Conditionally Resampling from Ensemble Streamflow Forecasts Using a *k*-Nearest Neighbor Bootstrapping Approach

Brennan Middleton and Balaji Rajagopalan
Department of Civil, Environmental and Architectural Engineering, University of Colorado, Boulder, Colorado, USA

Shane Coors
Precision Water Resources Engineering, Loveland, Colorado, USA

Abstract: Ensemble streamflow forecasts produced by the National Weather Service (NWS) are essential to helping decision makers formulate informed and effective reservoir operating strategies. These forecasts are used to characterize uncertainty in seasonal streamflow predictions based on the current conditions and past climatology of a basin. Typically, each trace in an ensemble forecast is weighted the same, which makes the implicit assumption that climatology from any year in a historical record is equally as likely to occur. This study presents a generalized framework for weighting traces using a *k*-nearest neighbor bootstrapping approach, and compares the skill of this approach to that of a set of ensemble reforecasts produced for spring streamflow in the Truckee River Basin. Results show that conditionally resampling from a set of traces based on large-scale climate information can greatly improve forecast skill, and has the potential to narrow the spread of ensembles produced at longer lead times.

Introduction

Predictions of seasonal streamflow are integral to decision support for mid-term reservoir operations and planning. Runoff volume forecasts for the upcoming spring or water year allow reservoir managers to plan for future conditions by allocating or storing water as efficiently as possible. Skillful forecasts at longer lead times have also made possible more flexible operating strategies that respond directly to current and forecasted conditions. To this end, operators rely heavily on short to long-term forecasts provided by NWS River Forecast Centers (RFCs). While each RFC typically offers hydrologic guidance for a variety of time scales, and employs a suite of hydrologic models and procedures in doing so, ensemble forecasts are an approach that are often used for long range, seasonal forecasts. Ensemble forecasts better characterize forecast uncertainty by supplying a *range* of possible values. This method of presenting forecasts probabilistically, rather than deterministically, has allowed reservoir managers to better manage the uncertainty and risk inherent in operations.

The Colorado Nevada River Forecast Center (CNRFC), located in Sacramento, California, uses the Hydrologic Ensemble Forecast Service (HEFS) to issue probabilistic forecasts out to one year. Each ensemble is made up of a number of "traces," or members, which are initialized using current conditions of the basin and near-term precipitation and temperature forecasts. Past climatology is then used to extend these forecasts, where each trace corresponds to observed conditions from one year in a historical record. For example, a historical record of 1950-2008 would result in 59 unique traces, or possible outcomes. In current practice, each trace is assumed to be equally as likely to occur, which assumes no a priori knowledge of the future state of the system. However, previous studies have found that certain climate variables serve as valuable predictors for spring streamflow.

In one particular study, *Grantz et al.* (2005) developed a regression-based seasonal forecasting model to predict spring streamflow in the Truckee-Carson River Basin at longer lead times. Large-scale climate information such as snow water equivalent (SWE), 500 mbar geopotential height and sea surface temperatures (SST) were found to be significantly correlated with spring streamflow in this basin and were used as predictors for the ensemble forecast model. Results from this study indicated that incorporating basin specific climate indices in spring streamflow forecasts greatly improved forecast skill, especially for fall and early winter months in which there have yet to be any observations of snowpack accumulation.

In addition to this, there exists a substantial body of literature in which climate indices have been applied to the construction or weighting of ensemble forecasts (*Hamlet and Lettenmaier,* 1999; *Werner et al.,* 2004; *Regonda et al.,* 2006; *Wood and Lettenmaier,* 2006; *Moradkhani and Meier,* 2010; *Najafi et al.,* 2012). *Bradley et al.* (2015) applied a Bayesian approach to the weighting of Blue Nile flood season flow forecasts, weighting ensembles based on the conditional distribution between ensemble members and a single ENSO index. This approach, which requires the estimation of a likelihood function, was found to perform similarly to both kernel and *k*-nearest neighbor (K-NN) methods.

K-NN bootstrapping simplifies the approach of simulating from the conditional distribution by avoiding any prior assumptions of a joint PDF. K-NN bootstrapping employs an algorithm to simulate future conditions by conditionally resampling from historic observations. The resampling is conditional in that observations are weighted based on how closely they resemble the conditions of the modeled timestep, as measured by a set of variables known as a "feature vector." K-NN bootstrapping has been applied to many timeseries analyses, especially in simulating future weather or streamflow patterns (*Lall and Sharma,* 1996; *Rajagopalan and Lall,* 1999; *Yates et al.,* 2003; *Gangopadhyay et al.,* 2005; *Prairie et al.,* 2005; *Apipattanavis et al.,* 2007, *Gangopadhyay et al.,* 2009). Past studies have also directly applied K-NN to the weighting of ensemble members. For instance, *Werner et al.* (2004) applied a "distance-sensitive nearest-neighbor weighting" scheme to weight ensembles of Colorado spring runoff based on the Niño-3.4 index. *Najafi et al.* (2012) underwent a similar analysis, although in this case a principal component analysis (PCA) was performed on various climate signals and the first PC used as an index. Both studies found that post-processing ensembles using K-NN improved forecast performance.

This research will further investigate the benefits of applying K-NN bootstrapping to the weighting of seasonal ensemble traces, specifically demonstrating whether this improves forecast skill at longer lead times, and identifying suitable climate indices for the Truckee River Basin. Weighting traces, as it is used in this context, implies constructing a modified ensemble by preferentially sampling from RFC forecasts.

Study Region: Truckee River Basin

A schematic of the study region is provided in **Figure 1**. The Truckee River Basin, which spans an area of approximately 3,060 square miles in the Sierra Nevada Mountains of both California and Nevada, is managed by the Bureau of Reclamation (Reclamation) Lahontan Basin area office with technical support provided by Precision Water Resources Engineering. The hydrology in this region is snowmelt-dominated, with the majority of precipitation arriving between November and April as snow, producing streamflow in the later spring months (April-July). The Truckee is a complex river basin in which the waters have been fully appropriated, further motivating the need for skillful forecasts of spring streamflow. The CNRFC currently produces ensemble forecasts for this basin for both April-July and water year (WY) streamflow volumes.

Data

Retrospective forecasts (reforecasts) produced using the HEFS were obtained for the Farad gauging station (FARC1), which is located on the Truckee River on the border of California and Nevada. These reforecasts were produced for each day between 1985-2010 for the purposes of computing forecast verification metrics. Inputs to the HEFS reforecasts included raw precipitation and temperature forecasts from the NCEP operational Global Ensemble Forecast System (GEFS) reforecasts for days 1 to 15 and from climatology for days 16 to 365 (*cnrfc.noaa.gov*,

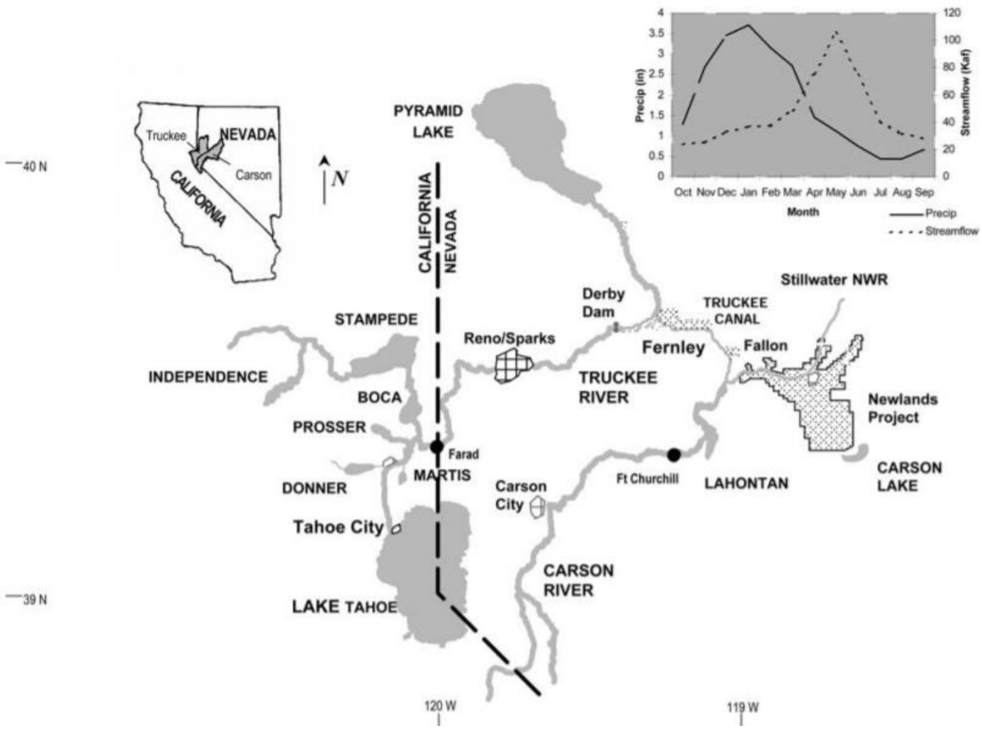


Figure 1. Truckee and Carson River Basins with Annual Average Hydrographs (*top right*) (*Grantz et al*, 2005)

n.d.). A period of record of 1950-2009 was used to construct ensembles of 60 traces. Each reforecast was aggregated into spring (April-July) accumulated volumes in units of thousand acre-feet (kaf). Reforecasts produced during the spring period, or after April 1st, were combined with observed totals to date to be consistent with how forecasts are produced in real-time. This dataset was used as a baseline for which to compare the performance of the methodology proposed in this paper.

In addition to the HEFS reforecasts, the following datasets were used in this analysis:

(1) Naturalized daily streamflow (cfs) for the Farad gauging station, spanning October 1st, 1900 to the current date (provided by Precision Water Resources Engineering)
(2) Monthly values of Kaplan SST (http://iridl.ldeo.columbia.edu/SOURCES/.KAPLAN/.EXTENDED/.v2/.ssta/)
(3) Monthly values of geopotential heights obtained from NCAR/NCEP Reanalysis on the Climate Diagnostics Center (CDC) webpage (http://www.cdc.noaa.gov)

Climate Diagnostics

A preliminary analysis was conducted to determine which large-scale climate variables would be appropriate to apply to the K-NN resampling. Findings from *Grantz et al.* (2005) suggested that the standard indices monitoring the ENSO and PDO phenomena did not show significant correlation with spring streamflow in the Truckee River. Rather, SSTs and 500 mbar geopotential heights were found to serve as suitable substitutes. **Figure 2** presents the correlations between spring streamflow and these variables. Values for these variables were averaged over the areas of highest correlation for each year over the period 1950-2009 (177.5°-182.5°E and 42.5°-47.5°N for SST, and 225°-235°E and 35°-45°N for geopotential height).

Winter SWE was not included as a climate variable in this analysis since this information is used to initialize model runs and thus is already "present"

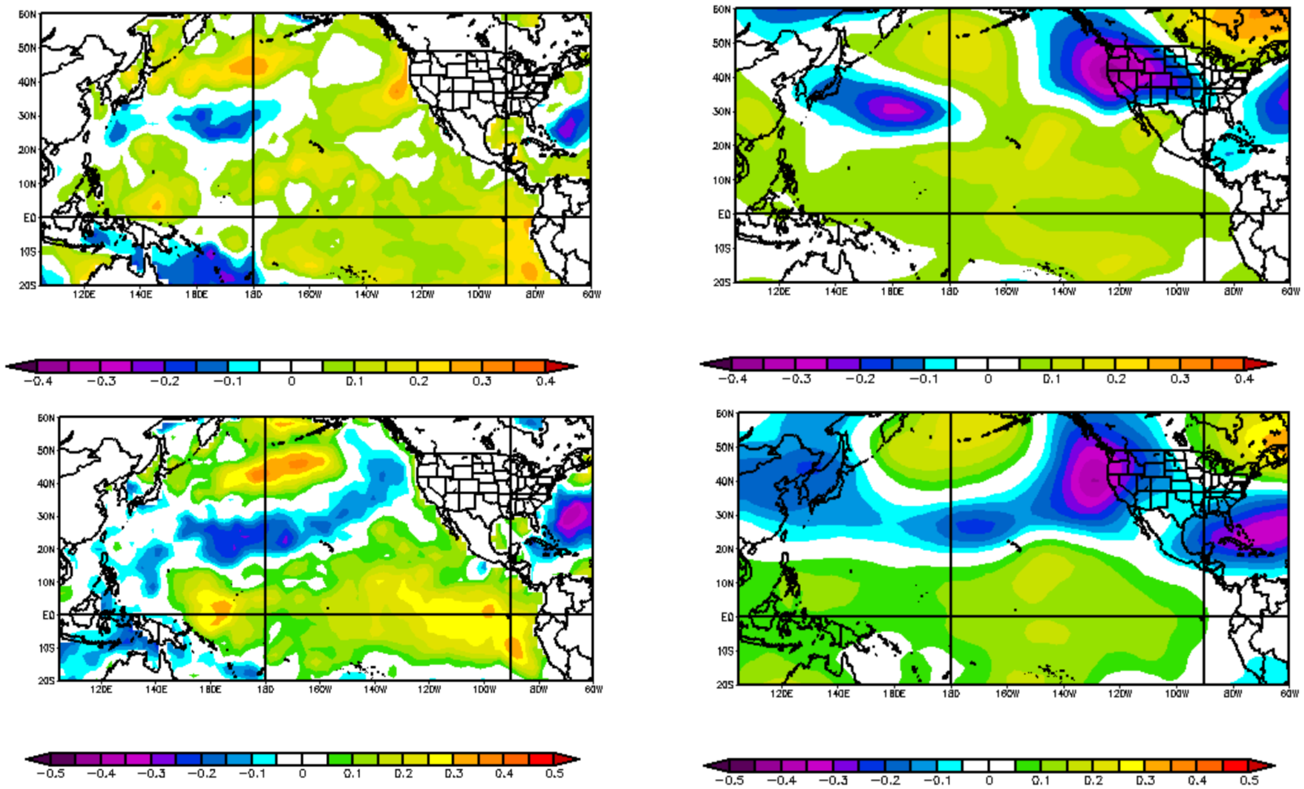


Figure 2. Correlations of Truckee River spring streamflow with (*top*) September-November and (*bottom*) December-February (*left*) SST and (*right*) 500 mbar geopotential heights

in the ensembles. Additionally, SWE observations for the Truckee Basin are limited to a period of record back to 1981, leaving no information for trace years between 1950 and 1980.

The K-NN Resampling Algorithm

A K-NN resampling algorithm was applied to preferentially sample traces from the HEFS reforecasts based on large-scale climate information. This is equivalent to simulating from the conditional PDF of the current forecast, given what we know about past trace years. In theory, this would lead to less variability or spread in the ensemble, and thus higher confidence in our prediction of seasonal streamflow.

For each year of the reforecast period (i.e. 1985-2010), the following steps were taken:

(1) Starting with the first forecast month of October, construct a "feature vector" which contains the spatially averaged values of SST and geopotential height for the preceding three months (i.e. September, August, and July).

(2) Construct a matrix containing the same information for each trace year, excluding the current forecast year (since there is overlap between the 1950-2009 period of record and the 1985-2010 reforecast period). These vectors are commonly referred to as the "successors" (*Lall and Sharma*, 1996).

(3) Compute the Euclidean distance between the feature vector and each successor using the following equation:

$$r_m = \sqrt{\left[\sum_{j=1}^{d} (x_{fj} - x_{sj})^2\right]}$$

Where x_{fj} represents the *j*th component of the feature vector and x_{sj} represents the *j*th component of the successor. This is repeated m times, where m is equal to the number of successors.

(4) Order the distances, and weight them using a decreasing kernel $K[j]$ as follows:

$$K[j] = \frac{1/j}{\sum_{j=1}^{k} 1/j}$$

Where j is the order number 1,2,...,k and $k = \sqrt{N}$ nearest neighbors (N being the sample size or total number of traces). This ensures that the successor, or trace, with climate variables most similar to the current year is weighted the heaviest, and the kth nearest neighbor is weighted the least.

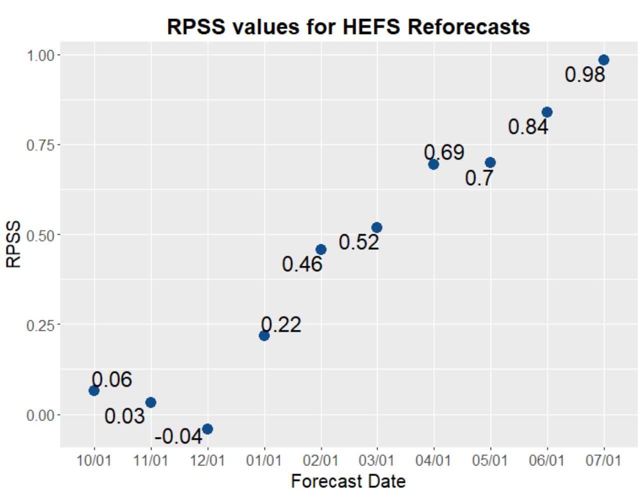
(5) Randomly select from the k nearest neighbors using the weights calculated from the previous step, and include the forecasted time series for this trace in the resampled ensemble. Repeat steps 1-5 until a new ensemble with 60 traces is generated.

(6) Repeat steps 1-6 for the remaining forecast months (November-July).

This K-NN resampling algorithm allows any one trace to be resampled multiple times, simulating the effect of weighting this trace more heavily in the modified ensemble. Adjusting the number of nearest neighbors to choose from, k, effects how many traces can be resampled, and consequently the magnitude of weights (i.e., selecting from fewer years results in overall higher weights). Although $k = \sqrt{N}$ was used initially, as proposed by *Lall and Sharma* (1996), various values of k were tested as can be seen in the results.

Results

The first step in assessing the performance of this proposed approach was to quantify the current skill in the HEFS reforecasts as a baseline. **Figure 3** presents the median forecast correlation with observations for each month (as produced by the CNRFC for verification purposes), as well as RPSS values for each month. RPSS, or ranked probability skill score, is a metric often used to quantify the skill of an ensemble by measuring the forecast's ability to capture the PDF relative to climatology. This is accomplished by measuring how ensemble traces are distributed amongst a set of categories, such as the 33rd and 66th observed streamflow percentiles, as compared to observations (which is 33% for each category in this case). An RPSS of one represents a perfect forecast, while a negative RPSS indicates the forecast is less accurate than climatology, or an average of observations.



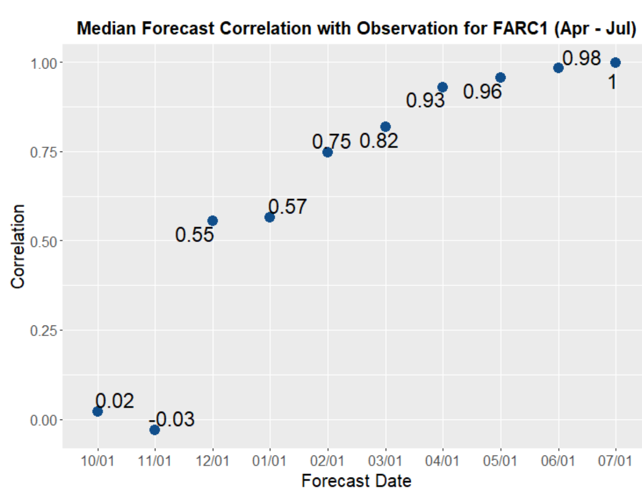


Figure 3. Baseline Skill Statistics for HEFS Reforecasts Produced for the Farad Gauging Station

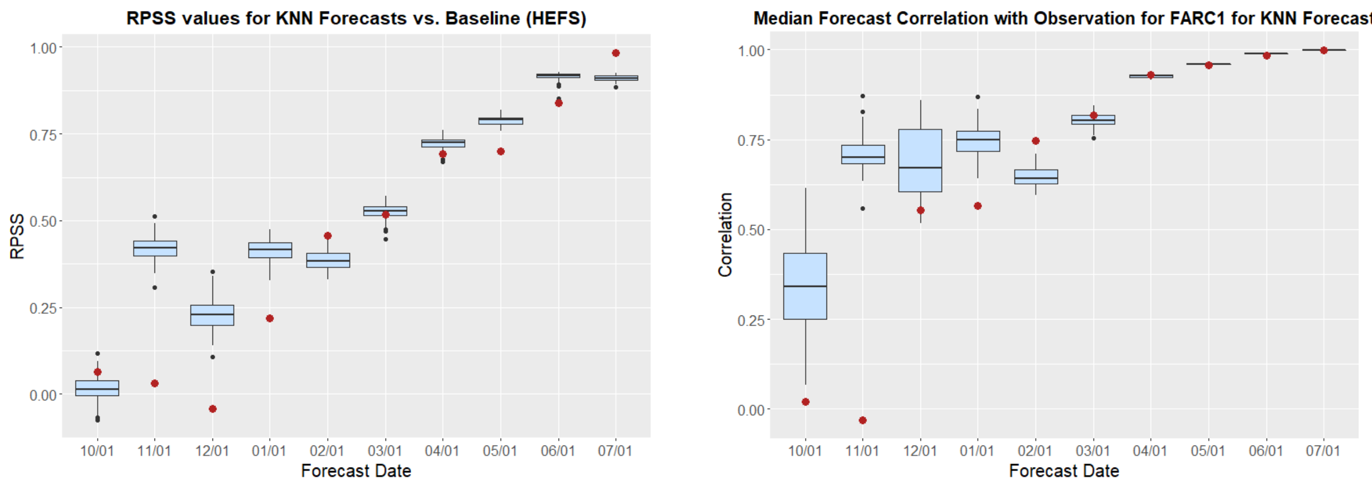


Figure 4. Skill Statistics for K-NN Resampled Forecasts vs. the Baseline (*shown as red points*) for $\boldsymbol{k} = \sqrt{\boldsymbol{N}}$

As can be seen from both plots, forecast skill increases over the course of the year. This is to be expected since more information is known with time, especially for months May to July in which some period of spring streamflow has already been observed. One observation is that in this case, a negative RPSS value for the December 1st forecasts does not correspond to a low correlation. This indicates that although the median forecasts for this month are relatively correlated with observations, the PDF, or distribution of values, is still not well represented. One of the main utilities of ensemble forecasts is their ability to produce PDFs, or exceedance probabilities of runoff thresholds (i.e. the probability that spring streamflow will exceed 200 kaf). Thus, forecasts with a low RPSS are not nearly as useful in providing decision support for reservoir operations.

Figure 4 (*above*) illustrates the same skill statistics for the K-NN produced ensembles against the baseline. Here, values are represented using boxplots since skill statistics were calculated for modified ensembles produced over 60 simulations. On average, weighting traces based on large-scale climate information improves forecast skill for the earlier, winter months. This is especially evident in

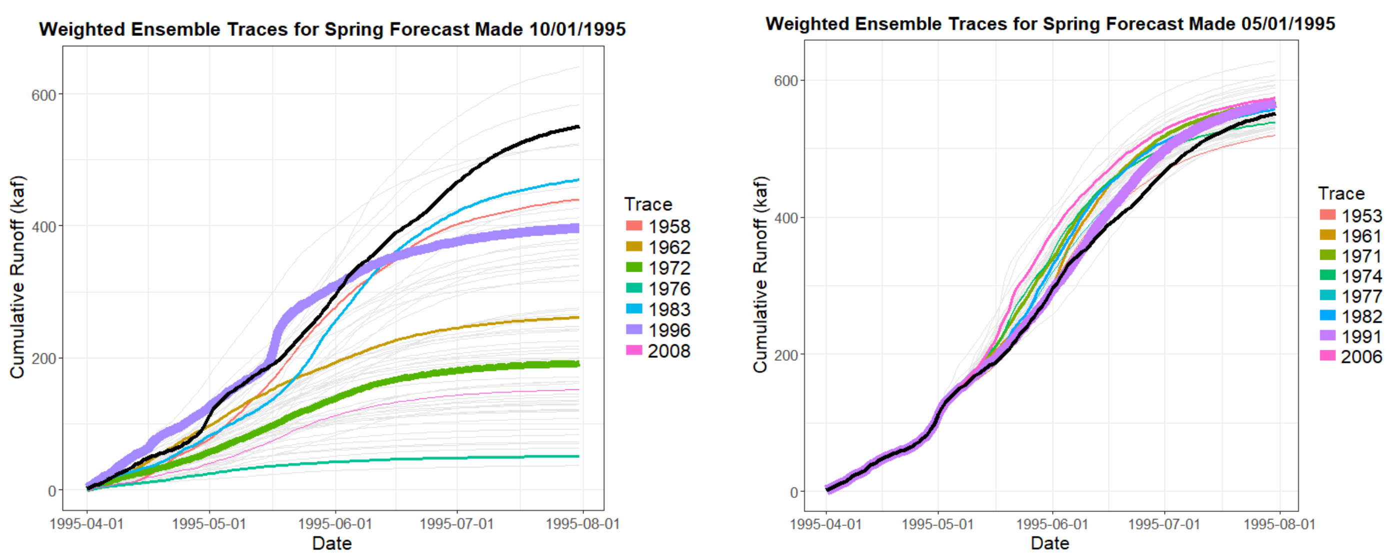


Figure 5. Weighted Traces Shown Against Full Ensemble (*grey*) and Observations (*black*) for October forecast (*left*) and May forecast (*right*), $\boldsymbol{k} = \sqrt{\boldsymbol{N}}$

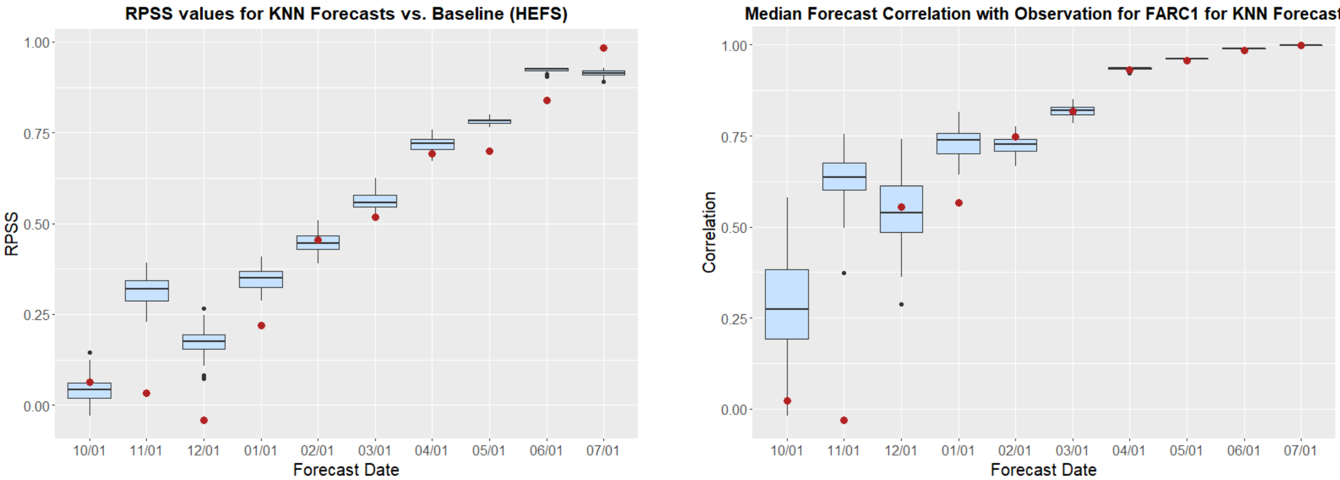


Figure 6. Skill Statistics for K-NN Resampled Forecasts vs. the Baseline (*shown as red points*) for ***k* = *N/2***

the correlation statistics—although the winter months exhibit a large spread, the median simulations (represented by the horizontal lines) show considerably higher correlation with streamflow observations. This means that by weighting ensemble traces, reservoir managers could more accurately predict spring streamflow volumes much earlier in the water year.

Figure 5 (*shown on pg. 6*) provides a visual representation of the new weighted ensemble for two monthly lead times when $k = \sqrt{N}$. The line weight of each trace is proportional to how many times that year was resampled in the modified forecast. In this case, a $k = \sqrt{N}$ means the modified ensemble is only sampling from a subset of 8 traces. This leads to a new ensemble in which fewer traces are resampled many times (for the forecast made October 1st, the year 1996 is resampled 26 times). Although this approach improves the skill of the forecast, and is optimal in terms of maximizing RPSS for the winter months, weighting any one trace this heavily may not be desirable. In order to obtain a finer weighting scheme, the value of *k* can be modified to include a larger sample of nearest neighbors.

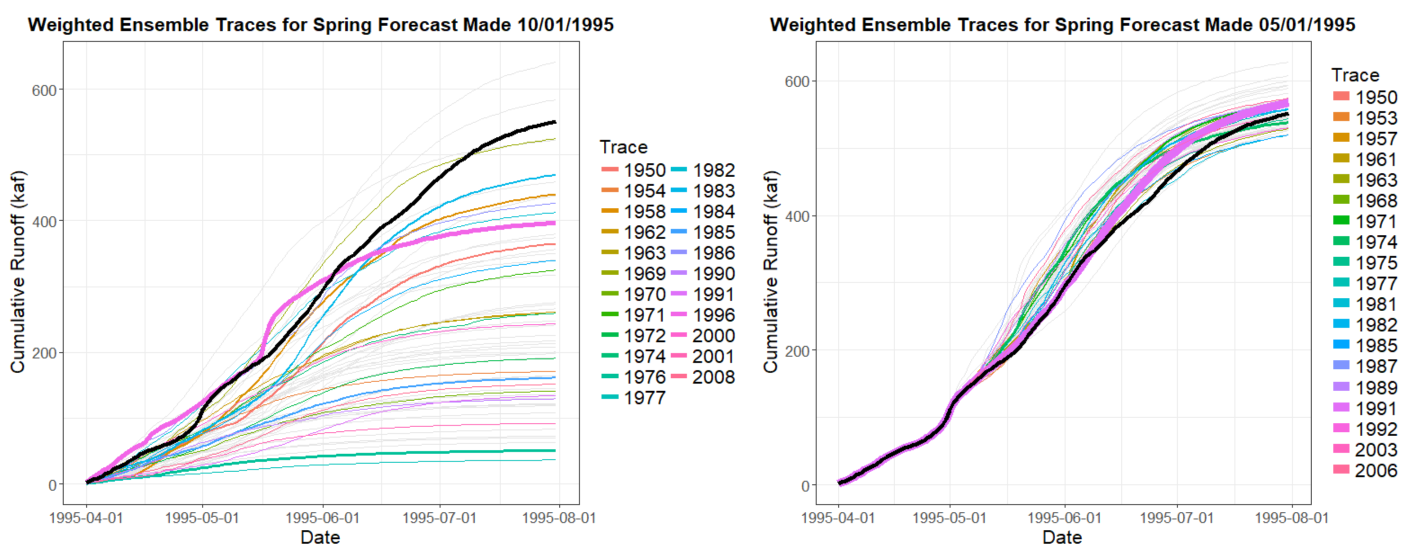


Figure 7. Weighted Traces Shown Against Full Ensemble (*grey*) and Observations (*black*) for October forecast (*left*) and May forecast (*right*), ***k* = *N/2***

Figure 7 (*shown on pg.* 7) presents a modified ensemble for the same forecast year and monthly lead times, with $k = N/2$. Here, the K-NN algorithm resamples from 30 nearest neighbors, producing a larger spread of traces. In other words, for the October 1st forecast 23 unique traces are included in the ensemble (as opposed to 7 for the previous scenario), with the most heavily weighted trace being resampled 10 times instead of 26. **Figure 6** (*shown on pg.* 7) shows the skill statistics when $k = N/2$, which are only slightly reduced from $k = \sqrt{N}$. Thus, there are clearly tradeoffs between a large and small value for k, and the exact value to use may depend on the situation at hand.

As previously mentioned, ensemble forecasts are useful for decision support in that they provide probability distributions that help answer risk-based questions. **Figure 8** provides PDFs produced from HEFS and K-NN ensemble forecasts made on November 1st. These PDFs were developed using a kernel density (nonparametric) estimation function, rather than from a standard family of distributions. In the left plot (which presents forecasts made for 1993 spring streamflow), the modified ensemble exhibits a peak at the observed value, whereas the HEFS ensemble under predicts spring runoff. Although the HEFS ensemble does a better job at predicting the observed value in 2004, the modified ensemble is able to predict the same value with higher certainty. It should be noted that these same results are not reflected for all forecasts made between the years 1985-2010, however, on average the PDF is better captured using this approach. This is what is represented by the increase in RPSS for months November through January.

Conclusions

This paper presents a generalized framework for weighting ensemble forecast members based on a *k*-nearest neighbor bootstrapping approach. The framework was applied to HEFS reforecasts produced for the Truckee River Basin, using spatially averaged SST and geopotential heights as climate indices. Using K-NN to preferentially sample traces from ensemble forecasts improved both RPSS and median forecast correlation with observations, especially for forecasts made at longer lead times. Although this research presents just one possible approach to weighting ensemble members, these results demonstrate the value of post-processing forecasts produced by the NWS. K-NN, as it has been used in this context, is both relatively easy to implement and to convey to a less technical audience, which is important when dealing with a diverse group of stakeholders. After further refinement, this approach could potentially lead to more effective mid-term operating strategies by mitigating the risk of over or under estimating spring runoff volumes.

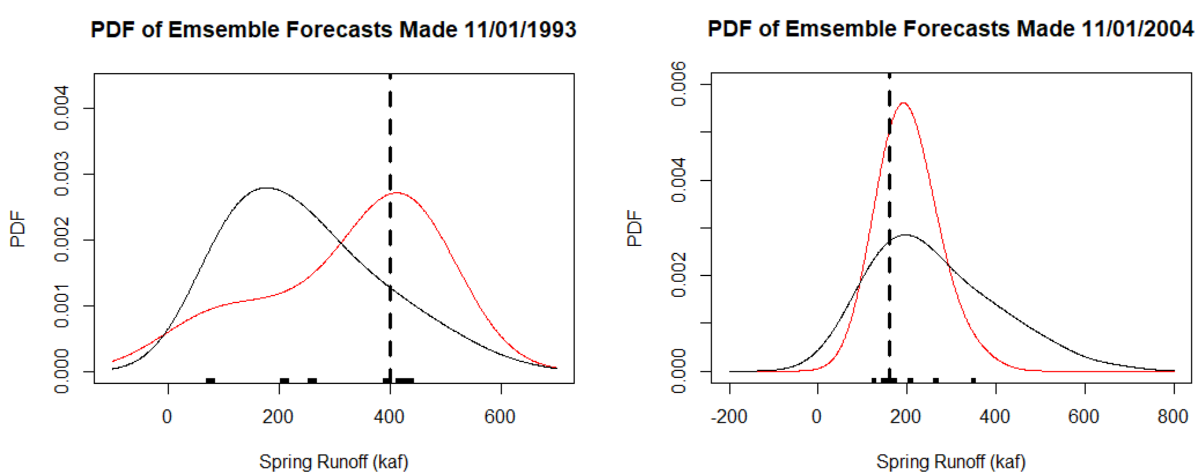


Figure 8. PDFs Produced from HEFS (*black*) and K-NN (*red*) Ensembles Against Observations (*dashed*)

References

Apipattanavis, S., Podestá, G., Rajagopalan, B. and Katz, R. (2007). A semiparametric multivariate and multisite weather generator. *Water Resources Research*, 43(11).

Bradley, A., Habib, M. and Schwartz, S. (2015). Climate index weighting of ensemble streamflow forecasts using a simple Bayesian approach. *Water Resources Research*, 51(9), pp.7382-7400.

Cnrfc.noaa.gov. (n.d.). *CNRFC - Ensemble Plots - Water Resources Forecast Verification Methodology*. [online] Available at: http://www.cnrfc.noaa.gov/WRverification_help.php [Accessed 15 Dec. 2017].

Gangopadhyay, S., Clark, M. and Rajagopalan, B. (2005). Statistical downscaling using K-nearest neighbors. *Water Resources Research*, 41(2).

Gangopadhyay, S., Harding, B., Rajagopalan, B., Lukas, J. and Fulp, T. (2009). A nonparametric approach for paleohydrologic reconstruction of annual streamflow ensembles. *Water Resources Research*, 45(6).

Grantz, K., Rajagopalan, B., Clark, M. and Zagona, E. (2005). A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts. *Water Resources Research*, 41(10).

Hamlet, A. and Lettenmaier, D. (1999). Columbia River Streamflow Forecasting Based on ENSO and PDO Climate Signals. *Journal of Water Resources Planning and Management*, 125(6), pp.333-341.

Lall, U. and Sharma, A. (1996). A Nearest Neighbor Bootstrap For Resampling Hydrologic Time Series. *Water Resources Research*, 32(3), pp.679-693.

Moradkhani, H. and Meier, M. (2010). Long-Lead Water Supply Forecast Using Large-Scale Climate Predictors and Independent Component Analysis. *Journal of Hydrologic Engineering*, 15(10), pp.744-762.

Najafi, M., Moradkhani, H. and Piechota, T. (2012). Ensemble Streamflow Prediction: Climate signal weighting methods vs. Climate Forecast System Reanalysis. *Journal of Hydrology*, 442-443, pp.105-116.

Prairie, J., Rajagopalan, B., Fulp, T. and Zagona, E. (2006). Modified K-NN Model for Stochastic Streamflow Simulation. *Journal of Hydrologic Engineering*, 11(4), pp.371-378.

Rajagopalan, B. and Lall, U. (1999). Ak-nearest-neighbor simulator for daily precipitation and other weather variables. *Water Resources Research*, 35(10), pp.3089-3101.

Regonda, S., Rajagopalan, B., Clark, M. and Zagona, E. (2006). A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River Basin. *Water Resources Research*, 42(9).

Werner, K., Brandon, D., Clark, M. and Gangopadhyay, S. (2004). Climate Index Weighting Schemes for NWS ESP-Based Seasonal Volume Forecasts. *Journal of Hydrometeorology*, 5(6), pp.1076-1090.

Wood, A. and Lettenmaier, D. (2006). A Test Bed for New Seasonal Hydrologic Forecasting Approaches in the Western United States. *Bulletin of the American Meteorological Society*, 87(12), pp.1699-1712.

Yates, D., Gangopadhyay, S., Rajagopalan, B. and Strzepek, K. (2003). A technique for generating regional climate scenarios using a nearest-neighbor algorithm. *Water Resources Research*, 39(7).