A Multi-Site Seasonal Ensemble Streamflow Forecasting Technique

Cameron Bracken\textsuperscript{1}, Balaji Rajagopalan\textsuperscript{2,3,*} and James Prairie\textsuperscript{4}

\textsuperscript{1}Department of Environmental Resources Engineering, Humboldt State University, Arcata, CA 95521 USA
\textsuperscript{2}Department of Civil Environmental and Architectural Engineering, University of Colorado, Boulder, CO 80309 USA
\textsuperscript{3}Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309 USA
\textsuperscript{4}Bureau of Reclamation, University of Colorado, Boulder, CO 80309 USA

\textbf{Water Resources Research}

March 2009

\textsuperscript{*}Corresponding Author
Balaji Rajagopalan
University of Colorado
428 UCB
Boulder, Colorado 80309-0421
Phone: 303.492.5968
Fax: 303.492.5317
Email: balajir@colorado.edu
Abstract

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 multi-model ensemble forecast technique (Regonda et al. 2006) and the nonparametric space-time disaggregation technique (Prairie et al. 2007). The four main components of the proposed framework are: (i) an index gauge streamflow is constructed as the sum of flows at all the desired spatial locations, (ii) potential predictors of the spring season (April-Jul) streamflow at this index gauge are identified from the large scale ocean-atmosphere-land system including snow water equivalent, (iii) the multi-model ensemble forecast approach (Regonda et al., 2006) is used to generate the ensemble flow forecast at the index gauge, (iv) 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. 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 ‘coordinated’ forecast – a combination of the ESP, Statistical Water Supply, a principal component regression technique, and modeler knowledge, and find that overall the proposed method is equally skilful while. The forecasts from this approach can be a valuable input for efficient planning and management of water resources in the basin.

1. Introduction and Background
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.

Nearly 80% of the streamflow in the UCRB and in many of western US river basins is due to snowmelt and as a result streamflow models have long been dominated by this information in particular, the April 1st 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 (Nov-Mar) requiring a skilful projection of the spring streamflow at a time when the snow information is incomplete. This poses a challenging problem of providing skilful basin-wide streamflow forecasts at longer lead time.

There is increasing evidence that large scale climate features in the Pacific have strong influence on the hydroclimatology of western US 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). These links enhance the prospects for long lead streamflow forecast. These links were identified and incorporated in a statistical modeling approach to generate skilful forecasts of spring streamflow at long lead times during early winter on several river basins in the western US, 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 (Stapleton et al., 2007).

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

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 multi-model ensemble streamflow technique that includes a principal component analysis on the basin streamflow, which 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. i. The forecasted flows at the spatial locations do not satisfy the summability criteria, i.e., 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. ii. 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 i.).
To overcome the drawbacks of the coordinated forecasts and the multi-site forecasting method of Regonda et al. (2006), we propose a framework that integrates the multi-model 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: (i) an index gauge streamflow is constructed as the sum of flows at all the desired spatial locations, (ii) potential predictors of the spring season (April-Jul) streamflow at this index gauge are identified from the large scale ocean-atmosphere-land system including snow water equivalent, (iii) multi-model ensemble forecast approach (Regonda et al., 2006) is used to generate the ensemble flow forecast at the index gauge and, (iv) 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. 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 multi-model 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

The CRB includes parts of seven states in the western United States with an area of approximately 250,000 square miles. The basin includes widely varying topography with elevations ranging from 200 to 14,200ft. Most of the flow in the basin is a result of snowmelt from the UCRB, while most of the water use occurs in the semi-arid 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 gage, 16 miles 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 Gage

The natural streamflow data for the Colorado River Basin are developed by the Bureau of Reclamation (Reclamation) and updated regularly\(^1\). 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 (Apr-Jul) from spring melt of winter snow, much like other river basins in the western US.

3.2. Large-scale climate data

Ocean-atmospheric circulation variables that capture the large-scale climate forcings are available from NOAA’s Climate Diagnostics Center web site\(^2\). In particular, the variables used were 500 mb geopotential height (GPH), zonal (ZW) and meridional 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

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

---

\(^1\) The natural flow data and additional reports describing these data are available at http://www.usbr.gov/lc/region/g4000/NaturalFlow/index.html

\(^2\) http://www.cdc.noaa.gov/
Conservation Service (NRCS) from their website\(^3\). Data was obtained at 10 sites in the UCRB and averaged to create a continuous record of monthly SWE for February, March and April 1st.

3.3. PDSI data

Antecedent summer and fall season land conditions can play an important role in the variability of the following spring streamflow. This was shown in 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).

4. Proposed Integrated Framework

The proposed framework applies three steps to the index gauge; (i) identify potential predictors of the spring season (April-Jul) streamflow at the index gauge from the large scale ocean-atmosphere-land system including snow water equivalent, (ii) use multi-model ensemble forecast approach (Regonda et al., 2006) to generate ensemble flow forecast at the index gauge and (iii) disaggregate the ensembles 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. The components of each step are described below.

4.1. Predictor Suite for Index Gage seasonal streamflow

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 in each variable and the spatial average of these regions are computed to create a suite of potential predictors. The on-line tool developed by

\(^3\) http://www.wcc.nrcs.usda.gov/snow
Physical Sciences Division, National Oceanic and Atmospheric Administration (NOAA) is used for this purpose\(^4\).

4.2. Multi-Model Selection for I Gage Seasonal Flow Forecast

The MME methodology consists of two distinct steps; (i) selection of the multi models each with its own subset of predictors and (ii) combing forecasts from the individual models. The general form of the forecast model is:

\[ Y = f(x) + \varepsilon \]  \hspace{1cm} [1]

Where \( Y \) the index gage spring season streamflow, \( x \) is is a suite of predictors and \( \varepsilon \) is the residual that is assumed to be Normally distributed with mean 0 and variance \( \sigma^2 \).

If \( f \) is a global function based on the entire data and linear, the resulting 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, (i) the assumption of a Normal distribution of the errors and the variables and (ii) 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.

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

\[^4\text{http://www.cdc.noaa.gov/Correlation/}\]
set, there are several ‘local fits’, each capturing the local features, hence the ability to model any arbitrary features (linear or nonlinear) that the data exhibits. 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, 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., 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 as Regonda et al. (2006) also did.

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.

1. Select a subset of predictors
2. Select a degree of polynomial to fit, typically, $p=1$ (linear fit) or 2 (quadratic) is quite adequate.
3. Select a size of the neighborhood, $K=\alpha*N$ (where $\alpha=(0,1)$ and $N=$number of observations).
4. For a point $x$ 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).
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, $\hat{Y}$
6. Repeat steps 3-4 for all the data points
7. Compute the objective criteria, Generalized Cross Validation Estimation (GCV)

$$GCV(K,p) = \frac{\sum_{i=1}^{N} e_i^2}{N} \left(1 - \frac{q}{N}\right)$$  \[2\]
Where $e_i$ is the model residual ($Y_i - \hat{Y}$) for the $i^{th}$ 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 Whaba, 1979).

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) where in, an objective function such as Mallow’s $C_p$ statistic, adjusted $R^2$, AIC (Akaike Information Criteria), 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.

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; hence, a multi-model approach is warranted. Recent studies show that multi-model 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).

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., multi-models). Combinations with predictor variables significantly correlated amongst each other (i.e., multicollinear) are removed from the multi-model pool, as they lead to over fitting and poor predictive skill (Wadsworth, 1990), thus resulting in a set of multi-models each with different predictor subsets.

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. Multi-model Ensemble Forecast of Index Gage Seasonal Streamflow

The multi-model ensemble forecast generation is described below, again following Regonda et al. (2006). Suppose we desire a multi-model ensemble forecast for a point \( x_j \), the steps are as follows.

1. From one of the multi-models identified in the preceding section, obtain the model prediction \( \hat{Y}_i \) and the estimate of the error variance (\( \sigma^2_{e_j} \)) (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 \( \hat{Y}_i \).

\[
z_{i,j} = \hat{Y}_i(x_j) + N(0,\sigma_{e_j}) \tag{3}
\]

where \( N(0,\sigma_{e_j}) \) is a normally distributed random variable with mean 0 and standard deviation \( \sigma_{e_j} \). This approach assumes Normality of the residuals, a standard assumption from regression theory.

2. Repeat step 1 for all the models in the multi-model suite. Thus, obtaining an ensemble forecast of size \( n \) from each model.

3. Weight each of the multi-models based on 1/GCV criteria. This way the model with the lowest GCV value is most heavily favored.

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

5. Randomly choose one of the \( n \) ensemble members.

6. Repeat steps 4-5 \( n \) times to obtain a multi-model ensemble forecast.

4.4. Spatial and Temporal Disaggregation of the Index Gauge Forecast

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 in Prairie et al. (2007). Below we provide a brief description of its implementation in the current application.
The disaggregation procedure can be thought of as sampling from the conditional probability density function (PDF), \( f(x|z) \), where \( x \) is a \( d \) dimensional vector of flows, \( z \) is the aggregate flow, but with the (additivity) constraint that the values in \( x \) add up to \( z \). This is achieved by an orthonormal rotation of the data \( x \) to \( Y \) 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.

1. The first step is to generate the orthonormal rotation matrix, \( R(d) \) using the Gram-Schmidt algorithm. Note that the \( d \times d \) matrix \( R \) is only a function of the dimension \( d \) and has the property \( R^T = R^{-1} \) 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.

2. The matrix \( X \) of the historical monthly streamflow at gauge I is rotated to \( Y \) by the rotation matrix \( R \) as,

\[
Y = RX. \quad [4]
\]

Both \( X \) and \( Y \) are of dimension \( N \) (number of years) rows by \( d \) (\( \approx 4 \) months in the season) columns.

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

\[
Y = [U, z']. \quad [5]
\]

3. For a given seasonal flow forecast say \( z_{nim} \), \( K \) nearest neighbors are identified from the historical seasonal flow values in \( z \). One of the \( K \) nearest neighbor is resampled using a weight function.
\[ W(k) = \frac{1}{\sum_{i=1}^{k} \frac{1}{i}} \] where \( k = 1, 2, \ldots, K \). \[6\]

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).

4. Suppose the selected neighbor corresponds to historic year \( j \), the new vector \( y_{sim} \) is constructed as

\[ y_{sim} = \left[ u_j, \frac{z_{sim}}{\sqrt{d}} \right] \] \[7\]

5. This is back transformed to the original space as

\[ x_{sim} = R^T y_{sim} \] \[8\]

The vector \( x_{sim} \) now contains \( d = 4 \) values, the disaggregated monthly flow of the seasonal total flow, \( z_{sim} \).

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 multi-model ensemble members results in a multi-model ensemble forecast for each month at the four locations.

4.5. Verification and Validation

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 multi-category
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 $(p_1, p_2, ..., p_k)$. The observational vector $(d_1, d_2, ..., d_k)$ is obtained for each forecast, where $d_k$ is unity if the observation falls in the $k^{th}$ category and zero otherwise. The ranked probability skill score (RPSS) is defined as follows:

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

$$ \text{RPSS} = 1 - \frac{RPS(\text{forecast})}{RPS(\text{climatolgy})} $$

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

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 three types of forecasts:

1. Leave-one out forecast. In this each year is dropped and the multi model ensemble for that year is generated from the rest, e.g., Grantz et al., 2005; Regonda et al., 2006; Singhattratna et al., 2005.

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 insight into the sensitivity of the forecast to sampling variability.
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

We chose four lead times, Nov 1st, Jan 1st, Feb 1st and Apr 1st, 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

The index gauge spring season streamflow was correlated with large-scale ocean, atmosphere and land variables from preceding months. Figure 2 shows the correlation with Feb-Mar GPH, Mar ZNW, Jan-Mar SST and Mar MDW, the combination of months that showed strong correlation. The correlation maps are consistent with prior findings for the river flows in western US (e.g., Grantz et al., 2006; Regonda et al., 2007). 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 Feb 1st.

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. Figure 3 shows the composite maps of Oct-Mar vector winds for the 6 wettest and driest years and the preceding fall (Sep-Nov) season PDSI for the same years. During wet years the anomalous wind propagation during the winter in the basin is from the ocean and from a south west 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 snow melt enabling
enhanced streamflow. This can be seen in the PDSI composite and vice-versa during dry years.

5.2. Multi Model Selection

Using the predictors identified in Table 1 and the methodology described in the previous section, multi-models were selected for different lead times and listed in Table 2. It is interesting to note that the number of multi-models decreases with lead time. This is intuitive, in that on Jan 1st 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 multi-model pool. However, on April 1st, 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 multi-models. Similar observations were found by Regonda et al. (2006).

5.3. Forecast Skill

The leave-one out cross validated ensemble forecast of the index gauge spring season streamflow issued on April 1st and Jan 1st are shown as boxplots in Figure 4. 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 Jan 1st forecast is noteworthy as this forecast is made entirely from climate information. These forecasts were disaggregated to ensemble forecasts at the four locations, Figure 5 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 (Fig 4) 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).
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 skilful, this can be seen in the boxplot of May monthly ensemble forecast at Lees Ferry in Figure 6. 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.

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 Apr 1st are shown as boxplots in Figure 7. 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 for the reasons mentioned above. Similar results were observed for the monthly forecasts.

One of the advantages of the disaggregation method is that it can capture the spatial correlation in a parsimonious manner. We evaluate this in Figure 8 with boxplots of the cross correlation between the leave-one out cross validated forecasted flows at the four locations compared with the historical values. The spatial correlations are very well captured.

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. Figure 9(a), (b) shows the forecasts of Lees Ferry seasonal streamflow issued on Jan 1st and Apr 1st, respectively, Figure 9(c) shows the forecasts of May streamflow at Lees Ferry issued on Apr 1st. In all these, the forecast from our approach is comparable to the coordinated forecasts. 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. Note that though we provide these comparisons, systematic comparisons need to be made.
with forecasts from all the techniques. We also suggest a multi-model ensemble combination approach to objectively blend these forecasts.

Summary

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; (i) a nonparametric multi-model ensemble forecast technique (Regonda et al. 2006) and (ii) 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 mutli-model 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.

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, multi-location 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.

Acknowledgements

The first author is thankful for a summer fellowship in 2007 funded by the NSF REU program that was instrumental in this research.
References

Brandon, D. G. (2005), Using NWSRFS ESP for making early outlooks of seasonal runoff volumes into Lake Powell, in *Hydrology of Arid and Semi-Arid Regions*.


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, Journal of Climatology, 13, 4196 - 4216.


Rajagopalan, B., U. Lall, and S. Zebiak (2002), Optimal categorical climate forecasts through multiple GCM ensemble combination and regularization, Monthly


Figure Captions

Figure 1: Study area.

Figure 2: Correlation maps of index gage spring (Apr-Jul) streamflow with Feb.-Mar. GPH (top left), March ZW (top right), Jan.-Mar. SST (bottom left) and Mar. MW (bottom right).

Figure 3: Composite maps of Oct.-Mar. 500 mb vector winds for 6 wettest years (top left) and 6 driest years (top right) years and preceding Fall PDSI for wet (bottom left) and dry (bottom right) years.

Figure 4: Leave-one CV forecasts of index gauge spring streamflow issued on (a) Apr 1st and (b) Jan 1st. The box corresponds to the interquartile range, the horizontal line inside the boxes is the median, whiskers extend to 5th and 95th percentile of ensemble and solid line joins the true value for each year. The dashed horizontal lines correspond to the 33rd, 50th and 66th percentiles of the data.

Figure 5: (a) same as Figure 4(a) but for Lees Ferry. (b) Same as Figure 4(b) but for Lees Ferry.

Figure 6: (a) same as Figure 5(a) but for May streamflow at Lees Ferry. (b) same as Figure 5(b) but for May streamflow

Figure 7: Boxplot of RPSS of spring streamflow forecast at the four locations and the index gauge. Forecasts are based on dropping 10% of the observations.

Figure 8: Boxplots of cross correlation of spring season streamflow forecasts issued on Apr 1st between sites: C=Cisco, B=Bluff, G=GRUT, LF=Lees Ferry. The solid line represents the observed statistics.
Figure 9: Spring flow forecast at Lees Ferry issued on (a) Jan. 1st and (b) Apr. 1st in a retroactive mode. Also, May flow forecast at Lees Ferry (c) after spatial and temporal disaggregation, issued on Apr. 1st. The dashed line represents the coordinated forecast and the dotted line represents the ESP forecast.

Table Captions

Table 1: Potential predictors and their regions in lon:lat

Table 2: Selected multi-models for each lead time. “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 an Apr 1st forecast).

*This Jan. 1st prediction used the Apr. 1st and Nov. 1st GPH as separate predictors.

Table 3: Skills of spring season streamflow forecast at different lead times
Figure 1: Study area (CBRFC 2006).
Figure 2: Correlation maps of index gage spring (Apr-Jul) streamflow with Feb.-Mar. GPH (top left), March ZW (top right), Jan.-Mar. SST (bottom left) and Mar. MW (bottom right).
Figure 3: Composite maps of Oct.-Mar. 500 mb vector winds for 6 wettest years (top left) and 6 driest years (top right) years and preceding Fall PDSI for wet (bottom left) and dry (bottom right) years.
Figure 4: Leave-one CV forecasts of index gauge spring streamflow issued on (a) Apr 1st and (b) Jan 1st. The box corresponds to the interquartile range, the horizontal line inside the boxes is the median, whiskers extend to 5th and 95th percentile of ensemble and solid line joins the true value for each year. The dashed horizontal lines correspond to the 33rd, 50th and 66th percentiles of the data.
Figure 5: (a) same as Figure 4(a) but for Lees Ferry. (b) Same as Figure 4(b) but for Lees Ferry.
Figure 6: (a) same as Figure 5(a) but for May streamflow at Lees Ferry. (b) same as Figure 5(b) but for May streamflow.

Figure 7: Boxplot of RPSS of spring streamflow forecast at the four locations and the index gauge. Forecasts are based on dropping 10% of the observations.
Figure 8: Boxplots of cross correlation of spring season streamflow forecasts issued on Apr 1st between sites: C=Cisco, B=Bluff, G=GRUT, LF=Lees Ferry. The solid line represents the observed statistics.
Figure 9: Spring flow forecast at Lees Ferry issued on (a) Jan. 1st and (b) Apr. 1st in a retroactive mode. Also, May flow forecast at Lees Ferry (c) after spatial and temporal disaggregation, issued on Apr. 1st. The dashed line represents the coordinated forecast and the dotted line represents the ESP forecast.
## Tables

### Table 1: Potential predictors and their regions in lon:lat

<table>
<thead>
<tr>
<th>Variable</th>
<th>Lead Time</th>
<th>Season</th>
<th>Negative Region</th>
<th>Positive Region</th>
<th>Max Cor.</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPH</td>
<td>Nov. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>oct</td>
<td>40,42N:137,140E</td>
<td>17,19N:178,182E</td>
<td>0.39</td>
</tr>
<tr>
<td>ZNW</td>
<td>Nov. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>oct</td>
<td>40,42N:-108,-106E</td>
<td>30,32N:132,135E</td>
<td>0.43</td>
</tr>
<tr>
<td>MDW</td>
<td>Nov. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>oct</td>
<td>40,42N:124,126E</td>
<td>40,42N:124,126E</td>
<td>0.31</td>
</tr>
<tr>
<td>SST</td>
<td>Nov. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>oct</td>
<td>-17,-15N:-96,-92E</td>
<td>38,40N:170,173E</td>
<td>-0.26</td>
</tr>
<tr>
<td>GPH</td>
<td>Jan. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>oct-dec</td>
<td>42,45N:-115,-110E</td>
<td>52,58N:-160,-155E</td>
<td>-0.56</td>
</tr>
<tr>
<td>ZNW</td>
<td>Jan. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>oct-dec</td>
<td>54,55.5N:-119,-114E</td>
<td>32,34N:-110,-106E</td>
<td>-0.58</td>
</tr>
<tr>
<td>MDW</td>
<td>Jan. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>oct-dec</td>
<td>45,50N:-135,-130E</td>
<td>42,45N:-91,-88E</td>
<td>-0.57</td>
</tr>
<tr>
<td>SST</td>
<td>Jan. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>oct-dec</td>
<td>28,31N:161,164E</td>
<td>46,49N:-166,-163E</td>
<td>-0.63</td>
</tr>
<tr>
<td>GPH</td>
<td>Feb. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>oct-jan</td>
<td>60,62N:128,132E</td>
<td>61,64N:-153,-150E</td>
<td>0.24</td>
</tr>
<tr>
<td>ZNW</td>
<td>Feb. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>oct-jan</td>
<td>72,74N:126,128E</td>
<td>28,31N:-119,-116E</td>
<td>0.42</td>
</tr>
<tr>
<td>MDW</td>
<td>Feb. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>oct-jan</td>
<td>66,68N:-108,-105E</td>
<td>38,42N:-90,-85E</td>
<td>0.30</td>
</tr>
<tr>
<td>SST</td>
<td>Feb. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>oct-jan</td>
<td>25,28N:160,163E</td>
<td>46,48N:-167,-164E</td>
<td>-0.64</td>
</tr>
<tr>
<td>GPH</td>
<td>Apr. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>feb-mar</td>
<td>40,42N:137,140E</td>
<td>65,70N:160,165E</td>
<td>-0.47</td>
</tr>
<tr>
<td>ZNW</td>
<td>Apr. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>mar</td>
<td>56,58N:172,176E</td>
<td>25,30N:155,160E</td>
<td>0.46</td>
</tr>
<tr>
<td>MDW</td>
<td>Apr. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>mar</td>
<td>10,12N:-177,-174E</td>
<td>56,60N:142,144E</td>
<td>-0.42</td>
</tr>
<tr>
<td>SST</td>
<td>Apr. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>jan-mar</td>
<td>24.5,26N:158,162E</td>
<td>48,51N:-165,-160E</td>
<td>-0.54</td>
</tr>
</tbody>
</table>

### Table 2: Selected multi-models for each lead time. “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 an Apr 1<sup>st</sup> forecast).

*This Jan. 1<sup>st</sup> prediction used the Apr. 1<sup>st</sup> and Nov. 1<sup>st</sup> GPH as separate predictors.*

<table>
<thead>
<tr>
<th>Lead Time</th>
<th>Predictors</th>
<th>PDSI</th>
<th>GPH</th>
<th>ZNW</th>
<th>MDW</th>
<th>SST</th>
<th>SWE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apr. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Feb. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Feb. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Feb. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Jan. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Jan. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Nov. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Nov. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 3: Skills of spring season streamflow forecast at different lead times

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RPSS</td>
<td>RPSS</td>
<td>RPSS</td>
<td>RPSS</td>
<td>MC</td>
<td>MC</td>
<td>MC</td>
<td>MC</td>
</tr>
<tr>
<td>Leave-one</td>
<td>Index</td>
<td>0.95</td>
<td>0.85</td>
<td>0.65</td>
<td>0.28</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Leave-one</td>
<td>Cisco</td>
<td>0.37</td>
<td>0.58</td>
<td>0.70</td>
<td>0.34</td>
<td>0.56</td>
<td>0.75</td>
<td>0.44</td>
<td>0.42</td>
</tr>
<tr>
<td>Leave-one</td>
<td>GRUT</td>
<td>0.63</td>
<td>0.49</td>
<td>0.12</td>
<td>0.21</td>
<td>0.76</td>
<td>0.77</td>
<td>0.44</td>
<td>0.37</td>
</tr>
<tr>
<td>Leave-one</td>
<td>Bluff</td>
<td>0.73</td>
<td>0.06</td>
<td>0.40</td>
<td>0.19</td>
<td>0.78</td>
<td>0.53</td>
<td>0.41</td>
<td>0.26</td>
</tr>
<tr>
<td>Leave-one</td>
<td>Lees Ferry</td>
<td>0.68</td>
<td>0.71</td>
<td>0.68</td>
<td>0.31</td>
<td>0.63</td>
<td>0.79</td>
<td>0.44</td>
<td>0.42</td>
</tr>
<tr>
<td>Retroactive</td>
<td>Index</td>
<td>0.80</td>
<td>0.85</td>
<td>0.29</td>
<td>0.10</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Retroactive</td>
<td>Cisco</td>
<td>0.97</td>
<td>0.73</td>
<td>0.62</td>
<td>0.78</td>
<td>0.81</td>
<td>0.72</td>
<td>0.31</td>
<td>0.52</td>
</tr>
<tr>
<td>Retroactive</td>
<td>GRUT</td>
<td>0.54</td>
<td>0.46</td>
<td>0.17</td>
<td>0.58</td>
<td>0.92</td>
<td>0.79</td>
<td>0.13</td>
<td>0.48</td>
</tr>
<tr>
<td>Retroactive</td>
<td>Bluff</td>
<td>0.37</td>
<td>0.40</td>
<td>0.15</td>
<td>0.15</td>
<td>0.41</td>
<td>0.58</td>
<td>0.52</td>
<td>0.35</td>
</tr>
<tr>
<td>Retroactive</td>
<td>Lees Ferry</td>
<td>0.87</td>
<td>0.73</td>
<td>0.62</td>
<td>0.60</td>
<td>0.88</td>
<td>0.77</td>
<td>0.33</td>
<td>0.57</td>
</tr>
</tbody>
</table>