WATER RESOURCES RESEARCH, VOL. 18, NO. 5, PAGES 1451–1459, OCTOBER 1982

Stochastic Daily Precipitation Models
1. A Comparison of Occurrence Processes

JOSÉ ROLDÁN

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

DAVID A. WOOLHISER

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

A first-order Markov chain and an alternating renewal process (ARP) with a truncated geometric distribution of wet day intervals and a truncated negative binomial distribution of dry day intervals are compared as models describing the occurrence of sequences of wet and dry days. Numerical optimization techniques are used to obtain approximate maximum likelihood estimates of the Fourier coefficients which describe the seasonal variation of the two Markov chain parameters and the three parameters in the alternating renewal process. For the four U.S. stations studied, the Markov chain model was superior to the ARP using the minimum Akaike information criterion.

INTRODUCTION

Most stochastic daily precipitation models are specified by a discrete occurrence process describing the sequence of wet and dry days, a continuous distribution function for the amount of precipitation given that the day is wet, and the required parameter set. Two methods have been used to account for seasonal variation in the precipitation process. In the first method, the parameters are assumed to be constant for a period varying from a few weeks to a few months. In the second method, finite Fourier series are used to describe the seasonal variability. *Woolhiser and Pegram* [1979] reviewed several models that used least squares estimates of Fourier coefficients and pointed out that this method has two major deficiencies: (1) the parameters estimated for each period (which are then fitted by the Fourier series) are statistics with differing sample variances, yet each value is given the same weight and (2) there is no sound statistical technique that can be used to decide if added harmonics are significant.

As an alternative, *Woolhiser and Pegram* [1979] proposed the use of direct numerical maximum likelihood estimates of the Fourier coefficients and a likelihood ratio test to determine the significance of added harmonics. They demonstrated the technique using a first-order Markov chain as the occurrence process for four U.S. rainfall stations and 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 occurrence process.

Before any such mapping is attempted, however, alternative occurrence processes should be compared to determine if any one process is superior. The objective of this paper is to provide such a comparison between the first-order Markov chain and another alternating renewal process. A comparison of distributions of precipitation amounts will be considered in a subsequent paper [*Woolhiser and Roldan*, this issue].

This paper is not subject to U.S. copyright. Published in 1982 by the American Geophysical Union.

Paper number 2W1225.

ALTERNATING RENEWAL PROCESSES

Daily precipitation is described by the stochastic process

$$Z_t = X_t \, Y_t \qquad t = 1, 2, \cdots$$

where $X_t = 1$ if the day is wet and $X_t = 0$ if it is dry. Y_t represents the amount of precipitation and is described by the distribution function $F_t(y) = P\,\{Y_t \leq y\}$. Therefore a realization of the precipitation occurrence process X_t is a sequence of zeros and ones.

A dry day interval of length k is defined as the sequence of k consecutive dry days bounded on each side by a wet day, and a wet interval is defined analogously; thus the occurrence process can also be considered as a sequence of alternating wet and dry intervals. A dry interval of length k beginning on the ith day is the event defined as follows:

$$\{X_{i-1} = 1, X_i = 0, X_{i+1} = 0, \cdots X_{i+k-1} = 0, X_{i+k} = 1\} \qquad (1)$$

Given that a dry interval begins on the ith day, the probability that the length L_0 is equal to k is [*Buishand*, 1977, p. 116]

$$P\{L_0 = k\} = P\{X_{i-1} = 1, X_i = 0, X_{i+1} = 0, \cdots X_{i+k-1} = 0, X_{i+k} = 1 | X_{i-1} = 1, X_i = 0\} \qquad (2)$$

Assuming that the sequence $\{X_t; t = 0, 1 \cdots\}$ is the first-order homogeneous Markov chain with the transition probability

$$p_{ij} = P(X_t = j | X_{t-1} = i) \qquad i, j = 0, 1 \qquad (3)$$

The distribution of the length of dry intervals is

$$P\{L_0 = k\} = p_{00}^{k-1}(1 - p_{00}) \qquad (4)$$

$$k = 1, 2 \cdots \qquad 0 \leq p_{00} \leq 1$$

where p_{00} is the probability of a dry day following a dry day and $(1 - p_{00})$ is the probability of a wet day following a dry day. The distribution of the length of wet intervals is

$$P\{L_1 = k\} = p_{11}^{k-1}(1 - p_{11}) \qquad k = 1, 2 \cdots \quad 0 \le p_{11} \le 1 \tag{5}$$

The distributions (4) and (5) are known as truncated geometric distributions or (TGD), with parameter $d_i = p_{ij} = 1 - p_{ii}$; $i, j = 0, 1; i \neq j$.

Buishand [1977, 1978] tested and recommended the use of the truncated negative binomial distribution (TNBD) to describe the distribution function for the length of wet and dry intervals for the precipitation occurrence process. The TNBD has the form [*Sampford*, 1965]

$$P\{L_i = k\} = \binom{k + r_i - 1}{k} \frac{b_i^{r_i}(1 - b_i)^k}{1 - b_i^{r_i}} \tag{6}$$

where $i = 0, 1, k = 1, 2 \cdots, r_i \ge -1$ and $0 < b_i \le 1$. If $r_i = 0$, we obtain the logarithmic series distribution, and if $r_i = 1$, we obtain the truncated geometric distribution.

To allow the parameters in (3), (4), (5), and (6) to vary daily in accordance with the method utilized by *Woolhiser and Pegram* [1979], we substitute for each parameter a_j the indexed parameter $a_j(n)$, where n is the day of the year $n = (t \text{ modulo } 365) + 1$. Each parameter is then expressed by the polar form of a finite Fourier series:

$$a_j(n) = A_j + \sum_{i=1}^{m} \left[C_{ji} \sin\left(\frac{2\pi ni}{365} + \phi_{ji}\right)\right] \tag{7}$$

where $j = 1, 2, \cdots, 6$, $n = 1, 2 \cdots 365$, C_{ji} and ϕ_{ji} are the amplitude and phase angle for the ith harmonic, m is the maximum number of harmonics, and A_j is the mean.

The TGD for a wet or dry interval beginning on day n can be written

$$P\{L_i(n) = k\} = d_i(n)[1 - d_i(n)]^{k-1} \qquad k = 1, 2 \cdots, i = 0, 1 \tag{8}$$

and the TNBD can be written as

$$P\{L_i(n) = k\} = \binom{k + r_i(n) - 1}{k} b_i(n)^{r_i(n)} \frac{[1 - b_i(n)]^k}{1 - b_i(n)^{r_i(n)}} \tag{9}$$

where $k = 1, 2, \cdots, i = 0, 1$, and $d_i(n)$, $r_i(n)$, and $b_i(n)$ are the parameter values for day n. When the subscript $i = 0$, we have the distribution of dry intervals and when $i = 1$, the distribution of wet intervals.

The general form of both TGD and TNBD implies that the parameters for a particular wet or dry interval are specified for a particular day of that interval, only. We assumed that the parameters are determined by the day on which the interval begins.

PROCEDURES FOR ESTIMATING PARAMETERS BY MAXIMUM LIKELIHOOD

Suppose that for N years of record, $M_i(n)$ wet or dry intervals begin on day n, and for each interval length k there are $m_i(k, n)$ intervals. The log-likelihood function for the $m_i(k, n)$ intervals of length k for the TGD can be written as

$$m_i(k, n)\{\log d_i(n) + (k - 1) \log [1 - d_i(n)]\} \tag{10}$$

By summing this expression over the total number of interval lengths, s and for 365 days in a year, we obtain the log-likelihood function for the TGD:

$$\log L_{Gi} = \sum_{n=1}^{365} \left\{ M_i(n) \log d_i(n) + \log [1 - d_i(n)] \cdot \sum_{k=1}^{s} (k - 1) m_i(k, n) \right\} \tag{11}$$

A similar technique can be used to obtain the log-likelihood function for the TNBD:

$$\log L_{Bi} = \sum_{n=1}^{365} \left\{ M_i(n)[r_i(n) \log b_i(n) - \log (1 - b_i(n)^{r_i(n)})] + \log [1 - b_i(n)] \sum_{k=1}^{s} k m_i(k, n) + \sum_{k=1}^{s} m_i(k, n) \left[\sum_{j=1}^{k} \log \left(\frac{r_i(n) + j - 1}{k - j + 1} \right) \right] \right\} \tag{12}$$

where $i = 0, 1$, $m_i(k, n)$ is the total number of intervals of length k beginning on day n, and $M_i(n)$ is the total number of intervals ($i = 0$, dry; $i = 1$, wet) beginning on day n. Note that

$$M_i(n) = \sum_{k=1}^{s} m_i(k, n)$$

and s is the maximum interval length.

Let θ_{Gi} be a vector whose elements are the coefficients of the Fourier series describing $d_i(n)$ and θ_{Bi} be a vector whose elements are the Fourier coefficients describing $r_i(n)$, and $b_i(n)$. We wish to find the estimates θ_{Gi} and θ_{Bi}, which maximize L_{Gi} and L_{Bi}.

THE OPTIMIZATION PROCESS

Because we wish to compare our results with those of *Woolhiser and Pegram* [1979], similar methodology was used. Precipitation records were first subdivided into 14-day periods, with the first period beginning March 1. We assumed that an interval belongs to a particular period if it begins on a day within that period. The extra day in a normal year was included in period 26, and the extra day in a leap year was neglected. The parameters d_i for the TGD and r_i and b_i for the TNBD were estimated by maximum likelihood methods using the Newton-Raphson iteration procedure for the TNBD. Initial estimates for the iterative solution were obtained by the method of moments [*Brass*, 1958]:

$$b_i = \frac{\bar{X}_i}{s_i^2}\left(1 - \frac{n_{1i}}{n_i}\right) \tag{13}$$

$$r_i = \frac{b_i \bar{X}_i - n_{1i}/n_i}{1 - b_i} \tag{14}$$

where

- $\bar{X}_i$ mean length of wet or dry intervals;
- s_i^2 variance of length of wet or dry intervals;
- n_{1i} the number of sample dry or wet intervals of length one;
- n_i total number of dry or wet intervals.

TABLE 1. Fourier Coefficients for Alternating Renewal Process

Station	Parameter	Distribution	Method of Estimation*	Mean	C_1	ϕ_1†	C_2	ϕ_2	C_3	ϕ_3	Number of Parameters	log L‡	Total Number of Parameters	Total log L
Wichita, Kansas	b_0	TNBD	LS	0.175	0.0644	0.0490	NS§	NS	NS	NS				
			ML	0.182	0.0645	−0.0648	NS	NS	NS	NS				
	r_0		AML	1.020							4	−2578.22		
	d_1	TGD	LS	0.596	NS	NS	NS	NS	NS	NS				
			ML	0.596	NS	NS	NS	NS	NS	NS	1	−1148.71	5	−3726.93
Kansas City, Missouri	b_0	TNBD	LS	0.222	0.0830	0.419	NS	NS	NS	NS				
			ML	0.226	0.0753	0.253	NS	NS	NS	NS				
	r_0		AML	1.070							4	−3402.56		
	d_1	TGD	LS	0.588	0.0482	2.844	NS	NS	NS	NS				
			ML	0.586	0.0482	3.072	NS	NS	NS	NS	3	−1665.97	7	−5068.53
Tallahassee, Florida	b_0	TNBD	LS	0.310	0.0369	0.520	0.1121	2.701	0.0793	0.0317				
			ML	0.318	0.0692	−0.409	0.1159	2.922	0.0388	0.726				
	r_0		AML	1.534							8	−3033.66		
	d_1	TGD	LS	0.538	0.1134	2.256	0.0692	−0.553	NS	NS				
			ML	0.536	0.1137	2.418	0.0709	−0.233	NS	NS	5	−1811.09	13	−4844.75
Sheridan, Wyoming	b_0	TNBD	LS	0.231	0.1111	1.063	NS	NS	NS	NS				
			ML	0.249	0.0629	1.116	NS	NS	NS	NS				
	r_0		AML	1.198							4	−2709.88		
	d_1	TGD	LS	0.549	0.0717	−2.277	NS	NS	NS	NS				
			ML	0.544	0.0711	−2.231	NS	NS	NS	NS	3	−1467.02	7	−4176.90
Indianapolis, Indiana	b_0	TNBD	LS	0.289	0.0857	1.583	NS	NS	0.0452	−0.0440				
			ML	0.295	0.0718	1.285	NS	NS	0.0279	0.638				
	r_0		AML	1.169							6	−2797.02		
	d_1	TGD	LS	0.547	0.0583	−1.757	NS	NS	NS	NS				
			ML	0.543	0.0569	−1.657	NS	NS	NS	NS	3	−1670.39	9	−4467.41

*LS means least squares, ML, maximum likelihood, and AML, approximate maximum likelihood.
†Phase angles are in radians.
‡Log L is the numerical value of the log-likelihood function.
§NS means not significant at 0.01 level.

Brass pointed out that these initial estimates were not unbiased. Therefore it is not unusual to obtain inadmissable estimates (i.e., $b_i > 1$ and $r_i < -1$). Maximum likelihood estimates of TGD parameters, d_i, are the reciprocal of the distribution means. After the parameter estimates were obtained for each period, they were assumed to be correct for the middle of the period, and least squares estimates of the Fourier coefficients (equation (7)) were obtained as starting values for the numerical optimization.

A multivariate, unconstrained optimization technique from the IMS (International Mathematical and Statistical) Library, called ZXMIN and based on a paper by *Fletcher* [1972], was used. Because the parameters to be optimized are bounded, a penalty function was added to the log likelihood functions (11) and (12). This function caused a large decrease in the likelihood whenever parameter values took excursions outside the appropriate range. In all cases studied, the maximum log likelihood occurred with parameter values well within the constraints. Parameters were estimated to a minimum of three significant digits. With the parameters estimated to this level of accuracy, the errors in the likelihood functions were of the order of 10^{-5} or less, according to the values of the gradient vector printed out by ZXMIN.

The likelihood functions for the wet intervals and the dry intervals were optimized independently using the following procedure. First, each parameter was considered constant throughout the year and equal to the mean maximum likelihood value calculated for the Fourier series. Then the first harmonic was added, and a three-parameter (mean, amplitude, and phase angle) optimization was performed. If the harmonic was significant, the second harmonic was added, and the procedure was repeated with a five-parameter optimization; otherwise, the first harmonic was neglected, the second was added, and a three-parameter optimization was performed.

Additional harmonics for each parameter were accepted or rejected by using the likelihood ratio test. In general, the likelihood ratio (LR) test is based on the statistic (see, for example, *Hoel* [1971] or *Mielke and Johnson* [1973])

$$\lambda = -2 \log_e \frac{L(x, \theta')}{L(x, \theta)} \tag{15}$$

Where $L(x, \theta')$ is the maximum likelihood function under

TABLE 2. Log-Likelihood Functions for Alternating Renewal Process and Markov Chain Process

Station	Alternating Renewal log L	Alternating Renewal Number of Parameters	Markov Chain log L	Markov Chain Number of Parameters
Kansas City	−5068.5	7	−5067.6	6
Tallahassee	−4844.8	13	−4846.8	10
	(−4856.6)*	11		
Sheridan	−4176.9	7	−4170.0	6
Indianapolis	−4467.4	9	−4467.5	8
	(−4472.4)*	7		

*Likelihood attained for a maximum of two harmonics.

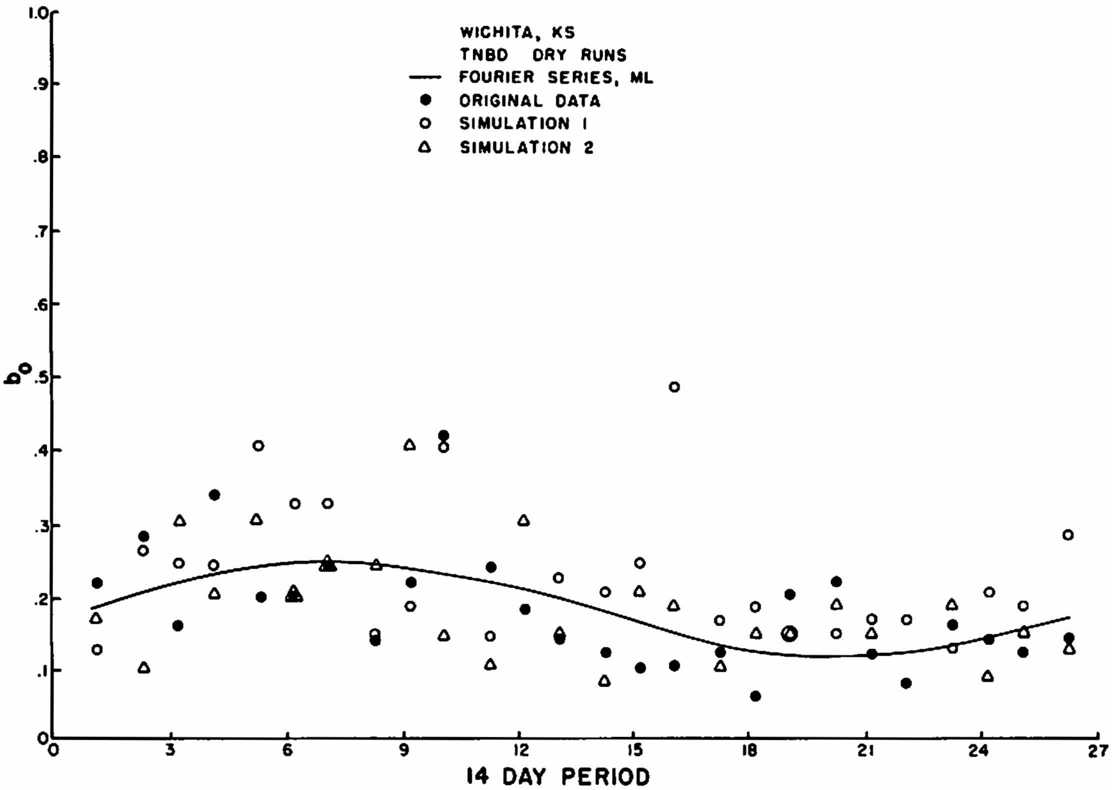

Fig. 1*a*

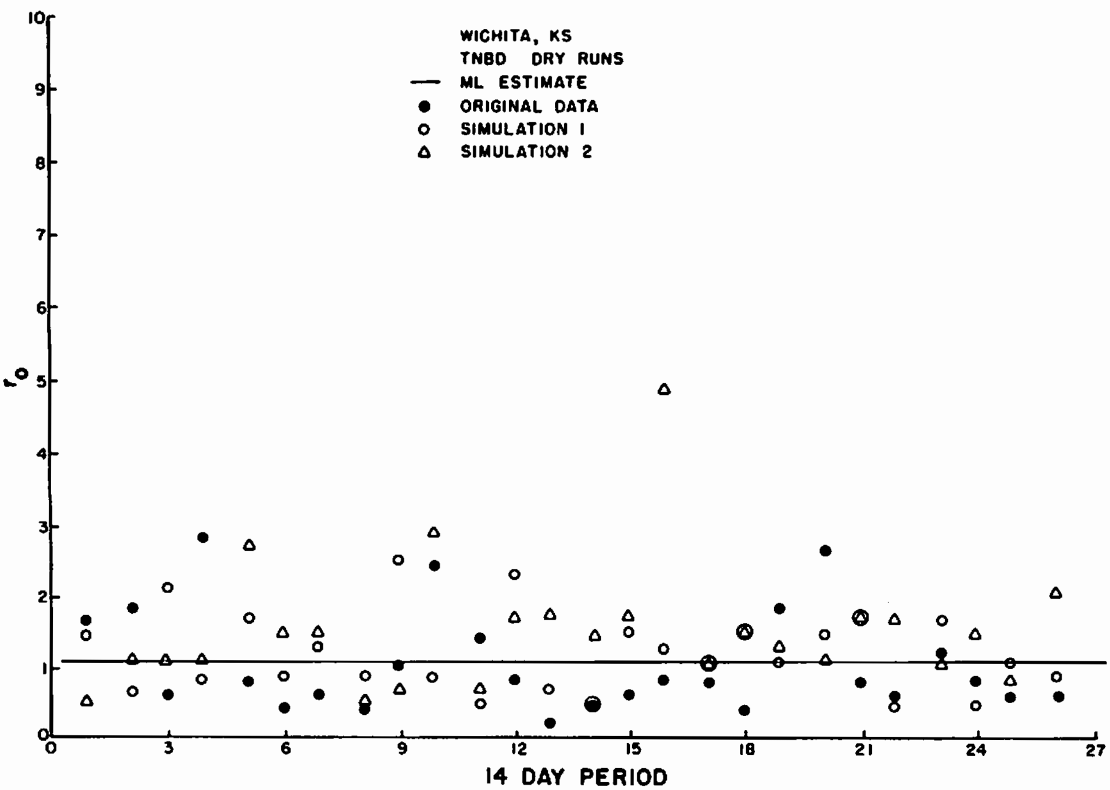

Fig. 1*b*

Fig. 1. Parameters from original and simulated precipitation series, Wichita, Kansas.

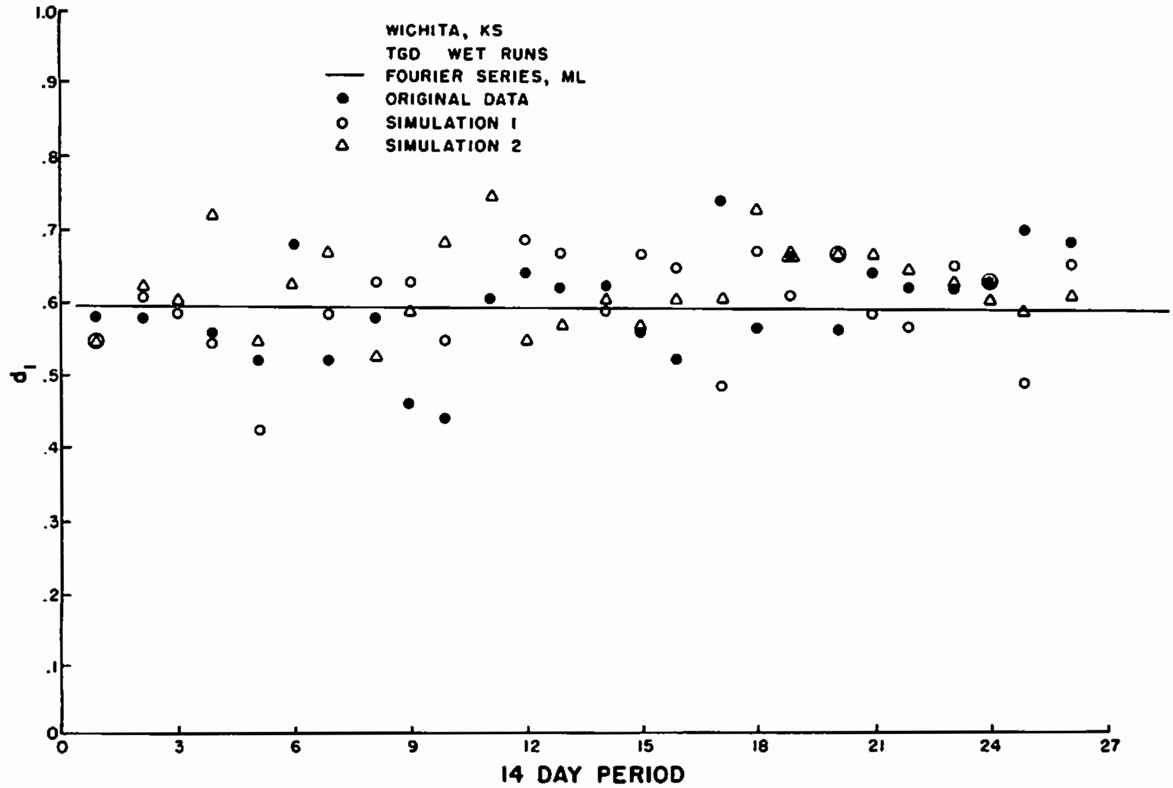

Fig. 1c

the Null hypothesis and $L(x, \theta)$ is the maximum likelihood function under the alternative hypothesis. For example, to test the significance of the first harmonic in the Fourier series describing the parameter $d_i(n)$ in the TBD, we test the Null parameter space (Null hypothesis) given by

$$H_0: \quad \theta_{Gi}' = (D_i', C_{1i} = 0, \phi_{1i} = 0)$$

against the alternative parameter space,

$$H_1: \quad \theta_{Gi} = (D_i, C_{1i}, \phi_{1i})$$

where D_i is the annual mean value of $d_i(n)$.

Under certain regularity conditions, the statistic λ has a distribution that approaches the chi square distribution for large sample size with degrees of freedom (df) equal to the number of parameters determined by H_0 (in this case, 2 df). To be consistent with *Woolhiser and Pegram* [1979], we rejected the Null hypothesis only if the probability of having a more extreme test statistic was less than 0.01. Because our procedures require repeated use of the likelihood ratio test and the likelihood ratios are 'approximate' maxima, the effective significance level is somewhat different. By using this same procedure, *Woolhiser and Pegram* [1979] found that the second harmonic was the highest significant harmonic for Markov chain parameters for the stations investigated here. Therefore we considered only the first four harmonics for the ARP.

Results of Data Analysis

Five National Weather Service daily stations were analyzed in this investigation: (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. These stations represent a range of climatic types from middle-latitude steppe (Sheridan) to humid subtropical (Tallahassee). Because they are located near average climatic boundaries, according to the Koppen system, the climatic classification of Wichita, Kansas City, and Indianapolis, may change from year to year [cf. *Trewartha*, 1954]. The last four stations were also studied by *Woolhiser and Pegram* [1979].

One of the fundamental assumptions of the ARP model is that successive wet and dry intervals are independent. We did not test the hypothesis that successive intervals are independent for these stations. However, *Buishand* [1977] and *Cole and Sherriff* [1972] concluded that they could not reject the independence hypothesis based upon correlations between the lengths of consecutive dry and wet spells for several European stations.

Our preliminary attempts to fit the TNBD to the wet intervals revealed some problems that had been previously identified by *Buishand* [1977] but which are more serious in this case because we have much smaller sample sizes because of the shorter records and the use of 14-day rather than monthly periods. First, we found that the likelihood equations had no solution within the parameter space for several periods (for example, this occurred for 10 periods for the Kansas City record). Specifically, the method of moments estimate given by (13) and (14) gave values of $r_i < -1$ and $b_i > 1.0$, so they could not be used as starting values. Under these circumstances, when acceptable values of r_i and b_i were used as starting values, the iterative solution quickly gave unacceptable values. Second, we found a very large and apparently random variability for the acceptable values of the parameter r_1 (for example, r_1 varied from 0.49 to 107 for the Indianapolis record). Part of this excessive variability

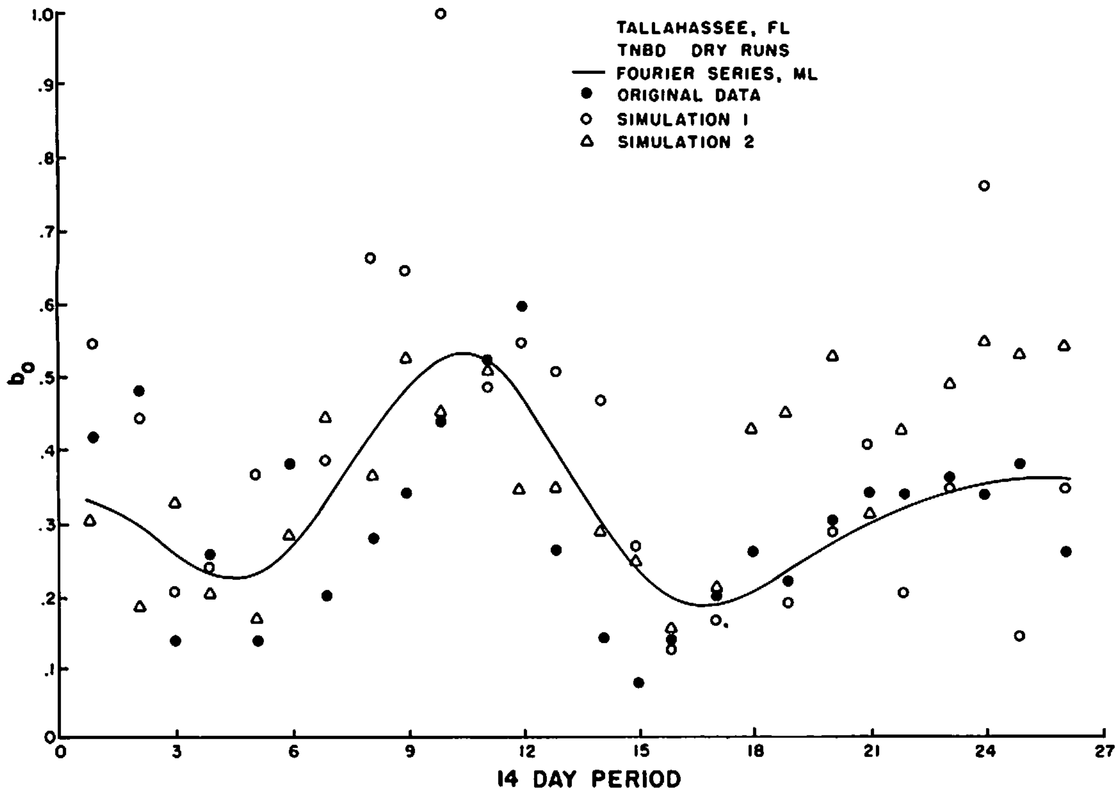

Fig. 2*a*

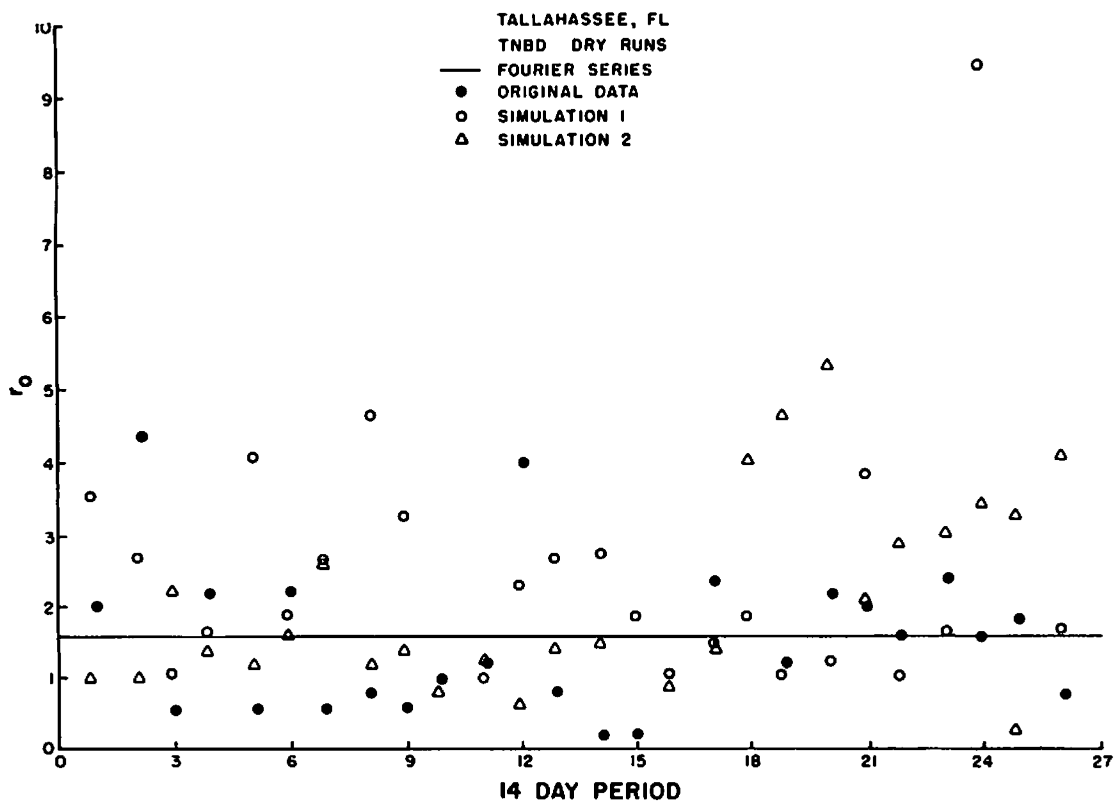

Fig. 2*b*

Fig. 2. Parameters from original and simulated precipitation series, Tallahassee, Florida.

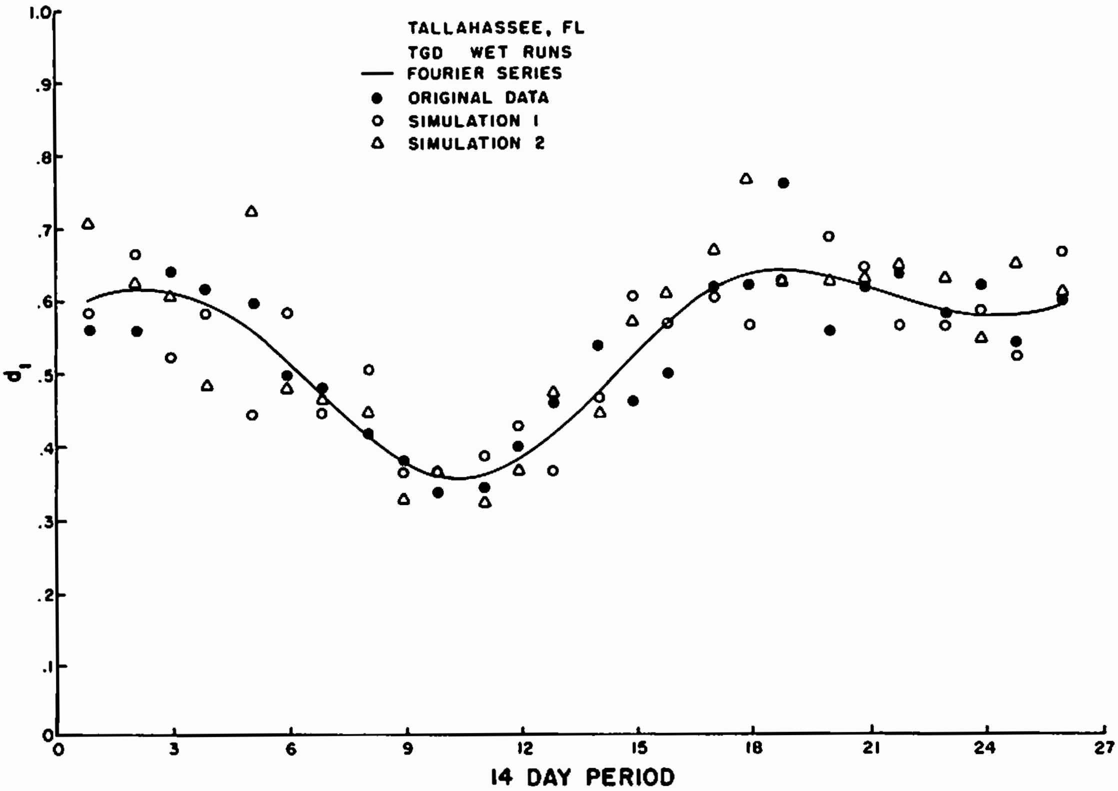

Fig. 2c

is due to the small sample size and part is due to the strong interaction between the parameters b_i and r_i. *Buishand* [1977] found that the simple correlation between b_i and r_i was greater than 0.9 from long (> 65 year) records in the Netherlands. Therefore the range of b_i and r_i values within the 95% confidence region can be quite large for a given period.

At this stage of the investigation, we had two options: (1) assume that r_1 was constant throughout the year and determine the value of r_1 and the 26 values of b_1 that maximized the sum of the 26 log-likelihood functions for each period, or (2) abandon the TNBD for the wet intervals and use the TGD (i.e., assume $r_1 = 1$). Because $r_1 = 1$ was almost invariably included in the 95% confidence region for the periods for which we could obtain maximum likelihood estimates of b_1 and r_1 and a choice of $r_1 = 1$ reduced the number of parameters, we chose to adopt the TGD for wet intervals.

For dry intervals, we followed *Buishand*'s [1977] procedure and assumed that r_0 was a constant throughout the year and equal to the average of the maximum likelihood estimates for each 14-day period. These estimates will be denoted as approximate maximum likelihood (AML) estimates because the likelihood is not fully maximized over the parameter space. Buishand found that the AML estimates of r_0 were always within the 95% confidence region of the ML estimate of r_0.

The parameters obtained by the optimization procedure for the TNBD and TGD are shown in Table 1. The row in Table 1 labeled LS denotes the Fourier coefficients obtained by least squares fitting to parameter values estimated for 14-day periods. The row labeled ML denotes the coefficients obtained using a numerical technique (ZXMIN) to maximize the log-likelihood functions. Only significant harmonics are shown. The climatic complexity of a station is revealed by the number of harmonics required for each parameter; Wichita, for example, requires only five parameters, while Tallahassee requires 13. A comparison of the ML and LS estimates of Fourier coefficients shows that there can be substantial differences. Therefore LS estimates may be far from optimal in the likelihood sense.

The optimized coefficients were utilized in a simulation model to generate two additional series equal to the length of record at each station. The simulated records were then used to estimate the parameters b_0 and r_0 for each 14-day period for the TNBD and the parameter d_1 for the TGD. The sample parameter values obtained from the original records and from two simulated records, along with the maximum likelihood Fourier representation of the parameters, are shown for Wichita and Tallahassee in Figures 1 and 2. The scatter about the Fourier series curve is quite similar for the original data and the simulated records, suggesting a satisfactory identification of the parameters of the alternating renewal process.

COMPARISON WITH MARKOV CHAIN

The distribution of wet intervals for the alternating renewal process described by the TGD is very similar to that for the two-state, first-order Markov chain. The only difference is that the parameter for the TGD is specified by the first day of the interval, while for the first-order Markov chain the parameter may change each day during the interval. For example, for a k-day sequence of wet days beginning on day n, we have for the TGD,

$$P[L_1(n) = k] = d_1(n)[1 - d_1(n)]^{k-1} \tag{16}$$

TABLE 3. AIC for Alternating Renewal and Markov Chain

Station	Alternating Renewal	Markov Chain
Kansas City	10151.0	10147.2*
Tallahassee	9715.6 (9735.2)†	9713.6*
Sheridan	8367.8	8352.0*
Indianapolis	8952.8 (8958.8)†	8951.0*

*Minimum AIC.
†Maximum of two harmonics.

and for the Markov chain,

$$P[L_1(n) = k] = [1 - p_{11}(n + k)] \prod_{j=1}^{k-1} p_{11}(n + j) \quad (17)$$

As a practical matter, the difference in the probabilities is very small. For instance, using data from Kansas City for $n = 130$ and $k = 3$, we obtain

TGD:

$$P[L_1(130) = 3] = 0.099056$$

Markov chain:

$$P[L_1(130) = 3] = 0.099058$$

If the parameter r_0 approaches one, the parameter b_0 in the TNBD is analogous to d_0 in the TGD. Therefore the TNBD approaches the TGD, and a comparison of the fit of the alternating renewal process with the Markov chain would only test the sensitivity of the likelihood function to the daily parameter variation within a run implicit in the Markov chain formulation. Otherwise, $b_0 \neq d_0 = 1 - p_{00}$, and differences between both likelihood functions are due to other factors, as well as to the daily parameter variation. For example, for Kansas City, when $n = 130$, $k = 3$, and $r_0 = 1.07$ (See Table 1), we obtain for the TNBD,

$$P[L_0(130) = 3] = 0.13528$$

by using (9) with $i = 0$, and for the Markov chain,

$$P[L_0(130) = 3] = 0.13327$$

by using (17) with $i = 0$. The difference between these probabilities is larger than those for the wet runs, although r_0 is very near to 1 and can lead to significant differences in the likelihood function.

Likelihood functions for the Markov chain (MC) process, obtained by *Woolhiser and Pegram* [1979] and the alternating renewal process (AR) for four stations, are shown in Table 2. For the Markov chain, no harmonics higher than the second were significant, while the third harmonic was significant for b_0 for both Indianapolis and Tallahassee. In Table 2, log-likelihood functions are shown in parentheses for both stations for a maximum of two harmonics. A comparison of the likelihood functions shows that the alternating renewal process does not significantly improve the likelihood function (For Sheridan and Kansas City, the AR process has a lower log-likelihood function with one more parameter, although this may be due to the fact that r_0 is not fully optimized over the parameter space.). A choice can be made between the two competing models by selecting the model which provides the minimum information theoretical criterion AIC [*Akaike*, 1974] where

$$\text{AIC} = (-2) \log (\text{maximum likelihood}) + 2k \quad (18)$$

where k is the number of independently adjusted parameters. As shown in Table 3, the Markov chain model results in the minimum AIC for each station. Because the maximum likelihood expressions for the alternating renewal process with the TNBD are much more complicated, the estimation of parameters takes considerably more computer time than for the Markov chain. Therefore economic as well as statistical considerations suggest that the Markov chain process is superior to the alternating renewal process with the TNBD for dry periods and the TGD for wet periods for the stations and for the record lengths investigated in this study.

Considering that a first-order Markov Chain results in intervals distributed as a TGD, this finding appears to conflict with the results of *Buishand* [1977, 1978], who found that the TNBD was superior to the TGD on the basis of chi square tests for monthly periods for rainfall records from the Netherlands, Germany, Belgium, India, Indonesia, Surinam, and Egypt. Possible reasons for the different conclusions include the following.

1. The U.S. stations may have precipitation occurrence characteristics different from those of the stations studied by Buishand, and tests identical to those he used could lead to the conclusion that the TGD is superior to the TNBD.

2. The flexibility introduced by allowing the Markov chain parameters to vary daily within a wet or dry run, as contrasted to the TNBD, where the parameters are associated with the first day of a run, may provide a superior fit.

3. The use of the likelihood statistic, rather than the chi square test, may lead to different conclusions.

4. The much shorter record length used in this study (20–25 years, as compared to 31–104-year records used by *Buishand* [1977]) and the shorter periods (14 days versus months) did not provide AML estimates of r_0 that are close to the ML estimates.

The first possibility seems unlikely, considering the wide range of climatic characteristics of the U.S. stations. The difference is probably due to a combination of factors 2, 3, and 4. Additional analyses would be required to test this hypothesis.

Factor 4, length of record, has been shown to be a significant factor in model selection by *Chin* [1977] and *Eidsvik* [1980]. Both investigators used the Akaike model identification procedure and found that higher order Markov chains were selected as the length of record increased. Eidsvik also found that the estimated order of the Markov chain increased with the length of the interval defining the season. He attributed this result to either low frequency variations and/or a methodological tendency to underestimate the order when the sample size is small. From these results it appears plausible that the more complicated ARP would have been selected had we used longer records.

On the other hand, Buishand did find that, for many periods, the likelihood equations for the TNBD had no solution within the parameter space for wet periods. We also found this to be true for the U.S. stations investigated. Although this problem can be overcome by assuming that the parameter r_1 is constant throughout the year, it does require different procedures than those used for dry intervals and makes the computer programing more difficult. The use of Fourier series to allow the parameters to vary daily, as well as parsimony with respect to the number of parameters, make it advantageous to choose one distribution only for all dry or wet periods.

SUMMARY AND CONCLUSIONS

A first-order Markov chain and an alternating renewal process with truncated geometric distribution of wet day intervals and a truncated negative binomial distribution of dry intervals were compared as models describing the occurrence of sequences of wet and dry days for five U.S. rainfall stations. The two Markov chain parameters and two of the three parameters of the alternating renewal process were allowed to vary daily, as described by a Fourier series. The third parameter in the alternating renewal process, r_0, was estimated by approximate maximum likelihood techniques, as were the Fourier coefficients. The likelihood functions for the alternating renewal process were not significantly greater than those for the Markov chain, and the Markov chain resulted in the minimum *Akaike* [1974] information criteria for the four stations that were compared. Because the numerical optimization technique to estimate Fourier coefficients by the method of maximum likelihood required considerably more computer time for the alternating renewal process than for the Markov chain, economic as well as statistical considerations suggest that the Markov chain model is superior for the stations considered for record lengths of 20 to 25 years. The likelihood ratio test at the 0.01 level and the Akaike information criterion gave similar results when used to determine the number of significant harmonics.

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

REFERENCES

Akaike, H., A new look at the statistical model identification, *IEEE Trans. Autom. Control, 19*(6), 716–723, 1974.
Brass, W., Simplified methods of fitting the truncated negative binomial distribution, *Biometrika, 45*, 59–68, 1958.
Buishand, T. A., Stochastic modeling of daily rainfall sequences, *Meded. Landbouwhogesch. Wageningen, 77-3*, 1977.
Buishand, T. A., Some remarks on the use of daily rainfall models, *J. Hydrol., 36*, 295–308, 1978.
Chin, E. H., Modeling daily precipitation occurrence process with Markov chain, *Water Resour. Res., 13*(6), 949–956, 1977.
Cole, J. A., and J. D. F. Sherriff, Some single- and multi-site models of rainfall within discrete time increments, *J. Hydrol., 17*, 97–113, 1972.
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.
Hoel, P. G., *Introduction to Mathematical Statistics*, 4th ed., John Wiley, New York, 1971.
Mielke, P. W., and E. S. Johnson, Three-parameter kappa distribution maximum likelihood estimates and likelihood ratio tests, *Mon. Weather Rev., 10*(9), 701–707, 1973.
Sampford, M. R., The truncated negative binomial distribution, *Biometrika, 42*, 58–69, 1955.
Trewartha, G. T., *An Introduction to Climate*, 3rd ed., McGraw-Hill, New York, 1954.
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*, 34–42, 1979.
Woolhiser, D. A., and J. Roldán, Stochastic daily precipitation models, 2, A comparison of distributions of amounts, *Water Resour. Res.*, this issue.

(Received January 29, 1982;
revised August 2, 1982;
accepted August 10, 1982.)