WATER RESOURCES RESEARCH, VOL. 29, NO. 7, PAGES 2335–2344, JULY 1993

Effects of an Index of Atmospheric Circulation on Stochastic Properties of Precipitation

RICHARD W. KATZ

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

MARC B. PARLANGE

Hydrologic Science, Department of Land, Air and Water Resources and Department of Agricultural Engineering, University of California, Davis

Conditional chain-dependent processes are fit to time series of daily precipitation amounts given an index of large-scale atmospheric circulation (i.e., either below or above normal monthly mean sea level pressure). Precipitation data for January at several locations in California are analyzed. The two conditional daily models differ both in terms of the parameters of the occurrence process and of the intensity process. Each of these daily effects contributes to changes in the distribution of monthly total precipitation associated with the circulation index. The process induced by combining the conditional daily models produces a variance for monthly total precipitation much closer to the observed value than that for the corresponding unconditional chain-dependent model.

1. INTRODUCTION

Much research has dealt with the relationship between large-scale atmospheric circulation and precipitation [e.g., *Cayan and Peterson*, 1989; *Glantz et al.*, 1991]. Most of this work has focused on precipitation totaled over relatively long time periods, such as a month, season, or year. Of course, for many hydrologic applications, information about the statistical characteristics of precipitation over much shorter time periods (e.g., days) is needed. Recently, some effort has been devoted to the problem of examining the relationship between circulation and time series of daily precipitation amounts. Usually, an empirical approach has been employed, in which the conditional statistical characteristics of daily precipitation, given a pattern of atmospheric circulation, are estimated. This research has been successful in demonstrating that the effects of circulation can be detected in not just monthly but daily precipitation statistics [*Bardossy and Plate*, 1991, 1992; *Wilson et al.*, 1991; *Woolhiser et al.*, 1993].

Like the traditional approach in hydrology of fitting (what might be termed in the present context "unconditional") stochastic models to time series of precipitation amounts, the research just cited has involved fitting conditional stochastic models to the same time series. However, little attention has been paid to how conditional models differ from the conventional, unconditional models [one exception is *Wilks*, 1989]. In particular, are these conditional models consistent with the unconditional one, or do they "induce" a different (possibly improved) overall model for the daily precipitation process? Further, how do these results concerning the conditional statistical characteristics of precipitation on a daily time scale compare with those originally obtained for, say, a monthly scale?

In the present paper we focus on these specific issues. A simplified index of large-scale atmospheric circulation (i.e., either below or above normal monthly mean sea level pressure) is used, known to be related to monthly or seasonal total precipitation [*Cayan and Peterson*, 1989]. A stochastic model for time series of daily precipitation, a chain-dependent process [*Katz*, 1977*a*, *b*], is applied whose theoretical properties are well established. In this way the effects on precipitation of conditioning on circulation can be studied in a relatively straightforward manner. As an application, time series of daily precipitation for a collection of sites within the state of California are analyzed, a region where water resources management has received much attention. The approach taken is very similar to that of *Wilks* [1989], except that he conditioned on the monthly total precipitation itself rather than on atmospheric circulation. Finally, this simple form of circulation index is especially of interest because monthly mean pressure could be potentially forecast several months in advance, whereas precipitation itself is difficult to forecast directly even a few days ahead.

2. SPECIFICATION OF MONTHLY PRECIPITATION

The nature of the relationship between patterns in large-scale atmospheric circulation and total precipitation (e.g., over a month or season) has been extensively studied for the western United States, in particular for the state of California during the winter wet season [*Cayan and Peterson*, 1989]. Some of the circulation measures that are most popular in teleconnection studies, such as the El Niño/Southern Oscillation phenomenon [*Glantz et al.*, 1991] or the Pacific–North Atlantic pattern [*Wallace and Gutzler*, 1981], have a weak link at best to wintertime precipitation over most of California [*Redmond and Koch*, 1991]. Nevertheless, patterns in atmospheric circulation closer to California (e.g., over the adjacent Pacific Ocean) do have a substantial ability to specify contemporaneous precipitation over much of the state [*Cayan and Peterson*, 1989; *Weare and Hoeschele*, 1983].

The mean monthly sea level pressure (SLP) at a single grid point (i.e., 40°N, 130°W) located in the Pacific Ocean off the

Paper number 93WR00569.
0043-1397/93/93WR-00569$05.00

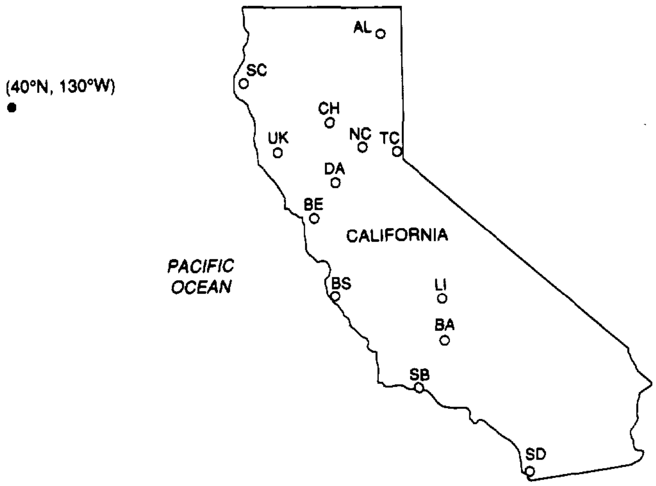

Fig. 1. Geographical location of the 13 stations (open circles) in California for which precipitation data are analyzed and of the grid point (solid circle) for sea level pressure on which the index of atmospheric circulation is based.

coast of California (Figure 1) is utilized to construct an index of atmospheric circulation [*Jenne*, 1975; *Trenberth and Paolino*, 1980]. The single month of January, the peak of the wet season in California, is treated. Formally, the index I is defined as follows:

$$I = 0 \quad \text{if monthly SLP} < \text{mean},$$
$$I = 1 \quad \text{if monthly SLP} > \text{mean}. \qquad (1)$$

Here "mean" refers to the long-term sample mean of the pressure record, and the January SLP averages 1018.6 mbar over the time period 1899–1988. We stress that this index (1) does not vary within a given month. Although such an index has the disadvantage of not using all of the temporal information about pressure variations, it has the advantage of being in a form that could be forecast (e.g., by use of teleconnections).

The work of *Cayan and Peterson* [1989] indicates that this index should have a significant, albeit imperfect, relationship with monthly total precipitation over much of California. They also provide some physical explanations related to regional teleconnection patterns to explain the existence of such a relationship. We analyze time series of daily and monthly precipitation amounts at 13 locations across California (Figure 1). These stations are selected both to cover a wide range of climate regimes and to have a relatively long historical record (i.e., ranging from 40 to 82 years (see Table 1)). In the present section, only the effects of circulation on monthly precipitation are considered. In section 3, how the same monthly circulation index is related to daily precipitation amounts is examined.

The mean and standard deviation of the unconditional and conditional distributions of January total precipitation for the 13 locations are summarized in Table 1. The unconditional mean (standard deviation) ranges from 25.3 mm (18.7 mm) at Bakersfield to 260.9 mm (151.8 mm) at Nevada City. The conditional monthly mean is always greater (usually a substantial difference) when $I = 0$ (i.e., lower than normal pressure). The conditional monthly standard deviation is also virtually always greater when $I = 0$, although the

TABLE 1. Descriptive Statistics for Unconditional and Conditional Distributions of January Total Precipitation at 13 Locations in California

	Mean, mm			Standard Deviation, mm		
Station	Unconditional	Conditional $\{I = 0\}$	Conditional $\{I = 1\}$	Unconditional	Conditional $\{I = 0\}$	Conditional $\{I = 1\}$
Alturas	39.1	45.8	33.0	26.6	27.2	25.0
	(53)	(25)	(28)	(53)	(25)	(28)
Bakersfield	25.3	33.7	17.0	18.7	17.4	16.2
	(50)	(25)	(25)	(50)	(25)	(25)
Berkeley	116.1	152.1	89.2	74.4	66.9	68.7
	(70)	(30)	(40)	(70)	(30)	(40)
Big Sur	219.0	280.2	151.3	135.5	123.9	116.2
	(40)	(21)	(19)	(40)	(21)	(19)
Chico	134.8	190.9	84.1	88.6	82.4	58.9
	(78)	(37)	(41)	(78)	(37)	(41)
Davis	86.4	122.3	61.0	58.8	54.4	47.8
	(70)	(29)	(41)	(70)	(29)	(41)
Lindsay	56.3	74.4	36.3	39.5	42.3	24.3
	(40)	(21)	(19)	(40)	(21)	(19)
Nevada City	260.9	344.5	182.9	151.8	134.9	123.8
	(58)	(28)	(30)	(58)	(28)	(30)
Santa Barbara	84.6	115.7	54.9	78.7	80.7	65.3
	(47)	(23)	(24)	(47)	(23)	(24)
San Diego	48.4	55.0	42.6	42.7	43.5	41.7
	(62)	(29)	(33)	(62)	(29)	(33)
Scotia	228.0	297.7	167.7	117.3	108.8	88.2
	(56)	(26)	(30)	(56)	(26)	(30)
Tahoe City	155.4	197.0	119.3	113.0	127.6	85.2
	(56)	(26)	(30)	(56)	(26)	(30)
Ukiah	202.0	279.5	131.7	130.5	127.9	86.0
	(82)	(39)	(43)	(82)	(39)	(43)

Number of years included in parentheses.

differences are not nearly as substantial as those for the mean. For example, at Chico the conditional mean (standard deviation) changes from 190.9 mm (82.4 mm) to 84.1 mm (58.9 mm) depending on whether $I = 0$ or 1. The unconditional mean is just the weighted average of the two conditional means. However, the unconditional standard deviation (or variance) is not simply a weighted average of the two conditional standard deviations (variances). In fact, for many of the locations both of the conditional standard deviations are smaller than the unconditional one. A theoretical explanation for this phenomenon is provided in section 4.

Because the distribution of monthly total precipitation is still positively skewed, it is not complete to only consider the mean and standard deviation. The effect of the circulation index on this distribution is also more directly examined by constructing box plots for each of the conditional distributions of total precipitation given the index I. Figure 2 includes the examples of Chico and Nevada City. Like the mean, the median and lower and upper quartiles are always greater when $I = 0$. The effect on the interquartile range (i.e., the height of the box) is not as clear as for the standard deviation, but it is still greater for most stations when $I = 0$. It might be anticipated that the conditional distribution of total precipitation given $I = 0$ would be less skewed (in conjunction with its higher mean), but this effect is difficult to ascertain from the box plots because of the limited number of years of record.

3. STOCHASTIC MODELS FOR DAILY PRECIPITATION

3.1. *Definition of Model*

Many different types of stochastic processes have been proposed to model the statistical characteristics of time series of daily precipitation amounts. For our purposes it is convenient to adopt a model with a relatively simple structure; namely, the so-called chain-dependent process [*Katz*, 1977*a*, *b*; *Todorovic and Woolhiser*, 1975]. Such a process has the desirable feature of requiring only a relatively small number of parameters, while still accounting for the most important statistical properties of precipitation time series. In particular, it allows for both the tendency of wet and dry spells to persist and for the positive skewness of the distribution of precipitation amounts on wet days.

Basically, the chain-dependent modeling approach involves dividing the precipitation process into two component processes: (1) the occurrence process (i.e., the sequence of dry or wet days) and (2) the intensity process (i.e., the sequence of precipitation amounts on wet days). Formally, the occurrence process $\{J_t : t = 1, 2, \cdots\}$ is defined as

$$J_t = 1 \quad \text{if } t\text{th day is wet,} \qquad J_t = 0 \quad \text{otherwise,} \tag{2}$$

where a "wet day" refers to one on which measurable precipitation occurs. It is assumed that this occurrence process J_t constitutes a two-state, first-order Markov chain [*Gabriel and Neumann*, 1962; *Waymire and Gupta*, 1981].

This stochastic model is completely characterized by the transition probabilities

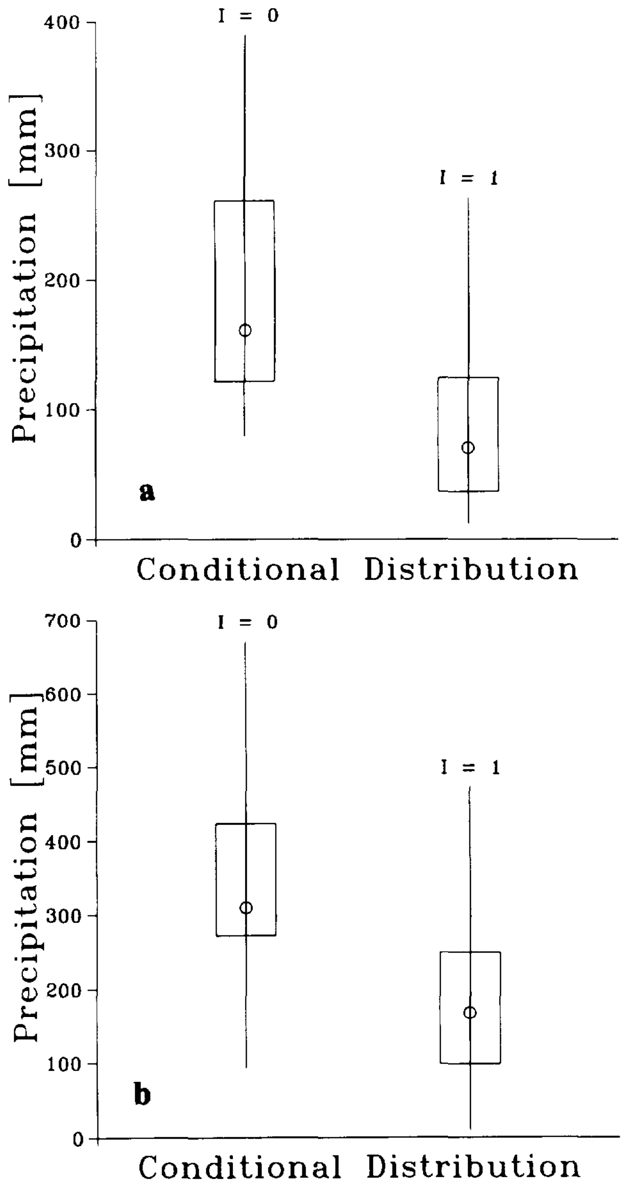

Fig. 2. Box plots of conditional distribution of January total precipitation given circulation index ($I = 0$ or $I = 1$) at (a) Chico and (b) Nevada City. Circles indicate median, top (bottom) of box indicates upper (lower) quartile, and top (bottom) of vertical line indicates maximum (minimum).

$$P_{ij} = \Pr\{J_{t+1} = j | J_t = i\}, \qquad i, j = 0, 1, \tag{3}$$

with the constraints that $P_{i0} + P_{i1} = 1$, $i = 0, 1$. It is convenient to reparameterize the Markov chain in terms of the probability of a wet day and the first-order autocorrelation coefficient (or "persistence parameter") of the J_t process (2), that is,

$$\pi = \Pr\{J_t = 1\}, \qquad \rho = \mathrm{Corr}\,(J_t, J_{t+1}). \tag{4}$$

Now these two parameters, π and ρ, can be expressed as simple functions of the transition probabilities (3) of the Markov chain; specifically,

$$\pi = P_{01}/(P_{01} + P_{10}),$$
$$\rho = P_{11} - P_{01} \quad (5)$$

[e.g., *Katz*, 1985; *Lloyd*, 1974]. This parameterization (5) will enable us to more simply interpret any effects on the transition probabilities of conditioning on the atmospheric circulation index.

The intensity process consists of a sequence of random variables $\{X_k : k = 1, 2, \cdots\}$, where X_k denotes the amount of precipitation on the kth wet day. These random variables are assumed to be independently and identically distributed, say with common mean μ and variance σ^2. This assumption can be interpreted as requiring that the sequence of daily amounts of precipitation is conditionally independent, given the sequence of states of the occurrence process [see *Katz*, 1977*b*]. Although the requirement that the intensities be conditionally independent is reasonable for a daily time scale [*Katz*, 1977*a*], results when this assumption is relaxed are briefly mentioned in section 4.2.

The daily precipitation intensity X_k is assumed to have a power transform distribution [*Box and Cox*, 1964]. That is, a transformation

$$X_k^* = X_k^p, \qquad 0 < p < 1 \quad (6)$$

exists, such that the transformed variable X_k^* has a normal distribution, say with mean μ^* and variance $(\sigma^*)^2$ (written $N[\mu^*, (\sigma^*)^2]$). In general, the mean and variance, μ and σ^2, of the untransformed intensities are complicated functions (whose form depends on the power transform parameter p) of both the mean and variance, μ^* and $(\sigma^*)^2$, of the transformed intensities. R. W. Katz and J. Garrido (Sensitivity of extreme precipitation events to climate change, submitted to *International Journal of Climatology*, 1993) discussed the application of this power transform approach in fitting distributions to monthly and seasonal total precipitation.

3.2. *Results of Model Fitting*

This specific form of chain-dependent process is fit to three different data sets of time series of daily precipitation amounts: (1) an unconditional model for all years, (2) a first conditional model for only years classified as low pressure (i.e., $I = 0$), and (3) a second conditional model for only years classified as high pressure (i.e., $I = 1$). That is, in cases (2) and (3), all of the individual parameters of the chain-dependent process (i.e., π, ρ, μ^*, and $(\sigma^*)^2$) are allowed to vary depending on the circulation index I (defined by (1)). Because preliminary analyses indicated that the power transform parameter p does not depend on the index I, it is constrained to remain the same as in the unconditional model for a given location and time period. As another simplifying approximation, all of the models are taken to be stationary within the month being fit (i.e., the model parameters are held constant from day to day).

The parameters of the chain-dependent process are estimated by the method of maximum likelihood. These estimators are included in the appendix. For the transition probabilities P_{ij} of the Markov chain model for the occurrence process, the maximum likelihood estimators $\hat{P}_{ij}$ are obtained from the transition counts (see (A2)). The appropriate power transform parameter p is identified using a method proposed by *Hinkley* [1977] that is based on a measure of skewness. That is,

$$d_p = (\text{mean} - \text{median})/(\text{standard deviation}), \quad (7)$$

where the mean, median, and standard deviation are calculated for the sample of pth power transformed data (6). Comparing this measure (7) for the trial values of p (i.e., $p = 1/2, 1/4, 1/8, \cdots$), the one that minimizes the degree of skewness (i.e., $d_p \approx 0$) is selected. Then the parameters μ^* and σ^* are estimated by the sample mean and standard deviation, $\bar{x}^*$ and s^* say, of the transformed intensities X_k^*.

Table 2 lists these parameter estimates for the single unconditional and the two conditional chain-dependent processes fit to the time series of daily precipitation amounts in January at the 13 locations in California. To facilitate comparisons, the sample mean and standard deviation of the untransformed intensities X_k, $\bar{x}$ and s say, are also included in the table. For the unconditional models the estimated probability of a wet day π ranges from 0.178 at Bakersfield to 0.484 at Scotia, and the sample mean of daily precipitation intensities ranges from 4.72 mm at Bakersfield to 22.46 mm at Big Sur. The probability π is always estimated to be greater when $I = 0$ (i.e., lower than normal pressure). For example, at Chico the estimates of π range from 0.413 when $I = 0$ to 0.253 when $I = 1$. No clear-cut pattern is apparent for the estimated persistence parameter ρ (see (4)). The selected value of the power transform parameter p is either 1/4 or 1/8. Like the effect on the probability of a wet day, the mean of the transformed intensities μ^* is always estimated to be greater when $I = 0$. For example, at Chico the estimates of μ^* range from 1.777 mm$^{1/4}$ when $I = 0$ to 1.585 mm$^{1/4}$ when $I = 1$. The effect on the standard deviation σ^* is not so obvious, although it is still more frequently estimated to be larger when $I = 0$.

It remains to identify which of the individual parameters of the chain-dependent process actually should be allowed to vary with the circulation index in order to produce the best fit to the time series of daily precipitation amounts. To make this comparison, the maximized likelihood function is obtained for different combinations of varying parameters. These candidate models are listed in Table 3. The techniques by which these likelihood functions can be obtained are described in the appendix. The maximized likelihood function can be viewed as a goodness of fit measure. To identify the appropriate form of model, this goodness of fit term should be penalized for the number of parameters required to be estimated by a given model. Two model selection criteria are employed, Akaike's information criterion (AIC) [*Akaike*, 1974] and the Bayesian information criterion (BIC) [*Schwarz*, 1978]. These criteria are defined as follows:

$$\text{AIC}(m) = -2 \ln M_m + 2k_m, \quad (8)$$

$$\text{BIC}(m) = -2 \ln M_m + k_m \ln n, \quad (9)$$

$m = 1, 2, \cdots$. Here M_m denotes the maximized likelihood function for the mth model, a model that requires the estimation of k_m parameters from a time series consisting of a total of n daily precipitation amounts. The model m is selected for which AIC(m) (or BIC(m)) is minimized.

Table 3 summarizes the results of the application of the AIC and BIC to the various forms of conditional stochastic models for daily precipitation amounts in January at the 13 sites in California. The AIC never chooses either the unconditional chain-dependent process or the fully conditional

TABLE 2. Estimated Parameters of Unconditional and Conditional Chain-Dependent Processes for Time Series of Daily Precipitation Amounts in January at 13 Locations in California

	Occurrence Process		Intensity Process		Transformed Intensity Process		
Station	π	ρ	μ, mm	σ, mm	p	μ^*, mm^p	σ^*, mm^p
Alturas							
Unconditional	0.271	0.272	4.75	6.23	1/8	1.151	0.139
$I = 0$	0.302	0.219	4.96	6.34	1/8	1.159	0.138
$I = 1$	0.244	0.319	4.51	6.12	1/8	1.141	0.140
Bakersfield							
Unconditional	0.178	0.316	4.72	4.80	1/8	1.158	0.136
$I = 0$	0.228	0.341	4.90	5.01	1/8	1.163	0.138
$I = 1$	0.128	0.244	4.42	4.41	1/8	1.149	0.135
Berkeley							
Unconditional	0.307	0.403	12.26	14.32	1/8	1.282	0.187
$I = 0$	0.393	0.377	12.67	12.47	1/8	1.296	0.181
$I = 1$	0.241	0.399	11.77	16.26	1/8	1.264	0.192
Big Sur							
Unconditional	0.317	0.454	22.46	29.44	1/8	1.355	0.229
$I = 0$	0.397	0.441	22.90	27.16	1/8	1.363	0.230
$I = 1$	0.230	0.426	21.62	33.51	1/8	1.342	0.227
Chico							
Unconditional	0.329	0.359	13.36	14.68	1/4	1.699	0.521
$I = 0$	0.413	0.334	15.16	15.50	1/4	1.777	0.515
$I = 1$	0.253	0.349	10.75	12.99	1/4	1.585	0.509
Davis							
Unconditional	0.274	0.342	10.29	11.86	1/8	1.253	0.182
$I = 0$	0.389	0.376	10.22	10.79	1/8	1.256	0.180
$I = 1$	0.190	0.233	10.39	13.27	1/8	1.248	0.184
Lindsay							
Unconditional	0.272	0.317	6.74	8.10	1/8	1.187	0.170
$I = 0$	0.310	0.320	7.89	9.31	1/8	1.208	0.179
$I = 1$	0.231	0.300	5.07	5.54	1/8	1.155	0.153
Nevada City							
Unconditional	0.387	0.481	22.13	23.90	1/4	1.920	0.605
$I = 0$	0.461	0.460	24.67	24.61	1/4	1.984	0.619
$I = 1$	0.319	0.482	18.73	22.50	1/4	1.833	0.575
Santa Barbara							
Unconditional	0.187	0.417	15.24	18.73	1/8	1.307	0.200
$I = 0$	0.233	0.471	16.84	19.00	1/8	1.331	0.198
$I = 1$	0.142	0.322	12.80	18.12	1/8	1.271	0.199
San Diego							
Unconditional	0.200	0.376	7.86	9.64	1/8	1.209	0.174
$I = 0$	0.206	0.353	8.81	10.36	1/8	1.228	0.178
$I = 1$	0.196	0.397	7.00	8.89	1/8	1.193	0.168
Scotia							
Unconditional	0.484	0.446	15.42	17.47	1/8	1.310	0.204
$I = 0$	0.565	0.412	17.28	19.55	1/8	1.330	0.207
$I = 1$	0.413	0.451	13.24	14.37	1/8	1.287	0.199
Tahoe City							
Unconditional	0.343	0.467	14.68	16.94	1/8	1.304	0.198
$I = 0$	0.406	0.438	15.96	17.54	1/8	1.322	0.198
$I = 1$	0.289	0.484	13.16	16.10	1/8	1.283	0.196
Ukiah							
Unconditional	0.419	0.477	15.66	17.62	1/4	1.779	0.522
$I = 0$	0.533	0.458	16.98	18.44	1/4	1.820	0.532
$I = 1$	0.315	0.441	13.62	16.09	1/4	1.717	0.501

The first row is for the unconditional model, the second row is for the conditional model given $I = 0$, and the third row is for the conditional model given $I = 1$.

model. It selects most often the model in which only the probability of a wet day π and the mean transformed intensity μ^* are varied with the circulation index. The selected model involves allowing variations in π for all but one station, frequently in μ^*, sometimes in the persistence parameter ρ, but only once in the standard deviation of the transformed intensities σ^*. Being more parsimonious, the BIC necessarily sometimes favors models with less parameters than those chosen by the AIC. For four locations (i.e., Alturas, Lindsay, Santa Barbara, and San Diego), all far removed from the index of atmospheric circulation (see Figure 1), the BIC selects the unconditional model. Otherwise, the selected model always varies π, sometimes μ^*, once ρ, and never σ^*.

In summary, it is evident that both changes in the daily occurrence and intensity processes are responsible for the observed effects of the circulation index on the distribution of January total precipitation (documented in section 2). Both the probability of wet day π and the mean transformed intensity μ^* need to be varied in most instances. In this

TABLE 3. Parameters Allowed to Vary in Model Selected by AIC and BIC for Time Series of Daily Precipitation Amounts in January at 13 Locations in California

Station	AIC	BIC
Alturas	π, ρ	none
Bakersfield	π, ρ	π
Berkeley	π, μ^*	π
Big Sur	π	π
Chico	π, μ^*	π, μ^*
Davis	π, ρ	π, ρ
Lindsay	π, μ^*, σ^*	none
Nevada City	π, μ^*	π, μ^*
Santa Barbara	π, ρ, μ^*	none
San Diego	μ^*	none
Scotia	π, μ^*	π, μ^*
Tahoe City	π, μ^*	π
Ukiah	π, μ^*	π, μ^*

AIC, Akaike's information criterion; BIC, Bayesian information criterion.

regard it is important to recall that varying μ^* (but not σ^*) still corresponds to changing both the mean and standard deviation, μ and σ, of the untransformed intensities. For simplicity in making comparisons among stations the analyses in the next section are still based on the fully conditional model, in which all of the parameters of the chain-dependent process are allowed to vary with the circulation index.

4. Induced Model for Monthly Precipitation

4.1. *Properties of Induced Model*

In section 3 the effects of an atmospheric circulation index on the time series of daily precipitation amounts have been modeled. One helpful way of interpreting these changes in the parameters of the stochastic model for daily precipitation is to convert them into terms of monthly total precipitation. For a chain-dependent process the statistical properties of a sum (e.g., representing monthly total precipitation) are well established [*Katz*, 1977*b*].

Let N be the number of wet days in a time period of length T days. Here N is a random variable, $0 \leq N \leq T$, whose distribution is determined by the Markov chain model (2)–(5) for the daily occurrence process J_t. The total precipitation, S_T say, over the time period T can be expressed as a random sum of the daily intensities

$$S_T = X_1 + X_2 + \cdots + X_N. \tag{10}$$

Applying the well-known expressions for the moments of a random sum [*Feller*, 1968, p. 301], the mean and variance of total precipitation S_T are given by

$$E(S_T) = T\pi\mu, \tag{11}$$

$$\text{Var }(S_T) \approx T\{\pi\sigma^2 + \pi(1-\pi)[(1+\rho)/(1-\rho)]\mu^2\} \tag{12}$$

[*Katz*, 1985]. The expression for the variance (12) includes a simplifying approximation for the term involving the persistence parameter ρ that is quite accurate for large time period T (e.g., as in the case of a monthly total).

Because the distribution of individual daily precipitation intensities is positively skewed (i.e., as governed by the power transform parameter p in (6)), the distribution of total precipitation S_T should still be somewhat positively skewed. The exact distribution of S_T for a chain-dependent process can be calculated by employing a recurrence relation [*Katz*, 1977*a*] or through simulation. Despite the temporal dependence of a chain-dependent process, the central limit theorem still holds for a random sum (10), implying that S_T has an approximately normal distribution for large time period T [*Katz*, 1977*b*]. But this normal approximation is not very accurate even for monthly totals [*Katz*, 1977*a*].

This paper has focused so far on stochastic models for time series of daily precipitation amounts, conditional on an index of atmospheric circulation. It is revealing to combine these conditional models, considering an induced model for daily (and monthly) precipitation. It might be anticipated that such an induced model would have exactly the same statistical properties as the unconditional chain-dependent process (i.e., fit to all of the daily precipitation amounts without any conditioning on atmospheric circulation). However, the induced model is a mixture of two chain-dependent processes with different parameters, and certain statistical characteristics could differ from those for the unconditional model.

The mean and variance of total precipitation can be expressed as functions of the conditional means and variances, given the circulation index I, as

$$E(S_T) = E[E(S_T|I)], \tag{13}$$

$$\text{Var }(S_T) = E[\text{Var }(S_T|I)] + \text{Var }[E(S_T|I)] \tag{14}$$

[*Feller*, 1971, p. 167]. Because the index I takes on only two possible values per month, these formulations (13) and (14) are particularly amenable to analysis. The expression for the mean (13) is quite intuitive and serves as no more than a check, because both the unconditional and induced models should generate monthly means that are essentially identical (depending on how the model parameters are estimated) to the actual sample mean of the monthly totals.

On the other hand, the expression for the variance (14) is not simply a weighted average of the conditional variances but includes a second term that reflects the variation of the conditional means (this phenomenon has already been noted in section 2). The induced model can produce a variance for total monthly precipitation that differs from that for the unconditional model, and neither of these variances necessarily agrees with the observed ("interannual") variance of monthly total precipitation. Because the induced variance of monthly total precipitation depends upon all of the parameters of the two chain-dependent processes (see (12)), it provides a stringent test of the performance of the model.

4.2. *Results for Induced Model*

The expressions (11) and (12) for the mean and variance of January total precipitation (i.e., $T = 31$ days) are explicit functions of the parameters of the chain-dependent process. Specifically, the mean total precipitation depends only on the probability of a wet day π and on the mean (untransformed) daily precipitation intensity μ. The variance of total precipitation depends on π and μ, as well as on the persistence parameter of the occurrence process ρ and on the variance of the (untransformed) daily precipitation intensity σ^2. By varying the individual parameters according to the conditional models for daily precipitation (Table 2), these expressions (11) and (12) quantify their contribution to the effects of the index of atmospheric circulation on the mean and variance of January total precipitation (already described in section 2).

For most stations the changes in the probability of a wet day π make a greater contribution to the effects on the mean of January total precipitation than do the corresponding changes in the mean daily precipitation intensity μ. Nevertheless, both π and μ are important in determining the magnitude of these effects. For example, if π is varied between its two conditional values (Table 2) while μ is held constant at its unconditional value, substitution into (11) yields a range of 66.3 mm for the mean of January total precipitation at Chico (or about 49% of the unconditional mean). On the other hand, if μ is varied while π is held constant, this mean has a range of 45.0 mm (or about 33% of the unconditional mean).

Likewise, the changes in the parameters of the occurrence process, π and ρ, and in the parameters of the intensity process, μ and σ^2, each make substantial contributions to the effects on the variance of January total precipitation. But for the variance the variations in the parameters of the intensity process are more important than those for the occurrence process at about half of the locations. For example, if π and ρ are varied (Table 2) while μ and σ^2 are held constant, (12) yields a range of 1,588 mm^2 for the variance of January total precipitation at Chico (or about 33% of the unconditional variance). If μ and σ^2 are varied while π and ρ are held constant, this variance has a range of 2,387 mm^2 (or about 50% of the unconditional variance).

These conclusions concerning the relative contribution of the occurrence and intensity processes to the effects of the circulation index on the mean and variance of monthly total precipitation probably cannot be easily generalized to other climates. Nevertheless, the important finding that does generalize is that both processes make substantial contributions.

Besides the mean and standard deviation, the effects of atmospheric circulation on the shape of the distribution of January total precipitation can be examined through a simulation approach (a chain-dependent process has a structure that is especially convenient for implementing simulation algorithms). Synthetic sequences, representing time series of daily precipitation amounts, are generated for both conditional chain-dependent processes and then summed to obtain an artificial monthly total precipitation. By repeating the simulation experiment (i.e., 1000 times for each model), empirical conditional distributions of total January precipitation, given the circulation index I, can be produced.

Figure 3 summarizes the outcome of this simulation study for Chico. In this case the parameter values (see Table 2) are $\pi = 0.413$, $\rho = 0.334$, $\mu^* = 1.777\ \text{mm}^{1/4}$, and $\sigma^* = 0.515\ \text{mm}^{1/4}$ when $I = 0$, and $\pi = 0.253$, $\rho = 0.349$, $\mu^* = 1.585\ \text{mm}^{1/4}$, and $\sigma^* = 0.509\ \text{mm}^{1/4}$ when $I = 1$. It is evident that the shape of the distribution is much less positively skewed when $I = 0$ (i.e., below normal pressure). Moreover, this shift in the distribution is consistent enough over the entire range of possible values of total precipitation that the one conditional distribution (i.e., $I = 0$ case) is clearly "stochastically larger" than the other. That is,

$$\Pr\{S_T > x | I = 0\} \geq \Pr\{S_T > x | I = 1\} \qquad (15)$$

for all $x > 0$ [e.g., *Ross*, 1983, p. 153]. Because of the limited number of years of observations of actual monthly total precipitation, these effects are more difficult to see in the

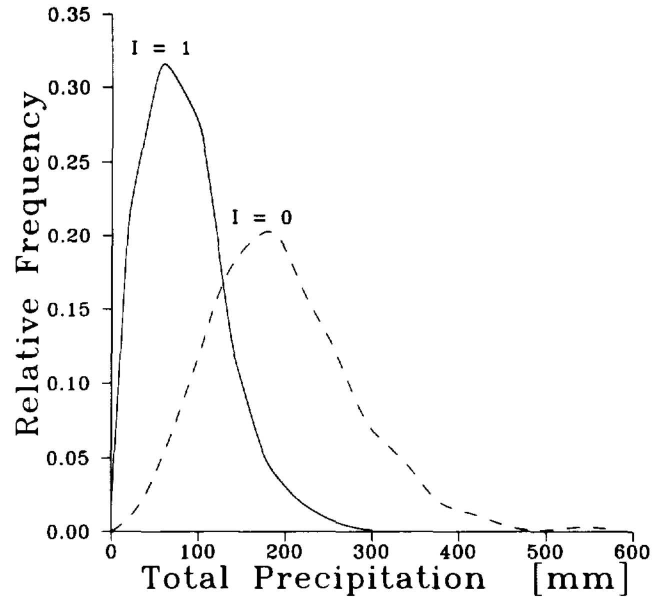

Fig. 3. Conditional distributions of January total precipitation at Chico given circulation index ($I = 0$ or $I = 1$), derived from simulations of time series of daily precipitation amounts based on chain-dependent models.

corresponding box plots or histograms for the real data (Figure 2). It would be straightforward to employ the same simulation approach to examine the behavior of precipitation statistics other than the monthly total (e.g., runs of consecutive days on which the intensity exceeds a specified threshold).

As mentioned earlier, several researchers have noted that stochastic models for time series of daily precipitation amounts tend to underestimate the observed ("interannual") variance of monthly total precipitation. Some have viewed this effect as a defect in the stochastic model which needs to be remedied [e.g., *Gregory et al.*, 1993; *Wilks*, 1989]. Others have labeled this difference in variance as a measure of the "potential predictability" of monthly total precipitation, for instance, as evidence of possible climate change [*Klugman*, 1983; *Madden and Shea*, 1982]. In any event, this discrepancy has important implications for stochastic weather generators (e.g., WGEN) that rely on a chain-dependent process to model daily precipitation [*Richardson and Wright*, 1984].

Table 4 gives the results of estimating the variance of January total precipitation from the different stochastic models for daily precipitation amounts at the 13 California sites. In agreement with previous studies, the unconditional chain-dependent process substantially underestimates the observed standard deviation of monthly total precipitation, sometimes by as much as 25% or more (Table 4 and Figure 4). For example, at Chico the estimated standard deviation based on the unconditional model is 69.2 mm or about 22% below the observed value. The two conditional chain-dependent processes produce monthly variances that are closer to the corresponding observed values, *but they still* have some tendency to underestimate. Using (14), the monthly variance for the induced model can be determined. This induced variance is always greater than that for the unconditional model, usually much closer to the observed value (Table 4 and Figure 4). For example, at Chico the

TABLE 4. Standard Deviation of January Total Precipitation as Estimated From Unconditional, Conditional, and Induced Models for Daily Precipitation at 13 Locations in California

Station	Unconditional Standard Deviation, mm	Induced Standard Deviation, mm	Conditional Standard Deviation, mm	
			$\{I = 0\}$	$\{I = 1\}$
Alturas	23.8	24.5	25.0	22.5
	(26.6)	(26.6)	(27.2)	(25.0)
Bakersfield	17.9	19.7	21.1	13.7
	(18.7)	(18.7)	(17.4)	(16.2)
Berkeley	65.4	72.1	67.2	61.7
	(74.4)	(74.4)	(66.9)	(68.7)
Big Sur	132.4	144.7	138.2	119.9
	(135.5)	(135.5)	(123.9)	(116.2)
Chico	69.2	86.8	80.8	52.2
	(88.6)	(88.6)	(82.4)	(58.9)
Davis	50.3	57.5	55.7	43.2
	(58.8)	(58.8)	(54.4)	(47.8)
Lindsay	33.0	38.4	40.4	22.0
	(39.5)	(39.5)	(42.3)	(24.3)
Nevada City	130.7	152.9	146.0	108.5
	(151.8)	(151.8)	(134.9)	(123.8)
Santa Barbara	68.5	76.4	83.5	51.5
	(78.7)	(78.7)	(80.7)	(65.3)
San Diego	35.4	36.1	38.8	32.2
	(42.7)	(42.7)	(43.5)	(41.7)
Scotia	96.9	115.4	110.2	78.3
	(117.3)	(117.3)	(108.8)	(88.2)
Tahoe City	84.8	93.4	93.5	74.1
	(113.0)	(113.0)	(127.6)	(85.2)
Ukiah	96.2	118.1	107.7	75.7
	(130.5)	(130.5)	(127.9)	(86.0)

Empirical estimates in parentheses.

estimated standard deviation based on the induced model is 86.8 mm or only about 2% less than the observed value.

It might be suspected that this improvement of the induced model over the unconditional model only reflects the fact that the sampling errors in the parameter estimates of the chain-dependent process are indirectly taken into account. However, when the individual Januarys are randomly classified into groups (i.e., ignoring the circulation index I), the standard deviation of January total precipitation estimated by the induced model is only slightly greater than that for the unconditional model. It might also be suspected that the underestimation of monthly variance is attributable to the requirement that the daily intensities be conditionally independent (see section 3.1). Although daily intensities do exhibit a slight degree of dependence (e.g., a first-order autocorrelation coefficient of about 0.17 at Chico), the estimated standard deviation of January total precipitation is again only slightly greater when this dependence is taken into account.

In summary, these results indicate that the unconditional chain-dependent process for daily precipitation has too simple a structure. In effect, the induced stochastic process, obtained by first conditioning on the atmospheric circulation and then combining these conditional models, provides a mechanism by which the consideration of more complex forms of models for daily precipitation could be justified.

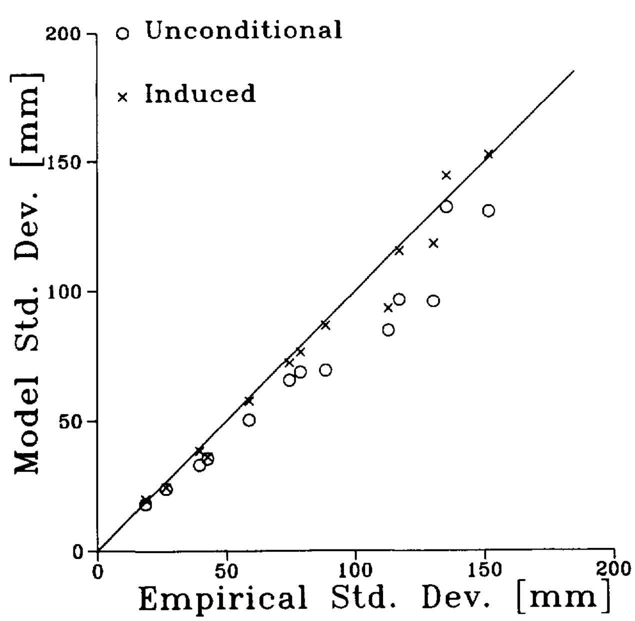

Fig. 4. Theoretical standard deviation of January total precipitation, as estimated from unconditional and induced models for daily precipitation, versus empirical standard deviation for 13 locations in California.

5. DISCUSSION

Unconditional and conditional chain-dependent processes have been fit to time series of daily precipitation amounts. Data for January at several locations in California are analyzed. The conditional models are dependent on whether the monthly mean sea level pressure is above or below normal, an index of large-scale atmospheric circulation. Through use of model selection criteria, it is established that the two conditional daily models generally differ both in terms of the probability of a wet day and the mean precipitation intensity on wet days. Loosely speaking, the higher mean of the conditional distribution of monthly total precipitation when pressure is below normal can be attributed to both more "storms" and more intense "storms." When these two conditional models are combined back into a single induced model for daily precipitation, a variance for monthly total precipitation is generated that is much closer to that actually observed than that generated by the original, unconditional daily model.

Although this relatively simple form of circulation index has the advantage of permitting an analytical treatment of the induced model for precipitation, the approach could naturally be extended to consider more complex schemes for classifying atmospheric circulation patterns. For instance, the index could have more than two states and could be permitted to vary within the month. Another issue concerns the operational prediction of the statistical properties of daily precipitation amounts. In this regard, techniques for predicting the index of atmospheric circulation a month or more in advance could be investigated.

Perhaps of more fundamental interest is the issue of in what respects the induced model for daily precipitation differs from the unconditional one. The theoretical properties of this induced stochastic process could be formally studied. In such a way an improved stochastic model for time series of daily precipitation amounts could arise. Moreover, this work would have implications concerning the related issue of long-range potential predictability of monthly total precipitation.

APPENDIX: LIKELIHOOD FUNCTIONS FOR MODEL IDENTIFICATION

This appendix outlines the general procedure for obtaining log likelihood functions for the various forms of chain-dependent processes fit to the time series of daily precipitation amounts. These likelihood functions constitute the goodness of fit component of the AIC and BIC (see (8) and (9)) for selecting the appropriate degree of conditioning on the atmospheric circulation index.

Occurrence Process

The logarithm of the maximized likelihood function, say M_1, for the Markov chain model of the occurrence process (neglecting term for initial state) satisfies

$$\ln M_1 = \sum_{i,j=0,1} n_{ij} \ln \hat{P}_{ij}. \quad \text{(A1)}$$

Here $\hat{P}_{ij}$ is the maximum likelihood estimator of P_{ij} and is given by

$$\hat{P}_{ij} = n_{ij}/(n_{i0} + n_{i1}), \quad \text{(A2)}$$

where n_{ij} denotes the number of times in the sample of observations of precipitation occurrences that a transition from state i to state j occurs (i.e., $J_t = i$ and $J_{t+1} = j$, for some day t), $i, j = 0, 1$ [e.g., *Billingsley*, 1961, p. 26].

Intensity Process

The logarithm of the maximized likelihood function, M_2 say, for the intensity process (neglecting constant terms) satisfies

$$\ln M_2 = -(N/2) \ln (s^*)^2. \quad \text{(A3)}$$

Here the sample mean and variance, $\bar{x}^*$ and $(s^*)^2$, of the transformed daily precipitation intensities X_k^*, $k = 1, 2, \cdots, N$, are the maximum likelihood estimators of the parameters of the normal distribution μ^* and $(\sigma^*)^2$ [e.g., *Lindgren*, 1968, pp. 282–283].

Chain-Dependent Process

The logarithm of the joint maximized likelihood function, M_3 say, for the chain-dependent process can be obtained simply by adding the individual log likelihoods for the occurrence and intensity processes. That is,

$$\ln M_3 = \ln M_1 + \ln M_2. \quad \text{(A4)}$$

Similarly, the logarithm of the joint maximized likelihood function, M_4 say, for the two conditional chain-dependent processes can be obtained by computing the log likelihoods separately for each conditional model and then adding them together. That is,

$$\ln M_4 = \ln M_3(0) + \ln M_3(1), \quad \text{(A5)}$$

where $M_3(0)$ and $M_3(1)$ denote the likelihood function (A4) for the conditional chain-dependent processes given the atmospheric circulation index $I = 0$ and $I = 1$, respectively. Finally, the same basic approach can be applied when some of the model parameters are constrained to not vary with the circulation index. For instance, the estimator of the common variance of the transformed intensities $(\sigma^*)^2$, when the mean of the transformed intensities is still allowed to vary with the circulation index, is the so-called "pooled variance" employed in the popular two-sample t test for the equality of means [e.g., *Lindgren*, 1968, p. 393].

Acknowledgments. We thank Mary Downton and Larry Rodriguez for help in obtaining data. The comments of two anonymous reviewers are also gratefully acknowledged. The sea level pressure data were retrieved from the data archives maintained by the Data Support Section, Scientific Computing Division, National Center for Atmospheric Research (NCAR). M.B.P. received support from NCAR's Environmental and Societal Impacts Group. NCAR is sponsored by the National Science Foundation.

REFERENCES

Akaike, H., A new look at the statistical model identification, *IEEE Trans. Autom. Control*, *19*, 716–723, 1974.

Bardossy, A., and E. J. Plate, Modeling daily rainfall using a semi-Markov representation of circulation pattern occurrence, *J. Hydrol.*, *122*, 33–47, 1991.

Bardossy, A., and E. J. Plate, Space-time model for daily rainfall using atmospheric circulation patterns, *Water Resour. Res.*, *28*, 1247–1259, 1992.

Billingsley, P., *Statistical Inferences for Markov Processes*, 75 pp., University of Chicago Press, Chicago, Ill., 1961.

Box, G. E. P., and D. R. Cox, An analysis of transformations (with discussion), *J. R. Stat. Soc., Ser. B*, *26*, 211–252, 1964.

Cayan, D. R., and D. H. Peterson, The influence of North Pacific atmospheric circulation on streamflow in the west, in *Aspects of Climate Variability in the Pacific and the Western Americas*, *Geophys. Monogr. Ser.*, vol. 55, edited by D. H. Peterson, pp. 375–397, AGU, Washington, D. C., 1989.

Feller, W., *An Introduction to Probability Theory and Its Applications*, vol. I, 3rd ed., 509 pp., John Wiley, New York, 1968.

Feller, W., *An Introduction to Probability Theory and Its Applications*, vol. II, 2nd ed., 669 pp., John Wiley, New York, 1971.

Gabriel, K. R., and J. Neumann, A Markov chain model for daily rainfall occurrence at Tel Aviv, *Q. J. R. Meteorol. Soc.*, *88*, 90–95, 1962.
Glantz, M. H., R. W. Katz, and N. Nicholls (eds.), *Teleconnections Linking Worldwide Climate Anomalies: Scientific Basis and Societal Impact*, 535 pp., Cambridge University Press, New York, 1991.
Gregory, J. M., T. M. L. Wigley, and P. D. Jones, Application of Markov models to area-average daily precipitation series and interannual variability in seasonal totals, *Clim. Dyn.*, in press, 1993.
Hinkley, D. V., On quick choice of power transform, *Appl. Stat.*, *26*, 67–69, 1977.
Jenne, R. L., Data sets for meteorological research, *NCAR Tech. Note TN/IA-111*, 194 pp., Natl. Cent. for Atmos. Res., Boulder, Colo., 1975.
Katz, R. W., Precipitation as a chain-dependent process, *J. Appl. Meteorol.*, *16*, 671–676, 1977*a*.
Katz, R. W., An application of chain-dependent processes to meteorology, *J. Appl. Probab.*, *14*, 598–603, 1977*b*.
Katz, R. W., Probabilistic models, in *Probability, Statistics, and Decision Making in the Atmospheric Sciences*, edited by A. H. Murphy and R. W. Katz, pp. 261–288, Westview, Boulder, Colo., 1985.
Klugman, M. R., Evidence of climate change in United States seasonal precipitation data, 1948–76, *J. Clim. Appl. Meteorol.*, *22*, 1367–1376, 1983.
Lindgren, B. W., *Statistical Theory*, 2nd ed., 521 pp., Macmillan, New York, 1968.
Lloyd, E. H., What is, and what is not, a Markov chain?, *J. Hydrol.*, *22*, 1–28, 1974.
Madden, R. A., and D. J. Shea, Potential long-range predictability of precipitation over North America, in *Proceedings of the Seventh Climate Diagnostics Workshop*, pp. 423–426, Climate Analysis Center, National Oceanic and Atmospheric Administration, Washington, D. C., 1982.
Redmond, K. T., and R. W. Koch, Surface climate and streamflow variability in the western United States and their relationship to large-scale circulation indices, *Water Resour. Res.*, *27*, 2381–2399, 1991.
Richardson, C. W., and D. A. Wright, WGEN: A model for generating daily weather variables, Agric. Res. Serv., *Rep. ARS-8*, 83 pp., Natl. Tech. Info. Serv., Springfield, Va., 1984.
Ross, S. M., *Introduction to Stochastic Dynamic Programming*, 164 pp., Academic, San Diego, Calif., 1983.
Schwarz, G., Estimating the dimension of a model, *Ann. Stat.*, *6*, 461–464, 1978.
Todorovic, P., and D. A. Woolhiser, A stochastic model of *n*-day precipitation, *J. Appl. Meteorol.*, *14*, 17–24, 1975.
Trenberth, K. E., and D. A. Paolino, Jr., The northern hemisphere sea-level pressure data set: Trends, errors, and discontinuities, *Mon. Weather Rev.*, *108*, 855–872, 1980.
Wallace, J. M., and D. S. Gutzler, Teleconnections in the geopotential height field during northern hemisphere winter, *Mon. Weather Rev.*, *109*, 784–812, 1981.
Waymire, E., and V. K. Gupta, The mathematical structure of rainfall representations, 1, A review of stochastic rainfall models, *Water Resour. Res.*, *17*, 1261–1272, 1981.
Weare, B. C., and M. A. Hoeschele, Specification of monthly precipitation in the western United States from monthly mean circulation, *J. Clim. Appl. Meteorol.*, *22*, 1000–1007, 1983.
Wilks, D. S., Conditioning stochastic daily precipitation models on total monthly precipitation, *Water Resour. Res.*, *25*, 1429–1439, 1989.
Wilson, L. L., D. P. Lettenmaier, and E. F. Wood, Simulation of daily precipitation in the Pacific Northwest using a weather classification scheme, *Surv. Geophys.*, *12*, 127–142, 1991.
Woolhiser, D. A., T. O. Keefer, and K. T. Redmond, Southern Oscillation effects on daily precipitation in the southwestern United States, *Water Resour. Res.*, *29*, 1287–1295, 1993.

R. W. Katz, Environmental and Societal Impacts Group, National Center for Atmospheric Research, Boulder, CO 80307.

M. B. Parlange, Hydrologic Science, Department of Land, Air and Water Resources and Department of Agricultural Engineering, University of California, Davis, CA 95616.

(Received August 12, 1992;
revised January 5, 1993;
accepted March 5, 1993.)