WATER RESOURCES RESEARCH, VOL. 22, NO. 4, PAGES 531–542, APRIL 1986

Continuous-Time Versus Discrete-Time Point Process Models for Rainfall Occurrence Series

EFI FOUFOULA-GEORGIOU[1] AND DENNIS P. LETTENMAIER

Department of Civil Engineering, University of Washington, Seattle

Several authors have had apparent success in applying continuous-time point process models to rainfall occurrence sequences. In this paper, it is shown that if rainfall occurrences are interpreted as the events of a point process (and not as a censored sample), the continuous-time point process methodology and estimation procedures are not directly applicable since they fail to account for the time discreteness of the sample process. This is demonstrated analytically by studying the effects of discretization on selected statistical properties of a Poisson process, a Neyman-Scott process, and a renewal Cox process with Markovian intensity. In general, the study of rainfall occurrences under the continuous-time point process framework may result in misleading inferences regarding clustering (dispersion), and consequently incorrect interpretations of the underlying rainfall generating mechanisms. For example, daily rainfall occurrence structures underdispersed relative to the Poisson process are usually overdispersed relative to the Bernoulli process (the discrete-time analogue of the Poisson). These findings are confirmed by the statistical analysis of six daily rainfall records representative of a range of U.S. climates, two of which are described in detail.

1. Introduction and Problem Statement

The stochastic structure of daily rainfall occurrences has been extensively studied over the past two decades. The models suggested have evolved from the alternating renewal models [*Green*, 1964; *Grace and Eagleson*, 1966], to Poisson models [*Todorovic and Yevjevich*, 1969; *Duckstein et al.*, 1972], Markov chains [*Gabriel and Neumann*, 1962; *Todorovic and Woolhiser*, 1974; *Smith and Schreiber*, 1973], discrete autoregressive moving average (DARMA) models [*Chang et al.*, 1984], and finally, to point process models [*Kavvas and Delleur*, 1981; *Smith and Karr*, 1983]. This paper concentrates only on the point process modeling approach. Other rainfall occurrence models have been reviewed elsewhere [*Waymire and Gupta*, 1981*a*; *Roldan and Woolhiser*, 1982]. For the general theory of point processes the reader is referred to *Cox and Lewis* [1978], *Çinlar* [1975], *Lawrance* [1972], and *Daley and Vere-Jones* [1972]. *Waymire and Gupta* [1981*b*, *c*] have presented a careful review of the theory of point processes and have illustrated their applicability to modeling rainfall and rainfall-driven hydrologic processes.

Rainfall is a continuous intermittent process, whose intensity we denote as $\xi(t)$. Rainfall measurements represent cumulative amounts over discrete time intervals such as minutes, hours, or days. Let $\{Y_k(\Delta)\}$, $k = 1, 2, 3, \ldots$ denote the discrete sequence of rainfall observations over an arbitrary time interval Δ. The continuous process $\xi(t)$ is related to the discrete process $\{Y_k(\Delta)\}$ by

$$Y_k(\Delta) = \int_{t_{k-1}}^{t_k} \xi(\tau)\, d\tau \qquad (1)$$

where $t_k - t_{k-1} = \Delta$ is the time scale of measurement. Figure 1 illustrates this point: the continuous process $\xi(t)$ is integrated over, say, daily time intervals to give the sequence of daily data $\{Y_k(\Delta)\}, \Delta = 1$ day.

In modeling daily rainfall occurrences as a point process, i.e., a stochastic process which is completely characterized by the position of its events, two interpretations of the sampled data are possible: (1) the occurrences represent all of the events of the point process and (2) the occurrences represent a filtered sample of an underlying point process in which multiple occurrences during a day are possible, but only one is recorded when one or more occur. If the first interpretation is implemented, the event takes the meaning of a rainy day (see Figure 1) and one has to deal with a discrete-time point process, i.e., a point process in which events can occur only at time marks integer multiples of the sampling interval. All previous studies on point process modeling of daily rainfall (except the recent work of *Rodriguez-Iturbe et al.* [1984]) have implemented the first interpretation. It will be shown in this paper that this approach fails to account for the time discreteness of the process.

A related issue that can present serious difficulties is model fitting. Under the first interpretation of rainfall occurrences, model parameters are estimated by fitting procedures based on direct comparison of the empirical properties of the discrete-time data with their theoretical continuous-time counterparts. This approach, which has been used in several previous studies [e.g., *Kavvas and Delleur*, 1981; *Smith and Karr*, 1983; *Ramirez-Rodriguez and Bras*, 1982], will be shown to introduce severe estimation biases. This will be demonstrated by studying the effects of discretization on selected statistical properties of three commonly used point process models for daily rainfall occurrences: a Poisson process, a Neyman-Scott process, and a renewal Cox process with Markovian intensity.

2. Statistical Background and Terminology

For the discussion which follows, it is necessary to introduce a few functions which describe the statistical properties of a point process. More details on these functions can be found in statistical texts such as *Cox and Lewis* [1978] or in the work by *Kavvas and Delleur* [1981].

Let m denote the rate of occurrence of a continuous-time stationary point process, and $F(x)$ denote the cumulative distribution of the interarrival times. The log-survivor function, ln $[R(x)]$, is defined as the logarithm of the probability of exceedance $R(x) = 1 - F(x)$. For a Poisson process ln $[R(x)] = -mx$. A concave log-survivor function indicates

[1]Now at St. Anthony Falls Hydraulic Laboratory, Department of Civil and Mineral Engineering, University of Minnesota, Minneapolis.

Paper number 5W4097.
0043-1397/86/005W-4097$05.00

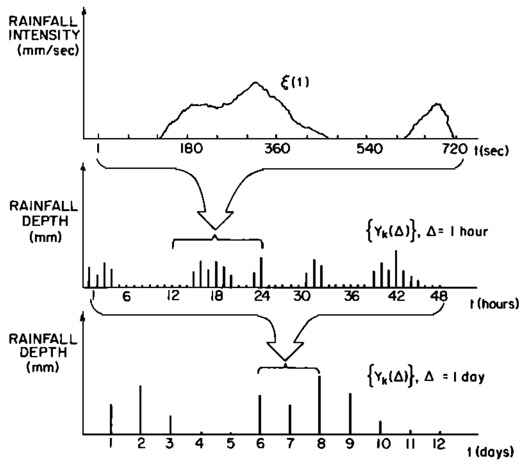

Fig. 1. Continuous rainfall process $\xi(t)$ and discrete hourly and daily rainfall sequences $\{Y_k(\Delta)\}$.

overdispersion (i.e., a tendency for clustering of rainfall events) relative to Poisson, whereas a convex log-survivor function is indicative of a process underdispersed relative to the Poisson.

The counting process, $\{N_t\}$, of a point process is defined as the number of events in $(0, t]$. The variance of N_t is a continuous function called the variance-time curve, $V(t) = \text{Var}(N_t)$. When divided by the mean number of events in $(0, t]$, $M(t)$, a function called the index of dispersion, $I(t) = V(t)/M(t)$ results. For a Poisson process, $M(t) = V(t) = mt$, and therefore $I(t) = 1$, $\forall\ t$. An index of dispersion $I(t) > 1$ (<1) indicates overdispersion (underdispersion) relative to the Poisson process. This property is analogous to the coefficient of variation c_v for the interarrival times, where similarly, $c_v > 1$ (<1) indicates overdispersion (underdispersion) relative to the Poisson process for which $c_v = 1$. The spectrum of the counting process $g_+(\omega)$, is defined as the Fourier transform of the covariance density of the differential counting process $\{\Delta N_t\}$, where ΔN_t is defined as the number of events in $(t, t + \Delta t]$; i.e., $N_{t+\Delta t} - N_t$. For a Poisson process $g_+(\omega) = m/\pi$, and the normalized spectrum of counts, defined as $g_+'(\omega) = \pi g_+(\omega)/m$, takes on the constant value of 1. Another important parameter of a continuous-time stationary point process is the conditional intensity function, $h(t)$, defined as

$$h(t) = \frac{P(dN_t = 1 \mid N\{0\} = 1)}{dt} \tag{2}$$

[*Cox and Lewis*, 1978, p. 73], where dN_t is the limit of ΔN_t as $\Delta t \to 0$, and $N\{0\}$ denotes "event at time 0." The interpretation of $h(t)dt$ is the probability of an event at time t, given an event at time 0. For a Poisson process $h(t) = m$, $\forall\ t$. Inferences about clustering of a point process can be made from the way $h(t)$ approaches the constant intensity m of the process. For more information on all these functions see, for example, *Cox and Lewis* [1978, chapter 4]. For the spectral analysis of a point process consult *Bartlett* [1963].

A discrete-time point process is a process in which events can only occur at time values $k = 1, 2, 3, \ldots$. Such an occurrence process can be viewed equivalently as a sequence of binary random variables $\{Z_k\}$, where Z_k takes on the values 1 and 0 depending upon whether an event did or did not occur at time k [*Lewis*, 1970]. Let m' denote the probability of occurrence of an event at an arbitrary time value k. All the statistical functions discussed previously can be extended in a straightforward way to the corresponding functions of a discrete-time point process. For example, $F(x)$ will be the cumulative distribution of the discrete probability mass function (pmf) of the interarrival times; N_k will be the number of events occurring within k time units; V_k will be the variance of the counting process $\{N_k\}$, etc. Similarly, the discrete-time analogue of the conditional intensity function is a sequence $\{h_k\}$ of conditional probabilities of occurrence, where h_k is defined as

$$h_k = P(Z_k = 1 \mid Z_0 = 1) \tag{3}$$

Note that h_k are probabilities, whereas $h(t)$ is a probability density. One major difference between continuous-time and discrete-time function definitions is the spectrum of counts, since the differential counting process $\{\Delta N_t\}$ is not defined for a discrete-time point process. The spectrum of counts is in that case defined as the Fourier transform of the autocovariance sequence $\{c_k\}$ of the binary time series $\{Z_k\}$. Helpful remarks on the spectral analysis of continuous-time versus discrete-time point processes can be found in *Lewis* [1970]. The reader is referred to *Guttorp* [1985] for more rigorous definitions of the statistical properties of a discrete-time point process.

3. Review of Continuous-Time Point Process Models

A Poisson cluster process, which has become known as the Neyman-Scott (N-S) process, was developed by *Neyman* [1939] for entomology and bacteriology population growth modeling. Subsequently, it was used by *Neyman and Scott* [1958] to model the spatial distribution of galaxies, and later by *LeCam* [1961] to model the areal distribution of rainfall. Based on the work of LeCam, *Kavvas and Delleur* [1981] applied a Neyman-Scott model to the rainfall occurrences on the time continuum, and found that such a process appeared to describe the clustering of daily rainfall occurrences in Indiana.

A N-S process is a two-level process. At the primary level, the rainfall generating mechanisms (RGM) occur according to a Poisson process with rate of occurrence h_0 (i.e., mean interarrival time $1/h_0$). Each RGM (also called a cluster center) gives rise to a group of rainfall events and each of these groups is called a cluster. Within each cluster, the events occur independently and their occurrence is completely specified by the distribution of the number of events and the distribution of their positions relative to their cluster center. *Kavvas and Delleur* [1981] assumed a geometric distribution, with parameter p, for the number of rainfall events in a cluster and an exponential distribution, with parameter θ, for the distances of events from their cluster centers. For these distributions, the observed process has a rate of occurrence $m = h_0/p$. To cope with the long-term trends and within-year seasonality Kavvas and Delleur applied a homogenization scheme to the data. This scheme consisted of fitting a time-varying function $\lambda(t)$ to the mean rate of occurrence and rescaling the original time increments of one day, Δt, to time-varying increments $\Delta\tau = \lambda(t)\Delta t$. In this way a stationary process (i.e., a process independent of the time origin) was obtained to which a N-S model was fitted. However, as Kavvas and Delleur comment, the fitted model cannot be used for simulation of daily rainfall occurrence sequences, since the inverse transformation, i.e., the deterministic transformation that will give the nonhomogeneous process from the homogeneous one is not valid for a non-Poisson process [*Çinlar*, 1975, chapter 4].

Smith and Karr [1983] introduced another point process, the renewal Cox process with Markovian intensity (RCM process), which belongs to the class of doubly stochastic Poisson processes (also called Cox processes). In an RCM process, the

TABLE 1. Statistical Properties of a Poisson and a Bernoulli Process

	Poisson	Bernoulli
Interarrival times, X_i	$f(x) = \lambda e^{-\lambda x}$ $\lambda > 0$ $E(X) = 1/\lambda$ $\text{Var}(X) = 1/\lambda^2$ $c_v = 1$ $c_s = 2$	$p(x) = p(1-p)^{x-1}$ $0 \le p \le 1$ $E(X) = 1/p$ $\text{Var}(X) = (1-p)/p^2$ $c_v = (1-p)^{1/2} < 1$ $c_s = \frac{2-p}{(1-p)^{1/2}} > 2$
Number of events, N_t	$p(N_t = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$ $E(N_t) = \lambda t$ $\text{Var}(N_t) = \lambda t$	$p(N_r = k) = \binom{k}{r} p^k (1-p)^{r-k}$ $E(N_k) = pk$ $\text{Var}(N_k) = p(1-p)k$
Conditional intensity function	$h(t) = \lambda$	$h_k = p$
Log survivor function	$\ln [R(x)] = -\lambda x$	$\ln [R(x)] = -\ln (1-p)x$
Variance time curve	$V(t) = \lambda t$	$V_k = p(1-p)k$
Index of dispersion function	$I(t) = 1$ $\forall t$	$I_k = 1 - p < 1$ $\forall k$
Spectrum of counts	$g_+(\omega) = \lambda/\pi$ $\omega \ge 0$	$g_+(\omega) = p(1-p)/\pi$ $\omega \ge 0$
Normalized spectrum of counts	$g_+'(\omega) = 1$ $\omega \ge 0$	$g_+'(\omega) = 1 - p < 1$ $\omega \ge 0$

Here λ, rate of occurrence; p, probability of success; c_v, coefficient of variation; and c_s, skewness coefficient. The properties of the discretized Poisson process can be obtained from the right-hand column by simply substituting the value of $\lambda' = 1 - e^{-\lambda}$ for the value of p.

rate of occurrence $\lambda(u)$ alternates between two states, one zero and the other positive. Retaining the nomenclature of Smith and Karr, let a_1 and a_2 be the parameters of the exponential sojourn distributions of the intensity $\lambda(u)$ in the states 1 (dry) and 2 (wet), respectively. This simply amounts to $1/a_1$ being the mean duration of a dry period and $1/a_2$ the mean duration of a wet period. During periods when the intensity is zero, no events can occur; during periods with positive intensity, events occur according to a Poisson process with rate of occurrence λ, and the sequence of states visited form a Markov chain. This process is a renewal one (i.e., interarrival times are independent) and Smith and Karr found it an adequate model of the summer season (July to October) daily rainfall occurrences in the Potomac River basin. It should be noted that both the Neyman-Scott process and the renewal Cox process with Markovian intensity are continuous-time point processes and clustered (overdispersed) relative to the Poisson process.

The problem of fitting and validating daily rainfall occurrence models has not received as much attention as the specification of the model form. *Kavvas and Delleur* [1981] used an iterative least squares estimation procedure, utilizing the estimated normalized spectrum of counts and the estimated log-survivor function, for the fitting of the N-S model to daily rainfall occurrences from Indiana. Problems were encountered in this procedure, however, since estimates of the implied variance of v, the number of events in a cluster, can become negative. To avoid this problem an iterative estimation procedure is required which involves a somewhat arbitrary assumption about the value of $E(v^2)/E(v)$. The problem of negative variance for the number of events in a cluster was also encountered by *Ramirez-Rodriguez and Bras* [1982], who fitted the N-S model to the daily rainfall occurrences of Denver, Colorado. Both of the above studies assessed the validity of the fitted model by the agreement of the theoretical and empirical spectrum of counts and log-survivor function, which were the same functions used for the fitting. However, the variance time curves inferred from the fitted parameters in both studies fail to reproduce the empirical ones, often dramatically.

We believe that the problems encountered by both *Kavvas and Delleur* [1981] and *Ramirez-Rodriguez and Bras* [1982] in fitting the N-S model are due to the inappropriate use of continuous-time point process models for daily rainfall occurrences, rather than shortcomings in the parameter fitting procedure. In that respect, even the maximum likelihood estimation (MLE) methods recently developed by *Smith and Karr* [1985] may cause difficulties when used for modeling discrete (e.g., daily) rainfall occurrences, since they use the likelihood function of the continuous-time process.

4. Inferences About Clustering of Daily Rainfall Occurrences

In the theory of continuous-time point processes, the existence and type of clustering in an occurrence process is often studied by comparing the process to an independent Poisson process with the same rate of occurrence. A clustered point process can be either overdispersed (i.e., more random occurrences) or underdispersed (i.e., more regular occurrences) relative to a Poisson process. However, it is important that the clustering of a discrete-time point process, such as the daily rainfall occurrence process, be compared with the independent Bernoulli process (the discrete-time analogue of the Poisson process). Most previous studies have treated the daily rainfall occurrence process as a continuous-time point process and have modeled it in a continuous-time point process framework; i.e., clustering of daily rainfall has been inferred by comparing the empirical properties of the observed occurrence series to the theoretical properties of the Poisson process. In this section it is shown that such an approach can result in incorrect inferences about clustering of the underlying process. In particular, it is shown that if indeed the daily rainfall occurrences were an independent process, i.e., a Bernoulli process, if modeled as a continuous-time point process they would be interpreted as underdispersed (relative to the Poisson process). On the other hand, daily rainfall occurrence series underdispersed relative to the Poisson process are, in fact, all shown to be overdispersed relative to Bernoulli. To illustrate these points, the statistical properties of a Bernoulli process are first studied.

Consider a sequence of independent repeated trials with two possible outcomes, say, success and failure. Let p denote the probability of success at each trial and N_r the number of

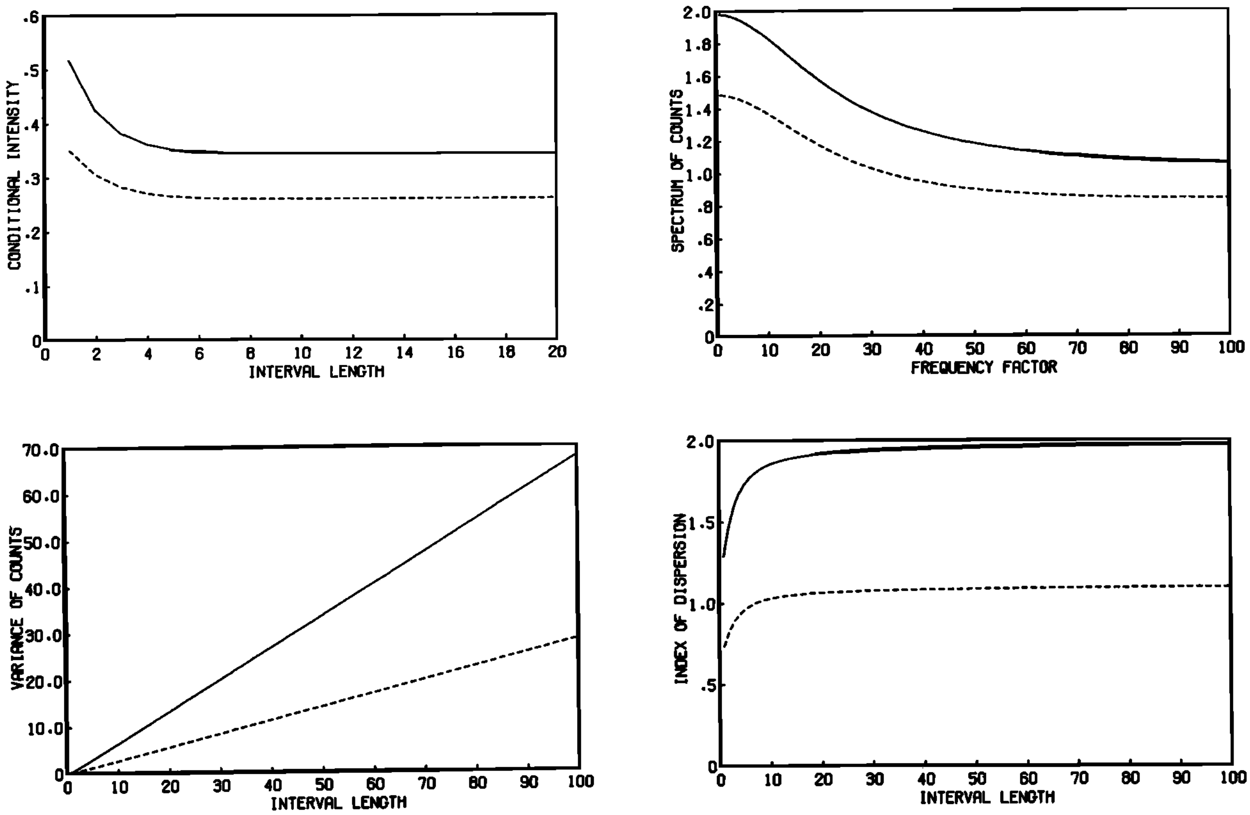

Fig. 2. Effects of discretization on a Neyman-Scott (N-S) process with parameters $h_0 = 0.23$, $p = 0.67$, and $\theta = 0.75$. Solid curve corresponds to the continuous-time process and the broken curve to the discrete-time process.

successes in r trials. Then N_r has a binomial probability distribution

$$P(N_r = k) = \binom{r}{k} p^k(1-p)^{r-k} \qquad k = 0, 1, 2, \ldots \qquad (4)$$

and the number of trials between the nth and $(n+1)$st success, X_n, has a geometric distribution

$$P(X_n = k) = p(1-p)^{k-1} \qquad k = 1, 2, \ldots \qquad (5)$$

for all n. Note that for the daily rainfall occurrences, the number of trials is interpreted as the number of discrete-time units, i.e., days. As was mentioned earlier, in the discrete-time point process terminology a success corresponds to the occurrence of an event (i.e., a rainy day), N_r to the counting process, that is the number of events in $\{0, \ldots, r\}$, and X_n to the time between events.

The Bernoulli process is the discrete-time analogue of the Poisson process, in the sense that it is characterized by independent intervals and independent counting increments. This lack of memory property is the result of the geometric distribution for the times between events, analogously to the exponential distribution for the Poisson (see *Feller* [1968, p. 329] for a proof). The statistical properties (i.e., mean, variance, and higher moments) of the geometric and binomial distributions are well known (see, for example, *Parzen* [1967]). For this work, some additional properties of the Bernoulli counting process are of interest, such as the spectrum of counts, log survivor function, and variance time curves. These properties can be easily computed (see, for example, *Foufoula-Georgiou* [1985]) and are summarized in Table 1 of this paper, together with the corresponding properties of a Poisson process with rate of occurrence λ. The significance of comparing these statistical properties is illustrated below.

Consider a sequence of daily rainfall occurrences. If a Bernoulli process is fit to the series, the maximum likelihood estimate (MLE) of its probability of success, p, is $\hat{p} = N_0/N$, where N_0 is the number of days on which rainfall is recorded and N is the total number of days. Similarly, if a Poisson process is fit (under the interpretation that rainfall events represent all of the events of the point process), the MLE of the rate of occurrence λ, is $\hat{\lambda} = N_0/N$. Thus $\hat{p} = \hat{\lambda}$. Notice, however, from Table 1 how different the other properties of the two processes are. In particular, the Bernoulli process has a coefficient of variation of interarrival times and an index of dispersion function always less than one, which imply underdispersion relative to Poisson. This means that inferences about over- and underdispersion of the daily rainfall occurrences would be different depending on whether the empirical functions of the process were compared to those of a Poisson or to those of a Bernoulli process. This is a simple but important observation and has immediate consequences in the interpretation of the statistical functions of the daily rainfall occurrence process. We suggest that a discrete-time point process model, such as daily rainfall occurrences, should be compared with the discrete-time independent Bernoulli process (and not with the continuous-time Poisson) if inferences about independence and clustering are to be made.

5. EFFECTS OF DISCRETIZATION ON CONTINUOUS-TIME POINT PROCESSES

In deriving a discrete-time occurrence series (binary series) from a continuous-time point process, two operations are performed: discretization and clipping. These operations are defined as follows: discretization is the grouping of events occurring within intervals of length equal to the time scale of measurement, and clipping is the assignment of the value of zero (or one) to each interval depending on whether or not at

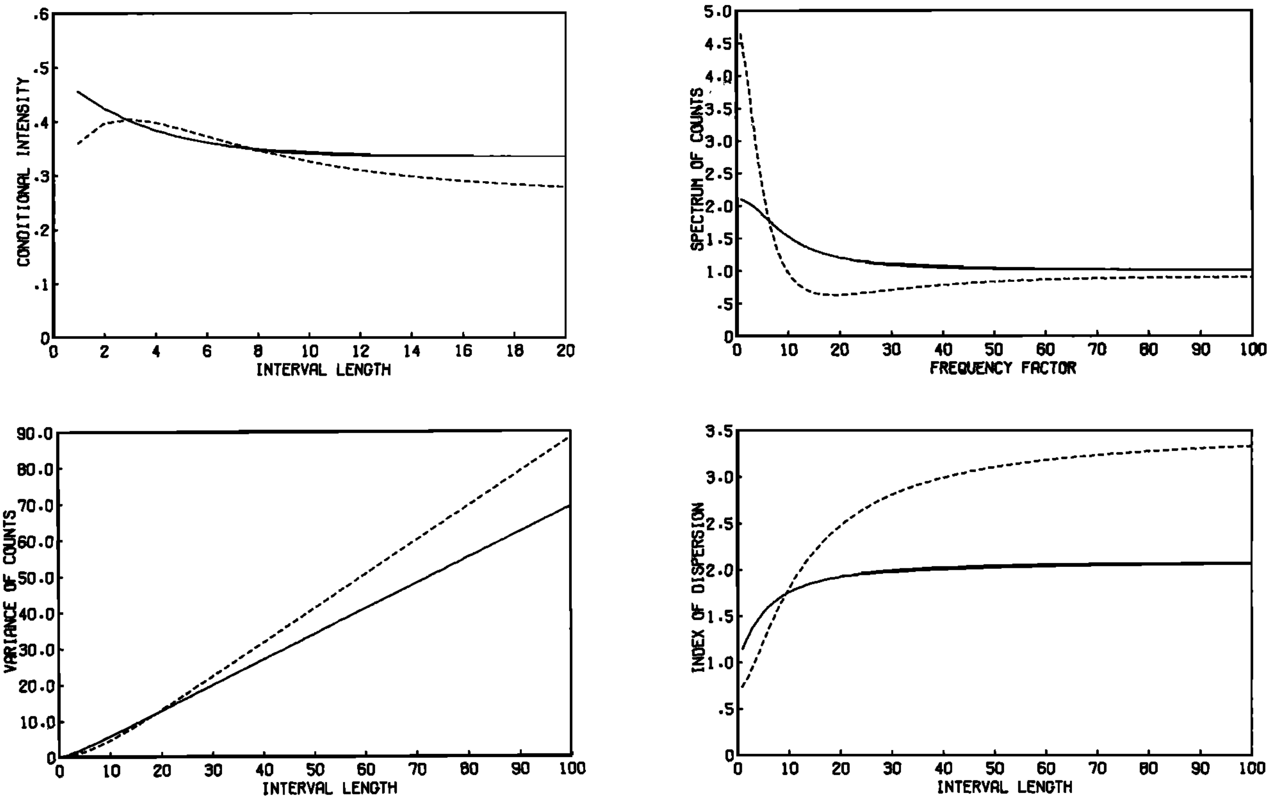

Fig. 3. Effects of discretization on a renewal Cox process with Markovian intensity (RCM) with parameters $a_1 = 0.2$, $a_2 = 0.1$, and $\lambda = 0.5$. Solid curve corresponds to the continuous-time process and the broken curve to the discrete-time process.

least one event occurred within the interval. In this section, the effects of discretization and clipping (usually referred to only as discretization) of the Poisson process, the Neyman-Scott process, and the renewal Cox process with Markovian intensity, all of which have previously been used for modeling daily rainfall occurrences, are studied.

5.1. *Poisson Process*

Let $F(x)$ denote the cumulative distribution function of the exponential pdf of the interarrival times in a Poisson process. The above discretization scheme is equivalent to replacing the continuous exponential distribution, $f(x)$, with a discretized one, $p(k)$, such that

$$p(k) = F(k) - F(k-1) = e^{-(k-1)\lambda}(1 - e^{-\lambda}) \quad k = 1, 2, \ldots \tag{6}$$

Note that the resulting discrete pmf, $p(k)$, is geometric with parameter

$$\lambda' = 1 - e^{-\lambda} \tag{7}$$

implying that the discretized Poisson process is a Bernoulli process with a probability of occurrence equal to λ', a value always less than λ. All the other properties of the discretized Poisson process can therefore be obtained by substituting the value of λ' for p in the right-hand column of Table 1.

It is seen, for example, that the discretized Poisson process has a coefficient of variation $c_v = (1 - \lambda')^{1/2} < 1$, an index of dispersion $I_k = 1 - \lambda' < 1$, and a normalized spectrum of counts $g_+'(\omega) = 1 - \lambda' < 1 \ \forall \ \omega$. This observation illustrates that if a daily rainfall occurrence process is modeled as a Poisson process, and if subsequently this Poisson process is used for simulation of rainfall, the resulting discrete-time rainfall occurrence series will have properties substantially different from those of the inferred continuous-time process. To better illustrate this point, consider a Poisson process with rate of occurrence $\lambda = 0.5$ (mean interarrival time of 2 days). If this process is used for simulation of daily rainfall occurrence, the synthetic arrival process will have a rate of occurrence $\lambda' = 1 - \exp(-0.5) = 0.393$ (mean interarrival time of 2.5 days) and a coefficient of variation equal to $0.626 < 1$. Further, if the simulated occurrence sequences is analyzed under the continuous-time point process framework, they will indicate a process underdispersed relative to the Poisson process, where, in fact, it is an independent Bernoulli process. In the above context, the improper use of the continuous-time point process framework for the analysis and synthesis of daily rainfall occurrences becomes apparent.

5.2. *Neyman-Scott Process*

The statistical properties of intervals and counts of a continuous-time Neyman-Scott process can be found in the work by *Kavvas and Delleur* [1981], while the corresponding properties of the resulting process after discretization are given in *Guttorp* [1985]. It is interesting to note that although the rate of occurrence of a Neyman-Scott process is $m = h_0/p$, the probability of occurrence of an event in the discretized process is [*Guttorp*, 1985]

$$m' = 1 - e^{-h_0}\left[1 - \frac{p}{1 - (1-p)e^{-\theta}}\right]^{h_0/\theta} \tag{8}$$

Observe that m' is a function not only of the rate of occurrence of the cluster centers (h_0) and the cluster size (p), but also of the dispersion of events within each cluster (θ).

Consider a Neyman-Scott process with parameters $h_0 = 0.23$, $p = 0.67$, and $\theta = 0.75$. These parameters correspond to a clustered (overdispersed) occurrence process and are approximately equal to those *Ramirez-Rodriguez and Bras* [1982]

TABLE 2. Information on the Six Daily Rainfall Stations Analyzed

Station Location	Station Identification	Years Analyzed	Latitude	Longitude	Elevation, ft	Observation Time
Snoqualmie Falls	45-7773	1948–1977	47°33′	121°51′	440	5 P.M.
Roosevelt	02-7281	1948–1977	33°40′	111°09′	2005	7 A.M.
Austin	41-0428	1948–1977	30°18′	97°42′	597	midnight
Miami	08-5663	1949–1978	25°48′	80°16′	12	midnight
Philadelphia	36-6889	1948–1977	39°53′	75°15′	10	midnight
Denver	05-2220	1949–1978	39°46′	104°52′	5286	midnight

One foot equals 30.48 cm.

found for daily rainfall occurrences at Denver, Colorado. This arrival process has a mean rate of occurrence $m = 0.34$ (mean interarrival time of approximately 3 days), while the arrival process resulting after discretization has a mean rate of occurrence $m' = 0.260$, that is, a mean interarrival time of approximately 4 days. Figure 2 shows the comparison of the conditional intensity function, normalized spectrum of counts, variance time curve, and index of dispersion for the continuous and discretized N-S process. It should be noted that the different levels in the conditional intensity function and spectrum of counts are the result of the different rate of occurrence of the two processes and do not affect the inferences about clustering, which only depends on the rate of decay of these functions. These two functions, together with the variance time curve and index of dispersion function, indicate that the resulting discrete-time arrival process is less clustered when compared with the continuous-time process it originated from. Also note that severe estimation biases are expected to result if the properties of the continuous (and not discretized) N-S process are used for the fitting.

5.3. *Renewal Cox Process With Markovian Intensity (RCM Process)*

The statistical properties of the RCM process are given in the work by *Smith and Karr* [1983], while the corresponding properties of the resulting process after discretization are given in the work by *Guttorp* [1985]. The parameters selected to illustrate the effects of discretization on an RCM process are $a_1 = 0.2$, $a_2 = 0.1$, and $\lambda = 0.5$. These parameters correspond to a daily rainfall occurrence process in which the dry periods, of an average duration of 5 days ($1/a_1$), are followed by wet periods, of an average duration of 10 days ($1/a_2$); during wet periods rainfall events occur on the average every 2 days ($1/\lambda$). The mean rate of occurrence of the continuous-time RCM process with the above parameters is $m = 0.333$, while the rate of occurrence of the discretized process is $m' = 0.264$. Figure 3 shows the comparison of the conditional intensity function (CIF), normalized spectrum of counts, variance time curve, and index of dispersion for the continuous and discretized RCM processes. It is important to observe that while the CIF of the continuous-time process decreases monotonically to the intensity (this is true for all RCM processes; see *Smith and Karr* [1983]), the CIF of the discretized process increases for lags up to 3 days and then starts decreasing to the intensity of the process m'. This implies that the discretized RCM process has an autocorrelation function r_k, which increases up to lag 3 and then starts decreasing. Such an autocorrelation function is highly atypical for rainfall occurrence series and without physical basis. Notice also from Figure 3 that the discrete-time occurrence process is more clustered than the continuous-time one when both are compared with the Poisson process. However, when the discretized RCM process is compared with a Bernoulli process, it seems to be about as clustered as the continuous RCM process.

6. STATISTICAL ANALYSIS OF DAILY RAINFALL OCCURRENCES

In the preceding sections, it has been shown that the common practice of testing the independence and degree of clustering of daily rainfall occurrences by studying deviations from a Poisson process can be highly misleading. The extent of the differences that can result are demonstrated in this section using daily rainfall occurrences from Snoqualmie Falls, Washington, and Miami, Florida. A complete statistical analysis of four other daily rainfall structures (from Arizona, Texas, Colorado, and Pennsylvania) is given in the work by *Foufoula-Georgiou* [1985]. (All six rainfall stations used for the analyses above are given in Table 2 of the present paper.)

Table 3 shows the mean, standard deviation, coefficient of variation, and skewness coefficient of the interarrival times of the daily rainfall occurrences for Snoqualmie Falls and Miami. It is observed that the coefficient of variation is not always greater than one, and this was the case for some of the other stations analyzed as well. In particular, the winter

TABLE 3. Statistics of Interarrival Times

Month	$\bar{x}$	s_x	c_v	c_s	Number of Events	c_v'
			Snoqualmie Falls			
Jan.	1.40	1.17	0.84	3.82	667	0.53
Feb.	1.50	1.33	0.88	3.93	557	0.58
Mar.	1.54	1.55	1.00	5.54	607	0.60
Apr.	1.72	1.59	0.92	3.06	530	0.65
May	2.21	2.50	1.13	3.19	429	0.74
June	2.69	4.35	1.62	4.94	375	0.79
July	4.26	6.08	1.43	2.91	205	0.87
Aug.	3.32	4.23	1.27	2.26	260	0.84
Sept.	2.71	3.76	1.39	4.27	332	0.79
Oct.	1.75	1.67	0.95	3.08	499	0.66
Nov.	1.44	1.23	0.85	4.22	613	0.55
Dec.	1.34	0.98	0.73	4.17	694	0.50
			Miami			
Jan.	4.64	4.89	1.05	2.24	197	0.89
Feb.	6.16	6.37	1.03	1.63	131	0.92
Mar.	5.80	6.20	1.07	1.87	162	0.91
Apr.	4.99	5.75	1.15	1.96	174	0.89
May	2.26	2.61	1.16	3.33	375	0.75
June	1.99	1.97	0.99	3.25	444	0.71
July	2.16	1.92	0.89	2.48	439	0.73
Aug.	1.83	1.47	0.80	2.40	502	0.67
Sept.	1.84	1.62	0.88	2.64	486	0.68
Oct.	2.65	3.07	1.16	3.24	366	0.79
Nov.	4.74	5.13	1.08	1.99	197	0.89
Dec.	4.91	5.52	1.12	2.16	195	0.89

Here $\bar{x}$, sample mean; s_x, sample standard deviation; c_v, sample coefficient of variation; c_s, sample coefficient of skewness; and c_v', coefficient of variation of the Bernoulli process with the same probability of occurrence as the rainfall series.

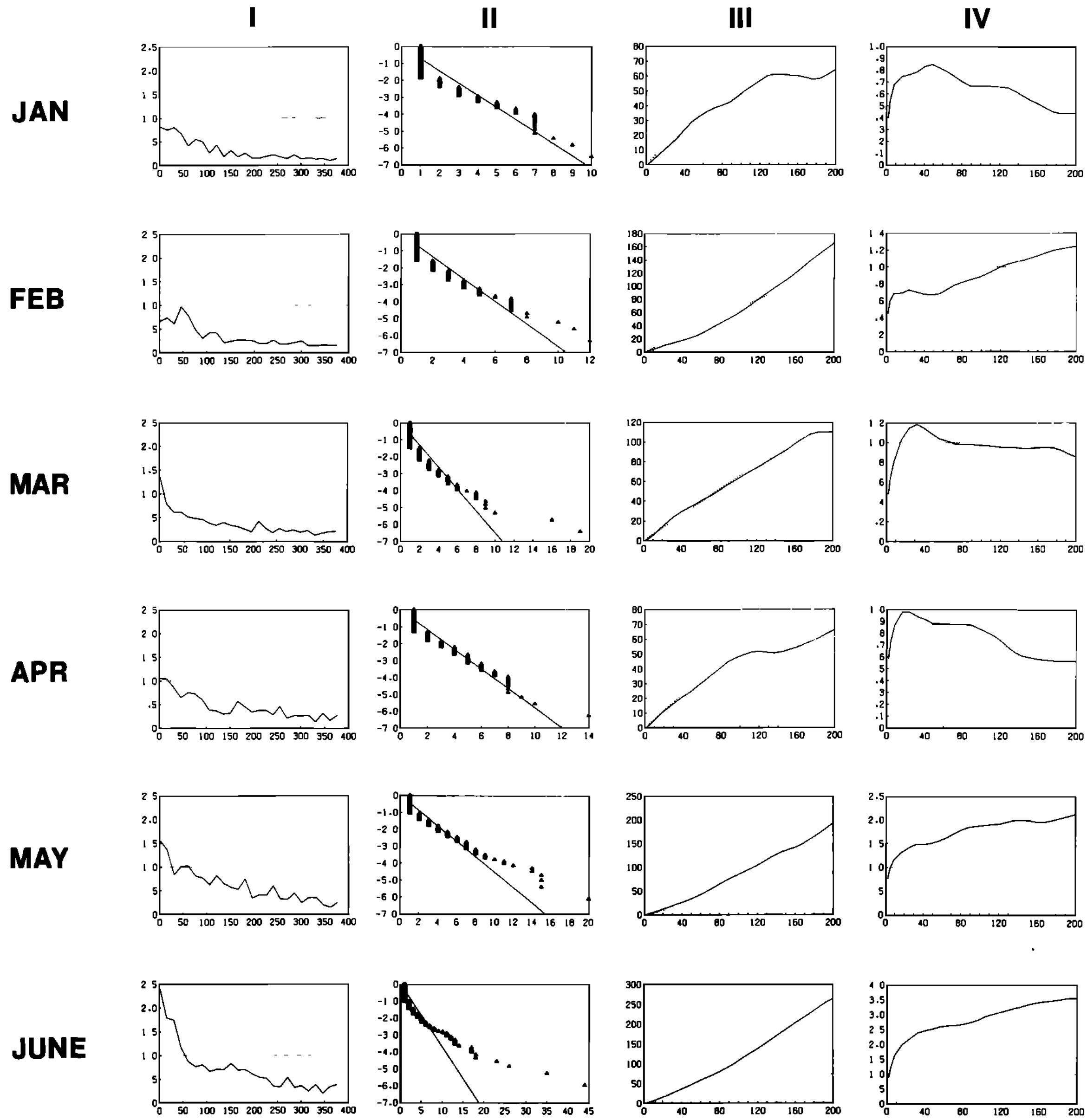

Fig. 4. Statistical properties of intervals and counts for daily rainfall occurrences at Snoqualmie Falls, Washington. I, normalized spectrum of counts versus frequency factor; II, log-survivor function versus interarrival time (days); III, variance of counts versus interval length (days); and IV, index of dispersion versus interval length (days).

months (October–February) for Snoqualmie Falls, the summer months (May and June) for Roosevelt, the summer months (June–September) for Miami, and most of the months (January–April, June, July, November, and December) for Philadelphia have a coefficient of variation less than one. Observe that this underdispersion is consistently shown in the spectrum of counts, log survivor function, variance time curve, and index of dispersion for these months and stations as well (see Figures 4 and 5). Therefore for these months comparison with the Poisson process would conclude that there is no clustering in rainfall and that rainfall events occur in a pattern more regular than that of an independent Poisson process. In fact, none of the available models would be able to accommodate such structures, since both the Neyman-Scott and renewal Cox process with Markovian intensity have a coefficient of variation greater than one. Previous studies have suggested that deterministic explanation for such regular occurrences should be sought. However, under the suggested approach, all these rainfall series appear to have a clustered structure. For example, Table 3 shows that the coefficient of variation of daily rainfall interarrival times is always greater than that of the Bernoulli process (the latter estimated as $c_v' = (1 - 1/\bar{x})^{1/2}$). It should be noted that for such structures the discretized Neyman-Scott or RCM processes may be appropriate models, since both can admit coefficient of variations less than one.

In reference to the spectrum of counts (see Figures 4 and 5), the following should be noted. For a continuous-time point process where theoretically, at least, events can occur arbitrarily close to each other, the spectrum of counts extends to $\omega = \infty$. For the daily rainfall occurrences, however, events cannot occur closer than one day apart and this introduces a cutoff frequency (Nyquist frequency) $\omega_N = \pi$, or equivalently, $f_N = 1/2$ days^{-1}. The value plotted on the abscissa of the spectrum of counts plots is called the frequency factor and is defined as $j = \omega T/2\pi$, where T is the total length of observation. Therefore the frequency factor corresponding to the Nyquist frequency is $j_N = T/2$, and this is the maximum value over

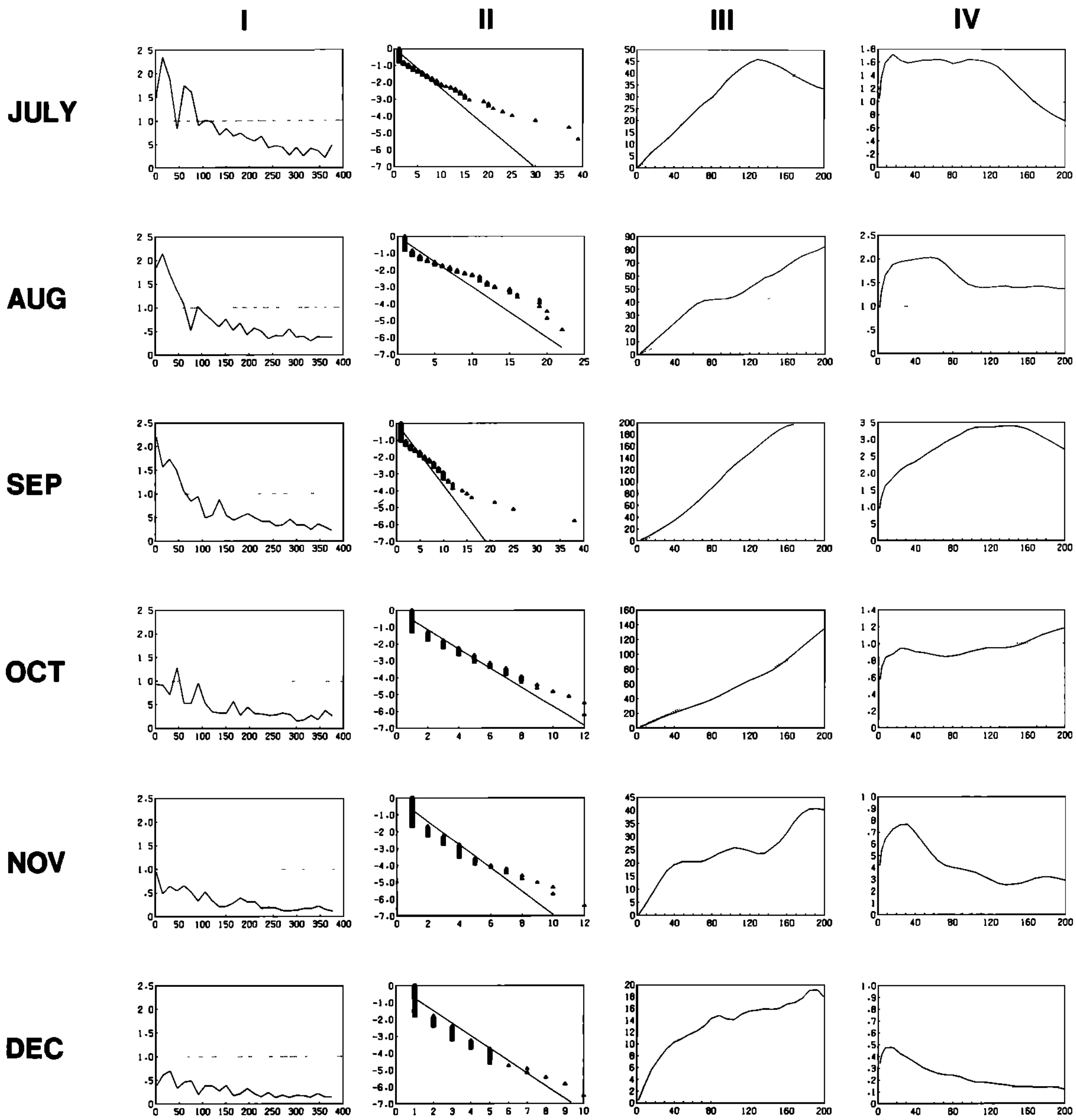

Fig. 4. (continued)

which the spectrum of counts should be computed. *Guttorp and Thompson* [1983] discuss aliasing of the spectrum of counts estimated from discrete sampled counting processes, and show that this can be severe, especially when the spectrum of counts does not decrease rapidly with respect to the sampling interval. For the daily rainfall occurrences, it was seen that the spectrum of counts began to rise at high frequencies, apparently due to aliasing, but the effects of aliasing introduced into lower frequencies cannot be easily assessed. The spectral estimates shown in Figures 4 and 5 were obtained using a uniform averaging over 15 nonoverlapping intervals. Notice that the normalized spectra of counts for most of the months decrease with increasing frequency to a value less than one and approximately equal to $1 - m$, where m is the estimated rate of occurrence. For the months that have coefficients of variation less than one, the spectrum of counts is either approximately constant (indicating an independent Bernoulli process) or increases slightly over a range of low frequencies and then decreases. Such spectra of counts are usually consistent with variance time curves below the one for the Poisson process, indicating underdispersion relative to Poisson. However, all these structures are overdispersed relative to the Bernoulli, since the variance time curve of the Bernoulli process has a slope equal to $m(1 - m) < m$.

The log survivor function has been plotted in such a way as to emphasize the discreteness of the interarrival times. For example, for an interarrival time $x = x_0$ multiple points (triangles) are shown on the plot to illustrate the number of ties, i.e., number of intervals of length x_0. To interpret the log-survivor function, i.e., concavity or convexity and slope, only the lowermost points (triangles) at each entry are needed. Also, the full length of interarrival times has been retained to illustrate extreme situations. These extreme points, however, are less reliable and should be given less weight if the log-survivor function is used for model fitting.

7. SAMPLING CONSIDERATIONS

The previous discussion was oriented primarily toward daily rainfall sequences. One would naturally wonder if the problems addressed herein are associated with the large sam-

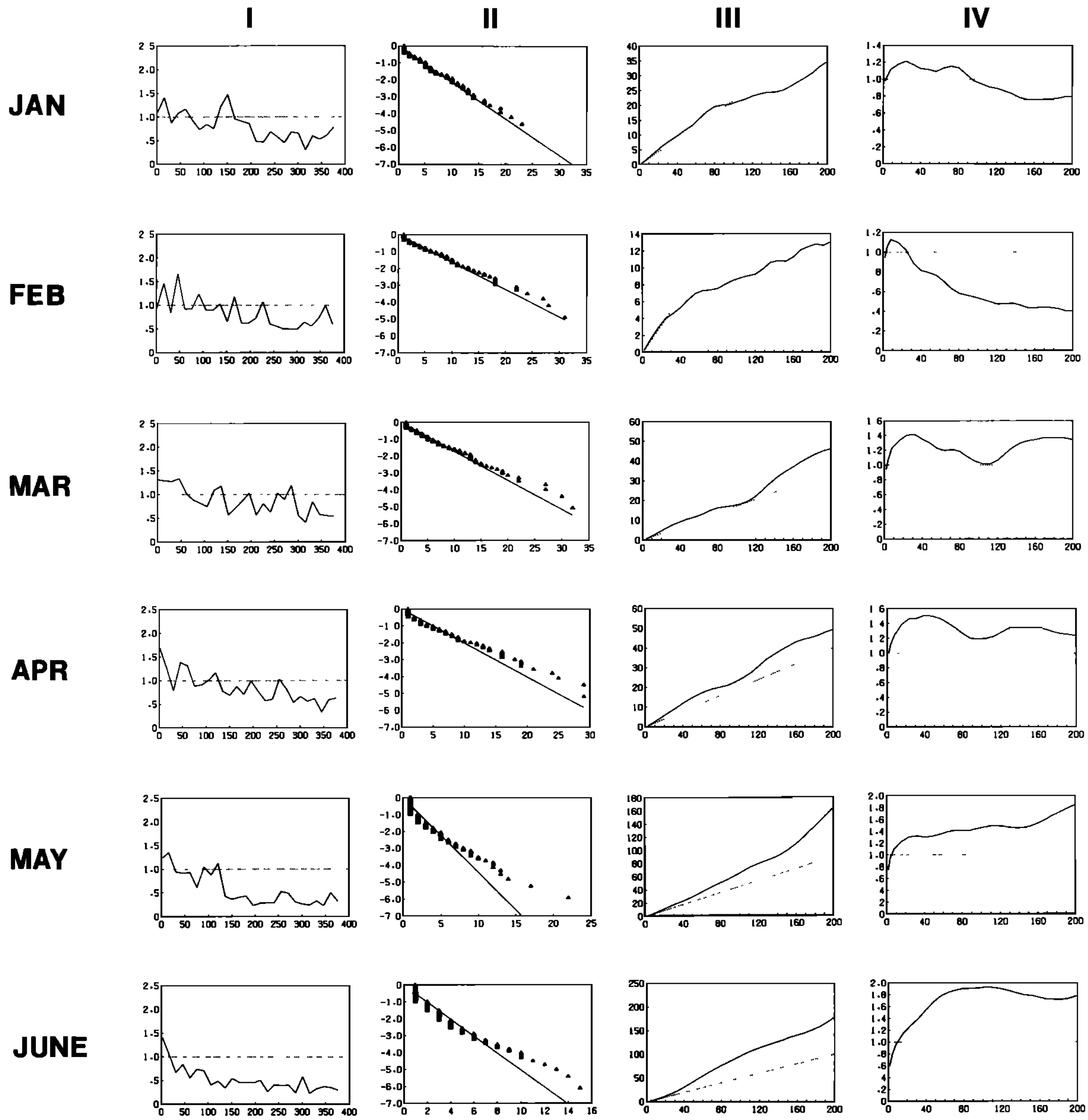

Fig. 5. Statistical properties of intervals and counts for daily rainfall occurrences at Miami, Florida. I, normalized spectrum of counts versus frequency factor; II, log-survivor function versus interarrival time (days); III, variance of counts versus interval length (days); and IV, index of dispersion versus interval length (days).

pling interval of 1 day and would not exist for, say, hourly rainfall. In this section it is shown that a sampled rainfall occurrence series cannot be considered as a continuous point process, even at sampling intervals as short as a few minutes.

Let Δ denote the sampling time interval in days. Then the rate of a Poisson process (m_P) and Neyman-Scott process (m_{NS}) are given as

$$m_P(\Delta) = (1 - e^{-\lambda\Delta})/\Delta$$

and

$$m_{NS}(\Delta) = \left[1 - e^{-h_0\Delta}\left(\frac{p}{1-(1-p)e^{-\theta\Delta}}\right)^{h_0/\theta}\right]\Big/\Delta$$

respectively [*Guttorp*, 1985], where all the parameters have been defined previously. Figure 6 shows $m_P(\Delta)$ for an arrival rate $\lambda = 2$ days^{-1}. We see that observations every 10 min induce only minor bias in the mean function, but hourly discretization has serious effects. On the same figure, the mean function of a N-S process $m_{NS}(\Delta)$ is shown for $h_0 = 0.1$, $p = 0.05$, and $\theta = 5.0$, approximately the parameters obtained by *Rodriguez-Iturbe et al.* [1984] for Denver rainfall data during the period of May 15 to June 16. This N-S process has the same rate of occurrence ($h_0/\theta = 2$ days^{-1}) as the Poisson process, but Figure 6 suggests that in this case observations every 30 s would be required in order to avoid serious bias in the mean function. Again, hourly data are completely unsatisfactory when regarded as observations from this continuous-time process. It is important also to observe from Figure 6 that the effects of discretization strongly depend on the degree of clustering. In general, the more clustered a process, the larger the loss of information through discrete sampling. Therefore it appears to be inadvisable to consider rainfall occurrences as the events of a point process and to use the continuous-time methodology for modeling, even when sampling intervals are as small as a few minutes. It should also be noted that the practical difficulties of measuring small rainfall accumulations, such as would occur over such short time in-

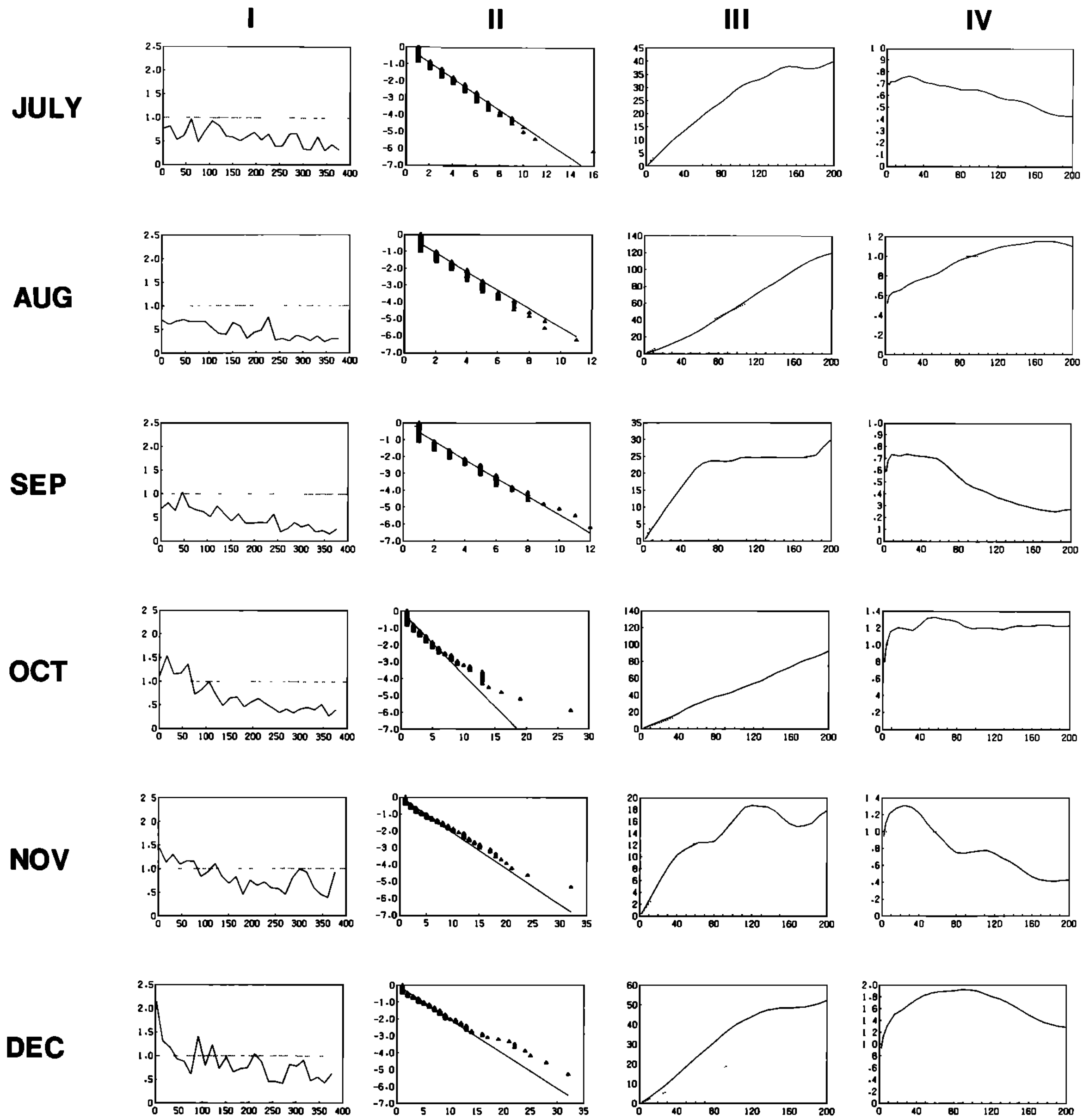

Fig. 5. (continued)

tervals, effectively preclude the application of the continuous-time point process framework to any real rainfall series of hydrologic interest.

8. SUMMARY AND CONCLUSIONS

Point process models for daily rainfall occurrences at a single station have been the subject of several earlier studies. Serious problems were encountered in fitting the Neyman-Scott model to daily rainfall occurrences in two of these studies. We suggest that these problems may be the result of the inappropriate use of continuous-time point process methodology for the daily rainfall occurrences. If daily rainfall occurrences are interpreted as (all of) the events of a point process, they form a discrete-time point process in which events can only occur at time marks integral multiples of the sampling interval apart. Although previous studies have implemented the above interpretation of rainfall occurrences, they have failed to account for time discreteness and have followed a continuous-time point process framework for modeling. In this paper, it has been shown that such an approach can induce severe biases in estimation and can result in misleading interpretations regarding rainfall clustering. We suggest that rainfall occurrences should be compared with the discrete-time independent Bernoulli process (and not with the continuous-time Poisson) if inferences about clustering and dependencies are to be made. We have shown that interpretations regarding clustering made under the continuous-time versus discrete-time framework differ, often critically. For example, daily rainfall structures which are underdispersed relative to Poisson (i.e., more regular occurrences than a Poisson process) are, in general, overdispersed relative to Bernoulli (i.e., more random occurrences than in a Bernoulli process).

The other issue addressed in this paper is the fitting of point process models to daily rainfall occurrences. Fitting procedures which directly use the parameters of the continuous-time point processes (and which have been extensively used in the hydrologic literature) have been shown to induce severe biases in the parameter estimates. Such procedures should not be used even in modeling hourly rainfall, since the sampling interval under which rainfall occurrence data may be con-

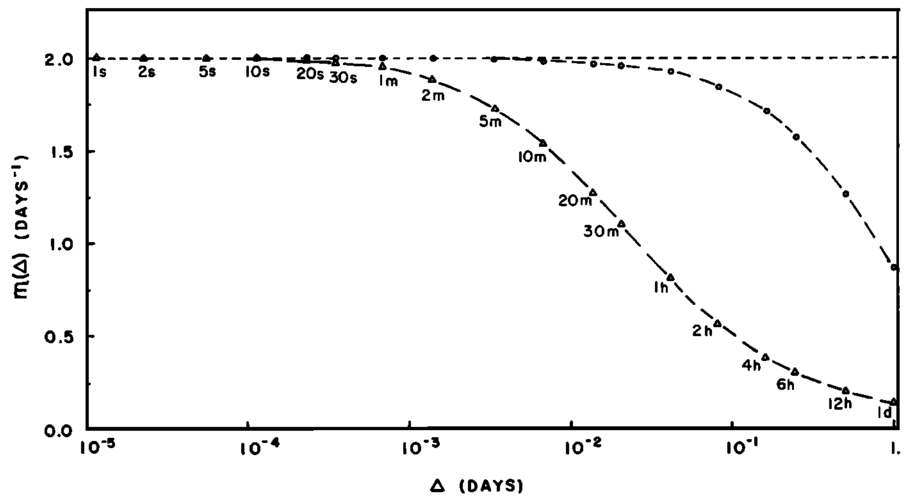

Fig. 6. Rate of occurrence of a discretized Poisson process with rate of occurrence $\lambda = 2$ days^{-1} (circles) and a discretized Neyman-Scott process with parameters $h_0 = 0.1$ days^{-1}, $p = 0.05$, and $\theta = 5.0$ days^{-1} (triangles) as function of the discretization time interval Δ (in days). Here s, m, h, and d denote seconds, minutes, hours, and days.

sidered approximately continuous is on the order of a few minutes. We recommend instead that the appropriate approach to modeling rainfall data using continuous-time point processes is the one followed by *Rodriguez-Iturbe et al.* [1984] and *Foufoula-Georgiou and Guttorp* [1985]. An alternate approach is to develop discrete-time point process models. *Foufoula-Georgiou* [1985] proposes the use of a semi-Markov model in this context. It seems, however, that the development of other classes of discrete-time point process models, as, for example, the discrete-time analogue of the Neyman-Scott process, will involve cumbersome closed form solutions, if they are feasible at all.

Acknowledgments. The authors gratefully acknowledge helpful discussions with P. Guttorp of the Statistics Department, University of Washington, and, in particular, his comments on sampling considerations (section 7). The critical comments of three anonymous referees are greatly appreciated.

REFERENCES

Bartlett, M. S., The spectral analysis of point processes, *J. R. Stat. Soc., Ser. B*, 25, 264–296, 1963.

Chang, T. J., M. L. Kavvas, and J. W. Delleur, Daily precipitation modeling by discrete autoregressive moving average processes, *Water Resour. Res.*, *20*(5), 565–580, 1984.

Çinlar, E., *Introduction to Stochastic Processes*, Prentice-Hall, Englewood Cliffs. N. J., 1975.

Cox, D. R., and P. A. W. Lewis, *The Statistical Analysis of Series of Events*, Chapman and Hall, London, 1978.

Daley, D. J., and D. Vere-Jones, A summary of the theory of point processes, in *Stochastic Point Processes: Statistical Analysis, Theory and Applications*, edited by P. A. W. Lewis, John Wiley, New York, 1972.

Duckstein, L., 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, John Wiley, New York, 1968.

Foufoula-Georgiou, E., Discrete-time point process models for daily rainfall, Ph.D. dissertation, Univ. of Fla., Gainesville, 1985.

Foufoula-Georgiou, E., and P. Guttorp, Compatibility of continuous rainfall occurrence models with discrete rainfall observations, *Tech. Rep. 25*, Dep. of Stat., Univ. of Br. Columbia, Vancouver, 1985.

Gabriel, K. R., and J. Neumann, A Markov chain model for daily rainfall occurrences at Tel Aviv, *Q. J. R. Meteor. Soc.*, *88*, 90–95, 1962.

Grace, R. A., and P. S. Eagleson, The synthesis of short-time-increment rainfall sequences, *Hydrodyn. Lab. Rep. 91*, Dep. of Civ. Eng., Mass. Inst. of Technol., Cambridge, Mass., 1966.

Green, J. R., A model for rainfall occurrence, *J. R. Stat. Soc., Ser. B*, *26*, 345–353, 1964.

Guttorp, P., On binary time series obtained from continuous-time point processes describing rainfall, *Tech. Rep. 9*, Dep. of Stat., Univ. of Br. Columbia, Vancouver, 1985.

Guttorp, P., and M. L. Thompson, Estimating point process parameters from sampled counting processes, Statistical Inference for Point Processes, *Tech. Rep. 35*, Dep. of Stat., University of Washington, Seattle, 1983.

Kavvas, M. L., and J. W. Delleur, A stochastic cluster model of daily rainfall sequences, *Water Resour. Res.*, *17*(4), 1151–1160, 1981.

Lawrance, A. J., Some models for stationary series of univariate events, in *Stochastic Point Processes: Statistical Analysis, Theory and Applications*, edited by P. A. W. Lewis, John Wiley, New York, 1972.

LeCam, L., A stochastic description of precipitation, in *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, University of California, Berkeley, 1961.

Lewis, P. A. W., Remarks on the theory, computation and application of the spectral analysis of series of events, *J. Sound. Vib.*, *12*(3), 353–375, 1970.

Neyman, J. E., On a new class of contagious distributions applicable in entomology and bacteriology, *Ann. Math. Stat.*, *10*, 35–57, 1939.

Neyman, J. E., and E. L. Scott, A statistical approach to problems of cosmology, *J. R. Stat. Soc., Ser. B*, *20*, 1–43, 1958.

Parzen, E., *Stochastic Processes*, Holden-Day, San Francisco, Calif., 1967.

Ramirez-Rodriguez, J. A., and R. L. Bras, Optimal irrigation control using stochastic cluster point processes for rainfall modeling and forecasting, *Tech. Rep. 275*, Ralph M. Parson Lab., Mass. Inst. of Technol., Cambridge, Mass., 1982.

Rodriguez-Iturbe, I., V. K. Gupta, and E. Waymire, Scale considerations in the modeling of temporal rainfall, *Water Resour. Res.*, *11*(20), 1611–1619, 1984.

Roldan, J., and D. A. Woolhiser, Stochastic daily precipitation models, 1, A comparison of occurrence processes, *Water Resour. Res.*, *18*(5), 1451–1459, 1982.

Smith, E. R., and H. A. Schreiber, Point processes of seasonal thunderstorm rainfall, 1, Distribution of rainfall events, *Water Resour. Res.*, *9*(4), 871–884, 1973.

Smith, J. A., and A. F. Karr, A point process model of summer season rainfall occurrences, *Water Resour. Res.*, *19*(1), 95–103, 1983.

Smith, J. A., and A. F. Karr, Statistical inference for point process models of rainfall, *Water Resour. Res.*, *21*(1), 73–80, 1985.

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

Todorovic, P., and V. Yevjevich, Stochastic processes of precipitation, *Hydrol. Pap. 35*, pp. 1–61, Colo. State Univ., Fort Collins, 1969.

Waymire, E., and V. K. Gupta, The mathematical structure of rainfall representations, 1, A review of stochastic rainfall models, *Water Resour. Res.*, *17*(5), 1261–1272, 1981*a*.

Waymire, E., and V. K. Gupta, The mathematical structure of rainfall representations, 2, A review of the theory of point processes, *Water Resour. Res.*, *17*(5), 1273–1286, 1981*b*.

Waymire, E., and V. K. Gupta, The mathematical structure of rainfall representations, 3, Some applications of the point process theory to rainfall processes, *Water Resour. Res.*, *17*(5), 1287–1294, 1981*c*.

E. Foufoula-Georgiou, St. Anthony Falls Hydraulic Laboratory, Department of Civil and Mineral Engineering, Mississippi River at 3rd Avenue S.E., Minneapolis, MN 55414.

D. P. Lettenmaier, Department of Civil Engineering, FX-10, University of Washington, Seattle, WA 98195.

(Received March 26, 1985;
revised September 18, 1985;
accepted November 12, 1985.)