A resampling procedure for generating conditioned daily weather sequences

Martyn P. Clark,[1] Subhrendu Gangopadhyay,[1,2] David Brandon,[3] Kevin Werner,[3] Lauren Hay,[4] Balaji Rajagopalan,[5,6] and David Yates[7]

Received 7 October 2003; revised 10 February 2004; accepted 27 February 2004; published 28 April 2004.

[1] A method is introduced to generate conditioned daily precipitation and temperature time series at multiple stations. The method resamples data from the historical record "nens" times for the period of interest (nens = number of ensemble members) and reorders the ensemble members to reconstruct the observed spatial (intersite) and temporal correlation statistics. The weather generator model is applied to 2307 stations in the contiguous United States and is shown to reproduce the observed spatial correlation between neighboring stations, the observed correlation between variables (e.g., between precipitation and temperature), and the observed temporal correlation between subsequent days in the generated weather sequence. The weather generator model is extended to produce sequences of weather that are conditioned on climate indices (in this case the Niño 3.4 index). Example illustrations of conditioned weather sequences are provided for a station in Arizona (Petrified Forest, 34.8°N, 109.9°W), where El Niño and La Niña conditions have a strong effect on winter precipitation. The conditioned weather sequences generated using the methods described in this paper are appropriate for use as input to hydrologic models to produce multiseason forecasts of streamflow. *INDEX TERMS:* 1833 Hydrology: Hydroclimatology; 1869 Hydrology: Stochastic processes; 1894 Hydrology: Instruments and techniques; *KEYWORDS:* hydroclimatology, prediction, stochastic hydrology

Citation: Clark, M. P., S. Gangopadhyay, D. Brandon, K. Werner, L. Hay, B. Rajagopalan, and D. Yates (2004), A resampling procedure for generating conditioned daily weather sequences, *Water Resour. Res.*, *40*, W04304, doi:10.1029/2003WR002747.

1. Introduction

[2] Accurate streamflow forecasts are vital for water managers to meet the competing demands for increasingly scarce fresh water resources. The U.S. National Weather Service (NWS) uses the extended streamflow prediction (ESP) procedure for streamflow forecasting [*Day*, 1985]. In the traditional implementation of this approach, a hydrologic model is driven with inputs of observed precipitation and temperature data up to the beginning of the forecast (e.g., 1 January) and is then run using inputs of precipitation and temperature for the same dates over the forecast lead time from all past years in the historical record. This provides an ensemble of possible outcomes given the modeled hydrologic conditions (e.g., soil moisture, water equivalent of the accumulated snowpack) at the start of the forecast. Forecast accuracy is entirely dependent on accurate specification of conditions over the basin at the start of the forecast and the influence of those conditions on the basin hydrologic response. The approach works well in river systems where significant lag times are introduced due to storage of water in snowpack or subsurface and groundwater reservoirs.

[3] Much effort has been devoted toward modifying the historical sequences of precipitation and temperature used in ESP in order to include information from meteorological forecasts and climate outlooks. As a first step in this direction, *Hamlet and Lettenmaier* [1999] modified the ESP approach by restricting attention to years (ensemble members) that were similar in terms of the phase of the El Niño-Southern Oscillation (ENSO) 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. However, when attention is restricted to a small subset of years, or when some years are weighted more heavily than others, the resultant probabilistic streamflow forecasts may be overwhelmed by unusual conditions in any of the selected years.

[4] The NWS Advanced Hydrologic Prediction Service [see, e.g., *Connelly et al.*, 1999] seeks to improve opera-

[1]Center for Science and Technology Policy Research, Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA.
[2]Also at Department of Civil, Environmental, and Architectural Engineering, University of Colorado, Boulder, Colorado, USA.
[3]Colorado Basin River Forecast Center, Salt Lake City, Utah, USA.
[4]Water Resources Division, United States Geological Survey, Lakewood, Colorado, USA.
[5]Department of Civil, Environmental, and Architectural Engineering, University of Colorado, Boulder, Colorado, USA.
[6]Also at Center for Science and Technology Policy Research, Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA.
[7]Research Applications Program, National Center for Atmospheric Research, Boulder, Colorado, USA.

0043-1397/04/2003WR002747$09.00

tional streamflow forecasts in the United States. NWS efforts to date have developed methods to use the official Climate Prediction Center (CPC) climate outlooks to preadjust historical precipitation and temperature time series that are used as input to the ESP system (J. Schaake, NWS Office of Hydrologic Development, personal communication, 2002) and methods to postprocess the climatological ESP traces (L. Rundquist, NWS Alaska-Pacific River Forecast Center, personal communication, 2003). The CPC forecasts are very conservative, and the departure from the climatological ESP forecasts are often very small.

[5] The family of stochastic methods used to generate synthetic sequences of weather [e.g., *Wilks and Wilby*, 1999, and references therein] provides an interesting alternative to the approaches above. Such methods, commonly known as ''weather generators,'' provide new weather sequences that compensate for inadequacies in the length of station records. Weather generator models typically contain separate treatments for precipitation occurrence and intensity [e.g., *Gabriel and Neumann*, 1962; *Todorovic and Woolhiser*, 1975; *Foufoula-Georgiou and Lettenmaier*, 1987; *Hay et al.*, 1991; *Wilks*, 1998]. For example, *Wilks* [1998] used a Markov chain process to generate the sequence of wet and dry days. Conditional probabilities are calculated for the cases (precipitation on day t, given no precipitation on day $t - 1$) and (precipitation on day t, given precipitation on day $t - 1$). In the generated weather sequence, the transition from wet-to-dry and dry-to-wet days is determined if a random number pulled from a uniform distribution is less than or equal to the appropriate conditional probability. That is, if day t is dry, day $t + 1$ is modeled as a wet day if the uniform random number is less than or equal to the conditional probability for the case (precipitation on day t | no precipitation on day $t - 1$). *Wilks* [1998] generated precipitation intensities on wet days by randomly selecting a value from a mixed distribution that is fitted for nonzero precipitation amounts.

[6] Another set of methods generates weather by resampling data from the historical record [e.g., *Young*, 1994; *Rajagopalan and Lall*, 1999; *Buishand and Brandsma*, 2001; *Yates et al.*, 2003]. These methods compare a vector of weather variables for day t against a vector of same variables from similar dates in the historical record. The k ($k = \sqrt{n}$) most similar days are taken as the k-nearest neighbors (n is the number of similar dates used in the comparison). One of these neighbors is randomly selected, and the day following the selected neighbor is taken as the next simulated day (day $t + 1$).

[7] Unfortunately, many existing weather generator methods have problems with under-prediction of precipitation when they are extended to multiple sites [e.g., *Jothityangkoon et al.*, 2000; *Buishand and Brandsma*, 2001; *Yates et al.*, 2003]. The objectives of this paper are twofold: (1) to introduce an approach for generating weather that preserves the mean, standard deviation, and skewness of the generated precipitation and temperature time series, while also preserving the temporal persistence, and intersite and intervariable correlations, and (2) to introduce methods for conditioning the weather generator on climate indices and probabilistic climate forecasts.

[8] The weather generator method presented in this paper is based on the resampling-type weather generators [*Young*, 1994; *Rajagopalan and Lall*, 1999; *Buishand and Brandsma*, 2001; *Yates et al.*, 2003] but uses the ensemble-member reordering method, which was recently introduced by *Clark et al.* [2004] for downscaling of numerical weather prediction model output, to reconstruct the observed space-time variability in generated weather sequences. The overall goal of this paper is to develop a robust method for generating weather sequences, representing estimates of future climate conditions, that can be used as input to hydrologic models to forecast streamflow.

2. Data

[9] This study uses daily precipitation and maximum and minimum temperature data from the National Weather Service (NWS) cooperative network of climate observing stations across the contiguous United States (Figure 1). These data were extracted from the National Climatic Data Center (NCDC) ''Summary of the Day'' (TD3200) data set by Jon Eischeid, NOAA Climate Diagnostics Center, Boulder, Colorado [*Eischeid et al.*, 2000]. Quality control performed by NCDC includes the procedures described by *Reek et al.* [1992] that flag questionable data based on checks for (1) extreme values, (2) internal consistency among variables (e.g., maximum temperature less than minimum temperature), (3) constant temperature (e.g., 5 or more days with the same temperature are suspect), (4) excessive diurnal temperature range, (5) invalid relations between precipitation, snowfall, and snow depth, and (6) unusual spikes in temperature time series. Records at most of these stations start in 1948 and continue through to the present. Attention is restricted to stations with less than 10% missing or questionable data during the period 1950–1999 (2307 stations; Figure 1).

3. The Weather Generator Method

[10] The method for generating weather sequences can be described in two main steps:

[11] 1. Data sequences from individual stations are resampled from the historical record ''nens'' times, where nens is the number of ensemble members. For a given simulated day (e.g., 14 January 2004), data are selected from days in the time series for which the Julian day is within $\pm x$ days of the simulated day (if $x = 7$, data are resampled from the period 7–21 January). Data may be selected from all years (for the generation of climatological weather sequences), or data may be preferentially selected from a subset of years (for the generation of conditioned weather sequences). Resampling of data is not dependent on previous simulated days (as in other weather generator methods), so it is not constrained at this stage to preserve the temporal persistence in station time series. Such bootstrap methods have been proven in many instances to preserve the moments (mean, standard

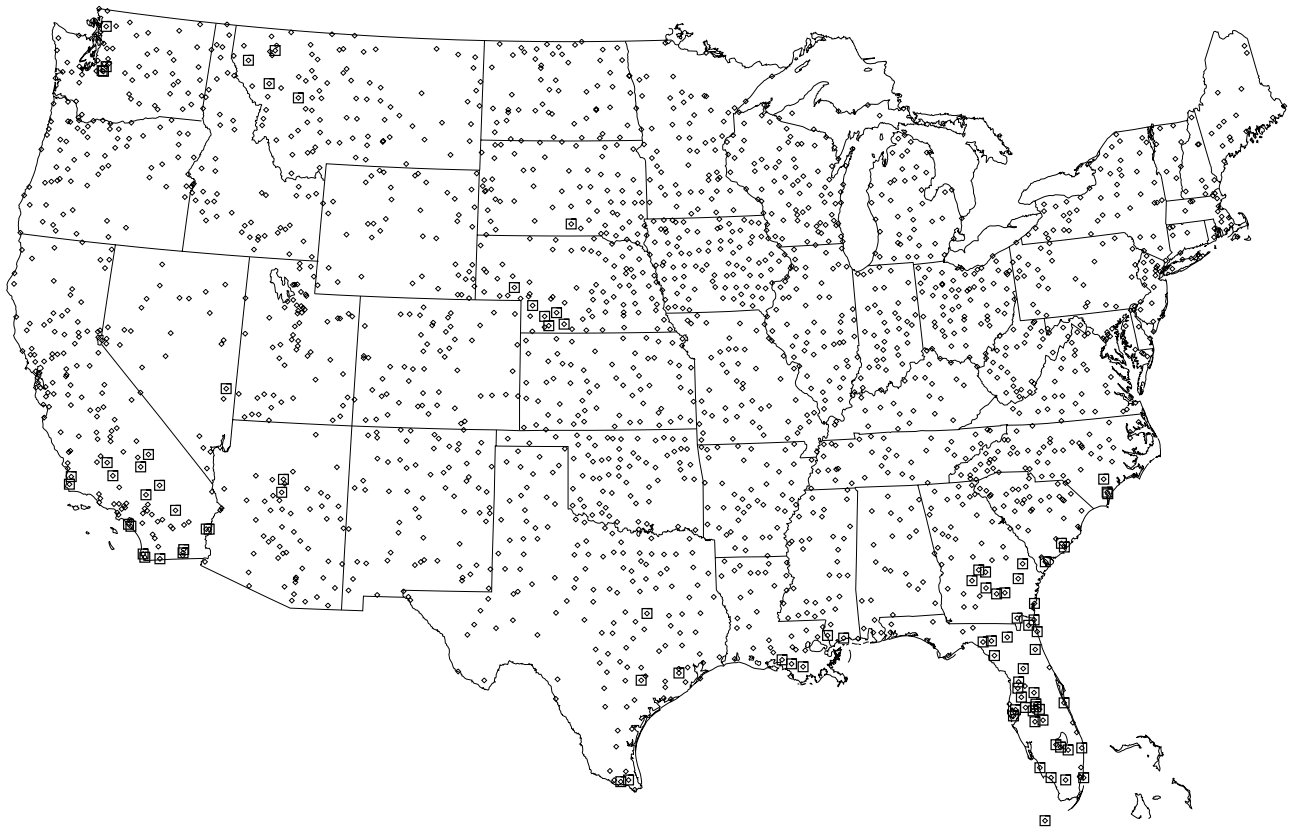


Figure 1. Location of stations used in this study. The squares depict stations used in the optimization study (see section 4.2.2 for more details).

deviation, and skewness) of station time series [e.g., *Efron*, 1979].

[12] 2. For a given generated day, the ensemble members are re-ordered so as to preserve the space-time variability in historical time series. The major difference between this method and other weather generator models is that the space-time variability is not preserved intrinsically, but is reconstructed as a post-processing step.

[13] The ensemble reordering methodology (''the Schaake shuffle'') was introduced recently by *Clark et al.* [2004] and requires further explanation. For a given simulated day, the starting point is a three-dimensional matrix of preferentially selected historical station observations $\boldsymbol{X}_{i,j,k}$, where i refers to each ensemble member, j refers to each station, and k refers to each variable. To correspond to the matrix $\mathbf{X}$, we construct an identically sized three-dimensional matrix $\mathbf{Y}_{i,j,k}$; also derived from historical station observations of the respective variables, where i refers to an index of dates in the historical time series, and, as in $\mathbf{X}$, j refers to each station and k refers to each variable.

[14] The $\mathbf{Y}$ matrix is used as a base to reconstruct the spatiotemporal variability of the preferentially selected historical station observations in $\mathbf{X}$. The differences between the $\mathbf{X}$ and $\mathbf{Y}$ matrices are (1) the observations used to populate the $\mathbf{Y}$ matrix are selected from all years in the historical record, whereas the observations in the $\mathbf{X}$ matrix may be preferentially selected from a subset of years (e.g., those reflecting future climate conditions); (2) for a given ensemble member i, the observations used to populate the $\mathbf{Y}$ matrix for the stations j and variables k are taken from the same dates in the historical record, whereas the observations used to populate the j and k dimensions of the $\mathbf{X}$ matrix can come from a mix of different dates; and (3) the dates used to populate the $\mathbf{Y}$ matrix are persisted for subsequent simulated days.

[15] For a given station j and variable k, the ensemble reordering method can be formulated as follows. Let $\vec{X}$ be a vector of n preferentially selected observations x and let $\vec{\chi}$ be the sorted vector of $\vec{X}$, that is,

$$\vec{X} = (x_1, x_2, \ldots, x_n), \tag{1}$$

$$\vec{\chi} = \left(x_{(1)}, x_{(2)}, \ldots, x_{(n)}\right), \qquad x_{(1)} \le x_{(2)} \ldots \le x_{(n)}. \tag{2}$$

Also, let $\vec{Y}$ be a vector of n historical observations that are selected from all years y, and let $\vec{\gamma}$ be the sorted vector of $\vec{Y}$, that is,

$$\vec{Y} = (y_1, y_2, \ldots, y_n), \tag{3}$$

$$\vec{\gamma} = \left(y_{(1)}, y_{(2)}, \ldots, y_{(n)}\right), \qquad y_{(1)} \le y_{(2)} \ldots \le y_{(n)}. \tag{4}$$

Furthermore, let $\vec{B}$ be the vector of indices describing the original observation number that corresponds to the values in the ordered vector $\vec{\gamma}$.

[16] As an example, preferential resampling maximum temperature for 10 ensemble members at a given station on

A. Ranked Ensemble of generated weather for 14th January 2004

Ens #	Stn 1	Ens #	Stn 2	Ens #	Stn 3
(5)	**7.5**	(2)	**6.3**	(9)	**12.4**
(7)	**8.3**	(9)	**7.2**	(3)	**13.5**
(3)	**8.8**	(4)	**7.5**	(4)	**14.2**
(6)	**9.7**	(3)	**7.9**	(7)	**14.5**
(10)	**10.1**	(7)	**8.6**	(2)	**15.6**
(9)	**10.3**	(1)	**9.3**	(6)	**15.9**
(2)	**11.2**	(6)	**11.8**	(10)	**16.3**
(4)	**11.9**	(10)	**12.2**	(1)	**17.6**
(8)	**12.5**	(5)	**13.5**	(5)	**18.3**
(1)	**15.3**	(8)	**17.7**	(8)	**23.9**

B. Randomly selected Observed Ensembles

Ens #	Date	Stn 1	Stn 2	Stn 3
1	8th Jan 1996	10.7	10.9	13.5
2	17th Jan 1982	9.3	9.1	13.7
3	13th Jan 2000	6.8	7.2	9.3
4	22nd Jan 1998	11.3	10.7	15.6
5	12th Jan 1968	12.2	13.1	17.8
6	9th Jan 1976	13.6	14.2	19.3
7	10th Jan 1998	8.9	9.4	12.1
8	19th Jan 1980	9.9	9.2	11.8
9	16th Jan 1973	11.8	11.9	15.2
10	9th Jan 1999	12.9	12.5	16.9

C. The Ranked Observed Ensembles

Ens #	Stn 1	Ens #	Stn 2	Ens #	Stn 3
(3)	**6.8**	(3)	**7.2**	(3)	**9.3**
(7)	**8.9**	(2)	**9.1**	(8)	**11.8**
(2)	**9.3**	(8)	**9.2**	(7)	**12.1**
(8)	**9.9**	(7)	**9.4**	(1)	**13.5**
(1)	**10.7**	(4)	**10.7**	(2)	**13.7**
(4)	**11.3**	(1)	**10.9**	(9)	**15.2**
(9)	**11.8**	(9)	**11.9**	(4)	**15.6**
(5)	**12.2**	(10)	**12.5**	(10)	**16.9**
(10)	**12.9**	(5)	**13.1**	(5)	**17.8**
(6)	**13.6**	(6)	**14.2**	(6)	**19.3**

D. Final Ensembles (shuffled output)

	E #	S 1	E #	S 2	E #	S 3
1	(10)	**10.1**	(1)	**9.3**	(7)	**14.5**
2	(3)	**8.8**	(9)	**7.2**	(2)	**15.6**
3	(5)	**7.5**	(2)	**6.3**	(9)	**12.4**
4	(9)	**10.3**	(7)	**8.6**	(10)	**16.3**
5	(4)	**11.9**	(5)	**13.5**	(5)	**18.3**
6	(1)	**15.3**	(8)	**17.7**	(8)	**23.9**
7	(7)	**8.3**	(3)	**7.9**	(4)	**14.2**
8	(6)	**9.7**	(4)	**7.5**	(3)	**13.5**
9	(2)	**11.2**	(6)	**11.8**	(6)	**15.9**
10	(8)	**12.5**	(10)	**12.2**	(1)	**17.6**

Figure 2. The ensemble reordering method for a hypothetical ensemble of 10 members, and for a given variable (e.g., maximum temperature) for 14 January 2004, showing (a) the ranked ensemble of generated weather for three stations (χ), (b) a random selection of historical observations for the three stations (Y), (c) the ranked historical observations (γ), and (d) the final reordered output (X^{SS}). See text for further details.

a given date ($\vec{X}$); and the corresponding selection of historical observations from all years ($\vec{Y}$), may provide vectors of $\vec{X}$, $\vec{\chi}$, $\vec{Y}$, $\vec{\gamma}$, and $\vec{B}$ that are

$$\vec{X} = (15.3, 11.2, 8.8, 11.9, 7.5, 9.7, 8.3, 12.5, 10.3, 10.1);$$

$$\vec{\chi} = (7.5, 8.3, 8.8, 9.7, 10.1, 10.3, 11.2, 11.9, 12.5, 15.3);$$

$$\vec{Y} = (10.7, 9.3, 6.8, 11.3, 12.2, 13.6, 8.9, 9.9, 11.8, 12.9);$$

$$\vec{\gamma} = (6.8, 8.9, 9.3, 9.9, 10.7, 11.3, 11.8, 12.2, 12.9, 13.6);$$

$$\vec{B} = (3, 7, 2, 8, 1, 4, 9, 5, 10, 6).$$

In this example, data from the first date selected from the historical record (10.7 in vector $\vec{Y}$) is ranked fifth lowest of the 10 ensemble members, as shown in the vectors $\vec{\gamma}$ and $\vec{B}$. Data from the second date are ranked third lowest (9.3 in vector $\vec{Y}$); and data from the third date selected from the historical record are ranked lowest of all 10 ensemble members (6.8 in vector $\vec{Y}$).

[17] Now the final step is to construct the reordered vector $\vec{X}^{SS}$, which is the final reordered output:

$$\vec{X}^{SS} = \left(x_1^{ss}, x_2^{ss}, \ldots, x_n^{ss}\right), \tag{5}$$

where

$$x_q^{ss} = x_{(r)}, \tag{6}$$

$$q = \vec{B}[r], \tag{7}$$

$$r = 1, \ldots, n. \tag{8}$$

Recall that the subscripts in parentheses refer to the elements in the sorted vector $\vec{\chi}$. Following through with the numbers from the example above provides

$$x_3^{ss} = x_{(1)} = 7.5,$$

$$x_7^{ss} = x_{(2)} = 8.3,$$

$$x_2^{ss} = x_{(3)} = 8.8,$$

and so on. Hence, in this example,

$$\vec{X}^{SS} = (10.1, 8.8, 7.5, 10.3, 11.9, 15.3, 8.3, 9.7, 11.2, 12.5).$$

[18] The ensemble reordering approach is demonstrated graphically in Figure 2 through an example. Figure 2

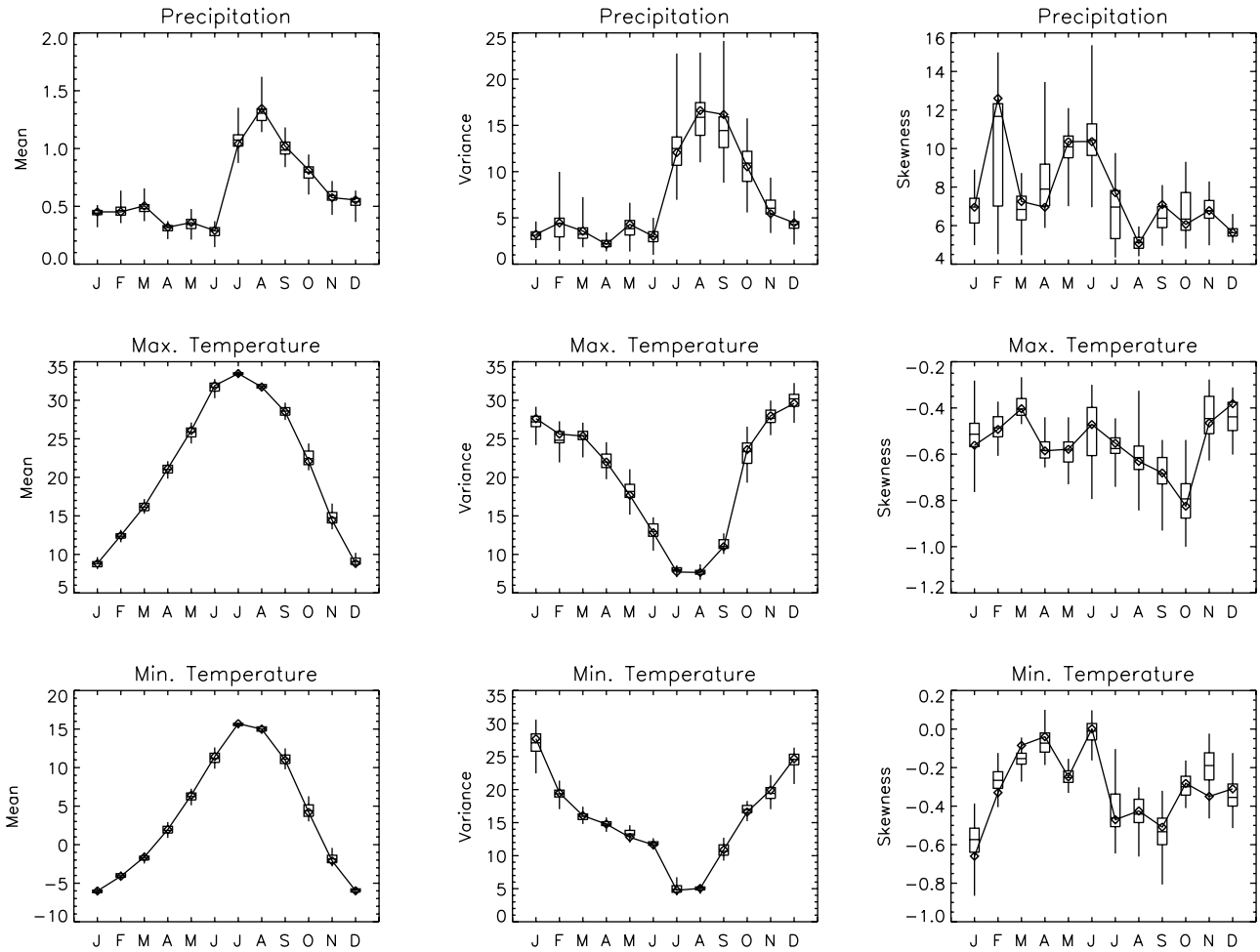


Figure 3. Comparison of the seasonal cycles of generated and observed precipitation (top row), maximum temperature (middle row), and minimum temperature (bottom row), for an example station in Arizona (Petrified Forest; 34.8°N, 109.9°W). The thin line with diamonds illustrates the observations, and the box-and-whiskers illustrate the minimum, lower quartile, median, upper quartile, and maximum of the generated weather sequences. Results are based on weather generated for the period 1950–1999.

outlines example results for three stations ($j = 3$) and one variable (e.g., maximum temperature) that are extracted from the matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ that were described at the beginning of this section. Figure 2a describes the ranked preferentially selected observations for the three example stations (the vectors $\vec{\chi}$ defined earlier). Figure 2b describes the observations selected from all years in the historical record (the vectors $\vec{Y}$ defined earlier), and Figure 2c shows the ranked historical observations (the vectors $\vec{\gamma}$ defined earlier). Also in Figure 2c is the vector $\vec{B}$, which is the index of the original ensemble member that corresponds to the values in the ordered vector $\vec{\gamma}$. Figure 2d describes the final shuffled output (the vectors $\vec{X}^{SS}$).

[19] The dark ellipses in Figure 2 correspond to the first ensemble member extracted from the historical record. When this is not sorted (i.e., the vector $\vec{Y}$), the values are 10.7, 10.9, and 13.5, for stations one, two, and three, respectively (Figure 2b). When these values are sorted with respect to all other ensemble members, the first observed ensemble is ranked fifth for station one, sixth for station two, and fourth for station three (Figure 2c). For the first ensemble member, the ranks 5, 6, and 4 are actually the values of the index (r) for the three stations (equation (6)); values for the first ensemble member, once resorted ($x_q^{ss} = x_{(r)}$), are 10.1 at station one ($x_1^{ss} = x_{(5)}$; see Figure 2a), 9.3 at station two ($x_1^{ss} = x_{(6)}$), and 14.5 at station three ($x_1^{ss} = x_{(4)}$). Also consider the second ensemble in Figure 2 (the shaded ellipses). Ensemble 2 from the historical record is ranked third, second, and fifth for stations one, two, and three, respectively (Figure 2c), such that the final shuffled output for ensemble 2 is 8.8 (station one), 7.2 (station two), and 15.6 (station three).

[20] This approach works because it preserves the Spearman Rank correlation structure between station pairs, and between climate variables. Consider first the correlation between station pairs. If observed data at two neighboring stations are similar (i.e., a high correlation between stations), then the observations at the two stations on a given (randomly selected) day are likely to have a similar rank. The rank of each preferentially selected observation (i.e., the resampled weather, in the vector $\vec{\chi}$) at the two stations is matched with the rank of each observation in the vector $\vec{\gamma}$, meaning that for all ensemble members, the rank of the resampled realizations will be similar at the two stations. When this process is repeated for all simulated days, the ranks of a given ensemble member will on average be

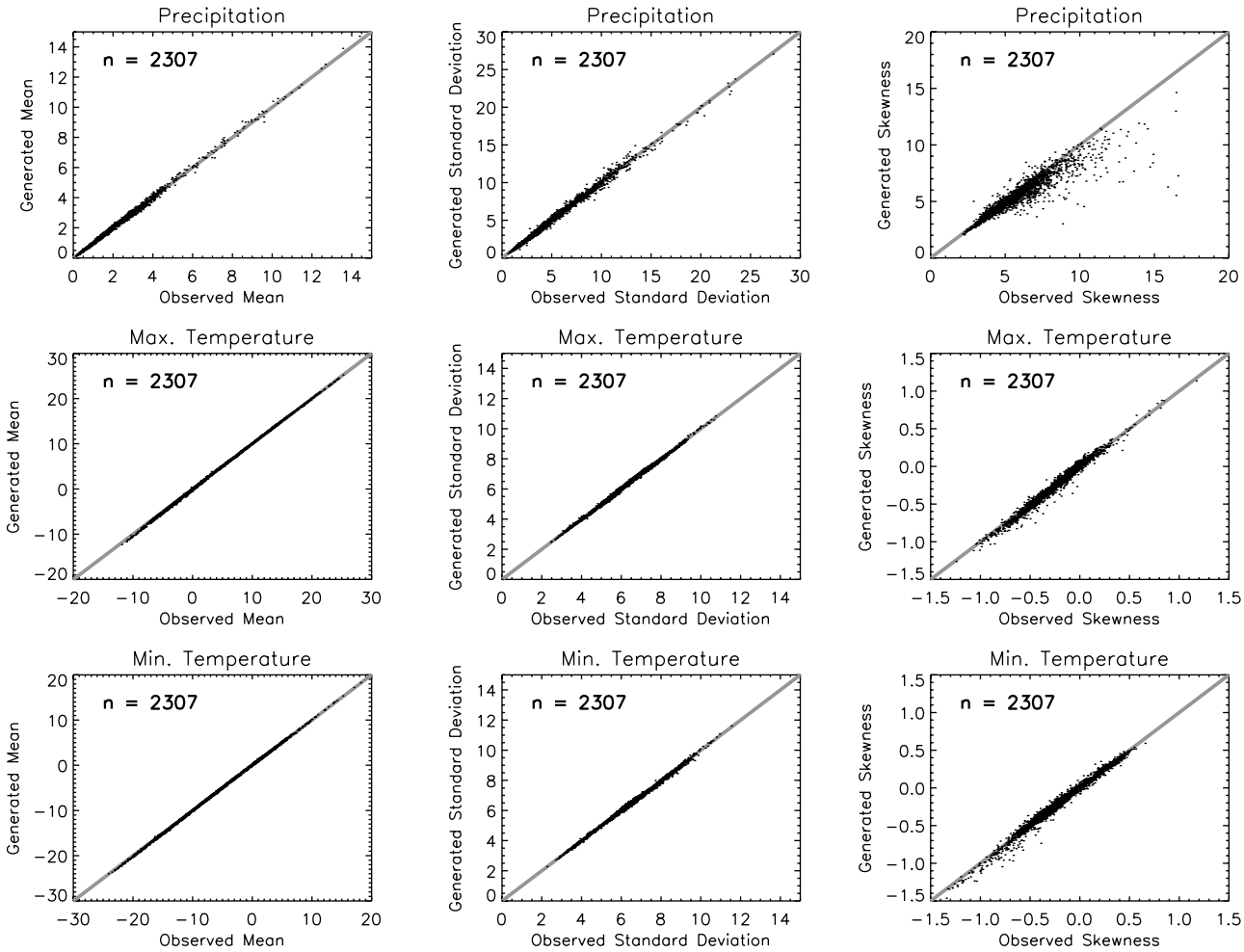


Figure 4. Comparison of generated and observed summary statistics for all stations in the contiguous United States (shown in Figure 1), for the month of January. The plots show comparisons of the station mean (left column), standard deviation (middle column), and skewness (right column), for precipitation (top row) maximum temperature (middle row), and minimum temperature (bottom row).

similar for the two stations, and the spatial correlation will be reconstructed once the vector $\vec{\chi}$ is resorted. This reasoning is identical for intervariable correlations.

[21] An ordered selection of dates from the historical record enables preservation of temporal persistence. The random selection of dates that are used to populate the matrix $\boldsymbol{Y}_{i,j,k}$ are only used for the first day in the generated sequence, and are persisted for subsequent forecast lead times. In the example presented in Figure 2, the dates for the next simulated day would be 9 January 1996 for the first ensemble member, 18 January 1982 for ensemble two, 14 January 2000 for ensemble three, and so on. High temporal persistence (e.g., as measured through lag-1 correlation statistics) means that the historical observations for subsequent days will, on average, have a similar rank. Because the (ranked) preferentially selected observations in the vector $\vec{\chi}$ are matched with the rank of successive observations in the vector $\vec{\gamma}$, the temporal persistence is reconstructed once the ensemble output is resorted.

[22] The main assumption in this approach is that the spatiotemporal correlation structure, as computed using all days in the historical record, is appropriate to reconstruct the spatiotemporal correlation structure for a subset of data. For unconditioned simulations this method generates sequences of weather that are similar (but not identical) to observed sequences. For conditioned simulations the method generates unique weather sequences (i.e., sequences of weather that have hitherto been unobserved).

4. Results

4.1. Summary Statistics of Generated Precipitation and Temperature Fields

[23] A necessary quality of any weather generator method is its capability to reproduce the summary statistics of the observed precipitation and temperature fields. Figure 3 shows box-and-whisker plots of the generated mean (left column), variance (middle column), and skewness (right column) for precipitation (top row), maximum temperature (middle row), and minimum temperature (bottom row), for all months of the year for an example station in central Arizona (Petrified Forest; 34.8°N, 109.9°W). The generated distribution is based on an ensemble of 50 weather sequences for the period 1950–1999. The observed mean, variance, and skewness are depicted in Figure 3 as the solid line with

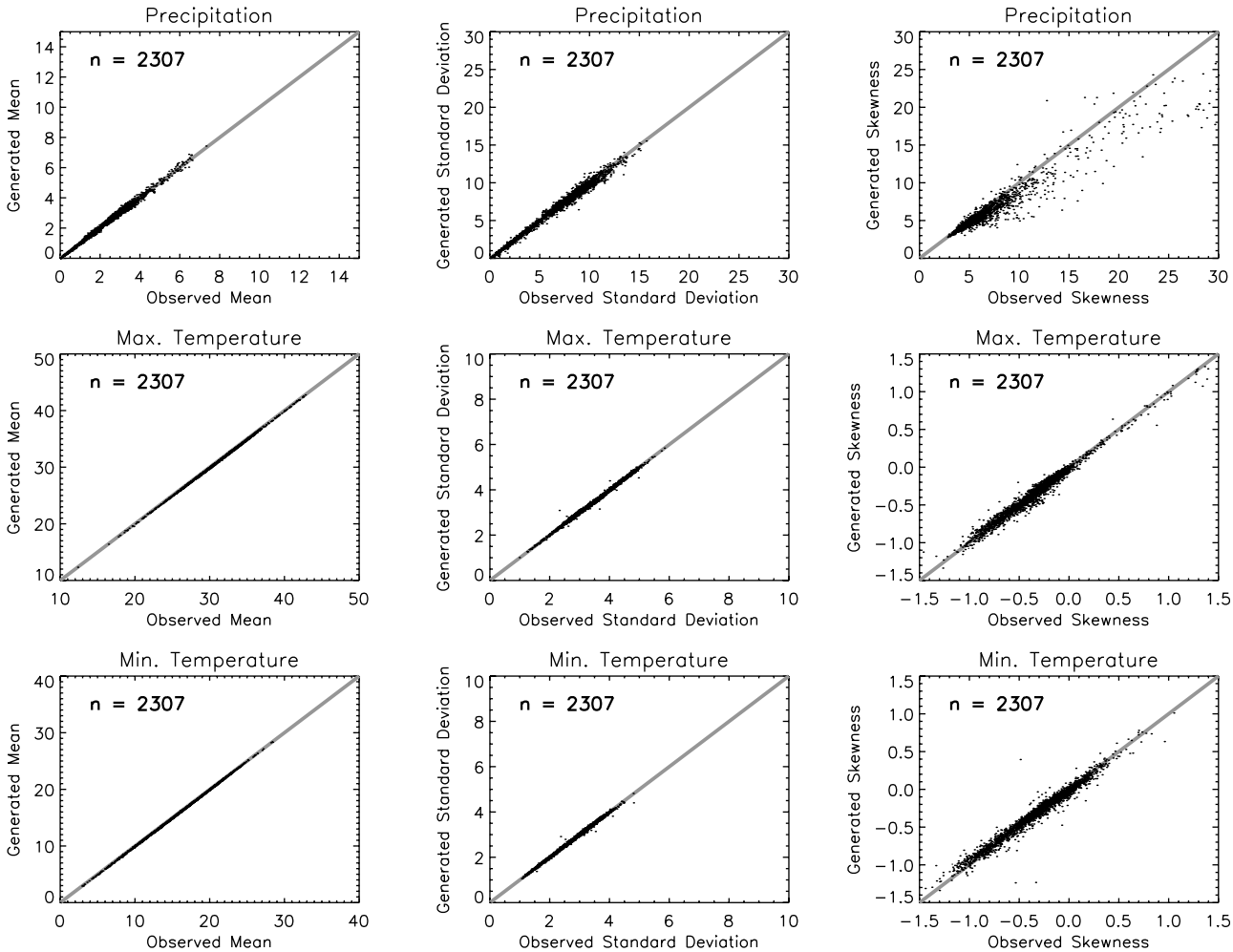


Figure 5. As in Figure 4, except for the month of July.

diamonds. The statistical moments at Petrified Forest are all produced well.

[24] Generating weather for a single station, however, is not a very robust test of the capabilities of the weather generator model. To extend this analysis, an ensemble of 50 daily weather sequences was generated for the period 1950–1999 for each station in Figure 1 (2307 stations). The mean, standard deviation, skewness, lag-1 correlation, and intervariable correlations were then computed for all generated ensemble members at each station. Spatial correlations of generated time series between station pairs were also computed for each ensemble member, but only for cases where the interstation separation is less than 500 km (spatial correlations were only computed for every fifth station, resulting in 47,105 valid station pairs). For distances greater than 500 km, the spatial correlation is very small. The median of the summary statistics of all ensemble members (mean, standard deviation, skewness, lag-1 correlation, and intervariable and intersite correlation) was computed for each station (or valid station pair) and compared against the observed summary statistic at that station or station pair.

[25] Figures 4 and 5 portray comparisons of the mean, standard deviation, and skewness from the generated and observed time series for the months of January and July, respectively, for all stations (Figure 1) in the contiguous United States. The top rows in Figures 4 and 5 present the comparison of the generated and observed precipitation time series. The mean (left column), standard deviation (middle column), and skewness (right column) are all adequately reproduced, although the skewness in the generated time series is lower than the corresponding statistics in the observed time series (Figures 4 and 5). The middle and bottom rows in Figures 4 and 5 present the comparison for the generated and observed maximum and minimum temperature time series. Similar to the results for Petrified Forest, the mean, standard deviation, and skewness of the temperature time series are all reproduced accurately.

[26] The results in Figures 3–5 are not altogether unexpected; unconditioned resampling of data from all years in the historical record should preserve the statistical moments of historical time series. A bigger challenge for the weather generator is an adequate depiction of the observed space-time variability in the generated precipitation and temperature fields. Figures 6 and 7 illustrate the observed and generated temporal (lag-1) Pearson correlations (top rows), and the spatial Pearson correlations between valid station pairs (bottom rows), for all stations in the contiguous United States. Results are shown for the months of January (Figure 6) and July (Figure 7). Correlations of generated precipitation are lower than the observed correlations (left columns of

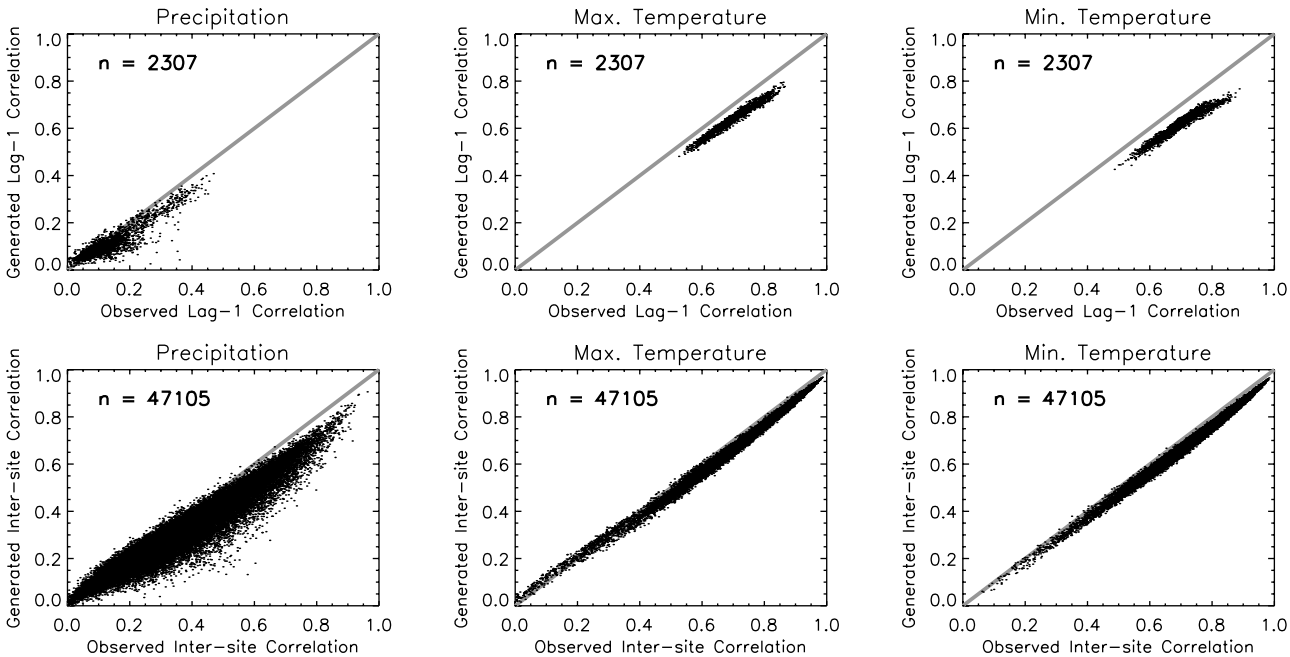


Figure 6. Comparison of generated and observed lag-1 correlations (top row) and intersite correlations (bottom row) for all stations in the contiguous United States (shown in Figure 1), for the month of January. The plots show comparisons for precipitation (left column), maximum temperature (middle column), and minimum temperature (right column).

Figures 6 and 7). This occurs because the ensemble reordering method has difficulties dealing with the intermittent properties of precipitation. When the ensemble of generated weather for a given day has less zero precipitation ensemble members than the ensemble from the persisted observed data, the generated ensemble members with zero precipitation days are matched with observed precipitation amounts, and their assignment to a given ensemble member is entirely random (see *Clark et al.* [2004] for more details).

[27] The generated maximum and minimum temperature time series reproduce the observed correlation structure fairly well (middle and right columns of Figures 6 and 7), although the generated lag-1 correlations are slightly lower than the observed correlations. This discrepancy could be due to bad data (e.g., when temperature sensors are "stuck" and the same temperature value is repeated for subsequent days) or to the lag-1 Pearson correlations being consistently lower than the Spearman Rank correlation (note that the ensemble-reordering method is only guaranteed to preserve the Spearman Rank correlation). Nevertheless, differences between observed and generated temperature correlations are very small.

[28] Figure 8 illustrates comparisons between generated and observed inter-variable correlations for all stations in the contiguous United States (shown in Figure 1), for the months of January (top row) and July (bottom row). The left and middle columns in Figure 8 portray correlations between precipitation and maximum temperature, and between precipitation and minimum temperature, respectively. The negative correlations indicate a tendency for lower temperatures on precipitation days (most pronounced for the precipitation and maximum temperature correlations in July), and this process is reproduced well by the weather generator model. The right column in Figure 8 portrays (high) correlations between maximum and minimum temperature, and these correlations are also reproduced well.

4.2. Conditioning on Climate Indices

4.2.1. Conditioning the Weather Generator on ENSO Indices

[29] Extending the weather generator to produce conditional weather sequences is fairly straightforward [e.g., *Yates et al.*, 2003]. Instead of resampling data from all years, one can rank the years (e.g., in terms of the similarity of a climate index), from most similar (highest rank) to least similar (lowest rank), and preferentially select certain years (*iyear*) based on weighting and selection criteria:

$$iyear = INT\left(u^{\lambda}\frac{N}{\alpha}\right) + 1 \tag{9}$$

In equation (9), $u \sim U(0, 1)$ is a random number selected from a uniform distribution ranging from zero to one, λ is a weighting parameter, α is a selection parameter, N is the number of years in the time series, and INT is the integer operator. Values of λ greater than 1 shift the uniform random number closer to zero, meaning years higher in the ranked list will be preferentially selected. Values of α greater than 1 will restrict the selection of years to a subset of years (e.g., if $\alpha = 5$, attention will be restricted to the top 20% of ranked years). The selection of years is unbiased if both λ and α are equal to 1. The probability of preferentially selecting different ranked years for selected α and λ values is illustrated graphically in Figure 9.

[30] This approach is implemented in this study as follows: (1) Years are ranked in terms of the similarity of the October value of the Niño 3.4 index, and (b) equation (9) is applied by selecting *iyear* 100 times, and using the selection of *iyears* to obtain the subset of days from which

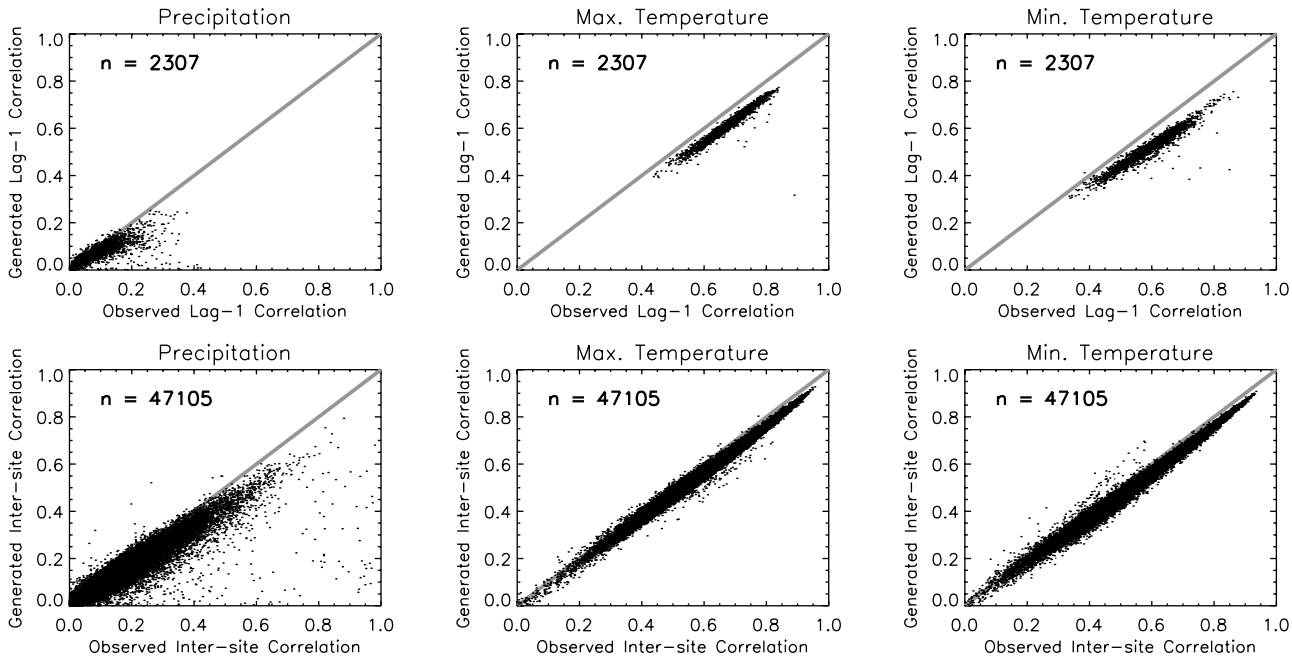


Figure 7. As in Figure 6, except for the month of July.

data are resampled. An additional random component is introduced by using a different random selection of years for each generated day. The Niño 3.4 index represents sea surface temperature (SST) anomalies in the central equatorial Pacific Ocean (5°N–5°S; 170°W–120°W); strong positive values of the Niño 3.4 index depict El Niño events, and strong negative values of the Niño 3.4 index depict La Niña events. The use of October Niño 3.4 conditions to rank the years means that the generated precipitation and temperature sequences are conditioned with respect to SST conditions at the beginning of winter; if the SST anomalies persist throughout winter, then they can affect the wintertime atmospheric circulation (and surface climate) over North America. The generated weather sequences thus can be considered forecasts initialized in October.

[31] Figure 10 illustrates 50 sequences of cumulative precipitation conditioned using equation (9) for water year

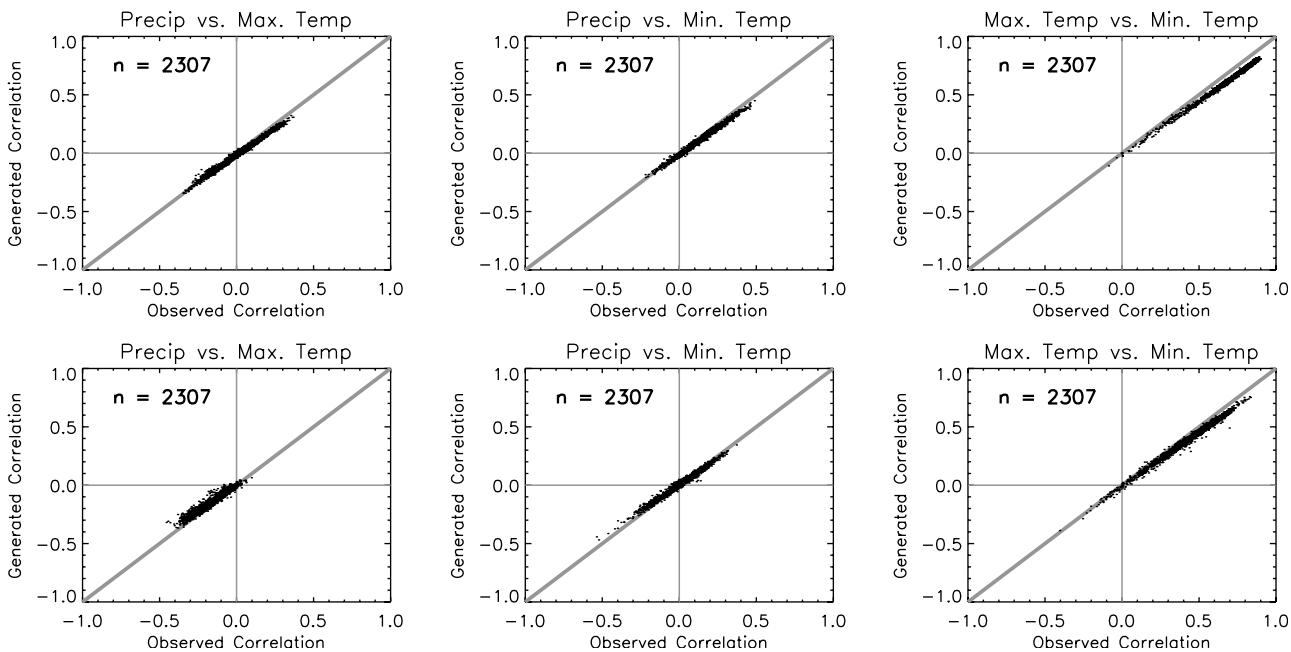


Figure 8. Comparison of generated and observed intervariable correlations, for all stations in the contiguous United States (shown in Figure 1). The plots show correlations between precipitation and maximum temperature (left column), between precipitation and minimum temperature (middle column), and between maximum temperature and minimum temperature (right column). The top plots depict results for January, and the bottom plots depict results for July.

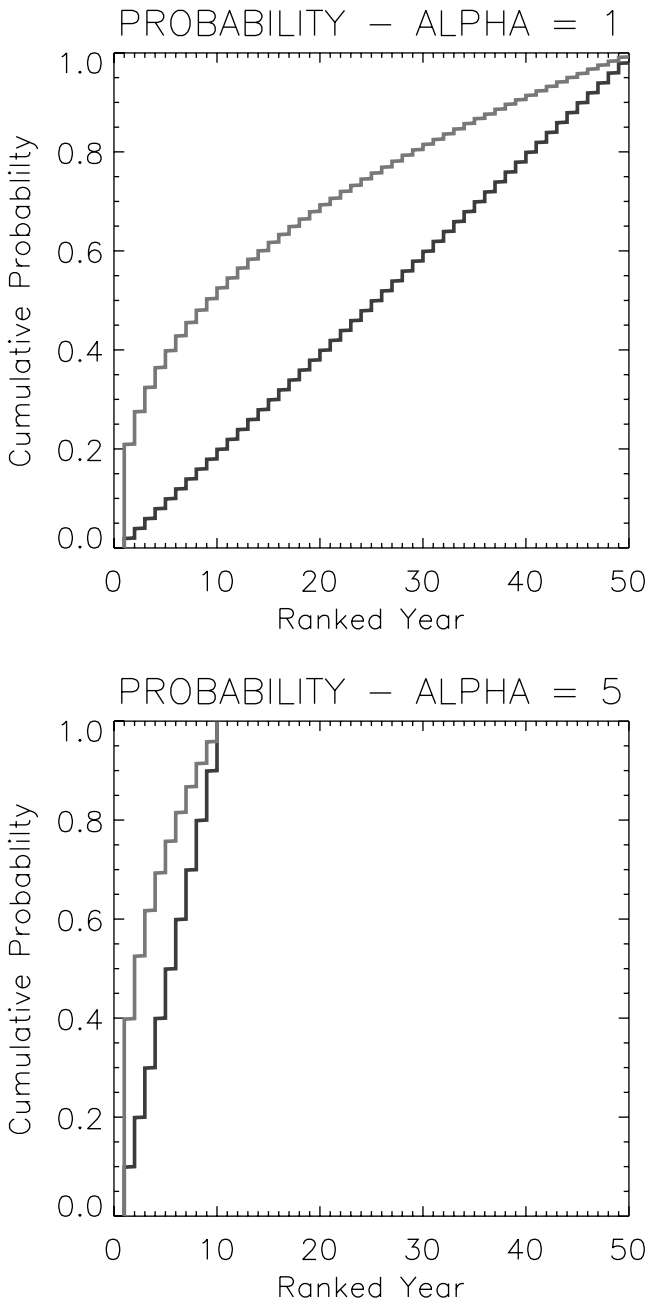


Figure 9. Cumulative probability of selecting years in a ranked list, based on different α and λ parameters. The top plot depicts probabilities for $\alpha = 1$, and the bottom plot illustrates results for $\alpha = 5$. In both plots the dark line (the light line) depicts probabilities for $\lambda = 1.0$ ($\lambda = 2.5$).

1988–1989, for our example station in Arizona (Petrified Forest; 34.8°N, 109.9°W), using different α and λ values. Dark lines in each panel depict the minimum (dashed), lower quartile, median, upper quartile, and maximum (dashed) cumulative precipitation, computed using all years, and the light lines illustrate the conditioned sequences of precipitation produced from the weather generator. The water year 1988–1989 was intentionally selected for this example because sea surface temperatures for the Niño 3.4 region were significantly lower than normal at this time (La Niña conditions), conditions which tend to favor lower than normal winter precipitation in Arizona [*Redmond and Koch*, 1991; *Cayan and Webb*, 1992; *Clark et al.*, 2001].

[32] Generated precipitation using α and λ values of 1.0 is shown in the top left panel of Figure 10. In this case the selection of years is unbiased, and, as expected, there is no tendency for the generated traces to favor low or high values. Generated precipitation using an α value of 1.0 and a λ value of 2.5 is shown in the upper right panel of Figure 10. These α and λ values mean that all years can be selected ($\alpha = 1.0$), but the years in which the October value of the Niño 3.4 index is most similar to the value for October 1988 will be preferentially selected. The generated traces in the upper right panel are indeed biased on the low side. The bottom two panels of Figure 10 illustrate generated traces for the α value of 5.0 where the selection of years is restricted to the top 20% of ranked years. The low bias in precipitation is more strongly apparent, especially for the λ value of 2.5 (bottom right panel of Figure 10) when more weight is given to years higher up in the ranked list.

[33] The generated ensembles in Figure 10 can be used to produce probabilistic forecasts of total winter precipitation. Figure 11 plots the cumulative probability, as computed from the generated traces in Figure 10, against the total winter precipitation (1 October through 31 March). The top panel in Figure 11 illustrates results for $\alpha = 1.0$, and the bottom panel in Figure 11 illustrates results for $\alpha = 5.0$ (the dark line (light line) in both plots illustrates results for $\lambda = 1.0$ ($\lambda = 2.5$)). The cumulative probability of total winter precipitation, the light line with diamonds, is shown for reference. As is expected from Figure 10, the cumulative probability distribution for α and λ values of 1.0 mirrors the cumulative probability distribution computed using all years; the cumulative probability distribution for $\alpha = 1.0$ and $\lambda = 2.5$ is biased slightly on the low side, and the cumulative probability distribution for $\alpha = 5.0$ is significantly drier than the climatological cumulative probability distribution. Figures 10 and 11 illustrate results for example values of the α and λ parameters; the next section will formalize the choice of these parameters.

4.2.2. Estimating the α and λ Parameters

[34] The values of the α and λ parameters can be specified a priori, but it may be possible to increase the probabilistic skill of the seasonal forecasts if the α and λ parameters are optimized. Figure 12 illustrates the ranked probability skill score (RPSS) for forecasts of total winter precipitation at Petrified Forest (the same station used in Figures 10 and 11), for different values of α and λ (more details on the RPSS are provided by *Wilks* [1995] and *Clark et al.* [2004]). In this experiment, the α and λ values were randomly selected within the range 1.0–10.0 (α) and 1.0–5.0 (λ), and conditioned sequences of weather were generated using these α and λ values. The top plot in Figure 12 illustrates the RPSS computed using all years, and the bottom plot illustrates the RPSS when computed for only ENSO years, defined as when the October Niño 3.4 anomalies were greater than 1.0°C or less than −1.0°C. Results show that situations where the closest analog year is preferentially selected (high values of both α and λ) actually produce negative skill at Petrified Forest (the squares in the top right corner of Figures 12a and 12b).

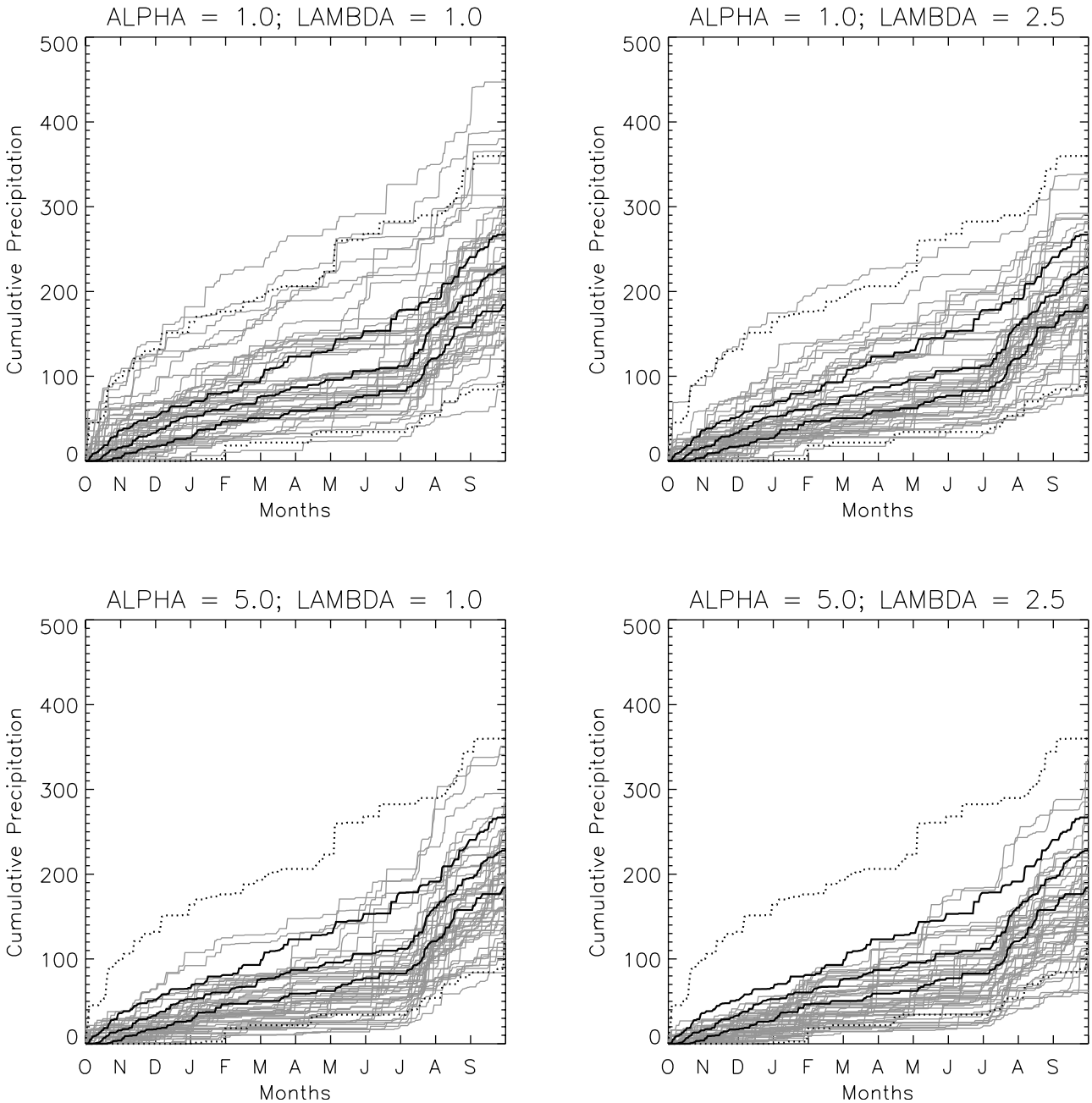


Figure 10. Fifty generated ensemble members of cumulative precipitation for water year 1988–1989, for the example station in Arizona used for Figure 3 (Petrified Forest; 34.8°N, 109.9°W), for different α and λ parameters.

Positive skill occurs in situations in which the weighting function is relaxed (Figure 12b). The shuffled complex evolution (SCE) method of *Duan et al.* [1992, 1993, 1994] was used to optimize the α and λ parameters by maximizing the RPSS (restricting attention to ENSO years). The SCE optimization provided parameter values of $\alpha = 5.550$ and $\lambda = 1.106$ (RPSS = 0.365).

[35] Attention is now directed to optimizing α and λ parameters for forecast applications. Figure 12 illustrates the dependence of probabilistic forecast skill on the strength of ENSO conditions; RPSS values are much higher when attention is restricted to ENSO years (El Niño plus La Niña). This is fairly intuitive. In years where there is weak ENSO forcing, assigning more weight to years that have similar values of the Niño 3.4 index is unlikely to result in an increase in probabilistic forecast skill. In other words, the α and λ parameters should depend on the strength of the Niño 3.4 index. The α and λ parameters should therefore be optimized locally, using years where the value of the Niño 3.4 index is similar to the year being forecast.

[36] The SCE optimization program [*Duan et al.*, 1992, 1993, 1994] was configured to maximize the RPSS value for the 10 years in which values of the Niño 3.4 index are most similar to the year being forecast. For a given water year, a function evaluation for a given α and λ parameter set

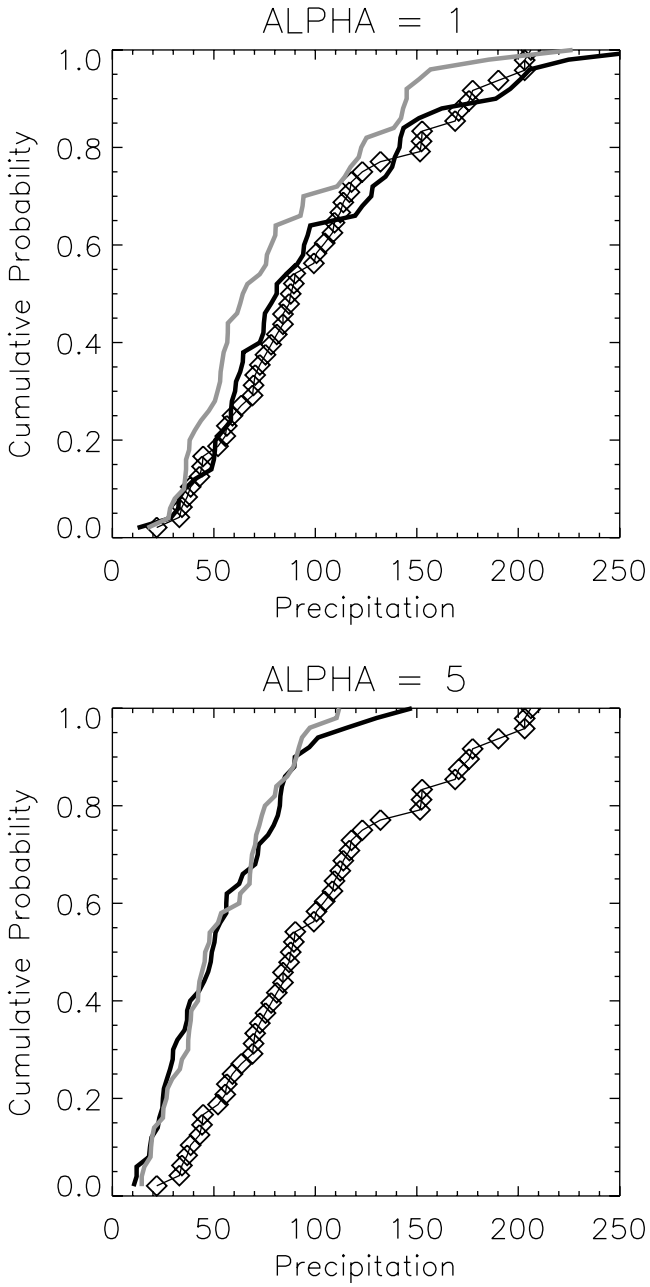


Figure 11. Probabilistic forecasts of total winter precipitation (1 October through 31 March) for water year 1988–1989, for Petrified Forest (34.8°N, 109.9°W), for different α and λ parameters. See text for further details.

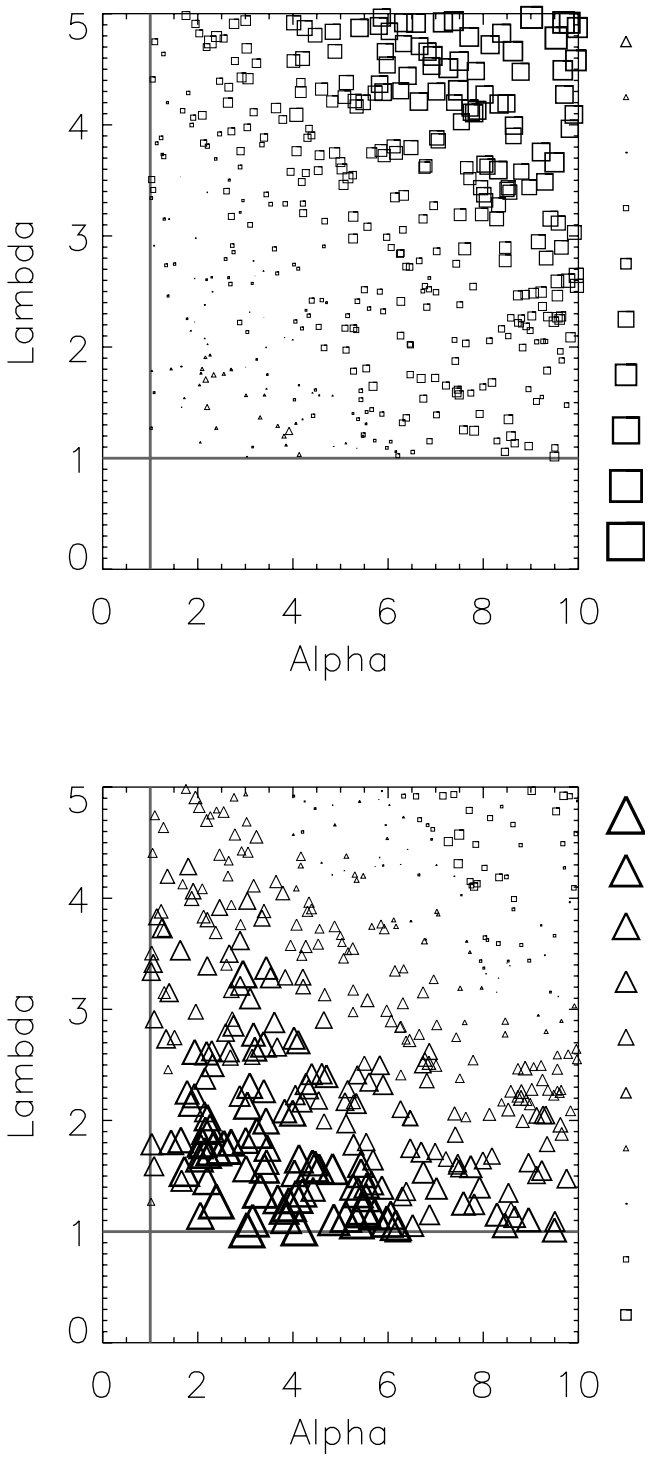
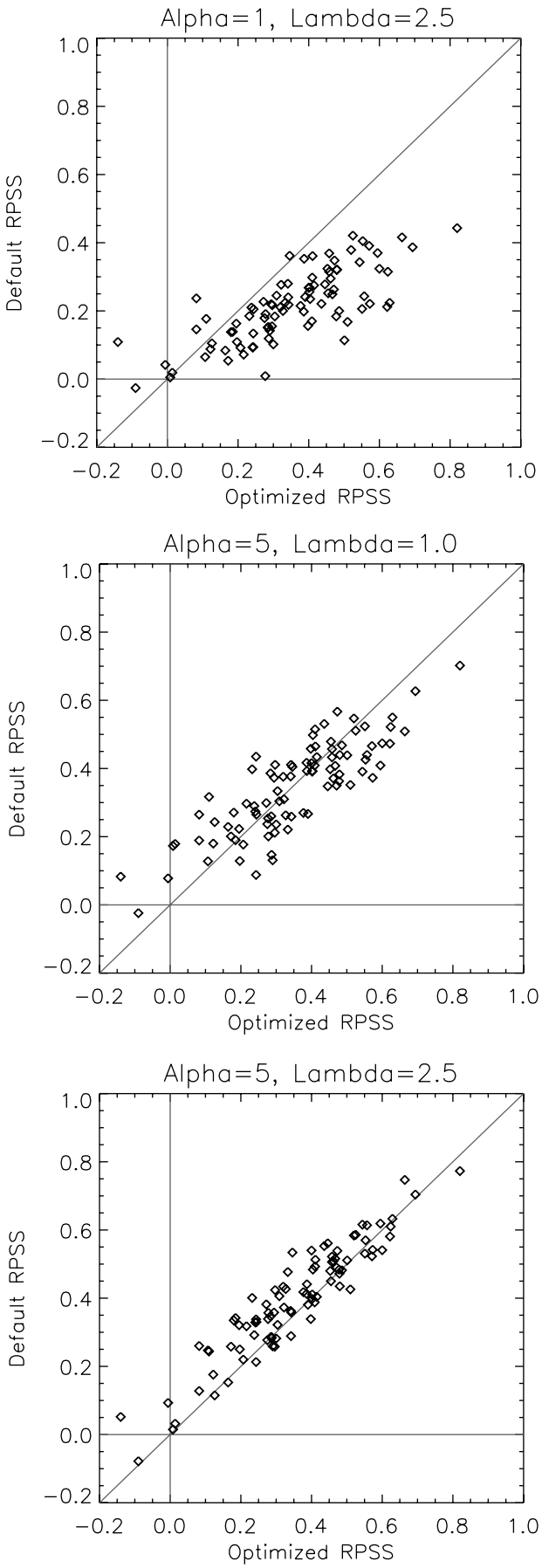


Figure 12. Topology of the ranked probability skill score (RPSS) score, using different α and λ parameters, for the Petrified Forest station in Arizona (34.8°N, 109.9°W). The top plot depicts RPSS values computed using all years; the bottom plot depicts RPSS values computed using El Niño-Southern Oscillation (ENSO) years (defined as when the October value of the Niño 3.4 index was >1.0°C or <−1.0°C).

involves (1) generating probabilistic forecasts of total winter precipitation, based on an ensemble of 50 generated weather sequences (1 October through 31 March) for each of the 10 years that have the most similar Niño 3.4 conditions, and (2) calculating the RPSS value for these 10 probabilistic forecasts. Function evaluations are repeated until the global optimal RPSS value is identified, or after 200 function evaluations, whichever occurs first. Results are cross validated; that is, data for a given water year are not used to estimate the α and λ parameters that are used to produce the optimized generated weather sequence for that water year. The optimization program has a high computational



burden, which was reduced by limiting the maximum number of function evaluations to 200, by restricting analysis to ENSO years, and by restricting attention to the stations in the contiguous United States where the correlation between the October value of the Niño 3.4 index and total 1 October through 31 March precipitation exceeds 0.5 (92 stations).

[37] Figure 13 compares the RPSS values using optimized α and λ parameters against RPSS values for the default α and λ parameters used in previous figures. Recall that analysis is restricted to ENSO years. Each symbol in Figure 13 illustrates the probabilistic forecast skill for a station in the contiguous United States that meets the selection criteria defined above. On the whole, the optimized α and λ parameters provide generated weather sequences with higher probabilistic skill than the default parameters of $\alpha = 1.0$ and $\lambda = 2.5$ (top plot in Figure 13). However, the default parameters of $\alpha = 5.0$ and $\lambda = 1.0$ and $\alpha = 5.0$ and $\lambda = 2.5$ (middle and bottom plots in Figure 13) provide generated weather sequences with equivalent probabilistic forecast skill to the optimized α and λ parameters. The lower optimized RPSS values can be interpreted as arising from situations where the sample size is insufficient to provide stable α and λ estimates. Results suggest parameters of $\alpha = 5.0$ and $\lambda = 2.5$ can be effectively used to condition the weather generator on ENSO indices.

5. Summary and Discussion

[38] A new approach for generating weather has been introduced that preserves the mean, standard deviation, and skewness of the generated precipitation and temperature sequences, while also preserving the temporal persistence and intersite and intervariable correlations. The method resamples data from the historical record "nens" times (nens is the number of ensemble members), and reorders the ensemble members to reconstruct the observed space-time variability in precipitation and temperature fields. The weather generator method has been applied to 2307 stations in the contiguous Unites States. When generated sequences are examined for all of these stations, results show that the weather generator reproduces the summary statistics (mean, standard deviation, skewness) very well. Intersite correlations from the weather generator are slightly lower than observed intersite correlations, due to difficulties in accurately reproducing the intermittent properties of precipitation [see also *Clark et al.*, 2004]. Intersite correlations for temperature are preserved very well. In summary, the weather generator reproduces the statistical moments at individual stations well and acceptably

Figure 13. Comparison of the RPSS when weather sequences are generated using optimized α and λ parameters, and when weather sequences are generated using default α and λ parameters. The plots illustrate results for the parameters $\alpha = 1.0$ and $\lambda = 2.5$, $\alpha = 5.0$ and $\lambda = 1.0$, and $\alpha = 5.0$ and $\lambda = 2.5$, for the top, middle, and bottom plots, respectively. Results are shown for the 92 stations in the contiguous United States where the correlation between the October value of the Niño 3.4 index and total winter precipitation exceeds 0.5.

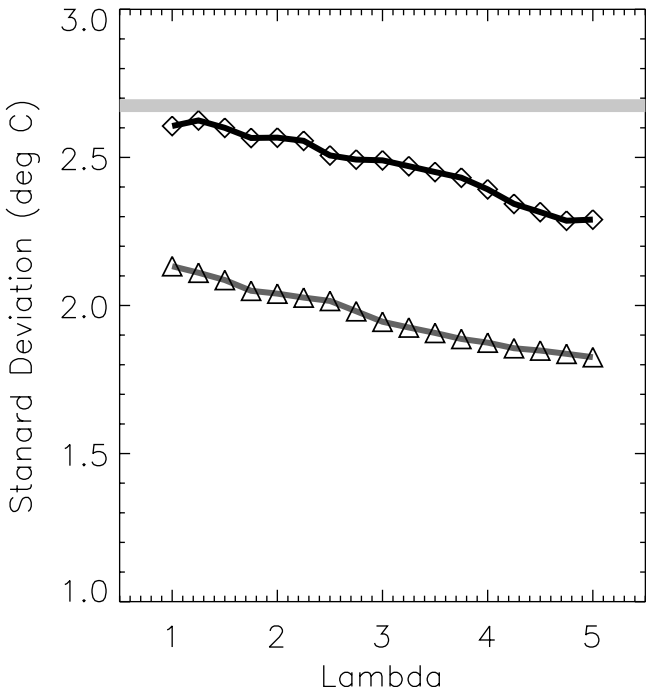


Figure 14. Interannual standard deviation of January maximum temperature at Petrified Forest, for different α and λ parameters, for the case where years at Petrified Forest are ranked from warmest to coldest. The plot illustrates results for λ values ranging from 1.0 to 5.0; the dark line with diamonds shows results for $\alpha = 1.0$, and the light line with triangles shows results for $\alpha = 5.0$. The thick horizontal light line depicts the observed interannual standard deviation in January maximum temperature at Petrified Forest.

duplicates the observed lag-1, intersite, and intervariable correlations.

[39] Methods are introduced to extend the weather generator model to generate weather sequences that reflect El Niño and La Niña conditions. Instead of resampling data from all years in the historical record, data are preferentially resampled from a set of years that are similar in terms of values of the Niño 3.4 index. The resultant set of years are identified using weighting and selection parameters. The weighting parameter gives more weight to years higher up in the ranked list, and the selection parameter restricts attention to a subset of years (e.g., the top 20% of years in the ranked list). The weighting and selection parameters are optimized, and default parameters are suggested for future studies. Example forecasts are provided for a station in Arizona (Petrified Forest; 34.8°N, 109.9°W) where El Niño and La Niña conditions have a strong effect on winter precipitation. The conditioned weather sequences generated using the methods described in this paper are appropriate for use as input to hydrologic models to produce multiseason forecasts of streamflow. Future work will assess the skill of multiseason hydrologic forecasts in numerous small river basins in the contiguous Unites States.

[40] It is possible to use the α and λ parameters discussed in this paper to produce climate change scenarios (e.g., by ranking years from warmest to coldest, as done by *Yates et al.* [2003]). However, it is recommended that this be done cautiously. When conditioning is based on a list of years that are ranked from warmest to coldest, the α and λ parameters will alter not only the mean of the distribution, but the standard deviation and skewness as well. This may not be desired. For example, Figure 14 shows the interannual standard deviation of January maximum temperature at Petrified Forest, constructed using different α and λ values. As expected, the standard deviation decreases when the selection of data is biased toward warm years. If the standard deviation and skewness are important attributes for the construction of climate change scenarios, then alternative methods for conditioning the weather generator must be developed and used. More research is needed in this area.

[41] **Acknowledgments.** This work is part of a collaborative effort between the University of Colorado and the Colorado Basin River Forecast Center that seeks to improve operational streamflow forecasts on intraseasonal to seasonal timescales. The work was funded by the National Oceanic and Atmospheric Administration (NOAA) Office of Global Programs (OGP) under awards NA16GP1587 and NA17RJ1229. The authors are grateful to Marina Timofeyeva for comments on an earlier draft of this manuscript.

References

Buishand, T. A., and T. Brandsma (2001), Multisite simulation of daily precipitation and temperature in the Rhine basin by nearest neighbor resampling, *Water Resour. Res.*, *37*, 2761–2776.

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*, edited by H. Diaz and V. Markgraf, pp. 29–68, Cambridge Univ. Press, New York.

Clark, M. P., M. C. Serreze, and G. J. McCabe (2001), Historical effects of El Niño and La Niña events on the seasonal evolution of the montane snowpack in the Columbia and Colorado River Basins, *Water Resour. Res.*, *37*, 741–757.

Clark, M. P., S. Gangopadhyay, L. E. Hay, B. Rajagopalan, and R. L. Wilby (2004), The Schaake shuffle: A method for reconstructing space-time variability in forecasted precipitation and temperature fields, *J. Hydrometeorol.*, *5*, 243–262.

Connelly, B. A., D. T. Braatz, J. B. Halquist, M. M. DeWeese, L. Larson, and J. J. Ingram (1999), Advanced hydrologic prediction system, *J. Geophys. Res.*, *104*(D16), 19,655–19,660.

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

Duan, Q., S. Sorooshian, and V. K. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, *28*, 1015–1031.

Duan, Q., V. K. Gupta, and S. Sorooshian (1993), A shuffled complex evolution approach of effective and efficient optimization, *J. Optim. Theory Appl.*, *76*, 501–521.

Duan, Q., S. Sorooshian, and V. K. Gupta (1994), Optimal use of the SCE-UA global optimization method for calibrating watershed models, *J. Hydrol.*, *158*, 265–284.

Efron, B. (1979), Bootstrap methods: Another look at the jackknife, *Ann. Stat.*, *7*, 1–26.

Eischeid, J. K., P. A. Pasteris, H. F. Diaz, M. S. Plantico, and N. J. Lott (2000), Creating a serially complete, national daily time series of temperature and precipitation for the western United States, *J. Appl. Meteorol.*, *39*, 1580–1591.

Foufoula-Georgiou, E., and D. P. Lettenmaier (1987), A Markov renewal model for rainfall occurrences, *Water Resour. Res.*, *23*, 875–884.

Gabriel, K. R., and J. Neumann (1962), A Markov chain model for daily rainfall occurrence at Tel Aviv, *Q. J. R. Meteorol. Soc.*, *88*, 90–95.

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

Hay, L. E., G. J. McCabe, D. M. Wolock, and M. A. Ayers (1991), Simulation of precipitation by weather-type analysis, *Water Resour. Res.*, *27*, 493–501.

Jothityangkoon, C., M. Sivapalan, and N. R. Viney (2000), Tests of a space-time model of daily rainfall in southwestern Australia based on nonhomogeneous random cascades, *Water Resour. Res.*, *36*, 267–284.

Rajagopalan, B., and U. Lall (1999), A *k*-nearest neighbor simulator for daily precipitation and other variables, *Water Resour. Res.*, *35*, 3089–3101.

Redmond, K. T., and R. W. Koch (1991), Surface climate and streamflow variability in the western United States and their relationship with large-scale circulation indices, *Water Resour. Res.*, *27*, 2381–2399.

Reek, T., S. R. Doty, and T. W. Owen (1992), A deterministic approach to the validation of historical daily temperature and precipitation data from the cooperative network, *Bull. Am. Meteorol. Soc.*, *73*, 753–762.

Todorovic, P., and D. A. Woolhiser (1975), A stochastic model of n-day precipitation, *J. Appl. Meteorol.*, *14*, 17,024.

Wilks, D. S. (1995), *Statistical Methods in the Atmospheric Sciences: An Introduction*, 467 pp., Academic, San Diego, Calif.

Wilks, D. D. (1998), Multi-site generalizations of a daily stochastic weather generation model, *J. Hydrol.*, *210*, 178–191.

Wilks, D. S., and R. L. Wilby (1999), The weather generation game: A review of stochastic weather models, *Prog. Phys. Geogr.*, *23*, 329–357.

Yates, D., S. Gangopadhyay, B. Rajagopalan, and K. Strzepek (2003), A technique for generating regional climate scenarios using a nearest neighbor algorithm, *Water Resour. Res.*, *39*(7), 1199, doi:10.1029/2002WR001769.

Young, K. C. (1994), A multivariate chain model for simulating climatic parameters from daily data, *J. Appl. Meteorol.*, *33*, 661–671.

D. Brandon and K. Werner, Colorado Basin River Forecast Center, 2242 W North Temple, Salt Lake City, UT 84116, USA.

M. P. Clark and S. Gangopadhyay, Center for Science and Technology Policy Research, Cooperative Institute for Research in Environmental Sciences, 1333 Grandview Avenue, Campus Box 488, University of Colorado, Boulder, CO 80309-0488, USA. (clark@vorticity.colorado.edu)

L. Hay, United States Geological Survey, Water Resources Division, Box 25046, MS 412, Denver Federal Center, Lakewood, CO 80225, USA.

B. Rajagopalan, Department of Civil, Environmental, and Architectural Engineering, Campus Box 428, University of Colorado at Boulder, Boulder, CO 80309-0428, USA.

D. Yates, Research Applications Program, National Center for Atmospheric Research, 3450 Mitchell Lane, Boulder, CO 80301, USA.