A Simple Stochastic Model of the Precipitation Process

HANS ALEXANDERSSON

Department of Meteorology, University of Uppsala, Uppsala, Sweden

(Manuscript received 6 May 1984; in final form 30 January 1985)

ABSTRACT

A simple and rather general model of the precipitation process is reviewed and some applications and comparisons are made using data from Sweden. This model has been used by several authors so the article is partly a survey of earlier works but also adds some new aspects, comparisons and practical techniques. The model used is a compound Poisson–exponential (Cpe) process. This is a continuous stochastic process, which is compounded from a Poisson process with a number parameter (the mean number of independent precipitation events) and an exponential distribution with an amount parameter (the mean amount at each event). The meaning and the limitations of these physical interpretations are discussed briefly. The basic process can be applied on, e.g., integrated precipitation amounts, the Cpe distribution, and on maximum amounts, the max Cpe distribution. The first application has been discussed in many works, the latter was developed recently and independently by Revfeim (1983a) and Alexandersson (1983). One advantage of this model is that it does not need to be extended or modified to handle periods with zero precipitation. Another advantage is that the parameters can be estimated from the series of monthly precipitation totals. It is important that these techniques do not involve too lengthy calculations which would considerably hamper the practical use. Thus a very fast way of deriving percentiles from a single table for a Cpe distribution is developed here.

1. Introduction

The number of different ways in which a more or less complicated stochastic process describing the precipitation process at a point can be formulated is very large. The main efforts seem to have been to simulate daily sequences of wet and dry days or hours using Markov chains and the precipitation amount at wet days or hours using for example a shifted Gamma distribution. In Buishand (1977) a broad review is made of these techniques and recently, also, Woolhiser and Roldàn (1982) and Roldàn and Woolhiser (1982) made a review of existing techniques and made comparisons using data from USA.

Here we have chosen to use a compound Poisson-exponential process as a model which, certainly quite crudely, should describe the gross features of the true precipitation process. The two parameters of this model can be interpreted as a number parameter, the mean number of independent events within e.g., a month, and an amount parameter, the mean amount of precipitation at each event during the same time interval. The application of a compound Poisson model on integrated precipitation was, perhaps, first proposed by Feller (1968); the first edition was published in 1950. It has also been dealt with in works by Fisher and Cornish (1960), Bernier and Fandeux (1970), Buishand (1977), Öztürk (1981), LeCam (1961), Revfeim (1982, 1983a,b), Alexandersson (1983) and probably elsewhere. In the Revfeim and Alexandersson works it has also been applied to maximum amounts.

Here we will discuss the basic assumptions of this "natural" model and the two main applications to integrated precipitation and maximum daily amounts. The maximum likelihood estimates of the parameters will be derived and a comparison which clearly favors the Cpe distribution will be summarized. The approximate maximum likelihood estimates derived by Öztürk (1981) will be given using the notations adopted here. A table with Cpe percentiles is also included and it will be seen that this greatly simplifies practical use of this distribution. Also, a modified form of the Cpe process, which also needs a constant (mean) rain intensity, will be used.

2. Fundamental assumptions and relationships

The Poisson process, which will be assumed to describe the occurrences of precipitation events, can be summarized in the following way where P is "probability of"

- P [no event on $(t, t + \Delta t)$] $\approx 1 - \nu\Delta t$
- P [one event on $(t, t + \Delta t)$] $\approx \nu\Delta t$
- P [more than one event on $(t, t + \Delta t)$] ≈ 0

These probabilities are independent of past occurrences and these Poisson postulates describing the occurrences of precipitation events are, of course, a simplification of the true precipitation process. If, for instance, a series of showers within an unstable airmass is received at a station, it is rather the sum of these showers that can constitute an event as the whole system of showers is a rather intimately connected unit. Also a sequence of warm and cold frontal rainbands within one specific low pressure circulation in middle latitudes could constitute one independent event.

There is also some tendency of a "bit longer" memory of the atmosphere–ocean system in many parts of the world. The most striking and discussed long memory in middle latitudes is related to so called blockings and the resulting long sequences of dry days. This long-term memory, although generally quite weak, reminds us of the fact that the Poisson process must be an idealization. A more refined but similar model could take this into account by using an intensity ν that is not independent of past occurrences.

The next fundamental assumption that has been adopted is that at each event the amount has an exponential distribution, which is a one parameter distribution. This may seem to be a crude oversimplification as, e.g., the shifted Gamma distribution fits daily amounts better than an exponential distribution. Especially small amounts will be underestimated of an exponential model. This can be seen in Fig. 1 where data from Särna (61°41′N, 13°8′E) in central Sweden has been used. The main part of the frequency points is well described by an exponential curve. The Cpe process, however, does not deal with daily amounts and it can be assumed that several of the small daily amounts can be included in larger sums of nearby days and then forming one (roughly) independent event. Also one rain storm may frequently last through more than one consecutive 24 hour period. In fact the division of the continuum of time into consecutive days is not necessarily the best strategy when we are dealing with the basically continuous precipitation process.

This second fundamental assumption can also to some extent be justified with theoretical arguments.

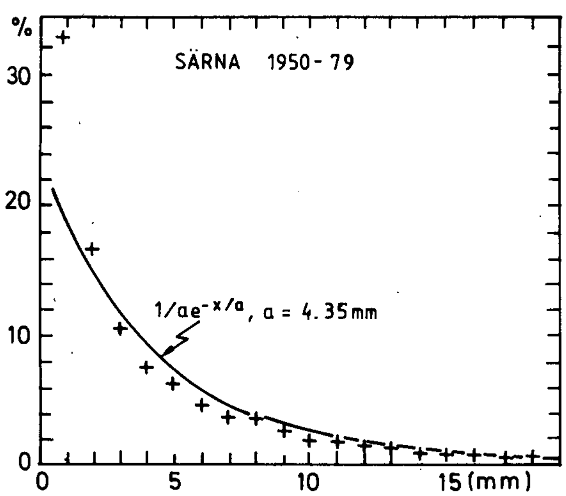

FIG. 1. Observed distribution for the daily amount of precipitation at Särna in central Sweden. Dry days (≤0.0 mm) are excluded. Also included is an exponential curve.

Suppose first that a rain event with a constant rain intensity r ends within a time interval τ with a probability proportional to the length of this interval and independent of the rain duration before this interval. Denote the proportionality constant by μ. From elementary statistical theory the integrated amount at this event will have an exponential distribution with mean value r/μ. Then, what is important for the integrated amount is the ratio r/μ so an event with large intensity r, like a heavy shower, will often also have a large μ and the ratio may well be about the same for heavy rain of short duration as for moderate but steady frontal rain.

Let us use a ($a = r/\mu$) for the mean amount at each event. For the integrated amount during a fixed time interval the number of occurrences of the Poisson process is given by a Poisson distribution with parameter $m = \nu t_0$. Thus we have for the integrated amount Y during a fixed time interval

$$Y = X_1 + \cdots + X_N, \qquad (1)$$

where, for N and X respectively, we have

$$p(n) = m^n e^{-m}/n!, \quad n \in (0, 1, 2, \cdots), \qquad (2)$$

and

$$f(x) = (1/a)e^{-x/a}, \quad x \geq 0. \qquad (3)$$

This is quite a unique two parameter distribution since there is a finite amount of probability of obtaining no precipitation at all during the time interval considered. Other widely used distributions, as the Gamma or the Weibull distributions, cannot immediately handle data where zeros occur. This can be circumvented by using censored samples but still it is very practicable to avoid further complications when we are dealing with climates where months sometimes occur with no precipitation at all.

The frequency function for the Cpe distribution is, unfortunately, not so simple but can be concentrated to the form

$$f(y) = \begin{cases} e^{-(y/a+m)} \sum_{n=1}^{\infty} (m/a)^n y^{n-1}/[n!(n-1)!], & y > 0, \\ e^{-m}, & y = 0. \end{cases} \qquad (4)$$

Alternatively a Bessel function can replace the sum (Öztürk, 1981) so that, at least for some values on y, the frequency function can be obtained from tables of the first order hyperbolic Bessel function I_1.

For practical applications we must of course have a method to estimate the parameters. Once these parameters are known we may use the asymptotically standard normal transform (Buishand, 1977, p. 17)

$$W = \sqrt{2/a}\,(\sqrt{Y} - \sqrt{ma}), \qquad (5)$$

which, however, is not appropriate for small m. So, as far as the Cpe distribution describes integrated precipitation with sufficient accuracy and m is not too near zero, the suggestion of transforming precipitation series to normal variables by taking square roots should be quite effective.

3. Moments and estimation procedures

We can use the moment-generating function to derive the moments of the Cpe distribution. Denoting this function by $\psi_Y(t)$ the general expression for a compound Poisson distribution is given by (Feller, 1968):

$$\psi_Y(t) = e^{m(\psi_X(t)-1)}. \quad (6)$$

Here X is assumed exponential so

$$\psi_X(t) = 1/(1 - at) \quad (7)$$

and then

$$\psi_Y(t) = e^{mat/(1-at)}. \quad (8)$$

The cumulants of the distribution can be obtained from successive differentiations at $t = 0$, from which it follows:

$$\left.\begin{aligned} E(Y) &= ma, && \text{mean value} \\ V(Y) &= 2ma^2, && \text{variance} \\ Sk(Y) &= 3(2m)^{-1/2}, && \text{skewness} \end{aligned}\right\} . \quad (9)$$

Thus the mean value of the integrated amount is obtained as the product of the mean number of events and the amount at each event. We can also note that the skewness is always positive and tends to zero as m approaches infinity. The coefficient of variation is, like the skewness, independent of a and given by $(2/m)^{1/2}$. In Revfeim (1982) an alternative derivation of the mean value and the variance was given.

Using the first two relations in (9) we can estimate the parameters m and a from the sample mean $\bar{y}$ and the sample variance s^2. This technique, the method of moments, gives

$$\left.\begin{aligned} m &= 2\bar{y}^2/s^2 \\ a &= s^2/(2\bar{y}) \end{aligned}\right\} . \quad (10)$$

However, as the Cpe distribution is significantly skewed for precipitation amounts for months and shorter periods it can be expected that the more satisfactory maximum likelihood method gives somewhat better estimates. Then we have to maximize the product of probabilities for the observed series:

$$L = \prod_{i=1}^{k} f(y_i). \quad (11)$$

Only a partial solution can be obtained when maximizing this product and that is (Alexandersson, 1983),

$$ma = \bar{y}. \quad (12)$$

We can obtain the complete solution by using this relation in Eq. (11) to get an equation with one independent variable. Then Newton–Raphson's formula can be applied to find the complete solution which was done by Buishand (1977) and Öztürk (1981). If we want to avoid the complicated derivative of Eq. (11) we can also proceed as follows. First three values of m are chosen: the solution of Eq. (10) and two adjacent values. Then Eq. (12) is used to find the corresponding values on the amount a. Then the product L, or rather the logarithm of L, is calculated for each set (m, a). A second order polynomial is then fitted to the points in the $(m, \ln L)$-plane and then maximized and a new pair $(m_x, \ln L_x)$ is obtained. The pair (m, a) with lowest $\ln L$-value in the original set is replaced by the new one and so on. This technique, which consumes much less computer time than the Newton–Raphson method, is illustrated in Fig. 2. In Öztürk (1981) a formula was derived from which approximate maximum likelihood estimates can be obtained even easier. With the notations used here his scheme becomes

$$\left.\begin{aligned} m &= \bar{y}/a \\ a &= 8A\sqrt{\bar{y}}/[n - k + [(n-k)^2 + 6AB]^{1/2}] \\ A &= n\sqrt{\bar{y}} - \sum_{i=1}^{n} \sqrt{y_i} \\ B &= \sum_{i=1}^{n} 1/\sqrt{y_i} \end{aligned}\right\} , \quad (13)$$

where the sum in B excludes zero values and where k is the number of zeros. These approximate estimates are very near the exact values.

The differences between the estimates for the two methods presented first (moments and maximum likelihood) are generally within 10%, the largest deviations occurring for the driest months and months with very low (sometimes zero) values in the observed series. In subsequent examples we use the maximum likelihood estimates.

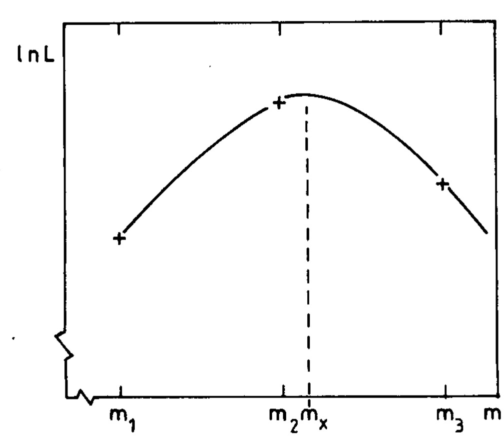

FIG. 2. A sketch of the iteration technique used to find the maximum likelihood estimates of the Cpe distribution.

4. Applications to monthly precipitation series at two stations

Here two stations are used representing quite different climates of the southern peninsula of Sweden (Götaland and parts of Svealand). One is Uppsala (59°52′N, 17°38′E) at the northeastern part of the peninsula; the other is Gångarebo (57°1′N, 12°55′E) in the southwestern area, Uppsala is situated on flat plains surrounding Lake Mälaren, an area which receives 500–550 mm yr^{-1} of precipitation on average. Gångarebo on the other hand is situated on the western slopes of the hilly inland area of Götaland and receives about 1100 mm yr^{-1} on average. Within this latter region the mean precipitation rises from about 700 mm at the coast to 1100–1150 mm at altitudes of 125–200 m and 15–20 km from the coast. This is by far the most pronounced local maximum in Sweden, outside the fell region. On a map in Fig. 3 these two stations are inserted.

In Fig. 4a, b the Cpe parameters are plotted as functions of month. Note that for Uppsala, 147 years of reliable measurements were available while for Gångarebo a series of 69 years can be obtained. From these plots we find that the number of events are almost the same and the difference in precipitation amounts is reflected mainly in the mean amount at each event. Even 147 years of measurements give discontinuities in these curves that could be explained partly by an insufficient sample size. In Revfeim (1982) the Poisson process rate ν and the exponential parameter a were fitted to single harmonic functions during the year. Thus the monthly values on m and a were smoothed to sinusoidal curves. This technique is based on the assumption that the solar cycle dominates the behavior of the precipitation process. In Sweden there is typically a maximum of convective precipitation in the summer months and a maximum of frontal rains in the autumn and winter months, and harmonic smoothing would probably too drastically simplify the true annual course.

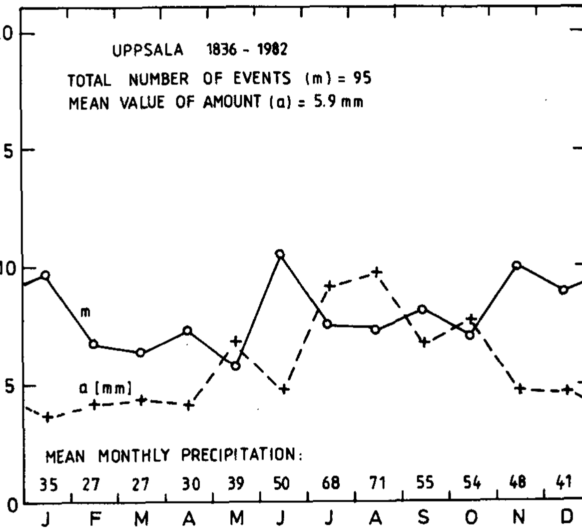

FIG. 4a. Annual variation of Cpe parameters, Uppsala (1836–1982).

We note that the total annual number of independent rain events are 95 and 89 for Uppsala and Gångarebo, respectively. This is, in a way, an estimate of the degrees of freedom for the precipitation climate; i.e., on average, the memory of the atmosphere concerning precipitation falls off to effectively zero after 365/90 ≈ 4 days. The number of rain events, 95 and 89, can be compared with the number of days with precipitation at the two stations which are for ⩾0.1 mm, 163 and 201, respectively, and for ⩾1.0 mm, 97 and 142, respectively (1962–79). Thus several of the just slightly wet days do not constitute independent rain events in the sense of the Cpe model.

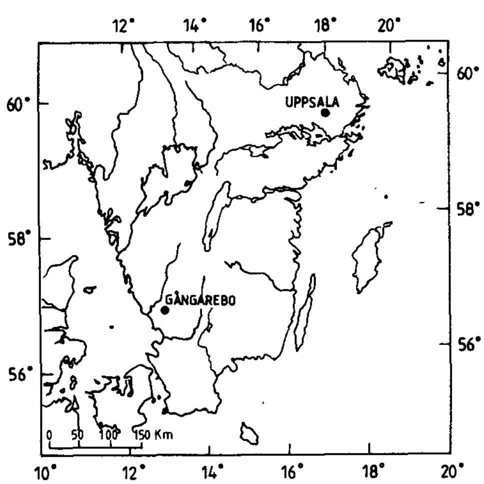

FIG. 3. A map of southern Sweden showing the position of Uppsala and Gångarebo.

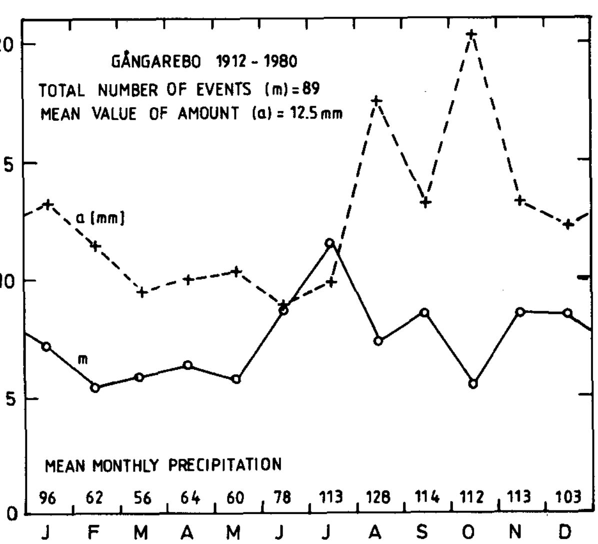

FIG. 4b. Annual variation of Cpe parameters, Gångarebo (1912–80).

To use the Cpe distribution in making one-point or spatial analyses of percentiles (e.g., the value exceeded by 5% of the population), probabilities, return periods etc., the parameters must, of course, be inserted in the expression for the frequency distribution. Alternatively, the approximation given by Eq. (5) could be used, or, better, the method given in section 5. Once this parent distribution is established one may also get distributions for extreme years, (p. 404 ff; Lindgren, 1968), e.g., for the largest integrated precipitation during an August month within 20 years. One example is shown in Fig. 5. August has been the wettest month in large parts of Sweden during the 20th century. From this plot one may read any probabilities or percentiles and classify into very dry, dry, normal, wet and very wet according to a chosen scheme of probability into each class.

The most crucial parts of the distributions are the ends representing unusually dry or wet conditions. Here a bias of the theoretical model chosen is very critical. In Table 1 theoretical and observed values of cumulative distributions are compared for $F(10)$, i.e., the probability of obtaining less than 10 mm. At Uppsala this probability reaches a maximum of about 10% in February and March but is as low as about 1% in the summer months. Particularly for Uppsala, with 147 observations, the agreement should be good between observations and the Cpe distributions of these more or less rare events.

On the average, no bias seem to be present here, an observation which is also supported by other investigations of the extreme tails of the Cpe distributions (Alexandersson, 1983). It should be noted that the observed values, $F^*(10)$, can take only certain values depending on the number of observations. So, for example, no observations below 10 mm have been observed at Gångarebo in June to August, while, quite naturally, the Cpe model predicts a small probability, i.e., it should occur but very rarely. On the other hand two October months have received less than 10 mm giving a $F^*(10)$ value of 0.029. The Cpe model, which is based on all observations, gives a smaller value. The Cpe model, therefore, slightly smooths the tails, which seems reasonable. In Buishand (1977) the Cpe distribution was successfully applied on data with a large fraction of zeros (Bangalore, India; the winter months).

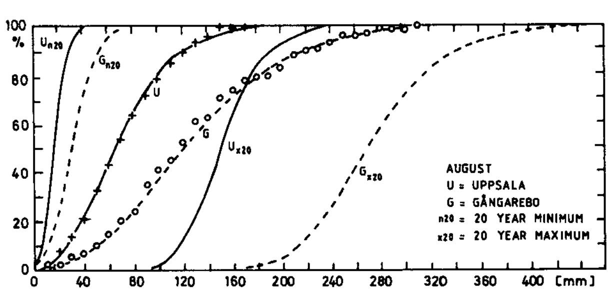

FIG. 5. Cumulative Cpe distributions and extreme value distributions at Uppsala and Gångarebo.

TABLE 1. Comparison between theoretical, Cpe[$F(10)$] and observed [$F^*(10)$] probabilities of obtaining less than 10.0 mm of precipitation during a month.

	Uppsala*		Gångarebo**	
Month	$F(10)$	$F^*(10)$	$F(10)$	$F^*(10)$
January	0.030	0.020	0.011	0.000
February	0.106	0.109	0.047	0.072
March	0.113	0.116	0.039	0.036
April	0.076	0.082	0.023	0.014
May	0.072	0.061	0.041	0.065
June	0.009	0.027	0.001	0.000
July	0.014	0.027	0.001	0.000
August	0.015	0.014	0.005	0.000
September	0.017	0.000	0.003	0.000
October	0.026	0.034	0.021	0.029
November	0.012	0.007	0.004	0.000
December	0.025	0.007	0.004	0.000
Mean value	0.043	0.040	0.017	0.018

* Period of record 1836–1982.
** Period of record 1912–80.

5. Practical applications of the Cpe distribution

Here we will try to make some more practical use of the Cpe distribution. When several stations within a limited region are analyzed, the number parameter often varies quite moderately. A data set for 57 stations within a region in southwest Sweden has been used here. The series were completed to a common period (1912–81) and apparently nonhomogeneous records were adjusted (Alexandersson, 1984). Here one month is selected: August. Estimating the Cpe parameters for these stations and averaging the number parameter gives $\bar{m} = 7.75$. The standard error of this mean, assuming 57 independent values, becomes 0.11; but this is not a true measure of the error in $\bar{m}$ because of dependence among the observed series. This regional smoothing of the number parameter simplifies the presentation of theoretical cumulative frequencies. As m is held fixed and $ma = \bar{y}$ we only need a map of the mean monthly precipitation (Fig. 6) and a nomogram with Cpe curves (Fig. 7) to present the distributions within this region. It should be mentioned that the mean values used to construct Fig. 6 are uncorrected and that a few stations with short series and questionably low mean values have been neglected. If we, for instance, are interested in the return periods for August months with more than 250 mm at two stations with mean amounts of 100 and 130 mm, these become 120 and 20 years respectively.

As the distribution of the precipitation totals involves fairly lengthy calculations a nomograph or a table of the cumulative probabilities would be valuable. We should first note a special characteristic of the Cpe distribution. We have

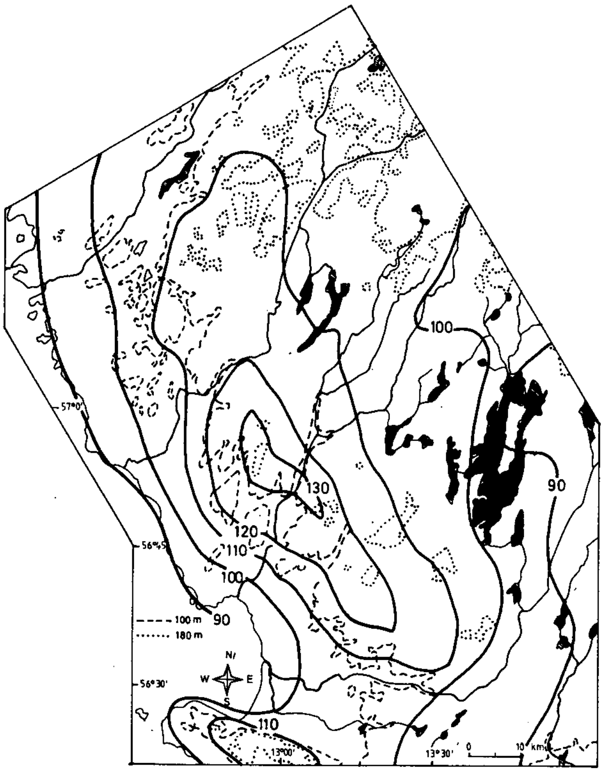

FIG. 6. Mean rainfall in August (1912–81) for a region in southwest Sweden.

$$F(y) = e^{-m} + \int_0^y (e^{-(x/a+m)} \sum_{n=1}^{\infty} \frac{(m/a)^n x^{n-1}}{n!(n-1)!} dx. \quad (14)$$

For an arbitrary value of n ($n = k$) the integral $[I_k(y)]$ becomes

$$I_k(y) = \frac{m^k e^{-m}}{k!(k-1)!} \int_0^y \frac{e^{-x/a} x^{k-1}}{a^k} dx, \quad (15)$$

or, after partial integration,

$$I_k(y) = \frac{e^{-m} m^k}{k!(k-1)!} [-(y/a)^{k-1} e^{-y/a} - (y/a)^{k-2} e^{-y/a} \ldots - e^{-y/a} + 1]. \quad (16)$$

So, for all k and also for $F(y)$, the parameter a only appears in the ratio y/a. Hence, if we double a, a doubling of y yields the same value of $F(y)$. Let us denote a specific value of $F(y)$ with α [%] and the corresponding value of y with y_α [mm]. Then we get

$$y_\alpha(a', m) = (a'/a) y_\alpha(a, m). \quad (17)$$

Thus, a single table with a set of m-values and some α-levels can be used to obtain the different percentiles y_α. In Table 2 the parameter a is fixed at 10 mm and the α-levels are 1, 2.5, 5, 10, 25, 50, 75, 90, 97.5 and 99%. Let us assume that we have estimated m as 9.60 and a as 3.64 mm (January, Uppsala). By interpolation in the column $\alpha = 5$ the 20-year return drought value becomes 33.5 mm for $a = 10$ mm. Then Eq. (17) gives 12.2 or rather 12 mm as an estimate of the 20-year return drought value in this case. Similarly the 20-year return wet value can be obtained from the column $\alpha = 95$ as 64 mm. Table 2 and Eq. (17) then offer a solution to the practical difficulties of the Cpe distribution. Also Öztürk's approximations, Eq. (13), make the estimation procedure easy.

If the actual m value is between a zero and a nonzero percentile, e.g., between four and five for $\alpha = 1\%$, the interpolation in m should be done from the upper m value to a m-zero value ($m_z = -\ln\alpha$). With $\alpha = 1\%$, m_z is 4.605 and then, if m equals 4.5, the 100-year return drought value is zero mm. But if m equals 4.9 and a equals 10 mm we get by interpolation 0.9 (4.9 − 4.605)/(5.0 − 4.605) ≈ 0.7 mm. Then any interpolations can be done from this Table when m does not exceed 20, a very large value for monthly data. The table is based on calculations of frequencies for integer values of mm.

6. Comparison with other distributions

Frequently Pearson's χ^2-statistic is used to compare the goodness of fit of distributions, but here we have used instead the likelihood function to compare three distributions, the Cpe, the gamma and the normal distribution, although the last one is defined below zero. All three are two-parameter distributions; thus it is fair to make a direct comparison of the value of the likelihood function when the maximum likelihood estimates are derived. Five Swedish stations were used and, as a few months with zero precipitation existed, the ordinary gamma distribution sometimes failed. In such cases the incomplete gamma distribution, including a third parameter which represents the probability at zero, can be used. This distribution has successfully been used to describe integrated precipitation amounts by e.g., Dingens and Steyeart (1971) and Gray (1976).

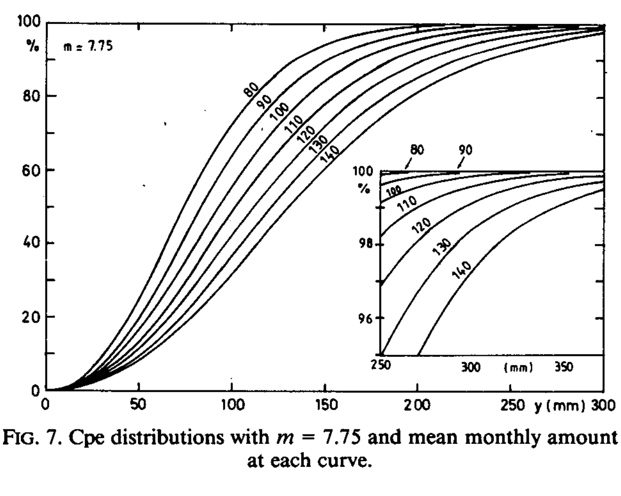

FIG. 7. Cpe distributions with m = 7.75 and mean monthly amount at each curve.

TABLE 2. Cpe-percentiles for m-values up to and including 20 and for a equal to 10 mm. With a in mm the unit of the percentiles is also mm.

	α (%)										
m	1	2.5	5	10	25	50	75	90	95	97.5	99
1	0.0	0.0	0.0	0.0	0.0	4.0	15.1	29.1	39.2	49.0	61.8
2	0.0	0.0	0.0	0.0	4.3	14.7	29.8	47.3	59.6	71.3	86.2
3	0.0	0.0	0.0	3.1	11.3	24.8	43.0	63.3	77.3	90.5	107.1
4	0.0	0.9	3.7	8.2	18.7	34.9	55.7	78.4	93.7	108.1	126.1
5	0.9	4.1	8.0	13.8	26.4	44.9	68.0	92.8	109.3	124.8	144.0
6	3.6	8.0	12.9	19.8	34.3	54.9	80.1	106.7	124.4	140.8	161.2
7	7.0	12.4	18.2	26.2	42.5	64.9	92.0	120.3	139.1	156.4	177.7
8	10.8	17.2	23.9	32.8	50.7	74.9	103.8	133.7	153.4	171.5	193.8
9	15.0	22.4	29.8	39.6	59.0	85.0	115.5	146.9	167.4	186.3	209.5
10	19.6	27.8	36.0	46.6	67.5	95.0	127.0	159.8	181.2	200.9	224.9
11	24.4	33.5	42.3	53.8	76.0	105.0	138.5	172.7	194.9	215.2	240.1
12	29.5	39.3	48.9	61.1	84.6	115.0	150.0	185.4	208.3	229.3	255.0
13	34.8	45.3	55.5	68.5	93.2	125.0	161.3	197.9	221.6	243.3	269.7
14	40.2	51.5	62.3	76.0	101.9	135.0	172.6	210.4	234.8	257.1	284.2
15	45.9	57.8	69.2	83.6	110.6	145.0	183.9	222.8	247.9	270.7	298.5
16	51.6	64.3	76.3	91.3	119.4	155.0	195.1	235.2	260.9	284.3	312.7
17	57.6	70.8	83.4	99.1	128.2	165.0	206.3	247.4	273.8	297.7	326.8
18	63.6	77.5	90.6	106.9	137.1	175.0	217.4	259.6	286.6	311.0	340.8
19	69.7	84.3	97.8	114.7	146.0	185.0	228.6	271.7	299.3	324.3	354.6
20	76.0	91.1	105.2	122.7	154.9	195.0	239.7	283.8	311.9	337.4	368.3

The maximum likelihood estimates of the gamma distribution have been derived by Thom (1958) and for the normal distribution they are simply equal to the estimates obtained from the method of moments.

Only in 5 out of 60 cases was the normal distribution best. Thus, we hereafter compare only the test results of the Cpe and the gamma distribution. Including the three months with zero values, the Cpe distribution was better in 41 out of 60 cases. This is highly significant according to a sign test. Excluding the three months with zero values, the Cpe distribution is better in 38 out of 57 cases which still is a significant difference. Thus even the incomplete gamma distribution cannot compete with the Cpe distribution although the latter one only has two parameters. An objection to this test could be that five stations within Sweden must be somewhat correlated. To avoid this the results can be averaged for each month. Then 11 out of 12 months (June being the exception) favor the Cpe distribution when compared with the gamma distribution. This still holds if months and stations with zero values, which occurred for one station in February and two in May, were excluded. The probability of obtaining 11 or more months favoring the Cpe distribution by mere chance is about 0.3% (one-sided test).

7. The distribution of maxima for the Cpe model

Let us make another application of the Cpe model. Instead of considering the sum as in Eq. (1) we may be interested in the largest of the individual rains which we can denote by Z:

$$Z = \max(X_1, \ldots, X_N). \tag{18}$$

Here, as before, X is assumed exponential with mean value a, and N is assumed of Poisson type with mean value m. To find the frequency distribution of Z one has to use the concept of conditional probability and the theory for maximum values. The result is a type of double exponential distribution and more precisely

$$f(z) = \begin{cases} (m/a)e^{-z/a}e^{-me^{-z/a}}, & z > 0 \\ e^{-m}, & z = 0, \end{cases} \tag{19}$$

and for the cumulative distribution

$$F(z) = e^{-me^{-z/a}}, \quad z \geq 0. \tag{20}$$

Henceforth we use max Cpe for this distribution in the text. Mean value and variance are approximately given by

$$\left.\begin{array}{l} E(Z) \approx a(\ln m + \gamma), \ \gamma = 0.5772 \text{ (Eulers constant)} \\ V(Z) \approx \pi^2 a^2/6 \end{array}\right\}. \tag{21}$$

The derivations of these formulas can be found in Alexandersson (1983) or in Revfeim (1983a) and the maximum likelihood estimates of the max Cpe distribution are given below. Equation (21) only holds when the probability of zero precipitation can be neglected. From Eq. (20) we can note that also here the parameter a only enters in z/a so Eq. (17) holds for both the Cpe and the max Cpe distribution. Here, however, Eq. (20) is very simple so no tables of percentiles need to be constructed.

One way of obtaining m and a, e.g., monthly, would be to use the values determined by application of the Cpe distribution to the series of integrated amounts.

However, in climatology one is more interested in 24 hour maxima than in maxima of amounts at certain events as defined by the Cpe model. These quantities are generally not identical. As a rule an independent event comprises more than 24 hours of precipitation (usually intermittent). This can be seen, for example, from the values of m, the number of events, which is clearly smaller than days with precipitation ($\geq$0.1 mm) (see section 4). This is also to be expected from the fact that rain may fall at both sides of the shift of the day, or rather at both sides of some special reading time.

To adapt the max Cpe distribution to describe 24 hour maxima we use observed series of these maxima and the maximum likelihood technique to find m and a. This technique is easier to apply on the max Cpe than on the Cpe distribution. In Revfeim (1983a) it was applied on maxima within 10 min to 72 h on a large data set from Kelburn, New Zealand.

We can also transform Z according to

$$Z' = \begin{cases} Z - c, & z > c \\ 0, & z \leq c, \end{cases} \tag{22}$$

where c is a lower "heavy rainfall" limit; i.e., if this value is not exceeded during a specific month or year we consider that month or year to be without any truly heavy rainfall. In Sweden, 10 mm could be an appropriate value to choose as a lower limit, especially if winter months are studied, because, then, 10 mm in 24 hours is rather high for lowland areas in Scandinavia. One effect of this transformation is that the number m of the max Cpe model now will be the number of really heavy rainfalls above c mm in 24 hours. The variable Z' may also be better than Z to describe very rare heavy rainfalls using a relatively high value for c. Then very large amounts (outliers) will be more important for the distribution. This technique is possible to use if we have very long series and may be better than using the variable Z if the very heaviest rains fall somewhat aside of the max Cpe curve.

Given a set of observations $\{z_i'\}$, which equals $\{z_i\}$ if $c = 0$, the maximum likelihood solutions can (after some calculations) be found from:

$$\left.\begin{aligned} & m = (n - k)/(k + \textstyle\sum e^{-z_i'/a}) \\ & n - k - (1/a) \textstyle\sum z_i'(k + \sum e^{-z_i'/a}) \\ & \qquad + ((n - k)/a) \textstyle\sum z_i' e^{-z_i'/a} = 0 \end{aligned}\right\} , \tag{23}$$

where k is the number of zeros of the total number n of (z_i':s) and all sums are evaluated for nonzero observations of z_i'. The second relation in Eq. (23) is an equation of one unknown a and can quite easily be solved by using Newton–Raphson's formula. Then m is found by inserting a in the first relation.

We also note that the maximum of (z_i':s) during h years is given by

$$F_X(z') = e^{-mhe^{-z'/a}}, \tag{24}$$

so it is still a max Cpe distribution although m in the parent distribution is replaced by mh.

In Fig. 8, two months, May and August, are analyzed and the corresponding observed cumulative distributions are also marked (Uppsala, 1866–1982). Also included is the number of days with more than c = 10 mm, a Poisson variable according to the Cpe model. In Fig. 9 the 24 hour maximum of the whole year is plotted together with 20 year minima and maxima for Uppsala. Comparing August with the whole year we find that they give almost the same probabilities above 50 mm while at 10 mm, the heavy rainfall limit chosen here, the cumulative distribution for August is 0.16 but 0.00 for the year. This is not surprising since the individual month August contributes to many of the largest annual values (11 out of 19 cases with >40.0 mm during 1866–1982) without contributing to the lowest annual values.

8. Simulations of hourly and daily values

For time intervals shorter than a month but still above about 15 days we can make use of the Cpe model thanks to its natural structure. Denoting the new period length by n' and the length of the month by n we can use the new primed estimates

$$\left.\begin{aligned} m' &= mn'/n \\ a' &= a \end{aligned}\right\} . \tag{25}$$

Thus we can immediately establish distributions for certain other length intervals. For longer periods, such as two months, with separate mean amounts, we can make a convolution of two Cpe distributions. But for really short periods the simplification using *instantaneous* amounts becomes too unrealistic. A rather straightforward way to generalize the Cpe model then

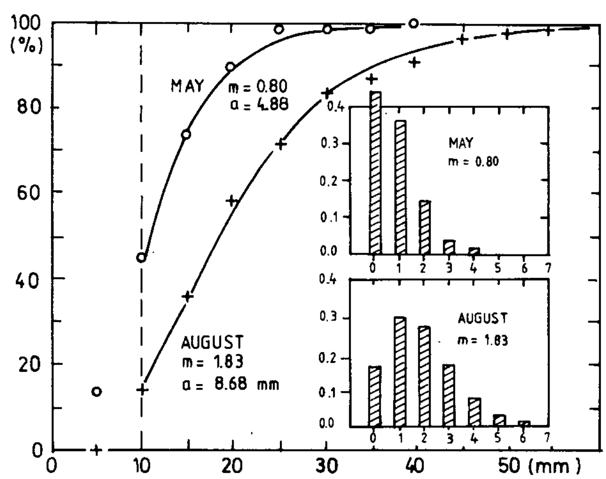

FIG. 8. Theoretical (max Cpe, c = 10 mm) and observed distribution for the 24 hour maximum precipitation in May and August, Uppsala (1866–1982). The distribution for the number of days with at least 10 mm is also inserted.

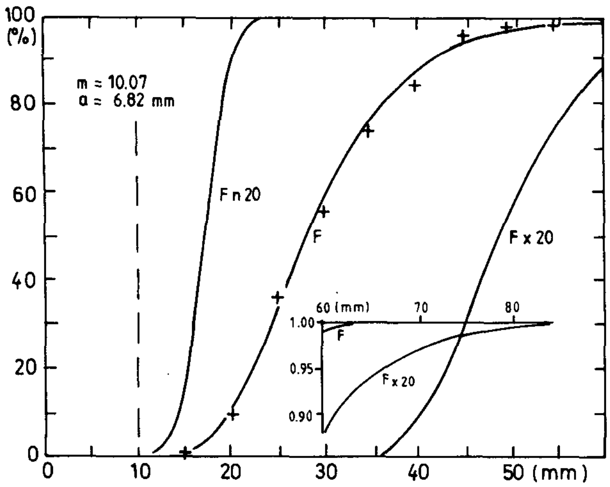

FIG. 9. The 24 hour maximum of the whole year (max Cpe, c = 10 mm) for Uppsala (1866–1982).

is to introduce a constant rain intensity r and the parameter μ as already discussed in section 2.

We may now use this modified Cpe model to see if the time structure, or more precisely the autocorrelation of series of daily precipitation, will be comparable to real data. Then the observed autocorrelations would merely result from the fact that a rain event lasts long enough to make nearby days correlated.

In Fig. 10 the Cpe model used heretofore and this slightly extended version are compared using two congruent realizations in order to clarify the difference.

The modified Cpe process used here is equivalent to a "birth-death" process with intensity matrix **A** defined as

$$A = \begin{bmatrix} -\nu & \nu & 0 & 0 & \cdots \\ \mu & -(\nu+\mu) & \nu & 0 & \cdots \\ 0 & 2\mu & -(\nu+2\mu) & \nu & \cdots \\ 0 & 0 & 3\mu & -(\nu+3\mu) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}. \quad (26)$$

We have, therefore, allowed two events (or more) to go on simultaneously and consequently the intensity of "death" is doubled. The average probability π_i of the process of being in state i (i events going on simultaneously) then becomes

$$\pi_i = (\nu/\mu)^i e^{-\nu/\mu} i!. \quad (27)$$

For this process we obtain the average state (Alexandersson, 1983)

$$E[x(t)] = \frac{r\nu}{\mu}, \quad (28)$$

and then

$$E(Y) = E\left(\int_0^{t_0} x(t)dt\right) = \nu t_0 \cdot r/\mu = ma, \quad (29)$$

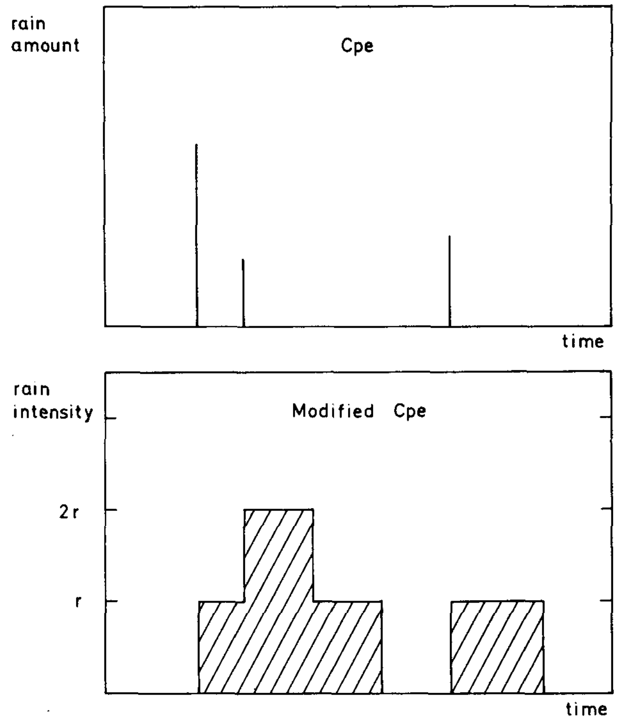

FIG. 10. Congruent distributions of the Cpe and the modified Cpe model.

so this modified Cpe process gives the same mean value for integrated amounts as the Cpe distribution.

The average time of precipitation (Uppsala) for winter months is about 21% and for summer months 11%. Thus, for Uppsala, average precipitation intensities of 0.22 and 0.83 mm h^{-1} are obtained for January and July respectively. For Gångarebo, and using 21 and 11%, the same intensities will be 0.61 and 1.39 mm h^{-1}. These rather low values arise from the quite large number of observations when just light precipitation falls, often not detectable by recording gauges.

Now we are ready to simulate the continuous modified Cpe process, or in practice, hourly values, from the m and a estimates using only monthly series and the constant rain intensity value. Hourly values were simulated for a number of Swedish stations and then mean values of autocorrelations of daily values were obtained for the two months. An observed series of daily values from Särna (61°41′N, 13°8′E) is published (annual, SMHI) and was used in the comparison. The theoretically derived autocorrelations are listed in Table 3 while Fig. 11 depicts the observed correlations of the

TABLE 3. Daily autocorrelations (ρ_i, i = time lag in days) as mean values for several stations in Sweden and simulated years using the modified Cpe model.

Month	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5
January	0.50	0.16	0.06	0.02	0.01
July	0.23	0.02	0.01	0.00	0.00

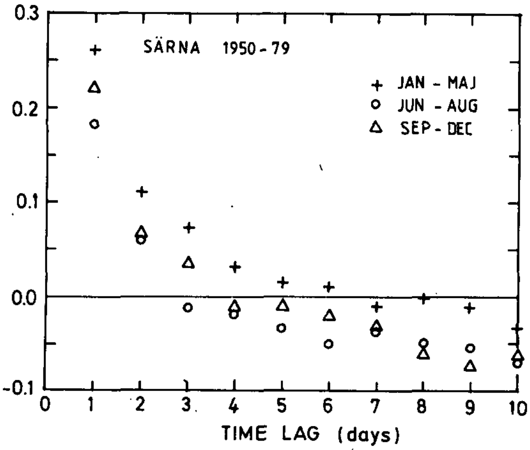

FIG. 11. Observed autocorrelations of daily precipitation, Särna (1951–80).

Särna series. In the simulations, 250 years were constructed which reduced the uncertainty of the autocorrelations to a minimum. For the much shorter observed series the months were grouped into three classes to get more stable values. Nevertheless we can see that the slightly modified or extended Cpe model gives reasonable time series of daily precipitation as far as autocorrelations and the gross time structure are concerned. The largest deviation occurs for ρ_1 in January compared with period 1 in Fig. 11. The large value according to the modified Cpe process is mainly a result of the low average intensity. During the winter months a lot of the precipitation in Sweden is in forms of snow grains, drizzle or very light snowfall. These types of precipitation do not contribute much to the integrated daily amounts.

It may seem somewhat surprising that this modified Cpe model gives reasonable autocorrelations between daily amounts but this is, of course, a result from the fact that the events now have a finite duration which may overlap more than one 24-hour period. This type of correlation, which results from reducing the continuous precipitation process into sequences of daily amounts, can therefore be called passive (Fitzpatrick and Krishnan, 1966).

9. Summary

The flexibility of the compound Poisson–exponential model makes it well suited for several applications in precipitation climatology. The basic assumptions are simple enough to get an analytical expression for integrated precipitation amounts, which may be quite important since the main amount of climatological data are published, and dealt with, in the form of monthly values. It then becomes possible to use the maximum likelihood technique to estimate the two parameters of the model, the mean number of precipitation events and the mean amount at each event. The Cpe model is also easy to use for the derivation of an extreme value distribution for the maximum amount at an individual event. This distribution is then used for the more concrete 24 hour maximum of a year or a month for which observed series often are available.

The Cpe model, which is a continuous stochastic process, links together models simulating time series of precipitation, hitherto as a rule using Markov chains applied to sequences of dry and wet days, and empirical distributions used for integrated amounts or extreme daily amounts. The model is based on a Poisson process. The assumption of independent events in this process is clearly an idealization. On the other hand the mean number of *independent* precipitation events is a sensible parameter related to the degrees of freedom of the true precipitation process at a station.

The basic assumptions of the simple Cpe model can be modified to allow for a more realistic description of the precipitation process without losing the natural parameters involved. In section 8 this was done by using a constant rain intensity. The modified Cpe process thus obtained gives a realistic time structure of the true precipitation process as far as daily autocorrelations are concerned. This continuous model can be extended in various other ways. So, for instance, we can introduce a memory (a decreasing autocorrelation) making nearby events correlated and no longer independent. We can also separate the precipitation process into two (or more) parts, one describing the convective precipitation, the other one the frontal or less intense precipitation. Such refinements will, however, give quite intractable equations and difficulties in estimating the parameters and for practical purposes they are perhaps less interesting.

This article is, in many parts, a review of published works about the Cpe model, but it is hoped that it will contribute to some spreading of the ideas and benefits of this simple model of the true precipitation process at a point. It is also hoped that the simple mathematical treatment and the applications given here will increase the practical use of the Cpe model.

Acknowledgments. The author wishes to thank Marianne Björklund and Aina Ekström for manuscript preparation.

REFERENCES

Alexandersson, H., 1983: A stochastic model of precipitation with applications, extensions and examples. Rep. 69, Dept. of Meteorology, University of Uppsala, 102 pp.

——, 1984: A homogeneity test based on ratios and applied to precipitation series. Rep. 79, Dept. of Meteorology, University of Uppsala, 55 pp.

Bernier, J., and D. Fandeux, 1970: Théorie du renouvellement. Application à l'étude statistique des précipitations mensuelles. *Revue de Statistique Appliquée,* **XVIII,** 75–87.

Buishand, T. A., 1977: *Stochastic modelling of daily rainfall sequences.* Mededelingen Landbouwhogeschool, Wageningen, 211 pp.

Dingens, P., and H. Steyeart, 1971: Distribution for K-day rainfall totals. *Bull. I.A.S.H. XCI,* **39,** 19–24.

Feller, W., 1968: *An Introduction to Probability Theory and its Applications, Vol I.* 3d ed., Wiley & Sons, 509 pp.

Fisher, R. A., and E. A. Cornish, 1960: The percentile points of distributions having known cumulants. *Technometrics,* **2,** 209–225.

Fitzpatrick, E. A., and A. Krishnan, 1966: A first-order Markov model for assessing rainfall discontinuity in central Australia. *Arch. Met. Geoph. Bioklima,* **B15,** 242–259.

Gray, B., 1976: Medium term fluctuations of rainfall in southeastern England. *Quart. J. Roy Meteor. Soc.,* **102,** 627–638.

LeCam, L., 1961: A stochastic description of precipitation. *Proc. 4th Berkeley Symp. on Mathematical Statistics and Probability,* J. Neyman, Ed., Univ. of California Press.

Lindgren, B. W., 1968: *Statistical theory,* 2d ed. Macmillan, 521 pp.

Öztürk, A., 1981: On the study of a probability distribution for precipitation totals. *J. Appl. Meteor.,* **20,** 1499–1505.

Revfeim, K. J. A., 1982: Comments "On the study of a probability distribution for rainfall totals." *J. Appl. Meteor.,* **21,** 1942–1945.

——, 1983a: On the analysis of extreme rainfalls. *J. Hydrol.,* **62,** 107–117.

——, 1983b: Stochastic process analysis of rainfall totals and extremes. *Proc. Second Int. Meeting on Statistical Climatology,* Inst. Nac. de Meteor. e Geofisica, 10.2.1–10.2.5.

Roldàn, J., and D. A. Woolhiser, 1982: Stochastic daily precipitation models, Part I, Comparison of occurrence processes. *Water Resour. Res.,* **18,** 1451–1459.

Thom, H. C. S., 1958: A note on the gamma distribution. *Mon. Wea. Rev.,* **86,** 117–122.

Woolhiser, D. A., and J. Roldàn, 1982: Stochastic daily precipitation models, Part 2, Comparison of distribution of amounts. *Water Resour. Res.,* **18,** 1461–1468.