An Extended Version of the Richardson Model for Simulating Daily Weather Variables

MARC B. PARLANGE

Department of Geography and Environmental Engineering, The Johns Hopkins University, Baltimore, Maryland

RICHARD W. KATZ

Environmental and Societal Impacts Group, National Center for Atmospheric Research, Boulder, Colorado*

(Manuscript received 7 December 1998, in final form 28 April 1999)

ABSTRACT

The Richardson model is a popular technique for stochastic simulation of daily weather variables, including precipitation amount, maximum and minimum temperature, and solar radiation. This model is extended to include two additional variables, daily mean wind speed and dewpoint, because these variables (or related quantities such as relative humidity) are required as inputs for certain ecological/vegetation response and agricultural management models. To allow for the positively skewed distribution of wind speed, a power transformation is applied. Solar radiation also is transformed to make the shape of its modeled distribution more realistic. A model identification criterion is used as an aid in determining whether the distributions of these two variables depend on precipitation occurrence. The approach can be viewed as an integration of what is known about the statistical properties of individual weather variables into a single multivariate model.

As an application, this extended model is fitted to weather data in the Pacific Northwest. To aid in understanding how such a stochastic weather generator works, considerable attention is devoted to its statistical properties. In particular, marginal and conditional distributions of wind speed and solar radiation are examined, with the model being capable of representing relationships between variables in which the variance is not constant, as well as certain forms of nonlinearity.

1. Introduction

Stochastic weather generators have been proposed as one technique for simulating time series consistent with the current climate as well as for producing scenarios of climate change (Wilks 1992). In particular, such simulations have been used in assessments of the effects of climate variability and change, primarily on managed environmental systems (e.g., Mearns et al. 1997). Because of this recent attention, an awareness of the limitations of these stochastic models is starting to develop (Johnson et al. 1996; Semenov et al. 1998). Among these limitations is the omission or inadequate statistical treatment of weather variables that are not normally distributed.

The reliance on one particular stochastic model, termed the Richardson model or WGEN (Weather Generator; Richardson 1981; Richardson and Wright 1984), has been prevalent in climate impact studies. For instance, Mearns et al. (1997) adapted this generator to produce climate change scenarios as input to crop–climate models. The Richardson model simulates daily time series of precipitation amount, maximum and minimum temperature, and solar radiation. Other weather generators that resemble the Richardson model have appeared, including one developed by Bruhn et al. (1980). All these generators attempt to integrate individual stochastic models for component weather variables, such as those for precipitation (e.g., Katz 1977; Todorovic and Woolhiser 1975) and temperature (e.g., Hansen and Driscoll 1977). Note that Young (1994) proposed an alternative approach based on resampling from the existing weather dataset (see also Rajagopalan et al. 1997).

Certain models of agricultural productivity and of natural ecosystems used in impact assessments require additional weather variables such as wind speed and relative humidity (Easterling et al. 1992; Neilson 1995). One of the extended versions of WGEN, known as WXGEN (Erosion/Productivity Impact Calculator weather generator; Nicks et al. 1990), does allow for the nonnormal distributions of these two variables, but wind speed is not linked to any of the other weather variables, and relative humidity is linked only to pre-

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author address: Dr. Richard W. Katz, Environmental and Societal Impacts Group, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000.
E-mail: rwk@ucar.edu

cipitation occurrence (Wallis and Griffiths 1995). Wind speed and dewpoint (from which relative humidity can be derived) are included in the weather generator GEM (Generation of Weather Elements for Multiple Applications; Hanson and Johnson 1998), with dependence being permitted among all the weather variables, but assuming normal distributions for these two variables.

In the current paper, another extension of the Richardson model is devised, effectively a hybrid of WXGEN and GEM, combining the individual improvements of these two stochastic weather generators. The additional weather variables daily mean wind speed and dewpoint are included in the model. Whether daily wind speed and dewpoint (as well as the other weather variables) ought to be linked to precipitation occurrence is determined, in part, through a model identification procedure. By using a power transformation to normality (i.e., the square root), this extension allows for the positively skewed distribution of wind speed while still permitting statistical dependence with the other variables. Instead of assuming normality as in WGEN, the distribution of daily solar radiation is reexamined for skewness, with transformations again being applied. The extended model is fitted to weather data in the Pacific Northwest, primarily in Eugene, Oregon. Still, this form of weather generator can be viewed as a general framework, applicable both to other regions and to additional weather variables.

2. Modeling methodology

a. Richardson model

By way of review, the basic elements of the Richardson model/WGEN for time series of daily weather variables are described briefly (Richardson 1981; Richardson and Wright 1984). As a slight simplification of the approach taken in WGEN, any annual cycles in the model parameters are ignored by restricting attention to a single month at a time. Some of the properties of this model also have been reviewed in Katz (1996) and Wilks (1992).

Because the other weather variables are modeled conditionally on precipitation occurrence, first the stochastic model for precipitation is specified. Let $\{J_t : t = 1, 2, \ldots\}$ denote the sequence of daily precipitation occurrence (i.e., $J_t = 1$ indicates that the tth day is ''wet''; $J_t = 0$ is a ''dry day'') at a given site. To represent the tendency of wet or dry weather to persist, it is assumed that this process is a first-order, two-state Markov chain (Katz 1977; Todorovic and Woolhiser 1975). This model is characterized completely by the transition probabilities

$$P_{ij} = \Pr\{J_t = j \mid J_{t-1} = i\}, \qquad i, j = 0, 1. \tag{1}$$

The amounts of daily precipitation on wet days are assumed to be conditionally independent given the sequence of occurrences of precipitation, with the gamma as the common distribution. Some enhancements of WGEN use other positively skewed distributions, such as the mixed exponential in USCLIMATE (Program for Daily Weather Simulation in the Contiguous United States; Johnson et al. 1996) or the ''skewed'' normal in WXGEN (Wallis and Griffiths 1995). Because this aspect does not enter into the manner in which the models link the other variables with precipitation, and because it already has received much study, it is not treated here.

Let $X_t(k)$, $k = 1, 2, \ldots, K$, denote the daily weather variables to be modeled conditionally on precipitation occurrence. In WGEN, K equals 3 with, by convention, $X_t(1)$ denoting the maximum temperature, $X_t(2)$ denoting the minimum temperature, and $X_t(3)$ denoting the total solar radiation. Given the precipitation occurrence state on the tth day, say $J_t = i$, the conditional distribution of $X_t(k)$ is assumed to be normal, with a mean and variance of

$$\mu_i(k) = E[X_t(k) \mid J_t = i], \quad \text{and}$$
$$\sigma_i^2(k) = \mathrm{Var}[X_t(k) \mid J_t = i],$$
$$i = 0, 1; \qquad k = 1, 2, \ldots, K. \tag{2}$$

As a simplification, in WGEN the conditional mean and variance of minimum temperature [i.e., $\mu_i(2)$ and $\sigma_i^2(2)$, $i = 0, 1$] actually are taken to be independent of the precipitation occurrence state i. This constraint is not necessarily imposed in the current paper, however. In fact, Hayhoe (1998) found a dependence between minimum temperature and precipitation occurrence for some locations in Canada, and Semenov et al. (1998) allow for this dependence in another weather generator, called the Long Ashton Research Station Weather Generator (LARS-WG).

Given the daily precipitation occurrence state $J_t = i$, one defines the randomly standardized variable (termed ''residual'' by Richardson 1981) as

$$Z_t(k) = [X_t(k) - \mu_i(k)]/\sigma_i(k), \qquad k = 1, 2, \ldots, K. \tag{3}$$

In other words, it is not known a priori (i.e., before the precipitation occurrence state is observed) which particular mean and standard deviation will enter into (3). By assumption, the $Z_t(k)$ time series have standard normal distributions [denoted by N(0, 1)].

To permit autocorrelation in the individual time series and cross correlations between the time series, a multivariate, first-order autoregressive [AR(1)] process is assumed as a statistical model of the randomly standardized variables $Z_t(k)$, $k = 1, 2, \ldots, K$. That is,

$$\mathbf{Z}_t = \mathbf{A}\mathbf{Z}_{t-1} + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathrm{MVN}(\mathbf{0}, \mathbf{S}). \tag{4}$$

Here, $\mathbf{Z}_t$ denotes the column vector of dimension K whose elements are $Z_t(k)$, $k = 1, 2, \ldots, K$; $\mathbf{Z}_{t-1}$ is the corresponding vector containing $Z_{t-1}(k)$; $\mathbf{A}$ is the $K \times K$ matrix of autoregression coefficients; and $\boldsymbol{\epsilon}_t$ is the column vector of dimension K of error terms $\epsilon_t(k)$. As indicated in (4), this error vector is assumed to have a multivariate normal distribution (MVN) with mean vector zero, denoted by **0**, and $K \times K$ variance–covariance matrix, denoted by **S**.

The matrices **A** and **S** satisfy the equations

$$\mathbf{AM}_0 = \mathbf{M}_1, \qquad \mathbf{S} = \mathbf{M}_0 - \mathbf{AM}_1^{\mathrm{T}}, \tag{5}$$

where $\mathbf{M}_0$ and $\mathbf{M}_1$ are the lag-zero and lag-one cross-covariance matrices [or, equivalently, cross-correlation matrices because the $Z_t(k)$ are standardized] and T denotes the matrix transpose operator (e.g., Brockwell and Davis 1991). That is, $\mathbf{M}_0$ is a symmetric matrix whose elements are of the form

$$\rho_{kl}(0) = \mathrm{Corr}[Z_t(k), Z_t(l)], \qquad k, l = 1, 2, \ldots, K, \tag{6}$$

and $\mathbf{M}_1$ is a matrix (not necessarily symmetric) whose elements are of the form

$$\rho_{kl}(1) = \mathrm{Corr}[Z_t(k), Z_{t-1}(l)], \qquad k, l = 1, 2, \ldots, K. \tag{7}$$

In particular, the diagonal elements of $\mathbf{M}_0$ are all unity, whereas those of $\mathbf{M}_1$ are the first-order autocorrelation coefficients of the randomly standardized variables $Z_t(k)$.

The so-called Yule–Walker estimators (Brockwell and Davis 1991) of the parameters of the multivariate AR(1) process (4) are obtained by first substituting the corresponding sample cross correlations into the matrices $\mathbf{M}_0$ and $\mathbf{M}_1$. An estimator of the autoregression matrix **A** then is obtained by solving the system of linear equations [first formula in (5)]. The estimator of the variance–covariance matrix **S** is obtained finally by substituting the estimator of **A** into the second formula in (5). This estimation technique has the property of reproducing all of the contemporaneous and lag-one cross correlations (including the first-order autocorrelations) among the randomly standardized variables $Z_t(k)$. Unlike WGEN, the matrices **A** and **S** are not constrained to be invariant with respect to geographical location and time of year. Instead, they are fitted for a particular location and assumed to be stationary within a given month. Our approach is consistent with Hayhoe (1998), who found that these matrices vary both seasonally and regionally for locations in Canada, and with Richardson (1982), who found seasonal and regional variations in the United States, especially for the cross-correlation coefficients (i.e., elements of $\mathbf{M}_0$).

b. Extension

In the extended version of the Richardson model, five daily weather variables (i.e., $K = 5$) are modeled conditionally on precipitation occurrence (see Table 1 for ordering convention). Daily mean dewpoint is treated, because relative humidity can be derived from this variable in combination with air temperature. Although minimum temperature sometimes is used as a substitute for dewpoint in estimating humidity, Kimball et al. (1997) showed that there can be substantial differences between these two variables, especially in arid or semiarid climates. Daily maximum and minimum temperature and dewpoint satisfy certain orderings (e.g., minimum ≤ maximum), not automatically preserved by the multivariate normality assumption in (4) (Jolliffe and Hope 1996). If desired, such orderings could be imposed on the simulated data through additional conditioning.

TABLE 1. Daily weather variables.

k	Variable	Transformation	Conditioning
1	Maximum temperature	None	Mean and std dev
2	Minimum temperature	None	Mean and std dev
3	Mean dewpoint	None	Mean and std dev
4	Mean wind speed	Square root	Mean only
5	Total solar radiation	Square root (Jan), reflected logarithm (Jul)	Mean only

It is well known that the distribution of hourly or daily mean wind speed is positively skewed, with theoretical distributions such as the Weibull having been fitted. It is difficult to allow, however, for the temporal or spatial correlations of wind speed time series as well in such an approach. With a transformation to normality, conventional multiple time series analysis still can be applied to treat autocorrelations as well as cross correlations with other variables. Brown et al. (1984), Carlin and Haslett (1982), and Haslett and Raftery (1989) all applied the square root transformation to hourly or daily mean wind speed before modeling temporal and/or spatial correlations. Moreover, assuming that the square-root-transformed wind speed is normally distributed effectively is equivalent to fitting a Weibull distribution to the original observations (Brown et al. 1984; Carlin and Haslett 1982).

The distribution of daily total solar radiation has not been examined closely in the development and application of the Richardson model, at least in part because of a lack of observations. In fact, WGEN (or WXGEN) sometimes is used to manufacture time series of daily radiation for locations without any measurements [although Hayhoe (1998) cautioned against using WXGEN for this purpose]. Daily radiation does not necessarily have a normal distribution (Bruhn et al. 1980; Semenov et al. 1998), and a transformation (square root or reflected logarithm) is applied in the current model.

3. Application

a. Data

The extended version of the Richardson model is fitted to daily weather data at Eugene and Portland, Oregon, for two months, January and July (in the midst of the wet and dry seasons, respectively), over a period of 30 yr, 1961–90. Precipitation occurs on relatively few

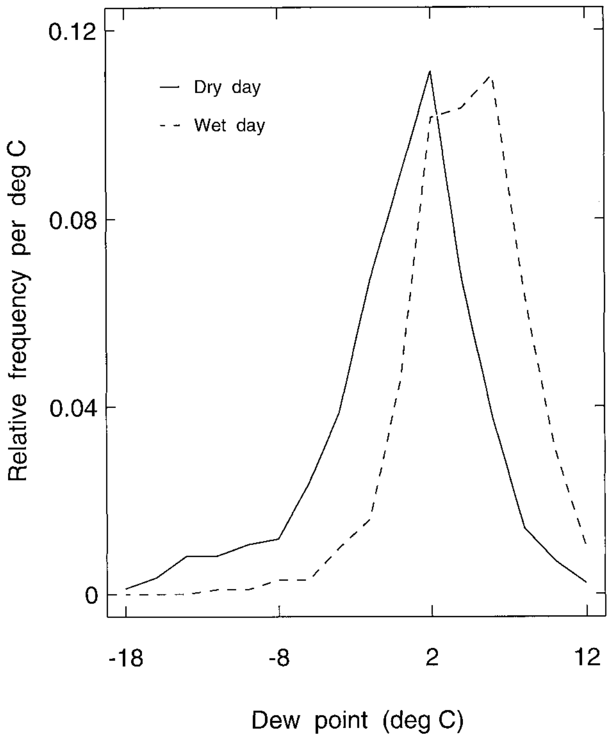

FIG. 1. Observed conditional distribution of daily mean dewpoint (°C) given a dry or wet day in Eugene in Jan.

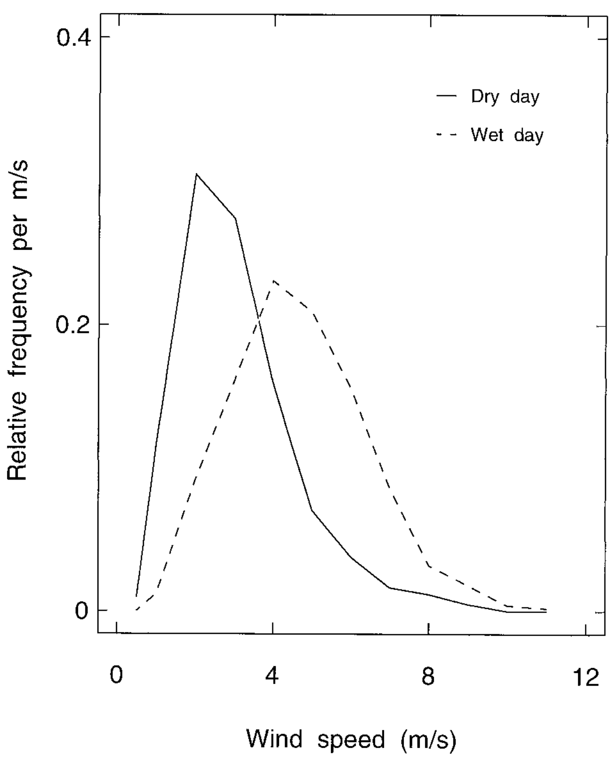

FIG. 2. Same as Fig. 1 but for wind speed (m s^{-1}).

days in July (e.g., on only 81 out of 930 days in Eugene), making the sample sizes small for certain conditional statistics. To simplify the presentation, for the most part the focus is on Eugene in January.

Some of the data, including solar radiation, were obtained from the National Solar Radiation Data Base (NSRDB) (National Renewable Energy Laboratory, 1992). Almost half of the radiation data for Eugene are measured, with the remaining half being modeled from observed or interpolated sky cover and derived aerosol optical depths, whereas only about 15% of the radiation data for Portland are measured. NSRDB gives the amount of solar radiation received during an hour in units labeled W h m^{-2} (1 W h m^{-2} = 3600 J m^{-2}). These hourly amounts are summed to obtain daily totals in units of W day m^{-2}.

b. Model identification

In this section, the form of transformation is specified for the daily weather variables, mean dewpoint, mean wind speed, and total solar radiation, to obtain a better approximation to the normal distribution. Whether the conditional mean and standard deviation of these three variables (as well as those of maximum and minimum temperature) ought to depend on precipitation occurrence also is considered. The form of transformation and nature of conditioning, as eventually determined, are summarized in Table 1.

Figures 1–3 show the observed conditional distributions of these three daily weather variables given a dry or wet day in Eugene in January (also July for radiation). A tendency for higher dewpoint on wet days is evident (Fig. 1). The shape of the two conditional distributions is reasonably symmetric, with perhaps a slight degree of negative skewness, especially given a dry day. As for maximum and minimum temperature in the Richardson model, no transformation will be applied to dewpoint. In July, there is not as much dependence of dewpoint on precipitation (figure not shown).

Higher wind speed tends to occur on wet days, with a substantial degree of positive skewness being present in the conditional distributions, especially given a dry day (Fig. 2). The square root transformation will be applied to wind speed, with a check on the appropriateness of this transformation being provided in section 4b. In July, the conditional distribution of wind speed similarly is positively skewed, but with not as much dependence upon precipitation occurrence (figure not shown).

A marked tendency for higher radiation on dry days is evident, both in January and July (Fig. 3). The conditional distribution given a wet day appears to be somewhat positively skewed in January (Fig. 3a). The square root transformation will be applied to radiation in January, with a check again being provided in section 4b. Note that Bruhn et al. (1980) also found evidence at some other geographical locations of the conditional

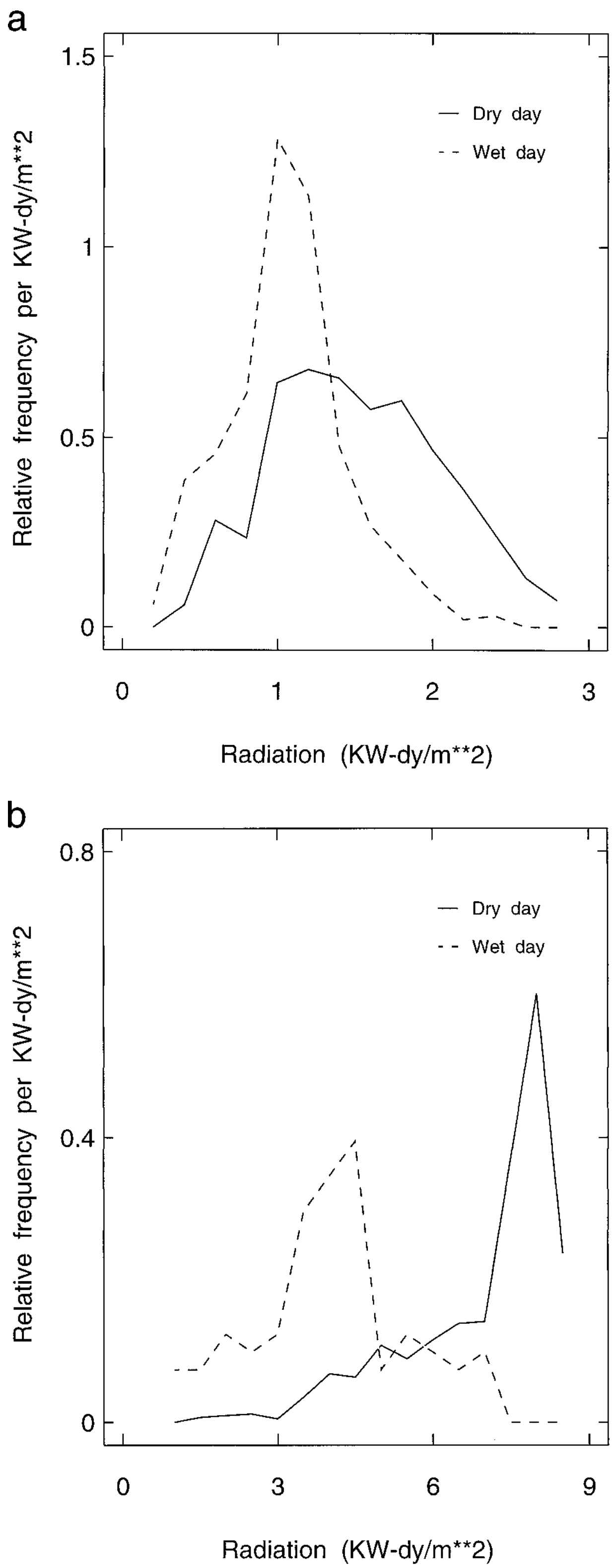

FIG. 3. Same as Fig. 1 but for daily total solar radiation (kW day m^{-2}) for (a) Jan and (b) Jul.

distribution of radiation given a wet day being positively skewed.

The shape of the conditional distribution of radiation in July given a dry day is very different from that in January, being substantially negatively skewed (Fig. 3b). A transformation of a form other than the square root is required. After some trial and error, a reflected logarithm was selected; that is, the transformation is of the form $\ln(c - x)$, where x denotes radiation, and c is a constant about which the logarithm is reflected. This constant could be determined to be the theoretical upper bound on daily total radiation, which is a function of the latitude of the location and the time of year (e.g., Brutsaert 1982). Instead, the value of $c = 9$ kW day m^{-2}, somewhat smaller than this upper bound, yields a better approximation to the normal distribution. Perhaps this unusual distribution shape could be attributed to precipitation occurrence not being an adequate surrogate for cloudiness during the dry summer season in the Pacific Northwest (in fact, morning cloudiness is common for days on which no precipitation occurs). Despite the differences in the sources of radiation data, the same forms of transformation were identified for Portland.

The appendix contains details about the model identification procedures, Akaike's information criterion (AIC) and the Bayesian information criterion (BIC), employed to determine whether the conditional mean and standard deviation of the daily weather variables that enter into (3) ought to vary with daily precipitation occurrence [see Katz and Parlange (1995) for a somewhat analogous use of these model selection criteria]. These criteria involve penalizing the goodness of fit of a model for the number of parameters required to be estimated, with the preferred model being the one with minimum AIC or BIC statistic (BIC is somewhat more parsimonious). Although these statistics are dimensionless, it is possible to convert the BIC values into approximate posterior probabilities.

For the following three candidate models, Table 2 gives the AIC and BIC statistics for the five weather variables for Eugene in January (number of observations $n = 930$):

model (i): $\mu_0(k) = \mu_1(k)$, $\sigma_0(k) = \sigma_1(k)$;

model (ii): $\mu_0(k) \neq \mu_1(k)$, $\sigma_0(k) = \sigma_1(k)$; and

model (iii): $\mu_0(k) \neq \mu_1(k)$, $\sigma_0(k) \neq \sigma_1(k)$.

In all cases, both AIC and BIC indicate a need to vary at least the conditional mean [i.e., model (i) of no dependence on precipitation occurrence is never selected]. Nevertheless, the results are not completely clearcut concerning whether the conditional standard deviation ought to be varied as well. The results are similar for Eugene in July, although the magnitude of the dependence on precipitation occurrence is less for most variables (but recall the paucity of wet days).

For the weather data from the Pacific Northwest analyzed in this paper, the following conditioning strategy

TABLE 2. Model identification statistics for Eugene in Jan.

k	Variable	Conditioning	No. of parameters	AIC	BIC
1	Maximum temperature	None	3	2212.2	2226.7
		Mean only	4	2157.4	2176.8
		Mean and std dev	5	2143.5*	2167.7*
2	Minimum temperature	None	3	2269.0	2283.5
		Mean only	4	2058.0	2077.3*
		Mean and std dev	5	2057.6*	2081.8
3	Dewpoint	None	3	2157.8	2172.3
		Mean only	4	1960.6	1979.9
		Mean and std dev	5	1917.5*	1941.6*
4	Wind speed (square root)	None	3	−1685.4	−1670.9
		Mean only	4	−1909.4*	−1890.0*
		Mean and std dev	5	−1906.6	−1882.4
5	Radiation (square root)	None	3	3617.5	3632.0
		Mean only	4	3427.3	3446.6
		Mean and std dev	5	3421.5*	3445.7*

* Denotes minimum value.

is adopted. Both the conditional mean and standard deviation are varied for all three (untransformed) temperature variables. For the transformed wind speed and radiation, only the conditional mean is varied with precipitation occurrence (i.e., the standard deviation is held fixed). This common conditional standard deviation, say $\sigma_D(k) \equiv \sigma_0(k) = \sigma_1(k)$, can be estimated directly through construction of a randomly normalized variable. Given a daily precipitation occurrence state $J_t = i$, one can define the variable

$$D_t(k) = X_t(k) - \mu_i(k). \quad (8)$$

It follows that the standard deviation of the $D_t(k)$ process is $\sigma_D(k)$. When constructing the $Z_t(k)$ time series, this common standard deviation is substituted into the denominator of (3). As will be seen in section 4b, constraining the conditional standard deviation of a transformed variable is not as restrictive as it might appear to be.

c. Fitted model

The parameter estimates of the extended version of the Richardson model are presented for Eugene in January. Using the so-called transition counts (e.g., Katz 1977), the estimated transition probabilities for daily precipitation occurrence [(1)] are

$$\hat{P}_{01} = 0.303 \quad \text{and} \quad \hat{P}_{11} = 0.748 \quad (9)$$

(i.e., the conditional probability of a wet day is about 0.45 higher, given a wet day).

Table 3 lists the estimated conditional mean and standard deviation given precipitation occurrence, denoted by $\hat{\mu}_i(k)$ and $\hat{\sigma}_i(k)$, respectively, for the other five daily weather variables for Eugene in January. The mean is higher on wet days than on dry days for the three temperature variables. In the case of maximum temperature, this result is not typical of other seasons or for other regions (e.g., for Eugene in July its mean on wet days is considerably lower, 21.4° vs 28.4°C). The conditional standard deviation of the three temperature variables always is higher when the conditional mean is smaller. For wind speed and radiation, the conditional mean and common conditional standard deviation listed in Table 3 are based on the square-root-transformed data.

The symmetric matrix of estimated cross correlations at lag zero [(6)] for Eugene in January, denoted by $\hat{\mathbf{M}}_0$, is given by

$$\hat{\mathbf{M}}_0 = \begin{bmatrix} 1.000 & 0.731 & 0.827 & 0.104 & 0.077 \\ & 1.000 & 0.909 & 0.196 & -0.266 \\ & & 1.000 & 0.031 & -0.263 \\ & & & 1.000 & 0.039 \\ & & & & 1.000 \end{bmatrix}. \quad (10)$$

In recollection of the earlier discussion about minimum temperature sometimes being employed as a substitute for dewpoint, the highest contemporaneous cross correlation is between these two variables (about 0.91). All the contemporaneous cross correlations between square root–transformed wind speed and the other variables are relatively small, the largest being with minimum temperature (about 0.20), which perhaps is consistent with higher winds lessening the effects of radiational cooling

TABLE 3. Estimated conditional mean and standard deviation for Eugene in Jan.

k	Variable	$\hat{\mu}_0(k)$	$\hat{\mu}_1(k)$	$\hat{\sigma}_0(k)$	$\hat{\sigma}_1(k)$
1	Maximum temperature (°C)	6.85	8.91	4.51	3.72
2	Minimum temperature (°C)	−1.44	2.73	4.29	3.94
3	Dewpoint (°C)	−0.04	4.15	4.94	3.60
4	Wind speed (m s^{-1})$^{1/2}$	1.67	2.10	0.414	0.414
5	Radiation (W day m^{-2})$^{1/2}$	38.3	31.9	6.64	6.64

[note that Hanson and Johnson (1998) found that correlations are higher between wind speed and the difference in maximum temperature between the current and previous day for some sites in complex terrain]. The estimated contemporaneous cross correlations for Eugene differ substantially with the season (e.g., the correlation between minimum temperature and dewpoint drops to only 0.69 in July), consistent with the findings of Richardson (1982).

The corresponding matrix of estimated cross correlations at lag one day [(7)] for Eugene in January, denoted by $\hat{\mathbf{M}}_1$, is given by

$$\hat{\mathbf{M}}_1 = \begin{bmatrix} 0.632 & 0.569 & 0.622 & -0.034 & -0.072 \\ 0.568 & 0.679 & 0.692 & 0.017 & -0.231 \\ 0.616 & 0.666 & 0.743 & -0.115 & -0.226 \\ 0.026 & 0.068 & -0.021 & 0.509 & 0.058 \\ -0.014 & -0.106 & -0.133 & 0.149 & 0.333 \end{bmatrix}. \quad (11)$$

The highest first-order autocorrelation [i.e., diagonal elements in (11)] is for dewpoint (about 0.74); the smallest is for square root–transformed radiation (about 0.33). The highest cross correlation [i.e., off-diagonal elements in (11)] is for dewpoint leading minimum temperature (about 0.69). Consistent with Richardson (1982), the seasonal dependence of the autocorrelations is less substantial than that of the contemporaneous cross correlations. Still, differences as great as 0.51 in July versus 0.68 in January arise for the estimated first-order autocorrelation of minimum temperature.

The matrix of estimated autoregression parameters for Eugene in January, denoted by $\hat{\mathbf{A}}$, is obtained through (5):

$$\hat{\mathbf{A}} = \begin{bmatrix} 0.438 & 0.179 & 0.093 & -0.116 & -0.029 \\ 0.091 & 0.353 & 0.280 & -0.068 & -0.068 \\ 0.104 & 0.103 & 0.553 & -0.161 & -0.055 \\ -0.021 & 0.014 & -0.023 & 0.507 & 0.038 \\ -0.030 & -0.002 & -0.027 & 0.141 & 0.322 \end{bmatrix}. \quad (12)$$

For instance, the equation governing the stochastic model of the randomly standardized maximum temperature variable [i.e., one of the components in (4)] is of the form

$$Z_t(1) = 0.438Z_{t-1}(1) + 0.179Z_{t-1}(2) + 0.093Z_{t-1}(3) - 0.116Z_{t-1}(4) - 0.029Z_{t-1}(5) + \epsilon_t(1). \quad (13)$$

In other words, among the five standardized weather variables on the previous day, the most weight is assigned to maximum temperature, the least (in absolute value) to radiation. In (13), $\epsilon_t(1)$ is the error term, a random variable whose distribution is next considered.

The symmetric estimated variance–covariance matrix for the error terms of the multiple AR(1) process, denoted by $\hat{\mathbf{S}}$, is also obtained through (5):

$$\hat{\mathbf{S}} = \begin{bmatrix} 0.557 & 0.292 & 0.349 & 0.143 & 0.142 \\ & 0.501 & 0.387 & 0.214 & -0.158 \\ & & 0.425 & 0.118 & -0.134 \\ & & & 0.739 & -0.051 \\ & & & & 0.868 \end{bmatrix}. \quad (14)$$

For instance, the error term that appears in (13) has a normal distribution with a mean of zero and a variance [i.e., one of the diagonal elements in (14)] of 0.557; that is, $\epsilon_t(1) \sim N(0, 0.557)$. This error term is contemporaneously correlated with the error terms for the four remaining variables, with the highest correlation being about 0.72 with the one for dewpoint [i.e., converting the off-diagonal elements in (14) from covariances to correlations].

4. Properties and use of model

a. Simulation

One of the primary purposes for the development of the Richardson model and its extensions concerns the production of artificial time series for daily weather variables. The simulation algorithm for the extended version of the Richardson model works in precisely the same fashion as for the conventional model. To run the simulations, the so-called Cholesky (or square root) decomposition of the variance–covariance matrix of the error terms in the multiple AR(1) model [(4)], $\hat{\mathbf{S}} = \mathbf{B}\mathbf{B}^{\mathrm{T}}$, is needed (Graybill 1969). Here the $K \times K$ matrix $\mathbf{B}$ is of lower triangular form.

Using the estimated variance–covariance matrix in (14), the estimated $\mathbf{B}$ matrix for Eugene in January is given by

$$\mathbf{B} = \begin{bmatrix} 0.747 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.391 & 0.590 & 0.000 & 0.000 & 0.000 \\ 0.467 & 0.347 & 0.295 & 0.000 & 0.000 \\ 0.192 & 0.236 & -0.180 & 0.784 & 0.000 \\ 0.190 & -0.393 & -0.293 & -0.061 & 0.767 \end{bmatrix}. \quad (15)$$

To generate error terms with the desired variance–covariance matrix, one sets

$$\boldsymbol{\epsilon}_t = \mathbf{B}\boldsymbol{\epsilon}_t^*, \qquad \boldsymbol{\epsilon}_t^* \sim \mathrm{MVN}(\mathbf{0}, \mathbf{I}), \quad (16)$$

where $\mathbf{I}$ denotes the $K \times K$ identity matrix. Because they are statistically independent, it is straightforward to generate the elements of the $\boldsymbol{\epsilon}_t^*$ vector.

The only remaining detail concerns how to generate the weather variables on the first day (i.e., $t = 1$). The transition probabilities of the Markov chain model for the daily occurrence of precipitation [(1)] can be converted into the corresponding unconditional probability of a wet day (e.g., Katz 1996),

$$\pi = P_{01}/(P_{01} + P_{10}). \tag{17}$$

The precipitation occurrence on the first day then simply is generated from the distribution $\Pr\{J_t = 1\} = \pi$, instead of using the transition probabilities.

The initial state of the multiple AR(1) process also needs to be simulated. In (4), set $\mathbf{Z}_0 = \mathbf{0}$ and generate a random error vector $\boldsymbol{\epsilon}_t \sim \mathrm{MVN}(\mathbf{0}, \mathbf{M}_0)$, that is, using the variance–covariance matrix for the vector of standardized weather variables rather than that for the vector of error terms. In particular, the square root decomposition of the matrix $\mathbf{M}_0$ would need to be determined, in the same manner as was just illustrated for the matrix $\mathbf{S}$.

b. Derived statistics

Both to help in understanding how a stochastic weather generator functions and for certain applications, it often is necessary to consider statistics other than those explicitly modeled. As has been seen, the Richardson model and its extensions are formulated in a conditional form that is convenient for performing simulation studies. The corresponding unconditional form can be preferable, for instance, in adjusting the model parameters to obtain scenarios of climate change with the desired statistical characteristics (Katz 1996).

In this extended version of the Richardson model, both wind speed and radiation have been transformed nonlinearly before being modeled statistically. The question naturally arises of how to determine the corresponding statistical properties for the original, untransformed variables. The square root transformation is applied to daily mean wind speed in both January and July and to daily total solar radiation in January for Eugene. In this case, let the original variable be denoted by $Y_t(k) = X_t^2(k)$. The modeled conditional mean and variance of the untransformed variable are related to the corresponding conditional mean and common conditional variance, $\sigma_D^2(k)$ [see (8)], of the square root–transformed variable by

$$E[Y_t(k)\,|\,J_t = i] = \mu_i^2(k) + \sigma_D^2(k), \quad \text{and}$$
$$\mathrm{Var}[Y_t(k)\,|\,J_t = i] = 2\sigma_D^2(k)[\sigma_D^2(k) + 2\mu_i^2(k)],$$
$$i = 0, 1, \tag{18}$$

respectively (e.g., Katz and Garrido 1994; Katz 1999). Analogous relationships hold for other transformations, such as the reflected logarithm applied to radiation in July.

To take wind speed at Eugene in January as an example, the modeled conditional mean of the original variable [(18)] is 2.95 and 4.60 m s^{-1} on dry and wet days, respectively. These values virtually are identical to the observed values of 2.96 and 4.59. The modeled conditional standard deviation for the untransformed variable [(18)] is 1.40 and 1.76 m s^{-1} on dry and wet days, respectively. Despite the constraint on the conditional standard deviation of transformed wind speed, these values compare favorably to those observed of 1.48 and 1.72, with the square root transformation capturing the effect of higher mean and variance on wet days.

For radiation for Eugene in January, the modeled conditional mean of the original variable is 1.51 and 1.06 kW day m^{-2} on dry and wet days, respectively, likewise essentially the same as the observed values of 1.52 and 1.06. The modeled conditional standard deviation is 0.51 and 0.43 kW day m^{-2} on dry and wet days, respectively, somewhat smaller in range than that observed (0.54 and 0.39), but still producing the effect of higher mean and variance on dry days. If the conditional standard deviation of transformed radiation were permitted to vary between dry and wet days, then this observed range in standard deviations for the untransformed variable effectively would be reproduced (i.e., modeled values of 0.55 and 0.40).

Next, the derivation of unconditional statistics for the daily weather variables is treated. For Eugene in January, the estimated transition probabilities give an unconditional probability of a wet day π [(17)] of 0.546. For the other daily weather variables that are not transformed (i.e., maximum and minimum temperature and dewpoint), the unconditional mean and variance, $\mu(k) = E[X_t(k)]$, and $\sigma^2(k) = \mathrm{Var}[X_t(k)]$, are related to the conditional mean and variance by (Katz 1996)

$$\mu(k) = (1 - \pi)\mu_0(k) + \pi\mu_1(k), \quad \text{and}$$
$$\sigma^2(k) = (1 - \pi)\sigma_0^2(k) + \pi\sigma_1^2(k)$$
$$+ \pi(1 - \pi)[\mu_1(k) - \mu_0(k)]^2. \tag{19}$$

That is, the unconditional mean is a weighted average of the two conditional means, whereas the unconditional variance is *not* simply a weighted average of the two conditional variances but includes another term reflecting the variation in the conditional means. In the case of transformed variables, the same sort of formulation in (19) holds for the conditional mean and variance of the untransformed variables (whose determination was discussed previously). For wind speed for Eugene in January, (19) yields an unconditional mean of 3.85 m s^{-1} and an unconditional standard deviation of 1.80 m s^{-1} (nearly the same as the observed values of 3.84 and 1.81, respectively). For radiation in January, the corresponding unconditional mean and standard deviation are 1.27 and 0.52 kW day m^{-2} (the same as the observed values).

The unconditional (or marginal) distribution of a daily weather variable in the Richardson model is a mixture of two conditional distributions. For the three temperature variables, this mixture consists of two conditional normal distributions. In the case of the square root transformation to normality, the unconditional probability

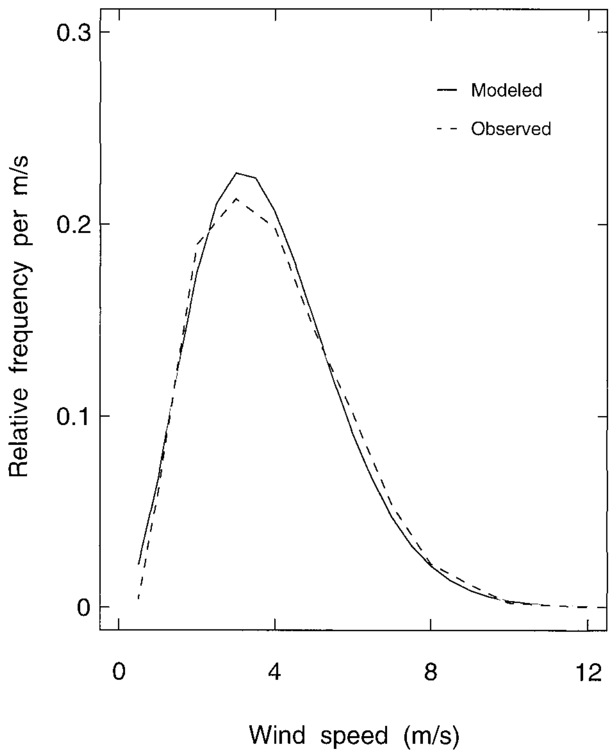

FIG. 4. Modeled and observed unconditional distribution of daily mean wind speed (m s^{-1}) for Eugene in Jan.

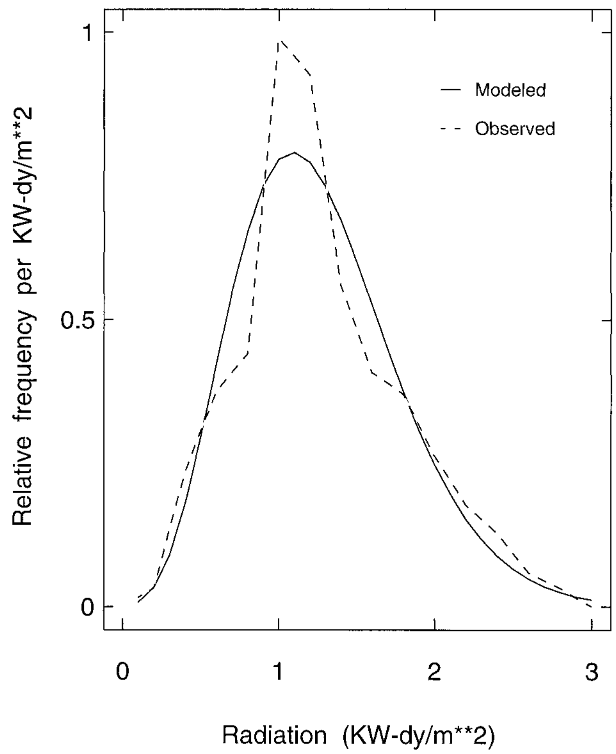

FIG. 5. Same as Fig. 4 but for daily total solar radiation (kW day m^{-2}).

density function of the untransformed variable can be expressed as (e.g., Katz and Garrido 1994)

$$f(y; k) = (1 - \pi) f_0[y; \mu_0(k), \sigma_D^2(k)] + \pi f_1[y; \mu_1(k), \sigma_D^2(k)],$$
$$f_i[y; \mu_i(k), \sigma_D^2(k)] = [(1/2)y^{-1/2}/\sigma_D(k)]\phi\{[y^{1/2} - \mu_i(k)]/\sigma_D(k)\}, \quad i = 0, 1. \quad (20)$$

Here ϕ denotes the standard normal $N(0, 1)$ density function. The application of (20) to produce modeled unconditional distributions of the original, untransformed variables is illustrated for wind speed (Fig. 4) and radiation (Fig. 5) for Eugene in January along with a comparison with the corresponding observed distributions. In both cases, the square root transformation adequately reflects the observed positive skewness, with the fit being somewhat closer for wind speed than for radiation. Note that the GEM weather generator (Hanson and Johnson 1998) would represent these distributions as being approximately normal (technically, a mixture of two normals).

Conditional distributions of daily mean wind speed and total radiation, given either the value of the same variable on the previous day or of another weather variable on the same day, also are considered, demonstrating the flexibility of the current modeling approach. In principle, the modeled conditional distributions could be determined analytically in a similar but somewhat more complex manner as that for the unconditional distributions. Instead, the simulation methodology described in section 4a is used. It is anticipated that the observed conditional distributions of wind speed and radiation will have a tendency for higher variability, the higher the mean (or median) is. Without the application of a power transformation, the modeled conditional distributions essentially would be normal, with a mean that is related linearly to (and a variance that is independent of) the value of the variable on which it is conditioned.

Figures 6–9 show certain statistics of the modeled and observed conditional distributions of daily mean wind speed and total solar radiation for Eugene in January. For simplicity, only three quartiles (i.e., lower quartile, median, and upper quartile) of the conditional distributions are plotted (with curves indicating the model statistics and points for the observations). In particular, Figs. 6 and 7 give the conditional distributions of wind speed (radiation) given wind speed (radiation) on the previous day. It is evident that the power transformation technique captures the tendency of the observed conditional distributions to be more variable (i.e., as measured by the interquartile range), the higher the median is. Figures 8 and 9 give the conditional distributions of wind speed (radiation) given minimum temperature on the same day. Again the relationship of higher variability

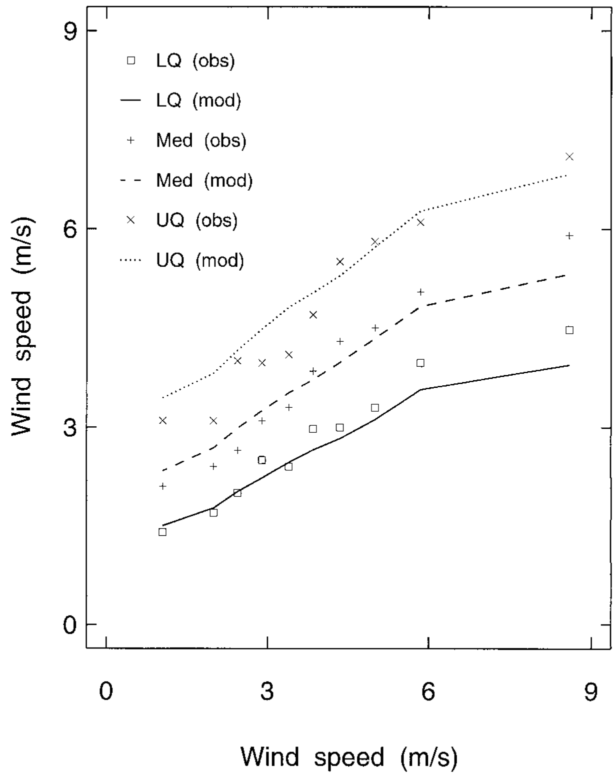

FIG. 6. Modeled (mod) and observed (obs) conditional distribution of daily mean wind speed (m s^{-1}) given wind speed on previous day for Eugene in Jan. Lower quartile = LQ, median = Med, and upper quartile = UQ.

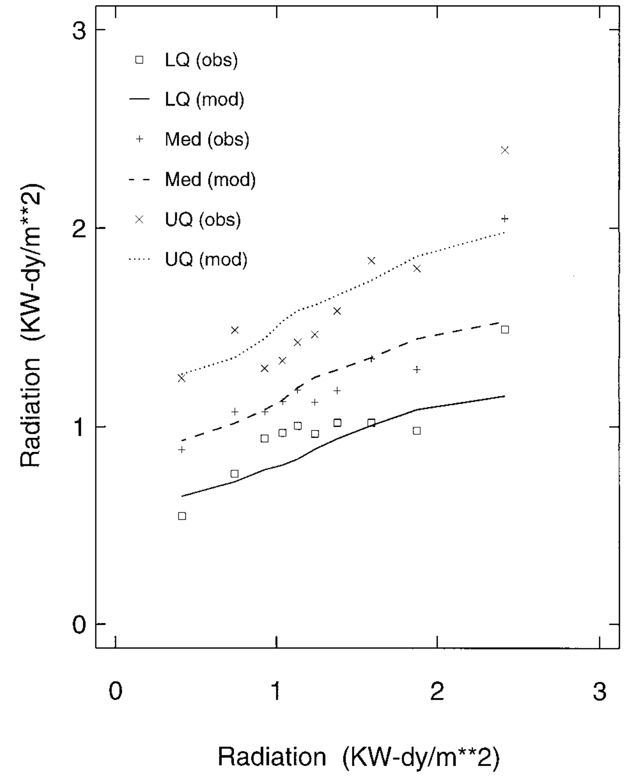

FIG. 7. Same as Fig. 6 but for daily total solar radiation (kW day m^{-2}) given radiation on previous day.

with higher median evident in the observations is captured by the model, with the somewhat nonlinear relationships between the conditional quantiles of both wind speed and radiation and the given value of minimum temperature being reasonably represented as well. Note that the approach to modeling wind speed and radiation in GEM (Hanson and Johnson 1998) would be incapable of representing such relationships.

5. Discussion

An extended version of the Richardson model/WGEN for time series of daily weather variables has been presented. The key to the extension is the transformation of variables that are not normally distributed. In this way, the serial and cross correlations with the other variables still are taken into account. Among other things, the flexibility of the technique in allowing for relationships between variables in which the variance is not constant, as well as for certain forms of nonlinearity, is demonstrated. The approach is in the same spirit as the original Richardson model, in that much of what is known about the statistical features of individual weather variables has been integrated into a single multivariate model.

This extended form of the Richardson model has been applied to weather data in the Pacific Northwest. As such, the specific results obtained cannot necessarily be generalized. Viewed as a general prescription for the development of stochastic weather generators, however, the same approach could be applied to other regions and to other weather variables. For example, the technique of transforming to achieve approximate normality still should work, but the form of transformation employed might differ.

Remaining issues include the treatment of the annual cycle by means of a Fourier series approach as in WGEN. For solar radiation, the shape of the distribution is such that a different type of transformation is required depending on the time of year in the Pacific Northwest, making the Fourier series approach problematic. Still, we conjecture that the simplification of applying the square root transformation at all times of the year would constitute an improvement over the current assumption of normality in WGEN. Moreover, this issue of seasonal dependence in the form of transformation may not arise in regions without any seasons that are extremely dry.

Some fundamental limitations to stochastic weather generators remain. In particular, such models sometimes have a marked tendency to underestimate the interannual variance of monthly, seasonal, or annual mean temperature and total precipitation (e.g., Hanson and Johnson 1998; Semenov et al. 1998). To what

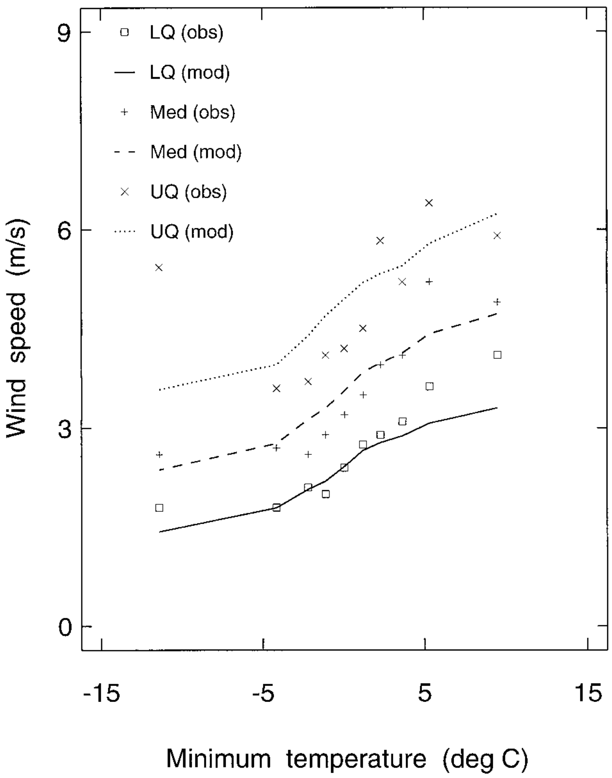

FIG. 8. Modeled and observed conditional distribution of daily mean wind speed (m s^{-1}) given minimum temperature (°C) for Eugene in Jan.

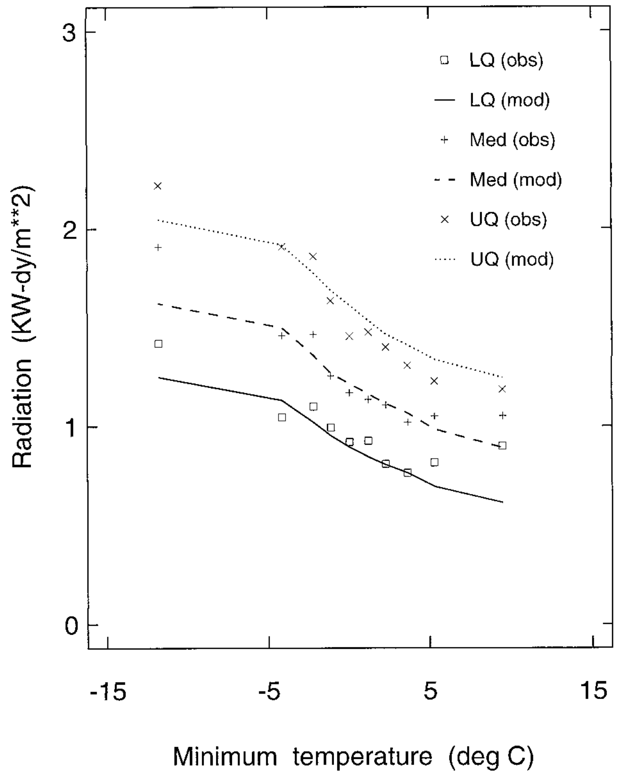

FIG. 9. Same as Fig. 8 but for daily total solar radiation (kW day m^{-2}).

extent this variance underestimation is attributable to an oversimplified model for high-frequency variations or to the lack of any provision for low-frequency variations is a topic of ongoing research (Katz and Parlange 1998; Katz and Zheng 1999). Because most impact assessments of climate change involve the weather across a region, another issue concerns the fact that most weather generators such as WGEN make no provision for the spatial dependence of daily weather variables. Some exceptions include recent work by Hutchinson (1995) and Wilks (1998, 1999).

Acknowledgments. We thank Sue Ferguson, Larry McDaniel, and Linda Mearns both for advice and for providing weather data, and also gratefully acknowledge the comments of two anonymous reviewers. This research was supported partially by NSF Grant DEB-9523611 through special funds for Methods and Models for Integrated Assessment.

APPENDIX

Model Identification Procedure

The details of the model identification procedure are similar to those in Katz and Parlange (1995). For purposes of model selection, each of the five weather variables $X_t(k)$, $k = 1, 2, \ldots, 5$, in the generalized Richardson model is treated separately as a univariate AR(1) model. The parameters to be estimated include the first-order autocorrelation coefficient $\rho_{kk}(1)$, and the conditional mean $\mu_i(k)$ and variance $\sigma_i^2(k)$, $i = 0, 1$, which possibly depend on precipitation occurrence. AIC and BIC are of the form

$$\mathrm{AIC} = -2 \ln L + 2m \quad \text{and} \quad \mathrm{BIC} = -2 \ln L + m \ln n. \tag{A1}$$

Here the first term in both AIC and BIC, involving the maximized log likelihood function $\ln L$, is a measure of the goodness of fit of the model. The second term, depending on the number of model parameters m that must be estimated and, in the case of BIC, the number of observations n, is a penalty for the complexity of the model.

For the three candidate models listed in section 3b, the details concerning the calculation of AIC and BIC are as follows.

a. Model (i)

In this case, m is equal to 3 [i.e., the parameters $\mu(k) \equiv \mu_0(k) = \mu_1(k)$, $\sigma^2(k) \equiv \sigma_0^2(k) = \sigma_1^2(k)$, and $\rho_{kk}(1)$ need to be estimated], and the goodness-of-fit component in (A1) can be expressed as

$$-2 \ln L \approx n \ln\{\hat{\sigma}^2(k)[1 - \hat{\rho}_{kk}^2(1)]\}. \quad \text{(A2)}$$

Here $\hat{\sigma}^2(k)$ is the variance of the weather variable for the entire sample (i.e., both dry and wet days). For this model (as well as for the other two), $\hat{\rho}_{kk}(1)$ is the sample first-order autocorrelation coefficient of the randomly standardized time series $Z_t(k)$ [see (3) and (7)].

b. Model (ii)

In this case, m is equal to 4 [i.e., the parameters $\mu_0(k)$, $\mu_1(k)$, $\sigma_D^2(k) \equiv \sigma_0^2(k) = \sigma_1^2(k)$, and $\rho_{kk}(1)$ need to be estimated], and the goodness-of-fit component in (A1) is of the same form as (A2) for model (i), except for $\hat{\sigma}_D^2(k)$ being substituted in place of $\hat{\sigma}^2(k)$. Here $\hat{\sigma}_D^2(k)$ is the sample variance of the randomly normalized times series $D_t(k)$ [see (8)].

c. Model (iii)

In this case, m is equal to 5 [i.e., the parameters $\mu_0(k)$, $\mu_1(k)$, $\sigma_0^2(k)$, $\sigma_1^2(k)$, and $\rho_{kk}(1)$ need to be estimated], and the goodness-of-fit component in (A1) can be expressed as

$$-2 \ln L \approx n_0 \ln\{\hat{\sigma}_0^2(k)[1 - \hat{\rho}_{kk}^2(1)]\} + n_1 \ln\{\hat{\sigma}_1^2(k)[1 - \hat{\rho}_{kk}^2(1)]\}. \quad \text{(A3)}$$

Here, $\hat{\sigma}_0^2(k)$ is the variance of the weather variable for the subsample consisting of n_0 dry days, and $\hat{\sigma}_1^2(k)$ is the variance for the n_1 wet days.

REFERENCES

Brockwell, P. J., and R. A. Davis, 1991: *Time Series: Theory and Methods.*, 2d ed. Springer-Verlag, 577 pp.

Brown, B. G., R. W. Katz, and A. H. Murphy, 1984: Time series models to simulate and forecast wind speed and wind power. *J. Climate Appl. Meteor.,* **23,** 1184–1195.

Bruhn, J. A., W. E. Fry, and G. W. Fick, 1980: Simulation of daily weather data using theoretical probability distributions. *J. Appl. Meteor.,* **19,** 1029–1036.

Brutsaert, W., 1982: *Evaporation into the Atmosphere.* Reidel, 299 pp.

Carlin, J., and J. Haslett, 1982: The probability distribution of wind power from a dispersed array of wind turbine generators. *J. Appl. Meteor.,* **21,** 303–313.

Easterling, W. E., N. J. Rosenberg, M. S. McKenney, C. A. Jones, P. T. Dyke, and J. R. Williams, 1992: Preparing the erosion productivity impact calculator (EPIC) model to simulate crop response to climate change and the direct effects of CO_2. *Agric. For. Meteor.,* **59,** 17–34.

Graybill, F. A., 1969: *Introduction to Matrices with Applications in Statistics.* Wadsworth, 372 pp.

Hansen, J. E., and D. M. Driscoll, 1977: A mathematical model for the generation of hourly temperatures. *J. Appl. Meteor.,* **16,** 935–948.

Hanson, C. L., and G. L. Johnson, 1998: GEM (Generation of weather Elements for Multiple applications): Its application in areas of complex terrain. *Hydrology, Water Resources and Ecology in Headwaters,* K. Kovar, U. Tappeiner, N. E. Peters, and R. G. Craig, Eds., International Association of Hydrological Sciences (IAHS) Press, 27–32.

Haslett, J., and A. E. Raftery, 1989: Space–time modeling with long-memory dependence: Assessing Ireland's wind power resource (with discussion). *Appl. Stat.,* **38,** 1–50.

Hayhoe, H. N., 1998: Relationship between weather variables in observed and WXGEN generated data series. *Agric. For. Meteor.,* **90,** 203–214.

Hutchinson, M. F., 1995: Stochastic space–time weather models from ground-based data. *Agric. For. Meteor.,* **73,** 237–264.

Johnson, G. L., C. L. Hanson, S. P. Hardegree, and E. B. Ballard, 1996: Stochastic weather simulation: Overview and analysis of two commonly used models. *J. Appl. Meteor.,* **35,** 1878–1896.

Jolliffe, I. T., and P. B. Hope, 1996: Bounded bivariate distributions with nearly normal marginals. *Amer. Stat.,* **50,** 17–20.

Katz, R. W., 1977: Precipitation as a chain-dependent process. *J. Appl. Meteor.,* **16,** 671–676.

——, 1996: Use of conditional stochastic models to generate climate change scenarios. *Climatic Change,* **32,** 237–255.

——, 1999: Moments of power transformed time series. *Environmetrics,* **10,** 301–307.

——, and J. Garrido, 1994: Sensitivity analysis of extreme precipitation events. *Int. J. Climatol.,* **14,** 985–999.

——, and M. B. Parlange, 1995: Generalizations of chain-dependent processes: Application to hourly precipitation. *Water Resour. Res.,* **31,** 1331–1341.

——, and ——, 1998: Overdispersion phenomenon in stochastic modeling of precipitation. *J. Climate,* **11,** 591–601.

——, and X. Zheng, 1999: Mixture model for overdispersion of precipitation. *J. Climate,* **12,** 2528–2537.

Kimball, J. S., S. W. Running, and R. Nemani, 1997: An improved method of estimating surface humidity from daily minimum temperature. *Agric. For. Meteor.,* **85,** 87–98.

Mearns, L. O., C. Rosenzweig, and R. Goldberg, 1997: Mean and variance change in climate scenarios: Methods, agricultural applications, and measures of uncertainty. *Climatic Change,* **35,** 367–396.

National Renewable Energy Laboratory, 1992: User's Manual: 1961–1990 National Solar Radiation Data Base, Version 1.0. NSRDB-Vol. 1. 243 pp. [Available from NREL Document Distribution Service, 1617 Cole Blvd., Golden, CO 80401-3393.]

Neilson, R. P., 1995: A model for predicting continental-scale vegetation distribution and water balance. *Eco. Appl.,* **5,** 362–385.

Nicks, A. D., C. W. Richardson, and J. R. Williams, 1990: Evaluation of the EPIC model weather generator. Erosion/Productivity Impact Calculator, 1. Model Documentation, A. N. Sharpley and J. R. Williams, Eds., USDA-ARS Tech. Bull. 1768, 235 pp. [Available from U.S. Government Printing Office, Washington, DC 20402.]

Rajagopalan, B., U. Lall, D. G. Tarboton, and D. S. Bowles, 1997: Multivariate nonparametric resampling scheme for generation of daily weather variables. *Stochastic Hydrol. Hydraulics,* **11,** 65–93.

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

——, 1982: Dependence structure of daily temperature and solar radiation. *Trans. Amer. Soc. Agric. Eng.,* **25,** 735–739.

——, and D. A. Wright, 1984: WGEN: A model for generating daily weather variables. USDA Publication ARS-8, 83 pp. [Available from National Technical Information Center, 5285 Port Royal Rd., Springfield, VA 22161.]

Semenov, M. A., R. J. Brooks, E. M. Barrow, and C. W. Richardson, 1998: Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates. *Climate Res.,* **10,** 95–107.

Todorovic, P., and D. A. Woolhiser, 1975: A stochastic model of *n*-day precipitation. *J. Appl. Meteor.,* **14,** 17–24.

Wallis, T. W. R., and J. F. Griffiths, 1995: An assessment of the weather generator (WXGEN) used in the erosion/productivity impact calculator (EPIC). *Agric. For. Meteor.,* **73,** 115–133.

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

——, 1998: Multisite generalization of a daily stochastic precipitation generation model. *J. Hydrol.,* **210,** 178–191.

——, 1999: Simultaneous stochastic simulation of daily precipitation, temperature, and solar radiation at multiple sites in complex terrain. *Agric. For. Meteor.,* **96,** 85–101.

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