VOL. 9, NO. 4 WATER RESOURCES RESEARCH AUGUST 1973

Point Processes of Seasonal Thunderstorm Rainfall

1. Distribution of Rainfall Events

R. E. Smith and H. A. Schreiber

USDA Southwest Watershed Research Center, Agricultural Research Service
Tucson, Arizona 85705

The hot rainy season marked by local scattered thunderstorms from June to September is typical of most of the lower elevations of the Sonoran and Chihuahuan regions of southwestern North America. This rainy season is analyzed by using long-term historical daily records to obtain insight concerning the underlying stochastic process. By using historical data from three scattered points in this region, we computed the discrete series of daily Bernoulli parameters and daily first-order Markov transition probabilities. The hypothesis of sequential independence versus a first-order Markov dependence hypothesis is tested by comparison of analytically derived distributions for dependent random variables generated by the nonstationary processes. These include wet and dry run lengths, occurrence of the first wet day in the season, number of runs per season, and total number of rainfall days per season. The comparative analysis of historical data indicates that (1) the Markov chain model is generally significantly superior to the Bernoulli model (which extends results of similar analyses by others from regions of largely frontal-type storms) and (2) year-to-year variations in the process require additional probabilistic descriptions, indicated by annual variance in number of rain days and significant annual changes in autocorrelation properties.

A hot rainy season marked by local scattered thunderstorms from June to September is an important hydrologic period in the lower elevations of southeastern Arizona and other parts of the Sonoran and Chihuahuan regions. This is also the growth period for most of the local plant species. In a previous paper [*Schreiber and Sutter*, 1972], rainfall in this area was analyzed to develop a model predicting hypothetical periods of 'available' soil water on the assumption of knowledge of several concurrent processess: infiltration, time of rain within a day, depth of plant rooting, and evaporation. Basically, it was an accounting process with no special characterization or examination of rainfall itself. The same records of daily precipitation are used in this paper to study in detail the stochastic input of water isolated from other aspects of the hydrology. A subsequent paper (part 2) will present the distribution of daily rainfall amounts, the distribution of rainfall amounts within arbitrary periods, and the effect of depth truncation on the stochastic process.

Rainfall Models

Rainfall is governed by physical laws and complex atmospheric processes. The fact that these causative processes are extremely complex and spatially and temporally dependent makes prediction of rainfall practically impossible. The complexity of the process, however, allows a probabilistic description of a local variable such as rainfall depth, and statistical analysis of this random variable provides a prediction of statistical properties of future rainfall.

The construction of a probabilistic rainfall model should reflect how the model will be used. Variations in small rainfall amounts are necessarily important to most rangeland plant species, whereas these variations might be unimportant in a model designed to predict large basin water yield or a stochastic model of runoff. The 'threshold' of 'significant' rains used by some [*Lane and Osborn*, 1972; *Duckstein et al.*, 1972] eliminates many plant important rains and presents an incomplete picture of rainfall distribution (discussed in part 2).

In this work we will treat thunderstorm rainfall probabilistically by proposing an underlying stochastic process and deriving several dependent random variables that are consequences of the process, such as number of crossings (dry and wet runs) and number of events per season. The distribution of each dependent random variable will be tested against historical records, and the underlying stochastic model will thereby be evaluated.

STOCHASTIC PROCESS OF RAINY DAY OCCURRENCES

Previous studies of rainfall in the semiarid region indicated above [e.g., *Fogel and Duckstein*, 1969] have been concerned primarily with rainfall associated with observed runoff (a parameter only partially dependent on the rainfall process) and have dealt with areal distribution, storm center location, and other areal properties. These investigations have been based on relatively short (12 to 15 year) records from dense point samples within an isolated area. The analysis reported here is concerned solely with properties of point rainfall, which is the perspective of a plant or small plot, and takes advantage of relatively long point records (55–73 years) available through U.S. Weather Bureau data. The location of the three gages used is indicated in Figure 1. Table 1 lists other descriptive data for these three locations.

Daily total rainfall of 0.01 inch or more is considered here as a rainfall event; distinction is not made between two storms in 1 day. From unpublished data for the Walnut Gulch watershed at Tombstone (D. L. Chery, personal communication, 1972), multiple-storm days apparently constitute only 10–15% of the 'rainy' days. Nevertheless, one should not consider this study to apply in all details to individual storm events. Because summer storms occur almost exclusively in the afternoon or late evening, a daily gage read in the morning cannot be considered to divide one storm into 2 days.

The period of summer thunderstorms in the study area typically includes June, July, and August. Most of the analyses discussed in the following section could apply to the entire year, but since approximately 70% of the average annual rainfall at Tombstone occurs in these months (and produces most of the forage), a 122-day period beginning June 1 will be the subject of our discussion.

Consider the sample frequency f_i of point rainfall event occurrence e on a day i, $i = 1, 2, \cdots, 122$, defined as

$$f_i = P(e_i = w) = 1 - P(e_i = d) \quad (1)$$

in which w represents a wet day and d a dry day. Figure 2 illustrates the variation in f_i for the 122-day period for Tombstone from 73 years of record. If rainy days are assumed to occur independently in time, the daily value of f_i from Figure 2 will describe the parameter for a nonstationary binomial process or uncondi-

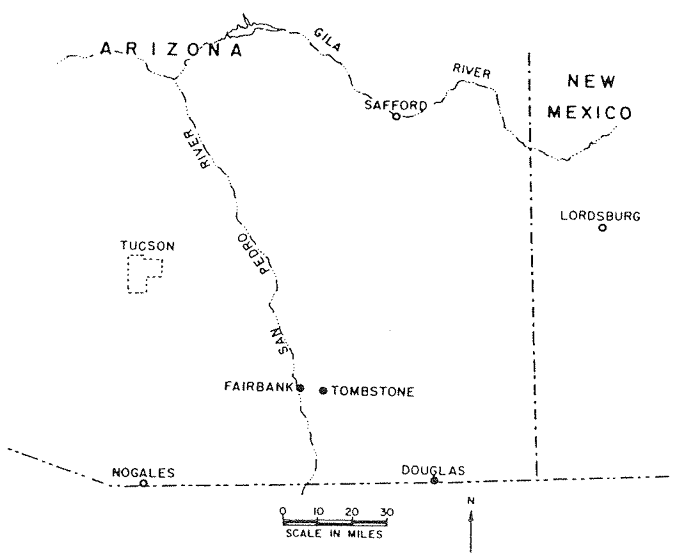

Fig. 1. Location of sampling sites (solid dots) of daily precipitation records used.

TABLE 1. Data for Sampling Stations

Station	Latitude	Longitude	Elevation, feet	Years of Record Used
Fairbank, Ariz.	31°43′	110°11′	3862	55
Tombstone, Ariz.	31°43′	110°04′	4540	73
Douglas, Ariz. (smelter)	31°21′	109°35′	3973	68

tional (Bernoulli) probability model for rainy day occurrence.

This is perhaps the simplest stochastic model that one might conceive for this process, but it remains to be shown if it is an accurate model. The simplest alternative is a simple Markov chain, in which the probability of a rainy day is a function of the occurrence or nonoccurrence of rain on the previous day. This is expressed by transition probabilities

$$f_{w_i} = P(e_i = w \mid e_{i-1} = w)$$

and

$$f_{d_i} = P(e_i = d \mid e_{i-1} = d)$$

These values form a matrix of transition probabilities:

$$\begin{array}{c|cc} & \multicolumn{2}{c}{e_i} \\ e_{i-1} & w & d \\ \hline w & f_w & 1 - f_w \\ d & 1 - f_d & f_d \end{array} \qquad (2)$$

Feller [1959] discusses Markov chains in detail. Various writers, including *Weiss* [1964], *Gabriel*

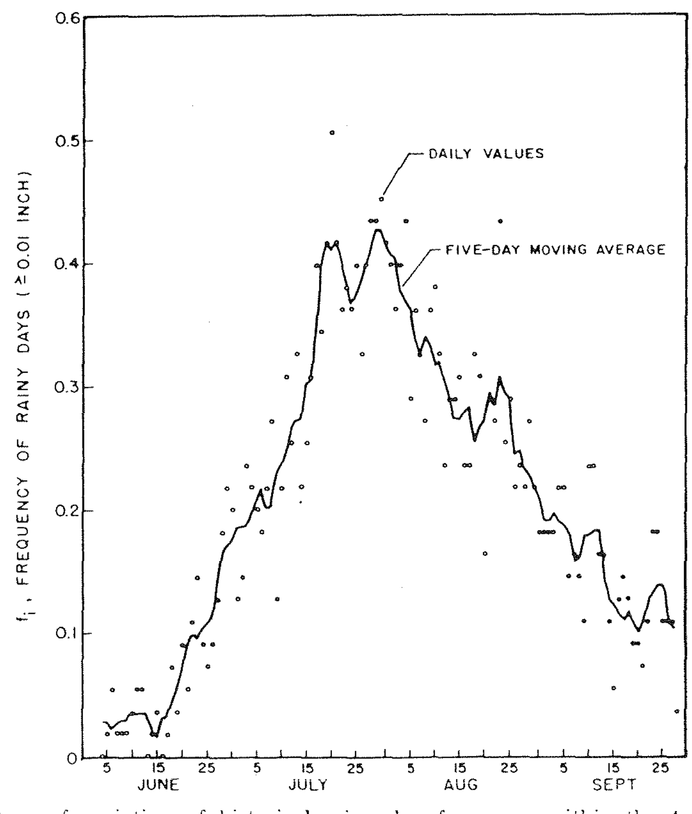

Fig. 2. Pattern of variation of historical rainy day frequency within the 4-month season at Tombstone.

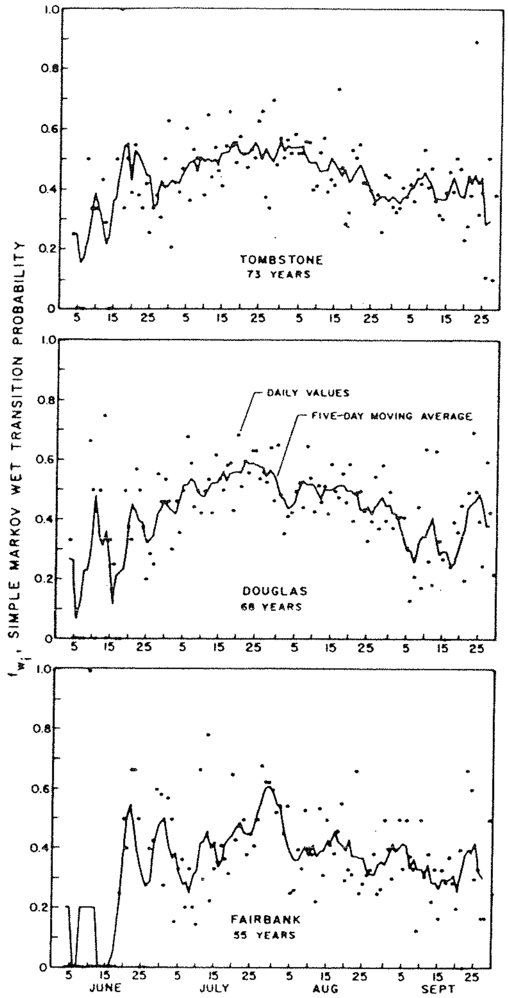

Fig. 3. Seasonal pattern of wet transition probability f_{w_i} of simple Markov chains for all three stations studied.

and Neumann [1962], and *Hershfield* [1970], have found that what may be called frontal storm rainfall may be described by a Markov chain. It has not been established how well such a chain applies to air mass thunderstorm rainfall. The f_{w_i} in matrix 2 is determined from historical records as

$$f_{w_i} = Y_{w_{i,i-1}} / Y_{w_{i-1}} \quad (3)$$

where $Y_{w_{i,i-1}}$ is the number of years both day i and day $i - 1$ were wet and $Y_{w_{i-1}}$ is the number of years day $i - 1$ was wet. Likewise, for transitions from dry day to dry day,

$$f_{d_i} = Y_{d_{i,i-1}} / Y_{d_{i-1}} \quad (4)$$

where $Y_{d_{i,i-1}}$ is the number of years both day i and day $i - 1$ were dry and $Y_{d_{i-1}}$ is the number of years day $i - 1$ was dry. The sequences f_{d_i} and f_{w_i}, $i = 1, 2, \cdots, 122$ are presented in Figures 3 and 4 for the three sampling sites used.

STOCHASTIC PROCESS—DEPENDENT RANDOM VARIABLES

Distribution of wet and dry run lengths. The alternative models (sequentially independent and dependent) for rainy day occurrence are first compared by deriving distributions for wet run lengths and dry run (drought) lengths under each assumption and comparing these predictions with historical data. *Yevjevich* [1972] suggests such a method of investigation and presents several references regarding run length statistics.

On the assumption of sequential independence of events (the Bernoulli model) it is easily shown that the probability of occurrence of a wet run w of length k days under constant f_i from (1) for the period beginning at day i is

$$p_i(w = k) = f_i^{k-1}(1 - f_i) \quad (5)$$

Likewise, for a drought d of length k given a constant f_i beginning at day i, the probability of occurrence would be

$$p_i(d = k) = f_i(1 - f_i)^{k-1} \quad (6)$$

For significantly changing f over the k days following day i the wet run probability would be

$$p_i(w = k) = (1 - f_{i+k}) \prod_{j=i+1}^{i+k-1} f_j \quad (7)$$

and the drought probability would be expressed as

$$p_i(d = k) = f_{i+k} \prod_{j=i+1}^{i+k-1} (1 - f_j) \quad (8)$$

An equation analogous to (5) presented by *Gabriel and Neumann* [1962] applies to the Markov model:

$$p_{w_i}(w = k) = (1 - f_{w_{i+k}}) f_{w_i}^{k-1} \quad (9)$$

and, similarly for droughts,

$$p_{d_i}(d = k) = (1 - f_{d_{i+k}}) f_{d_i}^{k-1} \quad (10)$$

When significant changes in f_d and f_w over the period k are considered, (9) and (10) become

$$p_{w_i}(w = k) = (1 - f_{w_{i+k}}) \prod_{j=i+1}^{i+k-1} f_{w_j} \qquad (11)$$

and

$$p_{d_i}(d = k) = (1 - f_{d_{i+k}}) \prod_{j=i+1}^{i+k-1} f_{d_j} \qquad (12)$$

A simple average over the total season was used to compute f, f_w, and f_d for each day of the period. These values were used in (5), (6), (9), and (10) for a simple comparison of the two stochastic models. Results are shown in Figures 5 and 6 for wet and dry run lengths, respectively.

The same equations were used, the season being divided into more closely homogeneous periods, and run length distributions were compared for runs whose center fell in each period. Goodness of fit to historical data was tested by using a χ^2 test; results of these tests are shown in Table 2 for both wet and dry runs. Variation in wet run length distribution during the season is naturally somewhat opposite to that in dry run length distribution, the longest observed lengths of drought occurring in the early part of the season and the longest wet runs occurring in the middle of the season.

Table 2 shows first that the χ^2 test comparing historical dry run length distributions to predictions is insufficiently sensitive to discriminate between the two models tested. This problem is due primarily to the small values for $1 - f_{d_i}$ and the resulting more extended distribution as well as to the often comparable values for f_i and $1 - f_{d_i}$ in (6) and (10). In most cases the statistical test accepted both distributions for dry runs at the 1% level. Comparison of wet run length distributions is quite conclusive, however. In all but two cases (Tombstone) the Markov chain was accepted and the Bernoulli model rejected at the 1% level. Furthermore, since the χ^2 parameter is the normalized sum of the squares of the deviations, one can compare magnitudes of this parameter with some meaning. In all but a few cases the χ^2 parameter is smaller for the Markov chain than for the Bernoulli (independent) model by an order of magnitude or more. These comparisons leave little doubt about the general inadequacy of the assumption [e.g., *Duckstein et al.*, 1972; *Lane and Osborn*, 1972] that thunderstorm rainfall is a sequentially independent phenomenon.

Further comparison can be made by considering the mean number of runs (wet or dry) in a season as predicted by each model, and the results are again striking. For a sequentially independent binomial process the expected value for the number of wet runs $r_{w,b}$ for uniform probability $\bar{f}$ is

$$E(r_{w,b}) = \bar{f}(1 - \bar{f})N \qquad (13)$$

in which N is the sample size (days of season). By considering the probability of transition, one

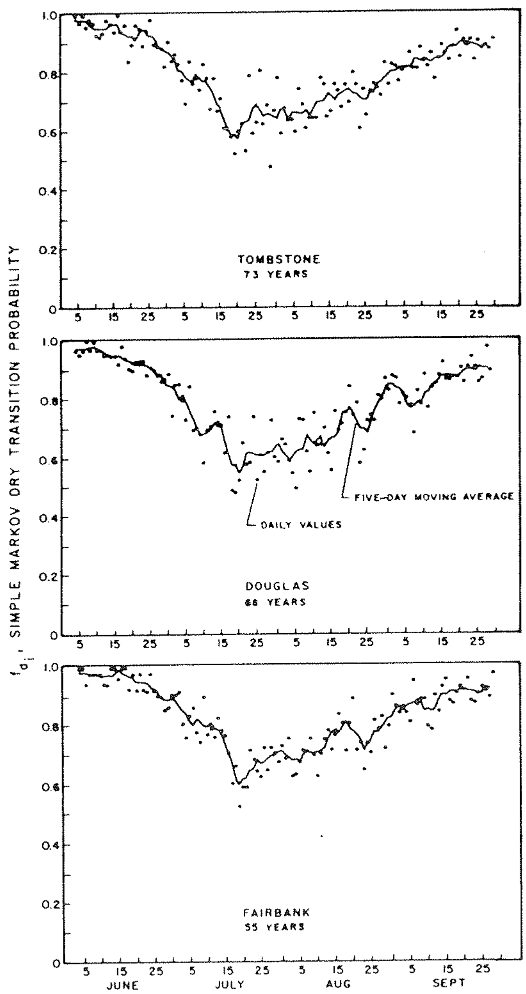

Fig. 4. Seasonal pattern of dry day transition probability f_{d_i} for a simple Markov chain for all three stations studied.

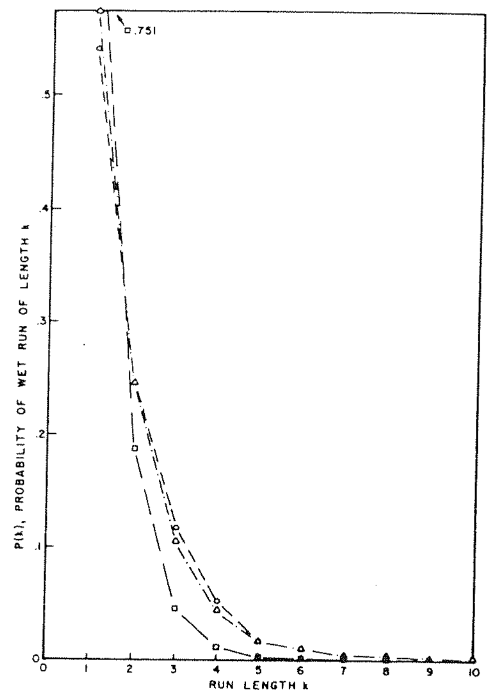

Fig. 5. Distribution of wet day run lengths from record and as predicted by (5) and (9) by using seasonal average values of f_i and f_{w_i}. Circles represent frequency from historical data. The season-averaged models are represented by squares (probability predicted by a sequentially independent, or Bernoulli, model) and triangles (probability predicted by a simple Markov chain).

may estimate the expected values of run numbers for a simple Markov chain for dry runs $r_{d,m}$ and wet runs $r_{w,m}$, whose value should be asymptotically equal, as follows:

$$E(r_{d,m}) = \bar{f}(1 - \bar{f}_w)N \quad (14)$$

$$E(r_{w,m}) = (1 - \bar{f})(1 - \bar{f}_d)N \quad (15)$$

As an example, for the Tombstone gage, $\bar{f}$ is 0.25, $\bar{f}_d$ is 0.795, and $\bar{f}_w$ is 0.424. Thus (13) predicts $\bar{r}_w = 23$ runs per year. Equations 14 and 15 predict 17 and 18 runs/year, respectively. The value of $\bar{r}_w$ for the historical data is 16.2.

Start of season. Given the defined 122-day period with nonhomogeneous unconditional and transition probabilities, the day i, $i = 1, 2, \cdots, 122$, on which the first rain falls will define a start of season. The day on which the season starts (and the amount of rain involved) could have significance to those plants depending on the thunderstorm season for their growth. It has been observed (Southwest Watershed Research Center, Agricultural Research Service, unpublished data, 1959–1972) that failure to receive rain in Tombstone until early August, or about the 70th day of the defined period, can mean that perennial grass plants die. The data indicate a small probability for this catastrophe. An interaction between first rains and their amounts can be presumed to exist because a

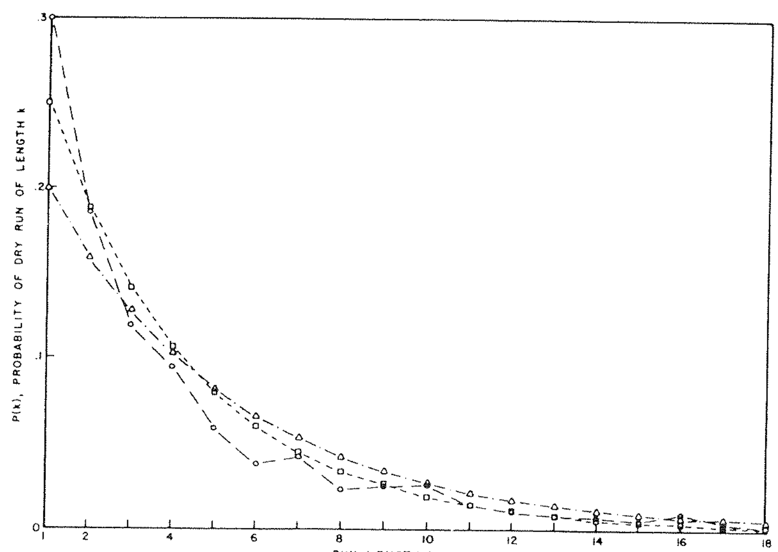

Fig. 6. Distribution of dry day run lengths from record and as predicted by (6) and (10) by using seasonal average values of f_i and f_{d_i}. Circles represent frequency from historical data. The season-averaged models are represented by squares (probability predicted by a sequentially independent, or Bernoulli, model) and triangles (probability predicted by a simple Markov chain).

TABLE 2. Computed χ^2 Values for Assumed Distributions of Run Lengths

Season Day Period	Fairbank						Douglas						Tombstone					
	Dry			Wet			Dry			Wet			Dry			Wet		
	Ind.	Dep.	d.f.	Ind.	Dep.	d.f.	Ind.	Dep.	d.f.	Ind.	Dep.	d.f	Ind.	Dep.	d.f.	Ind.	Dep.	d.f.
2–21	0.405*	0.780*	3	0.194	0.0206	1	0.337*	0.582*	5	0.709	*0.000027*	1	*0.922*	1.49*	5	1.506	0.0076	1
22–31	*0.155*	*0.312*	5	1.15	*0.0204*	2	0.225	0.211	6	0.372	0.0404	2	*0.190*	*0.213*	8	2.66	*0.00071*	2
32–41	*0.083*	*0.102*	6	0.053	0.0064	1	*0.0198*	*0.0426*	4	0.0603	0.0296	2	*0.133*	*0.105*	5	0.227	*0.0164*	2
42–51	*0.0339*	*0.0590*	3	0.031	*0.0000074*	2	0.118	*0.0420*	3	0.162	0.0262	2	*0.0256*	*0.0413*	3	*0.0638*	*0.0032*	3
52–61	*0.0567*	*0.0590*	4	0.062	*0.0078*	2	*0.0723*	*0.0033*	3	0.122	*0.0090*	3	*0.0438*	*0.0030*	3	0.1346	*0.0091*	3
62–71	*0.105*	*0.074*	4	0.054	*0.006*	2	*0.0270*	*0.0322*	3	0.0456	*0.0199*	2	*0.0361*	*0.0160*	3	*0.0934*	*0.0253*	3
72–81	*0.103*	*0.063*	4	0.122	*0.0112*	2	*0.089*	*0.0339*	4	0.0797	*0.00071*	2	*0.106*	*0.0181*	5	0.118	*0.0023*	2
82–91	*0.080*	*0.107*	4	0.0011	0.024	1	*0.0455*	*0.0334*	5	0.1768	*0.0109*	2	*0.193*	*0.0798*	4	0.0317	*0.0054*	2
92–101	*0.045*	*0.064*	9	0.601	*0.0144*	2	*0.0804*	*0.0531*	6	0.123	*0.0167*	2	*0.0684*	*0.0417*	6	0.216	*0.0065*	2
102–111	*0.108*	*0.145*	7	0.180	*0.000056*	1	*0.035*	*0.0501*	6	0.212	*0.00032*	2	*0.280*	*0.115*	8	1.151	*0.0194*	2
112–121	0.654*	1.08*	3	0.213	0.0284	1	0.116*	0.424*	5	2.69	*0.0026*	2	0.291*	0.591*	4	0.634	0.0216	2

Ind. represents daily sequential independence.
Dep. represents daily dependence of a simple Markov chain.
The italic data indicate tests with less than 1% chance of random correspondence.
* Data include a significant number of truncated run lengths resulting from the limits of the season.

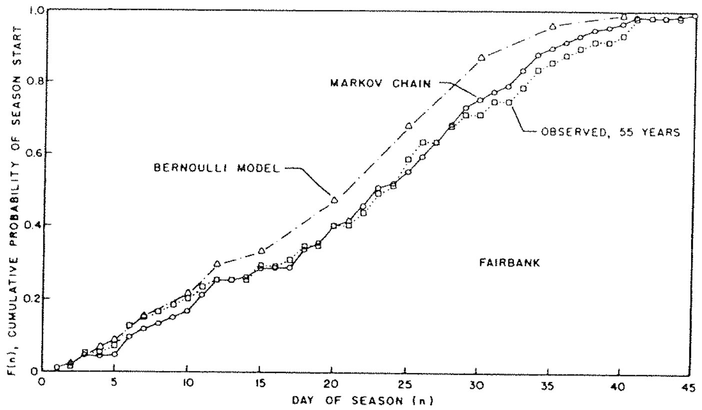

Fig. 7. Predicted and observed cumulative distribution of the first wet day of the season at Fairbank.

threshold minimum amount probably must be received to initiate cell differentiation and growth.

The probability of the first rainfall day is directly dependent on the model assumed for the underlying stochastic process. On the assumption of sequential independence the probability of starting day $s = n$, $n = 1, 2, \cdots, 122$, can be expressed as

$$p(s = n) = f_n \prod_{i=1}^{n-1} (1 - f_i) \tag{16}$$

The cumulative distribution for this random variable is

$$P(s \leq n) = \sum_{k=1}^{n} f_k \prod_{i=1}^{k-1} (1 - f_i) \tag{17}$$

One may easily show that the corresponding probability of the first wet day in the season from a Markov chain, given that day $i = 0$ is dry, is

$$p(s = n) = (1 - f_{d_n}) \prod_{i=1}^{n-1} f_{d_i} \tag{18}$$

and the cumulative distribution function is

$$P(s \leq n) = \sum_{k=1}^{n} (1 - f_{d_k}) \prod_{i=1}^{k-1} (1 - f_{d_i}) \tag{19}$$

Application of these functions using the series f_i and f_{d_i} for each of the three locations is illustrated in Figures 7–9. Prediction by the Markov chain is in good agreement with measured starts for most of the range of observed starting days on all stations. Some deviation occurs for all station predictions near the late-start range. By contrast, the independent model consistently overpredicts probabilities of starting day by several days.

Probability of number of rainfall days per season. As given by *Gabriel and Neumann* [1962] and *Todorovic and Woolhiser* [1973], a probability description of the number of wet days n in any period of length $i = m$ can be derived from the period-averaged Markov chain transition probabilities f_w, f_d, and a value for f at the start of the period (day 0). For each value of period length m, separate probabilities are calculated for the event that day 0 was dry ($e_0 = d$) and the event that day 0 was wet ($e_0 = w$).

Given a dry initial condition $e_0 = 0$, define [*Gabriel and Neumann*, 1962]

$$\begin{aligned} c_0 &= m + \tfrac{1}{2} - |2n - m - \tfrac{1}{2}| \qquad n > 0 \\ c_0 &= 0 \qquad n = 0 \end{aligned} \tag{20}$$

and further define

$$\begin{aligned} a &= \{\inf k;\, k \geq \tfrac{1}{2}(c - 1)\} \\ b &= \{\inf k;\, k \geq \tfrac{1}{2}c\} \end{aligned} \tag{21}$$

The probability of having n wet day in an m day sequence given a previous dry day is

$$p_0(n, m) = f_w{}^n f_d{}^{m-n} \sum_{c=1}^{c_0} \binom{n-1}{b-1} \binom{m-n}{a} \cdot \left(\frac{1 - f_w}{f_d}\right)^a \left(\frac{1 - f_d}{f_w}\right)^b \tag{22}$$

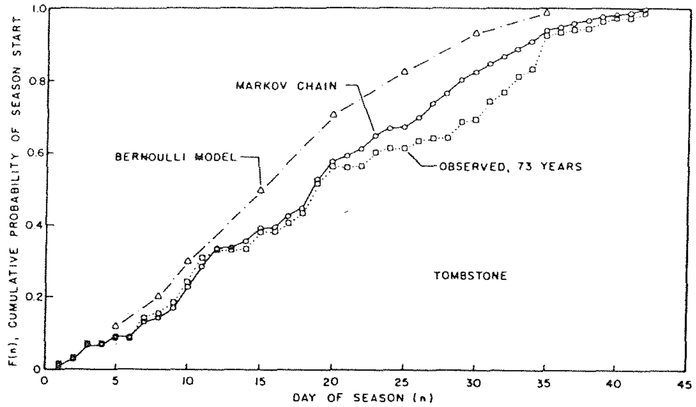

Fig. 8. Predicted and observed cumulative distribution of the first wet day of the season at Tombstone.

The other possibility, $e_0 = w$, requires a summation limit somewhat different from that given by (20):

$$c_1 = m + \tfrac{1}{2} - |2n - m + \tfrac{1}{2}| \qquad n < m$$
$$c_1 = 0 \qquad n = m \tag{23}$$

By use of a and b from (21) the probability of n wet days in a sequence of m days after an initial wet day ($e_0 = w$) is given as

$$p_1(n, m) = f_w^{\,n} f_d^{\,m-n} \sum_{c=1}^{c_1} \binom{n}{a}\binom{m-n-1}{b-1} \cdot \left(\frac{1-f_w}{f_d}\right)^b \left(\frac{1-f_d}{f_w}\right)^a \tag{24}$$

From (22) and (24) the unconditional probability of n wet days out of an m-day sequence is

$$p(n, m) = f_0[p_1(n, m)] + (1 - f_0)[p_0(n, m)] \tag{25}$$

The above probability functions require a period m in which f_w and f_d are relatively unchanging. Application of this relation was first attempted, however, by using a 122-day (seasonal) average for f_w and f_d at all three test stations to predict the distribution of the number of rainy day events per season, f_0 being taken as f_1. Figures 10–12 demonstrate the inadequacy of such averaging in a significantly nonhomogeneous period.

It is not difficult, however, to divide the season into small increments in which the f_w

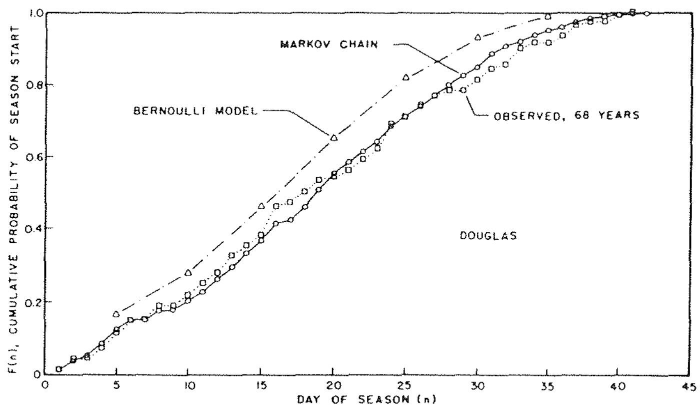

Fig. 9. Predicted and observed cumulative distribution of the first wet day of the season at Douglas.

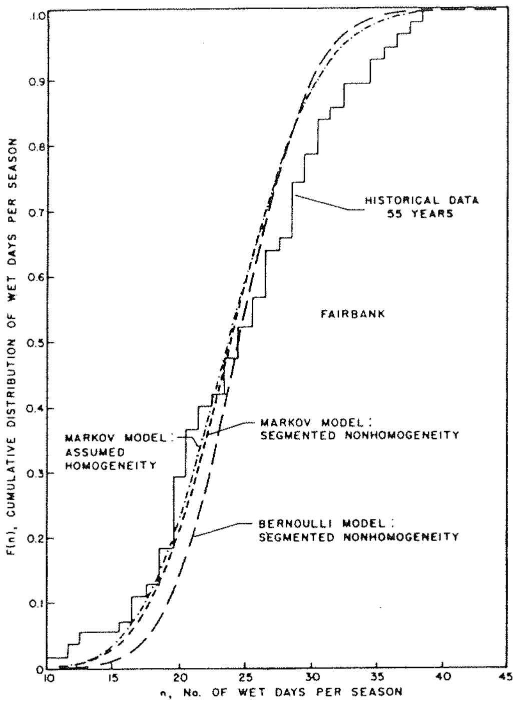

Fig. 10. Predicted and observed cumulative distribution of the number of wet days per season at Fairbank.

and f_d are much more uniform. Equation 25 may be applied to predict the probability of the number of events in each period. From this the probability of season total events may be derived by discrete convolution as follows.

Assume that the season is divided into N periods, each period j having a length $l = l(j)$ and a discrete probability density $p_j(n, l)$ for the number of events ($n = 0, 1, \cdots, l(j)$). The probability for the total number of events n in two periods a and b of lengths $l(a)$ and $l(b)$ involves convolution as follows [*Feller*, 1959]:

$$p_{a+b}[n, l(a) + l(b)] = p_a[0, l(a)]p_b[n, l(b)] + p_a[1, l(a)]p_b[n-1, l(b)] + p_a[2, l(a)]p_b[n-2, l(b)] + \cdots + p_a[n, l(a)]p_b[0, l(b)]$$

This operation may be indicated simply as

$$p_{a+b}[n, l(a+b)] = p_a[n, l(a)] * p_b[n, l(b)]$$

Thus the distribution for the season divided into N parts may be calculated as

$$p(n, m) = p_1[n, l(1)] * p_2[n, l(2)] * \cdots * p_N[n, l(N)] \qquad (26)$$

in which $m = \sum_{j=1}^{N}(j)$.

In this manner the theoretical distribution of the number of wet days per season was calculated for each location, the season being divided into 11 parts. Results are presented in Figures 10–12 along with the historic sample frequency and the distribution predicted by using (26), in which (25) is replaced by the appropriate expression for the Bernoulli model, which is a simple binomial distribution.

This comparison basically involves a test of annual variability, since the information used in (25) and (26) assumes an average year. Thus it is no surprise that the sample variance in Figures 10–12 is greater than that predicted by (26). The theory predicts distribution, implicitly assuming that all years have an equal pattern of f, f_d, and f_w. Obviously, the meteorological conditions vary from year to year, and some inadequacy in such a uniform stochastic model would be expected. An additional description for annual variance seems required.

One way to represent annual variance is to consider that each year the mean pattern of transition probabilities (Figures 3 and 4) is multiplied by a scale factor that is a random variable with a specified variance and zero mean value. A second method, more intuitively consistent with observed weather patterns, would be to specify

$$f_{w_i} = \bar{f}_{w_i}(1 + \epsilon a) \qquad 0 < f_{w_i} < 1$$

and

$$f_{d_i} = \bar{f}_{d_i}(1 + \epsilon a) \qquad 0 < f_{d_i} < 1$$

in which $\bar{f}_{w_i}$ and $\bar{f}_{d_i}$ are daily mean values obtained from Figures 3 and 4, a is a scale factor, and ϵ is a random variable with zero mean and high daily serial correlation.

It is significant to note that the Markov chain model is consistently better than the Bernoulli model in this derived distribution, as it is in other tests, although neither can properly predict annual variance in parameters of the stochastic process. The apparent greater degree of agreement between independent and dependent models in this distribution than in

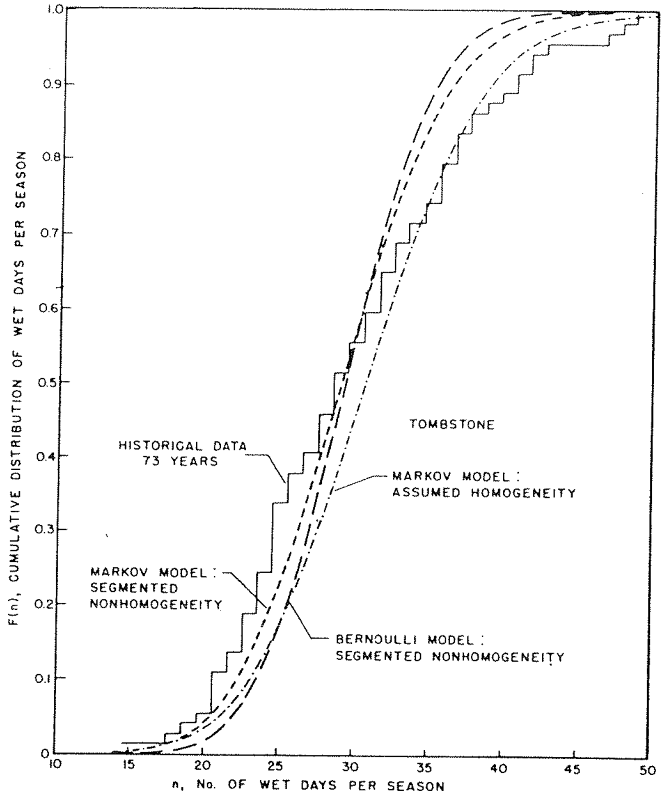

Fig. 11. Predicted and observed cumulative distribution of the number of wet days per season at Tombstone.

other distributions is expected in that the asymptotic law of large numbers applies. *Gabriel and Neumann* [1962] indicated that (22) and (24) are in fact asymptotic to a normal distribution. It is also interesting to note the geographical correspondence to the increasing number of rainy days per year, indicating that on the average the more southeastern gage receives rain more often.

Cyclicity and annual variations. A purely subjective hypothesis prompted a cursory investigation into the existence of short-term cyclicity within the seasonal stochastic process. In this area of the Southwest, summer thunderstorm occurrence is dependent on the input of tropical moist air from either the Gulf of Mexico or the Gulf of California. From study of Figures 3 and 4 it was questioned whether the movement of such moist air into the region involves wave phenomena with a dominant frequency or significant periodicities.

Lag autocorrelation coefficients were determined for all three stations for lags up to 12 days. The results demonstrated more than anything else the variation in weather patterns from year to year, already shown above in the analysis of rainy days per season. Some sample results are shown in Figure 13, indicating a great variation in sample autocorrelations from year to year. Also shown are the results using all seasons of data in a continuous trace. No significant dominant frequency was found.

Similar negative results were obtained in a harmonic analysis of the trace in Figure 4 for f_w and f_d. Tests of significance [*Hartley*, 1949] are dubious at best under nonhomogeneous conditions. Several harmonics (variable between 4 and 11 days) other than the very dominant season period were indicated to be significant in an F test on each gage. However, no one return period appeared statistically significant in all gages.

The degree of variation in rainfall day patterns from year to year is perhaps best demon-

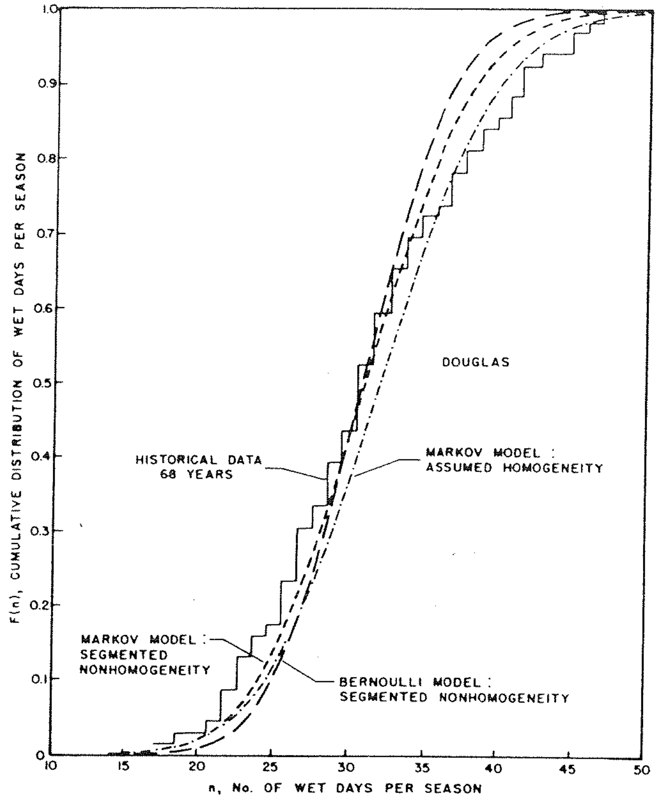

Fig. 12. Predicted and observed cumulative distribution of the number of wet days per season at Douglas.

strated by a presentation of the basic data. Figure 14 represents data from the Fairbank gage; 0 represents dry days, and 1 represents rainfall days.

Discussion and Conclusions

The several comparisons reported here appear conclusive in demonstrating the relative inadequacy of Bernoulli, or independent, event

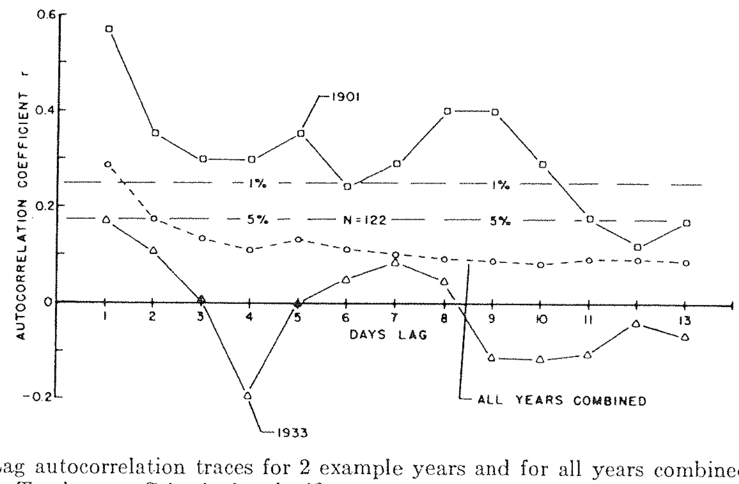

Fig. 13. Lag autocorrelation traces for 2 example years and for all years combined from the data at Tombstone. Criteria for significance at the 1% and 5% levels are also shown.

assumption for the stochastic description of rainfall in the thunderstorm season of the semiarid Southwest. The probability distributions for run lengths, drought lengths, starting day or first-season rainfall, and the number of rainy days per season have been derived from stochastic models based on a nonhomogeneous independent (Bernoulli) process and on a nonhomogeneous first-order dependent (Markov chain) process. In each case the dependent model is a far superior prediction of the sample distribution for three test stations with long-term records.

We do not wish to conclude that thunderstorm season daily rainfall occurs as a simple Markov chain. It has not been demonstrated, for example, that the order of dependency may not change within the season (as defined) or that a second- or higher-order chain may not be superior to a simple chain. Since lengths of successive runs are independent [*Gabriel and Neumann*, 1962], it is additionally possible that dry sequences occur more or less as an independent daily occurrence process but that wet sequences may best be described by a dependent stochastic process. Furthermore, it has been shown that the Markov chain model, to be accurate from year to year, requires an annual variance applied to the nonhomogeneous transition probabilities, whose sample mean is given in Figures 3 and 4.

It should be emphasized, however, that the simple Markov chain dependence model appears to be quite good for describing many if not most of the stochastic dependent properties of the southwestern thunderstorm rainfall process. This finding provides an extension of the positive results using a Markov chain obtained for several types of climatic locations across the United States by *Hershfield* [1970] as well as others, including *Feyerherm and Bark* [1965] in the Midwest, *Gabriel and Neumann* [1962]

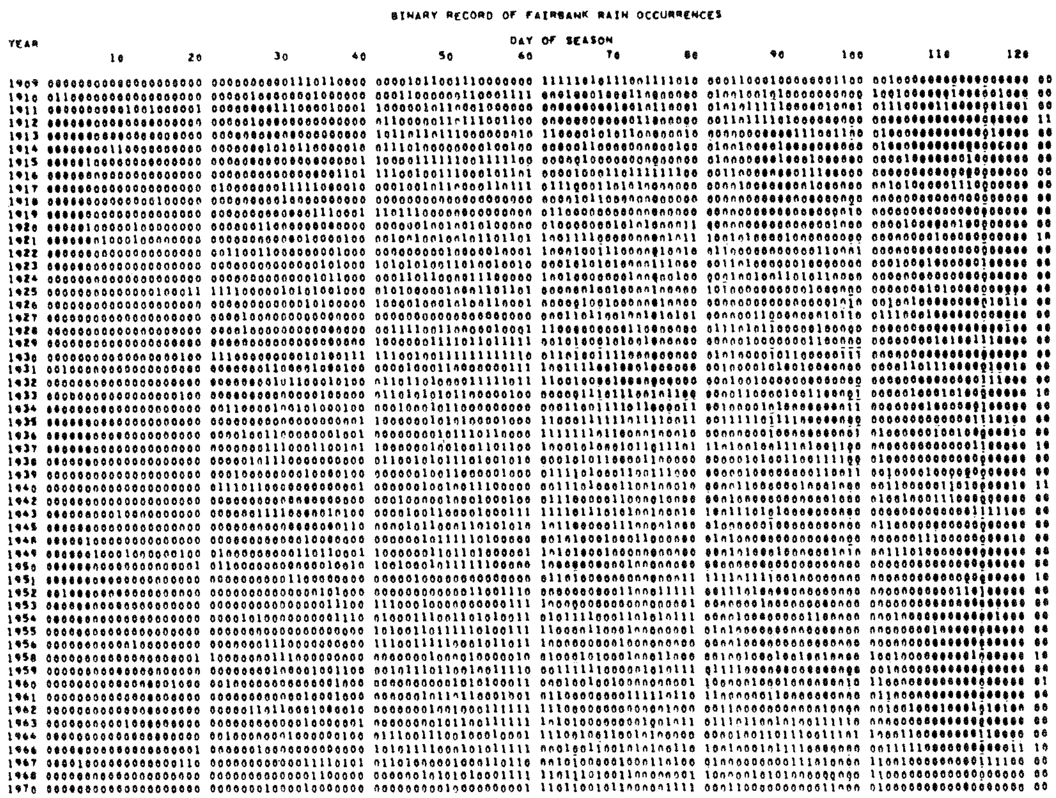

Fig. 14. Sample of historical data from Fairbank, Arizona; 1 indicates a wet day, and 0 represents a dry day.

for Tel Aviv, *Caskey* [1963] for Denver, and *Woolhiser et al.* [1972] for northeastern Colorado.

REFERENCES

Caskey, J. E., A Markov chain model for the probability of precipitation occurrence in intervals of various lengths, *Mon. Weather Rev., 91*(6), 298–301, 1963.

Duckstein, L., M. M. Fogel, and C. C. Kisiel, A stochastic model of runoff-producing rainfall for summer type storms, *Water Resour. Res., 8*(2), 410–421, 1972.

Feller, W., *An Introduction to Probability Theory and Its Applications,* vol. 1, pp. 268, 338–379, John Wiley, New York, 1959.

Feyerherm, A. M., and L. D. Bark, Statistical methods for persistent precipitation patterns, *J. Appl. Meteorol., 4,* 320–328, 1965.

Fogel, M. M., and L. Duckstein, Point rainfall frequencies in convective storms, *Water Resour. Res., 5*(6), 1229–1237, 1969.

Gabriel, K. R., and J. Neumann, A Markov chain model for daily rainfall occurrence at Tel Aviv, *Quart. J. Roy. Meteorol. Soc., 88,* 90–95, 1962.

Hartley, H. O., Tests of significance in harmonic analysis, *Biometrika, 36,* 194–201, 1949.

Hershfield, D. M., A comparison of conditional and unconditional probabilities for wet- and dry-day sequences, *J. Appl. Meteorol., 9,* 825–827, 1970.

Lane, L. J., and H. B. Osborn, Hypotheses on the seasonal distribution of thunderstorm rainfall in southeastern Arizona, paper presented at 2nd International Symposium in Hydrology, Colo. State Univ., Fort Collins, Sept. 11–16, 1972.

Schreiber, H. A., and N. G. Sutter, Available soil water: Time distribution in a warm-season rangeland, *J. Hydrol., 15,* 285–300, 1972.

Todorovic, P., and D. A. Woolhiser, Stochastic model of daily rainfall, *U.S. Dep. Agr. Misc. Publ.,* in press, 1973.

Weiss, L. L., Sequences of wet or dry days describe a Markov chain probability model, *Mon. Weather Rev., 92*(4), 167–176, 1964.

Woolhiser, D. A., E. Rovey, and P. Todorovic, Temporal and spatial variation of parameters for N-day precipitation, paper presented at 2nd International Symposium in Hydrology, Colo. State Univ., Fort Collins, Sept. 11–16, 1972.

Yevjevich, V., *Stochastic Processes in Hydrology,* pp. 174–211, Water Resources Publications, Fort Collins, 1972.

(Received February 21, 1973.)