Incorporating Climate Signals for Ensemble Streamflow Prediction Post Processing

Portland State UNIVERSITY

Mohammad Reza Najafi (najafim@cecs.pdx.edu), Hamid Moradkhani (hamidm@cecs.pdx.edu)
Department of Civil and Environmental Engineering, Portland State University

INTRODUCTION

In this study we propose a post-processing procedure that assigns weights to streamflow ensemble members using the large scale climate signals. Analysis is performed over the rainfall dominated Leaf River basin and snow dominated East River basin to improve the spring ensemble streamflow volume forecast. We employ Fuzzy C-Means clustering, Formal Likelihood, Informal Likelihood and K-Nearest Neighbors methods for the weighting and it is found that Principle Component Analysis (PCA) improve the accuracy of the weighting scheme considerably. The presented objective method can be applied to enhance the final ESPs; nevertheless the user expertise may change any of the process steps. The current predictions based on simple average or the median of the ensemble members may come with the weighted ensemble forecasts to better provide possible ranges and uncertainty bounds.

STUDY AREAS

The snow dominated East River basin is a tributary of the Gunnison River in the Colorado River basin. it is 754 km^2 with the elevation ranges from 2445 to 3898m. The data are obtained from the NWS CBRFC and consists of the six hourly precipitation and temperature data for each of the three elevation bands along with the daily streamflow in the outlet. The duration of the streamflow data is 1975-2005 and the forcing data 1960-2005.
The Leaf river basin is a rainfall dominated basin with an area of 1944km^2, which is located in the north of Collins, Mississippi, United States. The Precipitation, Temperature and Streamflow data were available from 1948-1988.

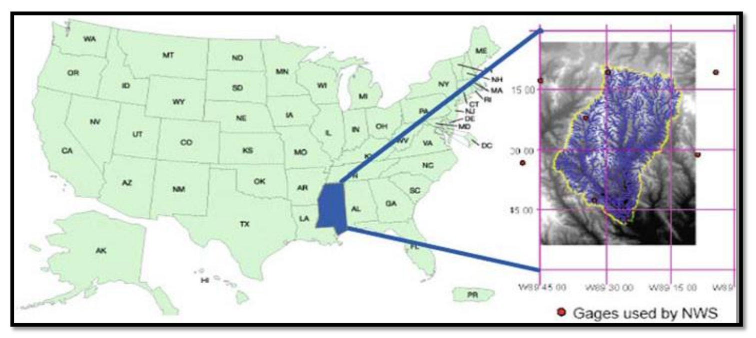


Figure 1. Leaf River basin

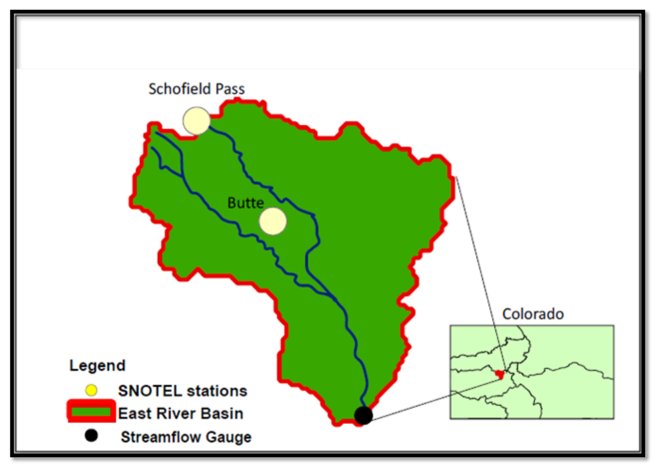


Figure 2. East River basin

Models

We combine SNOW-17 model with the Sacramento Soil Moisture Accounting Model (SAC-SMA) to predict the streamflow in snow dominated basin. SNOW-17 model is driven by temperature and precipitation as the forcing data and it simulates the Snow Water Equivalent (SWE), Snow Cover Area (SCA), and Rain plus Melt (RM) at each time step. RM then becomes the input to the SAC-SMA along with the Potential EvapoTranspiration (PET) to estimate the runoff. For Leaf rainfall dominated basin it solely runs the SAC-SMA model.

Ensemble Streamflow Prediction

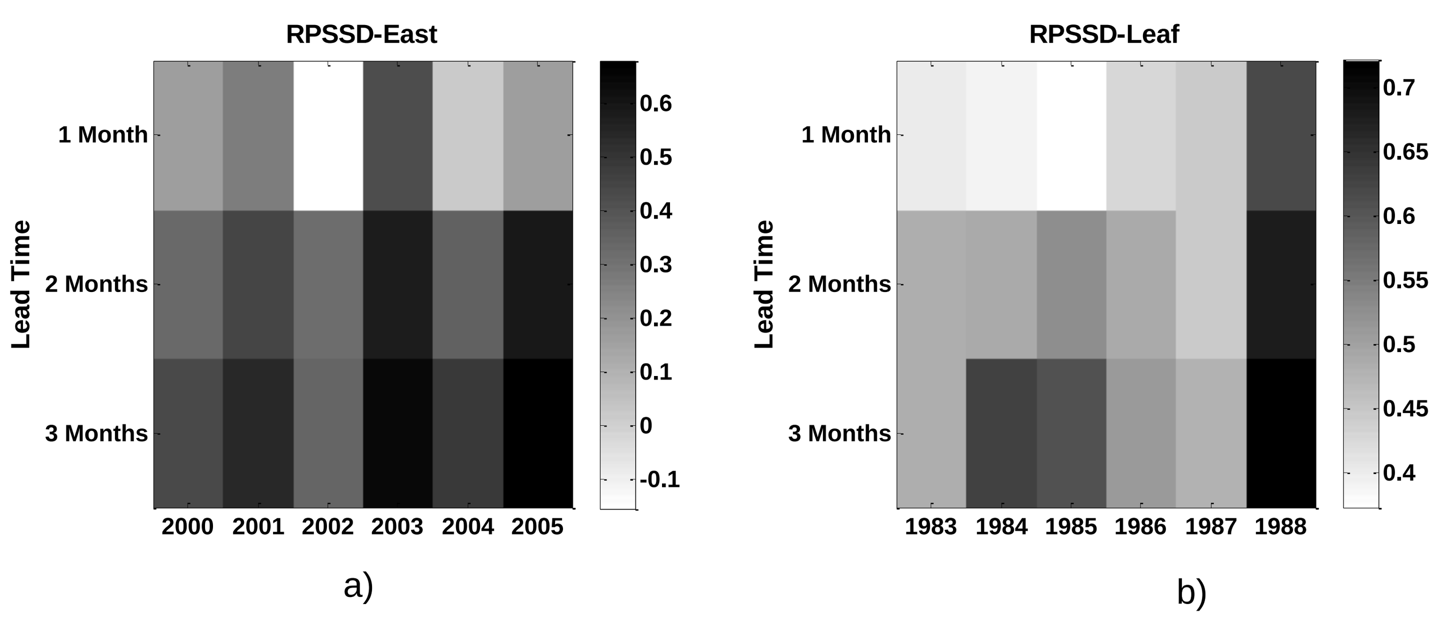


Figure 3. RPSSD for a) East river basin and b) Leaf river basin.

Selecting the Climate Signals

We use the information content in the climate signals in weighting the ensemble trajectories for the spring (April, May, June) streamflow volume forecasts. For this purpose we select the climate signals using the spearman correlation between each monthly signal (starting from the January in the previous year until March in the forecast year) and the spring runoff volume. For the hypothesis test, a correlation with a P-value of less than 0.05 is considered significant. In order to perform the ESP weighting, this process is done for each forecast year separately.

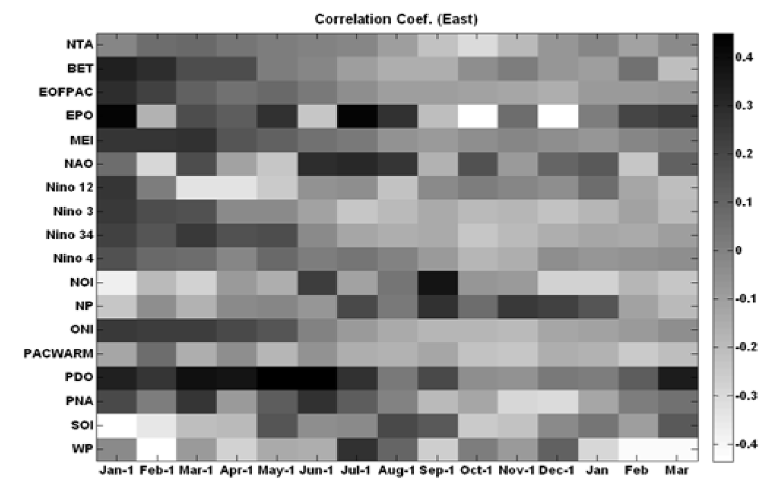


Figure 4: Spearman correlation coefficient between various climate signals and the spring runoff volume for East river basin in 2005.

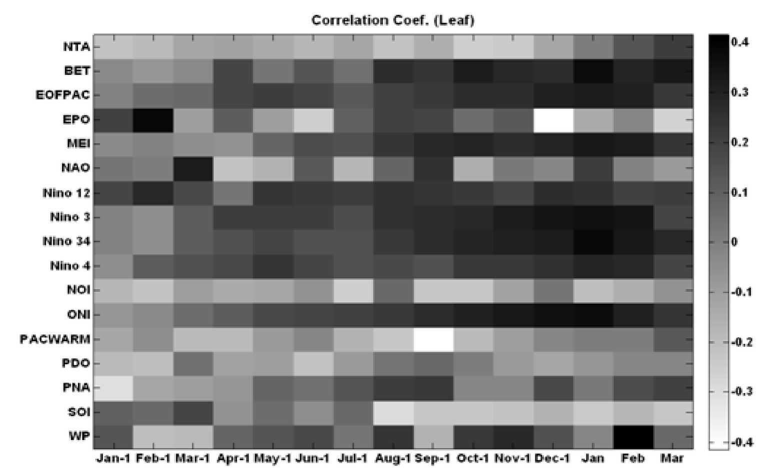


Figure 5: Spearman correlation coefficient between various climate signals and the spring runoff volume for Leaf river basin in 1983.

Some climate signals show significant negative correlation with the spring runoff volume such as West Pacific patter (WP) in February and March.

Provision of Uncorrelated Climate Signals

PCA is a procedure that removes the co-linearity between the variables and minimizes the redundancy in the information by obtaining principal components (PCs) that are linear combinations of the original variables.

The first PC accounts for the highest possible variance in the data set. The second PC which is uncorrelated with the first PC contributes maximum to the residual variance. This is continued until the total variance is accounted for.

We make the signals uncorrelated using PCA to reduce this redundancy and to increase the efficiency of the weighting methods. In order to determine the number of PCs to retain, we consider two approaches of the so called "NRCS" and "PRESS" methods:

1. In the NRCS method, PCs are added to the regression model one-by-one. A statistical t-test is performed each time a PC is added. If the addition of the PC is significant compared with a critical t-value of 1.6, it would be retained, else the current number of PCs are chosen.
2. In the Prediction Residual Sum of Squares (PRESS) method, a leave-one-out cross validation is performed and the value of PRESS is calculated each time a value is extracted. The optimum number of PCs is obtained based on the minimum value of PRESS.

Ensemble Post-Processing

Weights are assigned to each streamflow ensemble member based on the similarity of the climate signal PCs in the forecast year to the ones in the previous years using the following methods:

- Fuzzy C-Means clustering algorithm: (Bezdek and Ehrlich, 1984)
- Formal Likelihood: Involves a Multivariate Normal distribution with the selected climate signals in the forecast year as the mean and the covariance of the climate signals in the historical years as the covariance of the distribution.
- Informal Likelihood: Using the function presented by (Beven and Binley, 1992):

$$l_j = (\sigma_e{}^2)^{-N}$$

where σ_e is the variance of the residuals between the climate signals in the fore cast year and the ones in the historical year.

- K-Nearest Neighbor (KNN): The k number of nearest climate signal PCs in the previous years to the ones in the forecast year is chosen and weighted based on their Euclidean distance.
- 1D-KNN: Similar approach to KNN is adopted by (Werner et al., 2004).

Results

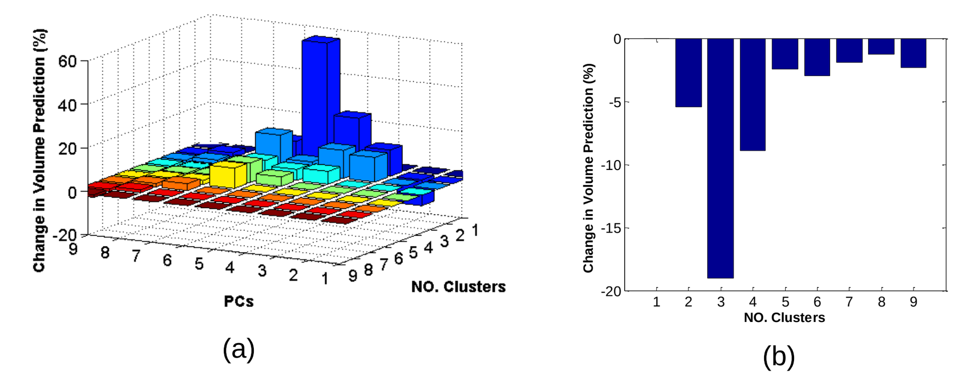
$$\%change = \frac{\Delta_s - \Delta_w}{\Delta_s} \times 100$$

$$\Delta_s = |\bar{V} - V_{obs}|$$

$$\Delta_w = |\bar{V_w} - V_{obs}|$$

In which $V_{obs}, \bar{V}, \bar{V_w}$ represent the observed spring streamflow, simple mean and weighted mean of the ensemble streamflow predictions respectively.

Figure 6. Weighting scheme based on Fuzzy C-Means Clustering using a) the principle components of the selected climate indices and b) all the selected climate indices for East River basin in 2005.



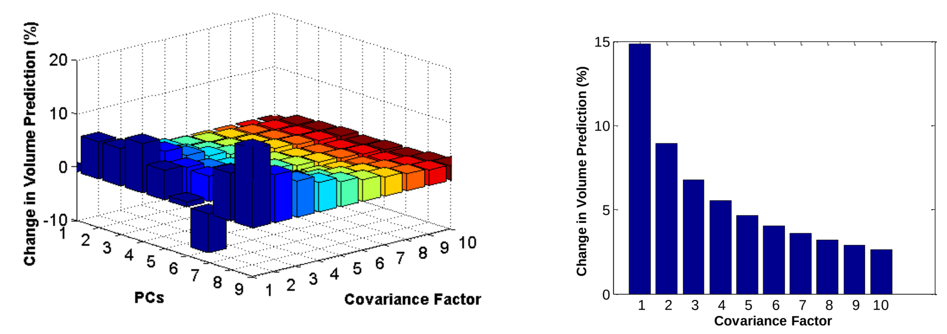


Figure 7. Weighting scheme based on Formal Likelihood approach using a) the principle components and b) all the selected climate indices for East River basin in 2005.

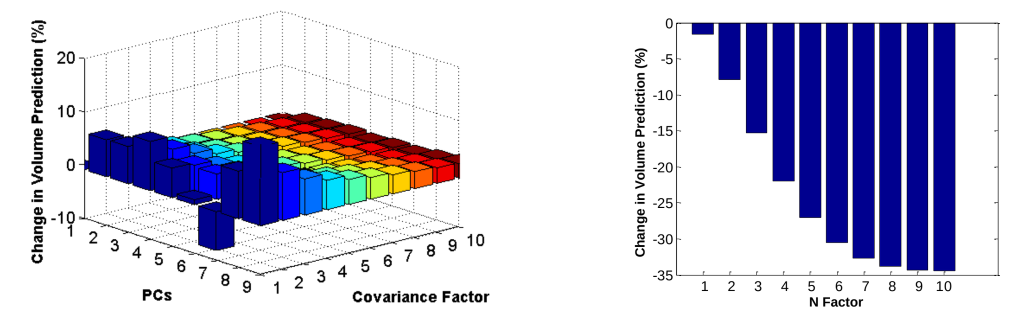
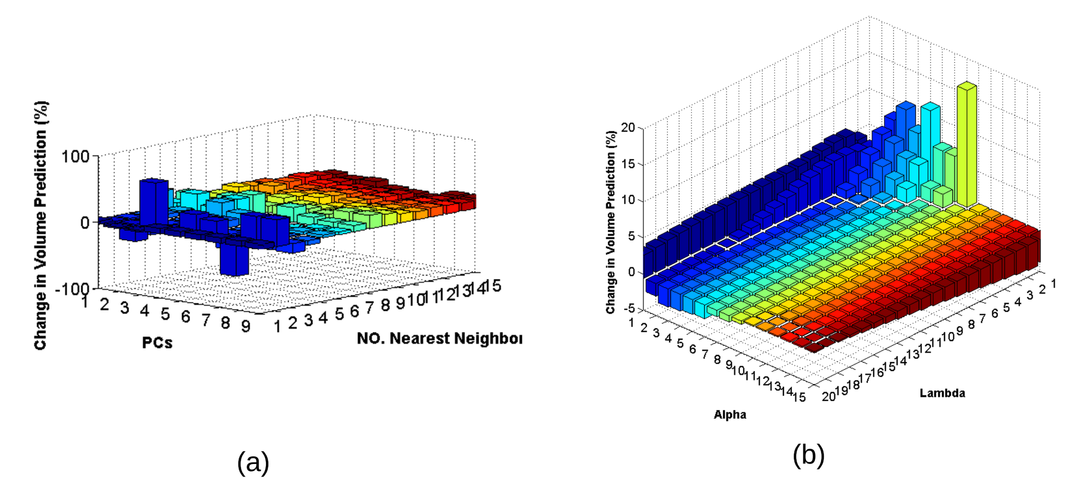


Figure 8. Weighting scheme based on Informal Likelihood approach using a) the principle components and b) all the selected climate indices for East River basin in 2005.

Figure 9. Weighting scheme based on a) KNN and b) 1D KNN for East River basin in 2005.



CONCLUSIONS

- Climate Signals show significant Correlation with the spring streamflow in both basins.
- The Principle Component Analysis effectively reduces the redundancy in the selected signals and enhances the overall weighting scheme considerably.
- The post processing weighting schemes improves the accuracy of the spring runoff volume prediction.
- Care should be given to the choice of parameters of each method and number of PCs. An objective method to select the number of PCs is presented in this study.
- The weighting methods are simple and efficient and the results are promising. The whole process can be automatically performed in an objective way.

REFERENCES

Beven K, Binley A. 1992. The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes* **6**: 279-298.
Bezdek JC, Ehrlich R. 1984. FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences* **10**: 191-203.
Werner K, Brandon D, Clark M, Gangopadhyay S. 2004. Climate index weighting schemes for NWS ESP-based seasonal volume forecasts. *Journal of Hydrometeorology* **5**: 1076-1090.