmospheric Research, Christchurch, New Zealand.

[4]Center for Advanced Decision Support for Water and Environmental Systems and Department of Civil, Environmental and Architectural Engineering, University of Colorado, Boulder, Colorado, USA.

casts/projections of the successive 24-month period are used to drive a decision support system model of the Aspinall unit. The system decision variables obtained from this 24-month outlook are used to make decisions, including a reservoir release schedule.

[6] The 24-month projection is issued by Bureau of Reclamation (BOR), and it typically uses the Colorado Basin River Forecast Center's (CBRFC) April to July forecast from January through July, and the CBRFC 3 month forecast the remainder of the year, followed by climatology (or monthly average) flows. CBRFC issues 10%, 50%, and 90% exceedance forecasts using primarily extended streamflow prediction (ESP) and statistical regression that utilizes snow, streamflow, precipitation and climate indices. Thus a single trace of the 24-month flows is developed in association with one of CBRFC's single forecasts and BOR's subjective criteria, and is used in the decision support model. The key drawbacks of BOR's forecast set up are (1) a single 24-month trace does not capture any uncertainty and (2) the ESP method produces ensembles by driving a hydrologic model with past weather scenarios and current initial conditions, e.g., if there are 20 years of data available at a location the ESP method can only produce 20 ensembles. Although the input time series of the ESP method is adjusted with monthly and seasonal forecasts of climate variables (e.g., temperature, precipitation, teleconnection indices), the range of uncertainty captured is quite limited. Furthermore, there is growing evidence of large-scale climate features' influence in the regional hydrology of the western United States that could be incorporated in the forecast methodology for improved skill [e.g., *Grantz et al.*, 2005, and references therein]. Particularly, large-scale climate information offers predictability for issuing forecasts of spring streamflow during December through February, when the snowfall information is limited at best [e.g., *Grantz et al.*, 2005].

[7] With this motivation, the main goal of the present study is to develop a framework for issuing long-lead ensemble forecasts of spring streamflow at several locations in the Gunnison River Basin (GRB) incorporating large-scale climate information. The paper is organized as follows: A background on the role of large-scale climate features in the variability of western United States hydroclimatology is first presented to motivate the need to incorporate climate information in the forecasting framework. The basin details and the data sets used are then described, followed by a detailed description of the proposed multimodel ensemble framework. Application to the GRB streamflows is then presented, concluding with a summary and discussion of the results.

2. Large-Scale Climate and Hydroclimatology in the Western United States

[8] There is a rich literature documenting the connection between the variability of western U.S. hydroclimatology and large-scale climate forcings especially, El Nino Southern Oscillation (ENSO), Pacific North American (PNA) pattern, and Pacific Decadal Oscillation (PDO). Below we provide a short summary and refer the readers to our comprehensive description by *Grantz et al.* [2005] and *Rajagopalan et al.* [2005].

[9] Interannual variability in the western U.S. hydroclimatology is driven largely by ENSO. During El Niño events (warm sea surface temperature anomalies in the central and eastern equatorial Pacific ocean) the subtropical jet over the southwestern United States (part of the PNA pattern) especially during winter strengthens [e.g., *Bjerknes*, 1969; *Horel and Wallace*, 1981] and consequently, below-normal precipitation results in the Pacific Northwest and above-normal precipitation in the desert Southwest [e.g., *Redmond and Koch*, 1991; *Cayan and Webb*, 1992; *Dettinger et al.*, 1998; *Cayan et al.*, 1998]. Generally opposing signals are evident during La Niña events (cooler sea surface temperature in the central and eastern equatorial Pacific), but some nonlinearities are present [*Hoerling et al.*, 1997; *Clark et al.*, 2001]. Similar ENSO teleconnection patterns have been observed in the interannual variability of winter snow water equivalent [*Clark et al.*, 2001; *Cayan*, 1996], surface temperatures [*Redmond and Koch*, 1991; *Higgins et al.*, 2002; *Gershunov and Barnett*, 1998], and streamflows [*Kahya and Dracup*, 1993, 1994; *Dracup and Kahya*, 1994; *Piechota et al.*, 1997; *Maurer et al.*, 2004]. On decadal timescales the hydrologic variability is thought to be driven primarily by the PDO [*Mantua et al.*, 1997; *McCabe and Dettinger*, 1999, 2002; *Hidalgo and Dracup*, 2003; *Brown and Comrie*, 2004], even though there is debate on the independence of PDO from ENSO [*Newman et al.*, 2003]. Recent studies also indicate a significant shift in the seasonal cycle of the western U.S. hydroclimatology [*Cayan et al.*, 2001; *Regonda et al.*, 2005a; *Stewart et al.*, 2005] modulated by the ENSO and the general warming trend, especially, the early occurrence of spring warming and consequent early spring time peak flows due to early snowmelt. The large-scale forcings, particularly ENSO and consequently, PNA pattern, are highly persistent and thus have enabled long-lead hydroclimate predictions of United States surface temperatures [*Higgins et al.*, 2004], Columbia River streamflows [*Hamlet and Lettenmaier*, 1999; *Clark et al.*, 2001] and Truckee and Carson River flows [*Grantz et al.*, 2005].

[10] Despite the strong links between large-scale climate forcings and western U.S. hydroclimatology, there is substantial variability in the strength of the teleconnections from one river basin to another [*McCabe and Dettinger*, 2002]. Moreover, relatively minor shifts in large-scale atmospheric patterns can result in large differences in surface climate [e.g., *Yarnal and Diaz*, 1986]. Therefore the typical indices that capture ENSO or PNA may not be useful in explaining the variability of the hydroclimatology and nonstandard indices have to be developed [e.g., *Grantz et al.*, 2005]. Such is the case in the GRB, which is situated in the interior west and lies outside the regions most strongly affected by ENSO variability [e.g., *Cayan and Peterson*, 1989; *Klein et al.*, 1965; *Weare and Hoeschele*, 1983, *Cayan et al.*, 1998; *Dettinger et al.*, 1998]. Several researchers [*Changnon et al.*, 1993; *McCabe*, 1994; *McCabe and Legates*, 1995; *McCabe*, 1996; *Robertson and Ghil*, 1999] have identified links between large-scale atmospheric circulation features (e.g., 700 mbar heights) and the winter precipitation in the western United States, and consequently streamflow. These circulation features are quite different from the typical PNA pattern. Therefore appropriate large-scale climate features that modulate the
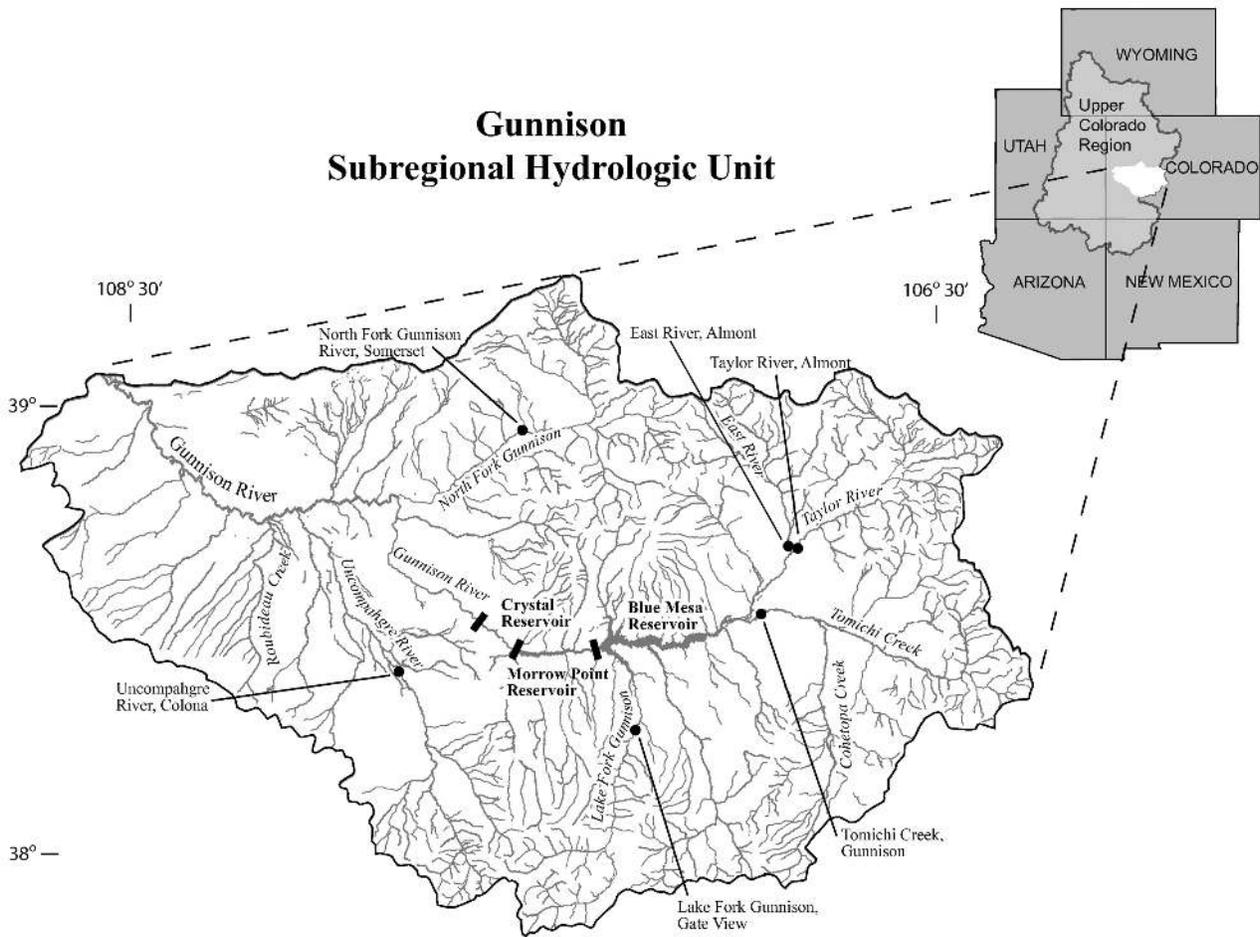
Figure 1. Map of the Gunnison River Basin and six key streamflow locations (shown as circles). Map was provided by James Pasquotto, University of Colorado, Boulder.

streamflow in the GRB have to be identified for use in an ensemble streamflow forecasting framework.

3. Study Basin

[11] The GRB is located in the southwestern part of Colorado (Figure 1). Six key streamflow locations in the basin are also shown in Figure 1. With a drainage area of approximately 7930 mi^2 (20,618 km^2) and basin elevations ranging from 1387 to 4359 m, the GRB extends from the continental divide to Grand Junction where it joins the Colorado River [McCabe, 1994]. Climate in this basin varies greatly due to its broad range of elevation with most of the annual precipitation falling as snow. The Gunnison River is a major tributary of the Colorado River and contributes approximately 42% of the streamflow for the Colorado River [Ugland et al., 1990]. The average annual hydrograph of the GRB is shown in Figure 2. It can be seen that almost all (greater than 70%) of the annual flow occurs during the spring (April–July). Consequently, spring flow plays a vital role in the decision and management of water resources and water managers wish to have accurate long-lead forecasts.

4. Data Sets

[12] A brief description of the data sets to be used to develop a long-lead streamflow forecast system is presented below.
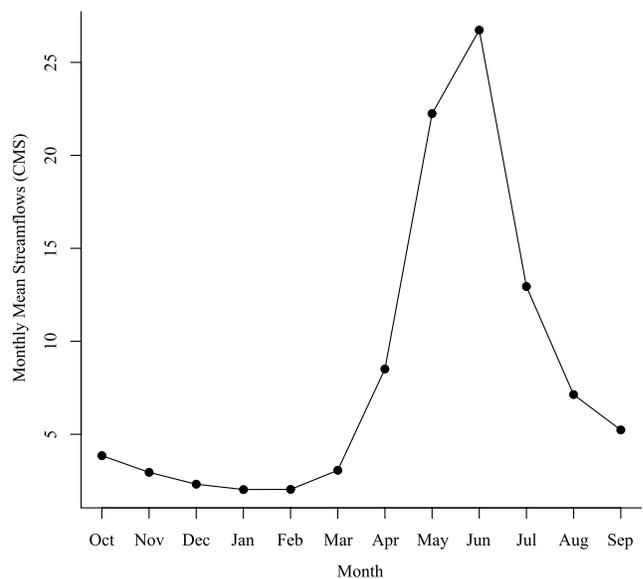
Figure 2. Annual average hydrograph of the Gunnison River Basin. Monthly values (m^3/s (cms)) are the average across the six locations.

Table 1. Gunnison River Basin Streamflow Data Information

Station Number	Site Number (USGS)	Site Name	Elevation, m	Drainage Area, 10^6 m^2
1	09110000	Taylor River at Almont, CO	2442	1235
2	09112500	East River at Almont, CO	2440	749
3	09119000	Tomichi Creek at Gunnison, CO	2325	2748
4	09124500	Lake Fork at Gate View, CO	2386	865
5	09132500	North Fork Gunnison River near Somerset, CO	1914	1362
6	09147500	Uncompahgre River at Colona, CO	1926	1160

4.1. Streamflow

[13] Streamflow locations are selected from the Hydro Climate Data Network (HCDN) developed by the USGS [*Slack and Landwehr*, 1992]. For the GRB HCDN provided six locations (shown in Figure 1) that have continuous flow record from 1949 onward, station details are presented in Table 1. The selected streamflow locations are situated in five of the six cataloging units (GRB is divided into six cataloging units) covering the entire basin. Four of the six stations contribute inflows to the water storage structures comprising the Aspinall unit. Spring flow, the summation of daily flows from April through July, is computed for each of the locations for further analysis.

4.2. Snow

[14] Snow water equivalent (SWE) data are obtained from snow course surveys conducted by the Natural Resource Conservation Service (NRCS). SWE measurements are generally taken on or about the beginning of each month especially at the beginning of April, which is representative of peak SWE in many regions. Stations with at least 80% of their measurements available from 1949 to 2002 for 1 February, 1 March, or 1 April are selected. Three, four, and six stations met the criteria for the months February, March and April, respectively.

4.3. Large-Scale Climate Variables

[15] Ocean-atmospheric circulation variables that capture the large-scale climate forcings are available from NOAA's Climate Diagnostics Center Web site (http://www.cdc.noaa.gov). In particular, some of the variables considered are geopotential height, surface air temperature (SAT), sea surface temperature (SST), and zonal and meridional winds. These variables are provided on a $2.2° \times 2.2°$ grid spanning the globe from the NCEP-NCAR reanalysis project [*Kalnay et al.*, 1996] for the period 1949 to present.

5. Methodology

[16] The objectives of the methodology are to provide an ensemble of spring streamflow forecasts at all the locations and at different monthly lead times incorporating large-scale climate information. Spring seasonal forecasts at all the locations need to be generated simultaneously so as to capture the spatial dependence among the locations.

[17] To this end, we propose a new modeling framework that produces multimodel ensemble forecasts of streamflows. The proposed methodology consists of four broad components: (1) principal component analysis (PCA) of the (spring season) streamflows at the locations to identify the dominant modes of variability, (2) climate diagnostics to obtain the large-scale ocean-atmosphere predictors of the

dominant modes, (3) forecasting framework and multimodel selection objective criterion to select a suite of models from the predictor set based on the generalized cross validation (GCV), and (4) multimodel ensemble streamflow forecast algorithm, in which ensembles of the dominant modes are predicted from the multimodels, and then combined, and translated into the original flow space to obtain the streamflow forecasts. These components will be described in detail in the following sections.

5.1. Principal Component Analysis

[18] Principal component analysis is widely used in climate research. This method decomposes a space-time random field (i.e., a multivariate data set such as the seasonal streamflows at the six locations in the GRB) into orthogonal space and time patterns using eigendecomposition [*Von Storch and Zwiers*, 1999]. The patterns are ordered according to the percentage of variance captured, i.e., the first space-time pattern (also called "mode") captures the most variance present in the data and so on. The temporal patterns are called principal components (PCs). Typically, the first few modes (PCs) capture most of the variance present in the data. This can also be thought of as a dimension reduction technique, where a large multivariate data set is effectively represented by a few PCs (i.e., smaller dimension). Furthermore, since the PCs are orthogonal they can be analyzed independently and combined to reconstruct the original data.

[19] The mathematical formulation is as follows:

$$[\vec{Z}]_{N \times M} = [Y]_{N \times M} [\vec{E}]_{M \times M} \tag{1}$$

$$[Y]_{N \times M} = [\vec{Z}]_{N \times M} [\vec{E}]_{M \times M}^T \tag{2}$$

where Z is the original flows, Y is the principal components, E is the eigenvectors, M is the number of streamflow locations, and N is the length of the data. Hence E can be considered a transformation matrix. The decomposition is obtained by minimizing the error $\varepsilon_i = E[Z - \hat{Z} E_i]^2$ (this maximizes variance) such that $EE^T = 1$ (orthonormal criteria).

5.2. Climate Diagnostics

[20] The next step is to search for indices that can be used to predict the dominant streamflow PCs. The leading PCs are correlated with global ocean and atmospheric circulation variables (i.e., surface temperatures, geopotential heights, winds) from preceding seasons. In this case, the leading PCs of the spring streamflows will be correlated with circulation variables from the preceding fall and winter months. From the resulting correlation maps, regions that exhibit strong

correlations are used to develop potential predictors. These are typically the area-averaged values of the variables. Thus a suite of potential predictors can be obtained. Composite analysis will be performed to identify the physical mechanisms driving the variability of the streamflows and also the consistency of the predictors.

5.3. Forecast Framework

[21] The leading PCs and their potential predictors are incorporated into a statistical model, which is typically of the form:

$$y = f(x_1, x_2, x_3, \ldots x_R) + e \qquad (3)$$

where f is a function fitted to the R predictor variables (x_1, x_2, …, x_R), y is the dependent variable (in this case the leading PC of the spring streamflows) and e is the errors assumed to be normally (or Gaussian) distributed with a mean of 0 and variance σ. Traditional statistical methods fit a linear function that minimizes the squared errors, known a linear regression. The theory behind this approach, the procedures for parameter estimation, and hypothesis testing are very well developed [e.g., *Helsel and Hirsch*, 1995; *Rao and Toutenburg*, 1999] and are widely used. However, they do have some drawbacks: (1) the assumption of a Gaussian distribution of the errors and the variables and (2) fitting a global relationship (e.g., a linear equation in the case of linear regression) between the variables. If the linear model is found inadequate, higher-order models (quadratic, cubic, etc.) have to be considered, which can be difficult to fit in the case of short data sets. Also if the variables are not normally distributed, which is often the case in practice, suitable transformations have to be obtained to transform them to normal distribution. All of this can make the process unwieldy. Thus a more flexible framework would be desirable.

[22] Local estimation methods (also known as nonparametric methods) provide an attractive alternative. In this, the function f is fitted to a small number of neighbors in the vicinity of the point at which an estimate is required. This is repeated at all the estimation points. Thus, instead of having a single equation that describes the entire data set, there are several "local fits," each capturing the local features. This provides the ability to model any arbitrary features (linear or nonlinear) that the data exhibits.

[23] There are several approaches for local functional estimation applied to hydrologic problems [see *Lall*, 1995]. Of these, the locally weighted polynomial regression (LWP) is simple and robust and has been used in a variety of hydrologic and hydroclimate applications with good results for streamflow forecasting on the Truckee and Carson river basins [*Grantz et al.*, 2005], salinity modeling on the upper Colorado river basin [*Prairie et al.*, 2005a], forecasting of Thailand summer rainfall [*Singhrattna et al.*, 2005], and spatial interpolation of rainfall in a watershed model [*Hwang*, 2005]. Given these experiences, we adopt the LWP method in this research.

[24] A brief description of the method is provided here and for specific details we refer the readers to *Loader* [1999] and *Grantz et al.* [2005]. In the LWP method a small number, $K = \alpha*N$ (where $\alpha = (0,1]$ and N = number of observations) neighbors of the point of estimate, $\mathbf{x}*$ are

identified from the data set. To the K neighbors, a polynomial of order p is fitted (equation (3)) by a weighted least squares method where the K observations are weighted inversely to their distances to $\mathbf{x}*$ using a weight function. The fitted polynomial is then used to obtain the mean estimate for the dependent variable, y* at the point of estimate. The standard regression theory also provides an estimate of the error variance (σ_{1e}^2) corresponding to the predicted value of the dependent variable, y* [*Loader*, 1999]. Random normal deviates with this variance and zero mean, when added to predicted value, y*, provide an ensemble. This assumes that the errors are normally distributed around the predicted value. *Prairie et al.* [2005b] developed a residual resampling approach to generate ensembles. In this the residuals of the LWP fit are resampled and added to the mean estimate. This can better capture the local error structure without assuming normality, but on short data sets the variety of ensembles produced is limited. Note that the LWP approach collapses to a traditional linear regression, when α is set to 1, p is set to 1, and each observation is weighted the same.

[25] The two key parameters in this method, the size of the neighborhood, K, and the order of the polynomial, p are obtained by minimizing an objective criterion, generalized cross validation (GCV) which is a good estimate of predictive risk of the model, unlike other functions which are goodness of fit measures [*Craven and Whaba*, 1979]. This is given as:

$$GCV(K,p) = \frac{\sum_{i=1}^{N} \frac{e_i^2}{N}}{\left(1 - \frac{q}{N}\right)^2} \qquad (4)$$

where e_i is the model residual, N is the number of data points, q is the number of parameters in the local polynomial model. Typically, several combinations of (K, p) are used and the combination that results in a minimum GCV value is selected.

5.4. Multimodel Selection

[26] The predictor variables used in the regression equations tend to be highly correlated. If all the correlated predictors are used in the model, it can lead to overfitting and poor skill in prediction; this is known as "multicollinearity." Thus the "best" subset of predictors needs to be identified. Typically, this is done using stepwise regression [e.g., *Rao and Toutenburg*, 1999; *Walpole et al.*, 2002] where in, an objective function such as Mallow's C_p statistic or adjusted R^2 or AIC (Akaike information criteria) or an F test, is calculated from the fitted model to several predictor combinations. The best subset is then selected based on the combination that gives the optimal value for the chosen objective function. For noisy data (i.e., most real data) the values of the objective functions for several predictor combinations tend to be very close, suggesting that several combinations (i.e., candidate models) might be admissible. Thus selecting the "best" subset might not be a good strategy, which warrants a multimodel approach. Recent studies show that multimodel ensemble forecasts tend to perform much better than a single model forecast [*Hagedorn et al.*, 2005; *Krishnamurti et al.*, 1999, 2000; *Rajagopalan et al.*, 2002].

[27] Here we propose the use of GCV for multimodel selection. The approach is as follows: (1) GCV is computed for all the possible combinations of the predictor set and parameter set (K, p). (2) All combinations with GCV values within a prescribed threshold are selected as admissible, constituting the pool of candidate models, i.e., multimodels. (3) Combinations with predictor variables significantly correlated amongst each other (i.e., multicollinear) are removed from the multimodel pool.

5.5. Multimodel Ensemble Forecast Algorithm

[28] The ensemble forecast from the multimodels identified above is summarized in the following six steps.

[29] 1. PCA is performed on the six streamflows in the GRB, and the leading PCs are selected.

[30] 2. All potential predictors of the leading PCs are identified from climate diagnostics.

[31] 3. These are passed through the forecasting framework and the multimodel selection. This results in a pool of multimodels, which includes the predictor set, and the corresponding parameter set.

[32] 4. Ensemble (say, 100) forecasts of the leading PCs are generated from each model of the multimodel pool. For each ensemble, the other nonleading PCs are randomly selected (i.e., bootstrapped) from those obtained from the PCA. Thus each ensemble forecast member consists of all PCs with the leading PCs obtained from the model forecast and the rest bootstrapped from the historic values. If ten models are selected in the multimodel pool the ensemble forecast will be a matrix of size (1000×6).

[33] 5. The ensemble forecast matrix of the PCs is multiplied by the eigenvectors to back transform into the streamflow space, i.e., $Z^* = Y^* * \vec{E}$.

[34] 6. Clearly, ensembles from all the models are not equal. The model with the least GCV value should be given more weight relative to the one with a higher GCV. To this end, a vector of probabilistic weights is created based on the GCV values (the weights are normalized to sum to unity); using this weight metric, a model is selected; then an ensemble member from this model is selected at random. This is repeated numerous times (say, 100), thus resulting in a final multimodel ensemble forecast.

[35] Steps 2 through 6 are repeated for each lead time; thus each lead time will have a different suite of multimodels with their own predictor and parameter sets.

5.6. Links to Model Averaging Literature

[36] The typical statistical modeling approach involves identifying a single best model based on an objective criterion. This neglects model inadequacy and increases the associated risk of model inferences. To alleviate this, a combination of several candidate models, also known a "model averaging" has been proposed [*Reid*, 1968; *Bates and Granger*, 1969; *Clemen*, 1989]. Another approach in a similar vein is to generate ensembles from several candidate models and combine them to provide a model average probabilistic forecast. This approach has been shown to perform better than any individual model ensemble, particularly in short-term and seasonal climate forecast [*Hagedorn et al.*, 2005; *Hou et al.*, 2001; *Krishnamurti et al.*, 1999, 2000; *Regonda et al.*, 2005b; *Tamea et al.*, 2005]. The candidate models are combined by weighting them, where the weights

are obtained from different approaches, e.g., regression, Bayesian framework, etc. Weighting the multimodels in a Bayesian framework is known as Bayesian model averaging [*Raftery et al.*, 1997; *Hoeting et al.*, 1999]. There are several variations within this framework [e.g., *Madigan and Raftery*, 1994]. The key message in all of these is that combining several models tends to perform better in terms of predictive skill and reliable uncertainty estimates than a single model.

[37] The multi model combination approach proposed in this paper is consistent with the model averaging framework and provides an attractive alternative approach for performing model averaging. The multimodels are selected based on the GCV criteria and they are combined using weights based on their GCV values as described in the previous sections.

5.7. Forecast Skill Evaluation

[38] Since we generate an ensemble forecast, i.e., the probability density function (PDF), the skill of the forecast needs to be evaluated in probabilistic terms. One such common measure is the ranked probability skill scores (RPSS) [*Wilks*, 1995]. Essentially, it measures the accuracy of multicategory probability forecasts relative to a climatological forecast. Typically, the flows are divided into k mutually exclusive and collectively exhaustive categories for which the proportion of ensembles falling in each category constitutes the forecast probabilities $(p_1, p_2, \ldots, p_k)$. The observational vector $(d_1, d_2, \ldots, d_k)$ is obtained for each forecast, where d_k equals one if the observation falls in the kth category and zero otherwise.

[39] The ranked probability skill score (RPSS) is defined as follows:

$$\mathrm{RPS} = \sum_{i=1}^{k} \left[\left(\sum_{j=1}^{i} p_j - \sum_{j=1}^{i} d_j \right)^2 \right] \qquad (5)$$

$$\mathrm{RPSS} = 1 - \frac{\mathrm{RPS(forecast)}}{\mathrm{RPS(climatology)}} \qquad (6)$$

In this research, the streamflows are divided into three categories, at the tercile boundaries, i.e., 33rd and 66th percentile of the historical observations. Values below the 33rd percentile represent "dry," above 66th percentile "wet" and "near normal" otherwise. Of course, the climatological forecast for each of the tercile categories is 1/3.

[40] The RPSS ranges from negative infinity to positive one. Negative RPSS values indicate the forecast accuracy to be worse than climatology, positive to be higher than climatology, zero to be equal to that of climatology, and a perfect categorical forecast yields an RPSS value of unity. In this application the RPSS is calculated for each year and the median value is reported.

6. Results

[41] The multimodel ensemble forecast framework developed in the previous section was applied to streamflows from the six locations in the GRB mentioned earlier. The results are described below.
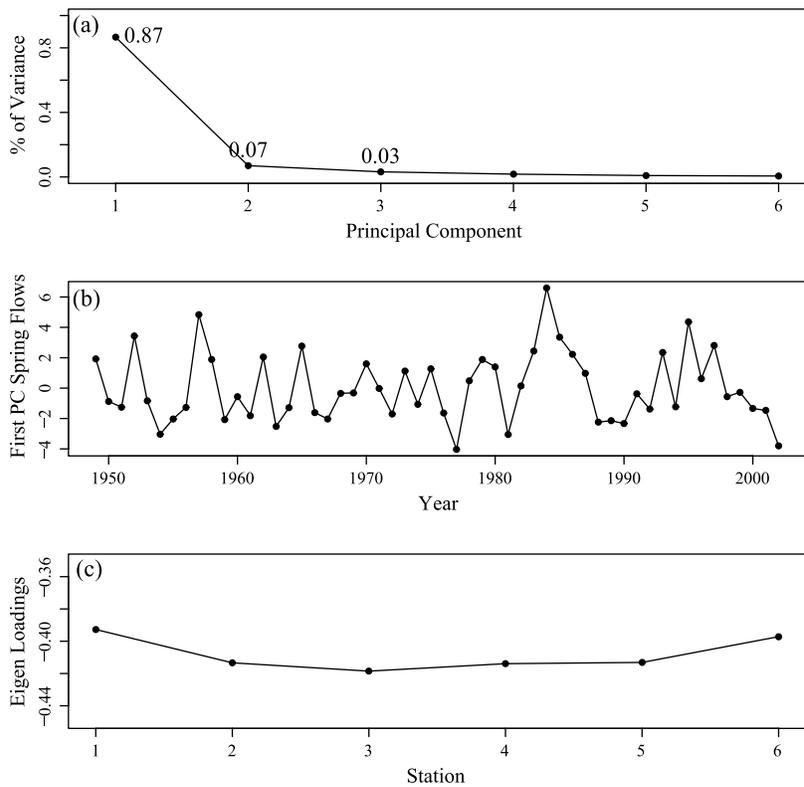
Figure 3. (a) Percentage variance explained by the six principal components (PCs), (b) time series of the first PC, and (c) eigenloadings of the first PC at the six streamflow locations.

6.1. Streamflow Characteristics

[42] PCA was performed on the normalized spring season streamflows from the six locations. The first PC explained most of the variance (around 87%) and the remaining five PCs together explained about 13% of the variance (Figure 3a). Clearly, the first PC (Figure 3b) is the dominant and leading "mode" containing the "signal," while the rest can be treated as "noise." The eigenvectors (i.e., loadings) at all the six
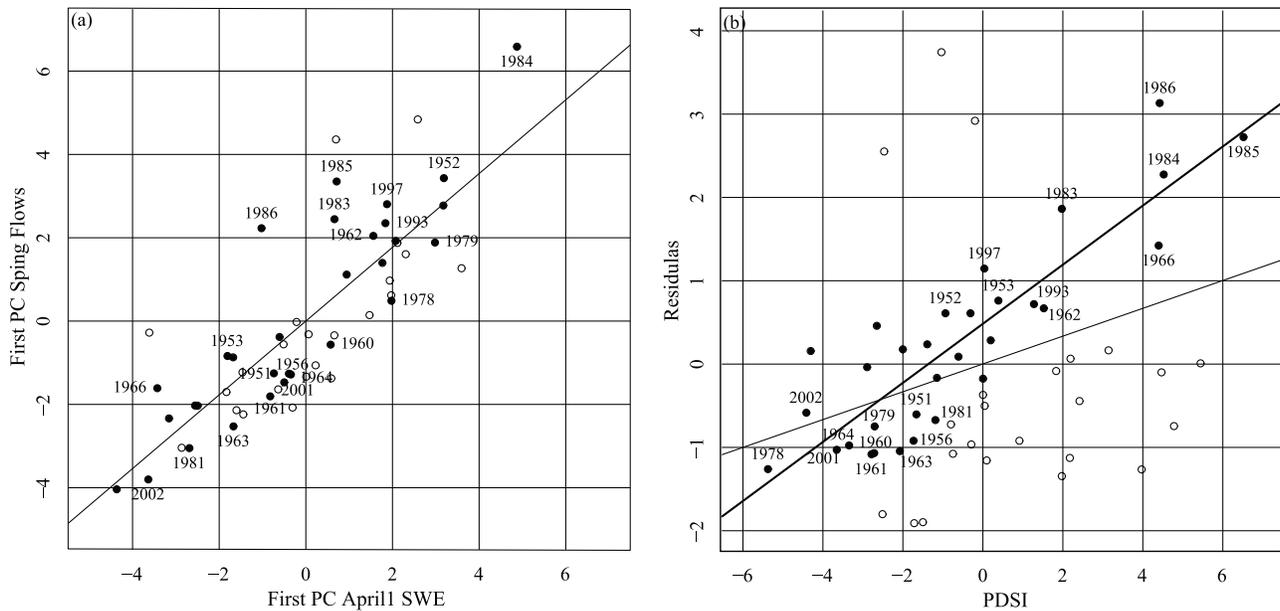
Figure 4. (a) Scatterplot of first PC of spring flow and 1 April SWE along with the best fit linear regression line and (b) scatterplot of residuals of the first PC of spring flow from the regression fit in Figure 4a with the preceding fall season PDSI. The thin line is the best fit line for the scatter, and the thick line is the best fit line of the solid circles.
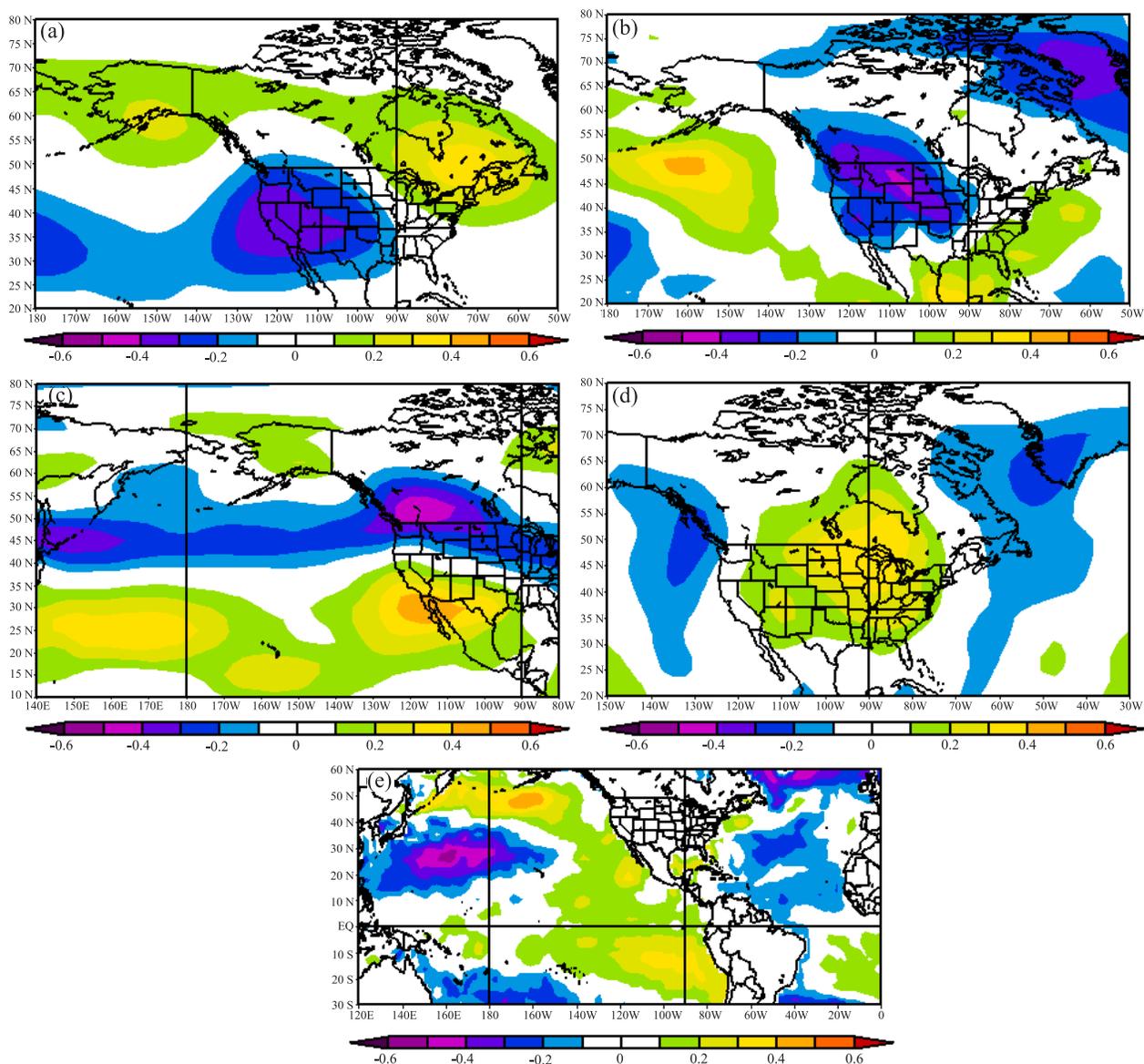
Figure 5. Correlation between the first PC of spring flow and November–March large-scale climate variables (a) geopotential height (700 mbar), (b) surface air temperature, (c) zonal wind (700 mbar), (d) meridional wind (700 mbar), and (e) sea surface temperature. Maps were generated from NOAA's Climate Diagnostic Center Website.

locations (Figure 3c) corresponding to the first mode are of similar magnitude and sign suggesting that the basin is homogenous in its variability. To corroborate this, the first PC correlates strongly (correlation coefficient >0.85) with the basin average spring flow.

[43] PCA is also performed on the SWE data separately for each month (1 February, 1 March, and 1 April). Here too, the first PC explained most of the variance in all the months. As expected, the leading PC of SWE and the leading PC of the spring streamflows were highly correlated, with the first PC of 1 April SWE having the highest correlation (0.82) with the flow PC (Figure 4a). It is interesting to note that there are few years (e.g., 1979, 1984, 1986, 2002 shown as solid circles in Figure 4a) in which flows are not proportional to the 1 April SWE. Our hypothesis is that if the preceding summer/fall is drier than normal and the following

winter is wetter than normal, then a large part of the following spring runoff is absorbed by the soil and lost to the atmosphere through evaporation, thus reducing the streamflow output. To test this, we computed the average preceding fall season Palmer drought severity index (PDSI) over the basin, a good surrogate for soil moisture [*Dai et al.*, 2004], and plotted them against the residuals of the linear regression between the first PC of the streamflow and SWE (Figure 4b). The correlation between the residuals and the fall PDSI is 0.35, which is low but statistically significant. However, for the years that do not follow the linear SWE-flow relationship (solid circles in Figures 4a and 4b) the residuals show a very strong correlation with the preceding fall PDSI (0.873, solid line in Figure 4b). This suggests the key role of the PDSI in modulating the streamflows especially during years with a drier than normal fall
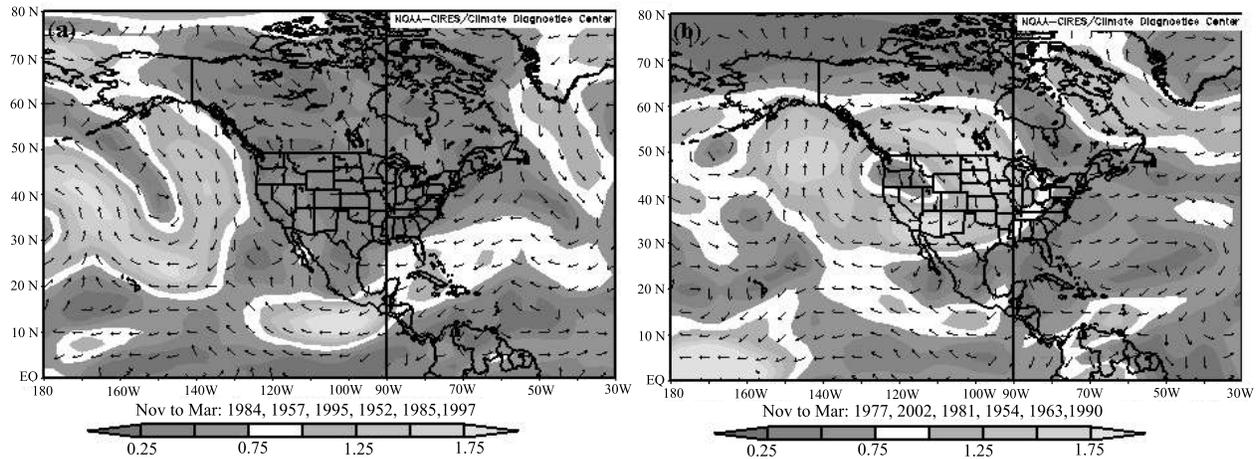
Figure 6. Composite maps of vector wind anomalies at 700 mbar for (a) wet years and (b) dry years. Maps were generated from NOAA's Climate Diagnostic Center Website.

season followed by a wet winter and spring, and the need for it to be included in the suite of predictors.

6.2. Climate Diagnostics: Predictor Selection

[44] As a first step, we correlated the first PC of spring streamflow with the standard large-scale climate forcing indices of ENSO, PDO and PNA, and we found no relationship. This is consistent with the past research results for the GRB [*Cayan*, 1996; *Clark et al.*, 2001; *Klein*, 1963; *Klein and Bloom*, 1987; *McCabe and Legates*, 1995], which suggest that climate signals (particularly ENSO and PNA) in the intermountain West tend to be weak due to the fact that the region is situated in the transition zone of the ENSO and PNA teleconnections.

[45] The first flow PC was correlated with the preceding seasons' atmospheric and oceanic circulation variables, i.e., 700 mbar geopotential heights, SAT, SST, zonal (700 mbar) and Meridional (700 mbar) winds. Correlation maps with the November–March season are presented in Figure 5. Negative correlations are observed between the spring flows and 700 mbar heights over the western United States (Figure 5a), indicating an above average spring streamflow in the GRB with negative geopotential height anomalies in this region, and vice versa. The negative height anomalies tend to direct the storm tracks into the basin resulting in increased SWE, and consequently, increased streamflows in the spring. The surface air temperatures are negatively correlated with the spring flows in this region (Figure 5b), consistent with Figure 5a. Correlations with zonal and meridional winds at 700 mbar (Figures 5c and 5d) are consistent with the 700 mbar geopotential heights in that winds over the southwestern United States bring moisture into the GRB basin, leading to above average spring streamflows. The SST correlation (Figure 5e) shows a weak ENSO pattern with positive correlation in the tropical Pacific and negative in the northern Pacific. The correlations in the regions described above are statistically significant. Composite maps of vector winds at 700 mbar corroborate the results from the correlation maps (Figure 6). In wet years (years with PC values above the 90th percentile) the average wind pattern (Figure 6a) is from the southwest direction, blowing in to the basin from the ocean bringing in

moisture, and consequently, more snow and streamflow and vice versa during dry years (Figure 6b). The SST composites (not shown) also indicated results consistent with the correlations. These observations are qualitatively in agreement with *McCabe and Legates* [1995].

[46] From the correlation maps we identified the regions of strong correlation to develop the predictors by area averaging over these regions. We take the difference between the regions of positive and negative correlations, in order to reduce the number of predictors and enhance the predictor signal. From Figure 4b we saw that the land surface information during the preceding fall plays a role in the following spring melt flows. Hence we also included the PDSI in the set of predictors. In addition, we included the first PC of the SWE. In all, the climate diagnostics suite of predictors (geopotential height anomalies, SAT, winds, PDSI, SWE) were developed for the 1 April forecast and they are detailed in Table 2.

[47] The correlation analysis is performed for all the lead times separately, e.g., for 1 March forecast, the first PC of spring flows is correlated with November–February climate variables and the predictors obtained from the maps and so on. We found the regions of strong correlation to be largely the same with slight differences among the lead times.

6.3. Multimodel Ensemble Forecast

[48] As mentioned at the outset, spring streamflow forecasts at several lead times before the spring flow period are required by the water managers; hence we issue forecasts of the upcoming spring streamflows on the first of every month starting 1 December through 1 April. The predictors and the first PC of spring flows are passed through the modeling framework described in the previous section. The residuals (or errors) from the model were all found to be normally distributed using a Kolmogorov-Smirnov test [*Wilks*, 1995]. The forecasts are made in a cross-validated mode, i.e., a year is dropped from the data set, the PCA analysis is performed on the remaining data, and an ensemble forecast is issued. This is repeated for all the years. Of course, for the forecasts issued on 1 December and 1 January, only the climate predictors are used (as the SWE information is not yet available). However, for the forecasts

Table 2. Suite of Predictors of GRB Spring Streamflow for 1 April Forecast

| Index Series | Climate Variable | Positively Correlated Regions | | Negatively Correlated Regions | |
		P1	P2	N1	N2
SAT-N1	surface air temperature	-	-	latitude: 35.0–55.0 longitude: 235.0–265.0	-
(GPH-P1) – (GPH-N1)	geopotential height at 700 mbar	latitude: 42.5–57.5 longitude: 272.5–300.0	-	latitude: 32.5–42.5 longitude: 230.0–250.0	-
(GPH-P2) – (GPH-N1)	geopotential height at 700 mbar	-	latitude: 55.0–62.5 longitude: 200.0–217.5	latitude: 32.5–42.5 longitude: 230.0–250.0	-
(MW-P1) – (MW-N1)	meridional wind at 700 mbar	latitude: 52.5–40.0 longitude: 255.0–280.0	-	latitude: 55.0–75.0 longitude: 230.0–250.0	-
(MW-P1) – (MW-N2)	meridional wind at 700 mbar	latitude: 52.5–40.0 longitude: 255.0–280.0	-	longitude: 300.0–320.0	latitude: 57.5–42.5 longitude: 222.5–232.5
(ZW-P1) – (ZW-N1)	zonal wind at 700 mbar	latitude: 35.0–25.0 longitude: 235.0–257.5	-	latitude: 55.0–47.5 longitude: 235.0–255.0	-
(ZW-P2) – (ZW-N1)	zonal wind at 700 mbar	-	latitude: 30.0–20.0 longitude: 145.0–185.0	latitude: 55.0–47.5 longitude: 235.0–255.0	-
(SST-P1) – (SST-N1)	sea surface temperature	latitude: 50.5–44.8 longitude: 187.5–204.4	-	latitude: 33.3–21.9 longitude: 144.4–185.6	-
(SST-P2) – (SST-N1)	sea surface temperature	-	latitude: –8.6–18.1 longitude: 247.5–275.6	latitude: 33.3–21.9 longitude: 144.4–185.6	-
PDSI[a]	Palmer drought severity index				
PC1 SWE[b]	April 1 SWE				

[a]Colorado region 2 (August–October seasonal values).
[b]From the Gunnison River Basin.

issued in subsequent months, the first PC of SWE is also included in the predictor mix.

[49] As described in the methodology section, several models (i.e., predictor and parameter combinations) yield similar GCV values. This can be noticed in Table 3 where the first 6 models are shown for the 1 April forecast, which illustrates the difficulty in selecting a single model uniquely and underscores the need for multimodel ensemble approach. We defined a threshold of 20% of the least GCV value and all the models within this range are selected. By inspection we found that models within this threshold tended to have GCV values clustered together (as seen in Table 3). This rule of thumb seemed to work well for the application presented. However, we recognize the need for a more objective criterion. The GCV based optimal values of K and p were found to be one, indicating that the relationship between the predictors and the flow is largely linear. However, the local functional estimation aspect of the method enables it to capture subtle nonlinearities between the variables that are present (figure not shown) with these parameter values.

[50] The number of models selected tended to decrease from December to April. This is intuitive, in that on 1 December SWE is not available and hence the forecasts have to be made only from climate information. Consequently, individual models have greater uncertainty, and more models qualify as candidates for the multimodel pool. However, on 1 April, SWE information is complete and it is the best predictor of the ensuing spring streamflow. Therefore fewer models with other predictor variables are necessary to capture the streamflow dynamics and its variability. Hence a smaller number of multimodels are selected.

[51] On the basis of the criteria for selecting models for the multimodel ensemble, we identified 15 models for the 1 December forecasts and 6 models for the 1 April forecasts. Ensemble forecasts for the streamflow location, Tomichi, issued on 1 January and 1 April are shown as box plots in Figure 7. The box represents the interquartile range, the whiskers are the 5th and 95th percentile of the ensembles, and the horizontal line within the box is the median. The dashed horizontal lines represent the 33rd, 50th, and 66th percentiles of the historic spring streamflow data at this location, and the historic values are shown as points connected by a solid line. Three observations can be made from Figure 7: (1) The ensembles are shifted in the right direction of the observed streamflow in almost all years. (2) The ensemble forecasts issued on 1 January (with only the large-scale climate information) seem to capture the observations quite well. (3) As to be expected, the ensemble forecasts issued on 1 April are very good. The median RPSS for the 1 January forecast is 0.51 and the 1 April is 0.77. This indicates that skillful predictions of the spring streamflows can be made for the GRB in the middle of winter even when the snow information is absent, which is quite significant for water resources management.

[52] In order to evaluate forecast performance for extreme flows, flows are sorted into three categories: dry years (flows less than 25th percentile of the data), wet years (flows greater than 75th percentile), and near normal (remaining) years. Box plots of the ensemble forecasts for the wet and dry years are shown in Figures 8 and 9, respectively. It can be seen that the ensembles do quite well

Table 3. Multimodel Combinations for 1 April Forecast[a]

Number of Predictors	SAT-N1	(GPH-P1) – (GPH-N1)	(GPH-P2) – (GPH-N1)	(MW-P1) – (MW-N1)	(MW-P1) – (MW-N2)	(ZW-P1) – (ZW-N1)	(ZW-P2) – (ZW-N1)	(SST-P1) – (SST-N1)	(SST-P2) – (SST-N1)	PDSI	PC1 SWE	GCV
1	0	0	0	0	0	0	0	0	0	0	1	2.07
2	0	0	0	0	0	0	0	0	0	1	1	2.14
2	0	0	0	1	0	0	0	0	0	0	1	2.15
2	0	0	0	0	1	0	0	0	0	0	1	2.32
3	0	0	0	1	0	0	0	0	0	1	1	2.34
3	0	0	0	0	1	0	0	0	0	1	1	2.51

[a]Presence and absence of predictors are indicated by "1" and "0," respectively.

forecasting flow extremes even on 1 January. This is very useful to the water managers as the extreme years place undue stress on the system operations.

[53] Categorical skill scores for several thresholds (5th percentile through 95th percentile) for the Brier skill score (BSS) [*Wilks*, 1995] were also computed and similar performance of the multimodel ensembles was observed.

[54] To test the utility of large-scale climate information, multimodel ensembles forecasts were generated at all lead times using just the large-scale climate predictors. The skills from this along with those from including the SWE (of course, this is only for forecasts issued from 1 February onward) are shown in Table 4. It can be seen that including large-scale climate information and SWE in the multimodel ensemble framework provides higher skill than just the climate predictors. Rank histograms [*Hamill*, 2001] of the multimodel ensembles (figure not shown) revealed a flattened (or uniform) distribution indicating that the uncertainty estimates are reliable in comparison to the single best model counterparts. As mentioned earlier, one of the advantages of the framework is to provide ensemble forecast at all the sites capturing the spatial

dependence. The average spatial correlation of the ensemble streamflow forecasts was found to be similar to that of the observations.

7. Summary and Discussions

[55] We presented a multimodel ensemble framework to forecast streamflows at several locations simultaneously. The framework has four main steps: (1) principal component analysis is performed on the spatial streamflows to identify the dominant modes of variability; (2) large-scale ocean-atmospheric predictors are identified for the dominant modes; (3) objective criterion, generalized cross validation (GCV) is used to select a suite of candidate models; and (4) ensembles of forecast of the dominant modes and consequently the spatial flows are issued from the candidate models. The forecast model is based on the LWP approach that is data driven. Application of this framework to the GRB showed skilful long-lead forecasts. Furthermore, the multimodel ensemble forecast including large-scale climate features and SWE rendered better performance in terms of skill and reliability of uncertainty estimates than the single
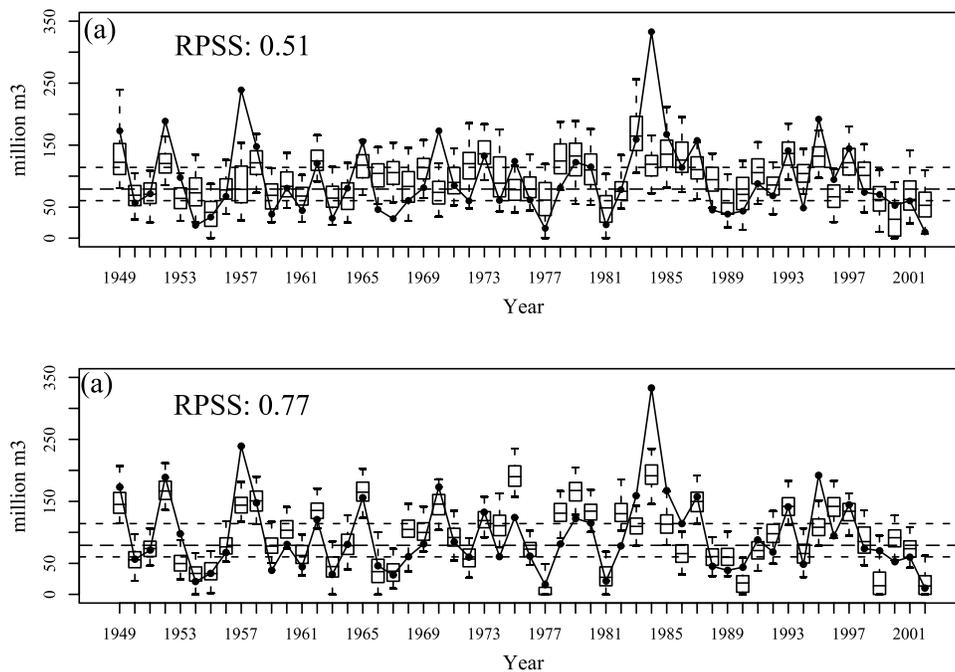
Figure 7. Box plots of spring streamflow (million m^3) at Tomichi issued on (a) 1 January and (b) 1 April. The dashed horizontal lines represent the 33rd, 50th, and 66th percentiles of the historical flow data, which are shown as points connected by the solid line.
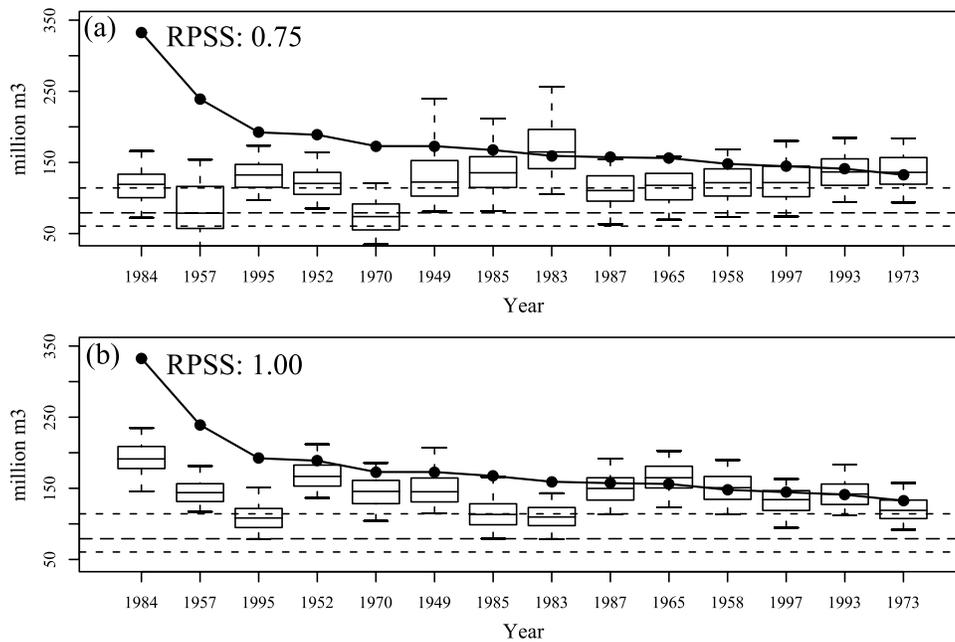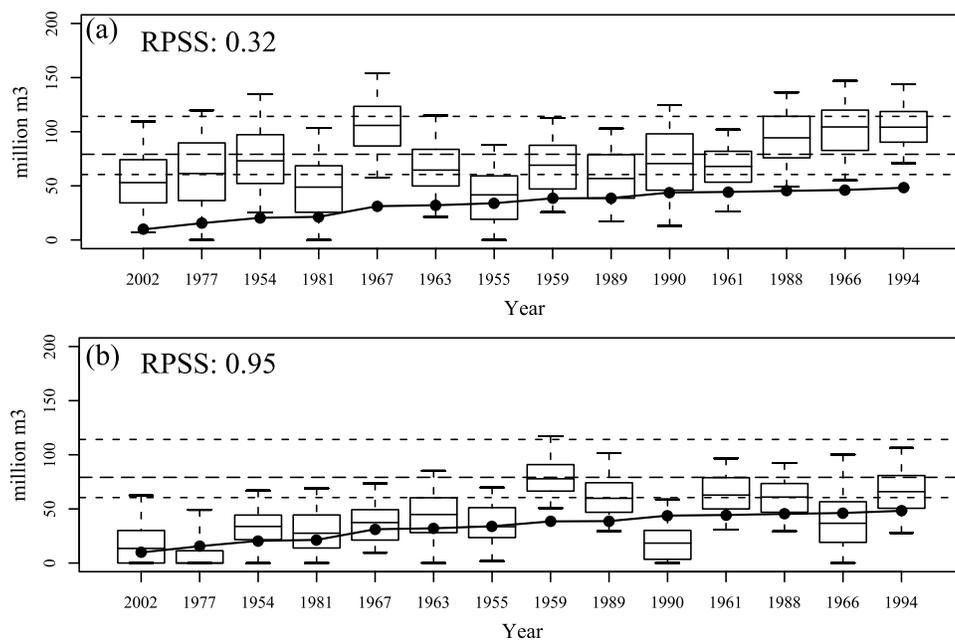
Figure 8. Same as Figure 7 but for forecast of wet years issued on (a) 1 January and (b) 1 April.

best model counterpart or a model with only the SWE information.

[56] We also observed an interesting land-surface effect on the streamflows: when the fall season is dry and following winter is wet, then the spring streamflows tend to be relatively lower compared to what would be expected from a wet winter. Therefore the soil moisture information of the preceding fall was included as a potential predictor, which improved the forecasts. This is corroborated by the presence of PDSI in several of the leading candidate models (see Table 3).

[57] Other factors such as vegetation feedbacks on the spring melt [e.g., *Wang et al.*, 2006] together with soil moisture effects (i.e., quantified by the PDSI), snowpack ripening, spring (daily) wind patterns, air temperature, relative humidity, melt patterns in association with topography and shading factors [*Lundquist and Flint*, 2006], and cloud cover need to be investigated to better understand the streamflow mechanism and incorporated in the framework to potentially improve the forecast skill. Objective methods for selecting the multimodels, combining their ensembles and dealing with multicollinearity of the predictors are some of the aspects of the modeling approach that need detailed investigation. Other nonlinear time series approaches that reconstruct the dynamics of the underlying process from observations for prediction [see e.g., *Porporato and Ridolfi*,

Figure 9. Same as Figure 7 but for forecast of dry years issued on (a) 1 January and (b) 1 April.

Table 4. Ranked Probability Skill Scores of Forecasts at Different Lead Times Using Climate and PDSI Predictors and Using Climate, PDSI, and SWE Information

River	Climate + PDSI					Climate + PDSI + SWE		
	1 Dec	1 Jan	1 Feb	1 Mar	1 Apr	1 Feb	1 Mar	1 Apr
Taylor								
All years	0.50	0.42	0.47	0.46	0.60	0.55	0.61	0.64
Wet years	0.79	0.91	0.74	0.93	0.97	0.94	0.96	0.98
Dry years	0.24	0.39	0.41	0.51	0.79	0.46	0.49	0.70
East								
All years	0.34	0.44	0.38	0.56	0.64	0.60	0.52	0.71
Wet years	0.42	0.79	0.56	0.82	0.86	0.74	0.90	0.97
Dry years	0.38	0.31	0.45	0.51	0.80	0.61	0.40	0.86
Tomichi								
All years	0.47	0.51	0.32	0.57	0.33	0.66	0.74	0.77
Wet years	0.52	0.75	0.39	0.83	0.92	0.88	0.92	1.00
Dry years	0.36	0.32	0.37	0.45	0.83	0.54	0.61	0.95
Lake Fork								
All years	0.29	0.43	0.20	0.35	0.40	0.39	0.48	0.71
Wet years	0.07	0.46	−0.02	0.42	0.58	0.39	0.53	0.96
Dry years	0.27	0.15	0.24	0.30	0.62	0.56	0.57	0.89
North Fork								
All Years	0.36	0.46	0.40	0.56	0.63	0.46	0.45	0.78
Wet years	0.52	0.81	0.66	0.81	0.90	0.64	0.82	0.97
Dry years	0.21	0.21	0.38	0.42	0.82	0.50	0.33	0.65
Uncompahgre								
All years	0.43	0.40	0.33	0.35	0.33	0.61	0.66	0.66
Wet years	0.40	0.74	0.46	0.79	0.84	0.84	0.81	0.98
Dry years	0.21	0.30	0.27	0.30	0.63	0.45	0.62	0.88

1997, 2001; *Regonda et al.*, 2005b; *Sivakumar et al.*, 2002; *Tamea et al.*, 2005] also offer interesting alternatives for consideration, although, they require a large number of data observations.

[58] Having a skillful forecast of the upcoming spring flows during early winter when the snow information is only partial is of significant importance to water managers in their planning and operation. Our preliminary results from driving a decision support model in the basin with these ensemble forecasts indicate similar skills in the decision variables. This is very promising and is being fully investigated further.

References

Bates, J. M., and C. W. J. Granger (1969), The combination of forecasts, *Oper. Res. Q.*, 20, 451–468.

Bjerknes, J. (1969), Atmospheric teleconnections from the equatorial Pacific, *Mon. Weather Rev.*, 97, 163–172.

Brown, D. P., and A. C. Comrie (2004), A winter precipitation "dipole" in the western United States associated with multidecadal ENSO variability, *Geophys. Res. Lett.*, 31, L09203, doi:10.1029/2003GL018726.

Cayan, D. R. (1996), Climate variability and snow pack in the western United States, *J. Clim.*, 9, 928–948.

Cayan, D. R., and D. H. Peterson (1989), The influence of North Pacific atmospheric circulation on streamflow in the west, in *Aspects of Climate Variability in the Pacific and the Western Americas*, *Geophys. Monogr. Ser.*, vol. 55, edited by D. H. Peterson, pp. 365–374, AGU, Washington, D. C.

Cayan, D. R., and R. H. Webb (1992), El Niño/Southern Oscillation and streamflow in the western United States, in *El Nino: Historical and Paleoclimatic Aspects of the Southern Oscillation*, edited by H. F. Diaz and V. Markgraf, pp. 29–68, Cambridge Univ. Press, New York.

Cayan, D. R., M. D. Dettinger, H. F. Diaz, and N. Graham (1998), Decadal variability of precipitation over western North America, *J. Clim.*, 11, 3148–3166.

Cayan, D. R., S. A. Kammerdiener, M. D. Dettinger, J. M. Caprio, and D. H. Peterson (2001), Changes in the onset of spring in the western United States, *Bull. Am. Meteorol. Soc.*, 82, 399–416.

Changnon, D., T. B. McKee, and N. J. Doesken (1993), Annual snowpack patterns across the Rockies: Long-term trends and associated 500-mb synoptic patterns, *Mon. Weather Rev.*, 121, 633–647.

Clark, M. P., M. C. Serreze, and G. J. McCabe (2001), Historical effects of El Nino and La Nina events on the seasonal evolution of the montane snowpack in the Columbia and Colorado River Basins, *Water Resour. Res.*, 37, 741–757.

Clemen, R. T. (1989), Combining forecasts: A review and annotated bibliography, *Int. J. Forecast.*, 5, 559–583.

Craven, P., and G. Whaba (1979), Optimal smoothing of noisy data with spline functions, *Numer. Math.*, 31, 377–403.

Dai, A., K. E. Trenberth, and T. Qian (2004), A global dataset of Palmer drought severity index for 1870–2002: Relationship with soil moisture and effects of surface warming, *J. Hydrometeorol.*, 5, 1117–1130.

Dettinger, M. D., D. R. Cayan, H. F. Diaz, and D. Meko (1998), North-south precipitation patterns in western North America on interannual-to-decadal time scales, *J. Clim.*, 11, 3095–3111.

Dracup, J. A., and E. Kahya (1994), The relationships between U.S. streamflow and La Nina events, *Water. Resour. Res.*, 30, 2133–2142.

Gershunov, A., and T. P. Barnett (1998), ENSO influence on intraseasonal extreme rainfall and temperature frequencies in the contiguous United States: Observations and model results, *J. Clim.*, 11, 1575–1586.

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

Hagedorn, R., F. J. Doblas-Reyes, and T. N. Palmer (2005), The rationale behind the success of multi-model ensembles in seasonal forecasting. Part I: Basic concept, *Tellus, Ser. A*, 57, 219–233.

Hamill, T. M. (2001), Interpretation of rank histograms for verifying ensemble forecasts, *Mon. Weather Rev.*, 129, 550–560.

Hamlet, A. F., and D. P. Lettenmaier (1999), Effects of climate change on hydrology and water resources in the Columbia River basin, *J. Am. Water Resour. Assoc.*, 35, 1597–1623.

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

Helsel, D. R., and R. M. Hirsch (1995), *Statistical Methods in Water Resources*, Elsevier, New York.

Hidalgo, H. G., and J. A. Dracup (2003), ENSO and PDO effects on hydroclimatic variations of the Upper Colorado River Basin, *J. Hydrometeorol.*, 4, 5–23.

Higgins, R. W., A. Leetmaa, and V. E. Kousky (2002), Relationships between climate variability and winter temperature extremes in the United States, *J. Clim.*, 15, 1555–1572.

Higgins, R. W., H.-K. Kim, and D. Unger (2004), Long-lead seasonal temperature and precipitation prediction using tropical pacific SST consolidation forecasts, *J. Clim.*, 17, 3398–3414.

Hoerling, M. P., A. Kumar, and M. Zhong (1997), El Nino, La Nina and the nonlinearity of their teleconnections, *J. Clim.*, 10, 1769–1786.

Hoeting, J. A., D. M. Madigan, A. E. Raftery, and C. T. Volinsky (1999), Bayesian model averaging: A tutorial (with discussion), *Stat. Sci.*, 14, 382–401.

Horel, J. D., and J. M. Wallace (1981), Planetary scale atmospheric phenomena associated with the Southern Oscillation, *Mon. Weather Rev.*, 109, 813–829.

Hou, D., E. Kalnay, and K. K. Droegemeier (2001), Objective verification of the SAMEX'98 ensemble forecast, *Mon. Weather Rev.*, 129, 73–91.

Hwang, Y. (2005), Impact of input uncertainty in ensemble streamflow generation, Ph.D. thesis, Univ. of Colo., Boulder.

Kahya, E., and J. A. Dracup (1993), U.S. streamflow patterns in relation to the El Niño/Southern Oscillation, *Water. Resour. Res.*, 29, 2491–2504.

Kahya, E., and J. A. Dracup (1994), The influences of type 1 El Niño and La Niña events on streamflows in the Pacific Southwest of the United States, *J. Clim.*, 7, 965–976.

Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, *77*, 437–471.

Klein, W. H. (1963), Specifications of precipitation from the 700-millibar circulation, *Mon. Weather Rev.*, *91*, 527–536.

Klein, W. H., and H. J. Bloom (1987), Specification of monthly precipitation over the United States from the surrounding 700 mb height field, *Mon. Weather Rev.*, *115*, 527–536.

Klein, W. H., C. W. Crockett, and J. F. Andrews (1965), Objective prediction of daily precipitation and cloudiness, *J. Geophys. Res.*, *70*, 801–813.

Krishnamurti, T. N., C. M. Kishtawal, T. E. LaRow, D. R. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran (1999), Improved weather and seasonal climate forecasts from multi-model superensemble, *Science*, *285*, 1548–1550.

Krishnamurti, T. N., C. M. Kishtawal, Z. Zhang, T. E. LaRow, D. R. Bachiochi, C. E. Williford, S. Gadgil, and S. Surendran (2000), Multi-model ensemble forecasts for weather and seasonal climate, *J. Clim.*, *13*, 4196–4216.

Lall, U. (1995), Recent advances in nonparametric function estimation: Hydraulic applications, *U.S. Natl. Rep. Int. Union Geod. Geophys. 1991–1995, Rev. Geophys.*, *33*, 1093–1102.

Loader, C. (1999), *Local Regression and Likelihood*, Springer, New York.

Lundquist, J., and A. Flint (2006), Onset of snowmelt and streamflow in 2004 in the western United States: How shading may affect spring streamflow timing in a warmer world, *J. Hydrometeorol.*, in press.

Madigan, D., and A. E. Raftery (1994), Model selection and accounting for model uncertainty in graphical models using Occam's window, *J. Am. Stat. Assoc.*, *89*, 1535–1546.

Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis (1997), A Pacific interdecadal climate oscillation with impacts on salmon production, *Bull. Am. Meteorol. Soc.*, *78*, 1069–1079.

Maurer, E. P., D. P. Lettenmaier, and N. J. Mantua (2004), Variability and potential sources of predictability of North American runoff, *Water Resour. Res.*, *40*, W09306, doi:10.1029/2003WR002789.

McCabe, G. J. (1994), Relationships between atmospheric circulation and snowpack in the Gunnison River Basin, Colorado, *J. Hydrol.*, *157*, 157–175.

McCabe, G. J. (1996), Effects of winter atmospheric circulation on temporal and spatial variability in annual streamflow in the western United States, *J. Hydrol. Sci.*, *41*, 873–887.

McCabe, G. J., and M. D. Dettinger (1999), Decadal variations in the strength of ENSO teleconnections with precipitation in the western United States, *Int. J. Climatol.*, *19*, 1399–1410.

McCabe, G. J., and M. D. Dettinger (2002), Primary modes and predictability of year-to-year snowpack variations in the western United States from teleconnections with Pacific Ocean climate, *J. Hydrometeorol.*, *3*, 13–25.

McCabe, G. J., and D. R. Legates (1995), Relationships between 700 hPa height anomalies and April 1 snowpack accumulations in the western USA, *Int. J. Climatol.*, *15*, 517–530.

Newman, M., G. P. Compo, and M. A. Alexander (2003), ENSO-forced variability of the Pacific Decadal Oscillation, *J. Clim.*, *16*, 3853–3857.

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

Piechota, T. C., H. Hidalgo, and J. Dracup (2001), Streamflow variability and reconstruction for the Colorado River Basin, paper presented at EWRI World Water and Environmental Resources Congress, Am. Soc. of Civ. Eng., Orlando, Fla., 20–24 May.

Porporato, A., and L. Ridolfi (1997), Nonlinear analysis of river flow time sequences, *Water Resour. Res.*, *33*, 1353–1368.

Porporato, A., and L. Ridolfi (2001), Multivariate nonlinear prediction of river flows, *J. Hydrol.*, *248*, 109–122.

Prairie, J., B. Rajagopalan, T. Fulp, and E. Zagona (2005a), Statistical nonparametric model for natural salt estimation, *J. Environ. Eng.*, *131*, 130–138.

Prairie, J., B. Rajagopalan, T. Fulp, and E. Zagona (2005b), A modified K-NN model for generating stochastic natural streamflows, *J. Hydrol. Eng.*, *11*, 371–378.

Raftery, A. E., D. Madigan, and J. A. Hoeting (1997), Bayesian model averaging for regression models, *J. Am. Stat. Assoc.*, *92*, 179–191.

Rajagopalan, B., U. Lall, and S. Zebiak (2002), Optimal categorical climate forecasts through multiple GCM ensemble combination and regularization, *Mon. Weather Rev.*, *130*, 1792–1811.

Rajagopalan, B., K. Grantz, S. K. Regonda, M. Clark, and E. Zagona (2005), Ensemble streamflow forecasting: Methods and applications, in *Advances in Water Science Methodologies*, edited by U. Aswathanarayana, chap. 7, Taylor and Francis, Philadelphia, Pa.

Rao, C. R., and H. Toutenburg (1999), *Linear Models: Least Squares and Alternatives*, Springer, New York.

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

Regonda, S., B. Rajagopalan, M. Clark, and J. Pitlick (2005a), Seasonal cycle shifts in hydroclimatology over the western US, *J. Clim.*, *18*, 372–384.

Regonda, S., B. Rajagopalan, U. Lall, M. Clark, and Y. Moon (2005b), Local polynomial method for ensemble forecast of time series, *Nonlinear Processes Geophys.*, *12*, 397–406.

Reid, D. J. (1968), Combining three estimates of gross domestic product, *Economica*, *35*, 431–444.

Robertson, A. W., and M. Ghil (1999), Large-scale weather regimes and local climate over the western United States, *J. Clim.*, *12*, 1796–1813.

Singhrattna, N., B. Rajagopalan, M. Clark, and K. Krishna Kumar (2005), Forecasting Thailand summer monsoon rainfall, *Int. J. Climatol.*, *25*, 649–664.

Sivakumar, B., A. W. Jayawardena, and M. K. G. T.Fernando (2002), River flow forecasting: Use of phase-space reconstruction and artificial neural networks approaches, *J. Hydrol.*, *265*, 225–245.

Slack, J. R., and J. M. Landwehr (1992), Hydro-Climatic Data Network: A U.S. Geological Survey streamflow data set for the United States for the study of climate variations, 1974–1988, *U.S. Geol. Surv. Open File Rep., 92-129*, 193 pp.

Stewart, I. T., D. R. Cayan, and M. D. Dettinger (2005), Changes toward earlier streamflow timing across western North America, *J. Clim.*, *18*, 1136–1155.

Tamea, S., F. Laio, and L. Ridolfi (2005), Probabilistic nonlinear prediction of river flows, *Water Resour. Res.*, *41*, W09421, doi:10.1029/2005WR004136.

Ugland, R. C., B. J. Cochran, M. M. Hiner, R. G. Kretschman, E. A. Wilson, and J. D. Bennett (1990), Water resources data for Colorado, water year 1990, volume 2, Colorado River Basin, *U.S. Geol. Surv. Water Data Rep., CO-90-2*, 344 pp.

Von Storch, H., and F. W. Zwiers (1999), *Statistical Analysis in Climate Research*, Cambridge Univ. Press, New York.

Walpole, R. E., R. H. Myers, S. L. Myers, K. Ye, and K. Yee (2002), *Probability and Statistics for Engineers and Scientists*, Prentice-Hall, Upper Saddle River, N. J.

Wang, W., B. T. Anderson, N. Phillips, R. K. Kaufmann, C. Potter, and R. B. Myneni (2006), Feedbacks of vegetation on summertime climate variability over the North American grasslands: 1. Statistical analysis, *Earth Interact.*, in press.

Weare, B. C., and M. A. Hoeschele (1983), Specification of monthly precipitation in the western United States from monthly mean circulation, *J. Clim. Appl. Meteorol.*, *22*, 1000–1007.

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

Yarnal, B., and H. F. Diaz (1986), Relationships between extremes of the Southern Oscillation and the winter climate of the Anglo-American Pacific coast, *J. Climatol.*, *6*, 197–219.

————————————————

M. Clark, NIWA, PO Box 8602, Christchurch, New Zealand.

B. Rajagopalan and S. K. Regonda, Department of Civil, Environmental and Architectural Engineering, University of Colorado, Campus Box 421, Boulder, CO 80309-0421, USA. (satish-kumar.regonda@colorado.edu)

E. Zagona, CADSWES, University of Colorado, Boulder, CO 80309-0488, USA.