A New Method for Generating Stochastic Simulations of Daily Air Temperature for Use in Weather Generators

KIMBERLY SMITH AND COURTENAY STRONG

Department of Atmospheric Sciences, University of Utah, Salt Lake City, Utah

FIRAS RASSOUL-AGHA

Department of Mathematics, University of Utah, Salt Lake City, Utah

(Manuscript received 21 March 2016, in final form 24 December 2016)

ABSTRACT

A stochastic harmonic autoregressive parametric (SHArP) weather generator is presented that simulates trended, nonstationary temperature values directly, circumventing the conventional approach of adding simulated standardized anomalies of temperature to a prescribed cyclostationary mean. The model mean makes autocorrelated transitions between wet- and dry-state values, and its parameters are determined by optimizing harmonic and trend terms. The precipitation-responsive autocorrelated transitions yield more realistic temperature behavior during frontal passage in comparison with prior models that switch abruptly between wet- and dry-state means. If the stochastic (noise) term is assumed to have constant amplitude, analytical results are available via maximum likelihood estimation (MLE) and are equivalent to least squares estimation (LSE). Where observations motivate a seasonally varying noise coefficient, MLE becomes nonlinear, and an analytical solution is formulated via LSE. For illustration, SHArP is shown to produce realistic representations of daily maximum air temperature at a single site, which for the study is the Salt Lake City International Airport (KSLC). SHArP reduces the temperature bias following frontal passages by over 2°C in three seasons. A method for generalizing the model to multiple variables at multiple sites is discussed.

1. Introduction

The drought-stricken western United States, including the Great Basin region of Utah, Wyoming, Idaho, Oregon, Nevada, and California, is facing an uncertain water future because of climate change. The northern half of the Great Basin, which includes northern Utah, is located in the center of the El Niño–Southern Oscillation (ENSO) dipole. ENSO is a well-known climatic teleconnection between sea surface temperatures and the atmosphere in the equatorial Pacific Ocean that affects global weather patterns (Troup 1965; Horel and Wallace 1981). The occurrence of precipitation in the Great Basin in any given year is dependent on both the phase of ENSO and the phase of the Pacific decadal oscillation (PDO), as the phase of the PDO shifts the ENSO dipole either north or south (Wise 2010; Brown 2011). Because of its complex terrain, the majority of the water used by those who live in the region is dependent on the snowpack that is stored in the mountains and released throughout the year via the reservoir system. This semiarid region is already experiencing inconsistent water availability throughout any given year because of the drastically different number of winter precipitation events from year to year. The ability to statistically model the occurrence of precipitation and air temperature is imperative to better forecast potential changes in future water availability as the climate changes. In this study, we introduce a stochastic harmonic autoregressive parametric (SHArP) weather generator, which statistically models meteorological variables (in this case, the occurrence and amount of precipitation and maximum air temperature). The model can be used to investigate how the future of the Great Basin may be impacted by climate change and to understand the meteorological extremes that are likely to play a part in that impact.

While the outputs of both statistically based stochastic weather generators (SWGs) and dynamically based

Corresponding author e-mail: Courtenay Strong, court.strong@utah.edu

DOI: 10.1175/JAMC-D-16-0122.1

global climate models (GCMs) are used in climate impacts studies, there are major differences between them. SWGs work on a point scale, or on a point scale expanded via multisite generalization to a basin scale, whereas GCMs work on a broad regional scale and can be downscaled to the basin or smaller scale. GCMs have difficulty capturing detail in areas of complex terrain, including the Great Basin, which is characterized by its basin-and-range topography (e.g., Thompson and Burke 1974). SWGs also have a faster computational time than GCMs, which can take upward of months to complete a single run. GCMs are very computationally expensive in comparison with SWGs, and thus there are not many GCM runs available for analysis. GCMs also have difficulty capturing the very low-frequency (century scale) connections between the Pacific Ocean and the Great Basin. The performance of state-of-the-art GCMs has been evaluated in terms of ability to capture the ''extremes'' in precipitation and temperature, and it has been found that GCMs poorly capture the extremes, though they perform better at temperature extremes than precipitation (Kiktev et al. 2007).

SWGs alleviate some limitations of GCMs and were introduced as a way to overcome a lack of observational meteorological data and problems associated with missing data both temporally and spatially (Wilks and Wilby 1999; Wilks 2008). In addition, they have been used to better understand the uncertainties associated with future climate (e.g., Wilks 1992; Forsythe et al. 2014). These statistical models generate synthetic time series of precipitation and in some cases also air temperature and solar radiation, which statistically resemble the data used to force the model—usually daily observational weather data (Wilks and Wilby 1999). There have been a multitude of early studies on SWGs that solely generate precipitation occurrence and amount because air temperature and other meteorological variables are affected by whether precipitation occurred.

The first studies using stochastic simulators of weather data employed two-state, first-order Markov chain frameworks regarding precipitation (Bailey 1964; Richardson 1981; Roldàn and Woolhiser 1982), meaning that the probability of precipitation occurrence on a given day is only dependent on whether precipitation occurred on the previous day. Precipitation amount was modeled separately, and maximum/minimum temperatures and solar radiation were modeled as a function of precipitation occurrence. Other studies involving SWGs considered a two-state, second-order Markov chain process (Stern and Coe 1984; Wilks 1999a). Markov chains of higher order have been found to better capture dry spells than first-order Markov chains, thus providing more accurate results for most areas of the western United States where dry spells are common, such as the semiarid Great Basin.

One limitation of the common SWGs is the ability to successfully capture nonstationary variability. Previous studies have found that over the western United States, El Niño results in a wetter Southwest and a drier Northwest, while La Niña results in the opposite (Ropelewski and Halpert 1986; Dettinger et al. 1998; Woolhiser 2008). In addition, the PDO also has significant impacts on precipitation in the western United States. The PDO is linked to ENSO, which in turn affects how the different phases of ENSO will impact the western United States (Gershunov and Barnett 1998; Gershunov et al. 1999; Mauget 2003). Woolhiser (2008) introduced the idea of adding nonstationarity to the stochastic framework in order to capture the effects these major oceanic oscillations have on western U.S. precipitation. Essentially, perturbations given as time series of the oscillations were linearly added to the probability of precipitation, and the coefficients associated with each perturbation give information on the sensitivity of each of the oscillations (Woolhiser 2008). We employ this method in this study and also include a trend to account for the changing climate.

In the SWG literature, simulation of daily maximum and minimum air temperature is usually conditioned on whether the day is wet or dry. The most widely used method for simulating temperature is the method used by Richardson (1981). This method involves generating the standardized residual time series of temperature (maximum and minimum temperature; the study also included solar radiation) and using the multivariate generation model as described by Matalas (1967). These standardized residuals are assumed normally distributed, and the coefficients in the generating model are matrices containing the cross correlations and autocorrelations between the residuals (Matalas 1967). After generating the synthetic residuals, the wet- or dry-state means and standard deviations that were initially removed are reintroduced to yield daily values of the variables. The means and standard deviations depend on whether the day was wet or dry; they are assumed to be cyclostationary and are determined by fitting harmonics of the annual cycle to observations (Richardson 1981).

In addition to the common parametric SWGs described thus far, including the SWGs introduced by Matalas (1967) and Richardson (1981), recent studies have employed nonparametric SWGs and generalized linear models (GLMs). These SWGs are data driven and involve either kernel density estimation (e.g., Rajagopalan et al. 1997; Harrold et al. 2003) or resampling via k-nearest-neighbor (k-NN) bootstrapping (e.g., Rajagopalan and Lall 1999; Caraway et al. 2014). These models do not rely on the statistical relationships applied in the parametric SWGs. They offer an alternative to the standard linear models presented in the parametric SWGs, which are unable to capture the nonlinear relationships between meteorological variables. The use of GLMs in SWGs, first introduced by Stern and Coe (1984), has also been increasing in popularity because they can easily model discrete variables and variables with nonnormal distributions (Furrer and Katz 2007). In addition, GLMs are especially useful tools because of their ability to treat ENSO and other major oceanic modes of variability as continuous variables (e.g., Chandler 2005). More details behind GLMs can be found in McCullagh and Nelder (1989).

A limitation of the widely used Richardson model is that its mean and standard deviation switch abruptly between wet- and dry-state values prescribed in advance of the simulation, and temperature is not simulated directly but rather through its residuals. This method inaccurately captures what occurs in reality, which instead are smooth, autocorrelated transitions between wet- and dry-state values. In this study, we introduce the mathematics and present illustrative results for a SHArP weather generator that is based on the Richardson model but that simulates temperature values directly with a mean that makes autocorrelated transitions between wet-and dry-state temperature values. Because of this innovation, the method described here better captures the temperature transitions between days with different precipitation states, including following frontal passages.

2. Data and study area

We chose to illustrate the SHArP weather generator using observations from the Salt Lake City International Airport (KSLC), which is located in the Great Basin. Its precipitation depends largely on a combination of the state of ENSO and the state of the PDO (Wise 2010; Brown 2011). The precipitation and temperature data used to force SHArP are daily observational data recorded at KSLC (40.78°N, 111.97°W) from 1 January 1948 to 31 December 2010 via the Global Historical Climatology Network (GHCN-Daily) provided by the National Centers for Environmental Information (obtained from http://www.ncdc.noaa.gov; Menne et al. 2012a,b). In addition, we obtained GHCN-Daily precipitation and temperature data for four climatologically similar surrounding sites to illustrate the autocorrelated transitions during frontal passages. The domain map (see Fig. 1) shows the location of KSLC in addition to the four surrounding sites: Boise Air Terminal (KBOI) and Pocatello Regional Airport (KPIH) in Idaho, Elko Regional Airport (KEKO) in Nevada, and Grand Junction Regional Airport (KGJT) in Colorado.

Future precipitation and temperature output used to force SHArP are daily 0.125° gridded bias correction constructed analog (BCCA) projections from the CCSM4 model, which was part of phase 5 of the Coupled Model Intercomparison Project (CMIP5) multimodel ensemble (Maurer et al. 2007; Brekke et al. 2013). We use the high-emissions scenario (RCP 8.5) data, and they span from 1 January 2006 to 31 December 2100. We use the data starting from 1 January 2011 following the end of the observational data.

A day was considered "wet" and given value $\chi = 1$ if the total precipitation on that day reached at least 0.25 mm (approximately 0.01 in.), corresponding to the minimum depth recorded by rain gauges. Otherwise, the day was considered dry and given value $\chi = 0$. The χ vector was determined from the precipitation time series, and this provided the precipitation occurrence needed to model temperature with SHArP. In this study, we use and generate only maximum surface air temperature at a single site. Generalization to multiple variables at multiple sites has been completed, and the formulation will be presented in a future paper.

3. Simulation of maximum air temperature and precipitation

The method introduced here is based on the Richardson (1981) method described in the introduction. The Richardson method is a linear equation given by

$$\boldsymbol{\chi}_{p,i}(j) = \mathbf{A}_R\boldsymbol{\chi}_{p,i-1} + \mathbf{B}_R\boldsymbol{\epsilon}_{p,i}(j),$$

where $\boldsymbol{\chi}_{p,i}(j)$ is a 3×1 matrix containing the residuals for day i of year p and $\boldsymbol{\chi}_{p,i-1}(j)$ is a 3×1 matrix containing the residuals for day $i-1$ of year p; j refers to the variable of interest (Richardson simulated three: maximum temperature, minimum temperature, and solar radiation). The $\boldsymbol{\epsilon}_{p,i}(j)$ is a 3×1 matrix of normally distributed, independent random noise with 0 mean and a variance of 1. The $\mathbf{A}_R$ and $\mathbf{B}_R$ are 3×3 matrices that contain the correct serial and cross-correlation coefficients (subscript R refers to the Richardson method). The mean and standard deviation of the variables are removed, and the residuals are simulated. The model makes abrupt switches between wet- and dry-state values because of the prescribed means and standard deviations prior to simulation, which are then used to determine the true values after simulation.

SHArP is based on the observation shown below that temperature makes autocorrelated transitions between wet- and dry-state means with characteristic annual cycles, while subject to random fluctuations associated with frontal passages. For maximum temperature at a single site, the linear model is

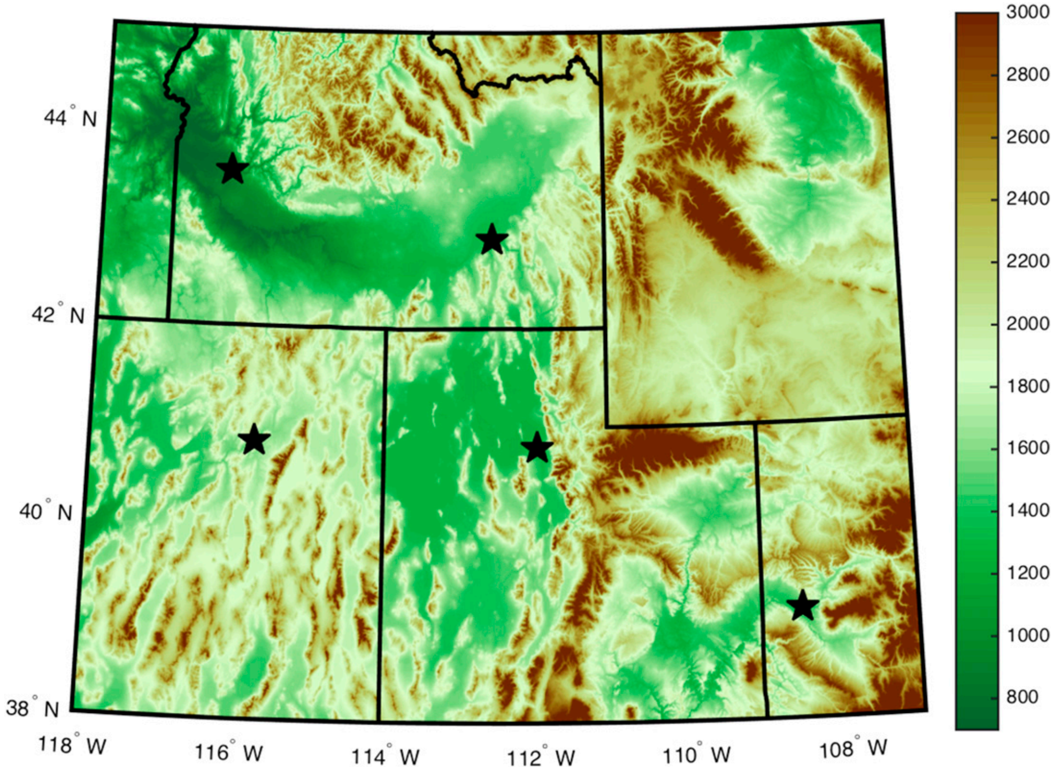

FIG. 1. The study area: the eastern half of the Great Basin (which includes northern and western Utah, extreme southwestern Wyoming, extreme southern Idaho, and Nevada) and the surrounding area. The stars indicate the location of KSLC and surrounding sites: KBOI, KPIH, KEKO, and KGJT. The color bar indicates elevation in meters above sea level.

$$T_{k+1} = aT_k + b_k + c_k\epsilon_k, \tag{1}$$

where a is a coefficient that is assumed to be constant and b_k and c_k are coefficients that depend on day k. The b_k coefficient captures the mean, annual cycle, and trend. Errors ϵ_k are independent and identically distributed (i.i.d.) random standard normals. The temperature on day $k+1$ is dependent on the temperature on day k, where k ranges from 0 to $K-1$ (K being the length of the simulation). We begin the simulations by taking the first temperature value from the training data as T_0, but this could also be drawn from an appropriate distribution.

a. Maximum likelihood estimation

We begin with a simplified case where c does not depend on k. We assume the temperature entries from T_1 to T_K are multivariate normals, and the joint density function is given by

$$f(T_1, \ldots, T_K) = \frac{1}{(2\pi)^{K/2}c}\exp\frac{-(\mathbf{DT}-\mathbf{B})'(\mathbf{DT}-\mathbf{B})}{2c^2}, \tag{2}$$

where $\mathbf{D}$ is the $K \times K$ matrix

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ -a & 1 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 & 0 \\ 0 & \cdots & 0 & -a & 1 \end{bmatrix},$$

and $\mathbf{B}$ and $\mathbf{T}$ are the $K \times 1$ vectors

$$\mathbf{B} = \begin{bmatrix} aT_0 + b_0 \\ b_1 \\ \vdots \\ b_{K-2} \\ b_{K-1} \end{bmatrix} \quad \text{and} \quad \mathbf{T} = \begin{bmatrix} T_1 \\ T_2 \\ \vdots \\ T_{K-1} \\ T_K \end{bmatrix}.$$

The mean is given by $\mathbf{D}^{-1}\mathbf{B}$, and the dry and wet day means are shown with their corresponding composite annual cycles from the KSLC training data in Fig. 2. Note the higher variability associated with wet days versus dry days. To restrict the model to a reasonable number of parameters, we give structure to the b_k values by giving them a trend and harmonics:

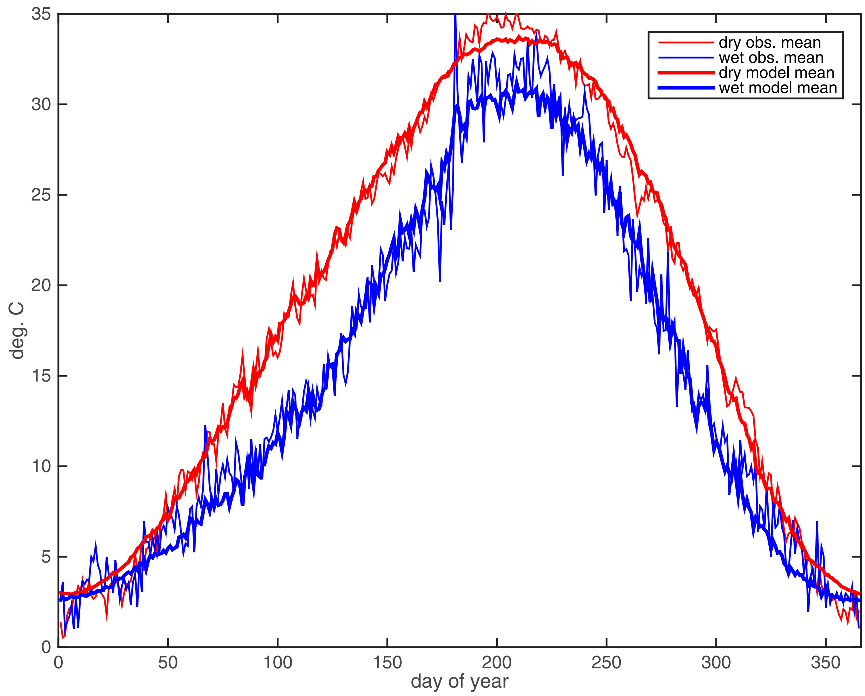

FIG. 2. Annual composite of the observation (thin lines) and model (thick lines) means for dry (red) and wet days (blue). Results are based on KSLC observations for years 1948–2010.

$$b_k = \gamma_{\chi_{k+1}} + \alpha k + \beta_{\chi_{k+1}} \cos(2\pi k/\tau) + \beta'_{\chi_{k+1}} \sin(2\pi k/\tau) + \delta_{\chi_{k+1}} \cos(4\pi k/\tau) + \delta'_{\chi_{k+1}} \sin(4\pi k/\tau), \quad (3)$$

where τ is the period, assumed to be 365 days. We include two harmonics to illustrate how b_k can be generalized to include any number of harmonics. A log-likelihood ratio test can be performed to determine statistical significance of additional harmonics.

We applied a maximum likelihood estimate (MLE) to the joint density function, which involves maximizing (2) or minimizing its negative log:

$$c^{-2}(\mathbf{DT} - \mathbf{B})'(\mathbf{DT} - \mathbf{B}) + 2K \log c. \quad (4)$$

We first minimize $(\mathbf{DT} - \mathbf{B})'(\mathbf{DT} - \mathbf{B})$ to get the MLEs for the $\mathbf{D}$ matrix and $\mathbf{B}$ vector. This returns the sum of squared errors:

$$(\mathbf{DT} - \mathbf{B})'(\mathbf{DT} - \mathbf{B}) = \sum_{k=0}^{K-1} (aT_k + b_k - T_{k+1})^2, \quad (5)$$

where b_k is given in (3).

Taking derivatives in (5) with respect to a and each of the parameters in b_k and setting them equal to zero gives the following 12 equations:

$$\sum_{k=0}^{K-1} T_k(aT_k + b_k - T_{k+1}) = 0,$$

$$\sum_{k=0}^{K-1} k(aT_k + b_k - T_{k+1}) = 0,$$

$$\sum_{k=0}^{K-1} (aT_k + b_k - T_{k+1})\mathbf{1}\{\chi_{k+1} = 0\} = 0,$$

$$\sum_{k=0}^{K-1} (aT_k + b_k - T_{k+1})\mathbf{1}\{\chi_{k+1} = 1\} = 0,$$

$$\sum_{k=0}^{K-1} \cos(2\pi k/\tau)(aT_k + b_k - T_{k+1})\mathbf{1}\{\chi_{k+1} = 0\} = 0,$$

$$\sum_{k=0}^{K-1} \cos(2\pi k/\tau)(aT_k + b_k - T_{k+1})\mathbf{1}\{\chi_{k+1} = 1\} = 0,$$

$$\sum_{k=0}^{K-1} \sin(2\pi k/\tau)(aT_k + b_k - T_{k+1})\mathbf{1}\{\chi_{k+1} = 0\} = 0,$$

$$\sum_{k=0}^{K-1} \sin(2\pi k/\tau)(aT_k + b_k - T_{k+1})\mathbf{1}\{\chi_{k+1} = 1\} = 0,$$

$$\sum_{k=0}^{K-1} \cos(4\pi k/\tau)(aT_k + b_k - T_{k+1})\mathbf{1}\{\chi_{k+1} = 0\} = 0,$$

$$\sum_{k=0}^{K-1} \cos(4\pi k/\tau)(aT_k + b_k - T_{k+1})\mathbf{1}\{\chi_{k+1} = 1\} = 0,$$

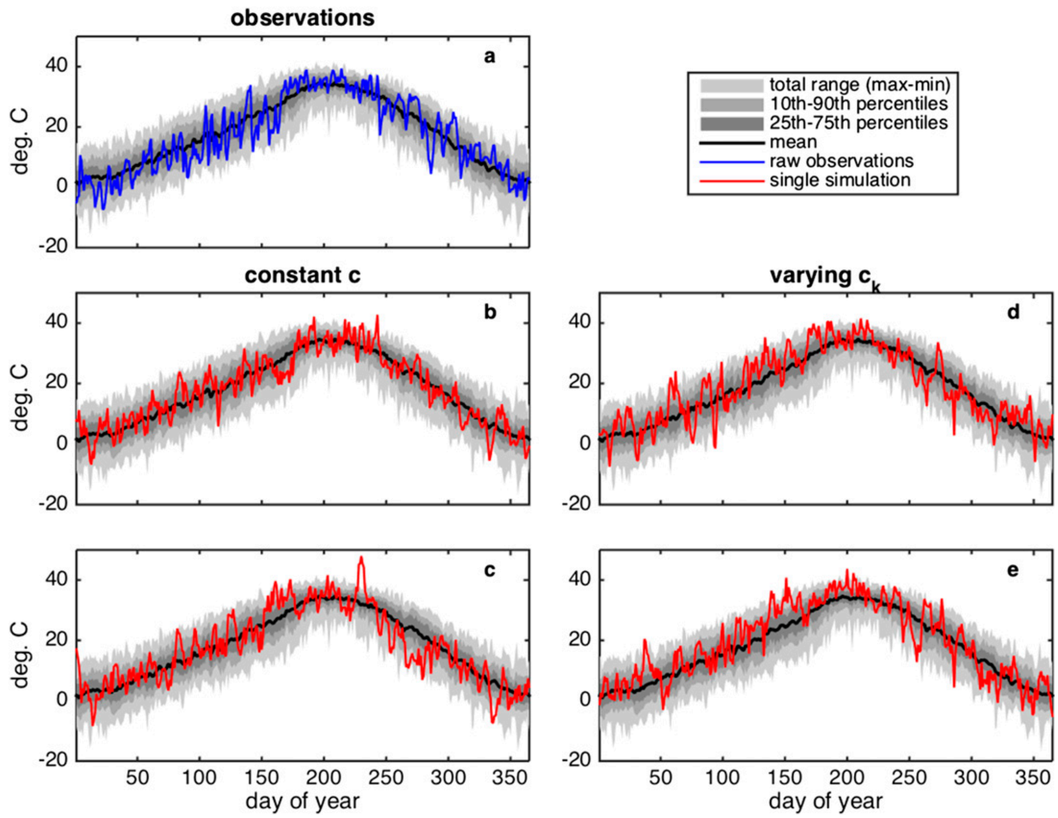

FIG. 3. Illustration of the SHArP weather generator with (a) input observational data for comparison. The blue curve (raw observations) shows 2008 as an example year, and shading in each panel corresponds to percentiles of the historical data for 1948–2010. Two simulations (red) of the temperature model with (b),(c) a constant c and (d),(e) a seasonally varying c_k.

$$\sum_{k=0}^{K-1} \sin(4\pi k/\tau)(aT_k + b_k - T_{k+1})\mathbf{1}\{\chi_{k+1} = 0\} = 0, \quad \text{and}$$

$$\sum_{k=0}^{K-1} \sin(4\pi k/\tau)(aT_k + b_k - T_{k+1})\mathbf{1}\{\chi_{k+1} = 1\} = 0,$$

where $\mathbf{1}$ is an indicator function that takes the value of 1 if the condition in brackets is met and 0 otherwise. This is a linear system of 12 equations and 12 unknowns, which we solve numerically.

We then minimize (4) as a function of c. Taking a derivative in c yields

$$-2c^{-3}(\mathbf{DT} - \mathbf{B})'(\mathbf{DT} - \mathbf{B}) + 2Kc^{-1}. \qquad (6)$$

The derivative has a unique point at which it vanishes:

$$c = \sqrt{K^{-1}(\mathbf{DT} - \mathbf{B})'(\mathbf{DT} - \mathbf{B})}, \qquad (7)$$

which is both the MLE value and least squares estimation (LSE) value. However, constant c tends to overestimate the variance in the summer and underestimate it in the winter (see Figs. 3b,c), motivating a seasonally varying c denoted by c_k, as in (1). The seasonally varying c_k makes the MLE nonlinear in the parameters, so we proceed by taking an LSE approach where linear analytical expressions can be obtained.

b. Least squares estimation with varying c_k

When c_k does not depend on k, the LSE for the parameters in b_k and a is equivalent to the MLE and system of 12 equations in section 3a. Now, we assume that c_k^2 has a cyclostationary structure similar to b_k but without a trend. Its formulation is given by

$$c_{k,0}^2 = \rho_0 + \epsilon_0 \cos(2\pi k/\tau) + \epsilon_0' \sin(2\pi k/\tau) + \kappa_0 \cos(4\pi k/\tau) + \kappa_0' \sin(4\pi k/\tau) \qquad (8)$$

for dry days and

$$c_{k,1}^2 = \rho_1 + \epsilon_1 \cos(2\pi k/\tau) + \epsilon_1' \sin(2\pi k/\tau) + \kappa_1 \cos(4\pi k/\tau) + \kappa_1' \sin(4\pi k/\tau) \qquad (9)$$

for wet days. Here, k also varies from 0 to $K - 1$. However, because we assume that c_k is cyclostationary with no trend, it is sufficient to specify $c_{k,0}$ and $c_{k,1}$ only for $k = 0, \ldots, \tau - 1$. Our strategy to estimate $c_{k,0}$ and $c_{k,1}$ is to align the data by day of year $j = 0, \ldots, \tau - 1$ and segregate it according to the precipitation sequence. This yields the MLE (and LSE) estimators

$$\hat{c}_{j,0} = \sqrt{N_{j,0}^{-1}(\mathbf{DT} - \mathbf{B})'_{j,0}(\mathbf{DT} - \mathbf{B})_{j,0}} \quad \text{and}$$
$$\hat{c}_{j,1} = \sqrt{N_{j,1}^{-1}(\mathbf{DT} - \mathbf{B})'_{j,1}(\mathbf{DT} - \mathbf{B})_{j,1}}, \tag{10}$$

where $(\mathbf{DT} - \mathbf{B})_{j,0}$ is the $K/(\tau \times 1)$ vector populated with $(\mathbf{DT} - \mathbf{B})_k$ if $\chi_{k+1} = 0$ and with 0 if $\chi_{k+1} = 1$, where $k = j, j + \tau, j + 2\tau, \ldots, j + K - \tau$. Similarly, $(\mathbf{DT} - \mathbf{B})_{j,1}$ is the $K/(\tau \times 1)$ vector populated with $(\mathbf{DT} - \mathbf{B})_k$ if $\chi_{k+1} = 1$ and with 0 if $\chi_{k+1} = 0$, where $k = j, j + \tau, j + 2\tau, \ldots, j + K - \tau$. The quantity $N_{j,0}$ is the number of times $\chi_{k+1} = 0$, and $N_{j,1}$ is the number of times $\chi_{k+1} = 1$.

Once we have estimated $c_{j,0}$ and $c_{j,1}$, we use the LSE method to estimate the parameters in (8) and (9). Specifically, we minimize

$$\sum_{j=0}^{\tau-1} [\rho_0 + \epsilon_0 \cos(2\pi j/\tau) + \epsilon'_0 \sin(2\pi j/\tau) + \kappa_0 \cos(4\pi j/\tau) + \kappa'_0 \sin(4\pi j/\tau) - \hat{c}_{j,0}^2]^2 \quad \text{and} \tag{11}$$
$$\sum_{j=0}^{\tau-1} [\rho_1 + \epsilon_1 \cos(2\pi j/\tau) + \epsilon'_1 \sin(2\pi j/\tau) + \kappa_1 \cos(4\pi j/\tau) + \kappa'_1 \sin(4\pi j/\tau) - \hat{c}_{j,1}^2]^2. \tag{12}$$

Taking derivatives in each of the parameters in (11) and (12) and setting them equal to zero yields equations that are familiar from Fourier analysis. For dry days, we have

$$\hat{\rho}_0 = \tau^{-1} \sum_{j=0}^{\tau-1} \hat{c}_{j,0}^2,$$
$$\hat{\epsilon}_0 = \frac{2}{\tau} \sum_{j=0}^{\tau-1} \hat{c}_{j,0}^2 \cos(2\pi j/\tau),$$
$$\hat{\epsilon}'_0 = \frac{2}{\tau} \sum_{j=0}^{\tau-1} \hat{c}_{j,0}^2 \sin(2\pi j/\tau),$$
$$\hat{\kappa}_0 = \frac{2}{\tau} \sum_{j=0}^{\tau-1} \hat{c}_{j,0}^2 \cos(4\pi j/\tau), \quad \text{and}$$
$$\hat{\kappa}'_0 = \frac{2}{\tau} \sum_{j=0}^{\tau-1} \hat{c}_{j,0}^2 \sin(4\pi j/\tau),$$

and for wet days, we have

$$\hat{\rho}_1 = \tau^{-1} \sum_{j=0}^{\tau-1} \hat{c}_{j,1}^2,$$
$$\hat{\epsilon}_1 = \frac{2}{\tau} \sum_{j=0}^{\tau-1} \hat{c}_{j,1}^2 \cos(2\pi j/\tau),$$
$$\hat{\epsilon}'_1 = \frac{2}{\tau} \sum_{j=0}^{\tau-1} \hat{c}_{j,1}^2 \sin(2\pi j/\tau),$$
$$\hat{\kappa}_1 = \frac{2}{\tau} \sum_{j=0}^{\tau-1} \hat{c}_{j,1}^2 \cos(4\pi j/\tau), \quad \text{and}$$
$$\hat{\kappa}'_1 = \frac{2}{\tau} \sum_{j=0}^{\tau-1} \hat{c}_{j,1}^2 \sin(4\pi j/\tau).$$

The parameters are inserted back into (8) or (9) to generate the synthetic temperature series using the linear model (1). Example simulations with the seasonally varying c_k are shown in the right column of Fig. 3. Note how seasonally varying c_k better captures the low variability in the summer and high variability in the winter. The dry and wet c_k curves are shown with composite annual cycles of the standard deviation of the noise in Fig. 4. These curves highlight the larger variability associated with wet days as well as the larger variability associated with the transition seasons (spring and fall) featuring strong frontal temperature contrasts.

c. *Simulation of precipitation*

In this section, for completeness, we provide formulation for simulation of daily precipitation in a manner compatible with the temperature model introduced above. Our formulation largely follows Woolhiser (2008), except here we allow for trends in the Markov chain parameters. The probability of precipitation occurrence is determined with a two-state (wet or dry), second-order Markov chain, which means that the probability of precipitation on a given day depends on the precipitation state on the previous two days as follows:

$$p_{ij,0}(t) = P\{\chi_t = 0 \,|\, \chi_{t-1} = j, \chi_{t-2} = i\}; \quad t = 1, 2, \ldots, 365M, \tag{13}$$

where M is the number of years. If we assume cyclostationarity, then the p_{ij0} terms are periodic, meaning $p_{ij,0}(t + K365) = p_{ij,0}(t)$ for any integer K. To account for nonstationarity associated with low-frequency oceanic forcing plus any trend, we define perturbed versions of (13):

$$\tilde{p}'_{ij,0}(t) = \tilde{p}_{ij,0}(t) + b_0^{(ij,0)} + b_1^{(ij,0)} t + b_2^{(ij,0)} S_1(t - \tau_1) + b_3^{(ij,0)} S_2(t - \tau_2), \tag{14}$$

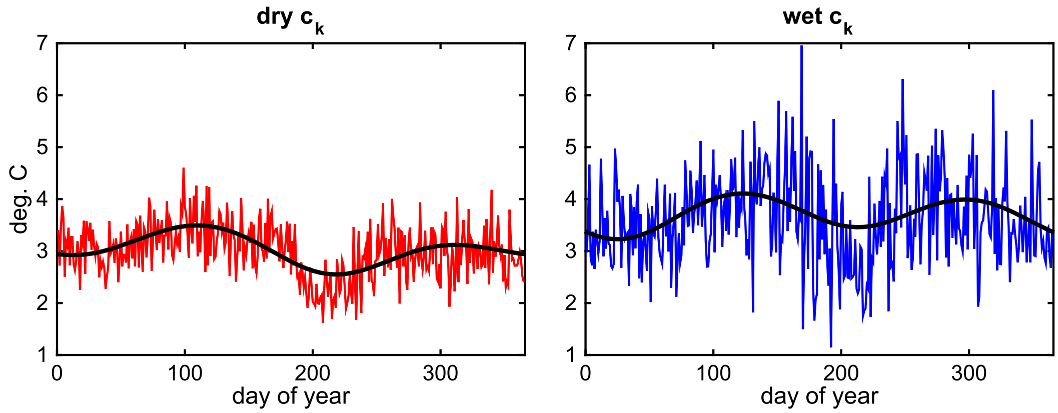

FIG. 4. Seasonally varying c_k curves for (left) dry and (right) wet days (black lines) and standard deviations of the noise (colored lines). Note the relatively higher variability in the transitional seasons and overall higher variability associated with the wet days.

where $\{b_0, b_1\}$ enable a trend, S_1 and S_2 are oceanic forcing with periodicity of 3–7 yr (ENSO) and 10–15 yr (PDO), respectively, and the τ terms are positive lags (i.e., variations in ENSO and PDO indices may lead their effects on precipitation by τ months).

4. Comparison with the Richardson method

Because the Richardson method of simulating stochastic temperature (referred to as the multivariate generation model) is the most widely used parametric method in the field and the one upon which SHArP builds, it is a useful point of comparison. The Richardson method is essentially an autoregressive process that simulates standardized residuals; the details of this method can be found in Richardson (1981) and Matalas (1967). The Richardson method prescribes the means and standard deviations of the data (for wet and dry days) prior to simulation via a harmonic fit and then reintroduces them after simulating standardized residuals. This causes the model mean and standard deviation to abruptly switch between wet- and dry-state values. The model we introduce here (1) also has wet- and dry-state harmonics (b_k) and noise amplitudes (b_k) prescribed in advance, but the mean of the model ($\mathbf{D}^{-1}\mathbf{B}$) and standard deviation make autocorrelated, and hence more realistic, transitions via the parameter a in $\mathbf{D}$.

We highlight the difference between the methods in Fig. 5, which compares the composite synthetic temperature simulated by the two models with the observational temperature for precipitation occurrence sequences of dry-dry-wet-wet-dry-dry for each season. The observational temperature reflects a typical cold frontal passage in each season (e.g., Shafer and Steenburgh 2008). In general, the observed maximum temperature increases shortly before the frontal passage because of southerly flow and warm air advection; on the first day of precipitation, the maximum temperature decreases modestly. On the second day of precipitation, the temperature continues to decrease, and it slowly rebounds following

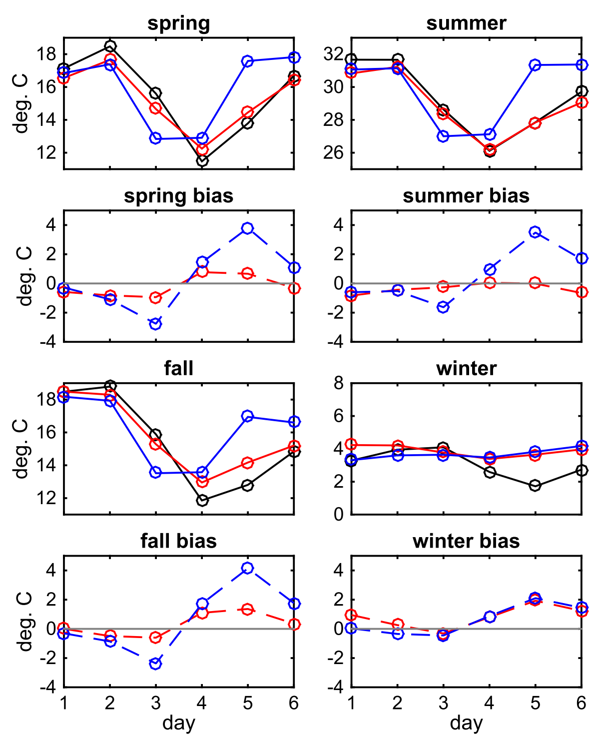

FIG. 5. Composite observation temperature (black lines) and composite synthetic temperature for sets of days that follow the precipitation occurrence sequence dry-dry-wet-wet-dry-dry, in each season. In addition, the bias for each season is also shown immediately below these composite panels. Composite is of each occurrence of this sequence at five climatologically similar sites (see Fig. 1). The red lines indicate SHArP, the model presented here, and the blue lines indicate the Richardson model. The number of samples in each set is approximately 500.

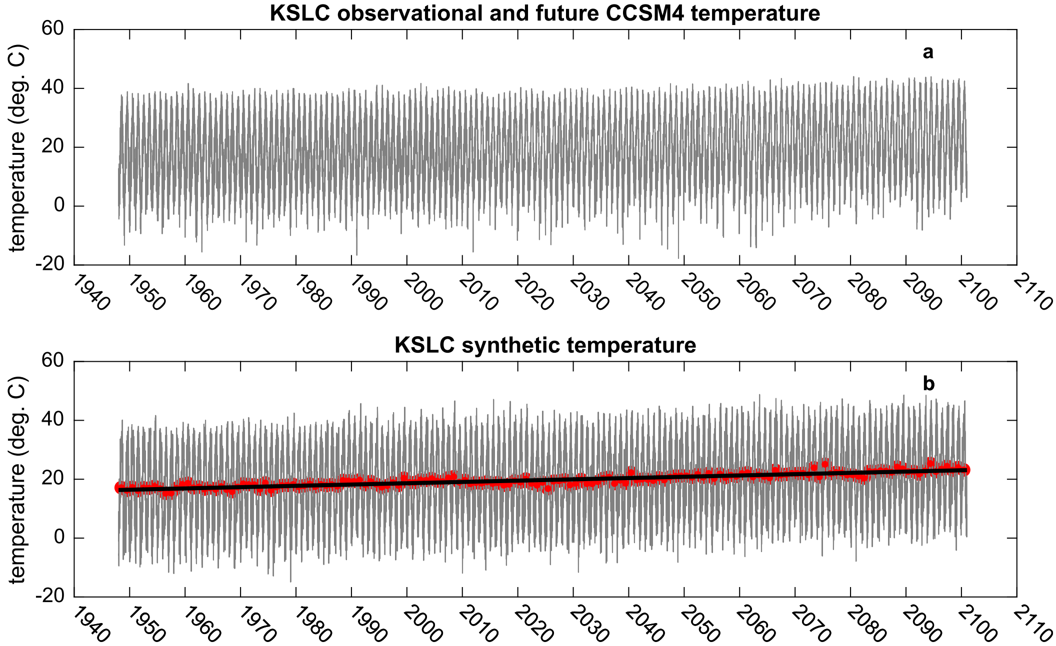

FIG. 6. (a) KSLC observation GHCN-Daily maximum temperature (1948–2010) and BCCA CCSM4 high-emissions (RCP8.5) maximum temperature output (2011–2100). (b) An example of a trended stochastic maximum temperature simulated from 1948 to 2100 for KSLC. The simulation was trained on the data shown in (a). The red dots indicate the average annual maximum temperature for each year of the simulation.

the precipitation event. SHArP is able to capture this overall pattern. In contrast, the abrupt switching between wet- and dry-state means in the Richardson model results in an unrealistically large decrease in temperature on the first day of precipitation, followed by minimal change on the second day (actually zero change with large enough sample). While there is little to no seasonal bias in the Richardson model, there is a bias in the temperature around frontal passages. The temperature bias in the Richardson model is up to 4°C following frontal passages, and SHArP is able to reduce that bias by 2°C in three seasons.

Although the Richardson framework as originally formulated does not contain a trend term, one could be added in principle. One approach would be to fit the trend by LSE and remove it prior to estimating the annual cycles of the mean and residual standard deviations, and then adding the trend back in after generating the simulated temperatures. In contrast to this multistep approach involving removing components, fitting, simulating, and reintroducing components, the model presented here involves only fitting and simulation because all variations are captured in the fit formulation, including a trend term that is incorporated into b_k. Trended output from observations (1948–2010) and future BCCA CCSM4 high-emissions scenario output (2011–2100) is shown in Fig. 6a with an example corresponding realization from the temperature model presented here shown in Fig. 6b.

5. Discussion and conclusions

This study presents a new linear model for simulating stochastic temperature realizations called SHArP, and the method was illustrated for maximum temperature at a single site within the Great Basin. We first considered a simplified version of the model with a constant noise coefficient c and applied MLE to obtain its parameters. However, this constant c compromised between the variance in the summer and the variance in the winter, which resulted in a simulation that did not adequately capture the seasonal variance found in the observations. A seasonally varying noise coefficient c_k rendered the MLE nonlinear, and we presented analytical solutions via LSE. The resulting temperature realization more closely matched that of observations, with increased wintertime variance and decreased summertime variance.

Further realism may also be possible by relaxing assumptions used here. For example, we assume the amplitude of noise c_k to be annually cyclostationary but without trend. It is possible for the noise to have similar nonstationarity because of ENSO and PDO. Curvilinearity (a trend) and variables related to ENSO and PDO could be added to the c_k^2 equations (if the area of interest is in a region where these oceanic modes play a major role) and solved using the same LSE method. A nonlinear trend could also be added to αk portion of the b_k equation, making it $\alpha_1 k + \alpha_2 k^2$. We also assume that temperature depends only on itself and precipitation occurrence, but precipitation amount and climate teleconnections that influence airmass trajectories may be additionally important.

Even though this study is focused on only maximum temperature at a single site, we have generalized the method described to include minimum temperature in addition to maximum temperature at multiple sites. The linear model remains the same, but the scalar computations become matrix computations. We extended ideas described in Wilks (1998) and Wilks (1999b), where the sites themselves have spatial correlation but are generated independently of each other, by introducing spatial correlations in the c_k matrices but not in the a matrix. However, this method introduced an increased number of parameters in the variance–covariance matrix that required a nontrivial technique to mitigate the issue, and this will be described in a future paper.

Acknowledgments. This material is based upon work supported by the National Science Foundation under Grants EPS-1135482, EPS-1135483, EPS-1208732, and DMS-1407574. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Provision of computer infrastructure by the Center for High Performance Computing at the University of Utah is gratefully acknowledged. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

REFERENCES

Bailey, N. T. J., 1964: *The Elements of Stochastic Processes with Applications to the Natural Sciences.* John Wiley and Sons, 249 pp.

Brekke, L., B. L. Thrasher, E. P. Maurer, and T. Pruitt, 2013: Downscaled CMIP3 and CMIP5 climate projections: Release of downscaled CMIP5 climate projections, comparison with preceding information, and summary of user needs. U.S. Department of the Interior Bureau of Reclamation Tech. Rep., 116 pp. [Available online at http://gdo-dcp.ucllnl.org/downscaled_cmip_projections/techmemo/downscaled_climate.pdf.]

Brown, D. P., 2011: Winter circulation anomalies in the western United States associated with antecedent and decadal ENSO variability. *Earth Interact.*, **15**, doi:10.1175/2010EI334.1.

Caraway, N. M., J. L. McCreight, and B. Rajagopalan, 2014: Multisite stochastic weather generation using cluster analysis and k-nearest neighbor time series resampling. *J. Hydrol.*, **508**, 197–213, doi:10.1016/j.jhydrol.2013.10.054.

Chandler, R. E., 2005: On the use of generalized linear models for interpreting climate variability. *Environmetrics*, **16**, 699–715, doi:10.1002/env.731.

Dettinger, M. D., D. R. Cayan, H. F. Diaz, and D. M. Meko, 1998: North–south precipitation patterns in western North America on interannual-to-decadal timescales. *J. Climate*, **11**, 3095–3111, doi:10.1175/1520-0442(1998)011<3095:NSPPIW>2.0.CO;2.

Forsythe, N., H. Fowler, S. Blenkinsop, A. Burton, C. Kilsby, D. Archer, C. Harpham, and M. Hashmi, 2014: Application of a stochastic weather generator to assess climate change impacts in a semi-arid climate: The upper Indus basin. *J. Hydrol.*, **517**, 1019–1034, doi:10.1016/j.jhydrol.2014.06.031.

Furrer, E. M., and R. W. Katz, 2007: Generalized linear modeling approach to stochastic weather generators. *Climate Res.*, **34**, 129–144, doi:10.3354/cr034129.

Gershunov, A., and T. P. Barnett, 1998: Interdecadal modulation of ENSO teleconnections. *Bull. Amer. Meteor. Soc.*, **79**, 2715–2725, doi:10.1175/1520-0477(1998)079<2715:IMOET>2.0.CO;2.

——, ——, and D. R. Cayan, 1999: North Pacific interdecadal oscillation seen as factor in ENSO-related North American climate anomalies. *Eos, Trans. Amer. Geophys. Union*, **80**, 25–30, doi:10.1029/99EO00019.

Harrold, T. I., A. Sharma, and S. J. Sheather, 2003: A nonparametric model for stochastic generation of daily rainfall amounts. *Water Resour. Res.*, **39**, 1343, doi:10.1029/2003WR002570.

Horel, J. D., and J. M. Wallace, 1981: Planetary-scale atmospheric phenomena associated with the Southern Oscillation. *Mon. Wea. Rev.*, **109**, 813–829, doi:10.1175/1520-0493(1981)109<0813:PSAPAW>2.0.CO;2.

Kiktev, D., J. Caesar, L. V. Alexander, H. Shiogama, and M. Collier, 2007: Comparison of observed and multimodeled trends in annual extremes of temperature and precipitation. *Geophys. Res. Lett.*, **34**, L10702, doi:10.1029/2007GL029539.

Matalas, N. C., 1967: Mathematical assessment of synthetic hydrology. *Water Resour. Res.*, **3**, 937–945, doi:10.1029/WR003i004p00937.

Mauget, S. A., 2003: Intra- to multidecadal climate variability over the continental United States: 1932–99. *J. Climate*, **16**, 2215–2231, doi:10.1175/2751.1.

Maurer, E. P., L. Brekke, T. Pruitt, and P. B. Duffy, 2007: Fine-resolution climate projections enhance regional climate change impact studies. *Eos, Trans. Amer. Geophys. Union*, **88**, 504–504, doi:10.1029/2007EO470006.

McCullagh, P., and J. Nelder, 1989: *Generalized Linear Models.* 2nd ed. Chapman & Hall, 532 pp.

Menne, M. J., and Coauthors, 2012a: Global Historical Climatology Network—Daily (GHCN-Daily), version 3. National Climatic Data Center, accessed 7 December 2015. [Available online at ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/.]

——, I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston, 2012b: An overview of the Global Historical Climatology Network-Daily database. *J. Atmos. Oceanic Technol.*, **29**, 897–910, doi:10.1175/JTECH-D-11-00103.1.

Rajagopalan, B., and U. Lall, 1999: A k-nearest-neighbor simulator for daily precipitation and other weather variables. *Water Resour. Res.*, **35**, 3089–3101, doi:10.1029/1999WR900028.

——, ——, and D. G. Tarboton, 1997: Evaluation of kernel density estimation methods for daily precipitation resampling. *Stochastic Hydrol. Hydraul.*, **11**, 523–547, doi:10.1007/BF02428432.

Richardson, C. W., 1981: Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resour. Res.*, **17**, 182–190, doi:10.1029/WR017i001p00182.

Roldàn, J., and D. A. Woolhiser, 1982: Stochastic daily precipitation models: 1. A comparison of occurrence processes. *Water Resour. Res.*, **18**, 1451–1459, doi:10.1029/WR018i005p01451.

Ropelewski, C. F., and M. S. Halpert, 1986: North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). *Mon. Wea. Rev.*, **114**, 2352–2362, doi:10.1175/1520-0493(1986)114<2352:NAPATP>2.0.CO;2.

Shafer, J. C., and W. J. Steenburgh, 2008: Climatology of strong intermountain cold fronts. *Mon. Wea. Rev.*, **136**, 784–807, doi:10.1175/2007MWR2136.1.

Stern, R. D., and R. Coe, 1984: A model fitting analysis of daily rainfall data. *J. Roy. Stat. Soc.*, **147A**, 1–34, doi:10.2307/2981736.

Thompson, G. A., and D. B. Burke, 1974: Regional geophysics of the basin and range province. *Annu. Rev. Earth Planet. Sci.*, **2**, 213–238, doi:10.1146/annurev.ea.02.050174.001241.

Troup, A. J., 1965: The 'southern oscillation.' *Quart. J. Roy. Meteor. Soc.*, **91**, 490–506, doi:10.1002/qj.49709139009.

Wilks, D. S., 1992: Adapting stochastic weather generation algorithms for climate change studies. *Climatic Change*, **22**, 67–84, doi:10.1007/BF00143344.

——, 1998: Multisite generalization of a daily stochastic precipitation generation model. *J. Hydrol.*, **210**, 178–191, doi:10.1016/S0022-1694(98)00186-3.

——, 1999a: Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. *Agric. For. Meteor.*, **93**, 153–169, doi:10.1016/S0168-1923(98)00125-7.

——, 1999b: Simultaneous stochastic simulation of daily precipitation, temperature and solar radiation at multiple sites in complex terrain. *Agric. For. Meteor.*, **96**, 85–101, doi:10.1016/S0168-1923(99)00037-4.

——, 2008: High-resolution spatial interpolation of weather generator parameters using local weighted regressions. *Agric. For. Meteor.*, **148**, 111–120, doi:10.1016/j.agrformet.2007.09.005.

——, and R. L. Wilby, 1999: The weather generation game: A review of stochastic weather models. *Prog. Phys. Geogr.*, **23**, 329–357, doi:10.1191/030913399666525256.

Wise, E. K., 2010: Spatiotemporal variability of the precipitation dipole transition zone in the western United States. *Geophys. Res. Lett.*, **37**, doi:10.1029/2009GL042193.

Woolhiser, D. A., 2008: Combined effects of the Southern Oscillation index and the Pacific decadal oscillation on a stochastic daily precipitation model. *J. Climate*, **21**, 1139–1152, doi:10.1175/2007JCLI1862.1.