A Multi-Site Seasonal Ensemble Streamflow Forecasting Technique

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

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

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 – 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) 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 the skillful forecast of spring seasonal streamflows at four locations in the Upper Colorado River Basin at different lead times. The skills from this simpler framework are comparable or better than the Colorado Basin River Forecast Center (CBRFC) predictions that are currently used. The forecasts from

this approach can be a valuable input for efficient planning and management of water resources in the basin.

Water Resources Research
November 2008

*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

1. Introduction and Background

The recent protracted dry period (2000-2004) 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. The Colorado River Basin Forecast Center (CBRFC) is the body charged with the task of predicting streamflows in the CRB. The current CBRFC models use multiple regression and Ensemble Streamflow Prediction (ESP) (Brandon 2005). 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 (Regonda et al. 2006). The ESP is included in the Bureau of Reclamation's (BOR) "24 month study". See Regonda et al. (2006) for a description of the BOR's forecast setup. They outline three main drawbacks: (1) no uncertainty is captured in a 24

month projection, (2) the ESP method contains the previously mentioned limitations and (3) the projection lacks the use of large-scale climate information.

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) that requires 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 winter snowfall in the Western US and UCRB in particular, and consequently, the spring melt (Grantz et al. 2005 and references therin), - enhancing the hopes 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 - Truckee and Carson river basins (Grantz et al. 2005), Gunnison river basin (Regonda et al., 2006) and the Yakima river basin (Stapleton et al., 2007).

Water resources management in a basin requires streamflow forecast at several locations on the river network. To address this, Regonda et al. (2006) developed a multi-model ensemble streamflow technique – in this, a principal component analysis is performed 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 back transformed to obtain ensemble forecast at multiple locations in the basin. They 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 is essential for management and 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 (i.e., the monthly forecasts should add up to the seasonal) and spatially (as mentioned in (i)).

To overcome these two key drawbacks, in this research, 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, (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 described which includes the algorithmic details of two main components - 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 303,450 square miles (Figure 1) (CBRFC 2006). 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 at the outflow of Lake Powell is the key gauge through which 90% of the Colorado River flow passes through also, it officially demarks the Upper and Lower basins for operational and management purposes.

Figure 1: Study area.

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 removing 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 1949-2005 period in this study. The flows at the four sites for each month are added to create the "index gauge", denoted as "I", monthly streamflow. The monthly average and the time series of the seasonal flow at I are shown in Figure 2 and Figure 3. As can be seen almost all of the annual flows occur during the spring (Apr-Jul) season. The index gauge is useful for disaggregation as mentioned before and will become clear in the following sections.

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

Figure 2: Index gage seasonal average hydrograph

Figure 3: Seasonal average (April-July) flows.

3.2. Large-scale climate data

Ocean-atmospheric circulation variables that capture the large-scale climate forcings areavailable from NOAA's Climate Diagnostics Center web site[2]. In particular, the variables used were 500 mb geopotential height (GPH), zonal (ZW) and meridonal winds (MW) and sea surface temperature (SST). These variables are provided on a $2.2^{\circ} \times 2.2^{\circ}$ 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 Conservation (NRCS) from their website[3]. Data was obtained at 10 sites in the UCRB and this was 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 following spring streamflow. This was shown in Regonda et al. (2006) in the Gunnison River Basin - where in 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

[2] http://www.cdc.noaa.gov/
[3] http://www.wcc.nrcs.usda.gov/snow

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 schematic of the framework proposed for multi-site ensemble streamflow forecast is schematically shown in Figure 4. The steps involved in these components are described below.

Figure 4: Framework

4.1. Predictor Suite for Index Gage seasonal streamflow

The index gauge spring streamflow is correlated with global ocean, atmosphere and land variables – GPH, SST, ZW, MW and PDSI, from preceding seasons. Regions of high correlations are identified and spatial average of these regions of the corresponding variables is computed to create a suite of potential predictors. The on-line tool developed by 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: (1) selection of the multi models each with its own subset of predictors and (2) forecast from the individual models and combining them. The general form of the model for obtaining forecast is.

$$Y = f(\mathbf{x}) + \varepsilon \qquad [1]$$

[4] http://www.cdc.noaa.gov/Correlation/

where *Y* is the index gage spring season streamflow**, x** is a suite of predictors and ε is the residual that is assumed to be Normally distributed with mean 0 and variance σ – N(0, σ). 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, the procedures for parameter estimation, and hypothesis testing are very well developed (e.g., Helsel and Hirsch, 1995; Rao and Toutenburg, 1999) and are widely used. However, they do have some drawbacks: (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 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 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 - 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. Regonda et al. (2006) also use the LWP method, which will be adapted here.

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 readers 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 α=(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 like Mahalanobis distance (Yates et al., 2003) can also be used.
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)

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

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) is 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 or adjusted R^2 or AIC (Akaike Information Criteria) or an F-test, is calculated from the fitted model to several predictor combinations. The GCV based model selection is better since it is a good 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. (ii) All combinations with GCV values within a prescribed threshold (typically within 5% of the least GCV value, this can be user specified) are selected as admissible – constituting the pool of candidate models i.e., *multi-models.*
(iii) 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 (e.g., Wadsworth, 1990).Thus resulting in a set of *multi models* each with different predictor subset.

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. Thus, LWP can be viewed as a more flexible approach where the traditional linear regression is a subset.

Figure 5: Model selection process

4.3. Multi-model Ensemble Forecast of I Gage Seasonal Streamflow

The schematic of the multi-model ensemble forecast generation of Regonda et al. (2006) is shown in Figure 6.Suppose we desire a multi-model ensemble forecast for a point $\mathbf{x}_j$, the steps are as follows.

1. From one of the multi-models obtained from the preceding section obtain the mean estimate $\hat{Y}_i$ and theestimate of the error variance ($\sigma^2{}_{e_j}$) corresponding to the predicted value $\hat{Y}_i$ (Loader, 1999). This is from standard regression theory. Assuming Normality of the residuals, random normal deviates with this variance and zero mean, when added to predicted value, $\hat{Y}_i$, provide an ensemble.

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

where $N(0,\sigma_{e_j})$is a normally distributed random variable with mean 0 and standard deviation σ_{e_j}. Generate an ensemble of size *n* (we chose, 250).

2. Repeat step 1 for all the models in the multi-model suite. Thus, obtaining 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 member.

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

5. Evaluate the skill of the model using the ranked probability skill score (RPSS). The RPSS is described in section 5.

Figure 6: Multimodel Ensemble forecast procedure

4.4. Spatial and Temporal Disaggregation of the I Gauge Forecast

The seasonal forecast of the I gauge generated from above needs to be disaggregated in space (to the four locations) and in time (to the four months of the season) – thus, resulting in ensemble forecast for each month at all the four locations. This is achieved by employing the nonparametric space-time disaggregation technique proposed by Prairie et al. (2007) to the seasonal forecast. The schematic of the disaggregation method as applied to this study is shown in Figure 7. The process as it was implemented here will be discussed for the benefit of the reader.

Figure 7: Schematic Representation of Disaggregation process.

All the motivational and technical details of the disaggregation method are comprehensively described in Prairie et al. (2007). Here we provide a brief description of its implementation.

The disaggregation procedure can be thought of as sampling from the conditional probability density function (PDF), $f(\mathbf{x}|z)$, where **x** is d dimensional vector of flows, z is the aggregate flow, but with the (additvity) constraint that the values in **x** add up to z. This is achieved by an orthonormal transformation of the data **x** to **Y** and the simulation is performed in this space and back transformed (Tarboton, et al., 1998; Prairie et al., 2007).

The steps are described below for a temporal disaggregation (seasonal to monthly at the I gauge), they are identical for the spatial disaggregation.

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 **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.

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

$$\mathbf{Y} = \mathbf{RX}. \qquad [4]$$

Both **X** and **Y** are of dimension N (number of years) rows by d (=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 $\mathbf{y}_d = \mathbf{z}/\sqrt{d} = \mathbf{z}'$, 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

$$\mathbf{Y} = [\mathbf{U}, \mathbf{z}']. \qquad [5]$$

3. For a given seasonal flow forecast say z_{sim}, 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

$$W(k) = \frac{1}{k\sum_{i=1}^{K}\frac{1}{i}} \text{ where } k = 1,2,\ldots,K. \qquad [6]$$

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 = N where N equals the sample size (Lall and Sharma, 1996), following the asymptotic arguments of Fukunaga (1990).

5. Suppose the selected neighbor corresponds to year 'j' in history, the new vector $\mathbf{y}_{sim}$ is constructed as

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

6. This is back transformed to the original space as

$$\mathbf{x}_{sim} = \mathbf{R}^T \mathbf{y}_{sim}. \qquad [8]$$

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

The above steps can be applied to all the forecast ensembles. Each monthly flow can be subjected to the same procedure to obtain flows at the four locations. This ensures the additivity criteria in that for each month the flows at the I gauge is a sum of the flows at the four locations.

4.5. Verification and Validation

Since we generate an ensemble forecast, i.e., the probability density function (PDF), the skill of the forecast needs to be evaluated in probabilistic terms. One such common measure is the Ranked Probability Skill Scores (RPSS) (Wilks, 1995). Essentially, it measures the accuracy of 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' equals to one if the observation falls in the k^{th} category and zero otherwise. The ranked probability skill score (RPSS) is defined as follows:

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

$$\text{RPSS} = 1 - \frac{\text{RPS(forecast)}}{\text{RPS(climatolgy)}} \qquad [10]$$

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

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, which is common.The forecast skills were computed for three types of forecasts

1. Leave-one out forecast. In this each year is dropped the multi model ensemble for that year is generated from the rest. This is quite common (e.g., Grantz et al., 2005; Regonda et al., 2006; Singhrattna et al., 2005).
2. Leaving one year at a time does not stress the model adequately. Here we drop 10% of observations and the 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, thus obtaining an ensemble of skill scores. This provides a good insight in the variability of the performance.
3. To be able to compare with forecasts from CBRFC, 'retroactive' forecasts have to be performed. In this, forecast for a give year is made using all the data prior to that year – similar to 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. Predictors are identified separately for each lead time and the above methodology applied to obtain the ensemble forecasts. Below we present a representative set of results.

Predictor Identification

The index gauge spring season streamflow was correlated with large-scale ocean, atmosphere and land variables from preceding months. Figure 8 shows the correlation with Feb-Mar GPH, Mar ZNW, Jan-Mar SST and Mar MDW. We explored the correlations for different combination of months and this combination suggested a 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 are identified as potential predictors – these are detailed in Table 1.

Figure 8: Correlation maps of I gage streamflow with GPH (top left), ZW (top right), SST (bottom left) and MW (bottom right) for the Apr 1 prediction.

Table 1: Correlation regions used to develop predictors. The regions are given in lon:lat.

To understand the physical relationship between the large scale variables and the spring streamflow, composite maps of wet and dry years are developed. Figure 9 shows the composite maps of Oct-Mar vector winds for 6 wettest and driest years and the preceding fall (Sep-Nov) season PDSI for the same years. It can be seen clearly that 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. In that, wetter conditions in the fall favor less infiltration during the spring snow melt and therefore, enable enhanced streamflow – this can be seen in the PDSI composite and, vice-versa during dry years.

Figure 9: Composite maps of 500 mb vector winds for wet (top left) and dry (top right) years and PDSI for wet (bottom left) and dry (bottom right) years.

For SWE the leading principal component (similar to a basin average) was used – this is available from Feb 1st.

Multi Model Selection

Using the predictors identified in Table 1 and the methodology described in the previous section, multi models were selected for the different lead times - they are 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 hence the forecasts have to be made only from climate information. Consequently, individual models have greater uncertainty, and more models qualify as candidates for the multi-model pool. However, on April 1st, SWE information is complete and thus, it is the best predictor of the ensuing spring streamflow – thus, fewer models with other predictor variables are necessary, hence, a smaller number of multi-models.

Table 2: Selected Multimodels for each lead time. "1" indicates the presence of a predictor and "0" indicates the absence of a predictor. *This Jan 1st prediction used the Jan 1st and Nov 1st GPH as separate predictors.

Forecast Skill

The leave-one out cross validated ensemble forecast of the index gauge spring season streamflow issued on April 1st is shown in Figure 10, as boxplots. The box height corresponds to the interquartile range, the whiskers 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, as attested by a very high RPSS score (of 0.95). Using the space-time disaggregation technique these forecasts were translated to ensemble forecasts at the four locations. Figure 11 shows the leave-one out cross validated ensemble forecast for Lees Ferry, indicating that forecast skill of the index gauge is well preserved to the spatial locations. The results were similar at other locations. The RPSS 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 forecasts to obtain ensembles of monthly forecasts also yielded good skills – this can be seen in the boxplot of the May monthly ensemble forecast at Lees Ferry in Figure 12. The RPSS of the monthly forecasts are also shown in Table 3. It needs to be underscored that the disaggregated forecasts provide significant skill at longer lead times – that is the water managers can obtain a good sense of the spring streamflow few months in advance which will be of immense use in efficient water resources management. As to be expected in all the forecasts the skills decrease with increase in lead time.

Figure 10: Leave-one CV predictions for the I gage Apr. 1st prediction. Whiskers extend to 5th and 95th percentile of each ensemble. The dashed horizontal lines correspond to the 33rd, 50th and 66th percentiles of the data.

Figure 11: Retroactive forecast for the Index gage, Apr. 1st prediction.

Figure 12: Retroactive forecast for Lees Ferry May flows after spatial and temporal disaggregation, Apr. 1st prediction. Forecasts are compared to the CBRFC predictions (dashed line).

To challenge the forecasting system, we performed forecasts by dropping 10% of the observations and repeating it 100 times. The RPSS skills on the dropped observations are shown as boxplots in Figure 13. There is considerable variablity in the skill scores due to sampling variability – but, the median skill scores are quite high at all the locations except Bluff for the reasons mentioned above.

Figure 13: Drop 10% Cross Validation Skill Scores.

One of the advantages of the disaggregation method is that it can capture the spatial correlation in a parsimonious manner. To evaluate this, boxplots of cross correlation between the forecasted flows at the four locations are computed and compared with the historical values in Figure 14. It can be seen that the leave-one out cross validated forecasts capture the spatial correlation very well.

Figure 14: Cross correlation between sites: C=Cisco, B=Bluff, G=GRUT, LF=Lees Ferry. The solid line represents the observed statistics and the boxplots represent the statistics reproduced with the disaggregation procedure.

In real life application forecasts are made one year at a time using all the prior data – 'retroactive forecast'. This also enables us to compare with the CBRFC forecasts, which are issued in this manner but based on a physical watershed model. We performed this for the 1990-2005 period for which the CBRFC forecasts are available for comparison. Figure 15 (a), (b) shows the forecasts of Lees Ferry seasonal streamflow issued on Jan 1st and Apr 1st, respectively. The forecast from our approach is comparable to the CBRFC forecasts and in some years, especially the wet years, better. Also, the proposed approach performs much better at longer lead time. Similar observations can be made for the May forecasts at Lees Ferry issued on Apr 1st (Figure 12).

Figure 15: Seasonal Forecast at Lees Ferry issued on (a) Jan. 1st and (b) Apr. 1st. Forecasts are compared to the CBRFC predictions (dashed line).

Summary

Acknowledgements

The first author performed this work during a summer fellowship funded by the NSF REU program, he is thankful for that.

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*.

CBRFC (2006), National Weather Service-Colorado Basin River Forecast Center, US Department of Commerce.

Craven, P., and G. Wahba (1979), Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross validation, *Numerical Mathematics*, 31, 377–403.

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

Fukunaga, K. (1990), *Introduction to Statistical Pattern Recognition*, Academic Press, New York.

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 Resources Research*, 41, 13 pp.

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. Series A, Dynamic Meteorology and Oceanography*, 57, 219 - 233.

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

Hwang, Y. (2005), *Impact of input uncertainty in ensemble streamflow generation*, Ph.D.

thesis, University of Colorado, Boulder.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. Mo, C. Ropelewski, J. Wang, R. Jenne, and D. Joseph (1996), The NCEP/NCAR 40-year reanalysis project, *Bulletin of the American Meteorological Society*, 77, 437–471.

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

Krishnamurti, T. N., C. M. Kishtawal, Z. Zhang, T. E. LaRow, D. R. Bachiochi, C. E. Williford, S. Gadgil, and S. Surendran (2000), Multimodel ensemble forecasts for weather and seasonal climate, Journal of Climatology, 13, 4196 - 4216.

Lall, U. (1995), Recent advances in nonparametric function estimation: Hydraulic applications, *Rev. Geophys*, 33, 1093 – 1102, National Report For The United States Of America International Union Of Geodesy And Geophysics 1991 - 1995.

Lall, U., and A. Sharma (1996), A nearest neighbor bootstrap for time series resampling, *Water Resources Research*, 32, 679–693.

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

Opitz-Stapleton, S., S. Gangopadhyay, and B. Rajagopalan (2007), Generating streamflow forecasts for the Yakima River Basin using large-scale climate predictors, *Journal of Hydrology*, 341, 131–143.

Prairie, J., and R. Callejo (2005), *Natural flow and salt computation methods*, Tech. rep., Bureau of Reclamation.

Prairie, J., B. Rajagopalan, U. Lall, and T. Fulp (2007), A stochastic nonparametric technique for space-time disaggregation of streamflows, *Water Resources Research*, 43, 10 pp.

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

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

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 Resources Research*, 42, 14 pp.

Singhrattna, N., B. Rajagopalan, M. Clark, and K. K. Kumar (2005), Forecasting Thailand summer monsoon rainfall, *International Journal of Climatology*, 25, 649–664.

Tarboton, D. G., A. Sharma, and U. Lall (1998), Disaggregation procedures for stochastic hydrology based on nonparametric density estimation, *Water Resources Research*, 34, 107–119.

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

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 Resources Research*, 39(7), 1199.

Figures

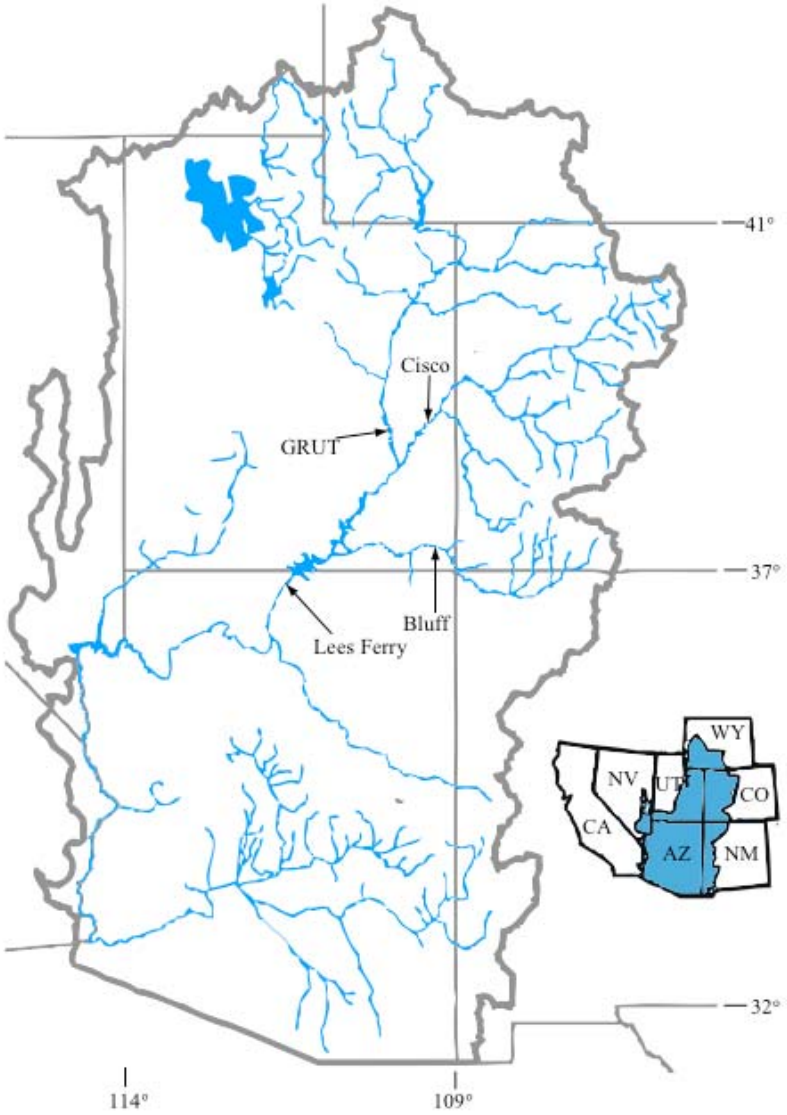

Figure 1: Study area.

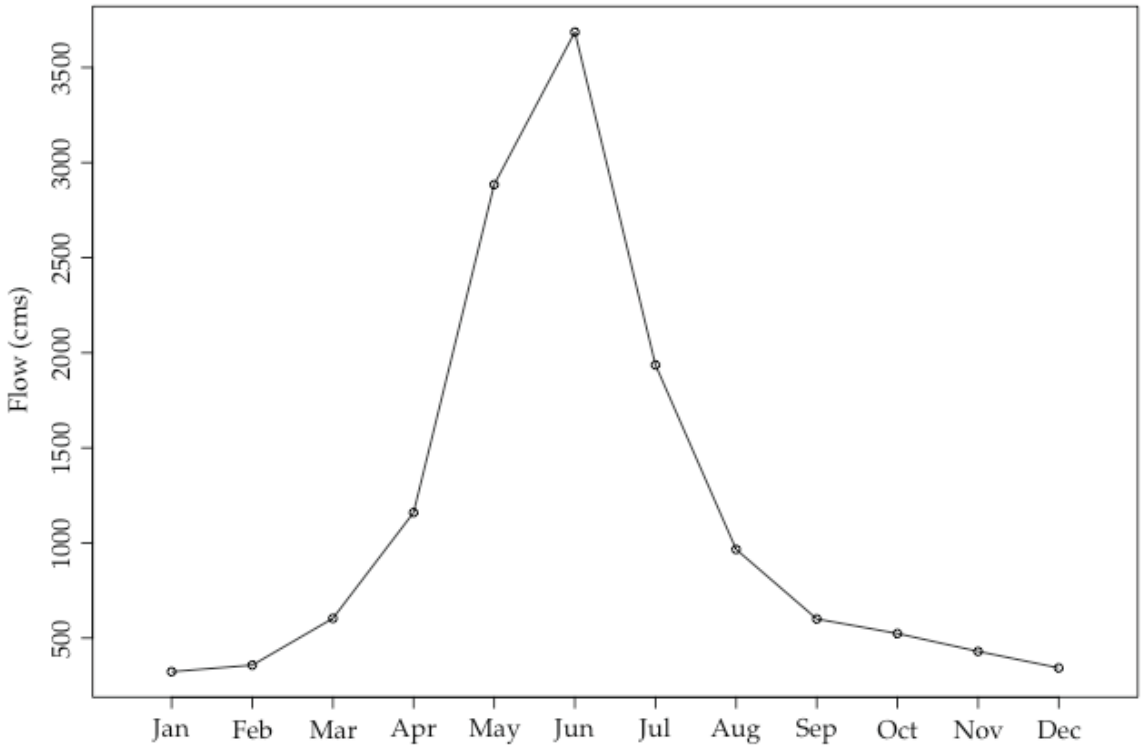

Figure 2: Index gage seasonal average hydrograph

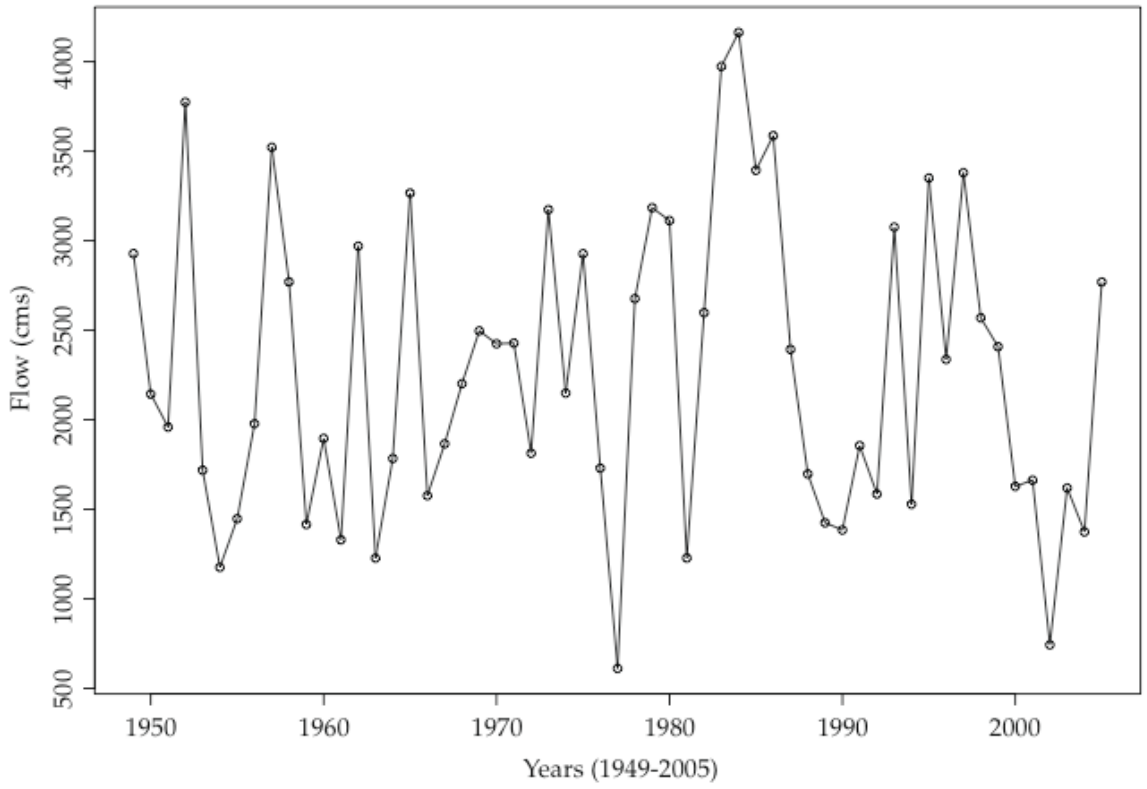

Figure 3: Seasonal average (April-July) flows.

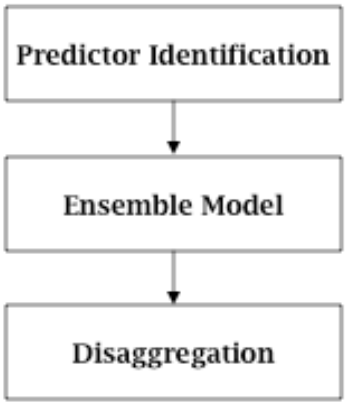

Figure 4: Framework

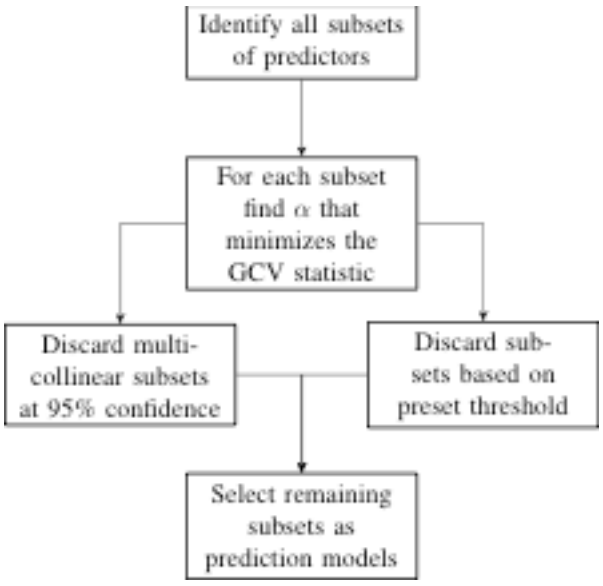

Figure 5: Model selection process

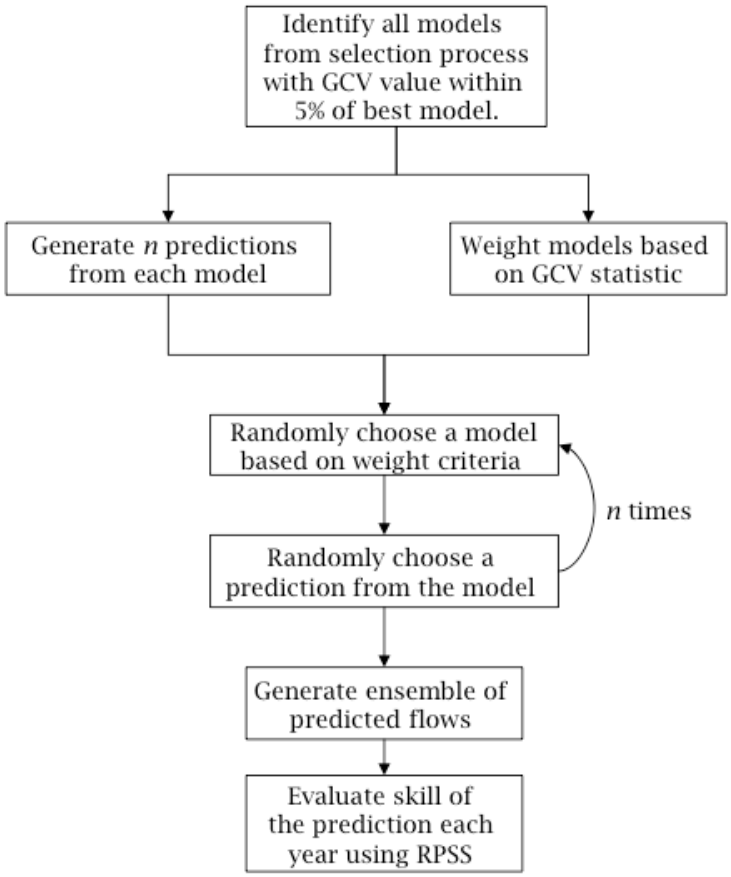

Figure 6: Multimodel Ensemble forecast procedure

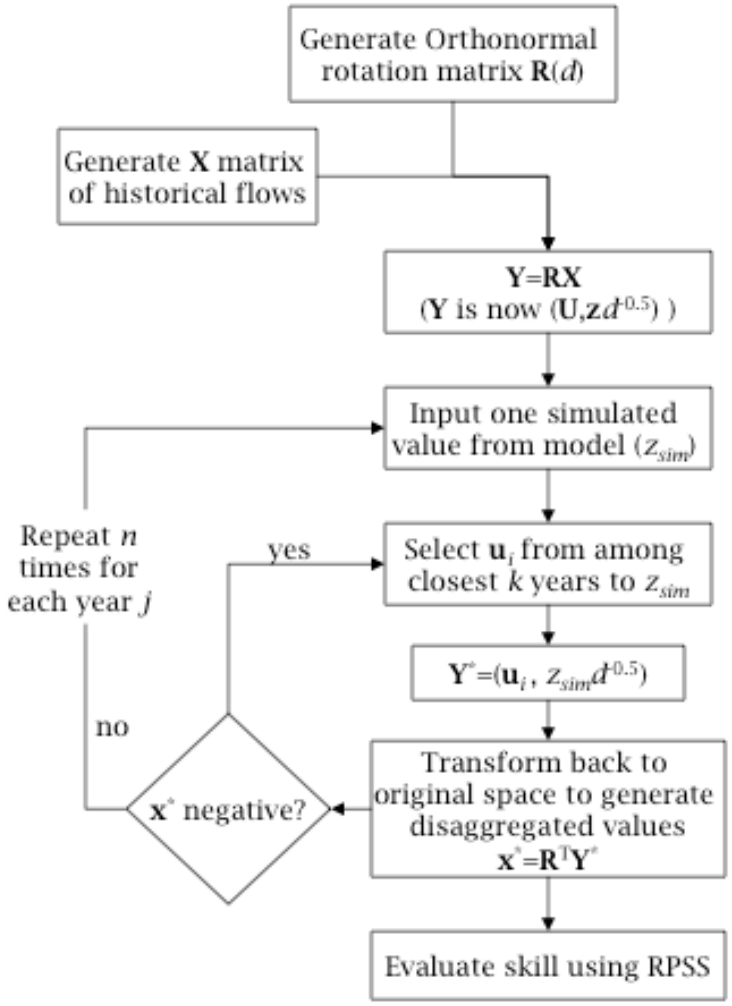

Figure 7: Schematic Representation of Disaggregation process.

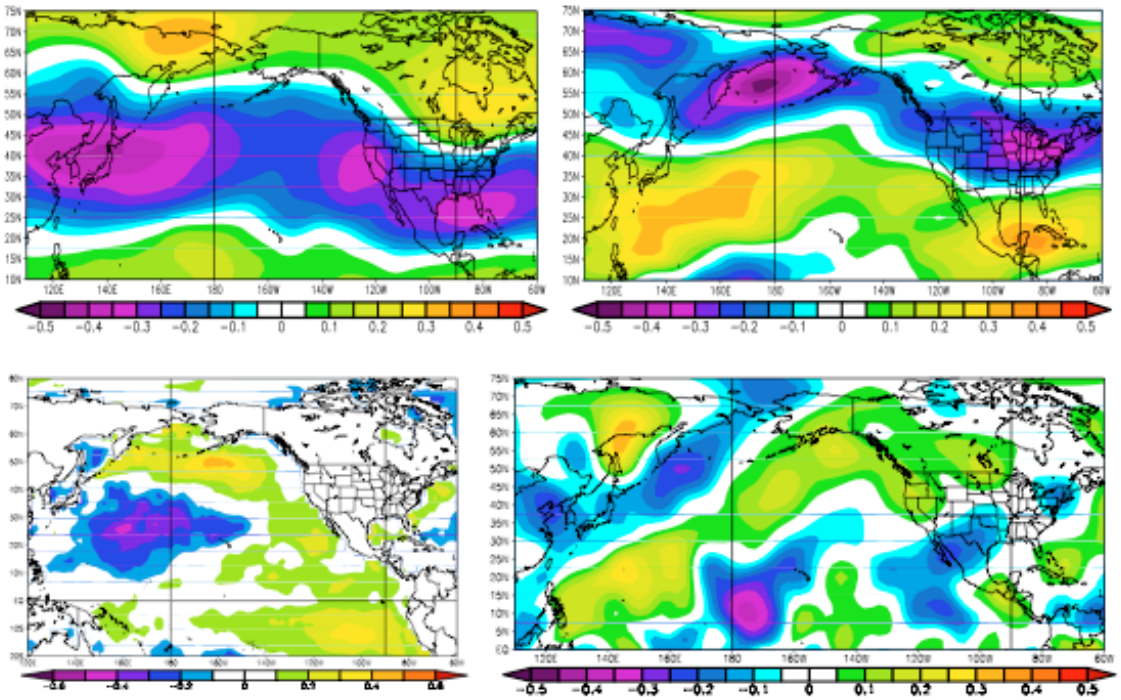

Figure 8: Correlation maps of I gage streamflow with GPH (top left), ZW (top right), SST (bottom left) and MW (bottom right) for the Apr 1 prediction.

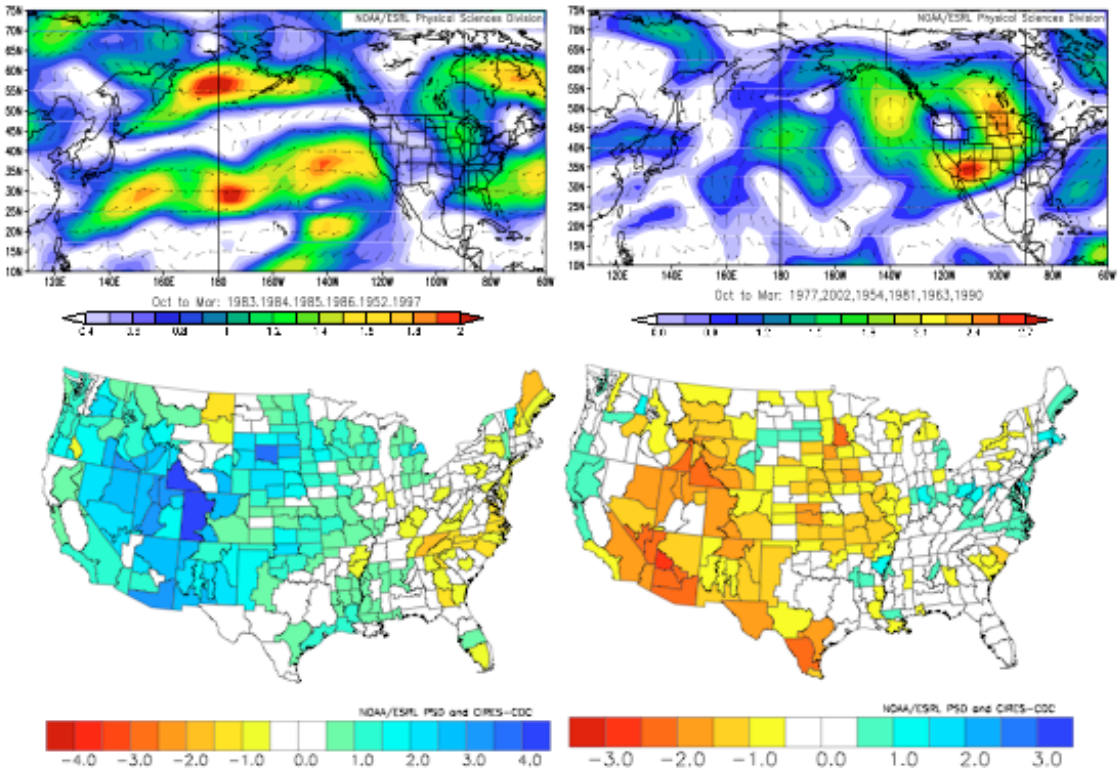

Figure 9: Composite maps of 500 mb vector winds for wet (top left) and dry (top right) years and PDSI for wet (bottom left) and dry (bottom right) years

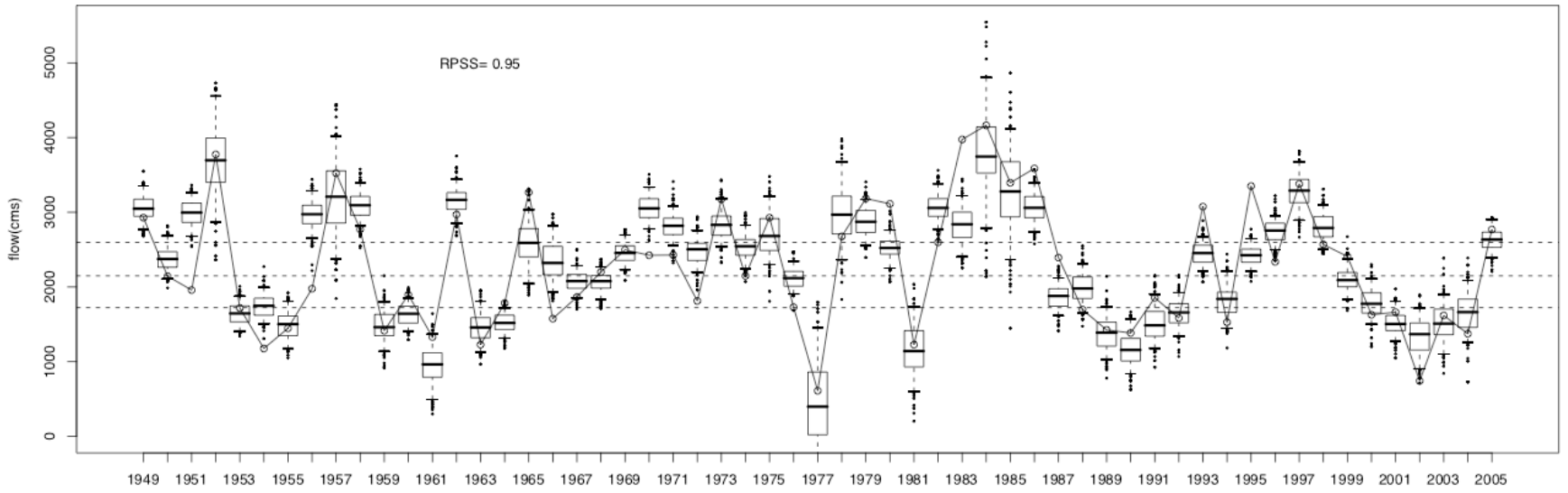
RPSS= 0.95
flow(cms)
0
1000
2000
3000
4000
5000
1949
1951
1953
1955
1957
1959
1961
1963
1965
1967
1969
1971
1973
1975
1977
1979
1981
1983
1985
1987
1989
1991
1993
1995
1997
1999
2001
2003
2005

Figure 10: Leave-one CV predictions for the I gage apr1 prediction. Whiskers extend to 5th and 95th percentile of each ensemble. The dashed horizontal lines correspond to the 33rd, 50th and 66th percentiles of the data.

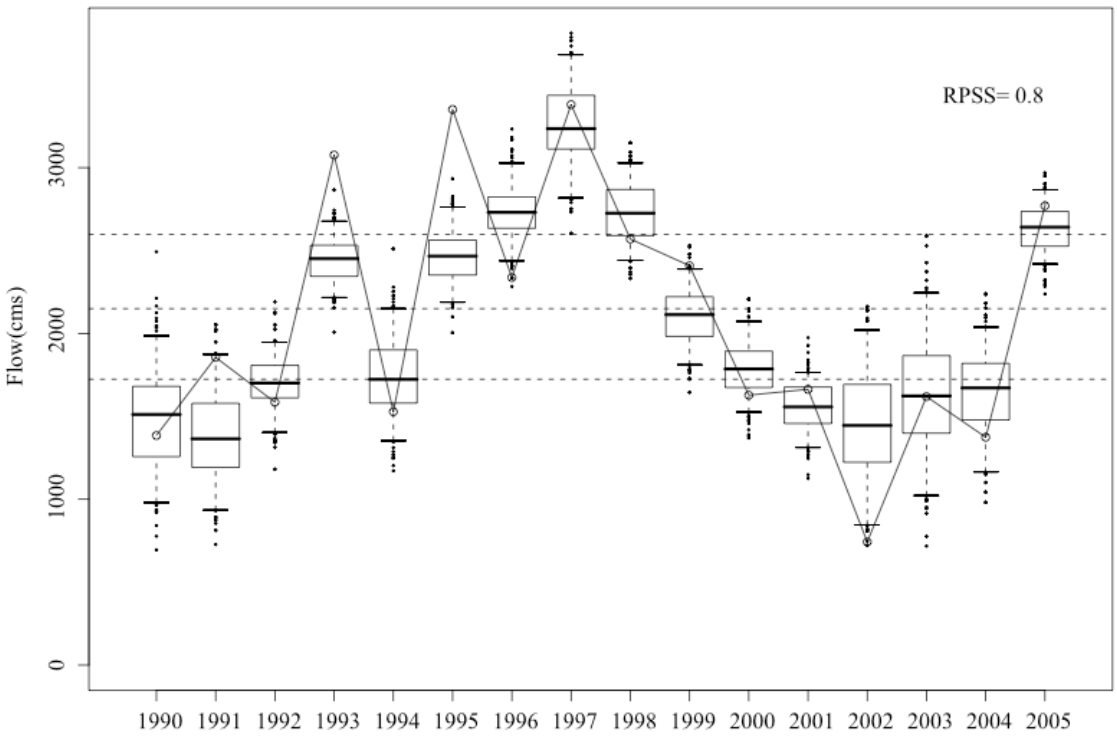

Figure 11: Retroactive forecast for the Index gage, Apr. 1st prediction.

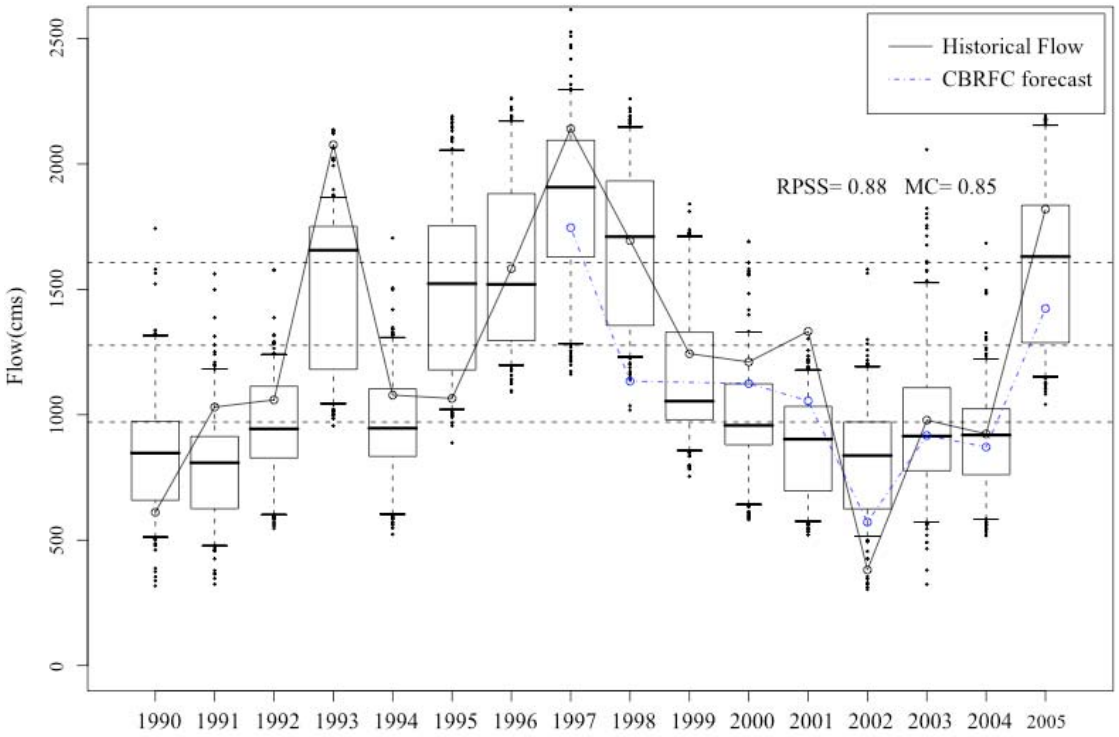

Figure 12: Retroactive forecast for Lees Ferry May flows after spatial and temporal disaggregation, Apr. 1st prediction. Forecasts are compared to the CBRFC predictions (dashed line).

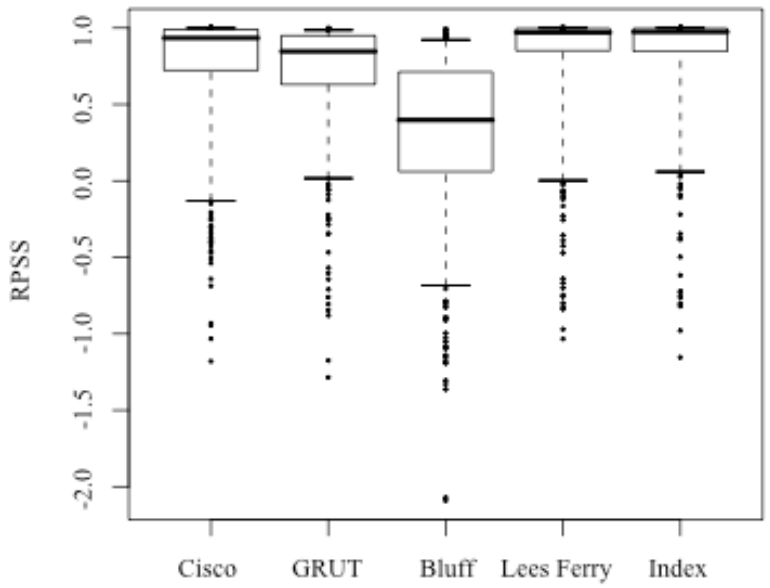

Figure 13: Drop 10% Cross Validation Skill Scores.

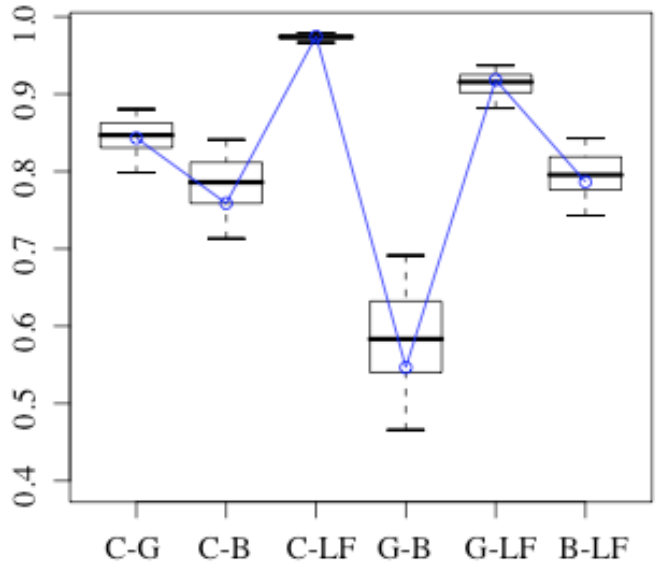

Figure 14: Cross correlation between sites: C=Cisco, B=Bluff, G=GRUT, LF=Lees Ferry. The solid line represents the observed statistics and the boxplots represent the statistics reproduced with the disaggregation procedure.

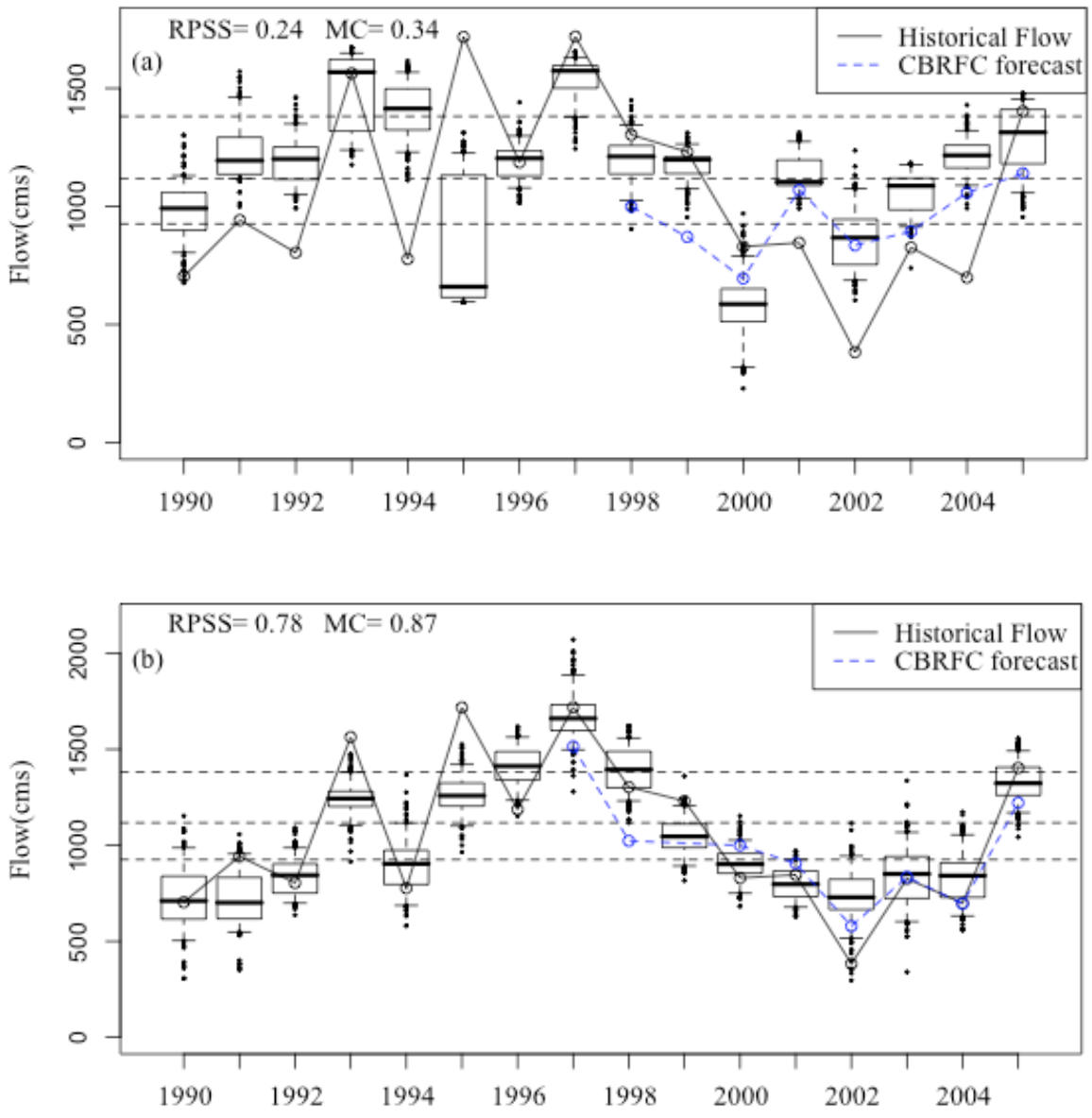

Figure 15: Seasonal Forecast at Lees Ferry issued on (a) Jan. 1st and (b) Apr. 1st. Forecasts are compared to the CBRFC predictions (dashed line).

Tables

Table 1: Correlation regions used to develop predictors. The regions are given in lon:lat.

Variable	Lead Time	Season	Negative Region	Positive Region	Max Cor.
GPH	Nov. 1st	oct	40,42N:137,140E	17,19N:178,182E	0.39
ZNW	Nov. 1st	oct	40,42N:-108,-106E	30,32N:132,135E	0.43
MDW	Nov. 1st	oct	40,42N:124,126E	40,42N:124,126E	0.31
SST	Nov. 1st	oct	-17,-15N:-96,-92E	38,40N:170,173E	-0.26
GPH	Jan. 1st	oct-dec	42,45N:-115,-110E	52,58N:-160,-155E	-0.56
ZNW	Jan. 1st	oct-dec	54,55.5N:-119,-114E	32,34N:-110,-106E	-0.58
MDW	Jan. 1st	oct-dec	45,50N:-135,-130E	42,45N:-91,-88E	-0.57
SST	Jan. 1st	oct-dec	28,31N:161,164E	46,49N:-166,-163E	-0.63
GPH	Feb. 1st	oct-jan	60,62N:128,132E	61,64N:-153,-150E	0.24
ZNW	Feb. 1st	oct-jan	72,74N:126,128E	28,31N:-119,-116E	0.42
MDW	Feb. 1st	oct-jan	66,68N:-108,-105E	38,42N:-90,-85E	0.3
SST	Feb. 1st	oct-jan	25,28N:160,163E	46,48N:-167,-164E	-0.64
GPH	Apr. 1st	feb-mar	40,42N:137 ,140E	65,70N:160,165E	-0.47
ZNW	Apr. 1st	mar	56,58N:172,176E	25,30N:155,160E	0.46
MDW	Apr. 1st	mar	10,12N:-177,-174E	56,60N:142,144E	-0.42
SST	Apr. 1st	jan-mar	24.5,26N:158,162E	48,51N:-165,-160E	-0.54

Table 2: Selected Multimodels for each lead time. “1” indicates the presence of a predictor and “0” indicates the absence of a predictor. *This Jan. 1st prediction used the Apr. 1st and Nov. 1st GPH as separate predictors.

Lead Time	Predictors	PDSI	GPH	ZNW	MDW	SST	SWE
Apr. 1st	2	1	0	0	0	0	1
Feb. 1st	2	0	0	0	0	1	1
Feb. 1st	2	1	0	0	0	0	1
Feb. 1st	3	0	0	0	1	1	1
Jan. 1st *	3	0	1	1	0	0	0
Jan. 1st	2	0	1	1	0	0	0
Nov. 1st	2	1	1	0	0	0	0
Nov. 1st	1	1	0	0	0	0	0

Table 3: Index gage forecast skill and spatially disaggregated forecast skills.

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