A multisite seasonal ensemble streamflow forecasting technique

Cameron Bracken,[1,2] Balaji Rajagopalan,[3,4] and James Prairie[5]

Received 10 March 2009; revised 20 October 2009; accepted 2 November 2009; published 30 March 2010.

[1] We present a technique for providing seasonal ensemble streamflow forecasts at several locations simultaneously on a river network. The framework is an integration of two recent approaches: the nonparametric multimodel ensemble forecast technique and the nonparametric space-time disaggregation technique. The four main components of the proposed framework are as follows: (1) an index gauge streamflow is constructed as the sum of flows at all the desired spatial locations; (2) potential predictors of the spring season (April–July) streamflow at this index gauge are identified from the large-scale ocean-atmosphere-land system, including snow water equivalent; (3) the multimodel ensemble forecast approach is used to generate the ensemble flow forecast at the index gauge; and (4) the ensembles are disaggregated using a nonparametric space-time disaggregation technique resulting in forecast ensembles at the desired locations and for all the months within the season. We demonstrate the utility of this technique in skillful forecast of spring seasonal streamflows at four locations in the Upper Colorado River Basin at different lead times. Where applicable, we compare the forecasts to the Colorado Basin River Forecast Center's Ensemble Streamflow Prediction (ESP) and the National Resource Conservation Service "coordinated" forecast, which is a combination of the ESP, Statistical Water Supply, a principal component regression technique, and modeler knowledge. We find that overall, the proposed method is equally skillful to existing operational models while tending to better predict wet years. The forecasts from this approach can be a valuable input for efficient planning and management of water resources in the basin.

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

1. Introduction and Background

[2] The recent protracted dry period (2000–2008) in the Upper Colorado River Basin (UCRB) has had various impacts on basin hydrology and management. For example, Lake Powell has seen its lowest levels since its filling in 1980 [*Brandon*, 2005]. In particular the recent drought has emphasized the need for accurate streamflow predictions at longer lead times than usual. Accurate forecasts at several spatial locations are also desirable for efficient basin-wide reservoir management.

[3] Nearly 80% of the streamflow in the UCRB and in many of western U.S. river basins is due to snowmelt and as a result streamflow models have long been dominated by this information. In particular, the 1 April snow water equivalent in the basin is a potent predictor of the subsequent spring snowmelt runoff. However, many important planning decisions are made during the winter (November–March) requiring a skillful projection of the spring streamflow at a time when the snow information is incomplete. This poses a challenging problem of providing skillful basin-wide streamflow forecasts at longer lead time.

[4] There is increasing evidence that large-scale climate features in the Pacific have strong influence on the hydroclimatology of western U.S. including the UCRB: on winter snow [*Clark et al.*, 2001; *Cayan*, 1996], surface temperatures [*Redmond and Koch*, 1991; *Higgins et al.*, 2002; *Gershunov and Barnett*, 1998], and streamflow [*Kahya and Dracup*, 1993, 1994; *Dracup and Kahya*, 1994; *Piechota et al.*, 1997; *Maurer et al.*, 2004; *McCabe and Dettinger*, 1999, 2002; *Hidalgo and Dracup*, 2003; *Brown and Comrie*, 2004; *Hunter et al.*, 2006; *Soukup et al.* 2009]. These links enhance the prospects for long lead streamflow forecast. These links were identified and incorporated in a statistical modeling approach to generate skillful forecasts of spring streamflow at long lead times during early winter on several river basins in the western U.S., e.g., Truckee and Carson river basins [*Grantz et al.*, 2005], Gunnison river basin [*Regonda et al.*, 2006], Columbia River [*Hamlet and Lettenmaier*, 1999; *Clark et al.*, 2001] and the Yakima river basin [*Opitz-Stapleton et al.*, 2007].

[5] The Colorado Basin River Forecast Center (CBRFC) and the Natural Resource Conservation Service (NRCS) are jointly charged with the task of predicting streamflows in

[1]Department of Environmental Resources Engineering, Humboldt State University, Arcata, California, USA.
[2]Now at Department of Civil Environmental and Architectural Engineering, University of Colorado at Boulder, Boulder, Colorado, USA.
[3]Department of Civil Environmental and Architectural Engineering, University of Colorado at Boulder, Boulder, Colorado, USA.
[4]Cooperative Institute for Research in Environmental Sciences, University of Colorado at Boulder, Boulder, Colorado, USA.
[5]Bureau of Reclamation, University of Colorado at Boulder, Boulder, Colorado, USA.

0043-1397/10/2009WR007965

the CRB. The current CBRFC models include Statistical Water Supply (SWS) and Ensemble Streamflow Prediction (ESP) [*Brandon*, 2005]. The SWS is a regression based method that relates observed data (precipitation, snow water equivalent, monthly flow volume, and climate indices) with future streamflow. The ESP is an empirically based method that includes antecedent streamflow, soil moisture, reservoir information, snowpack states and climate data in forecasting streamflows. The ESP produces ensemble forecasts from historical data based on current conditions thus ensemble size and scope is limited to that of the historical data; for example if 20 years of data are available then only 20 ensemble members can be generated. The NRCS model uses a principle components regression technique that includes snowpack states, fall and spring precipitation, base flow, and climate indices (e.g., Southern Oscillation Index) [*Pagano and Garen*, 2004]. Together, the CBRFC and NRCS issue a "coordinated forecast" based on their models and modeler knowledge. The coordinated forecast is used in the Bureau of Reclamation's (BOR) "24-Month Study." The lack of a rich variety in the ensemble forecasts and consequently the uncertainty estimation are some of the main shortcomings of this approach.

[6] Water resources management in a basin requires streamflow forecasts at several locations on the river network. To address this, *Regonda et al.* [2006] developed a multimodel ensemble streamflow technique that includes a principal component analysis on the basin streamflow. This method performs an orthogonal transformation of the streamflow at multiple locations in the basin. Predictors of the leading principal component in the land-ocean-atmosphere system are identified and a local polynomial based statistical method is used to generate an ensemble forecast of the leading principal component. These are lastly back transformed to obtain ensemble forecasts at multiple locations in the basin. The authors applied this technique for streamflow forecasting at six locations in the Gunnison River basin (a tributary of the Colorado River) and demonstrated significant skill at longer leads. While this approach is quite good it suffers from two key drawbacks. The first drawback is that the forecasted flows at the spatial locations do not satisfy the summability criteria; that is, the flow at any location should be the sum of flows at locations above it. This is important for water budgeting and essential for decision making when these flows are used to drive a decision support system). The second drawback is that the streamflow forecasts are for the spring season total, but the decision making process requires streamflow for all the months within the season that satisfy the summability criteria temporally (the monthly forecasts should add up to the seasonal) and spatially (as mentioned in first drawback.

[7] To overcome the drawbacks of the coordinated forecasts and the multisite forecasting method of *Regonda et al.* [2006], we propose a framework that integrates the multimodel ensemble forecast technique of *Regonda et al.* [2006] using large-scale climate information with a nonparametric space-time disaggregation technique to address the summability. The four components of the proposed framework are as follows. (1) an index gauge streamflow is constructed as the sum of flows at all the desired spatial locations; (2) potential predictors of the spring season (April–July) streamflow at this index gauge are identified from the large-scale ocean-atmosphere-land system, including snow water equivalent; (3) a multimodel ensemble forecast approach [*Regonda et al.*, 2006] is used to generate the ensemble flow forecast at the index gauge; and (4) the ensembles are disaggregated using a nonparametric space-time disaggregation technique [*Prairie et al.*, 2007] resulting in forecast ensembles at the desired locations and for all the months within the season. By using an index gauge in this manner we only need to develop predictors for a single "site." Not only is this approach parsimonious but spatial and temporal correlations are also preserved [*Prairie et al.*, 2007]. We demonstrate this framework to forecast monthly streamflow at four key sites in the UCRB. The paper is organized as follows. Preliminary information on the study area and data sets used are first described. Next, the proposed framework is explained, including the algorithmic details of two main components, the multimodel ensemble (MME) forecast and the disaggregation procedure. Results from application to the UCRB are then presented, followed by a summary and discussion.

2. Study Area

[8] The CRB includes parts of seven states in the western United States with an area of approximately 650,000 km^2. The basin includes widely varying topography with elevations ranging from 61 to 4328 m. Most of the flow in the basin is a result of snowmelt from the UCRB, while most of the water use occurs in the semiarid and desert regions of Lower Colorado River Basin (LCRB). To demonstrate the proposed forecast framework, we chose four key locations on the UCRB network shown in Figure 1: Colorado River near Cisco, Utah (Cisco); Green River at Green River, Utah (GRUT); San Juan River near Bluff, Utah (Bluff) and, Colorado River at Lees Ferry, Arizona (Lees Ferry). Lees Ferry gauge, 26 km below Glen Canyon Dam, is the key gauge through which 90% of the Colorado River flow passes through and is approximately a mile upstream of Lee Ferry the official demarcation point of the Upper and Lower Basins for operational and management purposes.

3. Data

3.1. Streamflow Data and Index Gauge

[9] The natural streamflow data for the Colorado River Basin are developed by the Bureau of Reclamation (Reclamation) and updated regularly. (The natural flow data and additional reports describing these data are available at http://www.usbr.gov/lc/region/g4000/NaturalFlow/index.html.) Naturalized streamflows are computed by correcting for anthropogenic impacts (i.e., reservoir regulation, consumptive water use, etc.) from the recorded historic flows. *Prairie and Callejo* [2005] present a detailed description of methods and data used for the computation of natural flows in the Colorado River Basin. We used the monthly natural streamflow at the four locations for the period 1949–2005. The flows at the four sites for each month are added to create the "index gauge" monthly streamflow. Almost all of the annual flow at these locations and in the basin occurs during the spring (April–July) from spring melt of winter snow, much like other river basins in the western U.S.

3.2. Large-Scale Climate Data

[10] 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, the variables used were 500 mb geo-

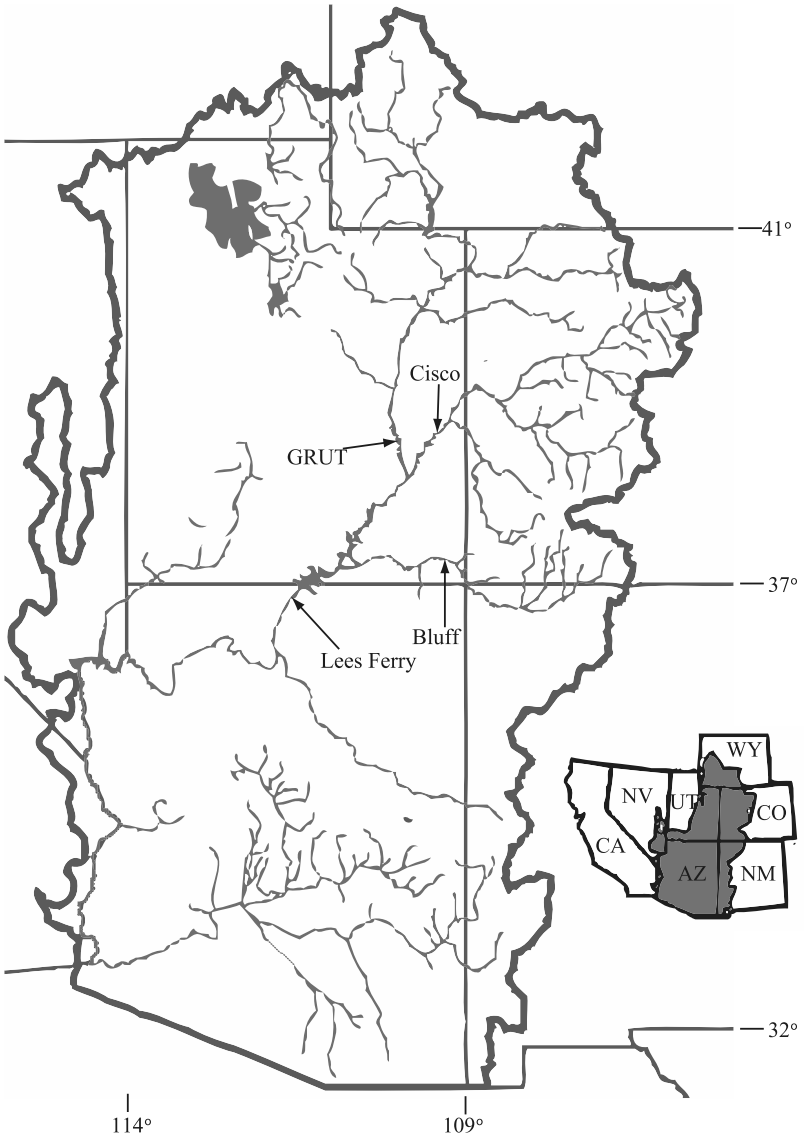

Figure 1. Study area.

potential height (GPH), zonal (ZW) and meridonal winds (MW) and sea surface temperature (SST). These variables are provided on a 2.2° by 2.2° grid spanning the globe from the NCEP-NCAR reanalysis project [*Kalnay et al.*, 1996] for the period 1949 to present.

3.3. SWE Data

[11] Snow water equivalent (SWE) data, which quantifies the amount of water present in a snowpack, is obtained from snow course surveys by the Natural Resources Conservation Service (NRCS) from their Web site (http://www.wcc.nrcs.usda.gov/snow). Data was obtained at 10 sites in the UCRB (for the years 1948–2005) for each monthly (1 February, 1 March and 1 April) SWE. Principal component analysis was performed for each month and the leading principal component (which explains most of the variance) was used as the "basin SWE" for each month. This is similar to using a spatial average but more robust; in this case, we found the leading principal component and the spatial average time series to be highly correlated.

3.4. PDSI Data

[12] Antecedent summer and fall season land conditions can play an important role in the variability of the following spring streamflow. This was shown by *Regonda et al.* [2006] for the Gunnison River Basin where they found a significant reduction in spring streamflow due to infiltration, relative to the snowpack, in years that succeed a dry summer and fall season and vice versa. Thus including this in the forecast can improve the skills, especially in such anomalous years. While soil moisture would be the best variable to capture the antecedent land surface conditions, PDSI is shown to be a good surrogate [*Dai et al.*, 2004]. Basin averaged PDSI data from October (for the years of 1948–2004) was used as a surrogate for antecedent moisture conditions.

4. Proposed Integrated Framework

[13] The proposed framework applies three steps to the index gauge: (1) predictor identification, (2) multimodel ensemble forecast, and (3) nonparametric spatial and temporal disaggregation resulting in forecast ensembles at the

desired locations and for all the months within the season. The components of each step are described below.

4.1. Predictor Suite for Index Gauge Seasonal Streamflow

[14] The index gauge spring streamflow is correlated with global ocean, atmosphere and land variables (i.e., GPH, SST, ZW, MW and PDSI from preceding seasons). Regions of high correlations are identified for each variable and the spatial average of data points in these regions are computed to create a suite of potential predictors. The online tool developed by Physical Sciences Division, National Oceanic and Atmospheric Administration (NOAA) is used for this purpose (see http://www.cdc.noaa.gov/Correlation/).

4.2. Multimodel Selection for Index Gauge Seasonal Flow Forecast

[15] The MME methodology consists of two distinct steps: (1) selection of the multimodels, each with its own subset of predictors, and (2) combining forecasts from the individual models. The general form of the forecast model is

$$Y = f(\mathbf{x}) + \varepsilon, \tag{1}$$

where Y is the index gauge spring season streamflow, $\mathbf{x}$ is a suite of predictors and ε is the residual that is assumed to be normally distributed with mean 0 and variance $\sigma^2 - N(0,\sigma)$.

[16] If f is a global function based on the entire data and linear, then the model is a traditional linear regression. The theory behind this approach, procedures for parameter estimation and hypothesis testing are very well developed [e.g., *Helsel and Hirsch*, 1995; *Rao and Toutenburg*, 1999] and widely used. However, they do have some drawbacks including (1) the assumption of a normal 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. 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 a normal distribution. All of this can make the process unwieldy. Thus, a more flexible framework would be desirable.

[17] Local estimation methods (also known as nonparametric methods) provide an attractive alternative. There are several approaches for local functional estimation applied to hydrologic problems, see *Lall* [1995]. Of these, the locally weighted polynomial (LWP) regression is simple and robust, and has been used in a variety of hydrologic and hydroclimate applications with good results, e.g., 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.*, 2005], 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 as *Regonda et al.* [2006] also did.

[18] The "best model" in this form is described by the size of the neighborhood, the order of the polynomial and the subset of predictor variables. This is identified using objective criteria. A brief description of the implementation steps is presented below, which is largely abstracted from *Regonda et al.* [2006]. For details on local polynomial estimation we refer the reader to *Loader* [1999]. The implementation steps are as follows.

[19] 1. Select a subset of predictors.

[20] 2. Select a degree of polynomial to fit, typically, $p = 1$ (linear fit) or 2 (quadratic) is quite adequate.

[21] 3. Select a size of the neighborhood, $K = \alpha * N$ (where $\alpha = (0,1)$ and N = number of observations).

[22] 4. For a point $\mathbf{x}$ identify the K nearest neighbors in the data; Euclidean distance is typically used, other metrics can be used such as Mahalanobis distance [*Yates et al.*, 2003].

[23] 5. Weighted least squares estimation is employed to fit a polynomial of order p to the K nearest neighbors. Use the fit to obtain the estimate, $\hat{Y}$.

[24] 6. Repeat steps 3–4 for all the data points.

[25] 7. Compute the objective criteria, generalized cross-validation estimation

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

where e_i is the model residual $(Y_i - \hat{Y})$ for the ith data point, N is the number of data points, q is the number of parameters in the local polynomial model. Generalized cross validation (GCV) provides a good estimate of predictive risk of the model, unlike other statistics, which are goodness of fit measures [*Craven and Wahba*, 1978].

[26] 8. Repeat steps 1 through 7, thus, obtaining the GCV values for a suite of predictor subset, K and p combination. The model with the least GCV score is selected as the "best" model. This is akin to the stepwise regression method in a traditional linear regression context [e.g., *Rao and Toutenburg*, 1999; *Walpole et al.*, 2002] wherein an objective function such as Mallow's C_p statistic, adjusted R^2, Akaike information criteria (AIC), or an F test is calculated from the fitted model for several predictor combinations. The GCV based model selection is preferable since it provides a better estimate of predictive risk as mentioned above.

[27] 9. All combinations with GCV values within a prescribed threshold (user specified as typically within 5% of the least GCV value) are selected as admissible constituting the pool of candidate models (i.e., multimodels). Recent studies show that multimodel ensemble forecasts which include multiple "best" models tend to perform much better than a single model forecast [*Hagedorn et al.*, 2005; *Krishnamurti et al.*, 1999, 2000; *Rajagopalan et al.*, 2002]. Combinations with predictor variables significantly correlated among each other (i.e., multicollinear) are removed from the multimodel pool, as they lead to over fitting and poor predictive skill [*Wadsworth*, 1990], thus resulting in a set of multimodels each with different predictor subsets.

[28] Note that if $\alpha = p = 1$ and if an ordinary least squares estimation is used to fit the model, it collapses to a traditional linear regression. Therefore, LWP can be viewed as a more flexible approach that includes traditional linear regression as a subset.

4.3. Multimodel Ensemble Forecast of Index Gauge Seasonal Streamflow

[29] The multimodel ensemble forecast generation is described below, again following *Regonda et al.* [2006].

Suppose we desire a multimodel ensemble forecast for a point $\mathbf{x}_j$, the steps are as follows.

[30] 1. From one of the multimodels identified in section 4.2, obtain the model prediction $\hat{Y}_i$ and the estimate of the error variance ($\sigma_{e_i}^2$) [*Loader*, 1999]. Ensembles z are then generated by sampling a given number (we chose 250) of random normal deviates and adding these to the prediction $\hat{Y}_i$:

$$z_{i,j} = \hat{Y}_i(\mathbf{x}_j) + N(0, \sigma_{e_j}), \tag{3}$$

where $N(0,\sigma_{e_j})$ is a normally distributed random variable with mean 0 and standard deviation σ_{e_j}. This approach assumes normality of the residuals, a standard assumption from regression theory. For the application here we found the residuals to satisfy this assumption of normality.

[31] 2. Repeat step 1 for all the models in the multimodel suite, thus obtaining an ensemble forecast of size n from each model.

[32] 3. Weight each of the multimodels based on 1/GCV criteria. This way the model with the lowest GCV value is most heavily favored.

[33] 4. Randomly choose a model based on the above set of weights.

[34] 5. Randomly choose one of the n ensemble members.

[35] 6. Repeat steps 4–5 n times to obtain a multimodel ensemble forecast.

4.4. Spatial and Temporal Disaggregation of the Index Gauge Forecast

[36] The seasonal forecast of the index gauge generated above needs to be disaggregated in space (to the four locations) and in time (to the four months of the season) resulting in an ensemble forecast for each month at all the four locations. This is achieved by employing a nonparametric space-time disaggregation technique proposed by *Prairie et al.* [2007] to the seasonal forecast. All the motivational and technical details of the disaggregation method are comprehensively described by *Prairie et al.* [2007]. Below we provide a brief description of its implementation in the current application.

[37] The disaggregation procedure can be thought of as sampling from the conditional probability density function (PDF), $f(\mathbf{x}|z)$, where $\mathbf{x}$ is a d dimensional vector of flows, z is the aggregate flow, but with the (additvity) constraint that the values in $\mathbf{x}$ add up to z. This is achieved by an orthonormal rotation of the data $\mathbf{x}$ to $\mathbf{Y}$ and the simulation is performed in the rotated space and back rotated [*Tarboton et al.*, 1998; *Prairie et al.*, 2007]. The steps are described below for a temporal disaggregation (seasonal to monthly at the index gauge), they are identical for the spatial disaggregation.

[38] 1. The first step is to generate the orthonormal rotation matrix, $\mathbf{R}(d)$ using the Gram- Schmidt algorithm. Note that the $d \times d$ matrix $\mathbf{R}$ is only a function of the dimension d and has the property $\mathbf{R}^T = \mathbf{R}^{-1}$ by definition, where T denotes transpose. This process is described in detail in the appendix of *Tarboton et al.* [1998], the reader is referred there for the details.

[39] 2. The matrix $\mathbf{X}$ of the historical monthly streamflow at gauge I is rotated to $\mathbf{Y}$ by the rotation matrix R as

$$\mathbf{Y} = \mathbf{RX}. \tag{4}$$

Both $\mathbf{X}$ and $\mathbf{Y}$ are of dimension N (number of years) rows by d (=4 months in the season) columns.

[40] The development of the $\mathbf{R}$ matrix is detailed in the appendix of *Tarboton et al.* [1998]. The rotated matrix $\mathbf{Y}$ has its last column $\mathbf{y}_d = \mathbf{z}/\sqrt{d} = \mathbf{z}'$, where $\mathbf{z}$ is the vector of the aggregate flows (i.e., the seasonal totals). If we denote the first $d-1$ columns as $\mathbf{U}$ then

$$\mathbf{Y} = \left[\mathbf{U}, \mathbf{z}'\right]. \tag{5}$$

[41] 3. For a given seasonal flow forecast say z_{sim}, K nearest neighbors are identified from the historical seasonal flow values in $\mathbf{z}$. One of the K nearest neighbor is resampled using a weight function

$$W(k) = \frac{1}{k\sum_{i=1}^{K}\frac{1}{i}}, \tag{6}$$

where $k = 1,2,\ldots,K$. This weight function gives more weight to the nearest neighbors and less to the farthest neighbors. For further discussion on the choice of the weight function readers are referred to *Lall and Sharma* [1996]. The number of nearest neighbors, K is based on the heuristic scheme $K = \sqrt{N}$ where N equals the sample size [*Lall and Sharma*, 1996], following the asymptotic arguments of *Fukunaga* [1990].

[42] 4. Suppose the selected neighbor corresponds to historic year j, the new vector $\mathbf{y}_{sim}$ is constructed as

$$\mathbf{y}_{sim} = \left[u_j, z_{sim}/\sqrt{d}\right]. \tag{7}$$

[43] 5. This is back transformed to the original space as

$$\mathbf{x}_{sim} = \mathbf{R}^T\mathbf{y}_{sim}. \tag{8}$$

The vector $\mathbf{x}_{sim}$ now contains $d = 4$ values, the disaggregated monthly flow of the seasonal total flow, $\mathbf{z}_{sim}$.

[44] The disaggregated monthly flows are subjected to the same procedure to obtain monthly flows at the four locations. This ensures the additivity criteria, for each month the flows at the index gauge are a sum of the flows at the four locations. Applying this to all the multimodel ensemble members results in a multimodel ensemble forecast for each month at the four locations. We also note that the sampling the residuals from a normal distribution coupled with the disaggregation method, provide the ability to generate values outside the range of the historical observations; this was shown by *Regonda et al.* [2006] and *Prairie et al.* [2007].

4.5. Verification and Validation

[45] Since we generate an ensemble forecast (i.e., a 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]. 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

Table 1. Potential Predictors and Their Regions in Longitude and Latitude

Variable	Lead Time	Season	Negative Region	Positive Region	Maximum Correlation
GPH	1 Nov	Oct	40,42N:137,140E	17,19N:178,182E	0.39
ZNW	1 Nov	Oct	40,42N:−108,−106E	30,32N:132,135E	0.43
MDW	1 Nov	Oct	40,42N:124,126E	40,42N:124,126E	0.31
SST	1 Nov	Oct	−17,−15N:−96,−92E	38,40N:170,173E	−0.26
GPH	1 Jan	Oct–Dec	42,45N:−115,−110E	52,58N:−160,−155E	−0.56
ZNW	1 Jan	Oct–Dec	54,55.5N:−119,−114E	32,34N:−110,−106E	−0.58
MDW	1 Jan	Oct–Dec	45,50N:−135,−130E	42,45N:−91,−88E	−0.57
SST	1 Jan	Oct–Dec	28,31N:161,164E	46,49N:−166,−163E	−0.63
GPH	1 Feb	Oct–Jan	60,62N:128,132E	61,64N:−153,−150E	0.24
ZNW	1 Feb	Oct–Jan	72,74N:126,128E	28,31N:−119,−116E	0.42
MDW	1 Feb	Oct–Jan	66,68N:−108,−105E	38,42N:−90,−85E	0.30
SST	1 Feb	Oct–Jan	25,28N:160,163E	46,48N:−167,−164E	−0.64
GPH	1 Apr	Feb–Mar	40,42N:137, 140E	65,70N:160,165E	−0.47
ZNW	1 Apr	Mar	56,58N:172,176E	25,30N:155,160E	0.46
MDW	1 Apr	Mar	10,12N:−177,−174E	56,60N:142,144E	−0.42
SST	1 Apr	Jan–Mar	24.5,26N:158,162E	48,51N:−165,−160E	−0.54

observational vector $(d_1,d_2,\ldots,d_k)$ is obtained for each forecast, where d_k is unity if the observation falls in the kth category and zero otherwise. The ranked probability skill score (RPSS) is defined as follows:

$$\mathrm{RPS} = \sum_{i=1}^{k}\left[\left(\sum_{j=1}^{i} p_j - \sum_{j=1}^{i} d_j\right)^2\right] \tag{9}$$

$$\mathrm{RPSS} = 1 - \frac{\mathrm{RPS(forecast)}}{\mathrm{RPS(climatolgy)}}. \tag{10}$$

In this research, the streamflows are divided into three categories, at the tercile boundaries, 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.

[46] The RPSS ranges from negative infinity to positive unity. 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. Correlation between the median value of the ensemble and the observed (MC) are also computed to test the performance of the median forecast. The forecast skills were computed for the following three types of forecasts.

[47] 1. In the leave-one-out forecast, each year is dropped and the multimodel ensemble for that year is generated from the rest [e.g., *Grantz et al.*, 2005; *Regonda et al.*, 2006; *Singhrattna et al.*, 2005].

[48] 2. Leaving one year out at a time may not stress the model adequately. Here we drop 10% of the observations and forecasts for the dropped years are made using the rest of the observations. The skill scores are computed for the forecasts. This is repeated a number of times, obtaining an ensemble of skill scores. This validation method provides a mechanism for sensitivity analysis.

[49] 3. To be able to compare with forecasts from CBRFC, "retroactive" forecasts have to be performed. In this, a forecast for a given year is made using all the data prior to that year imitating a real time situation.

5. Results

[50] We chose four lead times, 1 November, 1 January, 1 February and 1 April, to predict the spring streamflows. Separate predictors are identified for each lead time and the above methodology is applied to obtain the ensemble forecasts. Next, we present a representative set of results.

5.1. Predictor Identification

[51] The index gauge spring season streamflow was correlated with large-scale ocean, atmosphere and land variables from preceding months. The combination of months that showed strong correlation are February–March GPH, March ZNW, January–March SST and March MDW (see auxiliary material).[1] The correlation maps are consistent with prior findings for the river flows in western U.S. [e.g., *Grantz et al.*, 2005; *Regonda et al.*, 2006]. The regions with high correlations were identified and a spatially averaged time series was computed as potential predictors, these are detailed in Table 1. For SWE the leading principal component (similar to a basin average) was used, this is available from 1 February.

[52] To understand the physical relationship between the large-scale variables and the spring streamflow, composite maps of the 6 wettest and driest years are developed (see auxiliary material). During wet years the anomalous wind propagation during the winter in the basin is from the ocean and from a southwest direction. This brings moisture resulting in more snowfall in the basin and consequently, more spring streamflow. In the dry years it is the opposite, i.e., the basin experiences northerly dry winds; hence, less snow and consequently low spring streamflow. These composite map features are consistent with the correlation maps. The land conditions from antecedent fall season also plays a role in the spring streamflow. Wetter conditions in the fall favor less infiltration during the spring snowmelt enabling enhanced streamflow. This can be seen in the PDSI composite and vice versa during dry years.

[1] Auxiliary materials are available in the HTML. doi:10.1029/2009WR007965.

Table 2. Selected Multimodels for Each Lead Time[a]

Lead Time	Predictors	PDSI	GPH	ZNW	MDW	SST	SWE
1 Apr	2	1	0	0	0	0	1
1 Feb	2	0	0	0	0	1	1
1 Feb	2	1	0	0	0	0	1
1 Feb	3	0	0	0	1	1	1
1 Jan[b]	3	0	1	1	0	0	0
1 Jan	2	0	1	1	0	0	0
1 Nov	2	1	1	0	0	0	0
1 Nov	1	1	0	0	0	0	0

[a]Here "1" indicates the presence of a predictor and "0" indicates the absence of a predictor. A selected predictor refers to the most recent data available at the forecast time (e.g., May GPH for a 1 April forecast).
[b]This 1 January prediction used the 1 April and 1 November GPH as separate predictors.

5.2. Multimodel Selection

[53] Using the predictors identified in Table 1 and the methodology described in section 5.1, multimodels were selected for different lead times and listed in Table 2. It is interesting to note that the number of multimodels decreases with lead time. This is intuitive, in that on 1 January SWE is not available and 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 the best predictor of the ensuing spring streamflow; therefore, fewer models with other predictor variables are needed, necessitating a smaller number of multimodels. Similar observations were made by *Regonda et al.* [2006].

5.3. Forecast Skill

[54] The leave-one-out cross-validated ensemble forecast of the index gauge spring season streamflow issued on 1 April and 1 January are shown as box plots in Figure 2. The box height corresponds to the interquartile range, the whiskers depict the 5th and 95th percentiles and the horizontal line is the median. The true observations for each year are joined by a line. The dashed horizontal lines correspond to the 33rd, 50th and 66th percentiles of the data. The forecast ensembles capture the true values very well at both lead times, as attested by high RPSS scores. The high skill of the 1 January forecast is noteworthy as this forecast is made entirely from climate

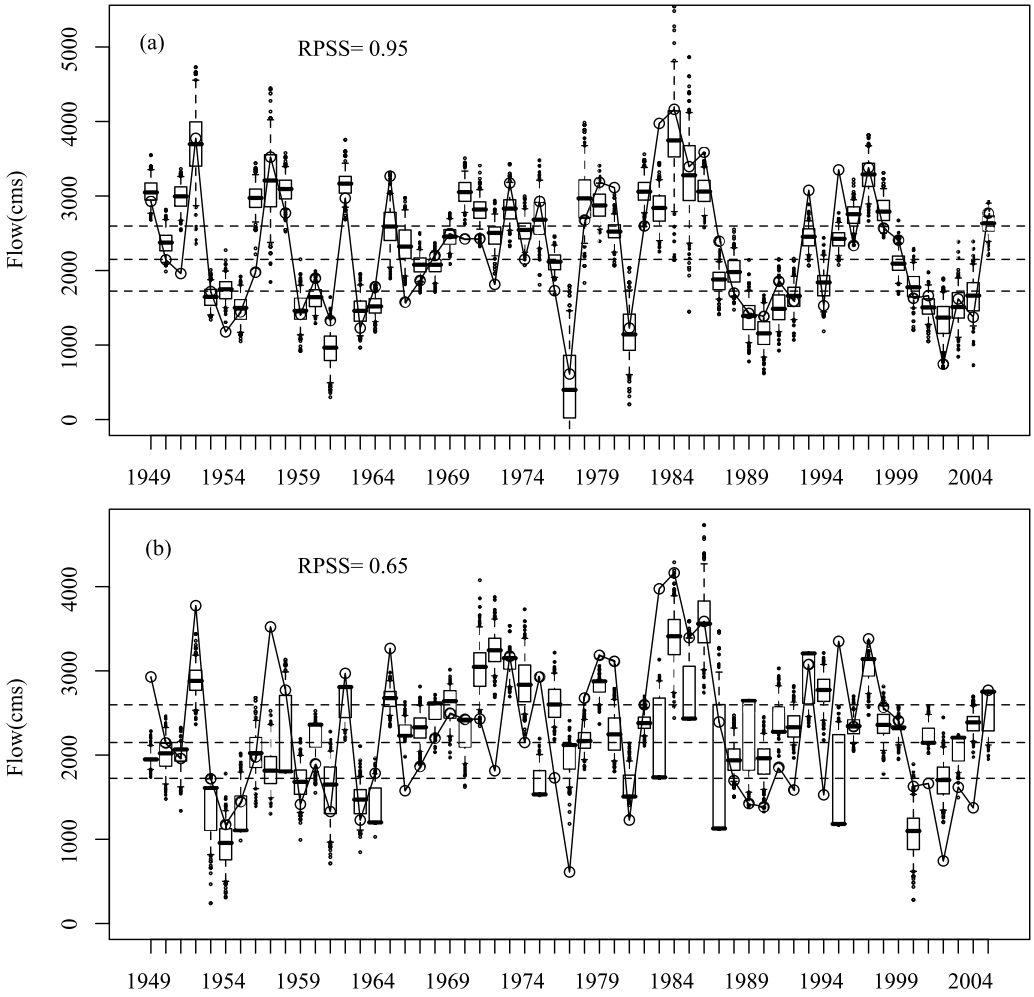

Figure 2. Leave-one CV forecasts of index gauge spring streamflow issued on (a) 1 April and (b) 1 January. The boxes correspond to the interquartile ranges, the horizontal line in each box is the median, whiskers extend to the 5th and 95th percentiles of ensemble, and the solid line joins the true values for each year. The dashed horizontal lines correspond to the 33rd, 50th, and 66th percentiles of the data.

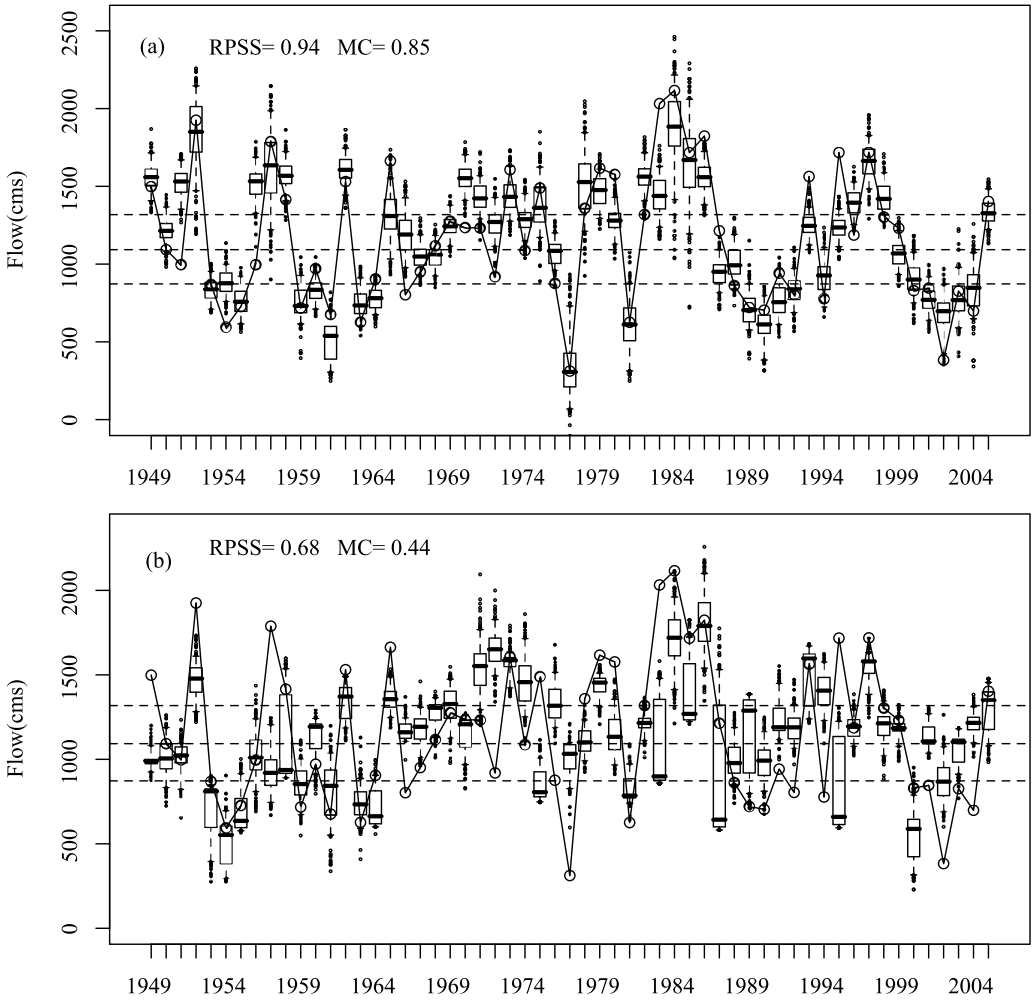

Figure 3. Same as Figure 2a but for Lees Ferry.

information. These forecasts were disaggregated to ensemble forecasts at the four locations, Figure 3 shows the ensemble forecast for Lees Ferry. We find that the forecasts capture the observed flows quite well and the high skill in forecasting the index gauge (Figure 2) is translated to the disaggregated forecasts. The results were similar at other locations. The skill scores of the seasonal forecast at different locations and lead times are shown in Table 3. The skill at the Bluff location is low in comparison to the rest. This is due to the fact that flows at Bluff are quite small compared to the other three locations and the disaggregation method tend to under perform in such situations, also seen by *Prairie et al.* [2007]. One alternative is to perform the disaggregation in two steps where the Bluff flows are generated in the second step. Temporal disaggregation of the seasonal forecasts to monthly forecasts was also skillful, this can be seen in the box plot of May monthly ensemble forecast at Lees Ferry in Figure 4. The noteworthy aspect is that the skills are retained through the spatial and temporal disaggregation process. Furthermore, the forecasts are significantly skillful at longer lead times; therefore, water managers can obtain a good idea of the spring streamflow months in advance which will be useful for efficient water resources management. As to be expected forecast skills decrease with increase in lead time but improve upon climatology, nonetheless.

[55] To challenge the forecasting system, we performed the forecasts by dropping 10% of the observations and repeating the forecast 100 times. The RPSS skills on the dropped observations for the seasonal flow forecast issued on 1 April are shown as box plots in Figure 5. There is considerable variability in the skill scores due to sampling, but the median skill scores are quite high at all the locations except Bluff

Table 3. Skills of Spring Season Streamflow Forecast at Different Lead Times

	RPSS				MC			
Site	1 Apr	1 Feb	1 Jan	1 Nov	1 Apr	1 Feb	1 Jan	1 Nov
Leave-One Validation Mode								
Index	0.95	0.85	0.65	0.28	-	-	-	
Cisco	0.37	0.58	0.70	0.34	0.56	0.75	0.44	0.42
GRUT	0.63	0.49	0.12	0.21	0.76	0.77	0.44	0.37
Bluff	0.73	0.06	0.40	0.19	0.78	0.53	0.41	0.26
Lees Ferry	0.68	0.71	0.68	0.31	0.63	0.79	0.44	0.42
Retroactive Validation Mode								
Index	0.80	0.85	0.29	0.10	-	-	-	-
Cisco	0.97	0.73	0.62	0.78	0.81	0.72	0.31	0.52
GRUT	0.54	0.46	0.17	0.58	0.92	0.79	0.13	0.48
Bluff	0.37	0.40	0.15	0.15	0.41	0.58	0.52	0.35
Lees Ferry	0.87	0.73	0.62	0.60	0.88	0.77	0.33	0.57

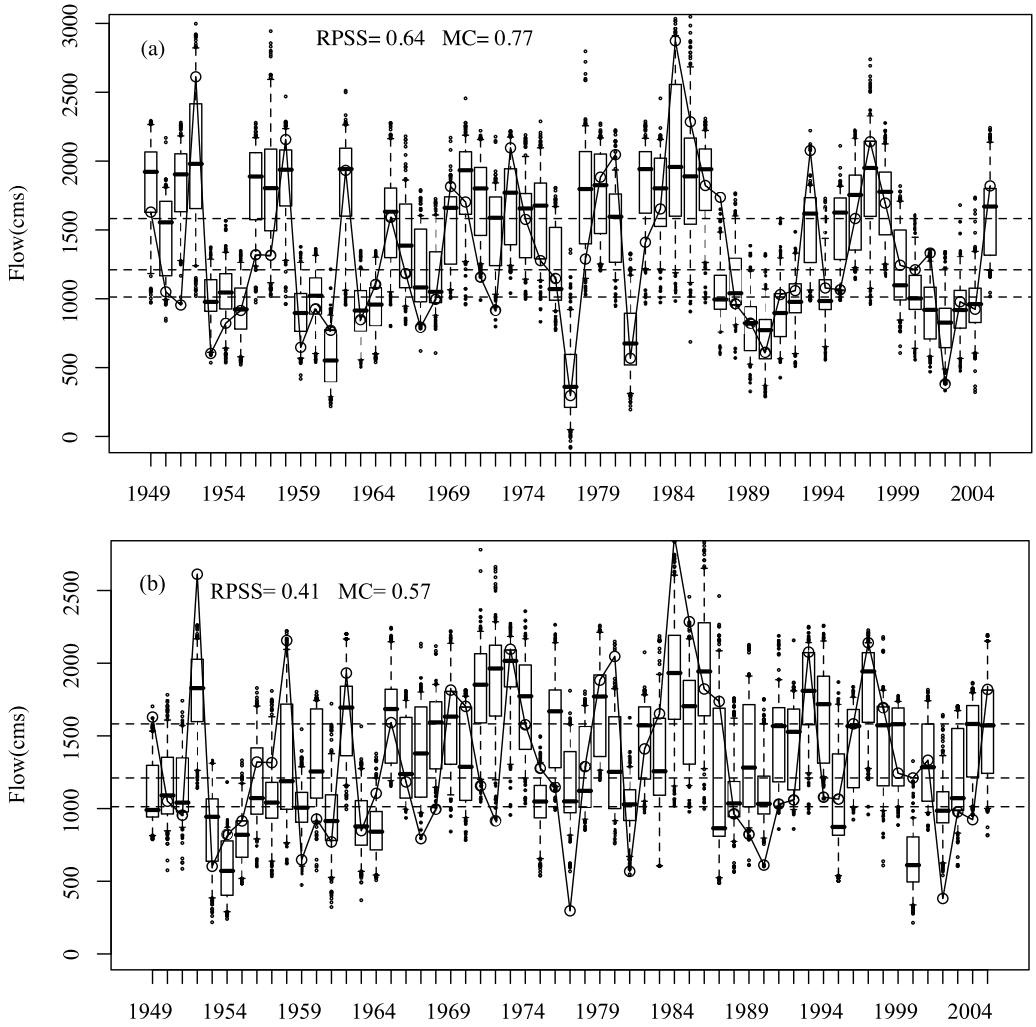

Figure 4. Same as Figure 3a but for May streamflow at Lees Ferry.

for the reasons mentioned above. Similar results were observed for the monthly forecasts.

[56] One of the advantages of the disaggregation method is that it can capture the spatial correlation in a parsimonious manner. We evaluate this by comparing the ensemble leave-one-out cross-validated forecasted flows with the historical values. The spatial correlations are very well captured (see auxiliary material).

[57] In practice forecasts are made one year at a time using all the prior data, i.e., a retroactive forecast. Creating a retroactive forecast enables us to compare with coordinated forecasts and the ESP forecasts, which are issued in this manner. We performed retroactive forecasts for the period 1990–2005 for which coordinated forecasts are available for comparison. Figures 6a and 6b show the forecasts of Lees Ferry seasonal streamflow issued on 1 January and 1 April, respectively, and Figure 6c shows the forecasts of May streamflow at Lees Ferry issued on 1 April. In all these, the forecast from our approach is comparable to the coordinated forecasts. The correlation between the historical flow and the coordinated forecast was computed for the years available and compared to the median ensemble forecast correlation for the corresponding year (Table 4). In wet years our approach tends to be better but does not capture extreme dry years such as 2002 as well as the coordinated forecast. In dry years land surface conditions play a substantial role, which is better captured by the physical models, resulting in increased skill in the CF relative to the statistical approach. This is corroborated

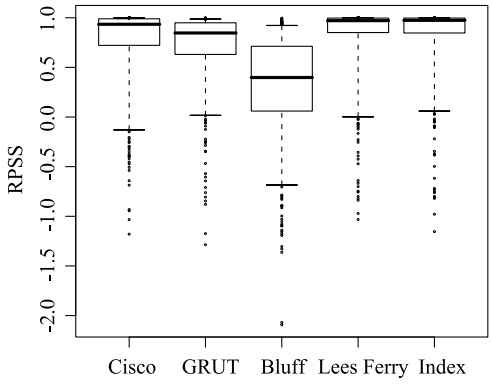

Figure 5. Box plot of RPSS of spring streamflow forecast at the four locations and the index gauge. Forecasts are based on dropping 10% of the observations.

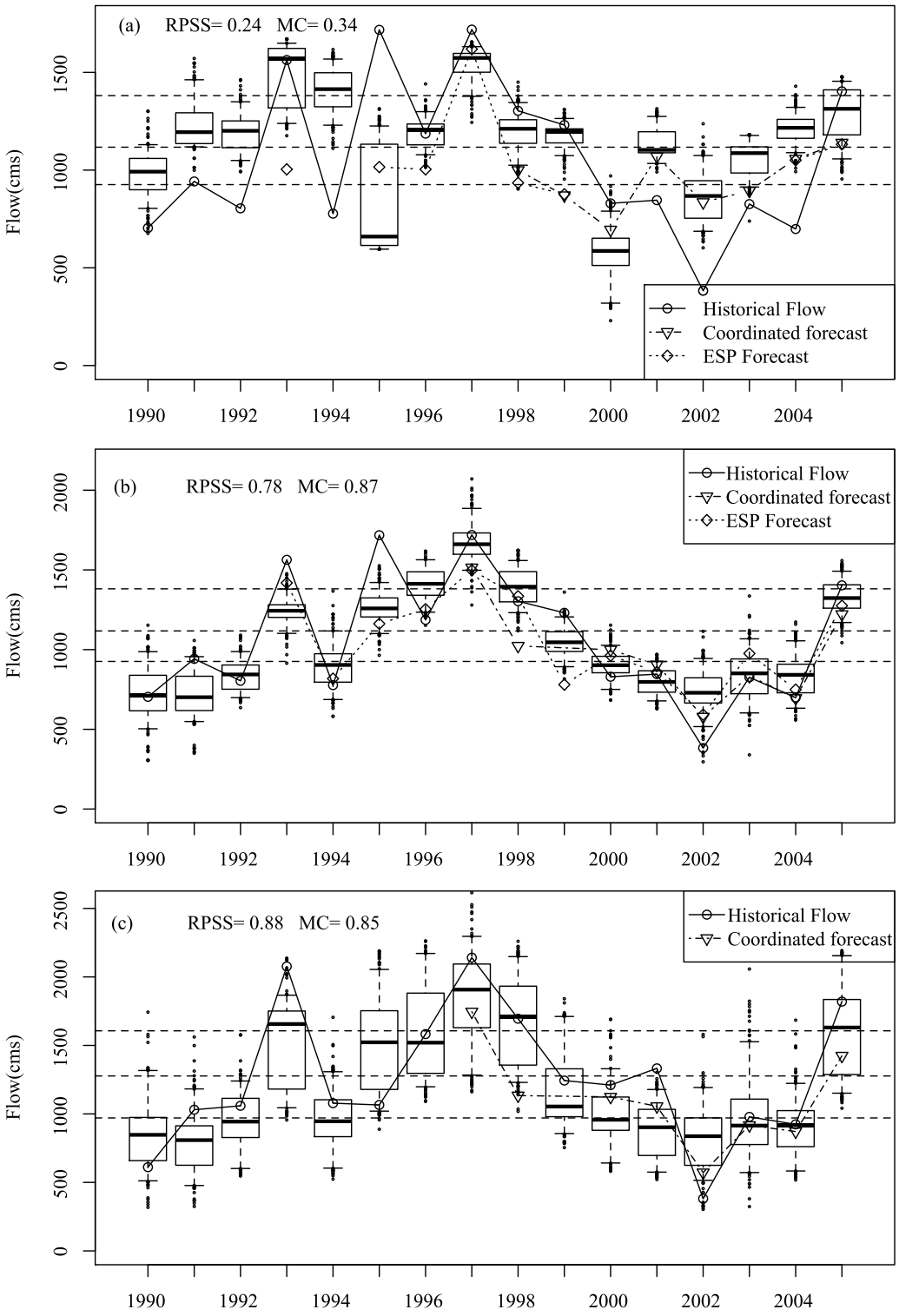

Figure 6. Spring flow forecast at Lees Ferry issued on (a) 1 January and (b) 1 April in a retroactive mode. (c) May flow forecast at Lees Ferry after spatial and temporal disaggregation, issued on 1 April. The dash-dotted line represents the coordinated forecast, and the dotted line represents the ESP forecast.

Table 4. Model Correlations With Historical Flow for Available Coordinated Forecast Years

Figure	Coordinated	Median Ensemble
6a	0.404	0.571
6b	0.955	0.962
6c	0.846	0.962

by *Regonda et al.* [2006]. Note that though we provide these comparisons, systematic comparisons need to be made with forecasts from all the techniques. We also suggest a multimodel ensemble combination approach to objectively blend these forecasts.

6. Summary

[58] We present a parsimonious framework to provide seasonal ensemble streamflow forecasts at several locations simultaneously in a river network that preserves summability. The framework is an integration of two recent approaches: (1) a nonparametric multimodel ensemble forecast technique [*Regonda et al.*, 2006] and (2) a nonparametric space-time disaggregation technique [*Prairie et al.*, 2007]. The approach generates an ensemble forecast for the seasonal streamflow of an "index gauge," which is constructed as the sum of all the spatial flows, using large-scale ocean-land-atmospheric features and a multimodel combination method. These forecasts are then disaggregated in space and time resulting in an ensemble forecast for all the months and at all the locations, thereby capturing the temporal and spatial correlations.

[59] We demonstrate the framework by generating streamflow forecast at four locations in the Upper Colorado River basin. The noteworthy finding is that the high-skill annual forecasts at the index gauge are translated to monthly, multilocation forecasts; achieving this was not obvious when we embarked on this study. The forecasts also showed significant skill at longer lead times; thus, providing crucial advance knowledge of the spring streamflow in the river basin enabling efficient planning and management. Furthermore, the skills from this simpler framework are comparable to the coordinated predictions that are currently used. The combination of these two forecasts is the logical next step to improve the forecast skills further.

[60] **Acknowledgments.** The first author is thankful for a summer fellowship in 2007 funded by the NSF REU program that was instrumental in this research. The authors would like to thank three anonymous reviewers and the editors for their helpful and insightful comments in improving the manuscript.

References

Brandon, D. G. (2005), Using NWSRFS ESP for making early outlooks of seasonal runoff volumes into Lake Powell, paper presented at Hydrology of Arid and Semi-Arid Regions, Am. Meteorol. Soc., San Diego, Calif., 9–13 Jan.

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. Climatol.*, *9*, 928–948, doi:10.1175/1520-0442(1996)009<0928:ICVASI>2.0.CO;2.

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, doi:10.1029/2000WR900305.

Craven, P., and G. Wahba (1978), Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross validation, *Numer. Math.*, *31*, 377–403, doi:10.1007/BF01404567.

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, doi:10.1175/JHM-386.1.

Dracup, J. A., and E. Kahya (1994), The relationships between U.S. streamflow and La Nina events, *Water Resour. Res.*, *30*, 2133–2142, doi:10.1029/94WR00751.

Fukunaga, K. (1990), *Introduction to Statistical Pattern Recognition*, Academic, San Diego, Calif.

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. Climatol.*, *11*, 1575–1586, doi:10.1175/1520-0442(1998)011<1575:EIOIER>2.0.CO;2.

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, doi:10.1111/j.1600-0870.2005.00103.x.

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, doi:10.1111/j.1752-1688.1999.tb04240.x.

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. Climatol.*, *15*, 1555–1572, doi:10.1175/1520-0442(2002)015<1555:RBCVAW>2.0.CO;2.

Hunter, T., G. A. Tootle, and T. C. Piechota (2006), Oceanic-atmospheric variability and western U.S. snowfall, *Geophys. Res. Lett.*, *33*, L13706, doi:10.1029/2006GL026600.

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

Kahya, E., and J. A. Dracup (1993), U.S. streamflow patterns in relation to the El Nino/Southern Oscillation, *Water Resour. Res.*, *29*, 2491–2504, doi:10.1029/93WR00744.

Kahya, E., and J. A. Dracup (1994), The influences of type 1 El Nino and La Nina events on streamflows in the Pacific southwest of the United States, *J. Climatol.*, *29*, 2491–2504.

Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, *77*, 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

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, doi:10.1126/science.285.5433.1548.

Krishnamurti, T. N., C. M. Kishtawal, Z. Zhang, T. E. LaRow, D. R. Bachiochi, C. E. Williford, S. Gadgil, and S. Surendran (2000), Multimodel ensemble forecasts for weather and seasonal climate, *J. Climatol.*, *13*, 4196–4216, doi:10.1175/1520-0442(2000)013<4196:MEFFWA>2.0.CO;2.

Lall, U. (1995), Recent advances in nonparametric function estimation: Hydraulic applications, *Rev. Geophys.*, *33*, 1093–1102, doi:10.1029/95RG00343.

Lall, U., and A. Sharma (1996), A nearest neighbor bootstrap for time series resampling, *Water Resour. Res.*, *32*, 679–693, doi:10.1029/95WR02966.

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

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., 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, doi:10.1002/(SICI)1097-0088(19991115)19:13<1399::AID-JOC457>3.0.CO;2-A.

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, doi:10.1175/1525-7541(2002)003<0013:PMAPOY>2.0.CO;2.

Opitz-Stapleton, S., S. Gangopadhyay, and B. Rajagopalan (2007), Generating streamflow forecasts for the Yakima River Basin using large-scale climate predictors, *J. Hydrol.*, *341*, 131–143, doi:10.1016/j.jhydrol.2007.03.024.

Pagano, T., and D. Garen (2004), Evaluation of official western U.S. seasonal water supply outlooks, 1922–2002, *J. Hydrometeorol.*, *5*, 896–909, doi:10.1175/1525-7541(2004)005<0896:EOOWUS>2.0.CO;2.

Piechota, T. C., J. A. Dracup, and R. G. Fovell (1997), Western US streamflow and atmospheric circulation patterns during El Nino–Southern Oscillation, *J. Hydrol.*, *201*, 249–271, doi:10.1016/S0022-1694(97)00043-7.

Prairie, J., and R. Callejo (2005), Natural flow and salt computation methods, technical report, Bur. of Reclam., Washington, D. C.

Prairie, J., B. Rajagopalan, T. Fulp, and E. Zagona (2005), Statistical nonparametric model for natural salt estimation, *J. Environ. Eng.*, *131*(1), 130–138, doi:10.1061/(ASCE)0733-9372(2005)131:1(130).

Prairie, J., B. Rajagopalan, U. Lall, and T. Fulp (2007), A stochastic nonparametric technique for space-time disaggregation of streamflows, *Water Resour. Res.*, *43*, W03432, doi:10.1029/2005WR004721.

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, doi:10.1175/1520-0493(2002)130<1792:CCFTRA>2.0.CO;2.

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, doi:10.1029/91WR00690.

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.

Singhrattna, N., B. Rajagopalan, M. Clark, and K. K. Kumar (2005), Forecasting Thailand summer monsoon rainfall, *Int. J. Climatol.*, *25*, 649–664, doi:10.1002/joc.1144.

Soukup, T., O. Aziz, G. Tootle, T. Piechota, and S. Wulff (2009), Long lead-time streamflow forecasting of the North Platte River incorporating oceanic-atmospheric climate variability, *J. Hydrol.*, *368*, 131–142, doi:10.1016/j.jhydrol.2008.11.047.

Tarboton, D. G., A. Sharma, and U. Lall (1998), Disaggregation procedures for stochastic hydrology based on nonparametric density estimation, *Water Resour. Res.*, *34*, 107–119, doi:10.1029/97WR02429.

Wadsworth, H. M. (Ed.) (1990), *Handbook of Statistical Methods for Engineers and Scientists*, McGraw-Hill, 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.

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

Yates, D., S. Gangopadhyay, B. Rajagopalan, and K. Strzepek (2003), A nearest neighbor bootstrap technique for generating regional climate scenarios for integrated assessments, *Water Resour. Res.*, *39*(7), 1199, doi:10.1029/2002WR001769.

C. Bracken and B. Rajagopalan, Department of Civil Environmental and Architectural Engineering, University of Colorado at Boulder, 428 UCB, Boulder, CO 80309, USA. (balajir@colorado.edu)

J. Prairie, Bureau of Reclamation, University of Colorado at Boulder, Boulder, CO 80309, USA.