A Multi-Site Streamflow Forecast Framework: Application to the Upper Colorado River Basin

Cameron Bracken
Advisor: Balaji Rajagopalan
Research Experience for Undergraduates
August 10, 2007

Abstract

We present a framework for providing ensemble streamflow forecasts at several locations simultaneously on a river network. The framework is an integration of two recent approaches Ð 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 this index gauge streamflow 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 disaggregation technique [Prairie et al. 2007] to forecast at all the desired locations. We demonstrate the utility of this framework in the skillful forecast of spring (Apr-Jul) seasonal streamflows at four locations in the Upper Colorado River Basin at several lead times. The skills 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 valueable input for efficient planning and management of water resources in the basin.

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 that usual. Accurate forecasts at several spatial locations are also desirable for efficient 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 is due to snowmelt and as a result streamflow models have long been primarily dominated by this [CBRFC 2006; Grantz et al. 2005]. Strong

evidence exists showing the utility of large-scale climate information in making skillful predictions in December-February when snowpack information is limited at best [Grantz et al. 2005]. In an application to the Truckee-Carson River system, Grantz et al. [2005] showed skill in November and December, as well as improvements over the snowpack-only predictions in January-April. Regonda et al. [2006] showed similar skills in an application to the Gunnison River Basin (GRB).

Previously we mentioned the need for seasonal forecasts at several locations as a tool for efficient reservoir management. Including large-scale climate factors presents another challenge in that making predictions at each site within a basin separately does not preserve the spatial interactions between sites, which are inherent in a river network. Regonda et al. [2006] used principal component analysis (PCA) to make predictions at 6 gages in the GRB. Two drawback of this method is that it does not preserve the summability criteria (i.e. downstream gage is the sum of upstream gages) or capture arbitrary nonlinearities which may be present in historical data. A K nearest neighbor (KNN) method of nonparametric disaggregation, which preserves the summability criteria and captures nonlinearities, has been implemented with success on the UCRB [Prairie et al. 2007; Prairie 2006]. The reader is referred to Prairie et al. [2007] and references therein for more information on parametric and nonparametric streamflow disaggregation.

A simple streamflow forecasting method is needed which takes into account large-scale climate variability and spatial interactions between sites within a single river basin. A methodology is presented here which includes the use of large-scale climate predictors and disaggregation in a linked framework to predict the seasonal and monthly streamflow at four key sites in the UCRB. This paper will first provide preliminary information on the study area and the data sets. Next, An outline of the framework is given followed by a description of the multi-model ensemble (MME) forecast and the disaggregation procedure. The application of the framework to the UCRB is given 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 mi2 (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 comes from snowmelt in the UCRB whereas most of the water use occurs in the lower colorado river basin (LCRB). This places particular importance on the skillful forecasts of streamflows in the upper basin. To demonstrate the forecast framework proposed in this research, we chose four key locations on the UCRB network. The four sites shown in Figure 1 are: (1) Colorado River near Cisco, Utah (Cisco); (2) Green River at Green River, Utah (GRUT); (3) San Juan River near Bluff, Utah (Bluff); and (4) Colorado River at Lees Ferry, Arizona (Lees Ferry). Lees Ferry gage at the outflow of Lake Powell is a very important gage primarily because it is the separation between the UCRB and the LCRB.

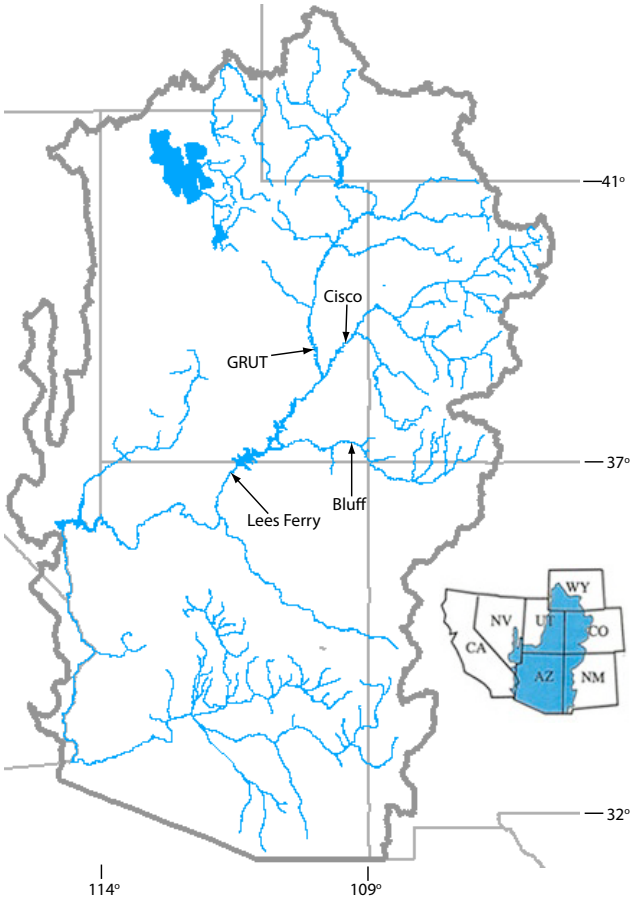


Figure 1: Study area.

3 Data

3.1 Streamflow Data and Index Gage

Monthly natural streamflows at the four locations are available for the 98 year period spanning 1906 Ð 2006. We used the 1949-2005 period for this study. Naturalized streamflows are computed by removing anthropogenic impacts (i.e., reservoir regulation, consumptive water use, etc.) from the recorded historic flows[1]. From this the Spring seasonal (April-July) flows were computed at each site. The seasonal averages were summed for all four of the sites creating an Ôindex gageÕ denoted as ÔIÕ, henceforth. The monthly average and the time series of the seasonal flow at I are shown in Figures and , respectively. Almost all of the annual flows occur during the spring months (Figure). The index gauge is useful for disaggregation, which will become clear in the followinig sections where the framework is described.

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

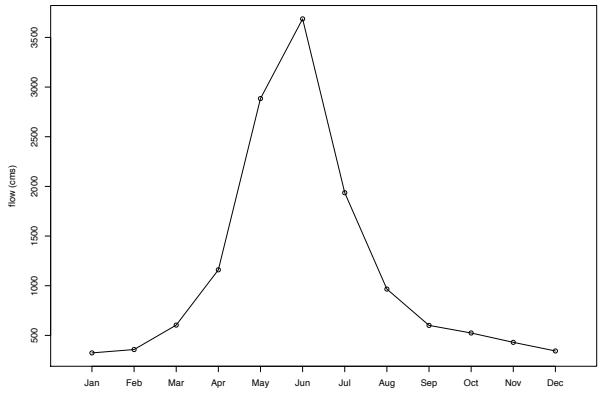


Figure 2: Index gage seasonal average hydrograph

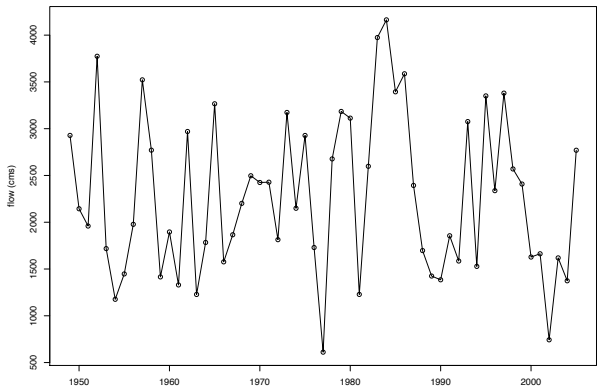


Figure 3: Seasonal average (April-July) flows.

3.2 Large-scale climate data

Ocean-atmospheric circulation variables that capture the large-scale climate forcings are available from NOAAÕs Climate Diagnostics Center Web site (`http://www.cdc.noaa.gov`. 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°×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 Conservation (NRCS) from their website `http://www.wcc.nrcs.usda.gov/snow`. 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.

3.4 PDSI data

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

The multi-site framework consists of (Figure 4)

1. Identifying large-scale climate predictors using climate diagnostics [Grantz et al. 2005],
2. Generating MME (probabilistic) forecasts [Regonda et al. 2006],
3. Spatially and temporally disaggregating the I gage predictions [Prairie et al. 2007].

The following section will describe each step of the framework in detail.

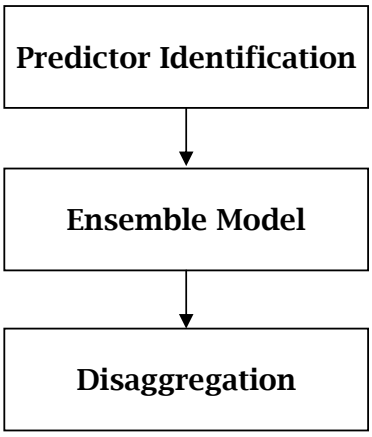


Figure 4: Framework

4.1 I Gage Predictor Identification

The National Oceanic and Atmospheric Administration (NOAA) linear correlation tool was used to identify regions of high correlation to the seasonal streamflow at the I gage. Area averages of the climate variables over these regions were used as potential predictors of seasonal flow. In this way a suite of potential predictors were identified.

4.2 Multi-Model Ensemble Forecast of I Gage Flows

The MME methodology consists of two distinct steps: (1) multi-model selection and (2) multi-model prediction. The general form of the model for every year t is

$$y_t = f(\mathbf{x}_t) + \varepsilon \tag{1}$$

where y_t is the streamflow, $\mathbf{x}_t$ is a suite of predictors and $\varepsilon \sim N(0, \sigma_t)$. If f is estimated globally and linearly then the model is traditional linear regression. Here we use a locally weighted polynomial (LWP) model.

4.2.1 Multi-Model Selection

Multi-models have been shown to preform better than single models in streamflow prediction scenarios [Regonda et al. 2006]. For this reason we adopt a multi-model approach. The multi-model selection process can proceed once all the predictors at a particular lead time have been identified (Figure 5). The general process involves identifying a feasible set of predictors and then narrowing down that set based on various objective criteria. The process is as follows:

1. Identify all subsets of predictors using the package `leaps` [Lumley 2007]. The function `leaps()` is set to identify up to 250 subsets with the same number of predictors (4 or less). The subsets are chosen to minimize Mallow's C_p statistic which is considered a goodness of fit measure [Wadsworth 1990].
2. Two parameters, the bandwidth, α (where $\alpha \in (0,1]$), and the degree, p (where $p \in 1,2$), of a locally weighted polynomial (LWP) fit are determined which minimize the generalized cross-validation (GCV) statistic. The bandwidth α determines the number of neighbors, K, used in the LWP model, where $\alpha = K/N$ and N is the sample size. The GCV statistic is given by
$$\mathrm{GCV}(\alpha, p) = N \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)}{(N - \nu)^2} \tag{2}$$
where ν is the fitted degrees of freedom. The GCV statistic is a good measure of predictive risk when working with large data sets [Loader 1999]. For more on LWP the reader is referred to Loader [1999], Prairie [2002], Grantz et al. [2005] and Regonda et al. [2006]. Note that if $\alpha = p = 1$, the LWP method colapses to traditional linear regression.
3. Once the optimal parameters (α, p) are determined for every subset the number of models must be reduced. Models are first discarded which are multicollinear (ie. predictors highly correlated among themselves) at 95% confidence. Multicollinearity can lead to overfitting and poor predictive skill [Wadsworth 1990].
4. Additionally models are discarded based on a preset threshold which allows only 15 models for any number of variables.
5. The remaining subsets of predictors (i.e. models) are selected as suitable for predictive use.

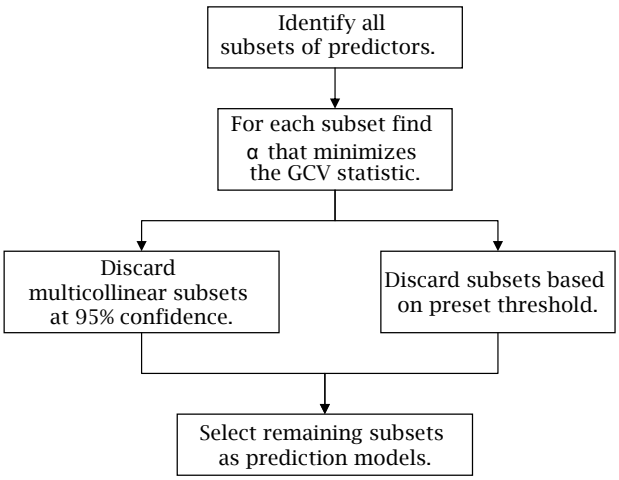


Figure 5: Model selection process

4.2.2 I Gage Ensemble Generation

Using the models chosen in the selection process, those models with a GCV value within 5% of the model with the lowest GCV are chosen as the final set of multi-models. Figure 6 gives an overview of the MME. The MME forecast proceeds as follows:

1. Generate n predictions from each model j by resampling (i.e. bootstrapping) the standard error of the LWP fit, σ_{e_j}, using the `locfit` package. Each simulated flow, $z_{i,j}$, is given by

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

   where $N(0, \sigma_{e_j})$ is a normally distributed random variable with mean 0 and standard deviation σ_{e_j}. The value $\hat{y}_i$ is the LWP model value at a given vector predictor values $\mathbf{x}_j$ and $i \in 1...n$.
2. Weight all the multi-models based 1/GCV criteria. This way the model with the lowest GCV value is most heavily favored.
3. Randomly choose a model based on the the weight criteria.
4. Randomly choose one of the n predictions from the model chosen in step 3. Repeat steps 3 and 4 n times to generate a ensemble of predictions.
5. Evaluate the skill of the model using the ranked probability skill score (RPSS). The RPSS is described in section 5.

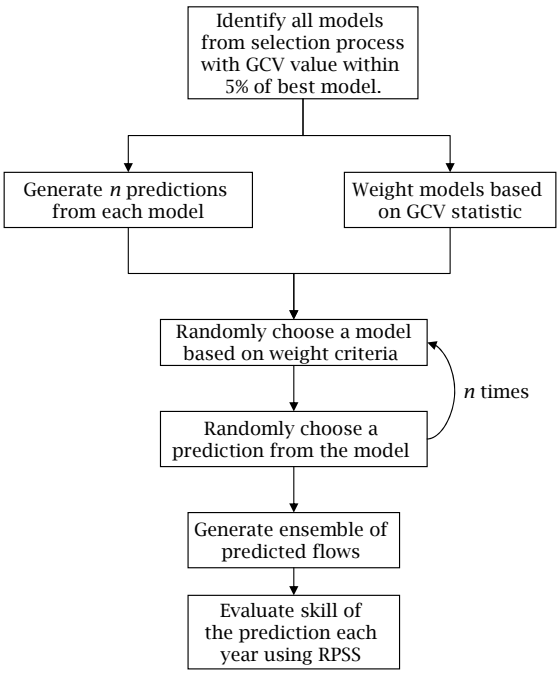


Figure 6: MME forecast

4.3 Disaggregation of the I Gage Forecasts

Once suitable ensembles are generated, temporal and spatial disaggregation can be preformed. Specifically, predictions are generated of the seasonal flows at the I gage which are used to generate seasonal predictions at the four sites, which in turn are used to generate monthly flows for April-July (Figure 7). The disaggregation procedure follows that of Prairie et al. [2007]. Again, Prairie et al. [2007] makes an extensive review of disaggregation techniques as well as showing simple example of the modified KNN disaggregation procedure. The process as it was implemented here will be discussed for the benefit of the reader.

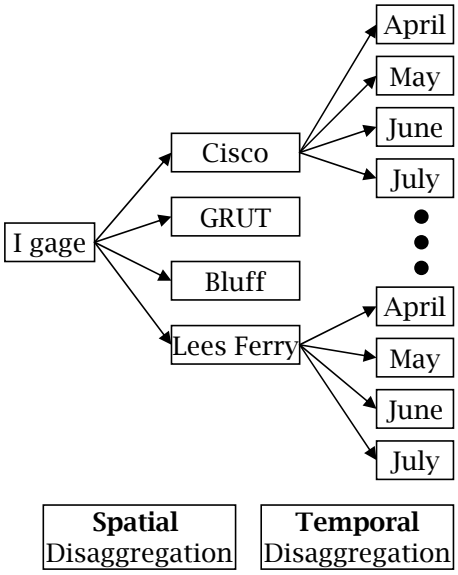


Figure 7: Disaggregation Schematic

The disaggregation procedure can be thought of as sampling from the conditional probability density function (PDF), $f(\mathbf{x}|z)$, where $\mathbf{x}$ is d dimensional a vector of flows, z is a simulated flow and $\sum_{i=1}^{d} x_i = z$. The conditional PDF is given by

$$f(\mathbf{x}|z) = \frac{f(\mathbf{x}, z)}{\int f(\mathbf{x}, z) d\mathbf{x}}. \tag{4}$$

In practice this PDF is not actually used because it involves constructing a d+1 dimensional function. Actually, all of useful information (that which preserves the additive constraint) is situated along the d-1 dimensional hyperplane given by

$$x_1 + x_2 + \cdots + x_d = z. \tag{5}$$

The hyperplane can be thought of as a slice thought the d-1 dimensional PDF, $f(\mathbf{x})$. The PDF can simulated from when $\mathbf{x}$ is rotated into the vector $\mathbf{y}$ which has its last coordinate aligned with the hyperplane given in equation 5. This process is very similar to a PCA in which the the data is rotated using the eigenloadings. In the rotated space, which preserves the additivity criteria by definition, the simulation is now done from the PDF, $f(\mathbf{y}|z)$.

The procedure is as follows:

1. The first step is to generate the orthonormal rotation matrix, $\mathbf{R}(d)$ using the Gram-Schmidt algorithm. Note that the $d \times d$ matrix $\mathbf{R}$ is only a function of the 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. Obtain the the matrix , $\mathbf{X}$, of historical flows and rotate into the matrix $\mathbf{Y}$ by

$$\mathbf{Y} = \mathbf{R}\mathbf{X} \tag{6}$$

where $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_t)$, $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_t)$ and t is the number of years in the simulation. For example if disaggregating to four spatial location, $d = 4$ and $\mathbf{X}$ will contain flows for each site down the rows and each year across the columns. $\mathbf{Y}$ is now $(\mathbf{U}, \mathbf{z}d^{-0.5})^T$. Where $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_t)$.

3. One simulated value, z_{sim}, from the MME to be disaggregated is selected.
4. One $\mathbf{u}_i$ corresponding to year i is randomly selected from among the closest K years to z_{sim} based on the weight function

$$W(k) = \frac{1}{k\sum_{i=1}^{K}\frac{1}{i}} \text{ where } k = 1, 2, ...K. \tag{7}$$

5. The new vector $\mathbf{y}^*$ is constructed as $\mathbf{y}^* = (\mathbf{u}_i, z_{sim}d^{-0.5})$.
6. Transform back to the original space and generate the vector of disaggregated flows (which preserve additivity) by

$$\mathbf{x}^* = \mathbf{R}^T\mathbf{y}^* \tag{8}$$

7. If any of the disagregated flow are negative, return to step 4 (the resampling) until all disaggregated values are positive or a maximum number of iterations are reached. If all the elements of $\mathbf{x}^*$ are positive or the maximum iterations are reached return to step 3 and repeat for all the simulated values.
8. Evaluate the skill of the disaggregation using the RPSS as before in the MME.

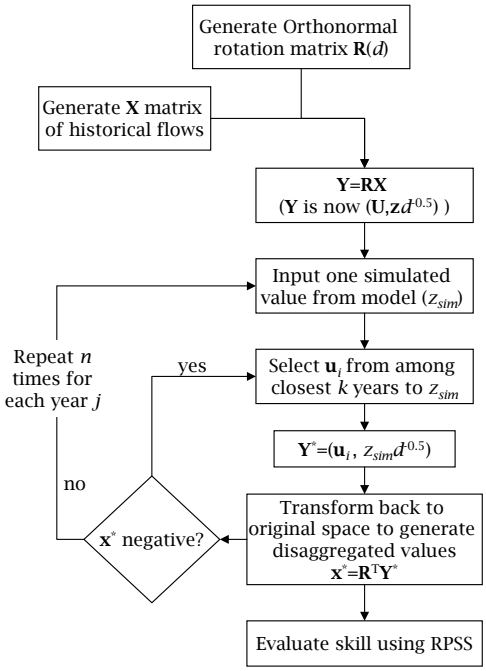


Figure 8: Disaggregation procedure

5 Verification and Validation

Many methods exist for verifying statistical models. For weather forecasting some commonly used methods are drop-one CV (CV), leave k out CV, retroactive forecasting, jacknifing, and bootstrapping [Mason 2004]. These same techniques are readily applied to streamflow forecasting. In this report we use these drop-one CV, retroactive forecasting and a modified leave k out CV. The general CV procedure as described by Mason [2004] is:

1. Leave at least one year out of the training sample.

2. Reconstruct the model using the new smaller training sample.
3. Forecast at least one of the years omitted.
4. Repeat at least step 3.

Drop-one CV is commonly used as a quick measure of the overall forecast skill and as the name suggests, involves dropping one year out of the forecast and then predicting that year. Drop-one CV is the method implemented by Grantz et al. [2005] and Regonda et al. [2006]. For a more rigorous test of the model, a modified leave k out CV which involves randomly dropping 10-15% (here we used 12.3% which translates to dropping 7) of the years, and predicting them as in the CV procedure. When large data sets are available, retroactive (i.e. realistic) forecasting can be used to make predictions using only data from previous years to predict the current year. Retroactive forecasting is desirable in the current context because the predictions made in this mode can be compared to past CBRFC predictions.

A commonly used metric in ensemble forecasting is the RPSS. The RPSS is a categorical measure of ensemble forecast skill in relation to the climatological prediction. The RPSS is calculated by first dividing the historical flows into k mutually exclusive and collectively exhaustive categories. For every year the proportion of predictions falling into each categories is calculated $(p_1, p_2, ...p_k)$, as well as the observational vector $(d_1, d_2, ...d_k)$, in which $d_i = 1$ if the historical flow for the given year falls in category i and $d_i = 0$ otherwise. The RPSS is then given by

$$\mathrm{RPS} = \sum_{j=1}^{k} \left[\left(\sum_{s=1}^{j} p_s - \sum_{s=1}^{j} d_s \right)^2 \right] \tag{9}$$

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

The RPSS may take any value (-∞,1], where 1 indicates a perfect forecast, 0 indicates no improvement over climatology and negative values indicate a worse prediction than climatology. For example an RPSS value of 0.87 represents an 87% improvement over a climatological prediction. Here we use three categories ($k = 3$) so the climatological prediction will always be 1/3 in each category.

Additionally we evaluate skill using the median correlation (MC). The MC is simply the correlation between the median ensemble forecast value of each year and the observed flow values for the prediction period (1949-2005).

6 Results

For this study, four lead times were chosen for prediction: November 1 (nov1), January 1 (jan1), February 1 (feb1), and April 1 (apr1). The nov1 and jan1 predictions only rely on large-scale climate predictors where the feb1 prediction relies on partial SWE data and the apr1 prediction can utilize the complete SWE information.

6.1 Climate Diagnostics and Predictor Identification

Large-scale climate variables were correlated with the I gage seasonal flow and regions of high positive and negative correlation were identified. Figure 9 shows the correlation maps for the apr1 prediction. Earlier lead times showed similar but weaker trends. Months prior to each lead time were used to develop predictors which were positively and negatively correlated for each variable and lead time. The positive predictors were subtracted from the negative predictors to obtain a more highly correlated predictor for a particular variable (Table 1).

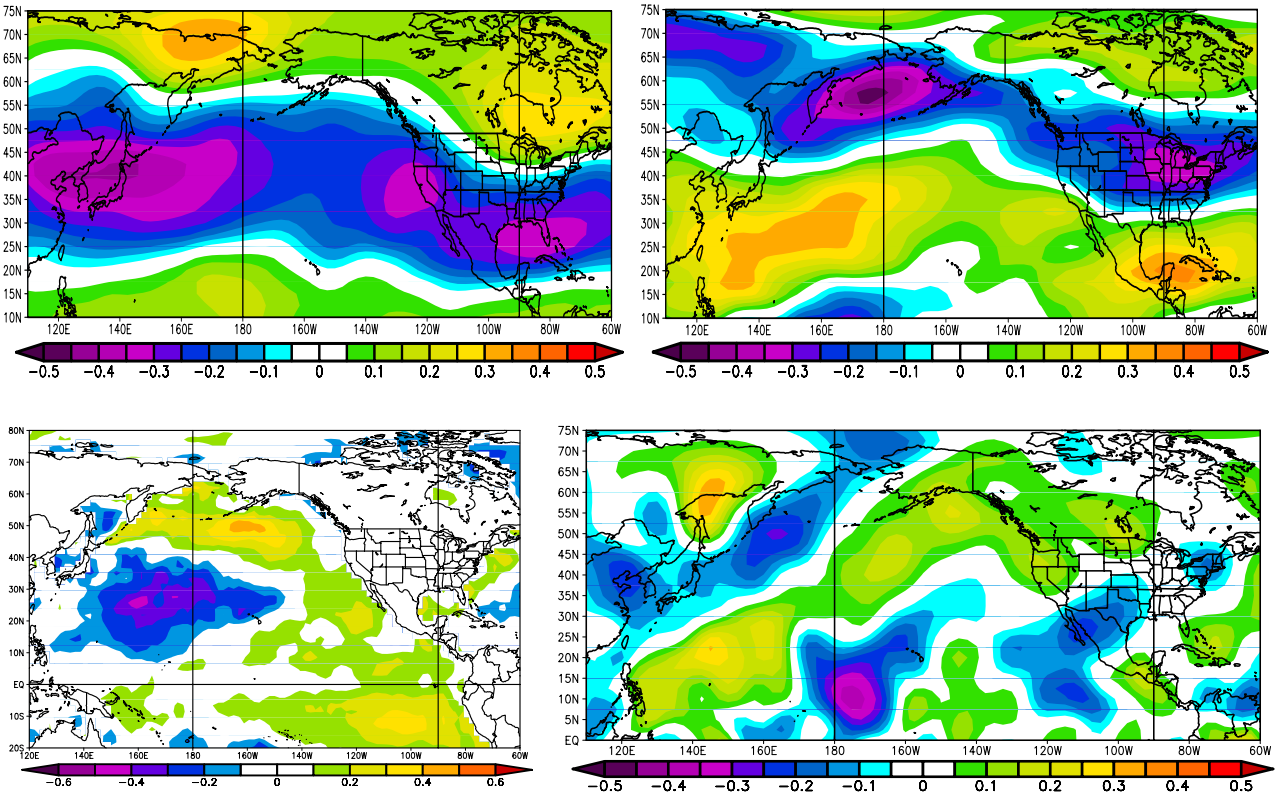


Figure 9: Correlation Maps of I gage streamflow with GPH (top left), ZW (top right), SST (bottom left) and MW (bottom right) for apr1 prediction.

Composite maps of wet and dry years give insight into the climate conditions which lead to high or low index gage flow (Figure 10). In the wet years winds approach the CRB from the ocean bringing moisture. Antecedent soil conditions tend to be more moist in wet years. In dry years winds approach primarily from the north bringing dry air. Antecedent soil conditions tend to be less moist in dry years.

6.2 Multi-Model Selection

The best models (subsets of predictors) were selected using the multi-model selection procedure. Table 2 gives the multi models which were selected. For the apr1 lead time PCA was used on the February-April SWE data so that the predictor used was actually the first PC of the SWE data. It is reasonable that the apr1 prediction only uses one model because the snowpack information at that date is complete and the PDSI information helps in years

Variable	Lead Time	Month/Season	Neg Region	Pos Region	Pos-Neg Cor.
GPH	nov1	oct	40,42N:137,140E	17,19N:178,182E	0.39
ZNW	nov1	oct	40,42N:-108,-106E	30,32N:132,135E	0.43
MDW	nov1	oct	40,42N:124,126E	40,42N:124,126E	0.31
SST	nov1	oct	-17,-15N:-96,-92E	38,40N:170,173E	-0.26
GPH	jan1	oct-dec	42,45N:-115,-110E	52,58N:-160,-155E	-0.56
ZNW	jan1	oct-dec	54,55.5N:-119,-114E	32,34N:-110,-106E	-0.58
MDW	jan1	oct-dec	45,50N:-135,-130E	42,45N:-91,-88E	-0.57
SST	jan1	oct-dec	28,31N:161,164E	46,49N:-166,-163E	-0.63
GPH	feb1	oct-jan	60,62N:128,132E	61,64N:-153,-150E	0.24
ZNW	feb1	oct-jan	72,74N:126,128E	28,31N:-119,-116E	0.42
MDW	feb1	oct-jan	66,68N:-108,-105E	38,42N:-90,-85E	0.30
SST	feb1	oct-jan	25,28N:160,163E	46,48N:-167,-164E	-0.64
GPH	apr1	feb-mar	40,42N:137 ,140E	65,70N:160,165E	-0.47
ZNW	apr1	mar	56,58N:172,176E	25,30N:155,160E	0.46
MDW	apr1	mar	10,12N:-177,-174E	56,60N:142,144E	-0.42
SST	apr1	jan-mar	24.5,26N:158,162E	48,51N:-165,-160E	-0.54

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

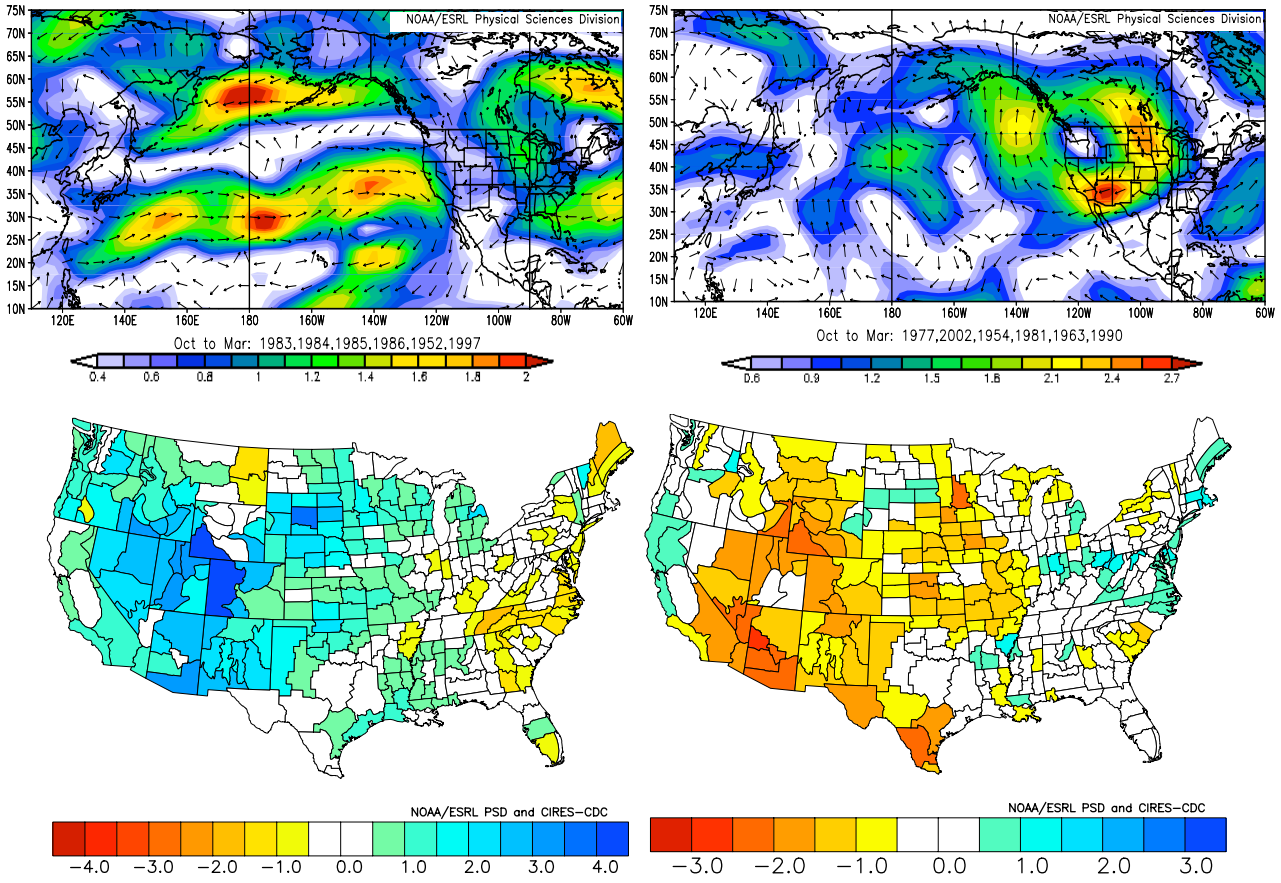


Figure 10: Composite maps of 500 mb vector winds for wet (top left) and dry (top right) years and PDSI for wet (bottom left) and dry (bottom right) years

which antecedent soil conditions are drier than normal and the following winter is wetter than normal. The other lead time use more models because the flow at spring flows at those times are less deterministic. One of the feb1 models is identical to the apr1 model because. SWE data at that point is incomplete so other models must be used to compensate. The jan1 and nov1 predictions must rely completely on climate variables.

Lead Time	Predictors	PDSI	GPH	ZNW	MDW	SST	SWE
apr1	2	1	0	0	0	0	1
feb1	2	0	0	0	0	1	1
feb1	2	1	0	0	0	0	1
feb1	3	0	0	0	1	1	1
jan1*	3	0	1	1	0	0	0
jan1	2	0	1	1	0	0	0
nov1	2	1	1	0	0	0	0
nov1	1	1	0	0	0	0	0

Table 2: Multimodels for each lead time. "1" indicates the presence of a predictor and "0" indicates the absence of a predictor. *This jan1 prediction used the jan1 and nov1 GPH as separate predictors.

6.3 MME forecasts

The RPSS for each year was computed and the median value is reported. Using leave-one CV the apr1 I gage prediction shows very high skills (Figure 11). Each boxplot represents an ensemble forecast with whiskers extending from 5th-95th percentiles and outliers lying beyond. Large connected points represent historical flow and the three dashed lines represent the 33rd, 50th and 66th percentiles of the historical data. All other lead times showed positive skills generally decreasing as lead time increased (Table 3). The nov1 predictions showed a showed a 28% improvement over climatology indicating somewhat skilful predictions can be made as early as November 1st. As mentioned before leave-one CV is generally not a very good indication of overall forecast skill. Retroactive forecasting showed very high skill though slightly reduced compared to the leave-one CV (Figure 12). This is to be expected seeing as more of the information is removed. In Figure 10 the dashed lines are the 33rd, 50th and 66th percentiles of the historical data before 1990.

Validation mode	apr1	feb1	jan1	nov1
Leave-one	0.95	0.85	0.65	0.28
Retroactive	0.80	0.85	0.29	0.10

Table 3: I gage predictions, median RPSS.

In some low and high flow years, the model failed to precisely capture the observed flow within the ensemble. The key is to recognize in which terciles the predictions fall and how the ensemble reacts as lead time decreases. The knowledge that a flow will be in the say, upper tercile is very valuable to water managers even if the exact value of flow is not predicted. The year 2002 was one of the lowest flow years on record. The model noticeably

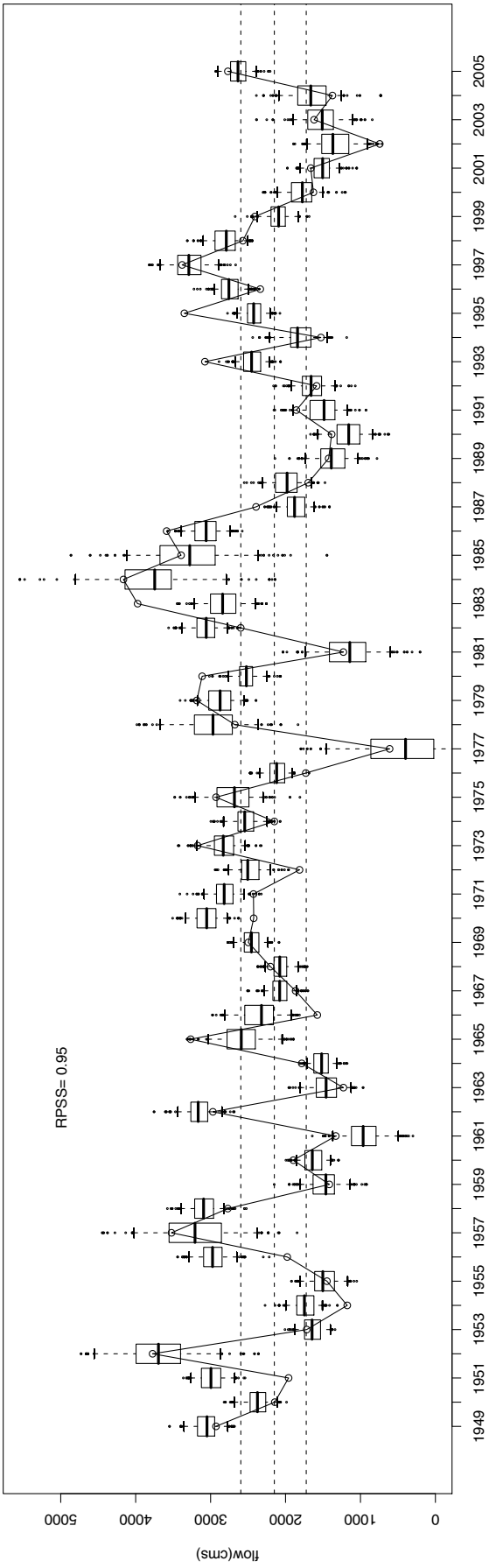


Figure 11: Leave-one CV predictions for the I gage apr1 prediction. Whiskers extend to 5th and 95th percentile of each ensemble.

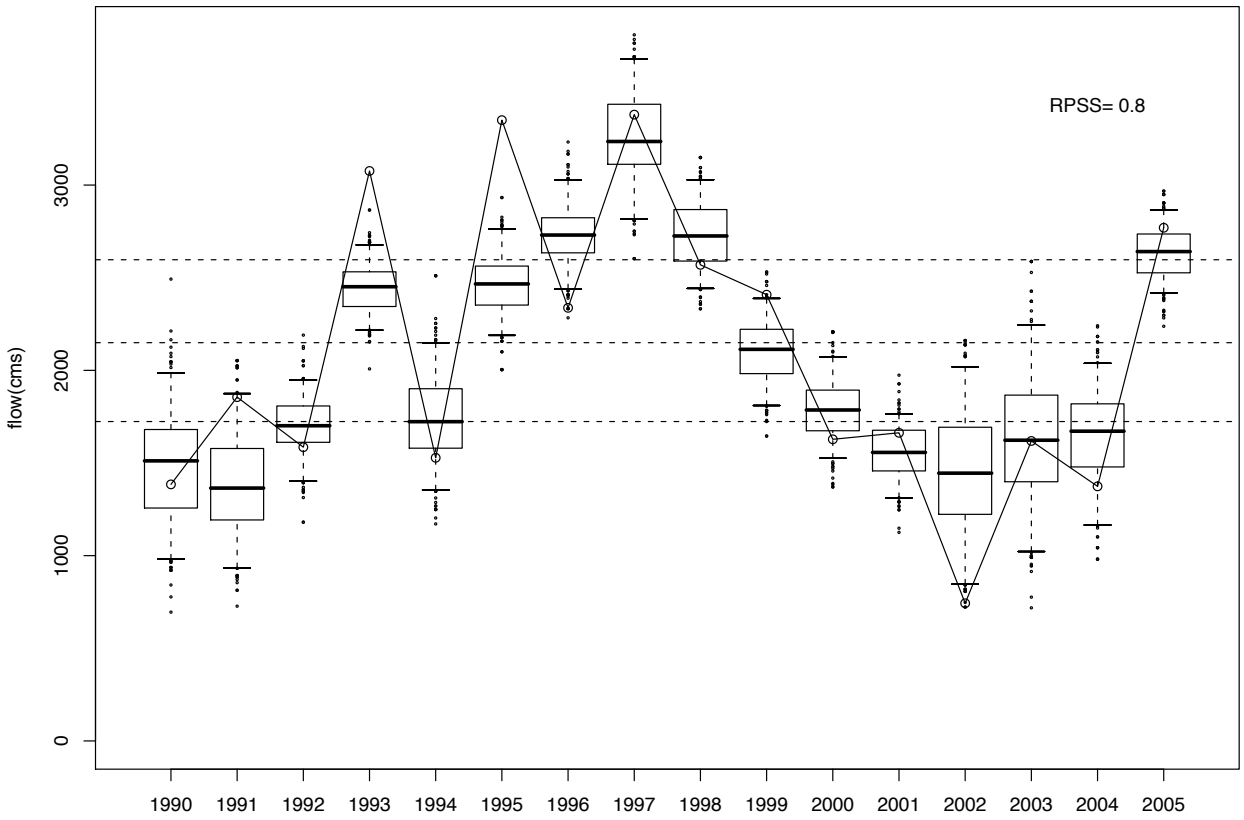
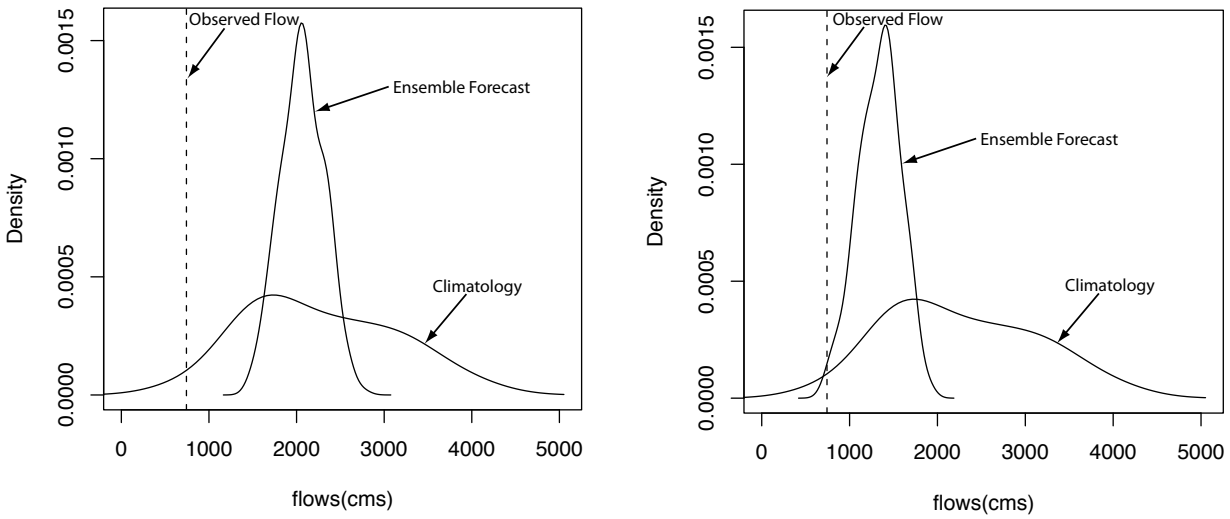


Figure 12: Retroactive forecast for the I gage on apr1 prediction.

reacts as lead time decreases by shifting the PDF much closer to the observed 2002 flow from November-April (Figure 13).



(a) nov1 prediction

(b) apr1 prediction

Figure 13: 2002 I gage predictions. The PDF produced in the ensemble model noteicably shifts toward the observed flow from November-April

6.4 Disagregation Verification

To verify that the spatial and temporal disaggregation reproduced the statistics of the historical distribution, the historical data was iteratively disaggregated as in Prairie et al. [2007]. The disaggregation successfully captured the mean, standard deviation, minimum, maximum, skew and cross correlation between the 4 sites. Figure 14The disaggregation also visually reproduced the shape of the historical distibution, including the famous bimodality in the may PDF at Fees Ferry.

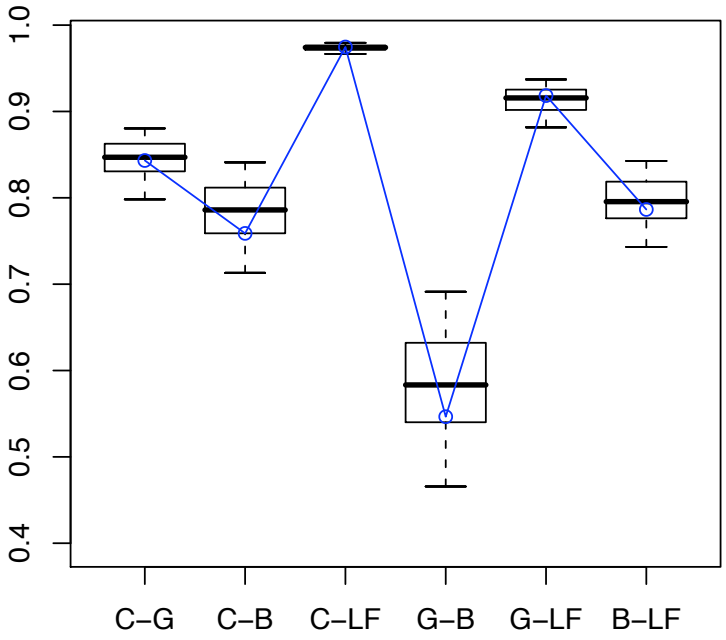


Figure 14: Cross correlation between sites: C=Cisco, B=Bluff, G=GRUT, LF=Lees Ferry. The solid line represents the observed statistics and the boxplots represent the statistics reproduced with the disaggregation procedure.

6.5 Disaggregation of the I Gage Forecasts

The disaggregation of the I gage streamflow produced skilful predictions at each of the four sites (Table 4). Figure 15 shows the ensemble disaggregation predictions at Lees Ferry compared to those of the CBRFC. Additionally, the disaggregation procedure was verified using a modified drop-k CV described in the methodology section. The boxplot of median RPSS values is an indication of the overall skill of the model (Figure 16). At all lead times similar trends were seen with the median RPSS value decreasing as lead time increased.

The difference between leave-one CV becomes apparent here. The reported RPSS values tended to vary wildly between the two modes of validation. With leave-one CV, Bluff showed the highest RPSS value on the apr1 prediction but with retroactive forecasting showed the lowest skill on apr1. The author would tend to agree with the retroactive forecast skills due to the fact that they put more strain on the predictive ability of the model.

Bluff was also the only site which negative simulated values were encountered though they accounted for 0.6% of the disaggregated values. This is perhaps due to the lower flows at Bluff in relation to the other sites. Extremely low RPSS values are due to the times when say,

Validation mode	Site	apr1 RPSS	feb1 RPSS	jan1 RPSS	nov1 RPSS	apr1 MC	feb1 MC	jan1 MC	nov1 MC
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	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	Lees Ferry	0.87	0.73	0.62	0.60	0.88	0.77	0.33	0.57

Table 4: Disaggregated I gage forecast skill.

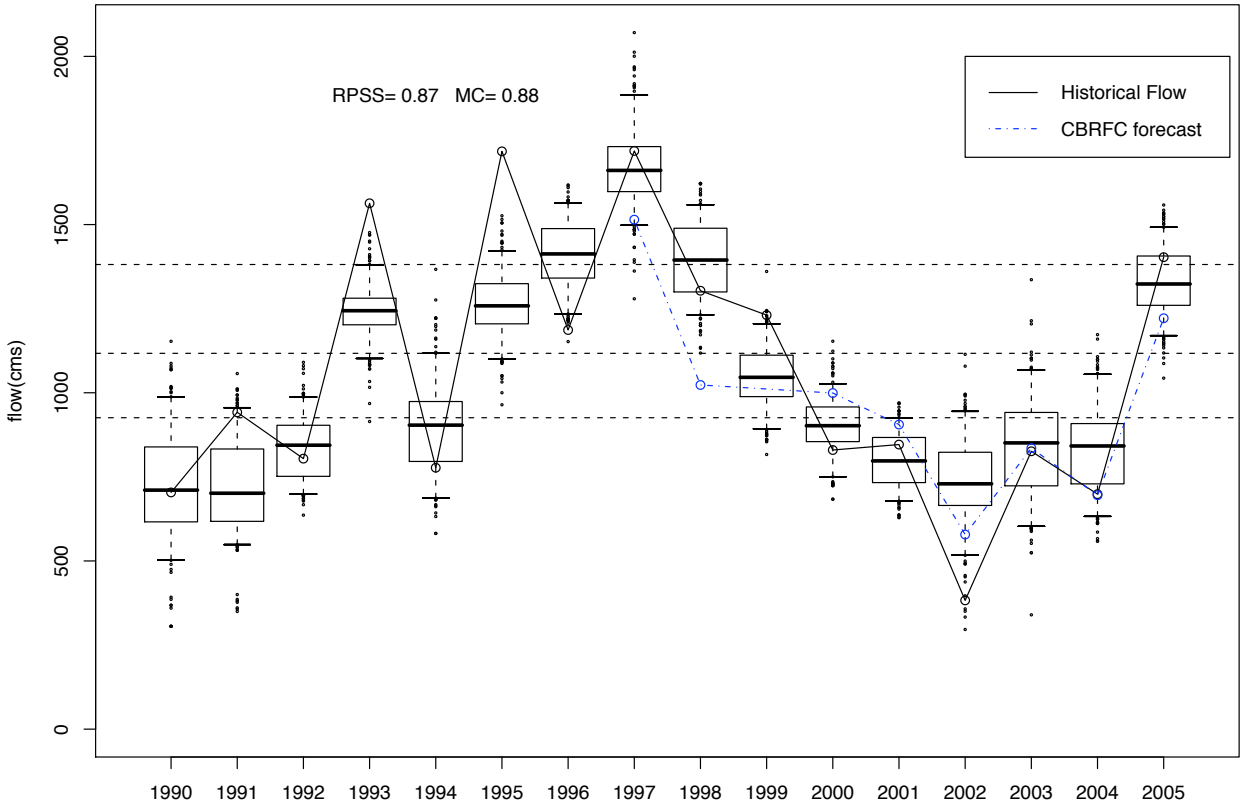


Figure 15: Retroactive forecast for Lees Ferry on apr1 prediction. forecasts are compared to the CBRFC predictions.

all the wettest years happened to be dropped and the model did not do well at predicting wet years.

6.6 Temporal Disaggregation

As previously mentioned, Lees Ferry is a very important gage for many reasons. A temporal disaggregation was carried out at Lees Ferry to predict monthly (April-July) flows. The temporal disaggregation at Lees Ferry was only verified using retroactive forecasting but

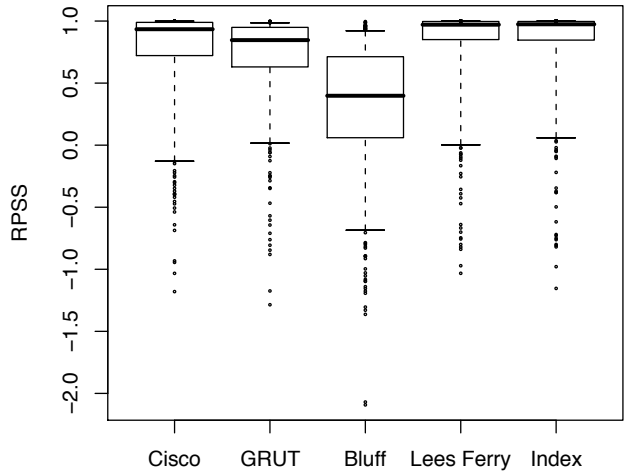


Figure 16: Modified drop-k CV

positive skills were observed (Table 5). Figure 17 shows the apr1 prediction of the May flows at Lees Ferry.

Less consistent trends in the RPSS values were seen than in the spatial disaggregation as lead time increased. For example, the nov1 july prediction showed more than twice the skill than the apr1 prediction. The MC values showed a more consistent decreasing trend as lead time increased. In the Lees Ferry temporal disaggregation, all except three monthly predictions showed skill over climatology.

Validation mode	Month	apr1 RPSS	feb1 RPSS	jan1 RPSS	nov1 RPSS	apr1 MC	feb1 MC	jan1 MC	nov1 MC
Retroactive	April	0.28	0.28	0.32	-0.05	0.68	0.86	0.47	0.31
Retroactive	May	0.88	0.74	0.17	0.26	0.85	0.86	0.60	0.53
Retroactive	June	0.24	-0.38	-0.16	0.41	0.79	0.66	0.25	0.56
Retroactive	July	0.31	0.14	0.63	0.64	0.58	0.40	0.0	0.53

Table 5: Temporal disaggregation of Lees Ferry

7 Summary and Discussion

The multi-site framework is a simple and parsimonious method for incorporating large-scale climate information into basin scale streamflow forecasts. The method is parsimonious because predictors need only be developed at one index gage, which is the sum of the seasonal flows at many spatial locations. Predictors were identified as in Grantz et al. [2005]. In an application to the UCRB, MME forecasts, in the fashion of Regonda et al. [2006], are made of the seasonal flows at the index gage. A KNN nonparametric disaggregation technique is implemented, as in Prairie et al. [2007], which provides seasonal forecasts at the four spatial locations and in turn monthly forecasts for the peak flow season (April-July).

Forecast skill were generally positive and high at that. RPSS values generally decrease as lead time increases because the atmosphere is less ordered. Very few negative RPSS values were encountered and all occurred during the second (temporal) disaggregation.

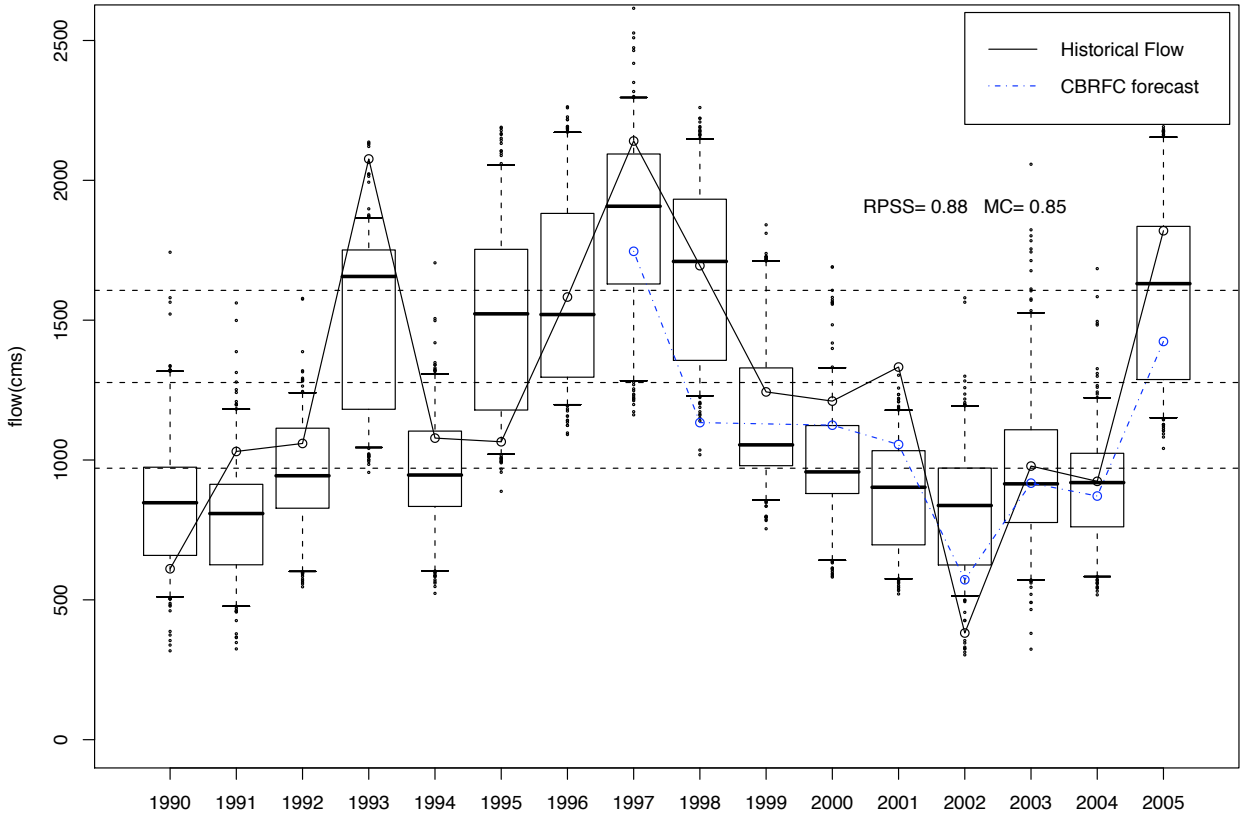


Figure 17: Retroactive forecast May monthly flow at Lees Ferry from apr1 temporal disaggregation. Predictions are compared to the CBRFC predictions.

The predictions made in retroactive forecast mode were comparable to the CBRFC predictions which are made using the ESP model. The earliest forecast of the ESP model is January 1 because of its heavy reliance on snowpack information. The multi-site framework has the ability to make skilful predictions as early as November 1. It is possible that the ESP model and the multi-site framework could be combined in a Bayesian context that could incorporate professional judgement.

Acknowledgements

References

David G. Brandon. Using nwsrfs esp for making early outlooks of seasonal runoff volumes into lake powell. In *Hydrology of Arid and Semi-Arid Regions*, 2005.

CBRFC. National weather service-colorado basin river forecast center. Technical report, US Department of Comerce, 2006.

Katrina Grantz, Balaji Rajagopalan, Martyn Clark, and Edith Zagona. A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts. *Water Resources Research*, 41, 2005.

Clive Loader. *Local Regression and Liklihood*. Springer, 1999.

Thomas Lumley. leaps: regression subset selection. Technical report, 2007. R package version 2.7, Fortran code by Alan Miller.

Simon J. Mason. Cross-validation and other Òout-of-sampleÓ testing strategies. In *AMS Short Course on Significance Testing, Model Evaluation and Alternatives*, 2004.

James Prairie. Long-term salinity prediction with uncertainty analysis: Application for colorado river above glenwood springs, colorado. Master's thesis, University of Colorado, Boulder, 2002.

James Prairie. *Stochastic Nonparametric Framework for Basin Wide Streamflow and Salinity Modeling: Application for the Colorado River Basin*. PhD thesis, University of Colorado, 2006.

James Prairie, Balaji Rajagopalan, Upmanu Lall, and Terrance Fulp. A stochastic nonparametric technique for space-time disaggregation of streamflows. *Water Resources Research*, 2007.

Satish Kumar Regonda, Balaji Rajagopalan, Martyn Clark, and Edith Zagona. A multimodel ensemble forecast framework: Application to spring seasonal flows in the gunnison river basin. *Water Resources Research*, 42, 2006.

David G. Tarboton, Ashish Sharma, and Upmanu Lall. Disaggregation procedures for stochastic hydrology based on nonparametric density estimation. *Water Resources Research*, 34:107–119, 1998.

Harrison M. Wadsworth, editor. *Handbook of Statistical Methods for Engineers and Scientists*. McGraw-Hill, 1990.