Climate Index Weighting Schemes for NWS ESP-Based Seasonal Volume Forecasts

KEVIN WERNER AND DAVID BRANDON

Colorado Basin River Forecast Center, Salt Lake City, Utah

MARTYN CLARK

Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado

SUBHRENDU GANGOPADHYAY

Cooperative Institute for Research in Environmental Sciences, and Department of Civil, Environmental, and Architectural Engineering, University of Colorado, Boulder, Colorado

ABSTRACT

This study compares methods to incorporate climate information into the National Weather Service River Forecast System (NWSRFS). Three small-to-medium river subbasins following roughly along a longitude in the Colorado River basin with different El Niño–Southern Oscillation signals were chosen as test basins. Historical ensemble forecasts of the spring runoff for each basin were generated using modeled hydrologic states and historical precipitation and temperature observations using the Ensemble Streamflow Prediction (ESP) component of the NWSRFS.

Two general methods for using a climate index (e.g., Niño-3.4) are presented. The first method, post-ESP, uses the climate index to weight ensemble members from ESP. Four different post-ESP weighting schemes are presented. The second method, preadjustment, uses the climate index to modify the temperature and precipitation ensembles used in ESP. Two preadjustment methods are presented. This study shows the distance-sensitive nearest-neighbor post-ESP to be superior to the other post-ESP weighting schemes. Further, for the basins studied, forecasts based on post-ESP techniques outperformed those based on preadjustment techniques.

1. Introduction and background

Ensemble streamflow forecasts are made routinely by the National Weather Service (NWS) for seasonal river volumes using Ensemble Streamflow Prediction (ESP), which is a component of the NWS River Forecast System (NWSRFS). ESP uses the current hydrologic model states as initial conditions and drives the model using historical temperature and precipitation (Day 1985). ESP produces a flow trace that corresponds to a particular year of historical weather. Taken together, the ensemble of flow traces may be transformed into a probabilistic forecast for any future variable. Current NWS methodology allows a user to choose different methods for transforming the ensemble values into a probabilistic forecast. The ensemble values can be used to define an empirical probability distribution or in fitting a probability distribution function (i.e., normal, weibull, etc.). However, used alone, this procedure does not account for any additional knowledge of the climate system,

such as the El Niño–Southern Oscillation (ENSO) state, that a forecaster may have. The ESP system includes two weighting methods to account for the current climate state or forecasted climate conditions. One method is a preadjustment technique that applies shifts to the temperature and precipitation inputs based on climate forecasts. The current NWS practice is to use climate forecasts produced at the Climate Prediction Center (CPC). The second method is a post-ESP technique that allows a user to weight the resulting flow traces based on user-defined weights. The current practice is to use a technique developed at the Alaska Pacific River Forecast Center (APRFC), known as the Alaska technique (L. Rundquist 1995, personal communication). The Alaska technique weights ESP-generated flow traces according to the correspondence of their forcing with a particular CPC forecast.

Several studies have demonstrated significant forecast improvements when climate information is used as part of the streamflow forecasting method. For example, Hamlet and Lettenmaier (1999) modified the ESP approach by restricting attention to years (ensemble members) that were similar in terms of the phase of ENSO

Corresponding author address: Kevin Werner, NWS/CBRFC, 2242 W. North Temple, Salt Lake City, UT 84116.
E-mail: Kevin.Werner@noaa.gov

and the phase of the Pacific decadal oscillation (PDO). In most cases this provided a set of ensembles that were more tightly clustered, and closer to observed runoff, than the full ensemble. On short forecast lead times of less than 2 weeks, Clark and Hay (2004) recently demonstrated significant forecast improvements in snowmelt-dominated river basins when downscaled output from a numerical weather prediction model was used to replace the historical ESP traces.

This study describes and compares four post-ESP weighting schemes to apply to river flow ensemble traces based on the observed relationships between flow and ENSO. Sixteen climate indices were examined as potential candidates to be used in the weighting scheme. The Niño-3.4 sea surface temperatures were determined to be the strongest candidate index. Each weighting scheme is described, including discussion of and optimization to any adjustable parameters. Each weighting scheme is optimized and applied to each of the test basins for comparison to other weighting schemes. The study also examines the preadjustment technique currently used in the NWS, using the Niño-3.4 index to develop synthetic climate forecasts. The ranked probability skill score (RPSS) is used to assess the skill of the various techniques (Wilks 1995).

2. Study areas and index choice

Three river basins in the Colorado Basin River Forecast Center (CBRFC) area were selected as study basins: The Green River Headwaters above Warren Bridge, Wyoming (WBRW4), in the northern part of the CBRFC area (268 mi^2), the Colorado River above Cameo, Colorado (CAMC2), in the middle part of the CBRFC area (8050 mi^2), and the Salt River above Chrysotile, Arizona (SLCA3), in the southern portion of the CBRFC area (2849 mi^2). Figure 1 shows these basins in relation to the Colorado River basin and the western United States.

Previous studies have shown an El Niño signal to exist in the southern portion of the CBRFC area (e.g., Clark et al. 2001 and Ropelewski and Halpert 1987). Therefore a significant a correlation was expected between runoff volumes for SLCA3 and tropical Pacific SST indices. Several extratropical indices where no signal was expected were also correlated with spring runoff volumes. Figure 2 shows lagged month-by-month correlations between seasonal spring runoff volumes for each of these basins against 18 climate indices (see Table 1 for a description of the indices). These indices include those that describe both extratropical teleconnection patterns (Barnston and Livezey 1987) as well as those that describe tropical SST and circulation anomalies associated with ENSO. They are commonly used in climate analyses. Spring runoff is the April through July volume for CAMC2 and WBRW4 and the February through May volume for SLCA3.

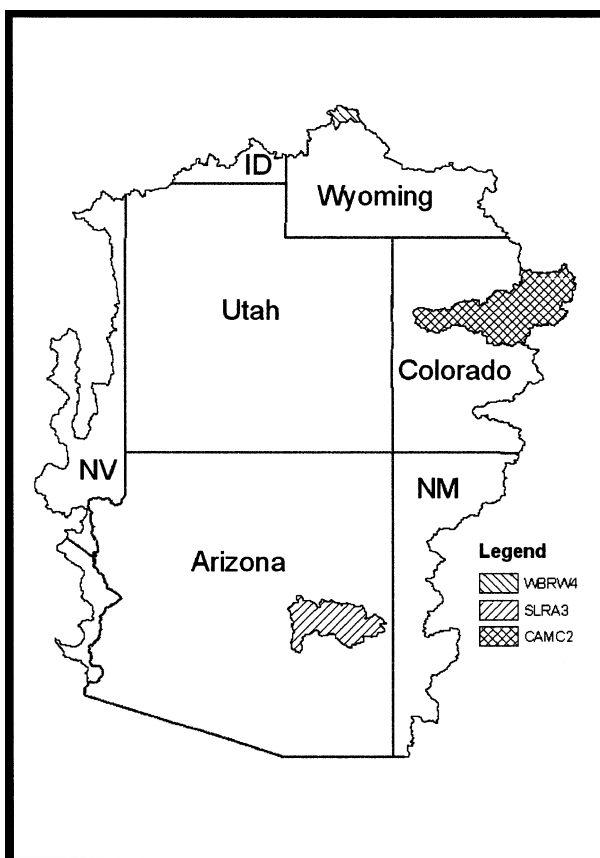Correlations are shown for the year prior to observed

FIG. 1. Location of test basins.

runoff and the year of the runoff. The 5% and 95% significance levels indicated by dotted lines in Fig. 2 were calculated with a bootstrapping technique (Shao et al. 1996). Correlations were computed between the observed runoff and a randomly reshuffled time series of the Niño-3.4 index. This was repeated 1000 times to establish the 5% and 95% significance levels. Most of the correlations with the extratropical climate indices are not statistically significant (at the 90% level). However, the correlations with the ENSO-related indices are statistically significant for various months for WBRW4 and SLCA3. Note correlations are opposite for these two basins, in support of previous work (e.g., Cayan and Webb 1992; Clark et al. 2001). The ENSO-related indices are less significant for CAMC2. The Pacific decadal oscillation may be important for WBRW4, but it shows less significance for the two more southerly basins.

SLCA3 was chosen as the initial test basin based on 1) its long period of record (1952–98) and 2) its known strong relationship to ENSO events (Fig. 2; Cayan and Webb 1992; Clark et al. 2001). The correlations between Niño-3.4 and SLCA3 volumes are shown at the bottom in Fig. 2. Note the highest correlations (about 0.45) exist for the November–December–January (NDJ) period
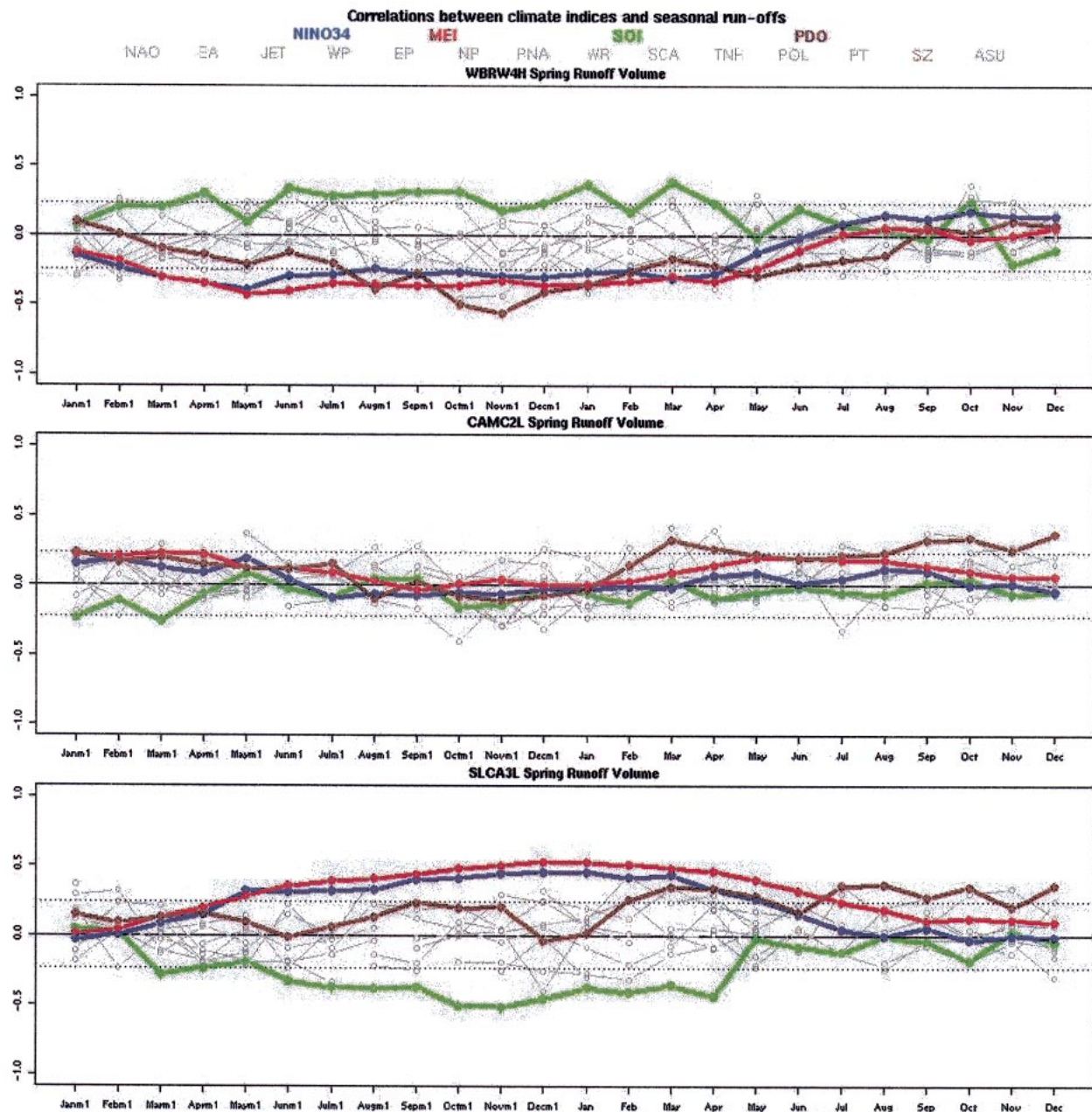
FIG. 2. Correlation between observed seasonal volumes for the three study basins and selected indices (bold) and patterns (light) for the months on the *x* axis; ''m1'' months refer to months from the year prior to the observed volume. Correlations were calculated for observed volumes between 1952–98 (WBRW4 and SLCA3) and 1978–98 (CAMC2). The 5% and 95% significance levels calculated with a bootstrapping method are depicted with a dashed line. Climate indices are listed in Table 1.

preceding the observed volume. Therefore a mean of those 3 months will be used as a basis for the weighting schemes.

NWSRFS was forced with station observations for the period 1952–98. The simulated hydrologic states (e.g., snow-water equivalent, soil moisture) were saved on 1 February for each year. The model estimates of these hydrologic states were used as initial conditions

for ESP reforecasts. By starting the reforecasts on 1 February, midseason snowpack information, which can provide some predictability (McCabe and Dettinger 2002), is implicitly included in the reforecast. However, traditional ESP forecasts weight the ESP ensemble member resulting from each year's historical observations equal to every other ESP ensemble member. In doing so, predictability is entirely derived by knowledge

Label	Pattern
Niño-34	Niño-3.4 index
MEI	Multivariate ENSO index (Wolter and Timlin 1993, 1998)
SOI	Southern Oscillation index
PDO	Pacific decadal oscillation (Mantua et al. 1997)
NAO	North Atlantic Oscillation
EA	East Atlantic pattern
JET	East Atlantic jet pattern
WP	West Pacific pattern
EP	East Pacific pattern
NP	North Pacific pattern
NAO	Pacific–North American pattern
WR	East Atlantic–west Russia pattern
SCA	Scandinavia pattern
TNH	Tropical–Northern Hemisphere pattern
POL	Polar–Eurasia pattern
PT	Pacific transition pattern
SZ	Subtropical zonal pattern
ASU	Asia summer pattern

of the initial hydrologic states rather than from a forecast of weather or climate conditions. By using the traditional ESP forecast as a benchmark for comparison, we may assess the additional predictability that is obtained by including climate information through various methods.

RPSS is used to evaluate the probabilistic forecasts derived from ESP (Epstein 1969; Murphy 1969, 1971; Hersbach 2000). The continuous version of the ranked probability score (RPS) upon which the RPSS is based is given by

$$RPS = \int_{-\infty}^{\infty} [P(x) - P_o(x)]^2 \, dx, \qquad (1)$$

where $P(x)$ is the forecasted exceedence probability for spring runoff volume, and $P_o(x)$ is the observed exceedence probability. For a continuous variable, such as the spring runoff volume in a particular year, the observed probability $P_o(x)$ will be either zero (if x is less than the observed volume) or unity (if x is greater than or equal to the observed volume). The RPS is distance sensitive in that it increasingly penalizes forecasts that contain forecasted probability farther away from the observed quantity. RPSS is based on the RPS and is given by

$$RPSS = 1 - \frac{RPS_f}{RPS_{ref}}, \qquad (2)$$

where RPS_f and RPS_{ref} are the forecast being evaluated and a reference forecast. The reference forecast is often taken to be climatology. RPSS values are less than or equal to unity. Positive RPSS values indicate percent improvement in forecast skill, while negative values in-

dicate the reference forecast is superior to the forecast being tested.

3. Post-ESP weighting scheme

Post-ESP weighting techniques weight ESP output (flow) traces based on information that may add predictive value to the forecast. In this case we use the Niño-3.4 index, averaged over the NDJ period immediately before the issue date of the forecast (i.e., 1 February). The post-ESP technique is described as follows:

1) Compute a vector **X** of absolute differences between the value of the NDJ Niño-3.4 index in the forecast year and the values of the NDJ Niño-3.4 index in all other years. Sort the vector **X** from lowest to highest. That is,

$$\mathbf{X} = (x_1, x_2, \ldots, x_n), \qquad (3a)$$

$$\aleph = [x_{(1)}, x_{(2)}, \ldots, x_{(n)}], \quad x_{(1)} \leq x_{(2)} \cdots \leq x_{(n)}. \quad (3b)$$

2) Compute a vector of weights; $\mathbf{W} = (w_1, w_2, \ldots, w_n)$, for all elements of the sorted vector $\aleph$:

$$w_i = \left[1 - \frac{x_{(i)}}{x_{(k)}}\right]^{\lambda-1}, \qquad x_{(i)} \leq x_{(k)} \qquad (4a)$$

$$w_i = 0, \qquad x_{(i)} > x_{(k)} \qquad (4b)$$

$$k = \text{NINT}\left(\frac{n}{\alpha}\right). \qquad (4c)$$

Here λ is a distance-sensitive weighting parameter, α is a parameter that defines the k nearest neighbors used, and NINT refers to the nearest integer operator. Higher values of λ give more weight to years (ensemble traces) in which the value of the NDJ Niño-3.4 index is closer to the Niño-3.4 index in the forecast year; higher values of α restrict attention to the n/α elements in the vector $\aleph$. The forecast year's trace is excluded by setting its weight to zero.

3) Finally, compute the modified probability assigned to each ESP trace

$$p_i = \frac{w_i}{\sum_{j=1}^{n} w_j}, \qquad (5)$$

where w_i is the weight for the year from which historical data were used to generate the ESP trace. The modified cumulative distribution function (CDF) (i.e., the resultant probabilistic forecast) is simply the integral of the probabilities assigned to each ESP trace. Figure 3 illustrates the post-ESP weights, p_i. Weights were calculated for $\lambda = 1, 3,$ and 10, and $\alpha = 1$ and 5 for the year 1972.

Four cases using this weighting scheme will be presented: equal weights, index difference weights, nearest-
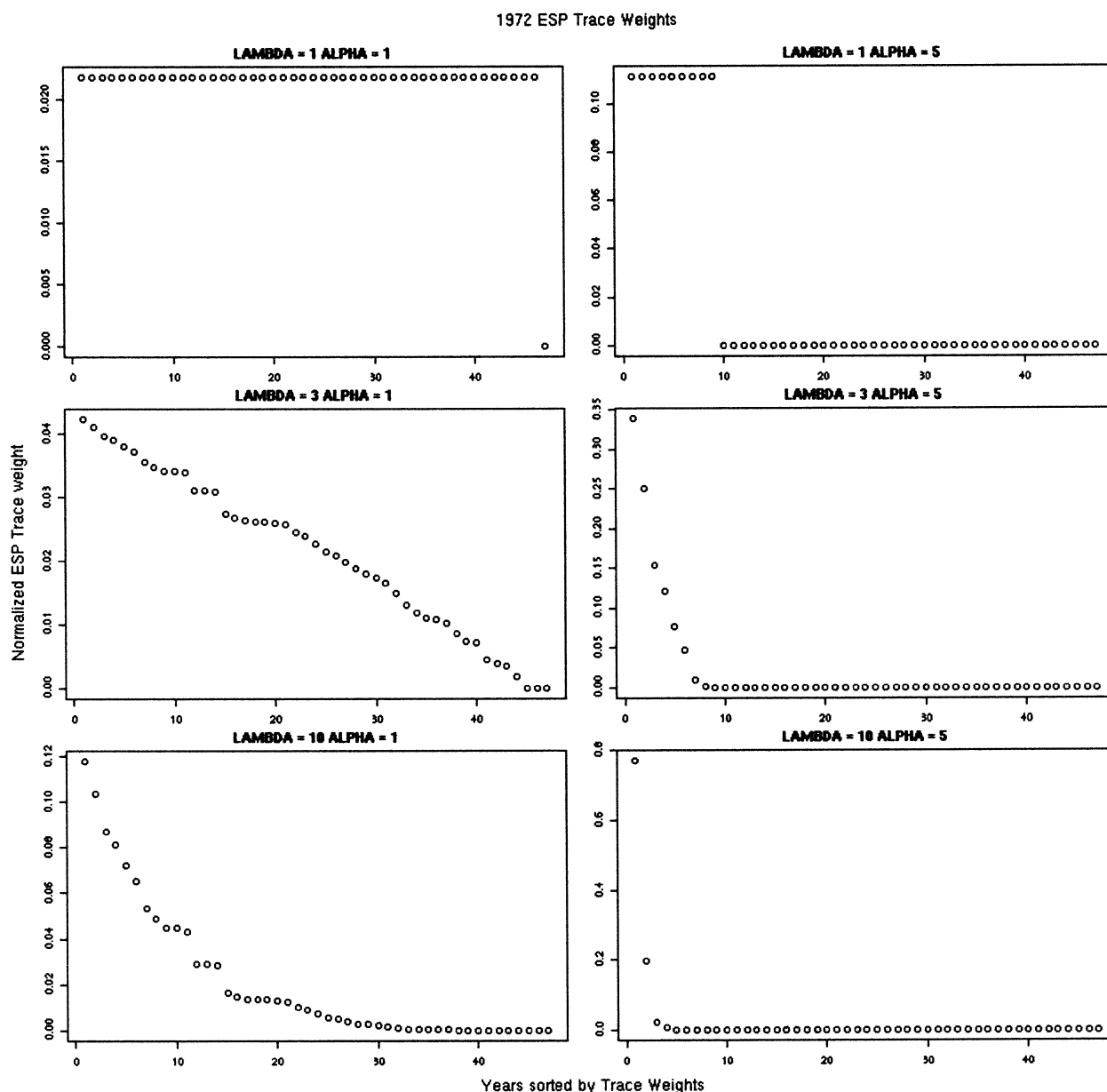
FIG. 3. 1972 post-ESP weights (i.e., p_i) for various α and λ combinations.

neighbor weights, and nearest-neighbors analog weights.

a. Case 1: Equal weights ($\alpha = \lambda = 1$)

The equal weighting (EW) scheme is the traditional weighting scheme applied to ESP forecasts. Each trace except the trace corresponding to the forecast year is given the same weight. As in every other weighting scheme, the trace corresponding to the forecast year is omitted by giving it zero weight. Mean RPSS values were computed for each year for the EW forecast for

SLCA3 with climatology as the reference forecast. The results are shown in Fig. 4. The median RPSS value for all years was 0.41 indicating a 41% improvement over forecasts made using climatology. Similar results were found for CAMC2 (39%; not shown). WBRW4 showed lower forecast improvement over climatology (15%; not shown).

Equal weighting forecasts represent current NWS methodology. To show improvement over the current methodology, EW forecasts will be used as the reference forecast for RPSS calculations for the other weighting schemes and methods presented here.
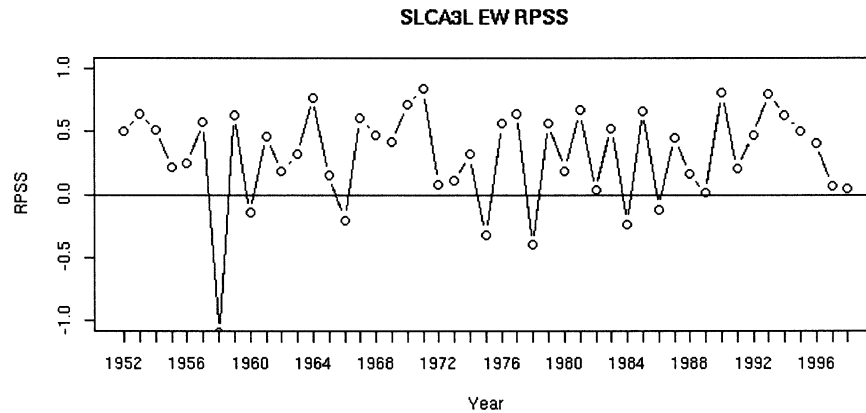
SLCA3L EW RPSS

FIG. 4. RPSS values for EW forecast with climatology as the reference forecast.

b. Case 2: Index difference weights (α = 1, vary λ)

The index difference weighting (IDW) scheme has nonzero weights for all ensemble traces (since $\alpha = 1$), but weights each trace differently based on the absolute difference in the climate index for the forecast year and the climate index for each year in the ensemble trace. Larger values of λ increasingly weight years with smaller absolute differences in Niño-3.4 values. A sensitivity test was conducted on the distance sensitive weighting parameter λ, which was varied between 1 and 40. The median RPSS from the IDW forecasts was computed

for each increment of λ, with the EW forecast used as the reference forecast. This is shown in Fig. 5. For SLCA3 weighted with the NDJ Niño-3.4, the optimal λ is about 20. A different λ value may yield higher skill scores if the IDW scheme is applied to a different basin, a different index, or even a different period of record.

Figure 6 shows RPSS values for each reforecasted year with years indexed according to their NDJ Niño-3.4 value. This ranking may illuminate a systemic improvement in a particular ENSO regime. Although there are a few years with large negative scores, the median

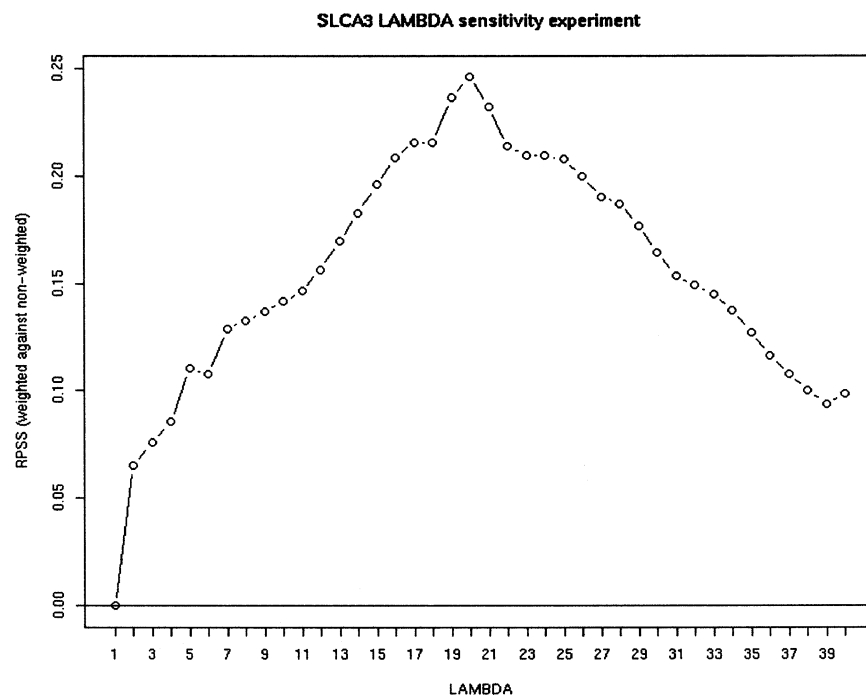**SLCA3 LAMBDA sensitivity experiment**

FIG. 5. Sensitivity test for case 2: index difference weighting. RPSS values are shown for different values of the post-ESP index difference-sensitive adjustable parameter λ for SLCA3 Feb–May volumes weighted with the NDJ Niño-3.4 index. RPSS values calculated using EW as the reference forecast.
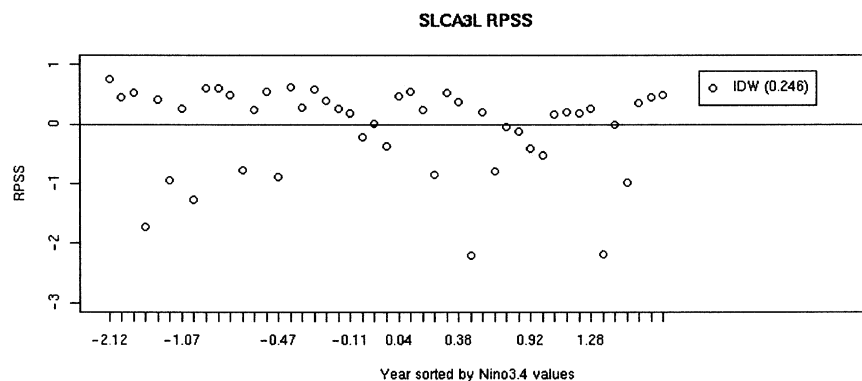
SLCA3L RPSS

FIG. 6. RPSS values using case 2: index difference weighting scheme for SLCA3 Feb–May volumes. The post-ESP scheme uses Niño-3.4 averaged over the preceding NDJ and λ set to 3.0. Years are ranked according to values of NDJ Niño-3.4 on the x axis. The median RPSS value is given in the legend; EW is the reference forecast for RPSS calculations.

RPSS is 0.246, indicating an overall improvement of forecasts with IDW by almost 25% over EW forecasts. No major systemic bias toward a particular ENSO regime was apparent.

c. Case 3: Nearest-neighbor weights ($\lambda = 1$, vary α)

The nearest-neighbor weighting (NNW) scheme uses only the k or N/α years with Niño-3.4 values closest to the forecast year. Each of the k years is weighted equally. Larger values of α include a decreasing number of analog years in the weighting scheme.

As with case 2, a sensitivity test was conducted. Here the nearest-neighbor selection parameter α was varied

from 1 to 10. Figure 7 shows the results of this test. The largest median RPSS values occur near $\alpha = 7$. This indicates only seven ESP traces (or one-seventh of the number of ESP traces) will be included in the post-ESP forecast. Note that as with case 2, this choice of α is unique to the dataset (basin) used in the sensitivity test.

Similar to Fig. 6, Fig. 8 shows RPSS values for each reforecasted year with years ranked according to their NDJ Niño-3.4 value. Here the median forecast improvement over EW is 26%. As with IDW, no systemic bias toward forecast skill improvement in a particular ENSO regime is apparent.

A modified version of this weighting scheme only uses in the forecast the ensemble traces corresponding
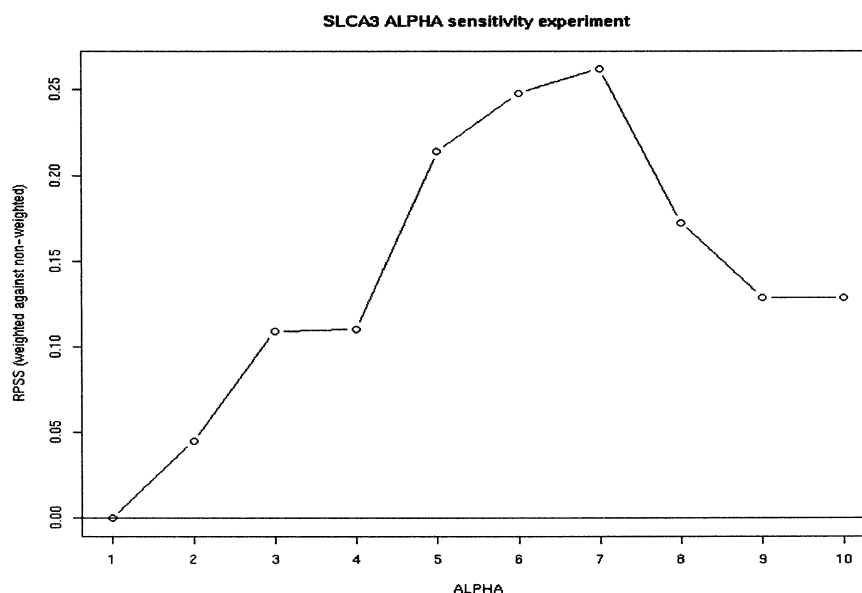
SLCA3 ALPHA sensitivity experiment

FIG. 7. Sensitivity test for case 3: nearest-neighbor weighting. RPSS values are shown for different values of the post-ESP nearest-neighbor selection parameter α for SLCA3 Feb–May volumes weighted with the NDJ Niño-3.4 index. RPSS values calculated using EW as the reference forecast.
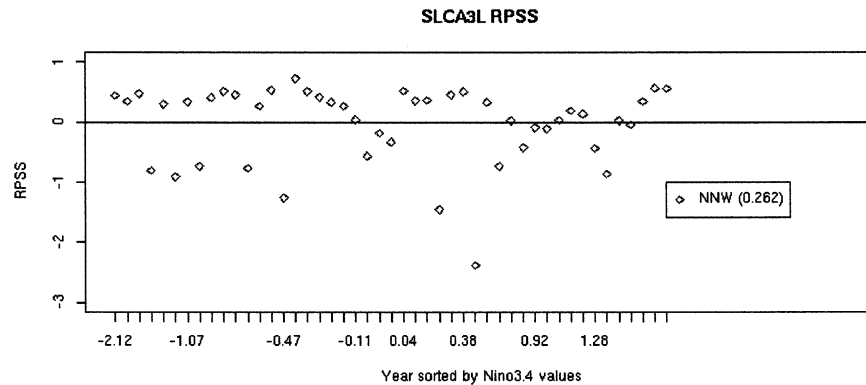
SLCA3L RPSS

FIG. 8. RPSS values using case 3: nearest-neighbor weighting scheme for SLCA3 Feb–May volumes. The post-ESP scheme uses Niño-3.4 averaged over the preceding NDJ and α set to 4.5. Years are ranked according to values of NDJ Niño-3.4 on the x axis. The median RPSS value is given in the legend; EW is the reference forecast for RPSS calculations.

to years with Niño-3.4 (or whatever other index is chosen) in the same tercile as the forecast year. For example, if a year is classified as "El Niño," attention is restricted to just El Niño years, and the El Niño traces are assigned equal weights in computing the forecast CDF. This approach is similar to that used by Hamlet and Lettenmaier (1999) and is sometimes used at NWS field offices when the ENSO state is known (i.e., El Niño, neutral, or La Niña). The difference between this method and the method presented above is that years close to the boundary of a given category will be compared to all other years in that category, and not to the k closest neighbors. Figure 9 shows the RPSS values computed with the modified nearest-neighbor weights or tercile analogue weights (TAW) relative to EW schemes. Here the forecasts have less skill than the reference EW forecasts, as indicated by a negative RPSS value (-0.04).

d. Case 4: Distance-sensitive nearest-neighbor weighing (both α and λ vary)

The distance-sensitive nearest-neighbor weighting (DSNNW) utilizes the concepts of both cases 2 and 3.

As with case 3, only k nearest neighbors are included in the post-ESP weighting scheme. Further, as in case 2, the weights assigned to each of the included years are dependent on the magnitude of the absolute difference in the climate index for the ESP trace year and the forecast year. As with other methods, a sensitivity test was conducted jointly on λ and α parameters where each was varied independently of the other between 1 and 10. The results were plotted in the contour plot shown in Fig. 10. The optimal combination of λ and α was taken to be the maximum point on the plotted RPSS surface. Finer-scale contour plots show the maximum median RPSS value occurs near $\lambda = 1.5$ and $\alpha = 6$. Note that although we used the global optimum here, several different parameter combinations will result in similar skill.

Figure 11 shows the RPSS score relative to EW for each year ranked according to the NDJ Niño-3.4 value. As expected from Fig. 9, the results are quite similar to NNW. The median RPSS indicates a nearly 28% median forecast skill improvement, which is slightly higher than that for IDW and NNW. Since λ is close to unity,
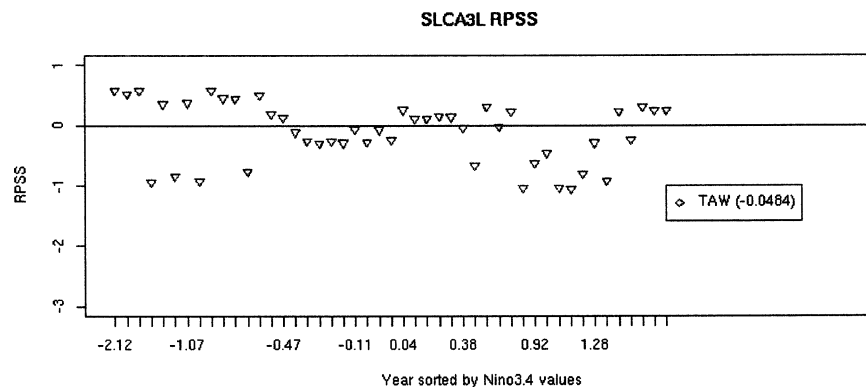
SLCA3L RPSS

FIG. 9. RPSS values using case 3a: tercile analog weighting scheme for SLCA3 Feb–May volumes weighted with Niño-3.4 averaged over the preceding NDJ. The median RPSS value is given in the legend; EW is the reference forecast for RPSS calculations.
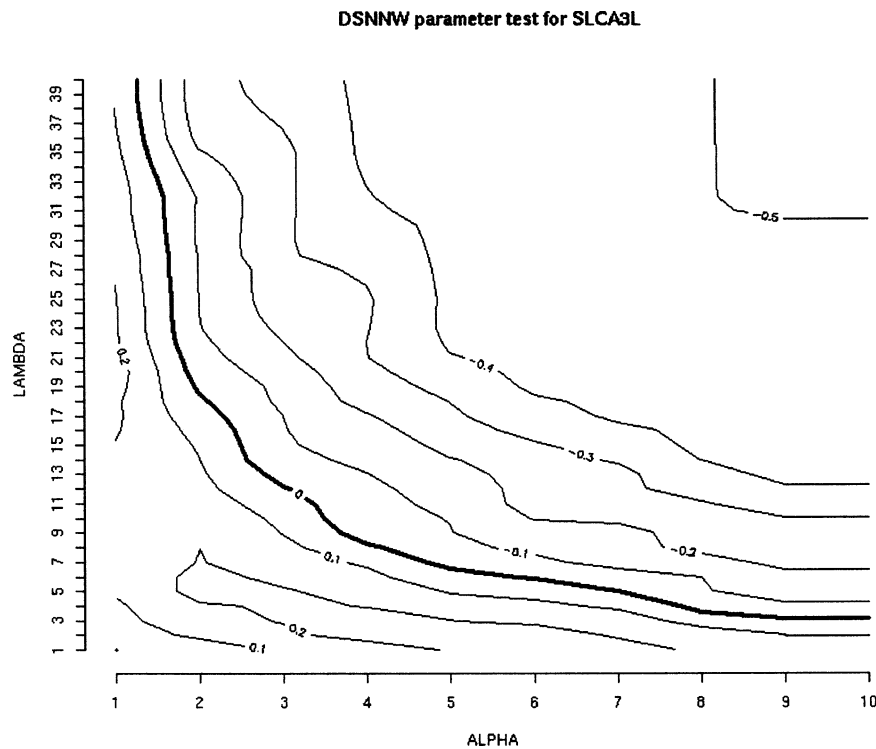
DSNNW parameter test for SLCA3L

FIG. 10. Sensitivity test on the post-ESP index difference-sensitive adjustable parameter λ and nearest-neighbor selection parameter α for SLCA3 Feb–May volumes weighted with Niño-3.4 averaged from the preceding Nov through Jan. Median RPSS over all years 1952–98 are contoured; EW is the reference forecast in RPSS calculations. The maximum RPSS value occurs near $\alpha = 6$ and $\lambda = 1.5$.

it is expected that the DSNNW results would be similar to the NNW. This suggests that for this basin and this forecast, the inclusion of the distance-sensitive weighting parameter λ does not add much forecast skill.

4. Weighting scheme comparisons

In order to assess strengths and weaknesses of the different weighting schemes, they are compared for the test basins roughly following a line of longitude in the CBRFC area described earlier.

a. SLCA3

Figure 12 shows RPSS scores for each of the weighting schemes, using the EW forecasts as the reference. The adjustable parameter(s) in all weighting schemes were optimized as described in section 3. For SLCA3,
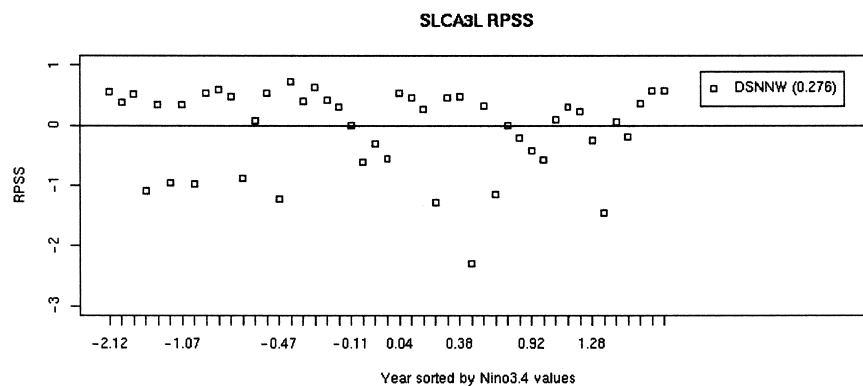
SLCA3L RPSS

FIG. 11. RPSS values using case 4: distance-sensitive nearest-neighbor weighting scheme for SLCA3 Feb–May volumes. The post-ESP scheme uses Niño-3.4 averaged over the preceding NDJ and $\alpha = 6$ and $\lambda = 1.5$. Years are ranked according to values of NDJ Niño-3.4 on the x axis. The median RPSS value is given in the legend; EW is the reference forecast for RPSS calculations.

SLCA3L RPSS
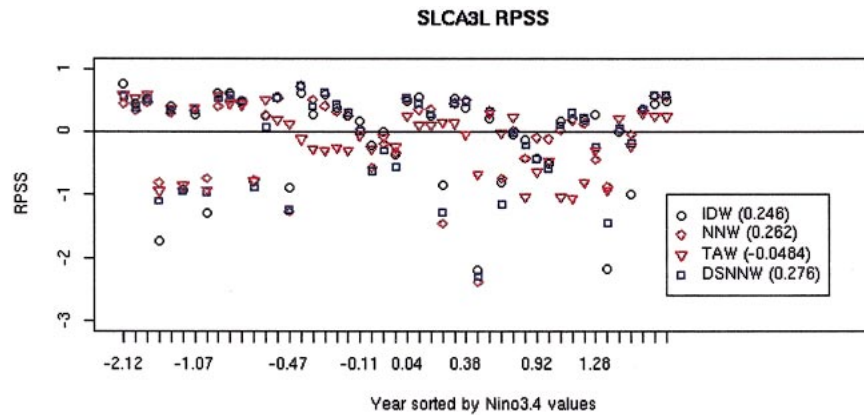

FIG. 12. RPSS for all weighting schemes for SLCA3 with equal weights as the reference forecast for RPSS calculation (RPSS = 0 = horizontal line for EW). Mean RPSS values are shown in parentheses on the legend. Weights calculated based on the preceding NDJ Niño-3.4 value. Adjustable parameters are the same as before. The median RPSS value is given in the legend.

DSNNW has the greatest median skill (0.276), followed by NNW (0.262) and IDW (0.246). For most years, the RPSS values for IDW, NNW, and DSNNW are quite similar. For this basin, the inclusion of both adjustable parameters [the nearest-neighbor selection parameter (α) and the distance sensitive parameter (λ)] in the post-ESP weighting scheme had only a small effect. In this case, the inclusion α or λ alone have similar effects in biasing to similar years, and neither parameter appears preferable.

b. CAMC2—Colorado River near Cameo, Colorado

CAMC2 is located on the Colorado River near Grand Junction, Colorado. It includes inflow from all the tributaries above. It is thought to be a nodal region where ENSO has little effect (i.e., areas to the north have opposite teleconnection patterns to ENSO as do areas to the south) (Cayan and Webb 1992; Clark et al. 2001). As such, we do not expect to gain much skill by applying the postweighting schemes based on Niño-3.4 indices. Indeed, the correlations between spring runoff and NDJ Niño-3.4, as shown in Fig. 2, were not statistically significant.

As was done in the previous section for SLCA3, a sensitivity test was conducted on the adjustable parameters. The values of the parameters maximizing the RPSS were determined to be $\lambda = 1$ and $\alpha = 1.75$. Therefore NNW and DSNNW provided the same solution. This choice of α and λ indicates the 27 nearest neighbors will be equally weighted. Figure 13 shows calculated RPSS for all weighting schemes applied to 1 February reforecasts of April through July water volume for CAMC2. As with SLCA3, the years are ordered by the NDJ Niño-3.4 value. Both the NNW and DSNNW schemes showed a median RPSS of just under
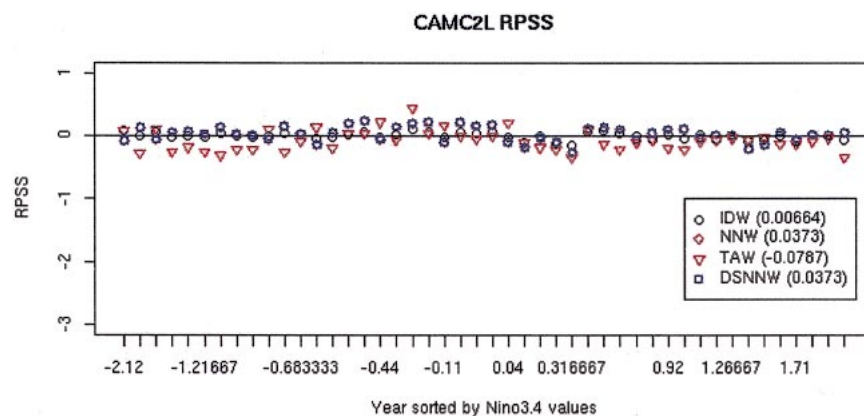
CAMC2L RPSS

FIG. 13. RPSS for all weighting schemes for CAMC2 with equal weights as the reference forecast for RPSS calculation (RPSS = 0 = horizontal line for EW). Mean RPSS values are shown in parentheses on the legend. Weights are based on the preceding NDJ Niño-3.4 values. The median RPSS value is given in the legend.
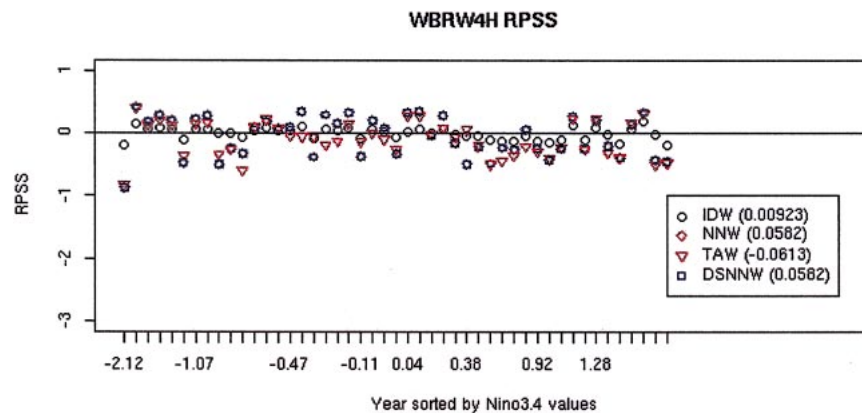
WBRW4H RPSS

FIG. 14. RPSS for all weighting schemes for WBRW4 with equal weights as the reference forecast for RPSS calculation (RPSS = 0 = horizontal line for EW). Mean RPSS values are shown in parentheses on the legend. Apr–Jul WBRW4 runoff volumes are weighted on NDJ Niño-3.4 values. The median RPSS value is given in the legend.

4%, indicating only a small improvement over EW. Unlike SLCA3, the inclusion of the distance-sensitive parameter, λ, contributed nothing to the improvement over EW.

c. WBRW4—Green River at Warren Bridge, Wyoming

WBRW4 is the headwater basin of the Green River near Daniel, Wyoming. It is the northernmost basin in the CBRFC area. The teleconnections with ENSO are of the opposite sign and weaker than for SLCA3. The maximum correlations between the Niño-3.4 monthly index and April through July seasonal discharges are about −0.3 for May of the year prior to and February of the year of the observed discharge. This value is a little over half the observed correlation for SLCA3. The correlation between the NDJ Niño-3.4 value and the spring runoff is about −0.3. Although the Pacific decadal oscillation index shows greater correlations with WBRW4 spring runoff, its variability is on a longer time scale than the historical record at WBRW4. Therefore, as with the other basins, the NDJ Niño-3.4 index is used here. Figure 14 shows RPSS calculations similar to the other two basins described. For WBRW4, 1 February ESP reforecasts were made for 1952 through 1998. As with the other basins, the ensemble traces were then weighted with the 3-month mean NDJ Niño-3.4 index from the period immediately preceding the runoff.

A sensitivity experiment on the adjustable parameters showed optimal values for $\lambda = 1$ and $\alpha = 3$. The RPSS resulting from the various weighting schemes is shown in Fig. 14. As with CAMC2, the DSNNW scheme is identical to NNW. Forecast skill improvements are nearly 6%, which is only slightly better than for CAMC2.

5. Pre-ESP

The NWSRFS has a capability to adjust the mean areal temperature (MAT) and mean areal precipitation (MAP) prior to input into ESP-based climate forecasts (i.e., pre-ESP). The current practice is to use forecasts from CPC using a technique in the NWS software known as the preadjustment technique (Perica 1998). The preadjustment technique modifies the MAT/MAP input into ESP. An additive adjustment is made to MATs while a multiplicative adjustment is made to MAPs based on a particular CPC forecast. These adjustments are calculated from "shifted anomalies" in the distributions of the forecasts. The adjustments are usually minor. The preadjustment technique is the most commonly used method by NWS River Forecast Centers to account for climate forecasts in long-term water supply forecasts.

Results from the post-ESP experiments in the previous section were compared to pre-ESP results. Since a long-term archive of CPC forecasts does not exist and the post-ESP methods are based directly on climate indices rather than CPC forecasts, two methods were used to emulate the existing preadjustment method. The first method used the observed linear regression relationships between the Niño-3.4 index and MAT/MAP anomalies. The second method used resampling methodology. As with post-ESP, both methods base ENSO relationships on the NDJ Niño-3.4 index.

a. Linear ENSO–"weather" relationships

Relationships between seasonal MAT and MAP anomalies and the NDJ Niño-3.4 index were computed with a linear regression analysis:

$$\text{MAT}_{\text{anom}} = a_T + b_T \text{Niño-3.4}, \quad (6)$$

$$\text{MAP}_{\text{anom}} = a_P + b_P \text{Niño-3.4}, \quad (7)$$

where MAT_{anom} and MAP_{anom} are the seasonal MAT and MAP anomalies for the forecast season. Niño-3.4 is the NDJ Niño-3.4 index for the forecast year. The a and b
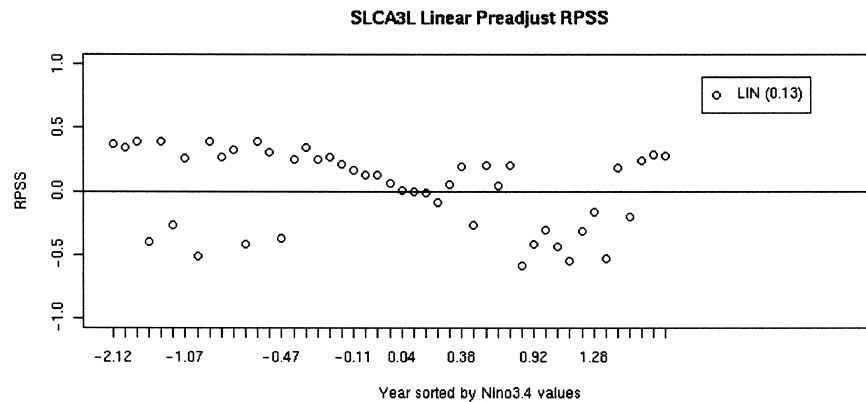
SLCA3L Linear Preadjust RPSS

FIG. 15. RPSS from SLCA3 with the NWSRFS preadjustment technique based the observed linear correlations between NDJ Niño-3.4 and MATs and MAPs. Results are sorted by NDJ Niño-3.4 values. The median RPSS value is given in the legend.

terms are found through a linear regression analysis minimizing the squared error function.

SLCA3 seasonal mean MATs exhibit a negative correlation with Niño-3.4 while seasonal total MAPs have a positive correlation with Niño-3.4. RPSS values were calculated using the EW post-ESP reforecasts from the previous section as the reference forecast. The resulting RPSS is shown in Fig. 15. Overall the median RPSS is 13%, indicating a small improvement of the forecast. However, there are many years, particularly La Niña years, with RPSS > 0. ENSO-neutral years showed very little difference in forecast skill relative to La Niña and El Niño years. It is speculated that the strong La Niña signal in Arizona is responsible for the generally higher RPSS values during La Niña years.

b. Resampling method

The resampling method is an alternative approach for creating synthetic CPC forecasts based on a climate index. First, the climate index value for a given reforecast year is used to find the historical years that are its nearest neighbors. Random resampling of monthly MAT and MAP values from the nearest neighbors is done to construct synthetic time using the same historical years as before, 1952–98. One thousand synthetic MAT and MAP years were created by resampling monthly MAT and MAP values at random from the 15 nearest neighbors according NDJ Niño-3.4 values. Running 3-month values were computed for each synthetic year. The number of synthetic years in each extreme tercile (i.e., warm or cold for MATs) was computed. Synthetic CPC reforecasts were set to the tercile with the most number of synthetic years if more than one-third of the synthetic years fell in that tercile. For example, for a particular forecast year, 450 of the synthetic MATs were in the cold tercile, 350 were in the near-normal tercile, and 200 were in the warm tercile. Therefore the synthetic CPC reforecast for that year would be for cold anomalies of $(450 - 333)/1000$, or 11.7% increased likelihood.

RPSS for SLCA3 reforecasts based on the resampling preadjustment technique are shown in Fig. 16. The over-
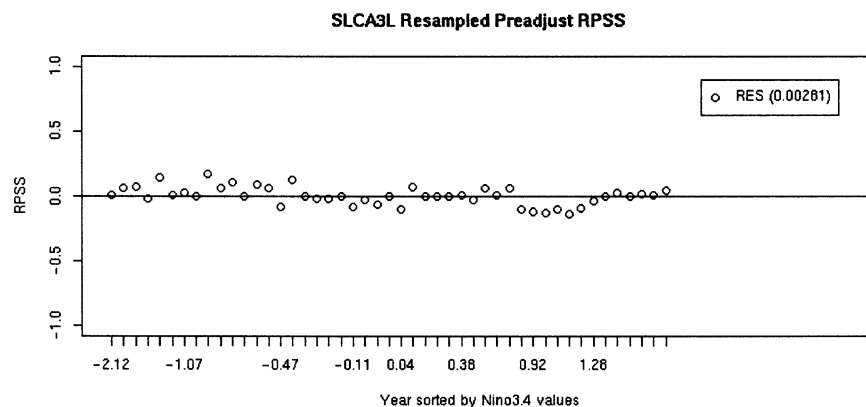
SLCA3L Resampled Preadjust RPSS

FIG. 16. RPSS from SLCA3 with the NWSRFS preadjustment technique based the resampling method to create synthetic CPC reforecasts. Results are sorted by NDJ Niño-3.4 values. The median RPSS value is given in the legend.
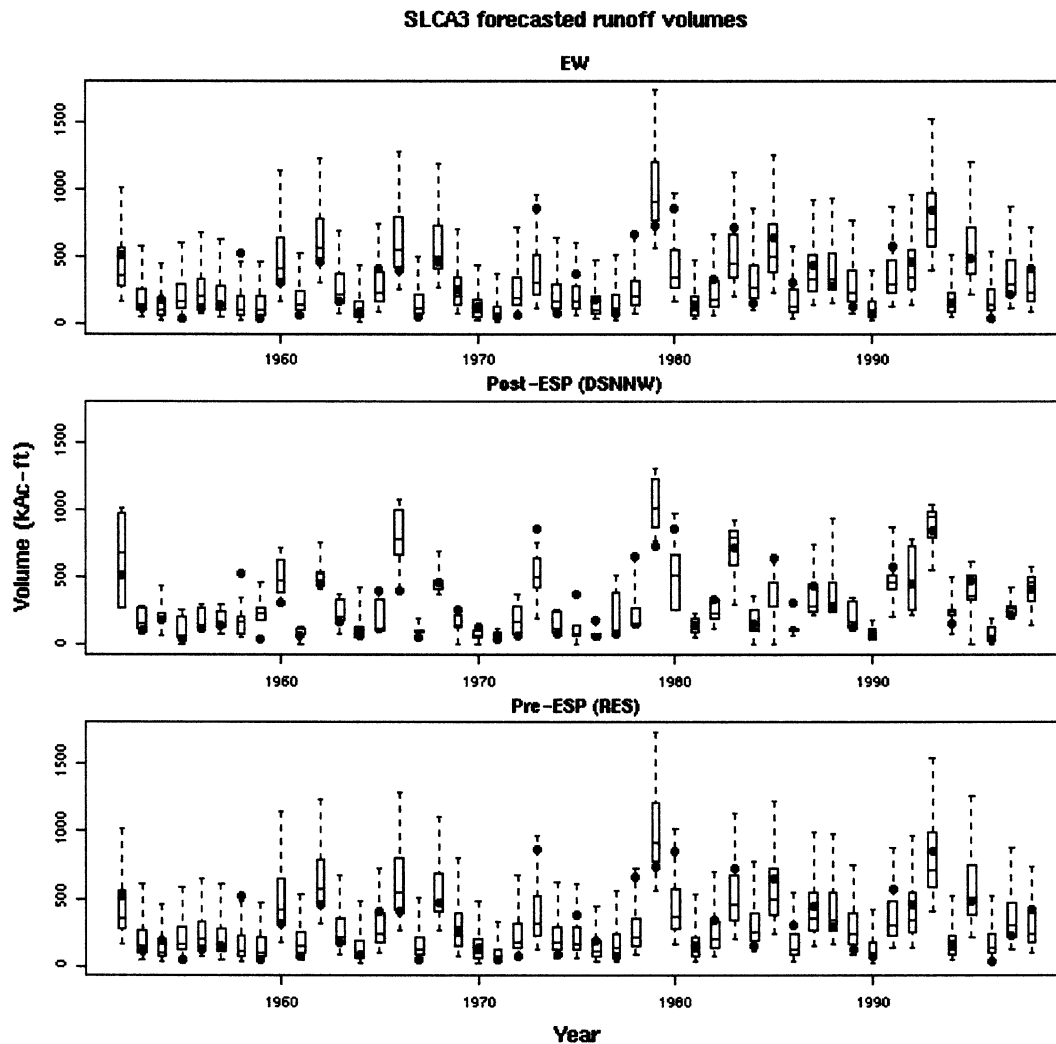
FIG. 17. Forecasted runoff volumes for SLCA3 using three different weighting methods: (top) EW, (middle) post-ESP DSNNW, and (bottom) pre-ESP resampling. Observed runoffs are shown with dots for each year.

all skill improvement was extremely minimal (less than 1%). Further, individual years showed very minimal differences in forecast skill relative to EW. This indicates that the preadjustment technique as applied here (i.e., with the resampling method to create synthetic MATs and MAPs) may not be capable of significantly improving an ESP forecast.

6. Forecasted runoff example

A series of retrospective runoff volume forecasts was made for SLCA3 using the weighting schemes described in this study. Figure 17 shows the "forecasts" made for each year using three of the methods presented: EW, post-ESP DSNNW, and pre-ESP resampling. It is apparent from the figure that the DSNNW reduces the spread of the forecast for most years (e.g., 1959) whereas pre-ESP does not. The previous RPSS results that

showed DSNNW superior to EW and pre-ESP is apparent here as well, although this result is not true for all years (e.g., 1985).

7. Summary and conclusions

Three basins in the Colorado River basin were used in the study, one each in the northern, middle, and southern parts of the Colorado basin. This was an attempt to capture and examine the well-known ENSO climate signal. The Niño-3.4 index was chosen as a teleconnection index that could introduce information on the climate state into the forecast process. The NWS Ensemble Prediction System was used to produce ensembles of streamflow forecasts in reforecast mode. The weighting schemes were applied in a post-ESP process. A separate evaluation was applied using a pre-ESP technique. Both techniques are available to NWS River Forecast Centers.

Improvements and differences between the techniques were evaluated using the RPSS. The goal was to find if and which weighting technique(s) would produce improved seasonal runoff forecasts in the Colorado basin. A secondary goal was to quantify the improvement by showing a percent improvement over the equal weights weighting scheme or climatology.

For SLCA3, the inclusion of both of the post-ESP parameters, the distance-sensitive parameter λ, and the nearest-neighbor selection parameter α in the DSNNW showed the best improvements in forecast skill. However, for WBRW4 and CAMC2 the distance-sensitive parameter λ was not important and DSNNW was the same as NNW. For SLCA3, DSNNW was only slightly better than NNW, suggesting that the distance-sensitive weighting parameter may not be very important for that basin either. As was expected, forecast improvements were more substantial (nearly 28%) for the basin with the stronger ENSO correlations (SLCA3) than for the basin with minimal ENSO correlation (CAMC2 showed only 4%).

The optimization of the post-ESP parameters presented in this paper was specific to seasonal runoff volume forecasts. The post-ESP weighting technique presented here would be applicable to any forecast quantity (such as the seasonal peak flow value or date). However, a different set of post-ESP parameters would be optimal for each different forecast quantity. This point suggests the applicability of the post-ESP technique to forecast problems outside the scope of this study.

The post-ESP techniques showed more forecast skill improvement than either of the pre-ESP methods used here. Both methods used to generate pre-ESP forecast showed only very minimal forecast skill improvement (0%–5%) for SLCA3, which showed forecast skill improvements of nearly 30% for post-ESP. The pre-ESP methods used here required a modification of the existing NWSRFS preadjustment method based on CPC forecasts since a long-term historical archive of such forecasts is not available. Therefore the preadjustment was based on the Niño-3.4 index instead of CPC forecasts.

It should be noted that preadjustment techniques are more computationally cumbersome than post-ESP techniques. Where only one set of ESP reforecasts is needed to optimize the post-ESP parameters, many sets of ESP reforecasts are necessary if a similar optimization of pre-ESP parameters (such as the number of nearest neighbors chosen in the resampling method) is to be done. This point alone argues strongly for post-ESP techniques in operational settings.

The results shown here demonstrate the potential value of using a post-ESP technique for forecasts of seasonal runoff, especially in regions with strong ENSO correlations. As these results are based on ensemble output from a physically based model (NWSRFS ESP), the approach can be extended to variables beyond seasonal runoff volumes that may be difficult to predict with statistical models alone. The magnitude and timing of the seasonal peak flow are examples of quantities that an ensemble-based physical model can easily simulate. Further research is necessary to explore the potential of the techniques presented here in the context of other variables, for example, the timing of the spring runoff peaks, which are difficult to capture using dynamical models. Further research will also apply these concepts to forecasts for other lead times and for other geographic areas.

REFERENCES

Barnston, A. G., and R. E. Livezey, 1987: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Mon. Wea. Rev.,* **115,** 1083–1126.

Cayan, D. R., and R. H. Webb, 1992: El Niño/Southern Oscillation and streamflow in the western United States, in El Niño. *Historical and Paleoclimatic Aspects of the Southern Oscillation,* H. Diaz and V. Markgraf, Eds., Cambridge University Press, 29–68.

Clark, M. P., and L. E. Hay, 2004: Use of medium-range weather forecasts to produce predictions of streamflow. *J. Hydrometeor.,* **5,** 15–32.

——, M. C. Serreze, and G. J. McCabe, 2001: Historical effects of El Niño and La Nina events on the seasonal evolution of the montane snowpack in the Columbia and Colorado River basins. *Water Resour. Res.,* **37,** 741–757.

Day, G. N., 1985: Extended streamflow forecasting using NWSRFS. *J. Water. Res. Plann. Manage.,* **111,** 157–170.

Epstein, E. S., 1969: A scoring system for probability forecasts of ranked categories. *J. Appl. Meteor.,* **8,** 985–987.

Hamlet, A. F., and D. P. Lettenmaier, 1999: Columbia River streamflow forecasting based on ENSO and PDO climate signals. *J. Water. Res. Plann. Manage.,* **125,** 333–341.

Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting,* **15,** 559–570.

Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis, 1997: A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull. Amer. Meteor. Soc.,* **78,** 1069–1079.

McCabe, G. J., and M. D. Dettinger, 2002: Primary modes and predictability of year-to-year snowpack variations in the western United States from teleconnections with Pacific Ocean climate. *J. Hydrometeor.,* **3,** 13–25.

Murphy, A. H., 1969: On the "ranked probability score." *J. Appl. Meteor.,* **8,** 988–989.

——, 1971: A note on the ranked probability score. *J. Appl. Meteor.,* **10,** 155–156.

Perica, S., 1998: Integration of meteorological forecasts/climate outlooks into extended streamflow prediction (ESP) systems. Preprints, *14th Conf. on Probability and Statistics in the Atmospheric Sciences,* Phoenix, AZ, Amer. Meteor. Soc., 130–133.

Ropelewski, C. F., and M. S. Halpert, 1987: Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. *Mon. Wea. Rev.,* **115,** 1606–1626.

Shao, J., D. Tu, and I. Olkin, 1996: *The Jackknife and Bootstrap.* Springer-Verlag, 516 pp.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 467 pp.

Wolter, K., and M. S. Timlin, 1993: Monitoring ENSO in COADS with a seasonally adjusted principal component index. *Proc. 17th Climate Diagnostics Workshop,* Norman, OK, NOAA/NMC/CAC, 52–57.

——, and ——, 1998: Measuring the strength of ENSO—How does 1997/98 rank? *Weather,* **53,** 315–324.