WATER RESOURCES RESEARCH, VOL. 19, NO. 1, PAGES 95–103, FEBRUARY 1983

A Point Process Model of Summer Season Rainfall Occurrences

JAMES A. SMITH

Interstate Commission on the Potomac River Basin, Rockville, Maryland 20850

ALAN F. KARR

Department of Mathematical Sciences, The Johns Hopkins University, Baltimore, Maryland 21218

A point process model of summer season rainfall occurrences is developed. The model, which is termed an RCM process, is a member of the family of Cox processes (Poisson processes for which the rate of occurrence of events varies randomly over time). Model development is based on counts and interarrival time statistics estimated from Potomac River basin rainfall data. The counting parameters used are the conditional intensity function, index of dispersion, and counts spectrum; the interarrival time parameters are the coefficient of variation and the autocorrelation function. Explicit results are presented for the counts and interarrival time parameters of RCM processes. Of particular importance in this paper is the interpretation of clustering suggested by the form of the RCM process. For the RCM process the rate of occurrence alternates between two states, one of which is 0, the other positive. During periods when the intensity is 0, no events can occur. The form of the intensity process suggests that clustering of summer season rainfall occurrences in the Potomac River basin results from the alternation of wet and dry periods. Computational results are presented for two extensions of the RCM process model of rainfall occurrences: a marked RCM process model of rainfall occurrences and associated storm depths and a bivariate RCM process model of rainfall occurrences at two sites.

1. Introduction

Kavvas and Delleur [1981] and *Waymire and Gupta* [1981] have discussed point process models of rainfall occurrences which possess cluster properties. The models they propose are members of the family of Poisson cluster processes. Kavvas and Delleur claim that the Poisson cluster processes provide not only a model that fits the statistical properties of rainfall occurrences in Indiana (over the entire year) but also a compelling interpretation of the process of clustering, based on frontal passages. In this paper we describe the development of a point process model of summer season rainfall occurrences for the Potomac River basin. Analysis of rainfall data for the Potomac [*Smith*, 1981] indicates that cluster properties are an important feature of summer season rainfall occurrences. The model that is developed here is a member of the family of Cox processes (also known as doubly stochastic Poisson processes) and is called an RCM process (for renewal Cox process with Markovian intensity). A Cox process is a Poisson process with a randomly varying rate of occurrence. The interpretation that we suggest for a Cox process model of rainfall occurrences is that the rate of occurrence of storms is determined by a randomly varying 'climatological process.' The form of the RCM process, and in particular, the presence of a '0' rate, suggests an interpretation of clustering based on alternation of wet and dry periods. *Namias* [1966] has related summer season dry periods in the Northeast to the frequent advection of cold air from Canada. *Smith* [1981] has shown that synoptic conditions during summer season drought periods in the Potomac River basin (1950–1970) correspond closely to the pattern described by Namias. We use this information along with the form of the RCM process to conclude that cluster properties of summer season rainfall occurrences in the Potomac River basin are drought-related phenomena.

During early phases of the study, three classes of stationary point processes (renewal processes, Cox processes, and Poisson cluster processes) were considered as models of summer Potomac rainfall occurrences. Model selection was based on two properties of the data: the counting statistics, which use numbers of events in specified intervals, and interarrival time statistics (or interval statistics), which use the times between events. The counting parameters which we employ are the conditional intensity function, index of dispersion, and counts spectrum. The interval parameters that are used are the coefficient of variation and autocorrelation function.

Model selection is based on two related criteria: consistency with the Potomac rainfall data analyzed by *Smith* [1981] (see also section 3 below) and computational tractability. The latter is necessary not only for application of the model but also for verification that key features of the data, which involve both counts and interval properties, are satisfied. All three classes of processes mentioned above are broad enough to represent diverse kinds of data, and all, but especially the Cox and Poisson cluster processes, have appealing physical interpretations. The three classes, however, have differing degrees of computational tractability, depending on what is to be computed. Renewal processes are tractable in terms of interval properties but not necessarily counts properties, while the reverse is true for Cox processes and Poisson cluster processes. Since one property of the data is the interval property of its being a renewal process, this suggests that one should seek a model that is simultaneously a renewal process and a Cox process or simultaneously a renewal process and a Poisson cluster process. Then the latter property will be exploited for analysis of counts properties. Therefore a key step in model selection is characterization of the intersections of the renewal processes and Cox processes and of the renewal

Paper number 2W1658.
0043-1397/83/002W-1658$05.00

processes and Poisson cluster processes. As described in more detail below, the latter intersection (which consists only of Poisson processes) is too small to reproduce 'cluster' properties of the data. However, a subclass of the renewal Cox processes, namely the RCM processes, is broad enough to contain processes exhibiting all the features of the data.

Another important feature of RCM processes from an applications viewpoint is computational tractability. In addition to obtaining the counts and interval properties of the RCM processes, we illustrate the ease with which the RCM process of rainfall occurrences can be extended to (1) a 'marked' RCM process of rainfall occurrences and associated storm depths and (2) a multivariate RCM process of rainfall occurrences at several sites, both of which are consistent with the data analyzed by *Smith* [1981].

2. DEFINITIONS AND NOTATION

Two classes of stochastic processes are drawn upon in this paper. In this section we present the definitions and properties of point processes and Markov processes that are required in this paper. For further details, consult *Waymire and Gupta* [1981], *Snyder* [1975], *Kallenberg* [1975], *Cinlar* [1975], and *Cox and Isham* [1980].

Let (Ω, F, P) be a probability space and let $R_+ = [0, \infty)$ denote the nonnegative half line. A point process over R_+ is a random process representing times of occurrences of (otherwise indistinguishable) events. Denote by $T(n)$ the time of the nth event, with $T(0) = 0$. Throughout we assume that the sequence $(T(n))$ represents a simple point process in the sense that $T(n) < T(n + 1)$ for each n and $\lim_{n\to\infty} T(n) = \infty$. The first condition prohibits coincident events, while the second ensures that every finite time interval contains only a finite number of events. The interarrival times $U(i) = T(i) - T(i - 1)$ represent times between events. The two sequences $(T(n))$ and $(U(i))$ each contain a full description of the process of events.

Yet another equivalent description of the point process is by the (real time) counting process $(N(t), t \geq 0)$ defined by

$$N(t) = n \qquad t \in [T(n), T(n + 1))$$

Then $N(t)$ is the random number of events occurring in the time interval $(0, t]$ and for $s < t$, $N(t) - N(s)$ is the number of events in the interval $(s, t]$. The fundamental relation between $(T(n))$ and $(N(t))$ is that for each n and t,

$$T(n) \leq t \qquad \text{iff } N(t) \geq n$$

We will use the notation $N(A)$ (i.e., N with an uppercase argument) for the number of events in a subset A of R_+. For our purposes, events will be rainfall occurrences.

We say that two point processes N and M have the same distribution provided that the finite dimensional distributions are identical, i.e.,

$$P(N(A_i) = k_i, i = 1, \cdots, n) = P(M(A_i) = k_i, i = 1, \cdots, n)$$

for all nonnegative integers $n, k_1, \cdots, k_n$ and all sets A_i. We write

$$N \overset{d}{=} M$$

for equality in distribution.

The Laplace functional L_N of a point process N is defined by

$$L_N(f) = E\left[\exp\left(-\int_0^\infty f(s)\, dN(s)\right)\right]$$

where f is a nonnegative function defined on R_+. *Waymire and Gupta* [1981] present a detailed discussion of the importance of such transforms in applications of point process models in hydrology. *Kallenberg* [1975] presents the following theorem, which indicates equivalent methods of specifying the distribution of a point process.

Theorem

Let N and M be simple point processes. Then the following are equivalent:

$$N \overset{d}{=} M \tag{1}$$

$$L_N(f) = L_M(f) \qquad \forall f \geq 0 \tag{2}$$

$$P(N(A) = 0) = P(M(A) = 0) \qquad \forall A \tag{3}$$

The choice of which method to use to specify the distribution of a point process is generally dictated by computational tractability. In general, transforms are easier to obtain than finite dimensional distributions; transforms yield moments by differentiation and facilitate description of asymptotic properties of the process; see *Waymire and Gupta* [1981] for additional details. The zero probability functional $Z(A) = P(N(A) = 0)$ shares computational advantages of the Laplace functional and has the additional advantage of yielding explicit probabilities; however (even though it uniquely defines the law of the process), much useful information (e.g., moments) is not easily obtained from it. Its relative, the zero probability function $Z(t) = P(N(t) = 0) = P(T(1) > t)$, does not uniquely define the distribution of N; however, both are of particular interest in applications pertaining to drought, since they are directly related to the duration of dry periods [see, e.g., *Gupta and Duckstein,* 1975].

Throughout this paper we impose the distributional assumption of stationarity. A point process N is said to be stationary if

$$P(N(I_j + t) = k_j; j = 1, \cdots, n) = P(N(I_j) = k_j; j = 1, \cdots, n)$$

for all choices of $n \geq 1$, nonnegative integers k_j, $t \geq 0$, and intervals I_j, where for an interval $I = (u, v]$, $I + t = (u + t, v + t]$ is the interval obtained by translating I by t units to the right. Stationarity implies that the probabilistic structure of the point process is unaffected by a change in the time origin and can be interpreted heuristically as meaning that the process began long enough ago to have reached a steady state (in which random fluctuations can still occur).

The first- and second-order moments of the counting process $N(t)$ and of the interarrival times $U(i)$ are used for analysis of the Potomac rainfall occurrence data. The counting parameters which are examined are the following.

Conditional intensity function:

$$m_f(t) = \frac{d}{dt} E[N(t)|N(0) = 1]$$

The expectation is defined with respect to the Palm function of N, $P(N(t) = k|N(0) = 1)$, which can be defined as follows:

$$P(N(t) = k|N(0) = 1) = \lim_{s \downarrow 0} P(N(t) - N(s) = k|N(s) \geq 1)$$

Index of dispersion:

$$I(t) = V(t)/mt$$

where $V(t) = \text{var}\,(N(t))$ is the variance time curve of N and

$$m = \frac{d}{dt} E[N(t)]$$

is the intensity of N.

Counts spectrum:

$$g_{+}(\omega) = (1/\pi)\int_{-\infty}^{\infty} \exp\,(i\omega t)\gamma(t)\,dt \qquad \omega > 0$$

where

$$\gamma(t) = \lim_{u \downarrow 0} \text{cov}\,((N(t + s + u) - N(t + s))/u,\ (N(s + u) - N(s))/u)$$

is the covariance density of N (for $t < 0$, $\gamma(t) = \gamma(-t)$; to extend $\gamma(t)$ to 0, $m\delta(t)$ is added to $\gamma(t)$, where $\delta(t)$ is the Dirac delta function). The moment properties used for the interarrival times are the standard parameters used in the analysis of time series, i.e., the mean, variance, coefficient of variation, and autocorrelation function.

Four classes of stationary point processes are discussed in this paper. Below we present definitions for Poisson processes, Cox processes, and renewal processes. *Waymire and Gupta* [1981] present a detailed discussion of Poisson cluster processes.

Poisson processes play a central role in the theory and applications of point processes. Extensive discussion of Poisson processes, from both applied and theoretical viewpoints, can be found in works by *Cox and Lewis* [1966], *Cinlar* [1975], and *Snyder* [1975].

A simple point process N is a homogeneous Poisson process if the following conditions are true.

1. For every finite collection of disjoint sets $A_1, \cdots, A_n$, the random variables $N(A_1), \cdots, N(A_n)$ are independent.
2. There is a constant $m \geq 0$ such that for every A, $N(A)$ has a Poisson distribution with parameter $m|A|$, where $|A|$ is the Lebesgue measure [*Rudin*, 1974] or length of A.

Thus, if $A = (s, s + t]$, $|A| = t$ and

$$P(N(A) = k) = \exp\,(-mt)\,(mt)^k/k! \quad k = 0, 1, 2, \cdots$$

The following are basic distributional properties of a Possion process with parameter m.

1. N is a stationary point process with intensity m.
2. $L_N(f) = \exp\,(-m \int_0^{\infty}(1 - e^{-f(s)})\,ds)$.
3. $Z(A) = \exp\,(-m|A|)$.
4. $Z(t) = \exp\,(-mt)$.
5. $m_f(t) = m$.
6. $I(t) = 1$.
7. $g_{+}(\omega) = m/\pi$.
8. $CV(U(i)) = 1$.
9. $COV(U(i), U(i + j)) = 0$, for $j \neq 0$.

Kingman [1967] refers to the independent increments property (condition 1 in the definition) as the 'completely random' property. *Khinchin* [1956] calls the Poisson process a point process that evolves 'without aftereffects.' *Smith* [1981] proposes a framework for interpreting clustering in point processes as departure from the Poisson process assumptions. The conditional intensity function plays an especially important role in interpreting cluster properties of point processes (as well as in development of the point process model of rainfall occurrences described in the following sections). Specifically, a conditional intensity function greater than the intensity indicates 'random clustering,' in which a point of the process is relatively more likely than an arbitrary point in time to be followed by additional points. On the other hand, a conditional intensity function less than the intensity corresponds to 'regular clustering,' where a point of the process is less likely than an arbitrary point in time to be closely followed by further points. The extreme in regular clustering is a process with equally spaced points. A point process with a constant conditional intensity exhibits neither form of clustering. Note that for a stationary point process the conditional intensity function converges to the intensity, i.e.,

$$\lim_{t \to \infty} m_f(t) = m$$

Thus for both of the cases described above, we are referring to behavior of the conditional intensity function in 'short' intervals following the origin.

The conditional intensity function of a Poisson process (property 5 above) is constant and equal to the intensity m. Thus for a Poisson process the rate of occurrence of events in an interval $(s, s + t]$ is unchanged by knowing that an event occurred at s, i.e.,

$$E[N(t + s) - N(s)|N(\{s\}) = 1] = mt = E[N(t + s) - N(s)]$$

This result has particularly strong connection with intuitive notions of nonclustered processes; for a Poisson process an arbitrary point of the process is neither more likely nor less likely than an arbitrary point in time to be followed by additional points.

Cox [1955] introduced a class of point processes which can be described as Poisson processes with randomly varying intensities. Processes of this type have been referred to as Cox processes [*Kallenberg*, 1975], doubly stochastic Poisson processes [*Bartlett*, 1963], and conditional Poisson processes [*Serfozo*, 1972]. Serfozo refers to the random intensity process as an 'environmental process,' with the interpretation that the occurrence process is operating in a randomly evolving 'environment.' *LeCam* [1961] describes a complicated three-stage model of precipitation that in some specific instances reduces to a Cox process; however, his model does not contain our model as a special case.

A Cox process is defined by means of a stochastic intensity process, which is a nonnegative stochastic process $(\lambda(u), u \geq 0)$. Let Λ be the random measure defined by

$$\Lambda(A) = \int_A \lambda(u)\,du$$

A point process N is a Cox process directed by Λ if the following are true.

1. For every finite collection of disjoint sets $A_1, \cdots, A_n$, the random variables $N(A_1), \cdots, N(A_n)$ are conditionally independent given Λ.
2. For all nonnegative integers k and all sets A,

$$P(N(A) = k|\Lambda) = \exp(-\Lambda(A))\Lambda(A)^k/k!$$

i.e., the conditional distribution of $N(A)$ given Λ is Poisson with parameter $\Lambda(A)$.

The basic distributional properties of the Cox processes are as follows.

1. A Cox process N is a stationary point process if and only if the intensity process $(\lambda(u))$ is a stationary stochastic process [*Karlin and Taylor*, 1975].
2. $L_N(f) = E[\exp(-\int_0^\infty(1 - e^{-f(s)})\lambda(s)\,ds)]$.
3. $Z(t) = E[\exp(-\Lambda(t))]$, where $\Lambda(t) = \int_0^t\lambda(u)\,du$.

Note that if $\lambda(u)$ is constant (i.e., both nonrandom and independent of time) and equal to m, then

$$L_N(f) = \exp\left(-m\int_0^\infty (1 - e^{-f(s)})\,ds\right)$$

Thus the Poisson processes are a subset of the Cox processes.

A point process N is a stationary renewal process if the interarrival times $U(i)$ are IID (independent and identically distributed) with distribution F and $T(1)$, the time to the first event, is independent of the interarrival times $U(i)$ with distribution

$$G(t) = \int_0^t (1 - F(s))\,ds \bigg/ \int_0^\infty (1 - F(s))\,ds$$

For a detailed discussion of renewal processes, consult *Cinlar* [1975].

We will need the following definitions and properties for Markov processes. A stochastic process $(\lambda(u), u \geq 0)$ is a (time homogeneous) Markov process with state space $E = (\lambda_1, \cdots, \lambda_n)$ if for every $t, s > 0, \lambda_i \in E$,

$$P(\lambda(t + s) = \lambda_i|\lambda(u), u \leq t) = P(\lambda(t + s) = \lambda_i|\lambda(t))$$
$$= P_s(\lambda(t), \lambda_i)$$

where $P_s(\ ,\ )$ is a Markov matrix. The family of matrices $(P_u(\ ,\ ), u \geq 0)$ is referred to as the transition function of the process, and satisfies the well-known Chapman-Kolmogorov equation:

$$P_{t+s} = P_t P_s$$

The limit distribution π of $(\lambda(u))$ is given by

$$\pi(i) = \lim_{u\to\infty} P(\lambda(u) = \lambda_i)$$

Conditions for the existence of π and methods of calculating π are given by *Cinlar* [1975].

A finite state Markov process can be characterized as follows [*Cinlar*, 1975]:

1. The sequence of states visited forms a Markov chain.
2. The sojourn times (i.e., the times spent in states) are conditionally independent given the states visited.
3. Each sojourn time has an exponential distribution with parameter dependent only on the state being visited.

3. DEVELOPMENT OF THE POINT PROCESS MODEL

The counts and interarrival time parameters described in the preceding section were estimated for summer season (July–October) rainfall occurrences at 15 sites in the Potomac River basin. (The period of record for each of the gages was 1950–1970. The sequences of arrival times were obtained by associating a single rainfall event with each day on which 0.01 inches of rainfall or greater were recorded. See *Smith* [1981] for details on the data and procedures; the latter are based on those of *Lewis et al.* [1969].) The principal qualitative features of the data are as follows.

1. The conditional intensity function begins much larger than the intensity, then decreases rapidly to the intensity.
2. The index of dispersion of the counting process is greater than one.
3. The counts spectrum decreases with increasing frequency.
4. The coefficient of variation of the interarrival times is greater than one.
5. The interarrival times are uncorrelated.

We also treat the data as a stationary point process. While the assumption of stationarity is seldom valid over the entire year, it is generally possible to break the year into seasons over which the assumption is appropriate. Data analysis by *Smith* [1981] suggests confirmation of the stationarity assumption for July–October rainfall occurrences in the Potomac River basin.

We first note that each of the properties 1–4 fails for Poisson processes. The general interpretation of these properties is that of clustering (i.e., departure from a Poisson process) and, more specifically, that of 'random' rather than 'regular' clustering. We wish to find a stationary point process model that possesses the five characteristics listed.

Our development of a point process model begins with property 5, which suggests (even though independence is a stronger condition than 0 correlation) a renewal process model for rainfall occurrences. A direct approach to obtaining a renewal process model would be to specify an interarrival time distribution F that leads to the properties 1–4 being satisfied. *Cox and Lewis* [1966] present the following formulas that relate counting parameters of a renewal process to the interarrival time distribution:

$$m_f^*(s) = f^*(s)/(1 - f^*(s)) \tag{4}$$

$$V^*(s) = m/s^2 + 2mm_f^*(s)/s^2 - 2m^2/s^3 \tag{5}$$

$$g_+(\omega) = (m/\pi)(1 + m_f^*(i\omega) + m_f^*(-i\omega)) \tag{6}$$

where m is the intensity of the point process,

$$f^*(s) = \int_0^\infty \exp(-st)f(t)\,dt$$

is the Laplace transform of the interarrival time probability density function f (assuming that it exists), and m_f^* and V^* are, respectively, the Laplace transforms of the conditional intensity function and variance time curve. To implement the direct approach, one must specify F, use (4)–(6) to obtain the conditional intensity function, index of dispersion, and counts spectrum, and then verify properties 1–4.

The difficulty one encounters in obtaining the counts properties from the interarrival time pdf's is, of course, that of inverting the Laplace transform in (4). Among the usual candidates: the Weibull, lognormal, and gamma distributions, we have been able to invert $m_f{}^*$ only for a special case of the gamma distribution, and in this case the conditional intensity function increases to the intensity, implying that property 1 is violated.

Rather than pursue indiscriminately a search for an interarrival time distribution with appropriate counts properties, we proceed instead by further restricting the form of the renewal process model. As discussed above, the Cox processes and the Poisson cluster processes are appealing in terms of physical interpretation and computational tractability of counts properties. Therefore we narrow our search for a model to the class of stationary point processes that are either simultaneously renewal processes and Poisson cluster processes or simultaneously renewal processes and Cox processes. It is then necessary to characterize these two intersections of classes of point processes.

Haberland [1975] showed that the only Poisson cluster processes that are also stationary renewal processes are the Poisson processes. Consequently, a Poisson cluster process cannot be found which satisfies both the counts and interval properties listed above; specifically, no Poisson cluster process satisfies properties 2 and 5.

Kingman [1964] characterized the class of Cox processes that are also renewal processes and showed this class to be strictly larger than the Poisson processes. Kingman showed that the intensity process $(\lambda(u))$ of a renewal Cox process is of the following form:

> ...$(\lambda(u))$ is equal to $\lambda > 0$ and 0 alternately on intervals whose lengths are independent random variables, the lengths of the intervals on which $\lambda(u) = \lambda$ having an exponential distribution and the lengths of the intervals on which $\lambda(u) = 0$ having an arbitrary distribution G.

In view of these results, we now seek a model that is a stationary point process, a renewal process, and a Cox process, simultaneously. By Kingman's characterization we are free to choose only the distribution G of the sojourn times of $(\lambda(u))$ in 0, which we do with two objectives: so that we have computational tractability and so that properties 1–4 of the data are satisfied. Fortunately, the natural choice for G, i.e., an exponential distribution (but with a different parameter from that for the state λ), attains both objectives, as we now proceed to discuss.

Note that an important consequence of the assumption that the distribution of sojourn times in 0 is exponential is that $(\lambda(u))$ is a Markov process. This follows from the characterization of Markow processes in the preceding section and Kingman's characterization of $(\lambda(u))$. We use the term RCM process to refer to a renewal Cox process for which the intensity process is a Markov process. In the remainder of this section we develop properties of RCM processes.

The following result is used to verify property 1 of the data. Let N be a Cox process with Markov intensity process $(\lambda(u))$. Suppose that $(\lambda(u))$ has state space $E = (\lambda_1, \cdots, \lambda_n)$ and transition function $(P_u(\ ,\ ), u \geq 0)$ and that a limit distribution π exists (see section 2). Suppose further that $\lambda(0)$ has distribution π, which implies that $(\lambda(u))$ is strictly stationary [see *Karlin and Taylor*, 1975] and hence that N is a stationary point process.

Proposition

Under the assumptions set forth in the preceding paragraph, the conditional intensity function of N is given by

$$m_f(t) = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \lambda_i P_t(\lambda_j, \lambda_i) \right) \left[(\pi(j)\lambda_j) \left(\sum_{k=1}^{n} \pi(k)\lambda_k \right)^{-1} \right] \tag{7}$$

The proof is given in the appendix.

In particular, if N is an RCM process, then the state space of $(\lambda(u))$ is $(0, \lambda)$ and the preceding proposition implies that

$$m_f(t) = \lambda P_t(\lambda, \lambda)$$

Let a_1 and a_2 be the parameters of the exponential sojourn distributions of $(\lambda(u))$ in 0 and λ, respectively. Then the transition function can be computed by methods described by *Cinlar* [1975], giving, in this case,

$$P_t(\lambda, \lambda) = a_1/(a_1 + a_2) + (a_2/(a_1 + a_2))e^{-(a_1+a_2)t} \tag{8}$$

Noting that the intensity m of an RCM process is given by $m = \lambda a_1/(a_1 + a_2)$, we have arrived at the following expression for the conditional intensity function of an RCM process:

$$m_f(t) = m + (\lambda a_2/(a_1 + a_2))e^{-(a_1+a_2)t} \tag{9}$$

Note, in particular, that the conditional intensity function decreases exponentially to the intensity, implying that property 1 holds for RCM processes.

The variance time curve $V(t)$ and counts spectrum $g_+(\omega)$ are obtained from (5) and (6). These results, along with the index of dispersion, $I(t) = V(t)/mt$, are presented below:

$$V(t) = mt + \frac{2m^2a_2}{a_1(a_1 + a_2)} \left[t - \frac{1}{(a_1 + a_2)} (1 - e^{-(a_1+a_2)t}) \right] \tag{10}$$

$$I(t) = 1 + \frac{2ma_2}{a_1(a_1 + a_2)} \left[1 - \frac{1}{t(a_1 + a_2)} (1 - e^{-(a_1+a_2)t}) \right] \tag{11}$$

$$g_+(\omega) = (m/\pi) \left(1 + \frac{2\lambda a_2}{\omega^2 + (a_1 + a_2)^2} \right) \tag{12}$$

These results imply that properties 2 and 3 hold for RCM processes. Thus we have shown that all the counts properties are satisfied. It remains only to show that the coefficient of variation of the interarrival times is greater than 1.

Kingman [1964] obtains a representation for the Laplace transform of the interarrival time distribution for renewal Cox processes. We adapt the result for RCM processes.

Proposition

The Laplace transform of the interarrival time distribution of an RCM process is given by

$$\Psi(s) = E[e^{-sU(i)}] = \frac{\lambda(s + a_1)}{s^2 + (a_1 + a_2 + \lambda)s + a_1\lambda} \tag{13}$$

Moments for the interarrival times can be obtained from the formula

$$(-1)^k \frac{d^k\Psi(0)}{ds^k} = E[U(i)^k]$$

giving for the RCM processes,

$$\text{var}\,(U(i)) = (1/m)^2 + 2a_2\lambda/(a_1\lambda)^2 \quad (14)$$

$$CV(U(i)) = m[(1/m)^2 + 2a_2\lambda/(a_1\lambda)^2]^{1/2} \quad (15)$$

Most importantly, we note that the coefficient of variation is greater than 1. Thus the final property that we require the point process model to satisfy has been verified.

Additional computational results for RCM processes are presented by *Smith* [1981]. Of particular importance are the zero probability function, $Z(t) = P(N(t) = 0)$, and the probability generating function, $\Phi(s, t) = E[s^{N(t)}]$. The zero probability function is of interest in applications pertaining to droughts, since it is directly related to the duration of dry periods. In the following section the probability generating function arises in obtaining distributional results for the 'marked' RCM process model of rainfall occurrences and associated storm depths. Calculations for $Z(t)$ and $\Phi(s, t)$ are based on methods presented by *Grandell* [1976]. The results are given below.

$$\Phi(s, t) = A_1(s)e^{-r_1(s)t} + A_2(s)e^{-r_2(s)t} \quad (16)$$

$$Z(t) = A_1(0)e^{-r_1(0)t} + A_2(0)e^{-r_2(0)t} \quad (17)$$

where

$$A_1(s) = 1 + \{[(a_1 + a_2)^2 - a_1\lambda(1 - s)] [(a_1 + a_2)^2((a_1 + a_2\lambda(1 - s))^2 - 4a_1\lambda(1 - s))^{1/2}]^{-1}\}$$

$$A_2(s) = 1 - A_1(s)$$

$$r_1(s) = (\tfrac{1}{2})[\lambda(1 - s) - [(\lambda(1 - s))^2 + 2(a_2 - a_1)\lambda(1 - s) + (a_1 + a_2)^2]^{1/2} + a_1 + a_2]$$

$$r_2(s) = -r_1(s) + a_1 + a_2 + \lambda(1 - s)$$

The form of the RCM process is of particular interest in modeling rainfall. During periods when the intensity process is in the state 0, there can be no rainfall events. Thus the 0 state in the intensity process can be interpreted as a 'rainfall-inhibiting' state. *Smith* [1981] presented evidence that associates historical periods of dry summer weather in the Potomac basin with the advection of cool, dry, stable air masses from Canada. These considerations strongly suggest an interpretation of the cluster properties of rainfall occurrences based on the alternation of wet and dry 'periods.' It should be noted that this interpretation of clustering is quite different from the interpretation that *Kavvas and Delleur* [1981] obtain from a Poisson cluster process, in which clusters of rainfall occurrences are associated with frontal passages. A major difference in the two modeling efforts is that Kavvas and Delleur consider rainfall occurrences over the entire year, while we restrict consideration to summer season rainfall occurrences.

4. MARKED POINT PROCESS MODELS

In this section we discuss extensions of the RCM process model of rainfall occurrences to models for which a storm depth $Y(i)$ is associated with a storm occurrence at the time $T(i)$. Processes of this type are referred to as marked point processes. Probabilistic properties of the process $(T(i), Y(i))$ were estimated for the Potomac rainfall data by *Smith* [1981]. These results indicate that a model of rainfall occurrences and storm depths for the Potomac basin should have the following properties (in addition to those previously listed for the process N of occurrence times).

1. The storm depths $(Y(i))$ are IID.
2. The occurrence process N and the storm depth process $(Y(i))$ are independent.

Given a specification of distributional properties of N, those of the process $(T(i), Y(i))$ are completely determined by specifying the distribution of the storm depths $Y(i)$. In applications, interest has focused on processes derived from the marked point process rather than on the process $(T(i), Y(i))$ itself. In this section we examine distributional properties of two processes derived from $(T(i), Y(i))$: the accumulated rainfall process and the process of storm occurrences with storm depths greater than a specified threshold.

The accumulated rainfall process and the process of storm occurrences greater than a threshold c can be expressed in terms of N and $(Y(i))$ as follows:

$$M(t) = \sum_{i=1}^{N(t)} Y(i)$$

$$N_c(t) = \sum_{i=1}^{N(t)} 1(Y(i) > c)$$

where

$$1(Y(i) > c) = 1 \qquad Y(i) > c$$
$$= 0 \qquad \text{otherwise}$$

To interpret, $M(t)$ is the total rainfall in the time interval $[0, t]$, while $N_c(t)$ is the number of rainfall events in $[0, t]$ for which the associated storm depth exceeds c.

The most important consequence of properties 1 and 2, above, is that M and N_c can be modeled as compound point processes. Let B, $B(1)$, $B(2)$, $\cdots$ be IID nonnegative random variables with distribution G which are independent of N. Then the process

$$S(t) = \sum_{i=1}^{N(t)} B(i)$$

is a G compound of N. The Laplace functional

$$L_S(f) = E\left[\exp\left(-\int_0^\infty f(s)\, dS\,(s)\right)\right]$$

for a G compound of a Cox process directed by Λ can be expressed as follows [*Kallenberg*, 1975]:

$$L_S(f) = E\left[\exp\left(-\int_0^\infty (1 - L_B(f(s)))\, d\Lambda\,(s)\right)\right],$$

where $L_B(u) = E[\exp\,(-uB)]$ is the Laplace transform of B.

Below we summarize distributional properties of the accumulated rainfall process $M(t)$ under conditions 1 and 2, above, and the assumption that N is an RCM process.

Proposition

Let $L_Y(s)$ be the Laplace transform of $Y(i)$, $u = E[Y(i)]$ and $\sigma^2 = \text{var}\,(Y(i))$. Then the following occur.

1. $E[e^{-sM(t)}] = \Phi(L_Y(s), t)$, where $\Phi(v, t)$ is the probability generating function of N, given in (16).
2. $E[M(t)] = umt$.
3. $\mathrm{var}\,(M(t)) = \sigma^2 mt + u^2 V(t)$, where $V(t)$ is given in (10).

The calculations appear in the work by *Smith* [1981].

Eagleson [1978] notes that accumulated rainfall in humid climates is approximately normally distributed for long time intervals. The following proposition shows that this is, in fact, the case for the process we have described.

Proposition

Let t_k, $k = 1, 2, \cdots$ be a sequence of positive real numbers increasing to infinity. Under the assumptions of the previous proposition,

$$\frac{M(t_k) - umt_k}{\mathrm{var}(M(t_k))^{1/2}} \xrightarrow{d} N(0, 1)$$

where $\xrightarrow{d}$ denotes convergence in distribution and $N(0, 1)$ denotes a standard normal random variable. The proof of this proposition is based on a result of *Karr* [1978].

The counts and interarrival time parameters for progressively thinned rainfall occurrences (i.e., N_c for $c = 0.03$ in., 0.05 in., 0.10 in., $\cdots$) were estimated by *Smith* [1981] for the Potomac basin rainfall data. It was determined that the thinned rainfall occurrences retained the counts and interarrival time properties presented in section 3. This implies that the form of the counts and interarrival time parameters of the thinned arrival process N_c should be the same as in (9), (11), (12), and (15), but that, roughly speaking, these parameters should be scaled to reflect the decreasing frequency of events. We now proceed to show that under properties 1 and 2, above, and the assumption that N is an RCM process, the properties of the thinned arrival process are consistent with the data analysis results.

A p thinning of a point process is a special form of compounding in which the random variables B, $B(1)$, $B(2)$, $\cdots$ can assume only the values 0 and 1 [see *Kallenberg*, 1975] and where $p = P(B = 1)$. The interpretation is that the p-thinning process results from independent trials in which each point of the original process is retained (in the same location) with probability p and deleted entirely with probability $1 - p$. Under assumptions 1 and 2, above, the process N_c of occurrences of storms with storm depth exceeding c is a p thinning of the original occurrence process N, with $p = P(Y(n) > c)$. In this case, $L_B(u) = 1 - p(1 - e^{-u})$. By substituting into the expression for the Laplace functional of a compound Cox process, we obtain the Laplace functional $L_{\tilde N}$ of a p-thinning $\tilde N$ of N:

$$L_{\tilde N}(f) = E\left[\exp\left(-\int_0^\infty p(1 - e^{-f(s)})\, d\Lambda\,(s)\right)\right]$$

$$= E\left[\exp\left(-\int_0^\infty (1 - e^{-f(s)})\, d\tilde\Lambda\,(s)\right)\right]$$

where $\tilde\Lambda(t) = p\Lambda(t)$. Thus the p thinning of a Cox process directed by Λ is itself a Cox process, directed by $p\Lambda$. In particular, a p thinning of an RCM process with intensity parameter λ is an RCM process with intensity parameter $p\lambda$. As a consequence, counts and interarrival time properties of a p thinning of an RCM process are immediately available (substitute $p\lambda$ for λ in (10) to (17)).

Since the thinned data sets exhibit the same qualitative properties (listed at the beginning of section 3) as the original data set, and in view of the discussion in the preceding paragraph, we see that the RCM model of rainfall occurrences is also entirely consistent with the characteristics of the thinned data.

5. A Bivariate RCM Process

In this section a bivariate RCM process model of rainfall occurrences at two sites is presented. The intensity process $(\lambda(u))$ provides the means through which site to site correlation is introduced into the bivariate RCM process model. An interpretation of the bivariate RCM process model described below is that rainfall at two sites is correlated because both sites are under the influence of the same 'climatological state.' We conclude this section with a discussion of extending this concept to a rainfall model that incorporates climatological observations.

We assume that the bivariate point process (N_a, N_b) satisfies the following.

1. N_a is an RCM process with intensity process $(\lambda(u))$.
2. N_b is an RCM process with intensity process $(\lambda(u))$.
3. N_a and N_b are conditionally independent, given $(\lambda(u))$.

The main result is the following.

Proposition

$\mathrm{Cov}\,(N_a(t), N_b(t)) = \mathrm{var}\,(N_a(t)) - mt$; $\mathrm{var}\,(N_a(t))$ (which equals $\mathrm{var}\,(N_b(t))$) is given in (10). The proof is given by *Smith* [1981].

In applications of Cox processes the intensity process is often explicitly related to a specific physical mechanism. For example, in Cox's study of stoppages of looms [*Cox*, 1955], the diameter of the yarn being fed into the loom is held to affect the rate of stoppages. The yarn occurs in homogeneous segments of random lengths that are 'tied' together. Thus the intensity process for loom stoppages is similar to ours: it remains constant for a period of time then shifts to a different level.

For the rainfall model this suggests the possibility of explicitly relating rainfall to climatological processes through a catalogue of climatological states. What we are suggesting is a Cox process model of rainfall occurrences for which the states of the intensity process correspond to observable climatological processes. A useful class of models to consider for this purpose is the Cox processes with Markovian intensities. Thus if there are k climatological states, the intensity process $(\lambda(u))$ is a Markov process with state space $(\lambda_1, \cdots, \lambda_k)$. A space-time rainfall model is being developed along these lines by the authors.

6. Summary and Conclusions

In this paper we have developed a point process model of summer season rainfall occurrences in the Potomac River basin. The model is an RCM process (renewal Cox process with Markovian intensity): a Cox process for which the intensity process $(\lambda(u))$ is a Markov process with two states, 0 and $\lambda > 0$. The principal steps in the development of the univariate point process model are the following:

1. The counts and interarrival time properties that we require the model to satisfy were obtained from a statistical analysis of Potomac rainfall data [*Smith*, 1981]. The requirements for the model are presented at the beginning of section 3.

2. From the interarrival time and counts properties we conclude that an appropriate model for the rainfall data cannot be found within the class of Poisson processes.

3. The interarrival time results for the rainfall data indicate that the model should be a renewal process. We are unable by direct specification of the interrenewal distribution to obtain the counts properties that are required because of computational difficulties.

4. The key step in model selection is characterizing the intersections of the renewal processes and Poisson cluster processes and renewal processes and Cox processes. The Poisson cluster processes are dropped from consideration due to the fact that the intersection of the Poisson cluster processes and renewal processes consists only of Poisson processes. The intersection of the renewal processes and Cox processes contains more than only Poisson processes. For reasons of computational tractability we restrict consideration to the renewal Cox processes with Markovian intensity process.

5. The most important computational result obtained in this paper is an expression for the conditional intensity function of a Cox process with Markovian intensity. This result enables us to compute the counts properties of an RCM process. We obtain the conditional intensity function, index of dispersion, and counts spectrum for RCM processes. The conditional intensity function decreases exponentially to the intensity, the index of dispersion is greater than 1, and the counts spectrum decreases with increasing frequency. The results are consistent with the counts properties outlined in section 3.

6. We show that the coefficient of variation of the interarrival times for an RCM process is greater than 1. This is the last of the counts and interarrival time properties that we required a point process model to satisfy. (The independence of the interarrival times follows directly from the fact that the process is a renewal process.)

7. In section 4 we extend the model to a marked RCM process model of rainfall occurrences and associated storm depths. The main assumptions are (1) the occurrence process N is independent of the storm depth process ($Y(i)$) and (2) the storm depths $Y(1)$, $Y(2)$, $\cdots$ are IID. We derive distributional properties of two processes derived from the marked point process model, as follows.

The accumulated rainfall process:

$$M(t) = \sum_{i=1}^{N(t)} Y(i)$$

The thinned rainfall occurrence process:

$$N_c(t) = \sum_{i=1}^{N(t)} 1(Y(i) > c)$$

8. A bivariate RCM process is described in section 5. The main assumption is that arrivals at one site are conditionally independent of arrivals at a second site, given the intensity process, but not independent of arrivals at the second site. Thus site to site correlation results from the two sites being under the influence of the same climatological state. We obtain the covariance function of the two counting processes. We conclude with a discussion of incorporating meteorological information into a Cox process model.

The form of the RCM process, in particular the 0 state in the intensity process, suggests that cluster properties of Potomac rainfall occurrences result from the alternation of wet and dry periods. *Namias* [1966] noted that during the drought of the 1960's,

> . . . over much of the Northeast the prevailing anomalous component of air flow was from the northwest. The Northeast area was characterized by frequent subsiding large-scale air motions. The frequent advection of cold air from Canada provided below normal temperatures as well as deficient precipitation.

A comparison of periods of low summer precipitation in the Potomac during the period 1950–1970 with synoptic conditions [*Smith*, 1981] supports the hypothesis that periods of deficit rainfall in the Potomac basin are most commonly associated with the invasion of cool, dry Canadian air masses. The RCM process provides a point process model of rainfall occurrences which not only is consistent with data analysis results for the Potomac River basin but also yields insight into the role of drought periods in the summer season rainfall process.

APPENDIX

Proposition (7) will be proven by a sequence of lemmas.

Lemma 1

$$P(N(t) = k|N(0) = 1) = \sum_{i=1}^{n} P(N(t) = k|\lambda(0) = \lambda_i) P(\lambda(0) = \lambda_i|N(0) = 1)$$

Proof

It will be recalled that the Palm function of N is defined by

$$P(N(t) = k|N(0) = 1) = \lim_{s \downarrow 0} P(N(t) - N(s) = k|N(s) \geq 1)$$

Now we compute $P(N(t) - N(s) = k|N(s) \geq 1)$:

$$\begin{aligned} P(N(t) - N(s) = k|N(s) \geq 1) &= \frac{P(N(t) - N(s) = k,\ N(s) \geq 1)}{P(N(s) \geq 1)} \\ &= \sum_{i=1}^{n} P(N(t) - N(s) = k,\ N(s) \geq 1,\ \lambda(s) = \lambda_i) \cdot [P(N(s) \geq 1)]^{-1} \\ &= \sum_{i=1}^{n} P(N(t) - N(s) = k|\lambda(s) = \lambda_i,\ N(s) \geq 1) \cdot P(\lambda(s) = \lambda_i,\ N(s) \geq 1)\ [P(N(s) \geq 1)]^{-1} \\ &= \sum_{i=1}^{n} P(N(t) - N(s) = k|\lambda(s) = \lambda_i) P(\lambda(s) = \lambda_i|N(s) \geq 1) \end{aligned}$$

Thus

$$\lim_{s \downarrow 0} P(N(t) - N(s) = k|N(s) \geq 1) = \lim_{s \downarrow 0} \sum_{i=1}^{n} P(N(t) - N(s) = k|\lambda(s) = \lambda_i) P(\lambda(s) = \lambda_i|N(s) \geq 1)$$

$$= \sum_{i=1}^{n} P(N(t) = k|\lambda(0) = \lambda_i)P(\lambda(0) = \lambda_i|N(0) = 1)$$

The step for which the Markov assumption on $(\lambda(u))$ is crucial is

$$P(N(t) - N(s) = k|\lambda(s) = \lambda_i, N(s) \geq 1)$$
$$= P(N(t) - N(s) = k|\lambda(s) = \lambda_i)$$

Corollary

$$E[N(t)|N(0) = 1] = \sum_{i=1}^{n} E[N(t)|\lambda(0) = \lambda_i] \cdot P(\lambda(0) = \lambda_i|N(0) = 1)$$

Lemma 2

$$P(\lambda(0) = \lambda_i|N(0) = 1) = \pi(i)\lambda_i\left[\sum_{j=1}^{n} \pi(j)\lambda_j\right]^{-1}$$

Proof

The process $(\lambda(u))$ is strictly stationary if and only if the initial distribution of $(\lambda(u))$ is the limiting distribution [*Cinlar*, 1975]. Therefore

$$P(\lambda(0) = \lambda_i|N(0) = 1) = \lim_{s \downarrow 0} P(\lambda(s) = \lambda_i|N(s) \geq 1)$$

$$= \lim_{s \downarrow 0} (P(N(s) \geq 1|\lambda(s) = \lambda_i)/s)P(\lambda(s) = \lambda_i)[P(N(s) \geq 1)/s]^{-1}$$

$$= \pi(i)\lambda_i\left[\sum_{k=1}^{n} \pi(k)\lambda_k\right]^{-1}$$

Lemma 3

$$E[N(t)|\lambda(0) = \lambda_i] = \int_0^t \sum_{j=1}^{n} \lambda_j P_u(\lambda_i, \lambda_j)\, du$$

Proof

$$E[N(t)|\lambda(0) = \lambda_i] = E[E[N(t)|\lambda(u), u \leq t]|\lambda(0) = \lambda_i]$$
$$= E\left[\int_0^t \lambda(u)\, du|\lambda(0) = \lambda_i\right]$$
$$= \int_0^t E[\lambda(u)|\lambda(0) = \lambda_i]\, du$$
$$= \int_0^t \sum_{j=1}^{n} \lambda_j P_u(\lambda_i, \lambda_j)\, du$$

This completes the proof of the proposition.

Acknowledgments. This research was supported by two matching funds grants. The first (Contract 14-34-0001-8089) was sponsored by the Interstate Commission of the Potomac River Basin (ICPRB), the State of Maryland, the Commonwealth of Virginia, and the Office of Water Research and Technology through the Water Resources Research Center of the University of Maryland. The second matching funds grant (Contract 14-34-0001-0407) was sponsored by the ICPRB, the State of Maryland, the Commonwealth of Virginia, and the Office of Water Research and Technology. The authors gratefully acknowledge the comments of Jared Cohon.

REFERENCES

Bartlett, M. S., The spectral analysis of point processes, *J. R. Stat. Soc. Sec. B*, *25*, 264–296, 1963.
Cinlar, E., *Introduction to Stochastic Processes*, Prentice-Hall, Englewood Cliffs, New Jersey, 1975.
Cox, D. R., Some statistical methods connected with series of events, *J. R. Stat. Soc. Sec. B*, *17*, 129–164, 1955.
Cox, D. R., and V. S. Isham, *Point Processes*, Chapman and Hall, London, 1980.
Cox, D. R., and P. A. W. Lewis, *The Statistical Analysis of Series of Events*, Methuen, London, 1966.
Eagleson, P. S., Climate, soil, and vegetation, 2, The distribution of annual precipitation derived from observed storm sequences, *Water Resour. Res.*, 14(5), 713–721, 1978.
Grandell, J., *Doubly Stochastic Poisson Processes*, Springer-Verlag, Berlin, 1976.
Gupta, V. K., and L. Duckstein, A stochastic analysis of extreme droughts, *Water Resour. Res.*, *11*(2), 221–228, 1975.
Haberland, E., Infinitely divisible stationary recurrent point processes, *Math. Nachr.*, *70*, 260–264, 1975.
Kallenberg, O., *Random Measures*, Akademie-Verlag, Berlin, 1975.
Karlin, S., and H. Taylor, *A First Course in Stochastic Processes*, Academic, New York, 1975.
Karr, A. F., Derived random measures, *Stochastic Processes Appl.*, 8, 159–169, 1978.
Kavvas, M. L., and J. Delleur, A stochastic cluster model of daily rainfall occurrences, *Water Resour. Res.*, *17*(4), 1151–1160, 1981.
Khinchin, A. Y., Sequence of chance events without after effects, *Theory Probabs. Its Appl. Engl. Transl.*, *1*, 1–15, 1956.
Kingman, J. F. C., On doubly stochastic Poisson processes, *Proc. Cambridge Philos. Soc.*, *60*, 923–932, 1964.
Kingman, J. F. C., Completely random measures, *Pac. J. Math.*, *21*, 59–78, 1967.
LeCam, L., A stochastic description of precipitation, *Proc. 4th Berkeley Symp.*, *3*, 165–186, 1961.
Lewis, P. A. W., A. M. Katcher, and A. H. Weis, SASE-IV: An improved program for the analysis of series of events, report, IBM, Yorktown Heights, N. Y., 1969.
Namias, J., Nature and possible causes of the northeastern drought during 1962–1965, *Mon. Weather Rev.*, *94*, 543–556, 1966.
Rudin, W., *Real and Complex Analysis*, 2nd ed., McGraw Hill, New York, 1974.
Serfozo, R., Conditional Poisson processes, *J. Appl. Prob.*, *9*, 288–302, 1972.
Smith, J. A., Point process models of rainfall, Ph.D. dissertation, The Johns Hopkins University, Baltimore, Md., 1981.
Snyder, D. L., *Random Point Processes*, Wiley-Interscience, New York, 1975.
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.

(Received April 16, 1982;
revised September 13, 1982;
accepted October 18, 1982.)