WATER RESOURCES RESEARCH, VOL. 21, NO. 1, PAGES 73–79, JANUARY 1985

Statistical Inference for Point Process Models of Rainfall

JAMES A. SMITH

Interstate Commission on the Potomac River Basin, Rockville, Maryland

ALAN F. KARR

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

In this paper we develop maximum likelihood procedures for parameter estimation and model selection that apply to a large class of point process models that have been used to model rainfall occurrences, including Cox processes, Neyman-Scott processes, and renewal processes. The statistical inference procedures are based on the stochastic intensity $\lambda(t) = \lim_{s\to 0, s>0} (1/s)E[N(t+s) - N(t)|N(u), u < t]$. The likelihood function of a point process is shown to have a simple expression in terms of the stochastic intensity. The main result of this paper is a recursive procedure for computing stochastic intensities; the procedure is applicable to a broad class of point process models, including renewal Cox process with Markovian intensity processes and an important class of Neyman-Scott processes. The model selection procedure we propose, which is based on likelihood ratios, allows direct comparison of two classes of point processes to determine which provides a better model for a given data set. The estimation and model selection procedures are applied to two data sets of simulated Cox process arrivals and a data set of daily rainfall occurrences in the Potomac River basin.

1. INTRODUCTION

The Poisson processes have played a predominant role in modeling rainfall occurrences [*Todorovic and Yevjevich*, 1969]. A particularly attractive feature of Poisson processes is availability of effective and well-understood procedures for parameter estimation and hypothesis testing. Application of the hypothesis testing procedures, however, has shown that for numerous rainfall occurrence data sets the "correct" model is not a Poisson process [*Kavvas and Delleur*, 1981; *Smith and Karr*, 1983]. As a consequence, a number of alternatives to the Poisson processes have been proposed for modeling rainfall occurrences. These models generally fall into three broad (and not disjoint) classes: Neyman-Scott cluster processes [*Kavvas and Delleur*, 1981; *Waymire and Gupta*, 1981; *Ramirez and Bras*, 1982], Cox processes [*Smith and Karr*, 1983], and renewal processes [*Grace and Eagleson*, 1966; *Rodhe and Grandell*, 1981]. In an assessment of precipitation research prepared by the Committee on Precipitation of the AGU Hydrology Section [*AGU*, 1984], it is noted that "model development has outpaced the development of inference procedures". As a consequence, development of parameter estimation and hypothesis testing procedures for stochastic rainfall models has become a "critical problem" of precipitation research.

In this paper we develop maximum likelihood procedures for parameter estimation and model selection that are applicable to the main point process models that have been used to model rainfall occurrences, including Neyman-Scott processes, Cox processes, and renewal processes. Of the three classes, only renewal processes have well-established estimation procedures. *Kavvas and Delleur* [1981] and *Ramirez and Bras* [1982] have used second-order moments of the counting process to estimate parameters of Neyman-Scott processes; both studies report that serious problems are encountered using this method. The model selection procedure developed in this paper differs from available methods which are designed to decide whether a specified model is the "correct" model for a given data set. The most commonly used model selection procedures have been those used to test whether the correct model for a given data set is a Poisson process. The procedure we describe, which is based on likelihood ratios, is designed to determine which of two competing classes of point processes provides a better model for a given data set. For example, a test can be formulated to determine whether a Neyman-Scott process or a renewal process with gamma-distributed interarrival times provides a better model.

Brémaud [1981] notes that in applications of point processes two schools can be distinguished. The first school "takes the moment point of view and aims at fitting a model to given moment functions estimated from collected data", while the second school "describes its models by means of a stochastic intensity which summarizes at a given instant the potential to generate an event in the near future, given some observations of the past including the complete record of all previous times of occurrences." The inference procedures developed in section 3 belong to the second school and are based on the stochastic intensity

$$\lambda(t) = \lim_{s\to 0, s>0} (1/s)E[N(t+s) - N(t)|N(u), u < t] \tag{1}$$

which formalizes the notion of stochastic intensity described above.

The inference procedures described in this paper were developed primarily for application to Cox processes and Neyman-Scott processes. These two classes are of special interest in modeling rainfall because their components have been related to specific physical mechanisms. *Kavvas and Delleur* [1981] have shown that the components of a Neyman-Scott process have a natural interpretation based on frontogenesis: cluster centers correspond to fronts, each of which has a random number of precipitation events associated with it. Cox processes can be interpreted as Poisson processes with randomly varying rates of occurrence. For modeling rainfall the random rate of occurrence has been interpreted as a "stochastic climatological process". In the Cox process model developed by *Smith and Karr* [1983] for summer season rainfall oc-

Paper number 4W1318.
0043-1397/85/004W-1318$05.00

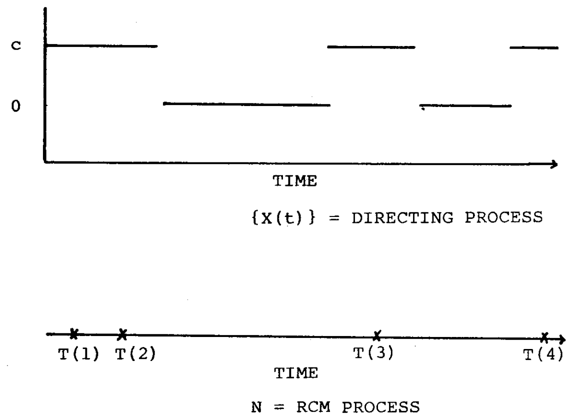

Fig. 1. Sample path for an RCM process and its associated directing process.

currences, the random rate process is related to the frequency and duration of the anticyclonic conditions.

Problems of statistical inference assume added importance when one deals with physically based models. Model selection in this context can potentially be used to distinguish between hypotheses concerning physical mechanisms. Model selection between Cox process and Neyman-Scott process models has been commonly used in plant and human geography to distinguish between "central location" and "random media" theories (see, for example, *Harvey* [1968]). The importance of parameter estimation for rainfall-runoff models has been demonstrated in numerous studies [e.g., *Sorooshian and Dracup*, 1980; *Kuczera*, 1982]. These studies demonstrate that one consequence of poor estimation procedures is that parameters of physically based models lose their physical significance.

Implementation of the estimation and model selection procedures is illustrated in section 4. The procedures are applied to two sets of simulated Cox process data and a data set of summer season rainfall occurrences in the Potomac River basin. The examples using simulated data illustrate the role of convergence rates of estimators for inference problems. Estimated parameters for the Cox process model developed by *Smith and Karr* [1983], using the historical data, are consistent with physical interpretations of model components.

2. DEFINITIONS AND NOTATION

A point process on the half line $[0, \infty)$ is a random process representing the times of occurrences of events. Denote by $T(n)$ the time of the nth event, with $T(0) = 0$. The interarrival times $U(i) = T(i+1) - T(i)$ represent times between events. The counting process $(N(t), t \geq 0)$ is defined by

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

We use the notation $N(A)$, i.e., N with an uppercase argument, for the number of events in a subset A of $[0, \infty)$.

Let $\{X(t)\}$ be a nonnegative stochastic process defined on $[0, \infty)$, and let

$$M(A) = \int_A X(u)\, du \qquad A \subset [0, \infty)$$

A point process N is a Cox process directed by X provided that

1. For disjoint sets $A_1, \ldots, A_k \subset [0, \infty)$, the random variables $N(A_1), \ldots, N(A_k)$ are conditionally independent given $\{X(t), t \geq 0\}$.
2. For every $A \subset [0, \infty)$ the conditional distribution of $N(A)$ given $\{X(t), t \geq 0\}$ is Poisson with parameter $M(A)$.

A renewal Cox process with Markovian intensity (RCM) process is a Cox process for which the directing process $\{X(t)\}$ is a two state Markov process with one of the states 0 and the other state $c > 0$ [see *Smith and Karr*, 1983]. The point process behaves as follows: when $\{X(t)\}$ is in state zero, no events can occur (for rainfall modeling this corresponds to a "dry" climatological state), and when $\{X(t)\}$ is in state c, events occur according to an ordinary Poisson process with rate c. An RCM process has three parameters: the rate c, and the exponential sojourn parameters a and b, representing the sojourn distributions of $\{X(t)\}$ in c and 0, respectively. Figure 1 illustrates a sample path of an RCM process and its associated directing process.

A Neyman-Scott process is constructed starting from an ordinary Poisson process $\tilde{N}$ of primary points ("cluster centers"), associated to each of which are a random number of randomly located secondary points ("cluster members") which are independent and identically distributed about their cluster center, and which constitute the events of the process. Furthermore, the distances from cluster members to their cluster center are independent of the number of members in the cluster. Figure 2 illustrates the structure of a Neyman-Scott process; detailed description can be found in the work by *Kavvas and Delleur* [1981].

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

$$G(t) = \int_0^t (1 - F(s))\, ds \Big/ \int_0^\infty (1 - F(s))\, ds \qquad (2)$$

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

The sample function density $p(n, s_1, \ldots, s_n)$ of a point process N with observations until time t is the joint probability density function of the number and times of events, i.e.,

$$P\{N(t) = n, T(1) \in A_1, \ldots, T(n) \in A_n\} = \int_{A_1} \cdots \int_{A_n} p(n, s_1, \ldots, s_n)\, ds_1 \ldots ds_n \qquad (3)$$

Heuristically, the sample function density $p(n, s_1, \ldots, s_n)$ can be viewed as the probability of n events occurring in the interval $(0, t]$ at the times $s_1, \ldots, s_n$. Typically, dependence of the

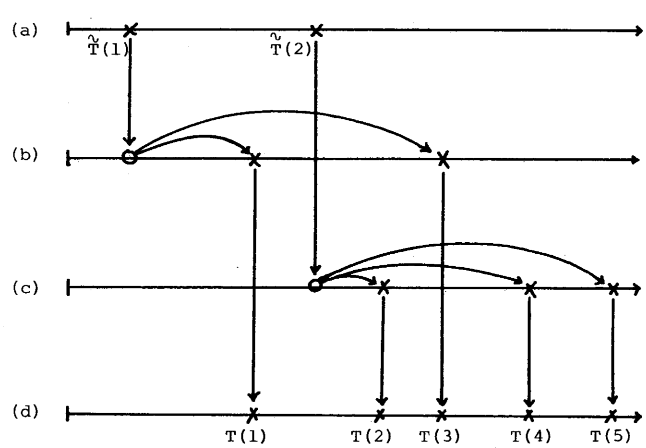

Fig. 2. Illustration of the structure of a Neyman-Scott cluster process. (*a*) $\tilde{N}$ is the cluster center process. (*b*) $\tilde{N}_1$ is the secondary process for cluster center $\tilde{T}(1)$. (*c*) $\tilde{N}_2$ is the secondary process for cluster center $\tilde{T}(2)$. (*d*) N is the Neyman-Scott cluster process.

sample function density on the model parameters θ is suppressed. The inference procedures developed in section 3 are based on the likelihood function

$$1(\theta|n, s_1, \ldots, s_n) = p(n, s_1, \ldots, s_n) \tag{4}$$

This notation emphasizes that in estimation and testing problems the data are treated as fixed and the parameters as variable. The log-likelihood function is the logarithm of the likelihood function and is denoted $L(\theta|n, s_1, \ldots, s_n)$.

The stochastic intensity $\{\lambda(t), t \geq 0\}$ of a point process N is defined as follows:

$$\lambda(t) = \lim_{s\to 0, s>0} (1/s)E[N(t+s) - N(t)|N(u), u < t] \tag{5}$$

Heuristically, the stochastic intensity describes the rate of occurrence of the point process at time t conditioned on the history of the point process strictly before time t.

3. Parameter Estimation and Model Selection

In this section we develop maximum likelihood procedures for parameter estimation and model selection that are applicable to a large class of point processes that have been used to model rainfall occurrences. The results of this section build upon two previous works. *Ogata* [1978] provides motivation for obtaining maximum likelihood procedures for point processes by showing that these procedures possess desirable optimality properties [see also *Konecny*, 1984]. Most importantly, maximum likelihood parameter estimators are consistent, i.e., they are asymptotically unbiased, and the variances converge to zero. This result guarantees that if sufficient data are available, parameters of physically based models can be estimated, so that their physical significance is preserved. Maximum likelihood estimators are also asymptotically efficient, implying that convergence rates are faster and variances smaller than for other estimation procedures. The importance of this result for applications is indicated in the following section where it is shown that for RCM model parameter estimators, convergence rates can be quite slow if the process is "nearly Poisson," a situation which arises in modeling the occurrence of "large" storms.

The second result we use is due to *Rubin* [1972], who showed that the sample function density of a point process has the following representation in terms of the stochastic intensity:

$$p(n, s_1, \ldots, s_n) = \exp\left\{-\int_0^t \lambda(u)\,du + \int_0^t \log(\lambda(u))\,dN(u)\right\} \tag{6}$$

The log-likelihood function for a point process N with parameters θ and observations over the interval $(0, t]$ is obtained directly from (6)

$$L(\theta|n, s_1, \ldots, s_n) = -\int_0^t \lambda(u)\,du + \int_0^t \log(\lambda(u))\,dN(u) \tag{7}$$

Dependence on θ is through the stochastic intensity λ. Parameter estimation based on (7) is standard and straightforward. The log-likelihood function is maximized (analytically or numerically) as a function of the unknown parameters. The parameter values that produce the maximum log-likelihood value are the maximum likelihood estimators.

The model selection procedure which we propose is not standard; it may, however, be more appropriate for many model selection problems faced in precipitation modeling than other available procedures (some of which are described in the work by *Cox and Lewis* [1978]). Our procedure is not used to decide whether a specific model is "correct," but simply to determine which of two given models is better for a given data set. The model that is selected may not be a "good" model. The model selection procedure in no way guarantees optimality; it is designed for situations in which the range of models can be restricted by considerations such as computational tractability or physical realism.

The procedure is based on the likelihood ratio

$$\Lambda(n, s_1, \ldots, s_n) = -\int_0^t (\lambda_0(u) - \lambda_1(u))\,du + \int_0^t \log(\lambda_0(u)/\lambda_1(u))\,dN(u) \tag{8}$$

where λ_0 and λ_1 are stochastic intensities of N under the null hypothesis H_0 and the alternative hypothesis H_1.

The procedure to choose between model H_0 and model H_1 is carried out in three steps: (1) compute maximum likelihood estimators for model H_0 and model H_1; (2) form the likelihood ratio using the maximum likelihood estimates obtained in step 1; (3) model H_0 is chosen if the log-likelihood ratio is greater than 0; otherwise model H_1 is chosen.

The procedures described above can be implemented only if stochastic intensities can be computed. For one of the three main classes of point processes that have been used to model rainfall occurrences, computation of the stochastic intensity is straightforward.

Example

Let N be a renewal process with interarrival time distribution F and interarrival time density f. Then the stochastic intensity is given by [*Brémaud*, 1981]

$$\lambda(t) = f(V(t))/(1 - F(V(t))) \tag{9}$$

where

$$V(t) = t - T(N(t)) \tag{10}$$

is the (left-continuous) backward recurrence time at t (the elapsed time at t since the most recent event). A Poisson process is a renewal process with exponentially distributed interarrival times; thus the stochastic intensity for a Poisson process is given by

$$\lambda(t) = m \tag{11}$$

where m is the parameter of the exponential distribution. Note that for Poisson processes the stochastic intensity is constant, implying that the rate of occurrence at a given time t is not affected by the history of arrivals prior to t.

The importance of intensity-based inference procedures is that they can be applied to a much wider class of models than renewal processes. *Ogata* [1978] and *Aalen* [1978] describe several classes of models for which the stochastic intensity has prescribed, tractable dependence on parameters. These models, however, have not been used to model rainfall occurrences and do not appear to provide any compelling physical basis for modeling rainfall.

In the remainder of this section we develop a procedure for computing stochastic intensities for several classes of models that have proven useful for modeling rainfall occurrences, including Neyman-Scott processes and RCM processes. In the following discussion two basic points to note are that (1) all of the point processes can be represented as Cox processes for which the directing process is a Markov process; and (2) the stochastic intensity of a Cox process directed by a Markov process can be computed "recursively."

We begin by showing that a large class of Neyman-Scott processes can be represented as Cox processes directed by a Markov process.

Proposition 1

Let N be a Neyman-Scott process of the following form:

1. The Poisson process $\tilde{N}$ of cluster centers has parameter c.
2. The distribution of cluster sizes is Poisson with parameter a.
3. The distances from each cluster center to its cluster members are IID exponentially distributed with parameter b.

Then the following hold:

1. N is a Cox process with directing process

$$X(t) = ab \int_0^t \exp(-b(t-u))\, d\tilde{N}(u) \tag{12}$$

2. The directing process $X(t)$ is a Markov process.

The first assertion follows from a more general result of *Bartlett* [1964]; the second assertion follows from direct calculation.

The following proposition shows that the stochastic intensity of a Cox process directed by $X(t)$ is simply the conditional expectation of $X(t)$ given the history of the point process up until time t (i.e., the times of arrivals up until t).

Proposition 2

The stochastic intensity of a Cox process N directed by $\{X(t)\}$ is given by

$$\lambda(t) = E[X(t)|N(u),\ u < t] \tag{13}$$

The proof is a straightforward conditioning argument [*Smith and Karr*, 1984].

Let N be a Cox process directed by a Markov process $\{X(t)\}$. Define the conditional moments of X given N as follows:

$$\hat{X}(t) = E[X(t)|N(u),\ u < t] \tag{14}$$

$$\hat{\Sigma}(t) = E[(\hat{X}(t) - X(t))^2|N(u),\ u < t] \tag{15}$$

$$\hat{\Phi}(t) = E[(\hat{X}(t) - X(t))^3|N(u),\ u < t] \tag{16}$$

Note from (13) that $\hat{X}(t) = \lambda(t)$; we will use $\hat{X}(t)$ and $\lambda(t)$ interchangeably, hereafter.

The generator A of the Markov process $\{X(t)\}$ is defined by

$$Af(x) = \lim_{s\to 0} (1/s)\{E[f(X(s))|X(0) = x] - f(x)\} \tag{17}$$

where f is a nonnegative function. We use the notation $A(X(t)^n)$ for $Af(X(t))$, where $f(x) = x^n$. Finally, we require $\hat{X}(t)$ to be strictly positive (which holds, for example, whenever $X(t)$ is strictly positive, but also more generally).

The following theorem provides a recursive procedure for computing stochastic intensities which can be applied to RCM processes (Corollary 1) and Neyman-Scott processes (Corollary 2).

Theorem

Under the above assumptions, $\hat{X}(t)$ and $\hat{\Sigma}(t)$ satisfy the stochastic differential equations

$$d\hat{X}(t) = E[A(X(t))|N(u),\ u < t]\, dt + \hat{\Sigma}(t)\hat{X}(t)^{-1}(dN(t) - \hat{X}(t)\, dt) \tag{18}$$

$$\begin{aligned} d\hat{\Sigma}(t) = \{&E[A(X(t)^2)|N(u),\ u < t] \\ &- 2\hat{X}(t)E[A(X(t))|N(u),\ u < t]\}\, dt \\ &+ \hat{\Phi}(t)\hat{X}(t)^{-1}(dN(t) - \hat{X}(t)\, dt) \\ &+ \hat{\Sigma}(t)^2\hat{X}(t)^{-2}\, dN(t) \end{aligned} \tag{19}$$

The proof is given in *Smith and Karr* [1984].

Corollary 1

The stochastic intensity of an RCM process with parameters given above satisfies the stochastic differential equation

$$d\hat{X}(t) = \{bc - (b + a)\hat{X}(t)\}\, dt + \{c - \hat{X}(t)\}(dN(t) - \hat{X}(t)\, dt) \tag{20}$$

Corollary 2

The stochastic intensity and state variance process for the Neyman-Scott process described above satisfy the stochastic differential equations

$$d\hat{X}(t) = \{abc - b\hat{X}(t)\}\, dt + \hat{\Sigma}(t)\hat{X}(t)^{-1}(dN(t) - \hat{X}(t)\, dt) \tag{21}$$

$$\begin{aligned} d\hat{\Sigma}(t) = \{&-2b\hat{\Sigma}(t) + a^2b^2c\}\, dt \\ &+ \hat{\Phi}(t)\hat{X}(t)^{-1}(dN(t) - \hat{X}(t)\, dt) \\ &+ \hat{\Sigma}(t)^2\hat{X}(t)^{-2}\, dN(t) \end{aligned} \tag{22}$$

The generator of the Neyman-Scott process can be computed from (12).

4. Implementation of Intensity-Based Inference Procedures

In this section the estimation and hypothesis testing procedures described (implicitly) in the previous section are illustrated in detail for RCM processes.

The basis of the inference procedures described in section 3 is recursive calculation of the stochastic intensity $\hat{X}(t)$. This is achieved for the RCM processes by representing (20) in finite difference form:

$$\Delta\hat{X}(t) = (bc - (b + a)\hat{X}(t))\Delta t + (c - \hat{X}(t))(\Delta N(t) - \hat{X}(t)\Delta t) \tag{23}$$

where

$$\Delta\hat{X}(t) = \hat{X}(t + \Delta t) - \hat{X}(t) \tag{24}$$

$$\Delta N(t) = N(t + \Delta t) - N(t) \tag{25}$$

To implement the procedure it is necessary to specify the time increment Δt and the initial state $\hat{X}(0)$; $\hat{X}(0)$ is obtained for the RCM processes by selecting the time origin to coincide with an event of the point process. In this case $\hat{X}(0)$ is equal to c, because events can only occur in the RCM process when the directing process is in the positive state, c. For precipitation data the natural time increments (i.e., the time increments at which precipitation is measured) are hourly and daily.

Maximum likelihood estimators for RCM processes cannot be obtained analytically from (23); they can be obtained numerically, however. A direct search procedure was used to obtain maximum likelihood estimators for two data sets of simulated RCM process arrivals. For the first data set the true parameters are $a = 0.12$, $b = 0.50$, and $c = 0.12$; for the second

TABLE 1. Maximum Likelihood Estimates for Simulated RCM Process Arrivals

Sample Size	a	b	c
100	0.04	1.10	0.08
200	0.06	1.05	0.08
500	0.16	1.00	0.11
900	0.14	0.95	0.11

True values are $a = 0.12$, $b = 0.50$, and $c = 0.12$.

TABLE 3. Likelihood Ratio Test Results for Simulated RCM Process Arrivals

Sample Size	Log-Likelihood Ratio	Result
100	−0.4	reject RCM model
200	−0.3	reject RCM model
500	−0.6	reject RCM model
900	−0.6	reject RCM model

Parameter values are $a = 0.12$, $b = 0.50$, and $c = 0.12$.

data set the true parameters are $a = 0.50$, $b = 0.12$, and $c = 0.80$. The estimators, for increasing sample size, are given in Tables 1 and 2.

The data sets of simulated RCM process arrivals were designed to represent two extreme cases: 1) a process very similar to a Poisson process and 2) a process very different from a Poisson process. For the first data set the coefficient of variation of the interarrival times is 1.05. Second-order moments are similar to those of rainfall occurrences for large storms [*Smith and Karr*, 1983]. Second-order moments of the second data set are not representative of rainfall occurrences in humid regions of the United States. The second data set represents a process which exhibits "clustering" of the following form. For long periods of time no events of the process can occur. These "dry periods" are interspersed with "wet periods" of short duration during which events occur at very high rate.

A striking feature of Table 1 is the slow rate of convergence for estimators of the nearly Poisson process. This suggests that if one desires to model occurrences of "large storms," it will not be possible to accurately estimate all of the parameters of the RCM model. It may be possible, however, to accurately estimate functions of the parameters that are of interest (the probability of no events in specified time intervals, for example). Table 2 illustrates that convergence rates are quite rapid for parameter estimators of the "clustered" process. These two examples indicate the range of convergence properties that can be expected for RCM models.

The model selection procedure described in the preceding section was applied to the two sets of simulated RCM process data to determine whether an RCM process or a Poisson process provides a better model. The log-likelihood ratio for this hypothesis testing problem is given by (see equation (8))

$$\Lambda(n, s_1, \dots, s_n) = -\int_0^t \hat{X}(u)\, du + \int_0^t \log(\hat{X}(u))\, dN(u) + N(t) - N(t)\log(N(t)/t) \tag{26}$$

where $\hat{X}(u)$ is the stochastic intensity of the RCM process.

Table 3 shows values of the log-likelihood ratio for the data set consisting of RCM process arrivals with parameters $a = 0.12$, $b = 0.50$, and $c = 0.12$. Surprisingly, the RCM model is rejected in favor of a Poisson process model for all sample sizes. This result points to the importance of convergence rates for parameter estimators in the model selection procedure. This point is further illustrated in Table 4, which presents values of the log-likelihood ratio for the data set of RCM process arrivals with parameters $a = 0.50$, $b = 0.12$, and $c = 0.80$. In this case the RCM process is accepted for all sample sizes. Note, also, that the margin by which the RCM process is accepted (for each sample size) is much larger than the margin by which the RCM process is rejected for the first data set.

The examples presented above indicate that sampling properties of maximum likelihood estimators and likelihood ratios can play a critical role in inference procedures for point process models of rainfall. These examples were designed to illustrate the range of convergence properties that could be expected in parameter estimation and model selection for rainfall models. The following example illustrates the performance of the estimation and model selection procedures in a more typical situation: modeling daily rainfall occurrences in a humid region.

The estimation and model selection procedures were applied to a data set of daily rainfall occurrences in the Potomac River basin. The data set consists of 10 years of daily rainfall occurrences (i.e., days for which 0.254 mm of rainfall or more were recorded) at Winchester, Virgina, for the summer season (July to September). The estimated parameters for this data set are $\hat{a} = 0.04$, $\hat{b} = 0.25$, and $\hat{c} = 0.31$.

These parameters yield the following interpretation of the RCM model; the directing process $\{X(t)\}$ alternates between wet periods, which last on the average 25 days ($\hat{a}^{-1}$), and dry periods, which last on the average 4 days ($\hat{b}^{-1}$); during wet periods rainfall events occur on the average every 3 or more days ($\hat{c}^{-1}$). *Smith and Karr* [1983] suggest that the dry state in an RCM model of summer season rainfall occurrences for the eastern United States is related to anticyclonic conditions, while the wet state is dominated by periods of convective activity. The estimated parameters for Winchester are not inconsistent with this interpretation. The estimators imply that the frequency of summer anticyclones in Virginia is approximately one per month while the mean duration is 4 days; these values are consistent with charts of tracks and frequencies of anticyclones presented by *Klein* [1957]. From

TABLE 2. Maximum Likelihood Estimates for Simulated RCM Process Arrivals

Sample Size	a	b	c
100	0.57	0.13	0.92
200	0.54	0.12	0.85
500	0.58	0.12	0.87
900	0.57	0.12	0.83

True values are $a = 0.50$, $b = 0.12$, and $c = 0.80$.

TABLE 4. Likelihood Ratio Test Results for Simulated RCM Process Arrivals

Sample Size	Log-Likelihood Ratio	Result
100	48.4	accept RCM model
200	95.0	accept RCM model
500	239.6	accept RCM model
900	436.2	accept RCM model

Parameter values are $a = 0.50$, $b = 0.12$, and $c = 0.80$.

these results, it can be concluded that the maximum likelihood procedure produces physically realistic parameter estimators using actual precipitation data; the stronger conclusion, that the estimators have precise physical interpretations, is worthy of further investigation.

The model selection procedure was applied to the Winchester data to determine whether an RCM model or a Poisson process model was more appropriate. The RCM model assumption was accepted. The log-likelihood ratio for 200 observations was 1.6. Comparing this result with Tables 3 and 4 it can be seen that the margin of acceptance is substantially larger than the margin by which the RCM assumption was rejected for the nearly Poisson data set and much smaller than the margin by which the RCM assumption was accepted for the second simulated data set.

The recursive procedure for Neyman-Scott processes requires simultaneous computation of the state variance process of (22). Note also that the equation for the state variance process includes the third conditional moment $\hat{\Phi}(t)$ of the state estimate. For implementation of the procedure we suggest that terms involving the third moment be ignored. This assumption is justified, for example, when errors for the state estimate are approximately normally distributed.

5. SUMMARY AND CONCLUSIONS

The main points of the paper are summarized below.

1. The inference procedures developed in this paper are based on the stochastic intensity

$$\lambda(t) = \lim_{s \to 0, s > 0} (1/s)E[N(t+s) - N(t)|N(u), u < t]$$

Equation (6) shows that the likelihood function for a point process has a simple expression in terms of the stochastic intensity. Thus maximum likelihood estimation can be applied whenever the stochastic intensity can be computed. Maximum likelihood procedures are particularly attractive for application to RCM and Neyman-Scott processes due to optimality properties of these estimators and due to the difficulties reported with moment-based estimation procedures for these classes. Consistency of maximum likelihood estimators guarantees that if sufficient data are available, the physical significance of parameters of conceptual models can be preserved.

2. A model selection procedure, based on the likelihood ratio presented in (8), is proposed in section 3. Traditional testing procedures that are used for model selection are designed to determine whether a specific model is the correct model or a good model. These procedures are largely restricted to Poisson and renewal processes. The procedure we propose is designed to allow direct comparison of two classes of models to determine which is better for a given data set. The only major restriction on application of the procedure is that stochastic intensities must be computable.

3. The main result of this paper is the theorem of section 3, which provides a recursive procedure for computing the stochastic intensity of a Cox process directed by a Markov process. This class of point processes contains RCM processes and a large class of Neyman-Scott processes. Proposition 1 shows that a Neyman-Scott process for which the cluster size distribution is Poisson and the distribution of (relative) locations of cluster members is exponential can be represented as a Cox process; furthermore, the directing process is a Markov process.

4. Application of the estimation and model selection procedures to simulated RCM process data illustrates the importance of sampling properties of maximum likelihood estimators and likelihood ratios. The first simulated data set represents a process that is nearly Poisson; for this data set convergence rates of parameter estimators are extremely slow. The second data set represents a process that is highly clustered; for it convergence rates are quite rapid.

5. The estimation and model selection procedures are also applied to a data set of daily rainfall occurrences in the Potomac River basin. Estimated parameters for this data set are consistent with the interpretation of RCM model components suggested by *Smith and Karr* [1983] based on the frequency and duration of anticyclones.

6. Application of intensity-based inference procedures has focused in this paper on stationary point process models. Generalization to nonstationary point processes is, in principle, straightforward. An important area in which generalization is computationally straightforward involves testing whether a nonstationary Poisson process provides a better model than an RCM or Neyman-Scott model.

7. The inference procedures developed in this paper cannot be generalized to point processes in the plane (or higher dimensions) or to space-time rainfall models such as those developed by *Gupta and Waymire* [1979]. Intensity-based inference procedures are intimately tied to ordering properties of the real line. These conclusions suggest that developing estimation and model selection procedures for space-time rainfall models will be difficult.

REFERENCES

Aalen, O., Nonparametric inference for a family of counting processes, *Ann. Stat.*, *6*(4), 701–726, 1978.

AGU Committee on Precipitation, A new interdisciplinary focus on precipitation research, *Eos Trans. AGU*, *65*(23), 377–380, 1984.

Bartlett, M. S., The spectral analysis of two-dimensional point processes, *Biometrika*, *51*(3), 299–311, 1964.

Brémaud, P., *Point Processes and Queues: Martingale Dynamics*, Springer-Verlag, New York, 1981.

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

Cox, D. R., and P. Lewis, *The Statistical Analysis of Series of Events*, Metheun, London, 1978.

Grace, R., and P. Eagleson, The synthesis of short-time increment rainfall sequences, *MIT Hydrodyn. Lab. Rep. 91*, MIT Press, Cambridge, Mass., 1966.

Gupta, V. K., and E. Waymire, A stochastic kinematic study of subsynoptic space-time rainfall, *Water Resour. Res.*, *15*(3), 637–644, 1979.

Harvey, D. W., Some methodological problems in the use of the Neyman Type A and negative binomial probability distributions in the analysis of spatial series, *Trans. Inst. Br. Geogr.*, *44*, 85–95, 1968.

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

Klein, W. H., Principal tracks and mean frequencies of cyclones and anticyclones in the northern hemisphere, *U.S. Weather Bur. Tech. Rep. 40*, Washington, D. C., 1957.

Konecny, F., Parameter estimation for point processes with partial observations: A filtering approach, *Syst. Control Lett.*, *4*, 218–226, 1984.

Kuczera, G., On the relationship between the reliability of parameter estimates and hydrologic time series data used in calibration, *Water Resour. Res.*, *18*(1), 146–154, 1982.

Ogata, Y., Asymptotic behaviour of the maximum likelihood estimators for the stationary point processes, *Ann. Inst. Stat. Math.*, *30*, 243–261, 1978.

Ramirez, J. A., and R. L. Bras, Optimal irrigation control using stochastic cluster point processes for rainfall modeling and forecasting, *MIT Ralph M. Parsons Lab. Rep. 275*, MIT Press, Cambridge, Mass., 1982.

Rodhe, H., and J. Grandell, Estimates of characteristic times for precipitation scavenging, *J. Atmos. Sci.*, *38*(2), 370–386, 1981.

Rubin, I., Regular point processes and their detection, *IEEE Trans. Inf. Theory, 18*, 547–557, 1972.

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, Inference for Markov-modulated Poisson processes, *ICPRB Tech. Rep.*, 1984.

Sorooshian, S., and J. Dracup, Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: correlated and heteroscedastic error cases, *Water Resour. Res., 16*(2) 430–442, 1980.

Todorovic, P., and V. Yevjevich, Stochastic process of precipitation, *Hydrol. Pap. 35*, Colo. State Univ., Fort Collins, 1969.

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.

A. F. Karr, Department of Mathematical Sciences, The Johns Hopkins University, Baltimore, MD 21218.

J. A. Smith, Interstate Commission on the Potomac River Basin, 6110 Executive Blvd., Suite 300, Rockville, MD 20852-3903.

(Received September 5, 1984;
revised October 11, 1984;
accepted October 24, 1984.)