The Schaake Shuffle: A Method for Reconstructing Space–Time Variability in Forecasted Precipitation and Temperature Fields

Martyn Clark

Center for Science and Technology Policy Research, Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado

Subhrendu Gangopadhyay

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

Lauren Hay

U.S. Geological Survey, Lakewood, Colorado

Balaji Rajagopalan

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

Robert Wilby

Environment Agency, Trentside Office, West Bridgford, Nottingham, United Kingdom

ABSTRACT

A number of statistical methods that are used to provide local-scale ensemble forecasts of precipitation and temperature do not contain realistic spatial covariability between neighboring stations or realistic temporal persistence for subsequent forecast lead times. To demonstrate this point, output from a global-scale numerical weather prediction model is used in a stepwise multiple linear regression approach to downscale precipitation and temperature to individual stations located in and around four study basins in the United States. Output from the forecast model is downscaled for lead times up to 14 days. Residuals in the regression equation are modeled stochastically to provide 100 ensemble forecasts. The precipitation and temperature ensembles from this approach have a poor representation of the spatial variability and temporal persistence. The spatial correlations for downscaled output are considerably lower than observed spatial correlations at short forecast lead times (e.g., less than 5 days) when there is high accuracy in the forecasts. At longer forecast lead times, the downscaled spatial correlations are close to zero. Similarly, the observed temporal persistence is only partly present at short forecast lead times.

A method is presented for reordering the ensemble output in order to recover the space–time variability in precipitation and temperature fields. In this approach, the ensemble members for a given forecast day are ranked and matched with the rank of precipitation and temperature data from days randomly selected from similar dates in the historical record. The ensembles are then reordered to correspond to the original order of the selection of historical data. Using this approach, the observed intersite correlations, intervariable correlations, and the observed temporal persistence are almost entirely recovered. This reordering methodology also has applications for recovering the space–time variability in modeled streamflow.

1. Introduction

Local-scale forecasts of precipitation and temperature are commonly obtained through statistical postprocess-

ing of output from numerical weather prediction models. A standard approach is model output statistics (MOS), in which predictors from the Numerical Weather Prediction model (NWP; e.g., temperature, humidity, and winds at different pressure levels) are used in a multiple linear regression model to forecast temperature and precipitation at individual stations (Glahn and Lowry 1972). In this approach, the regression equations (or other statistical transfer functions) are developed sep-

Corresponding author address: Martyn Clark, Center for Science and Technology Policy Research, CIRES, University of Colorado, Boulder, CO 80309-0488.
E-mail: clark@vorticity.colorado.edu

arately for each individual station of interest (Crane and Hewitson 1998; Murphy 1999; Wilby et al. 2002; Clark and Hay 2004) or for stations pooled together in a region (Antolik 2000). The latter approach (''pooled regression''), in which stations are stacked, one after the other, to provide a longer sequence of forecasts and observations, is favored in situations when the number of archived forecasts is relatively small, a situation that is very common because of the constant changes in operational forecast models. Both single-site and pooled regression techniques provide more accurate local-scale forecasts of precipitation and temperature than those obtained using raw output from forecast models (Antolik 2000; Clark and Hay 2004).

The problems with the MOS approach, as defined above, are that it does not preserve the spatial covariability between neighboring stations, nor does it preserve the observed temporal persistence in the predicted precipitation and temperature fields. Preserving the space–time variability in precipitation and temperature fields is critical for hydrologic applications; for example, if one is to use the downscaled precipitation and temperature forecasts as input to a hydrologic model to forecast streamflow. Consider the snowmelt-dominated river basins in the western United States. The high spatial coherence in daily temperature fields results in high spatial coherence in daily streamflow. If the spatial covariability in downscaled temperature fields is not preserved, then the resultant streamflow forecasts will misrepresent extreme values—higher runoff in one basin will offset lower runoff in a neighboring basin. Similar problems arise if the temporal persistence in precipitation and temperature fields is not preserved.

In this paper, a method is presented for reordering local-scale ensemble forecasts of precipitation and temperature to reconstruct the space–time variability of precipitation and temperature fields. As an example, results are shown in which predictors from the 1998 version of the National Centers for Environmental Prediction (NCEP) Medium-Range Forecast Model are used in a forward-screening multiple linear regression model to forecast temperature and precipitation at individual stations. Residuals are modeled stochastically to provide ensemble forecasts. In this example, a different statistical model is used for each station, which allows an assessment of the value of this approach for reconstructing the space–time variability of precipitation and temperature patterns based on local-scale forecasts from multiple models (e.g., a multimodel ensemble comprised of outputs from a mix of statistical and dynamical models). The remainder of this paper is organized as follows: the model output and station data are described in section 2, and methods are described in section 3. Results, including assessments of model bias, probabilistic forecast skill, intersite correlations, intervariable correlations, and temporal persistence, are described in section 4.

2. NWP model output and station data

a. The CDC forecast archive

The Climate Diagnostics Center (CDC) in Boulder, Colorado, has generated a ''reforecast'' dataset using a fixed version (circa 1998) of the NCEP operational Medium-Range Forecast Model (MRF). The dataset includes an archive of 14-day forecasts for the last 24 yr (1978–2001). The MRF has been running in near–real time since February 2002, allowing development of stable MOS equations using the archived forecasts and application of those MOS equations in near–real time.

Output variables used to develop MOS equations are (a) the accumulated precipitation for a 12-h period (e.g., 0000–0012 UTC), (b) 2-m air temperature, (c) relative humidity at 700 hPa, (d) 10-m zonal wind speed, (e) 10-m meridional wind speed, (f) total column precipitable water, and (g) mean sea level pressure. Previous work has demonstrated the importance of these variables when downscaling precipitation and temperature (e.g., Clark and Hay 2004).

b. Station data

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. These data were extracted from the National Climatic Data Center (NCDC) Summary of the Day Dataset by J. Eischeid, National Oceanic and Atmospheric Administration (NOAA) Climate Diagnostics Center (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 (a) extreme values, (b) internal consistency among variables (e.g., maximum temperature less than minimum temperature), (c) constant temperature (e.g., 5 or more days with the same temperature are suspect), (d) excessive diurnal temperature range, (e) invalid relations between precipitation, snowfall, and snow depth, and (f) unusual spikes in temperature time series. Records at most of these stations start in 1948 and continue through to the present.

Results are presented for four river basins used by Hay et al. (2002) and Clark and Hay (2004): (a) Cle Elum River in central Washington, (b) East Fork of the Carson River on the California–Nevada border, (c) Animas River in southwest Colorado, and (d) Alapaha River in southern Georgia (Fig. 1). Attention is restricted to the ''best stations'' in the cooperative network that are located within a 100-km search radius of the center of these four basins. These best stations are defined as those with less than 10% missing or questionable data during the period 1979–99. This provides 15 stations for the Animas basin, 16 stations for the Carson basin, 18 stations for the Cle Elum basin, and 10 stations for the Alapaha basin (Table 1).
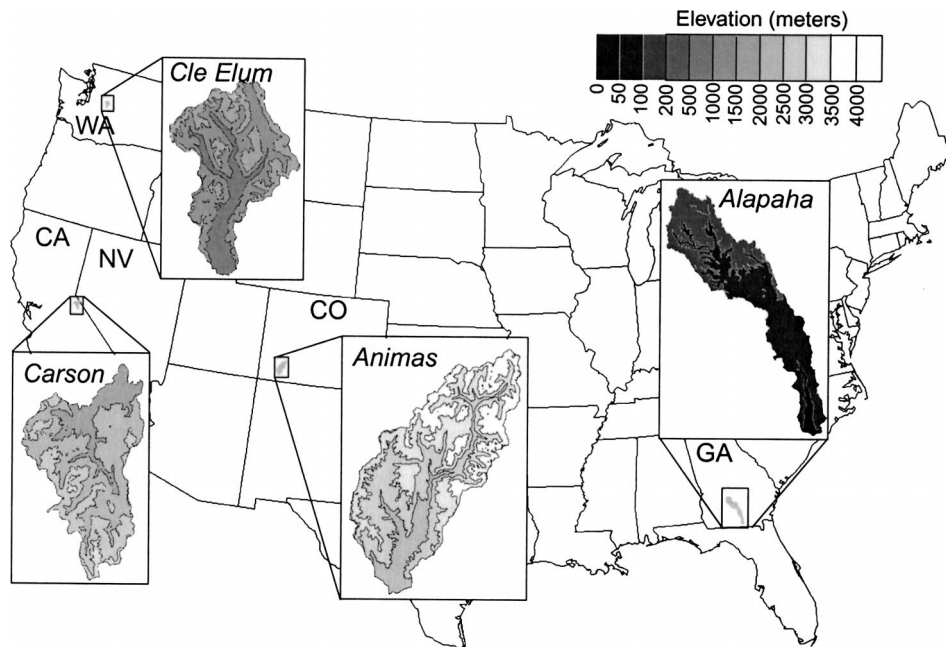
FIG. 1. Location and topography of the study basins.

3. Method

a. Multiple linear regression model

Multiple linear regression is used to develop the MOS equations (Glahn and Lowry 1972; Antolik 2000), in which predictors from the NCEP MRF model are used to forecast temperature and precipitation at individual stations. Separate regression models are developed for each variable, each station, each forecast lead time (1–14 days), and each month. Precipitation is disaggregated into occurrence and amounts—logistic regression is used to model precipitation occurrence, and ordinary least squares regression is used to model precipitation amounts.

The intermittent and skewed character of the daily precipitation data make it necessary to preprocess these data prior to training the regression equations. The station time series of precipitation are first disaggregated into a time series of occurrence (1 = wet days and 0 = dry days) and precipitation amounts (only wet days). The time series of occurrence is used as the response variable for the logistic regression model, and the time series of precipitation amounts is used as the response variable for the ordinary least squares model. For precipitation amounts, the station precipitation data (only wet days) are transformed to a normal distribution using a nonparametric probability transform (Panofsky and Brier 1963). For each data point, the cumulative probability of observed precipitation is computed. This is matched with the cumulative probability from a standard normal distribution (mean of zero and standard deviation of one), and the normal deviate corresponding to the cumulative probability in the standard normal dis-

tribution is used to replace the original precipitation value.

The regression models have the form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon, \tag{1}$$

$$p = 1 - \frac{1}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)}, \tag{2}$$

where Eq. (1) is for ordinary least squares regression and Eq. (2) is for logistic regression. In these equations, y is the response variable in the ordinary least squares model (e.g., maximum temperature, minimum temperature, or precipitation amounts at a station location), p is the response variable in the logistic regression model (e.g., the probability of precipitation at a station location), β_0 is the regression constant, β_1 is the slope coefficient for the first explanatory variable (x_1), β_2 is the slope coefficient for the second explanatory variable (x_2), β_k is the slope coefficient for the kth explanatory variable (x_k), and ε is the remaining unexplained noise in the data (the error). The explanatory variables (x_1, x_2, . . . , x_k) are forecasted outputs from the NCEP MRF model (e.g., 700-hPa relative humidity, mean sea level pressure). The solution of ordinary least squares equation was done using the singular value decomposition (SVD) algorithm (Press et al. 1992), and the logistic regression equation was solved iteratively using the method (and code) presented by Agterberg (1989).

The forward-selection approach is used to identify the variables used in the regression equations (Antolik 2000). This procedure first identifies the explanatory variable (e.g., 700-hPa relative humidity; x_1) that ex-

TABLE 1. List of stations used in each of the four basins.

Basin	Lat (°N)	Lon (°E)	Station ID	Station name
Animas	38.400	252.483	CO1609.COOP	CIMARRON 3 SE
	37.234	251.950	CO3016.COOP	FORT LEWIS
	37.767	252.867	CO3951.COOP	HERMIT 7 ESE
	38.025	252.685	CO4734.COOP	LAKE CITY
	37.199	251.512	CO5531.COOP	MESA VERDE NP
	38.132	251.714	CO6012.COOP	NORWOOD
	38.021	252.331	CO6203.COOP	OURAY
	37.243	252.983	CO6258.COOP	PAGOSA SPRINGS
	37.713	251.961	CO7017.COOP	RICO
	37.726	252.732	CO7050.COOP	RIO GRANDE RESERVOIR
	37.949	252.127	CO8204.COOP	TELLURIDE 4 WNW
	37.384	252.418	CO8582.COOP	VALLECITO DAM
	36.833	252.000	NM0692.COOP	AZTEC RUINS NATL
	36.936	253.000	NM2608.COOP	DULCE
	36.817	252.383	NM6061.COOP	NAVAJO DAM
Carson	39.383	239.900	CA0931.COOP	BOCA
	38.250	240.767	CA1072.COOP	BRIDGEPORT
	38.283	239.683	CA1277.COOP	CALAVERAS BIG TREES
	38.249	239.140	CA1428.COOP	CAMP PARDEE
	37.967	240.083	CA1697.COOP	CHERRY VALLEY DAM
	39.317	239.767	CA2467.COOP	DONNER MEMORIAL ST
	39.167	239.867	CA8758.COOP	TAHOE CITY
	38.446	239.501	CA8928.COOP	TIGER CREEK PH
	39.333	239.817	CA9043.COOP	TRUCKEE RS
	38.700	239.967	CA9105.COOP	TWIN LAKES
	37.757	240.412	CA9855.COOP	YOSEMITE PARK HDQTRS
	39.150	240.233	NV1485.COOP	CARSON CITY
	39.083	240.050	NV3205.COOP	GLENBROOK
	39.000	240.250	NV5191.COOP	MINDEN AIRPORT
	39.083	240.883	NV8822.COOP	WABUSKA 6 SE
	39.000	240.883	NV9229.COOP	YERINGTON
Cle Elum	47.767	238.517	WA0456.COOP	BARING
	47.167	238.000	WA0945.COOP	BUCKLEY 1 NE
	47.417	238.267	WA1233.COOP	CEDAR LAKE
	47.836	239.962	WA1350.COOP	CHELAN
	47.183	239.083	WA1504.COOP	CLE ELUM 1
	47.000	239.483	WA2505.COOP	ELLENSBURG
	47.383	238.033	WA4486.COOP	LANDSBURG
	47.133	237.733	WA5224.COOP	MC MILLIN RESERVOIR
	47.850	238.017	WA5525.COOP	MONROE
	47.150	238.067	WA5704.COOP	MUD MOUNTAIN DAM
	47.300	238.150	WA6295.COOP	PALMER 3 SE
	47.183	240.133	WA6880.COOP	QUINCY 3 S
	47.450	237.700	WA7473.COOP	SEATTLE TCOMA WSCMO
	47.542	238.164	WA7773.COOP	SNOQUALMIE FALLS
	47.871	238.283	WA8034.COOP	STARTUP 1 E
	47.430	239.689	WA9074.COOP	WENATCHEE
	47.402	239.795	WA9082.COOP	WENATCHEE
	46.568	239.457	WA9465.COOP	YAKIMA AIRPORT
Alapaha	31.583	275.833	GA0140.COOP	ALBANY
	31.533	277.483	GA0211.COOP	ALMA FAA AIRPORT
	31.183	275.800	GA1500.COOP	CAMILLA 3 SE
	31.967	276.217	GA2266.COOP	CORDELE WATER WORKS
	31.517	277.150	GA2783.COOP	DOUGLAS 2 NNE
	32.200	276.794	GA2966.COOP	EASTMAN 1 W
	31.717	276.750	GA3386.COOP	FITZGERALD
	31.033	277.200	GA4429.COOP	HOMER VILLE 3 WSW
	31.167	276.250	GA6087.COOP	MOULTRIE 2 ESE
	31.483	276.467	GA8703.COOP	TIFTON EXP STN

plains most of the variance of the response variable (y). It then searches through the remaining variables and selects the second explanatory variable (x_2) that reduces the largest portion of the remaining unexplained variance in combination with the variable already chosen. If the improvement in explained variance exceeds a given threshold (taken here as 1%), the variable is included in the multiple linear regression equation. The remaining variables are examined in the same way until no further improvement is obtained based on the correlation threshold. Cross-validation approaches are used to validate the regression equations. Each year in the time series is successively held out of the sample used to develop the regression equations, and the equations are used to predict data for the withheld year. This process is repeated for all years in the record.

Application of the regression model is most straightforward for forecasts of maximum and minimum temperature. The regression model is formulated using the method just described, and the value of y predicted from the regression equation ($\hat{y}$) is given for each data point in the time series [Eq. (1)]. The error term in the regression equation, ε, is then modeled stochastically to inflate the variance of the predicted values and generate an ensemble of forecasts:

$$y_{\text{iens}} = \hat{y} + N\sigma_e, \tag{3}$$

where y_{iens} is the predicted value for a given ensemble member, $\hat{y}$ is the value predicted by the regression equation, N is a random number from a standard normal distribution, and σ_e is the standard deviation of the regression residuals ($y - \hat{y}$). One hundred ensemble members were generated using this procedure.

The procedure is more complex for precipitation. To account for the intermittent properties of precipitation, precipitation is modeled in a two-stage process. Logistic regression is used to estimate precipitation occurrence [Eq. (2)], and ordinary least squares regression [Eq. (1)] is used to estimate precipitation amounts. Precipitation can then be modeled (in normal space) as follows:

$$y_{\text{iens}} = \begin{cases} 0, & \text{when } \hat{p} < u \\ \hat{y} + N\sigma_e, & \text{when } \hat{p} \geq u. \end{cases} \tag{4}$$

In Eq. (4), $u \sim U[0, 1]$ is a random number from a uniform distribution ranging from zero to one, and $\hat{p}$ is the probability of precipitation occurrence predicted from the logistic regression model [Eq. (2)]. If $\hat{p} < u$, then we assume there is no precipitation. If $\hat{p} \geq u$, precipitation is set to occur, and the precipitation amount is computed using Eqs. (1) and (4). When $\hat{p} \geq u$, the forecasted normal deviates from Eq. (4) are then transformed back to the original gamma-type distribution of observed precipitation using the nonparametric probability transform technique described above. The stochastic modeling of the regression residuals inflates the variance of precipitation and temperature forecasts, re-

ducing problems of variance underestimation that are typical of regression-based models.

b. The ensemble reordering method

The ensemble reordering method was described to the authors by Dr. J. Schaake, National Weather Service Office of Hydrologic Development, in October 2002 and is hereafter referred to as the Schaake Shuffle. For a given time, the starting point is a three-dimensional matrix of ensemble forecasts $\mathbf{X}_{i,j,k}$, 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 $\mathbf{Y}_{i,j,k}$ 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. The dates used to populate the matrix $\mathbf{Y}$ are selected so as to lie within 7 days before and after the forecast date (dates can be pulled from all years in the historical record, except for the year of the forecast). Populating the $\mathbf{Y}$ matrix in this way means that data from the same date is used for all stations (j) and variables (k).

For a given station (j) and variable (k), the Schaake shuffle method can be formulated as follows: Let $\mathbf{X}$ be a vector of n ensemble forecasts (x) and $\boldsymbol{\chi}$ be the sorted vector of $\mathbf{X}$, that is,

$$\mathbf{X} = (x_1, x_2, \ldots, x_n), \quad \text{and} \tag{5}$$

$$\boldsymbol{\chi} = (x_{(1)}, x_{(2)}, \ldots, x_{(n)}), \qquad x_{(1)} \leq x_{(2)} \cdots \leq x_{(n)}. \tag{6}$$

Also, let $\mathbf{Y}$ be a vector of n selected historical observations (y), and $\boldsymbol{\gamma}$ be the sorted vector of $\mathbf{Y}$, that is,

$$\mathbf{Y} = (y_1, y_2, \ldots, y_n), \quad \text{and} \tag{7}$$

$$\boldsymbol{\gamma} = (y_{(1)}, y_{(2)}, \ldots, y_{(n)}), \qquad y_{(1)} \leq y_{(2)} \cdots \leq y_{(n)}. \tag{8}$$

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

As an example, forecasts of maximum temperature for 10 ensemble members at a given station on a given date, and the corresponding selection of historical observations, may provide vectors of $\mathbf{X}$, $\boldsymbol{\chi}$, $\mathbf{Y}$, $\boldsymbol{\gamma}$, and $\mathbf{B}$ that are

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

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

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

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

and

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

In this example, data from the first date selected from the historical record (10.7 in vector $\mathbf{Y}$) is ranked fifth lowest of the 10 ensemble members, as shown in the

A. Ranked ensemble output

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

C. Ranked historical observations

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

B. Randomly selected historical observations

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

D. Final ensemble (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

FIG. 2. The ensemble reordering method for a hypothetical ensemble of 10 members and for a given variable (e.g., max temperature), showing (a) the ranked ensemble output for three stations, (b) a random selection of historical observations for the three stations, (c) the ranked historical observations, and (d) the final reordered output. See text for further details.

vectors γ and **B**. Data from the second date is ranked third lowest (9.3 in vector **Y**); and data from the third date selected from the historical record is ranked lowest of all 10 ensemble members (6.8 in vector **Y**).

Now, the final step is to construct the reordered vector $\mathbf{X}^{ss}$, which are the final forecast values:

$$\mathbf{X}^{ss} = (x_1^{ss}, x_2^{ss}, \ldots, x_n^{ss}), \tag{9}$$

where

$$x_q^{ss} = x_{(r)}, \tag{10}$$

$$q = \mathbf{B}[r], \tag{11}$$

$$r = 1, \ldots, n. \tag{12}$$

Recall that the subscripts in parentheses refer to the elements in the sorted vector χ. Following through with the numbers from the example above provides

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

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

$$x_2^{ss} = x_{(3)} = 8.8; \quad \text{and so on.}$$

Hence, in this example,

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

The Schaake Shuffle approach is demonstrated graphically in Fig. 2 through an example (pseudo-FORTRAN code illustrating the algorithm is given in the appendix).

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

The dark ellipses in Fig. 2 correspond to the first ensemble member extracted from the historical record. When this is not sorted (i.e., the vector **Y**), the values are 10.7, 10.9, and 13.5, for stations 1, 2, and 3, respectively (Fig. 2b). When these values are sorted with respect to all other ensemble members, the first observed ensemble is ranked fifth for station 1, sixth for station 2, and fourth for station 3 (Fig. 2c). For the first ensemble member, the ranks 5, 6, and 4 are actually the values of the index (r) for the three stations [Eq. (10)]—values for the first ensemble member, once resorted [$x_q^{ss} = x_{(r)}$], are 10.1 at station 1 [$x_1^{ss} = x_{(5)}$; see Fig. 2a], 9.3 at station 2 [$x_1^{ss} = x_{(6)}$], and 14.5 at station 3 [$x_1^{ss} = x_{(4)}$]. Also consider the second ensemble in Fig. 2 (the

light ellipses). Ensemble 2 from the historical record is ranked third, second, and fifth for stations 1, 2, and 3, respectively (Fig. 2c), such that the final shuffled output for ensemble 2 is 8.8 (station 1), 7.2 (station 2), and 15.6 (station 3).

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 downscaled ensemble member at the two stations is matched with the rank of each randomly selected observation, meaning that, for all ensemble members, the rank of the downscaled realizations will be similar at the two stations. When this process is repeated for all forecasts, the ranks of a given ensemble member will on average be similar for the two stations, and the spatial correlation will be reconstructed once the randomly selected days are resorted. This reasoning is identical for intervariable correlations.

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 **Y** are only used for the first forecast lead time (i.e., day+1) and persisted for subsequent forecast lead times. In the example presented in Fig. 2, the dates for the forecast with a lead time of 2 days would be 9 January 1996 for the first ensemble member, 18 January 1982 for ensemble 2, 14 January 2000 for ensemble 3, and so on. High temporal persistence (e.g., as measured through correlation statistics) means that the historical observations for subsequent days will, on average, have a similar rank. Because the ranked ensemble output is matched with the rank of successive observations, the temporal persistence is reconstructed once the ensemble output is resorted.

4. Results

a. Bias

The downscaling approach correctly preserves the mean seasonal cycles in precipitation and temperature. Figure 3 illustrates the forecasted and observed monthly climatologies of precipitation and maximum and minimum temperature for example stations in each of the study basins (forecast lead time is 5 days). The thin line (and small diamonds) depicts the observed climatologies, and the box and whiskers illustrate the climatologies of the minimum, lower quartile, median, upper quartile, and maximum of all forecasted ensemble members. In all cases the observed monthly mean values lie within the range of the forecast ensemble.

The results presented in Fig. 3 are not altogether unexpected. Even though all downscaled realizations are generated through a detailed cross-validation exercise, the regression equations are constrained to provide unbiased estimates (refer to section 3). The data "held out" in successive validation periods are not very different from the data used to estimate the coefficients in the regression equations. A more complete assessment of bias can be obtained through considering measures of reliability, or conditional bias.

The reliability diagram depicts the conditional probability that an event occurred (o_1) given different probabilistic forecasts (y_i), that is, $p(o_1 \mid y_i)$, $i = 1, \ldots , I$, where the subscript i denotes different probability categories (Wilks 1995). This can be implemented as follows: First, the ensemble output (100 ensemble members) is converted into probabilistic forecasts (i.e., the probability a specific event occurs). Next, the observed data are converted to a binary time series—a day is assigned "one" if the event occurs and "zero" if it does not. The above steps produce a set of probabilistic forecast–observation pairs for each variable, station, month, and forecast lead time. Finally, the forecasted probabilities are classified into I categories (i.e., probabilities between 0.0 and 0.1, between 0.1 and 0.2 . . . between 0.9 and 1.0), and for each category both the average forecasted probability (y_i) and the average of the observed binary data (o_1; i.e., the observed relative frequency) is calculated. In this case, the "event" is that the day is forecasted to lie in the upper tercile of the distribution, and the probability is simply calculated as all ensemble members in the upper tercile divided by the total number of ensemble members. The upper tercile is used in this example to focus attention on significant streamflow events, due to either snowmelt (temperature) or rainfall. This procedure provides a method to assess the conditional bias in probability space.

Results show that, on average, the forecasted probabilities match the observed relative frequencies remarkably well for all variables. Figure 4 shows reliability diagrams for precipitation and maximum and minimum temperature for each of the study basins. The dark line with triangles depicts the reliability diagram for the month of January, and the light line with squares depicts the reliability diagram for the month of July (again the forecast lead time is 5 days). The forecast–observation pairs are lumped together for all stations in each basin, so the reliability diagrams illustrate results for each basin as a whole. The major deficiency in the reliability diagrams is the tendency for a low observed relative frequency at high forecasted probabilities, most pronounced for precipitation in the Alapaha River basin. In other words, when a high probability of an event is forecasted, the actual occurrence of that event is less common. Note, however, that the sample size at high forecast probabilities is often very small. When all plots are considered there is very little conditional bias in the forecasted precipitation and temperature fields. Note that the biases are identical for shuffled and unshuffled ensemble members—the Schaake Shuffle just alters the
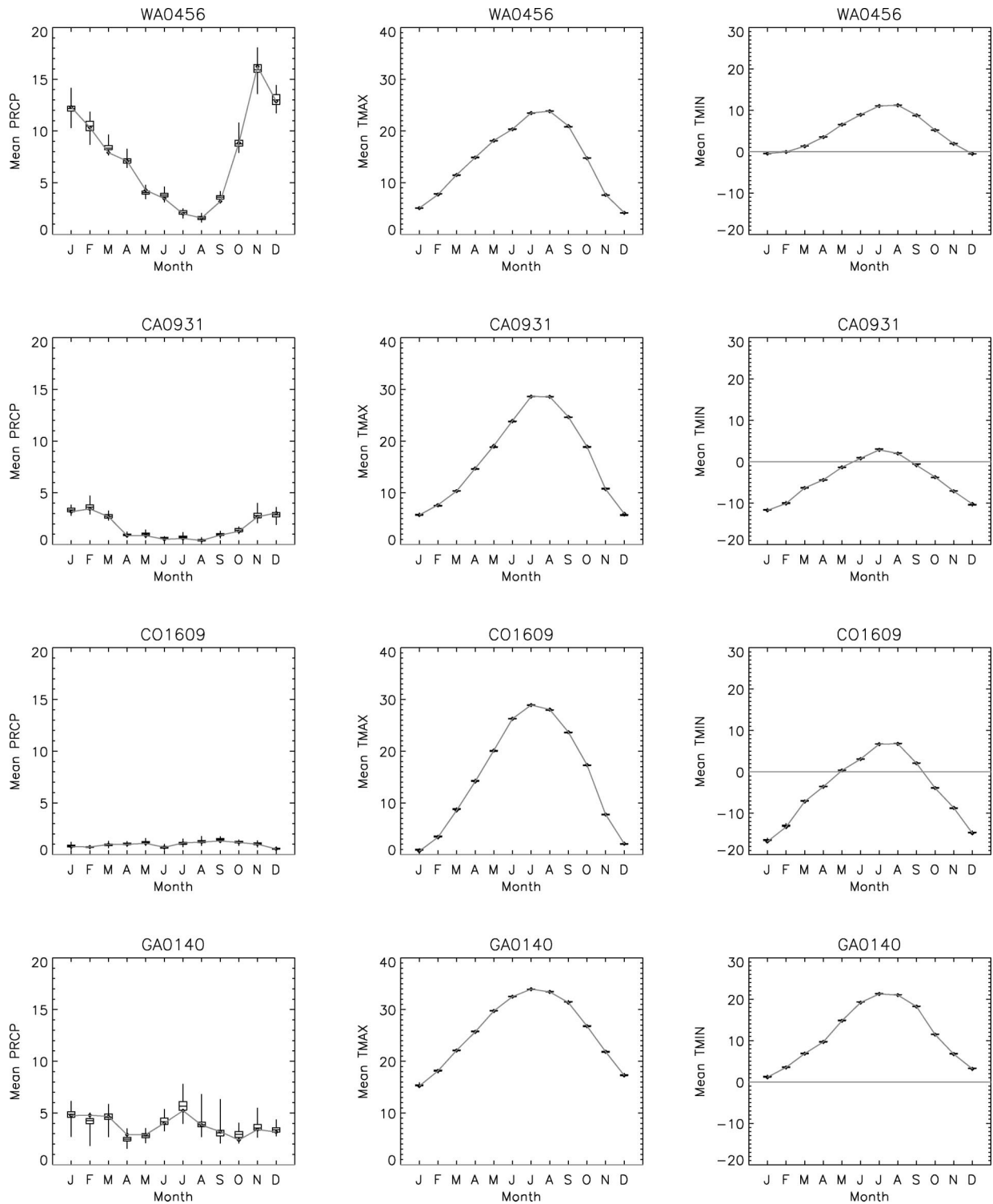
FIG. 3. Box-and-whisker plots illustrating the (left) observed and forecasted climatological mean precipitation, (middle) max temperature, and (right) min temperature for example stations in the (top) Cle Elum, (second from top) the East Fork of the Carson, (third from top) the Animas, and (bottom) the Alapaha basin. The observed climatology is depicted by the thin lines and small diamonds, and the box and whiskers illustrate the forecasted climatological mean for the lowest ensemble, the lower quartile, the median, the upper quartile, and the maximum ensemble.
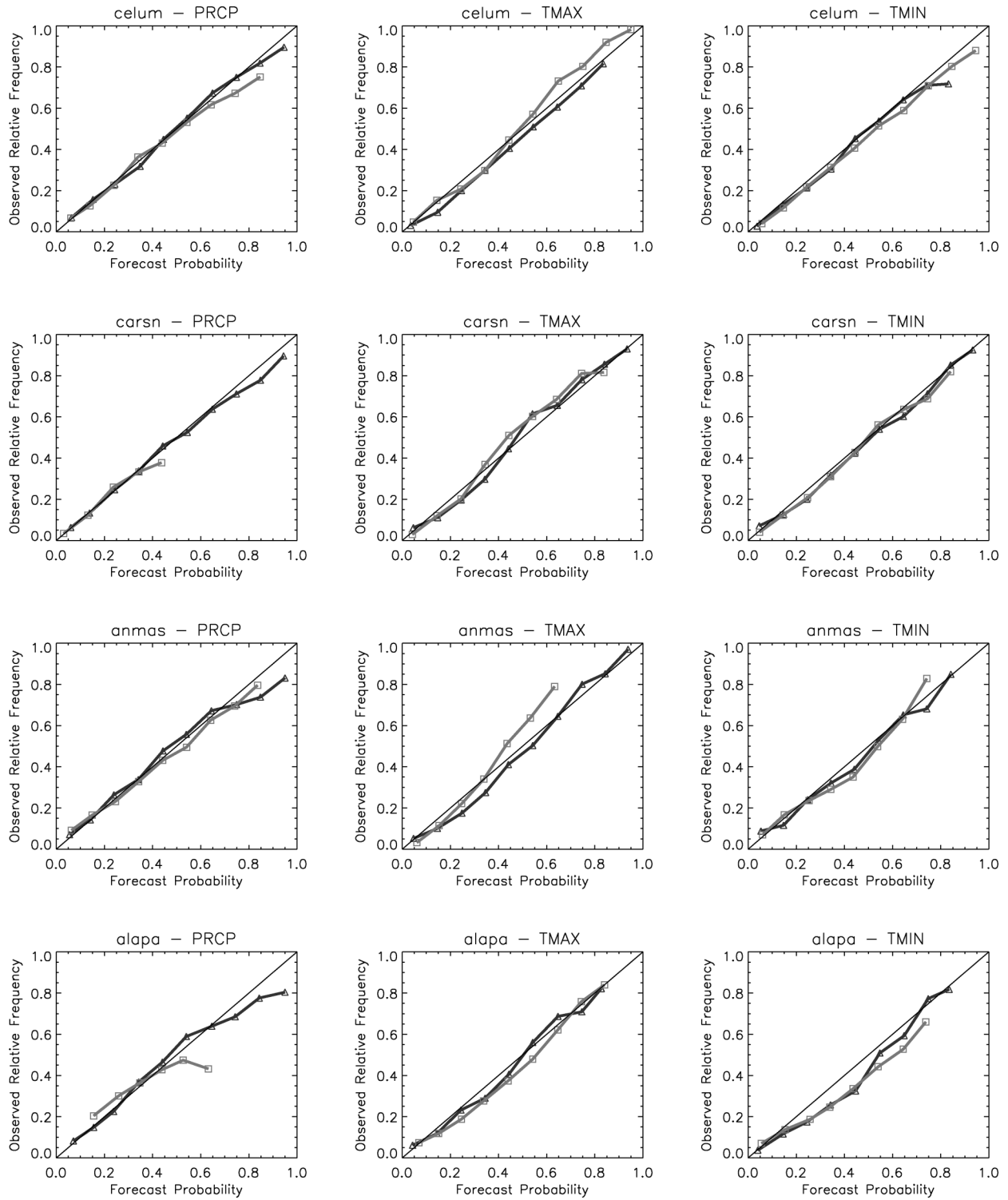
FIG. 4. Reliability diagrams (conditional bias) for the four study basins. The dark line with triangles illustrates the conditional bias for Jan, and the light line with squares illustrates the conditional bias for Jul. Data are plotted only if there are at least 30 cases in a given bin. Results are shown for the forecast lead time of 5 days. See text for further details.

order of the ensemble members, not their values (see section 3b).

b. Skill

The probabilistic skill of the downscaled precipitation and temperature forecasts is assessed using the ranked probability skill score (RPSS; Wilks 1995). The RPSS is based on the ranked probability score (RPS) computed for each forecast–observation pair:

$$\text{RPS} = \sum_{m=1}^{J} (Y_m - O_m)^2, \qquad (13)$$

where Y_m is the cumulative probability of the forecast for category m, and O_m is the cumulative probability of the observation for category m. This is implemented as follows: First, the observed time series is used to distinguish 10 (J) possible categories for forecasts of precipitation and temperature (i.e., the minimum value to the 10th percentile, the 10th percentile to the 20th percentile . . . the 90th percentile to the maximum value). These categories are determined separately for each month, variable, and station. Next, for each forecast–observation pair, the number of ensemble members forecast in each category is determined (out of 100 ensemble members), and their cumulative probabilities are computed. Similarly, the appropriate category for the observation is identified, and the observation's cumulative probabilities are computed (i.e., all categories below the observation's position are assigned ''0,'' and all categories equal to and above the observation's position are assigned ''1''). Now, the RPS is computed as the squared difference between the observed and forecast cumulative probabilities, and the squared differences are summed over all categories [Eq. (13)].

The RPSS is then computed as

$$\text{RPSS} = 1 - \frac{\overline{\text{RPS}}}{\overline{\text{RPS}}_{\text{clim}}}, \qquad (14)$$

where $\overline{\text{RPS}}$ is the mean ranked probability score for all forecast–observation pairs, and $\overline{\text{RPS}}_{\text{clim}}$ is the mean ranked probability score for climatological forecasts. For temperature, $\overline{\text{RPS}}_{\text{clim}}$ is computed using an equal probability in each of the m categories defined in Eq. 13 (i.e., $1/J$); for precipitation, the probability for the first category (zero precipitation) is taken as the observed probability of no precipitation, and the probability for all other categories is taken as $1/(J - 1)$ [see Eq. (13)].

The forecast ensemble has much higher probabilistic skill than the climatological for lead times up to about 7 days. Figure 5 plots the median RPSS values of all stations against forecast lead time for precipitation and maximum and minimum temperature in each of the study basins. The dark line with triangles depicts the median RPSS for January, and the light line with squares depicts the median RPSS for July. Most apparent in Fig.

5 is that the probabilistic skill of precipitation forecasts is much lower in July than in January in all basins. This is anticipated because of the increased prevalence of small-scale convective precipitation in July (see also Clark and Hay 2004). The decreases in RPSS with increased forecast lead time should be noted for the following discussion.

c. Intersite correlations

Accurate representation of the spatial covariability of precipitation and temperature fields is an important consideration in forecasting streamflow. For example, imagine a case in which precipitation in two subbasins within a watershed is highly correlated. Underestimating the spatial correlation structure will mean that forecasts of high precipitation in one subbasin (for a given ensemble member) will be canceled out by forecasts of moderate or low precipitation in the neighboring subbasin (for that ensemble member), and the resultant ensemble forecast of streamflow will have much fewer extreme values than would be expected if the spatial covariability in precipitation had been preserved perfectly. Herein lies the value of the Schaake Shuffle—the technique reorders the downscaled ensemble output so as to reconstruct the observed space–time variability in the downscaled precipitation and temperature fields.

The performance of the Schaake Shuffle for preserving intersite correlations is quite remarkable. Figure 6 illustrates the correlation between two example stations in the Cle Elum River basin for unshuffled and shuffled ensemble output for forecast lead times up to 14 days. The top plots show spatial correlations for precipitation and maximum and minimum temperature for the unshuffled forecast ensemble, and the bottom plots show the spatial correlations for the shuffled forecast ensemble. The thick gray line depicts the observed correlation between the two stations, and the box and whiskers depict the minimum, lower quartile, median, upper quartile, and maximum of the forecasted ensemble members. The spatial correlations for the shuffled ensemble are similar to the observed correlations. By contrast, the spatial correlations for the unshuffled ensemble output are much lower than the observed correlations, particularly at longer forecast lead times when there is little skill in the forecasts.

To evaluate the generality of the performance of the Schaake Shuffle method, the observed correlation is plotted against the median correlation from the shuffled and unshuffled ensemble output for each possible station pair for the month of January for the forecast lead time of 5 days (Fig. 7). The median unshuffled versus observed correlations are depicted as dark triangles, and the median shuffled versus observed correlations are depicted as light squares. Figure 7 demonstrates the capability of the Schaake shuffle to reconstruct the spatial variability in precipitation and maximum and minimum temperature over a wide range of cases—the shuffled
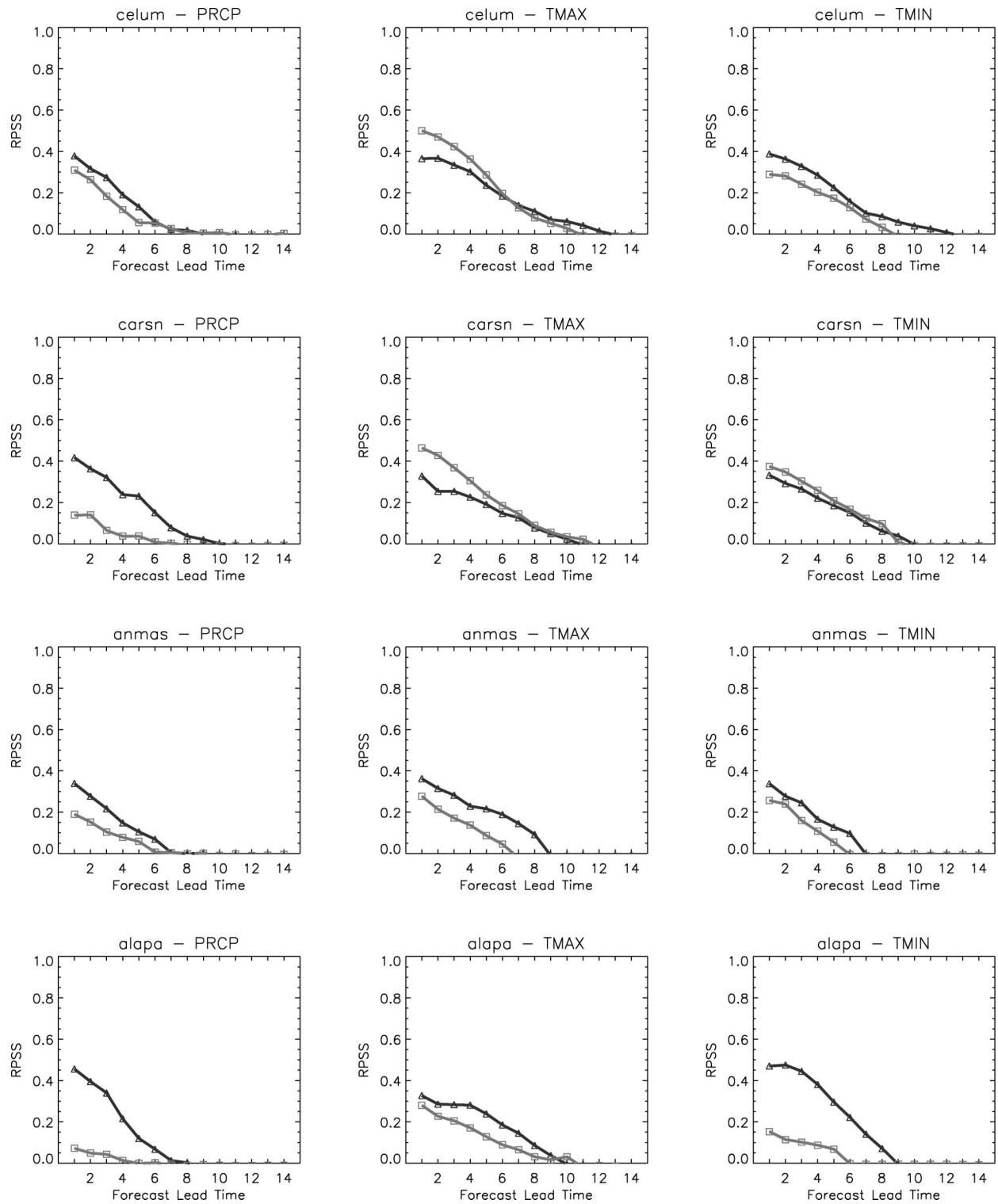
FIG. 5. The median RPSS of all stations in each of the four study basins. The dark line with triangles depicts the median RPSS for Jan, and the light line with squares depicts the median RPSS for Jul. See text for further details.
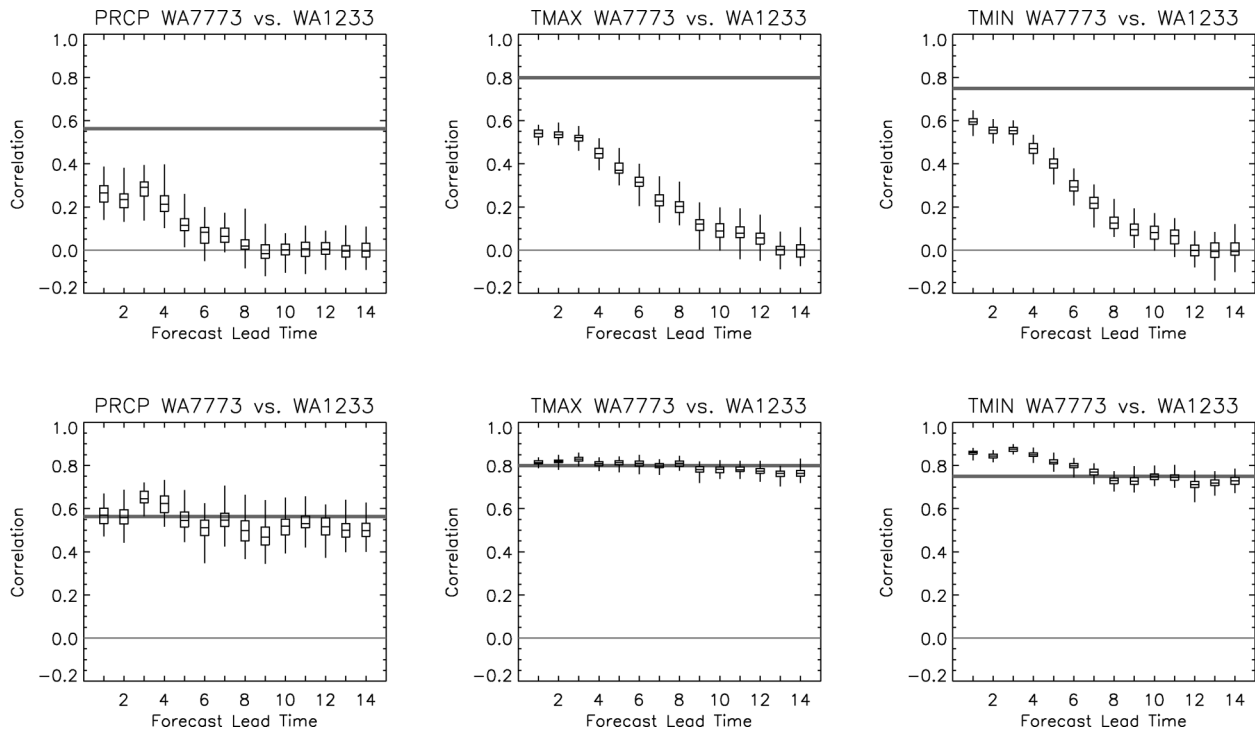
FIG. 6. The spatial correlation between two example stations in the Cle Elum basin in Jan for (top) raw ensemble output and (bottom) reordered ensemble output. The thick line depicts the observed correlation between the two stations, and the box and whiskers depict the minimum, lower-quartile, median, upper-quartile, and maximum correlations of all ensemble members.

ensemble output is much closer to the observed correlation than unshuffled output for all station pairs.

The Schaake shuffle technique does tend to slightly underestimate the observed spatial correlations for precipitation (left column in Fig. 7). This occurs because of the intermittent properties of precipitation time series. Consider the ranked "downscaled" and "observed" ensembles in Fig. 2. Both of these ensembles may include zero precipitation days. In the case that the downscaled ensemble has more precipitation days than the observed ensemble, some of the smaller downscaled precipitation amounts will be matched with zero precipitation days. These smaller precipitation amounts will consequently have an identical rank, and their assignment to a given ensemble member is random. The solution to this problem is to interpolate precipitation data from surrounding stations to fill in the zero precipitation days with trace precipitation amounts. Another case is that the downscaled ensemble has less precipitation days than the observed ensemble. In this case, zero precipitation days will be matched with observed precipitation amounts, and their assignment to a given ensemble member is entirely random. This latter problem is intractable. In spite of such problems with daily precipitation data, however, the Schaake shuffle does preserve most of the spatial correlation structure in precipitation fields and certainly provides a significant improvement over the unshuffled ensemble output.

Figure 8 shows correlations for all possible station pairs for the month of July. Results are fairly similar for the January plot in Fig. 7, except for precipitation in the Carson and Cle Elum River basins, where the shuffled output provides a poor representation of the observed spatial correlation structure. Precipitation is rare in the Carson basin in July, exemplifying the problems with matching zero days described above. Precipitation results for shuffled output in the Carson basin do, however, improve upon the unshuffled output. For other variables and basins, the Schaake Shuffle provides a credible portrayal of the observed spatial correlation structure in precipitation and temperature fields.

d. Intervariable correlations

The reconstruction approach should also be capable of preserving the correlation between variables (e.g., correlations between precipitation and temperature). Suppose if in Fig. 2 the three time series were named "precipitation," "maximum temperature," and "minimum temperature" instead of "station 1," "station 2," and "station 3." If the randomly selected observations between the two variables have a similar rank (i.e., high intervariable correlation), then this correlation structure should be reconstructed once the ensemble output is reordered. By using the same randomly selected dates for all stations and variables, both the spatial correlation structure and the consistency between variables should be preserved.
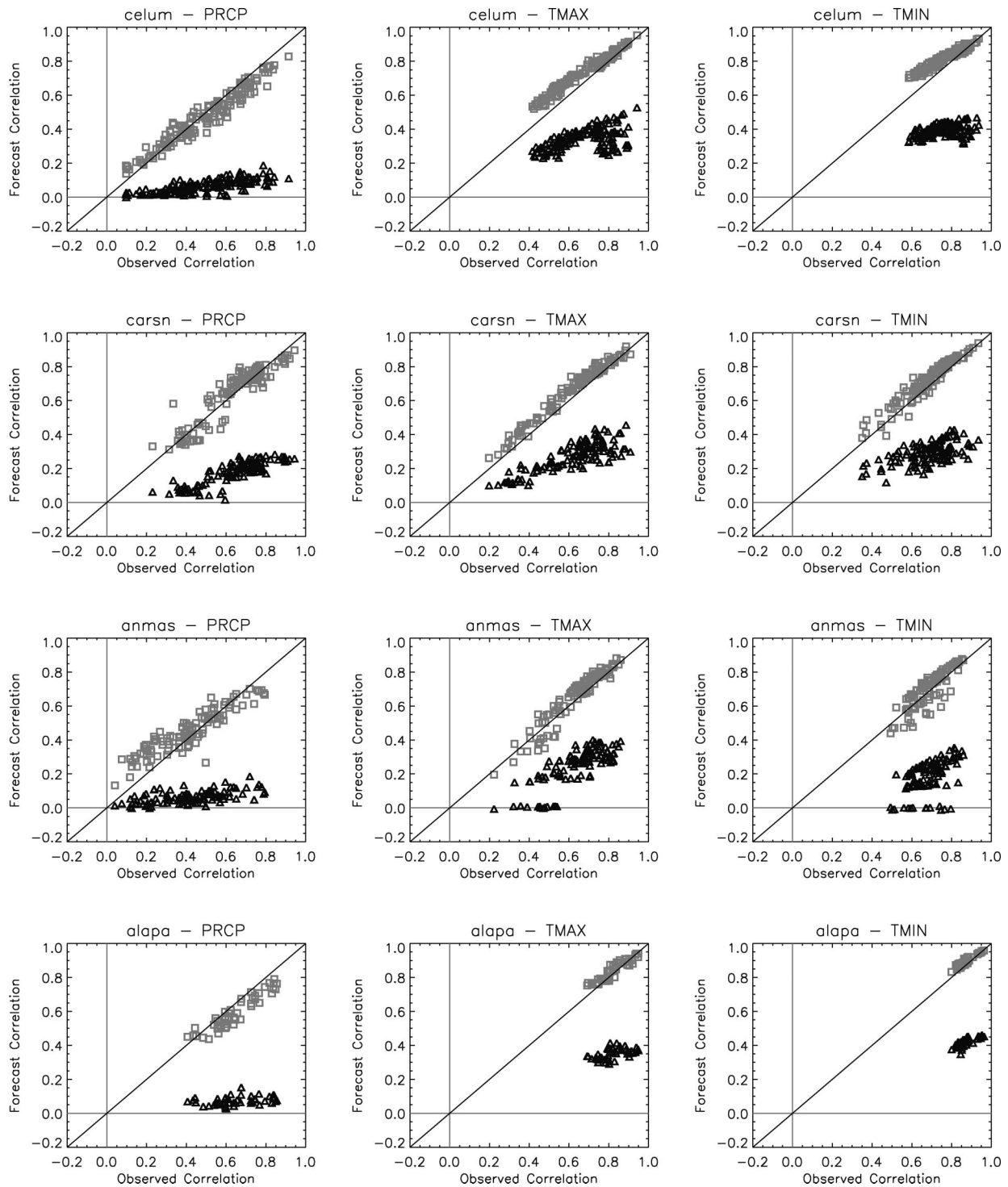
FIG. 7. The observed and forecasted intersite correlations for Jan for all station combinations in each of the four study basins, for both raw ensemble output (black triangles) and reordered ensemble output (gray squares). Results are shown for the forecast lead time of 5 days.

The comparison between the observed intervariable correlations and forecast intervariable correlations is shown for both shuffled and unshuffled ensemble output in Fig. 9. The presentation is similar to that in Figs. 7 and 8—the median intervariable correlation from the ensemble output for a given station is plotted against the observed intervariable correlation for that station. The dark (light) symbols depict unshuffled (shuffled)
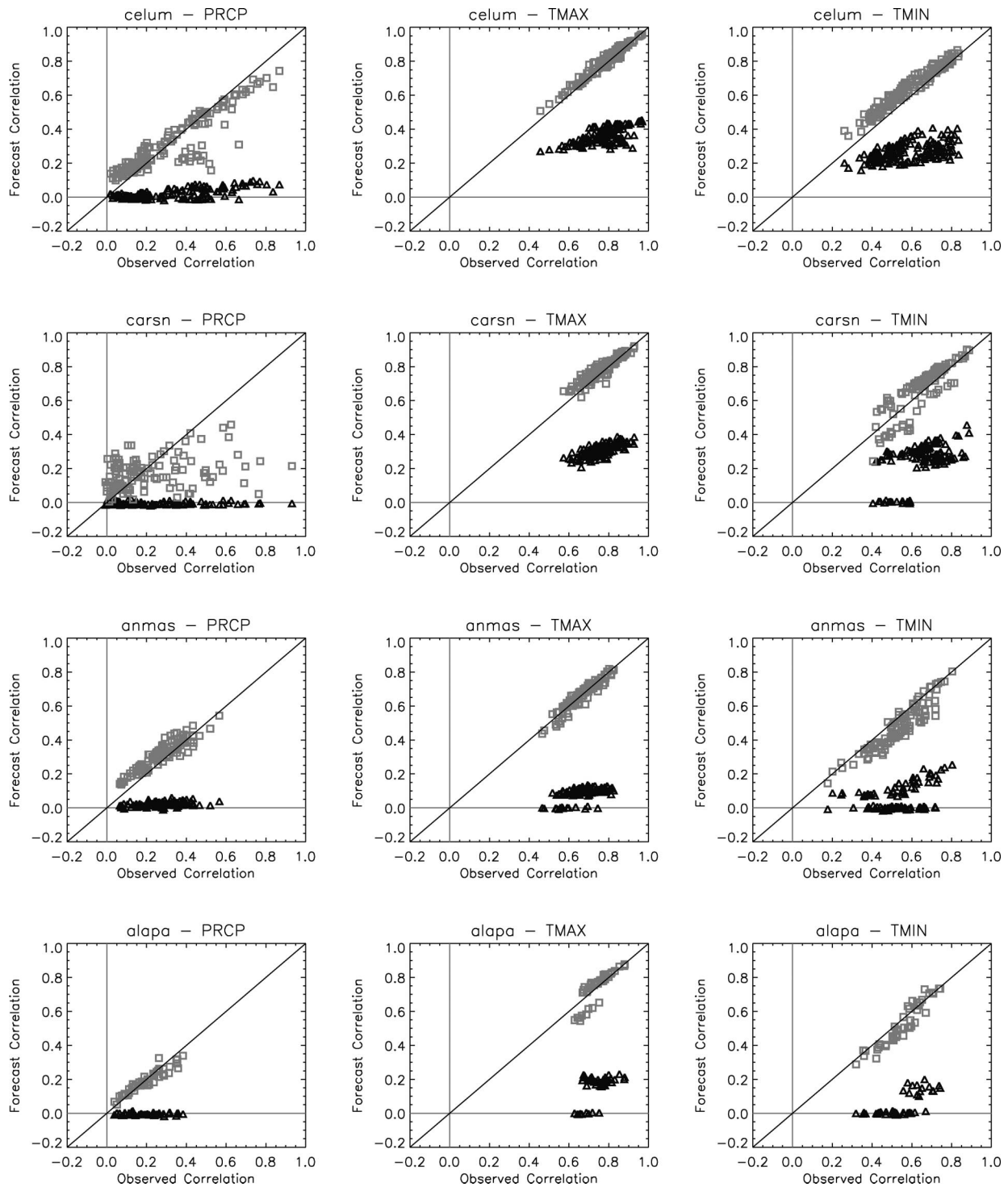
FIG. 8. Same as Fig. 7, except for Jul.

correlations, and the triangles (squares) depict results for the months of January (July). As expected from section 3, the observed intervariable correlations are preserved very well. An important feature in Fig. 9 is the negative correlations between precipitation and maximum temperature, most pronounced in July. This reflects lower temperatures on precipitation days, and although this feature is not present in the unshuffled output, it is reconstructed once the ensemble output has been shuffled. Another important feature in Fig. 9 is the
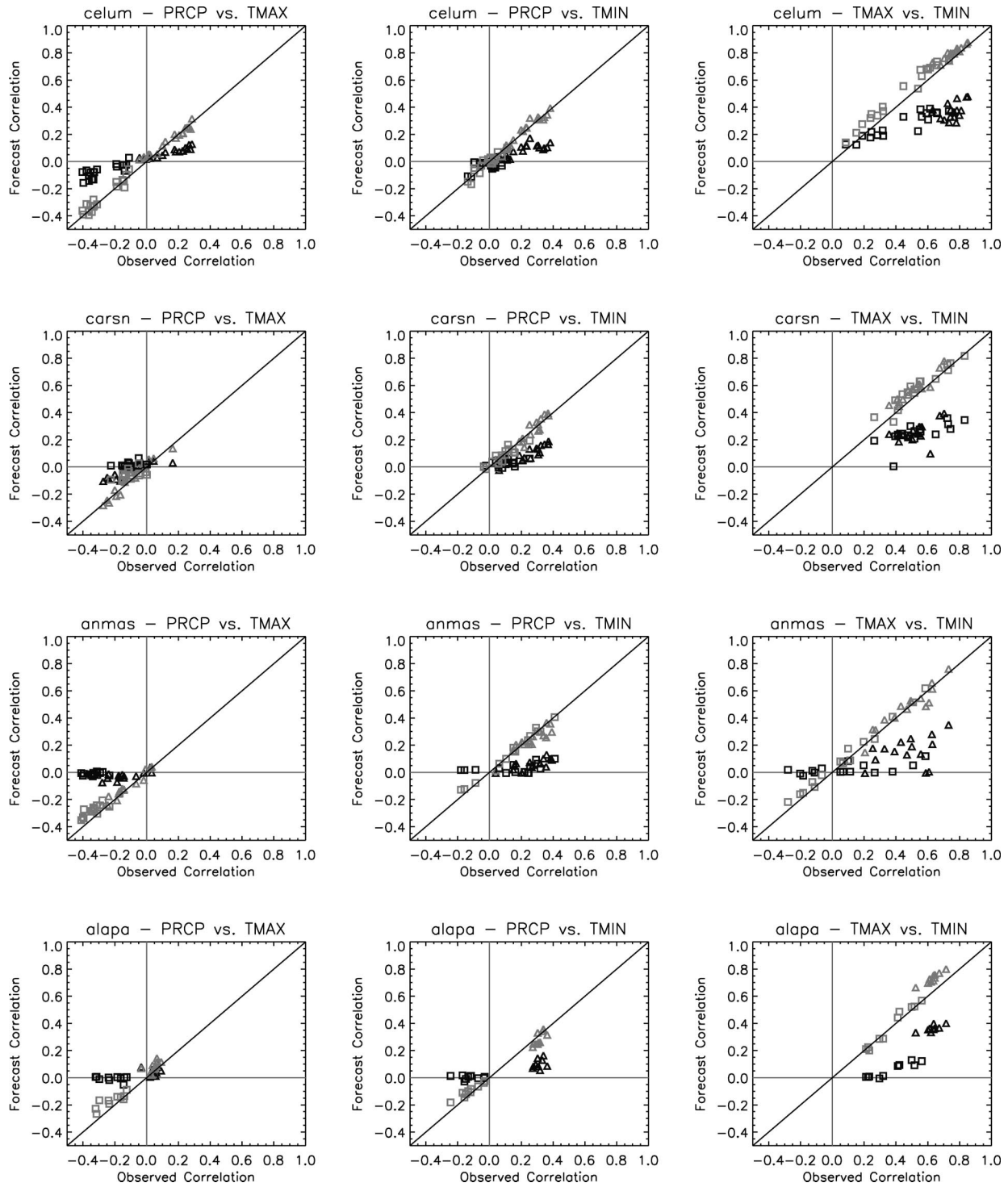
FIG. 9. The observed and forecasted intervariable correlations for all stations in each of the four study basins, shown for the forecast lead time of 5 days. The dark symbols depict results for unshuffled output, and the light symbols depict results for the shuffled output. For both unshuffled and shuffled output, diamonds depict results for the month of Jan and squares depict results for the month of Jul.
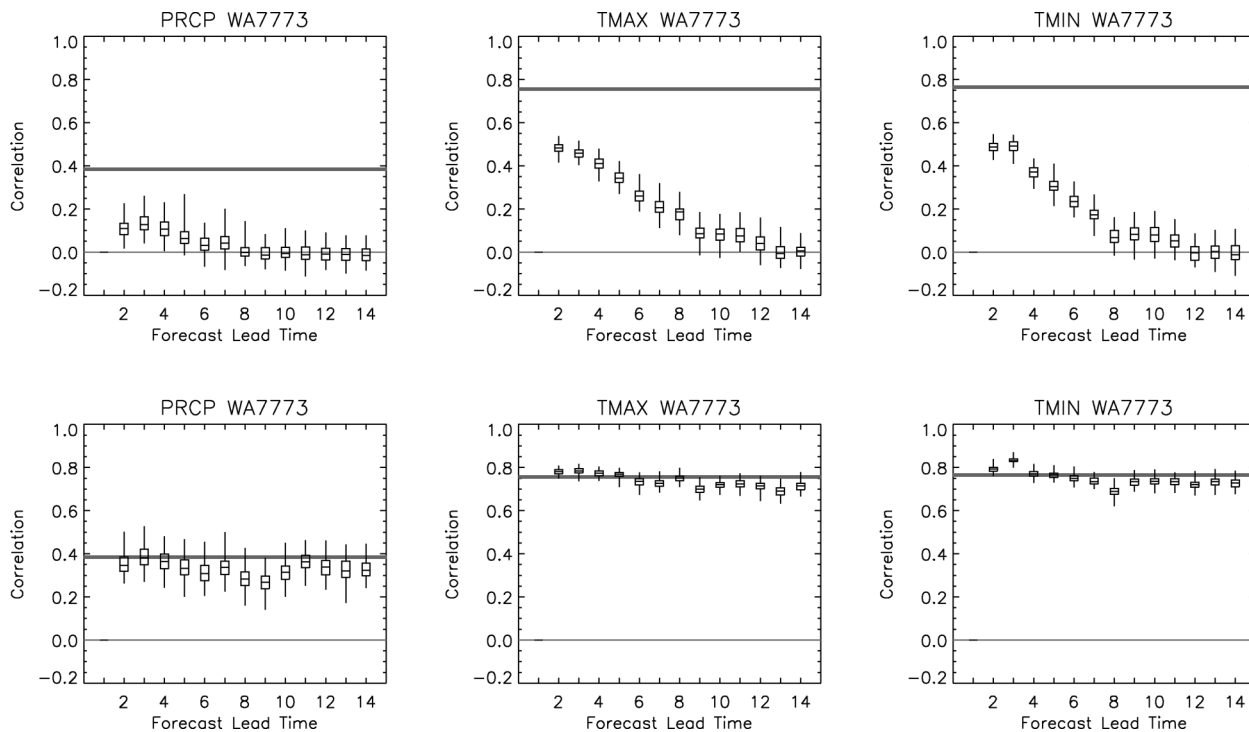
FIG. 10. The lag−1 correlation for an example station in the Cle Elum basin in Jan for (top) raw ensemble output and (bottom) reordered ensemble output. The thick line depicts the observed lag−1 correlation, and the box and whiskers depict the minimum, lower-quartile, median, upper-quartile, and maximum correlations of all ensemble members.

high correlation between maximum and minimum temperature, most pronounced in January. The observed correlation between maximum and minimum temperature is underestimated in the unshuffled output, but is reconstructed once the ensemble output is shuffled.

e. Temporal persistence

Accurate reproduction of the temporal persistence of precipitation and temperature time series also is an important consideration in forecasting streamflow. For example, if there is no temporal persistence in time series of temperature, warm days will be immediately followed by moderate-to-cold days, and any snowmelt will be abruptly halted. This scenario is in contrast to a situation closer to reality wherein warm days persist for several days, thereby creating rapid and persistent snowmelt. The Schaake Shuffle method should also be capable of reconstructing the temporal persistence in observed precipitation and temperature fields (see section 3b).

Figure 10 illustrates the temporal persistence (lag−1 correlation) from shuffled and unshuffled ensemble output for one example station in the Cle Elum River basin. Similar to Fig. 6, the top plots show temporal correlations for precipitation and maximum and minimum temperature for the unshuffled forecast ensemble, and the bottom plots show the temporal correlations for the shuffled forecast ensemble. In this example, day+2 forecasts are correlated against day+1 forecasts, day+3 forecasts are cor-

related against day+2 forecasts, and so forth. Results show similar patterns to those exhibited in Fig. 5. The unshuffled output preserves some (but by no means all) of the observed temporal persistence at short forecast lead times when there is skill in the forecasts. None of the temporal persistence is captured at longer forecast lead times when there is no forecast skill. By contrast, the shuffled ensemble output preserves the observed temporal persistence for all forecast lead times.

Extending these results, Fig. 11 presents lag−1 correlation statistics for all stations in each study basin. Results are presented in a similar manner to those in Figs. 7–9, where the median lag−1 correlation from the ensemble output for a given station is plotted against the observed lag−1 correlation for that station. Figure 11 demonstrates that the shuffling methodology is capable of reconstructing the observed temporal persistence for all stations and variables. In all cases the temporal persistence in the shuffled ensemble output is closer to observations than the unshuffled output.

Another feature important for hydrologic applications is the intermittency of precipitation, which is not described well by the lag−1 correlation plots. To address this issue, we calculate the probability that a dry day follows a wet day (P_{wd}), and the probability that a wet day follows a dry day (P_{dw}), for each day in the forecast cycle. These probabilities are termed "transition probabilities" and are widely used to construct Markov-type models of precipitation occurrence (e.g., Gabriel and
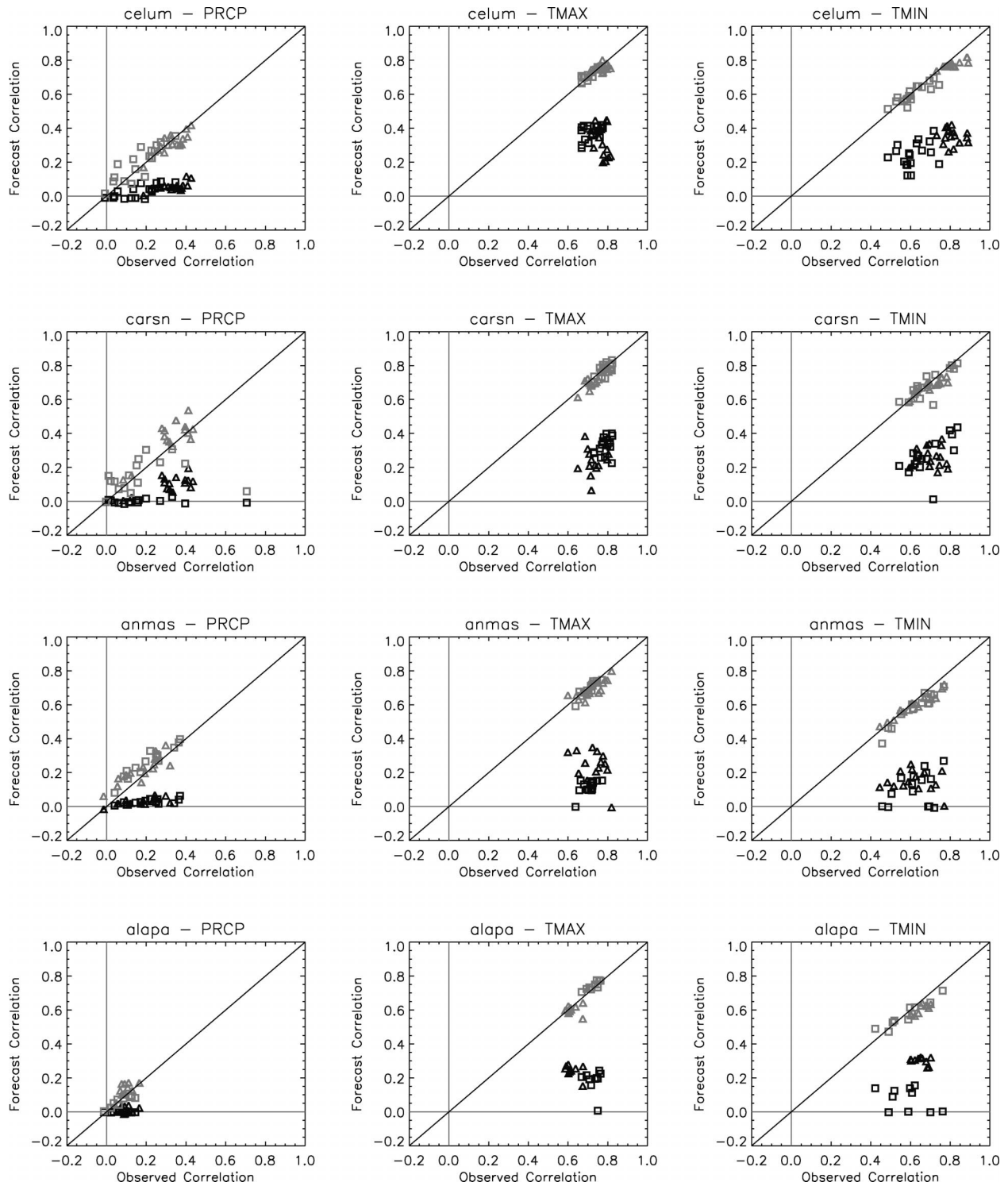
FIG. 11. The observed and forecasted lag−1 correlation for all stations in each of the four study basins, shown for the forecast lead time of 5 days. The dark symbols depict results for unshuffled output, and the light symbols depict results for the shuffled output. For both unshuffled and shuffled output, diamonds depict results for the month of Jan and squares depict results for the month of Jul.
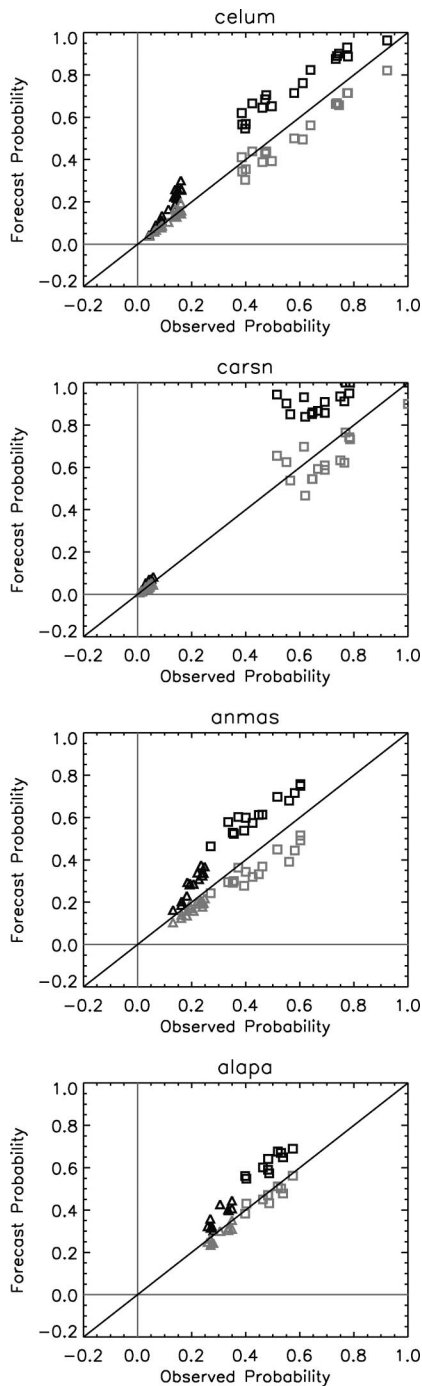
FIG. 12. The observed and forecasted transition probabilities for the month of Jul for the forecast lead time of 5 days. The dark symbols depict results for unshuffled output, and the light symbols depict results for the shuffled output. The probability that a dry day follows a wet day is illustrated with squares, and the probability that a wet day follows a dry day is illustrated with triangles.

Neumann 1962; Wilks 1998; Wilks and Wilby 1999). The comparison between observed and forecasted transition probabilities is presented in Fig. 12 for each station in the four basins. Results are presented only for

the month of July. The triangles (squares) illustrate the case for P_{dw} (P_{wd}), and, as in other figures, the dark (light) symbols depict unshuffled (shuffled) results. Note that if P_{dw} is similar to the climatological probability of precipitation occurrence (or if P_{wd} is similar to the probability of no precipitation), then the unshuffled ensemble output will mirror the transition probabilities quite well. The results in Fig. 12 show fairly small differences between the observed and forecasted transition probabilities for both unshuffled and shuffled ensemble output, but, on the whole, the shuffled transition probabilities are closer to the observed transition probabilities for all four basins.

5. Summary and conclusions

A method for reconstructing the space–time variability in forecasted precipitation and temperature fields is presented. In this approach, the ensemble output for a given forecast is ranked and matched with the rank of precipitation and temperature data from dates randomly selected from the historical record. The ensembles are then reordered to correspond to the original order of the selection of historical data.

To assess the value of this method, the space–time variability in local-scale ensemble forecasts of precipitation and temperature is examined in four river basins in the United States. Forecasts are generated using regression-based statistical downscaling of medium-range numerical weather prediction model output. Forecasts are essentially unbiased and have useful probabilistic forecast skill for all variables at short forecast lead times. The reconstruction methodology preserves the spatial correlation between all station pairs for all variables in each of the study basins. In some instances, the spatial correlation for precipitation is slightly underestimated due to the intermittency properties of precipitation time series, but the spatial correlations from the reconstruction methodology are closer to observed correlations than the raw ensemble output for all station pairs. Similarly, the reconstruction methodology preserves the observed temporal persistence in precipitation and temperature time series for all station pairs.

The main caveat with this approach is that it assumes stationarity in the spatiotemporal correlation structure. Put differently, the approach assumes that the spatial correlation from randomly selected dates in the historical record will be valid for forecasts in the future. This point was not addressed in the present paper, as all results (correlations) are computed using all years in the forecast archive (1979–98). As presented here, the Schaake shuffle approach will not preserve the spatial gradients in precipitation and temperature fields for individual forecasts. An extension to this climatological ensemble reordering approach is to preferentially select dates from the historical record that resemble forecasted atmospheric conditions and use the spatial correlation structure from this subset of dates to reconstruct the spatial variability for a specific forecast. This type of analog approach may

be capable of preserving spatial gradients for individual forecasts, at least for short forecast lead times. Further research is warranted on this topic.

It is of course possible to use statistical methods that preserve the space–time variability of downscaled fields intrinsically. One example of this type is from Wilby et al. (2003), where regression methods are applied to predict regionally averaged precipitation and spatially disaggregated by simply using data from individual sites for days that are similar to the regionally averaged predictions. In another approach, Charles et al. (1999) apply a nonhomogeneous hidden Markov model to simulate distinct patterns of multisite precipitation and amounts conditional on a set of synoptic-scale atmospheric variables. Both of these approaches preserve the observed intersite correlations very well. The value of the Schaake shuffle, however, is that it is applied as a postprocessing step. It is possible to use any statistical downscaling method (or combination of methods) that provide the best ensemble of predictions at individual stations, and then reconstruct the space–time variability after the fact.

The reordering methodology is a generic approach that can be applied in many different settings. As an example, consider ensemble simulations of streamflow generated using the multiobjective optimization algorithm (MOCOM-UA; Yapo et al. 1998) or the generalized likelihood uncertainty estimation (GLUE) methodology (Beven and Binley 1992), implemented for many subbasins within a large watershed. Now, if the models in each subbasin are calibrated separately, "ensemble 1" from a given subbasin will not necessarily correspond to "ensemble 1" from the neighboring subbasin. If these ensembles are combined to produce simulations of runoff for the entire watershed, then a canceling effect will ensue and the streamflow simulations for the watershed will underestimate both low and high extreme values. Some reordering of ensemble output will be required to circumvent this problem.

APPENDIX

Pseudo-FORTRAN Code to Perform the Schaake Shuffle

The *sort2* subroutine sorts two real vectors using the quick sort algorithm and can be used directly from Press et al. (1992).

```
BEGIN CODE
c       Variable declarations
            integer nens                        ! total number of ensemble members; equals length of observed data
            real b(nens)                        ! counter vectors; needs to be real for subroutine sort2
            real x(nens)                        ! unshuffled ensemble/model output of length nens
            real y(nens)                        ! observed data of length nens
            integer iens                        ! iteration variable
            integer ierr                        ! error code, 0 = no error; 1 = error
c
c       Create vectors, a and b, and read vectors x and y
        do iens = 1, nens
           b(iens) = float(iens)
           x(iens) = read unshuffled ensemble output
           y(iens) = read observed data
c       perform error check to ensure x(iens) and y(iens) have no missing value
           if ((x(iens) is missing).or.(y(iens) is missing)) ierr=1
        enddo !iens
c
c       Perform Schaake Shuffle
        if (ierr.eq.0) then
           call sort(nens, x)                   ! sort the ensemble output
           call sort2(nens, y, b)               ! sort the observed data
           call sort2(nens, b, x)               ! shuffle
        endif
END CODE
```

REFERENCES

Agterberg, F. P., 1989: Logdia—Fortran 77 program for logistic regression with diagnostics. *Comput. Geosci., 15,* 599–614.

Antolik, M. S., 2000: An overview of the National Weather Service's centralized statistical quantitative precipitation forecasts. *J. Hydrol., 239,* 306–337.

Beven, K., and A. Binley, 1992: The future of distributed models—Model calibration and uncertainty prediction. *Hydrol. Processes, 6,* 279–298.

Charles, S. P., B. C. Bates, and J. C. Hughes, 1999: A spatio-temporal model for downscaling precipitation occurrence and amounts. *J. Geophys. Res.,* **104,** 31 657–31 669.

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

Crane, R. G., and B. C. Hewitson, 1998: Doubled CO_2 precipitation changes for the Susquehanna basin: Downscaling from the Genesis GCM. *Int. J. Climatol.,* **18,** 65–76.

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. Meteor.,* **39,** 1580–1591.

Gabriel, K. R., and J. Neumann, 1962: A Markov chain model for daily rainfall occurrence at Tel Aviv. *Quart. J. Roy. Meteor. Soc.,* **88,** 90–95.

Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.,* **11,** 1203–1211.

Hay, L. E., M. P. Clark, R. L. Wilby, W. J. Gutowski, R. W. Arritt, E. S. Takle, Z. Pan, and G. H. Leavesley, 2002: Use of regional climate model output for hydrologic simulations. *J. Hydrometeor.,* **3,** 571–590.

Murphy, J., 1999: An evaluation of statistical and dynamical techniques for downscaling local climate. *J. Climate,* **12,** 1885–1908.

Panofsky, H. A., and G. W. Brier, 1963: *Some Applications of Statistics to Meteorology.* Mineral Industries Continuing Education, College of Mineral Industries, The Pennsylvania State University, 224 pp.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, 1992: *Numerical Recipes in Fortran.* 2d ed. Cambridge University Press, 963 pp.

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. Amer. Meteor. Soc.,* **73,** 753–762.

Wilby, R. L., C. W. Dawson, and E. M. Barrow, 2002: SDSM—A decision support tool for the assessment of regional climate impacts. *Environ. Modell. Software,* **17,** 145–157.

——, O. J. Tomlinson, and C. W. Dawson, 2003: Multi-site simulation of precipitation by conditional resampling. *Climate Res.,* **23,** 183–199.

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

——, 1998: Multi-site generalizations of a daily stochastic weather generation model. *J. Hydrol.,* **210,** 178–191.

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

Yapo, P. O., H. V. Gupta, and S. Sorooshian, 1998: Multi-objective global optimization for hydrologic models. *J. Hydrol.,* **204,** 83–97.