WATER RESOURCES RESEARCH, VOL. 18, NO. 5, PAGES 1461–1468, OCTOBER 1982

Stochastic Daily Precipitation Models
2. A Comparison of Distributions of Amounts

DAVID A. WOOLHISER

Southwest Rangeland Watershed Research Center, Agricultural Research Service, U.S. Department of Agriculture Tucson, Arizona 85705

JOSÉ ROLDÁN

Catedra de Hidráulica General y Agricola, Escuela Tecnica Superior de Ingenieros, Agronomos Universidad de Cordoba Cordoba, Spain

Chain-dependent and independent exponential, gamma, and mixed exponential distributions are compared as models for the distribution of daily precipitation. Parameters for each distribution are estimated by maximum likelihood techniques for 14-day periods. The Akaike information criterion is used to select the most appropriate distribution for each period and for the entire year. For the five U.S. stations studied, the independent mixed exponential distribution was the best on the basis of the Akaike information criterion, and the independent gamma and chain-dependent gamma ranked second and third, respectively. Fourier series are fit to the parameters by least squares to provide starting values for subsequent numerical maximum likelihood estimates of the Fourier coefficients. According to the Akaike information criterion, the Fourier series description of model parameters for the mixed exponential model is superior to the specification of parameters for each 14-day period.

INTRODUCTION

Woolhiser and Pegram [1979] used direct numerical maximum likelihood estimates of Fourier coefficients to describe the seasonal variation of parameters in a stochastic model of daily precipitation. They demonstrated the technique using a first-order Markov chain as the occurrence process and a mixed exponential distribution for the daily precipitation. They suggested that it may be possible to map the means, amplitudes, and phase angles for significant harmonics to provide a parsimonious regionalized model of the point precipitation process. In a previous paper [*Roldán and Woolhiser,* this issue] we used the techniques described by Woolhiser and Pegram to compare the Markov chain with another alternating renewal process. In this paper we compare the mixed exponential distribution with two other independent distributions and four chain-dependent distributions for 14-day time periods and on an annual basis. The overall objective of this work is to find a parsimonious model of the distribution of daily precipitation that is adapted to a Fourier series representation of the time variations in the parameters.

Several methods have been presented in the literature for modeling precipitation amounts on wet days. The most common approach is to assume that precipitation amounts on successive days are independent and to fit some theoretical distribution to the precipitation amounts [*Todorovic and Woolhiser,* 1974, 1975; *Woolhiser et al.*, 1973; *Smith and Schreiber,* 1974]. A second approach is to assume that precipitation amounts are independent but that the distribution function depends on whether the previous day was wet or dry (i.e., a chain-dependent process [*Katz,* 1977*a*]). *Buishand* [1977] distinguished between three different types of wet days, namely, solitary wet days, wet days bounded on one side by a wet day and on the other by a dry day, and wet days bounded on both sides by a wet day.

Theoretical distributions used include the exponential [*Todorovic and Woolhiser,* 1974, 1975; *Woolhiser et al.*, 1973], the mixed exponential [*Smith and Schreiber,* 1974; *Woolhiser and Pegram,* 1979], the gamma [*Ison et al.*, 1971; *Katz,* 1977*b*; *Buishand,* 1977], and the kappa [*Mielke,* 1973]. The seasonal variation of parameters has been handled by estimating parameters for discrete periods, by fitting Fourier series to the period values using least squares techniques [*Ison et al.*, 1971; *Buishand,* 1977], or by fitting Fourier series using direct maximum likelihood techniques [*Woolhiser and Pegram,* 1979].

Each of these models requires a different number of parameters, depending on dependence assumptions and the choice of the theoretical distribution. For example, the independent exponential requires one parameter, which makes it admirably suited to certain analytic results [*Todorovic and Woolhiser,* 1974], while the gamma distribution, with three types of wet days, requires six parameters. Because one of the goals of this approach is a parsimonious model, we have subjectively eliminated Buishand's approach from consideration.

DISTRIBUTIONS STUDIED

The precipitation process under consideration can be described by a bivariate sequence of random variables $\{(X_t, Y_t): t = 1, 2, \cdots\}$. If precipitation occurs on day t, it is said to be a wet day and $X_t = 1$; if not, it is a dry day and $X_t = 0$. Y_t is the amount of precipitation that falls on day t. The $\{Y_t\}$ process is a sequence of real-valued random variables which satisfies the following requirements:

1. The distribution of Y_t depends on X_{t-1} and X_t.
2. Given the $\{X_t\}$ process, the Y_t are independent.

We will consider two special cases for the sequence $\{Y_t\}$.

1. The Y_t are independent random variables with distri-

Paper number 2W1226.

TABLE 1. Distributions Compared

Process	Case	Distribution	Number of Parameters Required for Each Period*
1	2	Exponential (1)	1
2	2	Gamma (1)	2
3	2	Mixed exponential (1)	3
4	1	Exponential (2)+ Exponential (3)	2
5	1	Gamma (2)+ Gamma (3)	4
6	1	Exponential (2)+ Gamma (3)	3
7	1	Gamma (2)+ Exponenial (3)	3

Total population of wet days is indicatd by (1), (2) means population of wet days preceded by a wet day, and (3) is population of wet days preceded by a dry day.

*Data were subdivided into 14-day periods (see the section on analysis of data).

bution function

$$Fn_i(y) = P\{Y_t \leq y | X_t = 1, X_{t-1} = i\}$$

$$i = 0, 1 \qquad n = 1, 2, \cdots 365 \qquad 0 < y \tag{1}$$

where n is the calendar day [(t modulo 365) + 1].

2. The Y_t are independent random variables with distribution function

$$F_n(y) = F_{n0}(y) = F_{n1}(y) \tag{2}$$

For the first case we see that given a wet day on day t, the amount of precipitation which occurs on that day is chosen from one of two distributions for calendar day n, depending on whether the previous day was dry or wet. For the second case the amount of precipitation on day t is chosen from a single distribution for calendar day n, i.e., it is not affected by the state (wet or dry) of day $t - 1$.

If these distributions are independent of the calendar day within a season and if the sequence $\{X_t\}$ is a first-order Markov chain, both cases are so-called chain-dependent processes [*Katz,* 1977*a, b*]. Case 2 is similar to the process considered by *Todorovic and Woolhiser* [1975].

We will take some liberty with the terminology and refer to the sequence $\{Y_t\}$, described by (1), as chain-dependent, and the sequence $\{Y_t\}$, described by (2), as independent. Independence here is understood in the sense that

$$P\{Y_t \leq y | X_0, Y_1, X_1, Y_2 \cdots X_{t-1}\ Y_{t-1}\} = P\{Y_t \leq y | X_t = 1\} = F_n(y)$$

Although the process $\{X_t\}$ is frequently described by a Markov chain, the analyses in this paper does not place that restriction on $\{X_t\}$.

We will examine three distribution functions in this paper, the exponential, the gamma, and the mixed exponential. The probability density functions are as follows.

Exponential:

$$f_n(y) = \frac{1}{\lambda(n)} \exp[-y/\lambda(n)] \tag{3}$$

$$y > 0 \qquad \lambda(n) > 0 \qquad n = 1, 365$$

Gamma:

$$g_n(y) = \frac{y^{\delta(n)-1} \exp[-y/\gamma(n)]}{\gamma(n)^{\delta(n)} \Gamma[\delta(n)]} \tag{4}$$

$$y > 0 \qquad \gamma(n) \qquad \delta(n) > 0 \qquad n = 1, 365$$

Mixed exponential:

$$h_n(y) = \frac{\alpha(n)}{\beta(n)} \exp[-y/\beta(n)] + \frac{[1 - \alpha(n)]}{\theta(n)} \exp[-y/\theta(n)] \tag{5}$$

$$y > 0 \qquad 0 \leq \alpha(n) \leq 1 \qquad 0 < \beta(n) < \theta(n) \qquad n = 1, 365$$

Each of the three distributions specified by (3)–(5) were used for the 'independent process' (Case 2). The exponential and the gamma distributions were used for the chain-dependent processes. The number of parameters for each process studied is shown in Table 1. To account for seasonal variability, each of the parameters in (3)–(5) can be expressed in terms of the polar form of a finite Fourier series:

$$a_j(n) = A_j + \sum_{i=1}^{m} \left[C_{ji} \sin\left(\frac{2\pi ni}{365} + \phi\right)\right] \tag{6}$$

where $j = 1, k (k = 1, 2, 3,$ or 4, number of parameters), $n = 1, 2 \cdots 365$, m is the maximum number of harmonics, C_{ji} is the amplitude, ϕ_{ji} is the phase angle, and A_j is the mean of each parameter.

MAXIMUM LIKELIHOOD EQUATIONS

The coefficients A_j, C_{ji}, ϕ_{ji}, $j = 1, k$; $i = 1, m$ for each process can be estimated by direct numerical optimization techniques. The log likelihood functions are as follows.

Exponential:

$$\log L_f = \sum_{i=1}^{365} \sum_{j=1}^{p(i)} [-y_{ij}/\lambda(i) - \log \lambda(i)] \tag{7}$$

Gamma:

$$\log L_g = \sum_{i=1}^{365} \sum_{j=1}^{p(i)} (\delta(i) - 1) \log y_{ij} - y_{ij}/\gamma(i) - \delta(i) \log \gamma(i) - \log \Gamma(\delta(i)) \tag{8}$$

Mixed exponential:

$$\log L_h = \sum_{i=1}^{365} \sum_{j=1}^{p(i)} \left\{ \log \left[\frac{\alpha(i)}{\beta(i)} \exp \cdot [-y_{ij}/\beta(i)] + \left(\frac{1 - \alpha(i)}{\theta(i)}\right) \exp(-y_{ij}/\theta(i)) \right]\right\} \tag{9}$$

where $p(i)$ is the number of wet days for day i, $i = 1, 2, \cdots 365$ for the period of record, y_{ij} is the amount of precipitation for the jth wet day for day i, and each parameter is expressed in the form specified by (6). Similar expressions were used for the chain-dependent processes, except that the wet days were separated into two classes: wet days preceeded by a dry day and wet days preceeded by a wet day.

Let $\boldsymbol{\theta}_f$, $\boldsymbol{\theta}_g$, and $\boldsymbol{\theta}_h$ be vectors whose elements are the coefficients of the Fourier series describing the parameter set for each distribution. The objective is to find the estimate

$\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ that maximizes $\log L$ in (7), (8), and (9). Furthermore, the maximum likelihood function so obtained can be utilized in calculating the Akaike information criterion [*Akaike*, 1974] to choose the distribution which best fits the data for several stations when the parameters are allowed to vary with time, as specified by (6).

ANALYSIS OF DATA

Stations Investigated

Daily precipitation data were obtained from the National Weather Service for the following stations: (1) Wichita, Kansas, with 20 years of record, (2) Kansas City, Missouri, with 25 years, (3) Tallahassee, Florida, with 23 years, (4) Sheridan, Wyoming, with 20 years, and (5) Indianapolis, Indiana, with 20 years. The data were arranged so that day 1 was March 1, and the extra day during leap year was ignored. First, the data were subdivided into 14-day periods, and maximum likelihood estimates of parameters for each distribution were obtained assuming the parameters are constant within each period.

We did not test the degree of dependence between precipitation amounts on consecutive days. Other investigators have found that the hypothesis of no dependence either cannot be rejected or that the dependence is very weak [*Buishand*, 1977; *Katz*, 1977*a*].

Estimation of Parameters

The maximum likelihood estimates of parameters for the 14-day periods were obtained as follows.

Exponential Distribution.

$$\lambda_j = \frac{1}{n(j)} \sum_{i=1}^{n(j)} Y_{ji} \qquad j = 1, 26 \tag{10}$$

where $n(j)$ is the number of wet days in period j for the total period of record.

Gamma Distribution. The maximum likelihood equations for the gamma distribution were solved by Newton-Raphson iteration using the following approximate maximum likelihood estimates [*Choi and Wette*, 1969] as initial values:

$$\delta_j' = \left\{2\left[\log \bar{y}_j - \frac{1}{n(j)} \sum_{i=1}^{n(j)} \log y_{ji}\right]\right\}^{-1} \qquad j = 1, 26 \tag{11}$$

$$\gamma_j' = \bar{y}_j/\delta_j' \qquad j = 1, 26 \tag{12}$$

where $\bar{y}_j$ is the sample mean daily precipitation for period j.

Mixed Exponential Distribution. The maximum likelihood expressions for parameters in the mixed exponential distribution are also in implicit form, so Newton-Raphson iteration was also used to estimate the parameters. Initial values were estimated by the method of moments [*Rider*, 1961] and by selecting parameters corresponding to the maximum of the likelihood functions of the set of seven likelihood functions calculated by allowing α to range from 0.2 to 0.8 at intervals of 0.1, while β ranged from 0.2 $\bar{y}_j$ to 0.8 $\bar{y}_j$ at intervals of 0.1 $\bar{y}_j$. For each pair (α, β), θ was calculated from the following expression for the mean:

$$\bar{y}_j = \alpha\beta + (1 - \alpha)\theta \qquad j = 1,26 \tag{13}$$

Occasionally, the search procedure led to a local optimum when α approached 0 or 1, or β approached θ. Under these circumstances the mixed exponential approaches the single exponential.

Classical hypothesis testing procedures cannot be used to identify the best of the alternative models considered here. Although each of the models can be considered as a special case of the chain dependent mixed gamma distribution, the different restrictions on the parameter vector prevents the use of the likelihood ratio test for nine of the possible 21 pairwise combinations. *Akaike* [1974] has suggested an objective method of model identification designed to overcome this problem. The unknown distribution is a member of the parametric family $f(y|\boldsymbol{\theta})$. The parameter vector $\boldsymbol{\theta}$ is limited to some maximum dimension M. A candidate model to be tested has the dimension k. The Akaike information criterion (AIC) is

$$\mathrm{AIC}(k) = N^{-1}\left(-2 \log\left[\frac{L(y, \hat{\boldsymbol{\theta}}_k)}{L(y, \hat{\boldsymbol{\theta}}_M)}\right] + 2k - M\right) \tag{14}$$

where N is the number of observations and $L(y, \boldsymbol{\theta})$ represents the likelihood function. With M and N fixed, the constant factors in (14) may be omitted to give

$$\mathrm{AIC}(k) = -2 \log L(y, \hat{\boldsymbol{\theta}}_k) + 2k \tag{15}$$

As the number of independent parameters, k, increases, the likelihood $L(y, \hat{\boldsymbol{\theta}}_k)$ increases; however, the uncertainty of the model, which is proportional to $2k$, also increases. The best model is the one defined by the minimum AIC.

The likelihood ratio test can be used to evaluate alternative distributions when they belong to the same family. For example, if we wish to consider whether the use of the gamma distribution, rather than the exponential distribution, results in a significantly improved fit for period j, we assume that the density function has the form

$$f_j(y, \boldsymbol{\theta}_j) = \frac{y^{\delta_j - 1} \exp(-y/\gamma_j)}{\gamma_j^{\delta_j}\,\Gamma(\delta_j)} \tag{16}$$

and test the Null hypothesis:

$$H_0\colon \boldsymbol{\theta}_j = \boldsymbol{\theta}_j' = (\delta_j = 1, \gamma_j') \tag{17}$$

against the alternative:

$$H_1\colon \boldsymbol{\theta}_j = (\delta_j, \gamma_j). \tag{18}$$

Let $L(y, \hat{\boldsymbol{\theta}}')$ be the maximum likelihood function when H_0 is true and $L(y, \hat{\boldsymbol{\theta}})$ be the maximum likelihood function under the alternative hypothesis. Under certain regularity conditions, the statistic $-2 \log_e\{L(y, \hat{\boldsymbol{\theta}}')/L(y, \hat{\boldsymbol{\theta}})\}$ has a distribution that approaches the chi square distribution for large n with degrees of freedom equal to the number of parameters determined by H_0. For comparisons between distributions with the same number of parameters, for example, the mixed exponential and the three-parameter chain-dependent distributions, the distribution with the greatest likelihood function can be chosen. The AIC was calculated for each distribution for each 14-day period, and the distribution with minimum AIC was identified.

Evaluation Procedures

The number of periods out of a total of 26 for which a given distribution was the best, according to the Akaike information criterion, is shown in Table 2. According to this criterion, the mixed exponential distribution (MED) pro-

TABLE 2. Number of Periods of Best Fit for Each Distribution

Station	Distribution: Independent (Case 2) Exp (1)	Gam (1)	Mixed Exp (1)	Chain dependent (Case 1) Exp (2) + Exp (3)	Gam (2) + Gam (3)	Exp (2) + Gam (3)	Exp (3) + Gam (2)	Total
Wichita, Kansas	0	0	24	1	1	0	0	26
Kansas City, Missouri	0	0	25	0	1	0	0	26
Tallahassee, Florida	1	0	22	0	2	1	0	26
Sheridan, Wyoming	5	0	17	2	0	0	2	26
Indianapolis, Indiana	0	1	23	2	0	0	0	26

(1), (2) and (3) as in Table 1.

vides the best fit, ranking first for over half of the periods for all stations. Surprisingly, the chain-dependent processes are best for only 12 station periods out of a total of 130. A detailed presentation of the results for each period for each station would be too voluminous to present in this paper. However, the appropriate statistics for the alternative distributions considered are presented in Table 3 for period 14 for Indianapolis as an example and to demonstrate the relationship between the AIC and the classical likelihood ratio test. As *Akaike* [1974, p. 720] has noted,

> When the models are specified by a successive increase of restrictions on the parameter θ of $f(x|\theta)$, the MAICE (minimum information theoretical criterion estimate) procedure takes a form of repeated applications of conventional log-likelihood ratio tests of goodness of fit with automatically adjusted levels of significance defined by the terms $+\ 2k$.

To compare distributions on an annual basis, the 14-day period values of the log-likelihood functions were added together to form a composite likelihood function, and the Akaike information criterion was calculated. The values of the log-likelihood function and the AIC for each distribution are shown in Table 4. The mixed exponential distribution is the best for all stations, and the exponential is the worst. The independent gamma ranks second and the chain-dependent gamma third, according to the Akaike information criterion. This ranking is consistent for all stations except Sheridan, where the Exp(3) + Gam(2) ranks second, and the chain-dependent exponential third. From the statistics presented in Tables 3 and 4, it can be concluded that the mixed exponential distribution provides the best description of precipitation for 14-day periods for these stations.

Because it is possible that a distribution that provides the best fit for short time periods may not perform as well when the parameters are allowed to vary as specified by a Fourier series, we calculated Fourier coefficients for the first two harmonics for each parameter by least squares techniques and used daily parameter values derived from the series to calculate a likelihood function for the entire year for each case and for each station. The log-likelihood functions and the AIC are shown in Table 5. It can be seen that for the mixed exponential distribution, the AIC is always smaller when using the Fourier series representation than the individual values for the 14-day periods are, although there is a reduction in the likelihood. The AIC for the Fourier series representation was greater than the AIC for the composite of seasons only eight times out of the possible 35 shown. This suggests that the improvement in the likelihood function, using 14-day period estimates, is not statistically better than using the smaller number of Fourier coefficients. If maximum likelihood (rather than least squares) estimates of the Fourier coefficients had been used, the AIC would be even

TABLE 3. Statistics for Alternative Distributions, Period 14, Indianapolis, N = 41

Distribution	log L	AIC	Likelihood Ratio Statistic: Exp (1)	Gam (1)	Mixed Exp (1)	Exp (2) + Exp (3)	Gam (2) + Gam (3)	Exp (2) + Gam (3)	Gam (2) + Exp (3)
Exp (1)	3.433	−4.87		6.77*	10.64*	0.146	6.90	5.09	1.96
Gam (1)	6.820	−9.64			*A*	*L*	0.129	*A*	A
Mixed Exp (1)	8.757	−11.51				*A*	*A*	*L*	*L*
Exp (2) + Exp (3)	3.506	−3.01					6.76†	4.94†	1.81
Gam (2) + Gam (3)	6.884	−5.77						1.81	4.94*
Exp (2) + Gam (3)	5.977	−5.95							L
Gam (2) + Exp (3)	4.413	−2.83							

(1), (2), and (3) are as in Table 1. *A* indicates comparison by Akaite information criterion only, while *L* indicates comparison by likelihood function.
*Probability of greater test statistic <0.01.
†Probability of greater test statistic <0.05.

TABLE 4. Composite Log-Likelihood Functions and Akaike Information Criteria (AIC)

	Stations										
	Wichita, Kansas		Kansas City, Missouri		Tallahassee, Florida		Sheridan, Wyoming		Indianapolis, Indiana		Number of Param- eters
Distribution	Log L	AIC	Log L	AIC	Log L	AIC	Log L	AIC	Log L	AIC	
Exp (1)	244.62	−437.24	155.06	−258.13	−978.79	2009.58	2185.33	−4318.66	362.14	−672.28	26
Gam (1)	378.73	−653.46	272.27	−440.53	−799.38	1702.76	2236.08	−4368.15	482.73	−861.46	52
Mix Exp (1)	497.73	−839.46*	375.00	−594.00*	−702.38	1560.77*	2336.12	−4516.24*	585.16	−1014.32*	78
Exp (2) + Exp (3)	273.05	−442.10	197.01	−290.02	−932.99	1969.99	2245.99	−4387.98	389.23	−674.46	52
Gam (2) + Gam (3)	405.32	−602.65	307.60	−407.21	−760.88	1729.76	2291.46	−4374.91	507.22	−806.43	104
Exp (2) + Gam (3)	362.89	−569.78	266.84	−377.68	−834.79	1825.58	2264.18	−4372.36	445.96	−735.93	78
Exp (3) + Gam (2)	315.48	−474.97	237.77	−319.55	−859.09	1874.17	2273.26	−4390.53	450.48	−744.96	78

(1), (2), and (3) as in Table 1.
*Minimum AIC.

more favorable. With the Fourier series representation of parameters the chain-dependent gamma distribution appears to be slightly better than the independent gamma distribution. It is important to note that the likelihood function for the MED, using Fourier coefficients shown in Table 5, is higher than the likelihood function for the second-best distribution (chain-dependent gamma) evaluated for 14-day periods for three out of five stations. Although optimization of the Fourier coefficients and application of the likelihood ratio test or AIC to determine the number of significant harmonics will result in moderate increases in the log likelihood functions shown in Table 5 and in a significant reduction in the number of parameters, the substantial difference between the likelihood functions for the mixed exponential distribution and the chain-dependent gamma precludes any change in the first ranked distribution. For this reason we can confine our attention on the optimization of Fourier coefficients to the mixed exponential distribution.

Optimization of Fourier Coefficients for the Mixed Exponential Distribution

Woolhiser and Pegram [1979] demonstrated that the Fourier coefficients for the mixed exponential parameters are weakly dependent. For this reason the results obtained by sequential optimization of coefficients for each harmonic for each parameter will depend upon the order of computations. The more rigorous alternative of multivariate optimization with up to 15 variables (two harmonics for each parameter) required too much computer time. Therefore we examined three sequential optimization strategies.

Prior to the optimization the following screening procedure was utilized to determine which harmonics would be included for each parameter and the order of analysis. First, the log-likelihood function was calculated with parameters α, β, and θ constant; then a single harmonic was added for one parameter and the likelihood function was calculated again. That harmonic was considered for subsequent optimization if it was significant at the 0.05 level, according to a likelihood ratio test [*Mielke and Johnson,* 1973]. Those that were significant were entered into subsequent optimizations in an order determined by the incremental increase in the log-likelihood function due to that harmonic. The screening was also performed using the likelihood ratio test at the 0.10 level, but for the cases we studied, it had no effect on the final results.

The three optimization strategies used were (1) sequential optimization of each mean along with the amplitude and

TABLE 5. Log L and AIC Evaluated Daily Considering the Two First Harmonics for Each Parameter

	Stations										
	Wichita, Kansas		Kansas City, Missouri		Tallahassee, Florida		Sheridan, Wyoming		Indianapolis, Indiana		Number of Param- eters
Distribution	Log L	AIC	Log L	AIC	Log L	AIC	Log L	AIC	Log L	AIC	
Exp (1)	221.17	−432.35†	129.82	−249.64†	−997.39	2004.78	2155.72	−4301.43†	339.88	−669.75†	5
Gam (1)	353.57	−687.14	248.86	−477.72	−824.81	1669.63	2195.44	−4370.88	457.91	−895.82	10
Mix Exp (1)	461.72	−893.44*	344.54	−659.08*	−762.18	1554.37*	2290.51	−4551.02*	532.82	−1035.65*	15
Exp (2) + Exp (3)	235.56	−451.12	147.37	−266.75†	−973.85	1967.71	2192.71	−4365.41†	341.40	−662.80†	10
Gam (2) + Gam (3)	362.09	−684.18	259.02	−478.04	−812.56	1665.12	2217.48	−4394.96	457.91	−875.82	20
Exp (2) + Gam (3)	325.96	−621.92	217.54	−405.08	−879.23	1788.46	2205.27	−4380.53	396.54	−763.07	15
Exp (3) + Gam (2)	271.69	−513.38	184.85	−339.71	−907.18	1844.37	2204.92	−4379.85†	402.78	−775.55	15

(1), (2), and (3) as in Table 1.
*Minimum AIC.
†AIC for Fourier series representation greater than AIC for composite.

TABLE 6*a*. Estimated Coefficients for the Mixed Exponential Distribution, Tallahassee, Florida

Case	Parameter	Method of Estimation	Mean	Method of Estimation	Mean	Method of Estimation	C_1	ϕ_1, rad	C_2	ϕ_2, rad	Number of Coefficients	Log L	AIC	Computer Time, s
1	α			LS	0.451	LS								
				ML		ML	NS	NS	NS	NS				
	β			LS	0.143	LS								
				ML		ML	NS	NS	NS	NS				
	θ			LS	0.894	LS			0.158	0.156				
				ML	0.872	ML	NS	NS	0.151	0.421	5	−759.46	1528.92	73.49
2	α	LS	0.451	ML	0.370	LS								
		ML	0.370	ML		ML	NS	NS	NS	NS				
	β	LS	0.143	ML	0.110	LS								
		ML	0.110	ML		ML	NS	NS	NS	NS				
	θ	LS	0.894	ML	0.785	LS			0.158	0.156				
		ML	0.785	ML	0.803	ML	NS	NS	0.135	0.420	5	−755.71	1521.42	201.01
3	α	LS	0.451			LS								
		ML	0.370			ML	NS	NS	NS	NS				
	β	LS	0.143			LS								
		ML	0.110			ML	NS	NS	NS	NS				
	θ	LS	0.894			LS			0.158	0.156				
		ML	0.785			ML	NS	NS	0.124	0.465	5	−756.06	1522.12	168.00
No screening	α			LS	0.451	LS								
or order				ML		ML	NS	NS	NS	NS				
	β			LS	0.143	LS								
				ML		ML	NS	NS	NS	NS				
	θ			LS	0.894	LS			0.158	0.156				
				ML	0.872	ML	NS	NS	0.151	0.421	5	−759.46	1528.92	486.03

LS stands for least squares; ML, for maximum likelihood. NS means not significant at 0.01 level.

phase angle of each harmonic that passed the screening test, (2) simultaneous optimization of the means of α, β, and θ followed by sequential optimization of each mean along with amplitude and phase angle of each harmonic, and (3) simultaneous optimization of the three means followed by sequential two-parameter optimizations of the amplitude and phase angle of each harmonic. A maximum of two harmonics was considered, and only one harmonic was optimized at each stage. The first harmonic was analyzed for all parameters, followed by the second harmonic. If an optimized parameter was not significant at the 0.01 level according to the likelihood ratio test, it was dropped. Akaike's information criterion gave the same results.

Least squares procedures were used to estimate starting

TABLE 6*b*. Estimated Coefficients for the Mixed Exponential Distribution, Sheridan, Wyoming

Case	Parameter	Method of Estimation	Mean	Method of Estimation	Mean	Method of Estimation	C_1	ϕ_1, rad	C_2	ϕ_2, rad	Number of Coefficients	Log L	AIC	Computer Time, s
1	α			LS	0.550									
				ML		ML	NS	NS	NS	NS				
	β			LS	0.054	LS			0.012	−0.189				
				ML	0.058	ML	NS	NS	0.018	−0.455				
	θ			LS	0.264	LS	0.157	−0.045						
				ML	0.244	ML	0.140	−0.539	NS	NS	7	2295.65	−4577.3	225.36
2	α	LS	0.550	ML	0.618	LS								
		ML	0.618	ML		ML	NS	NS	NS	NS				
	β	LS	0.054	ML	0.058	LS			0.012	−0.189				
		ML	0.058	ML	0.064	ML	NS	NS	0.020	−0.477				
	θ	LS	0.264	ML	0.280	LS	0.157	−0.045						
		ML	0.280	ML	0.263	ML	0.154	−0.529	NS	NS	7	2292.85	−4571.7	303.24
3	α	LS	0.550			LS								
		ML	0.618			ML	NS	NS	NS	NS				
	β	LS	0.054			LS			0.012	−0.189				
		ML	0.058			ML	NS	NS	0.016	−0.472				
	θ	LS	0.264			LS	0.157	−0.045						
		ML	0.280			ML	0.169	−0.549	NS	NS	7	2289.20	−4564.4	210.15
	α			LS	0.550	LS	0.117	1.350	0.050	1.849				
No screening				ML	0.584	ML	0.150	2.508	0.122	2.701				
or order	β			LS	0.054	LS								
				ML		ML	NS	NS	NS	NS				
	θ			LS	0.264	LS	0.157	−0.045						
				ML	0.249	ML	0.066	−0.328	NS	NS	9	2293.14	−4572.28	600.12

LS stands for least squares; ML, for maximum likelihood. NS means not significant at 0.01 level.

TABLE 7*a*. Coefficient of Correlation Between the Mean Parameters and Between Each Mean and its Amplitude and Phase Angle, Sheridan, Wyoming

	α Mean	β Mean	θ Mean
α mean	1.0	0.963	−0.860
β mean		1.0	−0.958
θ mean			1.0
	θ Mean	$C_{1\theta}$	$\phi_{1\theta}$
θ mean	1.0	0.645	0.015
$C_{1\theta}$		1.0	0.564
$\phi_{1\theta}$			1.0
	β Mean	$C_{2\beta}$	$\phi_{2\beta}$
β mean	1.0	0.481	0.159
$C_{2\beta}$		1.0	0.401
$\phi_{2\beta}$			1.0

values for the Fourier coefficients for each parameter. These starting values were then used in a numerical optimization technique, ZXMIN, to find the Fourier coefficients that maximized the likelihood function. ZXMIN is a multivariate, unconstrained technique from the International Mathematical and Statistical (IMS) Library, based on a paper by *Fletcher* [1972]. A penalty function was added to the likelihood function to constrain the parameters, as specified by (9) and (6). In all cases the penalty function was only required in the early stage of the search; optimal values of the parameters were always well within the constraints.

The decision to limit the number of harmonics to two was made to conserve computer costs. Although higher harmonics may be significant for some parameters, it is unlikely that the ranking of optimization strategies would change if they were considered.

The results of the optimization, using strategies 1, 2, and 3 as described above, and a procedure without screening are shown for Tallahassee and Sheridan in Tables 6*a* and 6*b*. Column 4 in Tables 6*a* and 6*b* includes the initial (obtained by least squares) and final (obtained by maximum likelihood) values of the three parameter means when they are simultaneously optimized (cases 2 and 3). Column 6 includes the initial (obtained by least squares, case 1, and procedure without screening, or by maximum likelihood, case 2), and the final values (obtained by maximum likelihood) of three means when they are optimized along with amplitude and phase angle of each harmonic. If there are no significant harmonics, the final value of the parameter mean is equal to its initial value, and it is not shown. Columns 7–11 denote the initial (obtained by least squares) and final (obtained by maximum likelihood) values of both amplitude and phase angle of each harmonic. Insignificant harmonics are not shown.

Strategy 1 is directly comparable to the procedure that used no screening except for the order in which the parameters were optimized. For Tallahassee the results were identical because only one harmonic was significant. For Sheridan, strategy 1 resulted in a higher log-likelihood value with fewer coefficients. Strategy 1 required much less computer time for both stations. The final log-likelihood function attained and the numerical values of the amplitudes and phase angles differ for different strategies because of the dependence structure of the coefficients. The correlation between the significant coefficients was calculated by methods given by *Woolhiser and Pegram* [1979] and is shown for Sheridan and Tallahassee in Tables 7*a* and 7*b*. For Sheridan and for the other stations investigated, the correlations between the means of the parameters was higher than correlations between other coefficients. Tallahassee appears to be an exception. This suggests that procedure 3, in which the means are optimized simultaneously one time, might provide a good compromise between attaining the highest likelihood function and the procedure with minimum computer time.

The AIC statistics shown in Tables 6*a* and 6*b* are considerably smaller than the comparable statistics for Tallahassee and Sheridan in Tables 4 and 5. This suggests again that the 26-season representation is not superior to the Fourier series scheme and demonstrates the improvement that may be achieved by direct maximum likelihood optimization of Fourier coefficients, along with the rejection of insignificant harmonics.

DISCUSSION

Although it appears that the mixed exponential distribution is superior to the alternatives for the data analyzed in this study, it is possible that the ordering could change if longer records were used. Akaike's theory of model identification is applicable for a large sample. For the relatively short periods of record we used (20 to 25 years) and for 14-day periods within the year, the sample size ranged from 40 to 193. For the chain-dependent processes the sample size for precipitation, with the previous day wet, ranged from 14 to 128. *Eidsvik* [1980] found that the order of a Markov chain for the occurrence and nonoccurrence of daily precipitation, as identified by the Akaike criterion, increased as the sample size increased. With monthly intervals, he concluded that 30 years of data were not sufficient to obtain stable estimates of the Markov chain for three stations in Norway.

We have observed that increasing the record length from 20 to 40 years resulted in the identification of more harmonics in the Fourier series representation of parameters for several South Dakota stations, but budgetary and time constraints prevented an analysis of the effects of record length on model identification. This appears to be a fruitful area for further research.

CONCLUSIONS

A comparison of the chain-dependent and independent exponential, gamma, and mixed exponential distributions showed that, for five U.S. stations, the mixed exponential

TABLE 7*b*. Coefficient of Correlation Between the Mean Parameters and Between Each Mean and its Amplitude and Phase Angle, Tallahasse, Florida

	α Mean	β Mean	θ Mean
α Mean	1.0	0.021	−0.251
β Mean		1.0	−0.065
θ Mean			1.0
	θ Mean	$C_{2\theta}$	$\phi_{2\theta}$
θ Mean	1.0	0.322	−0.349
$C_{2\theta}$		1.0	−0.487
$\theta_{2\theta}$			1.0

described the distribution of daily precipitation best on the basis of the likelihood function and the Akaike information criterion for 14-day periods and on an annual basis. The independent gamma and the chain-dependent gamma distribution ranked second and third, respectively.

When seasonal variations in the parameters are fitted by finite Fourier series, the Akaike information criterion is reduced, suggesting that the composite likelihood function, using parameters estimated for 14-day periods, is not statistically better than the smaller likelihood function obtained using the much smaller number of Fourier coefficients. With Fourier representation of parameters the chain-dependent gamma distribution appears to be slightly better than the independent gamma.

The Fourier coefficients describing the seasonal variation of parameters for the mixed exponential distribution are mutually dependent. We found that a screening procedure to identify the harmonics that are likely to be significant, coupled with a sequential optimization process whereby the means are optimized simultaneously and followed by two parameter optimizations for significant harmonics, provided approximate maximum likelihood functions with acceptable computer time requirements. On the basis of the comparisons made in this paper, the mixed exponential distribution is recommended for development of regionalized models of point precipitation with seasonal parameter variation accounted for by Fourier series.

Acknowledgments. Contribution of the Agricultural Research Service, U.S. Department of Agriculture (USDA), in cooperation with the Colorado State University Experiment Station. J. Roldán received financial support from the Ministry of Education, Spain for postdoctoral training and research at Colorado State University and the Agricultural Research Service, USDA, Fort Collins, Colorado.

REFERENCES

Akaike, H., A new look at the statistical model identification, *IEEE Trans. Autom. Control, 19*(6), 716–723, 1974.

Buishand, T. A., Stochastic modeling of daily rainfall sequences, *Meded. Landbouwhogesch. Wageningen, 77-3*, 1977.

Choi, S. C., and R. Wette, Maximum likelihood estimation of the parameters of the gamma distribution and their bias, *Technometrics, 11*, 683–690, 1969.

Eidsvik, K. J., Identification of models for some time series of atmospheric origin with Akaike's information criterion, *J. Appl. Meteorol., 19*(4), 357–369, 1980.

Fletcher, R., Fortran subroutines for minimization by quasi-Newton methods, *Rep. AERE-R7125*, Atomic Energy Res. Establishment, Harwell, U.K., 1972.

Ison, N. T., A. M. Feyerherm, and L. Dean Bark, Wet period precipitation and the gamma distribution, *J. Appl. Meteorol., 10*(4), 658–665, 1971.

Katz, R. W., An application of chain-dependent processes to meteorology, *J. Appl. Prob., 14*, 598–603, 1977*a*.

Katz, R. W., Precipitation as a chain-dependent process, *J. Appl. Meteorol., 16*(7), 671–676, 1977*b*.

Mielke, P. W., Another family of distributions for describing and analyzing precipitation data, *J. Appl. Meteorol., 10*(2), 275–280, 1973.

Mielke, P. W., and E. S. Johnson, Three-parameter kappa distribution maximum likelihood estimates and likelihood ratio tests, *Mon. Weather Rev., 101*(9), 701–707, 1973.

Rider, P. R., The method of moments applied to a mixture of two exponential distributions, *Ann. Math. Stat., 32*, 143–147, 1961.

Roldán, J., and D. A. Woolhiser, Stochastic daily precipitation models, 1, A comparison of occurrence processes, this issue.

Smith, R. E., and H. A. Schreiber, Point processes of seasonal thunderstorm rainfall, 2, Rainfall depth probabilities, *Water Resour. Res., 10*(3), 418–423, 1974.

Todorovic, P., and D. A. Woolhiser, Stochastic model of daily rainfall, *U.S. Dep. Agric. Misc. Publ. 1275*, 232–246, 1974.

Todorovic, P., and D. A. Woolhiser, A stochastic model of *n*-day precipitation, *J. Appl. Meteorol., 14*(1), 17–24, 1975.

Woolhiser, D. A., and G. G. S. Pegram, Maximum likelihood estimation of Fourier coefficients to describe seasonal variations of parameters in stochastic daily precipitation models, *J. Appl. Meteorol., 18*(1), 34–42, 1979.

Woolhiser, D. A., E. Rovey, and P. Todorovic, Temporal and spatial variation of parameters for the distribution of *n*-day precipitation, in *Floods and Droughts: Proceedings of the Second International Symposium on Hydrology*, pp. 605–614, Water Resources Publications, Ft. Collins, Colorado, 1973.

(Received March 8, 1982;
revised August 2, 1982;
accepted August 10, 1982.)