COMMUN. STATIST. -SIMULA., 20(2&3), 777-809 (1991)

# MODELING AND SIMULATION OF A NONHOMOGENEOUS POISSON PROCESS HAVING CYCLIC BEHAVIOR

Sanghoon Lee

Department of Industrial Engineering
Kyung Won University
Seongnam, KOREA

James R. Wilson

School of Industrial Engineering
Purdue University
West Lafayette, Indiana 47907, USA

Melba M. Crawford

Department of Mechanical Engineering
The University of Texas at Austin
Austin, Texas 78712, USA

*Key Words and Phrases: Nonhomogeneous Poisson process; exponential-polynomial-trigonometric rate function; simulation; thinning.*

## ABSTRACT

In this paper we develop a unified approach to modeling and simulation of a nonhomogeneous Poisson process whose rate function exhibits cyclic behavior as well as a long-term evolutionary trend. The approach can be applied whether the oscillation frequency of the cyclic behavior is known or unknown. To model such a process, we use an exponential rate function whose exponent includes both a polynomial and a trigonometric component. Maximum likelihood estimates of the unknown continuous parameters of this function are obtained numerically, and the degree of the polynomial component is determined by a likelihood ratio test. If the oscillation frequency is unknown, then an initial estimate of this parameter is obtained via spectral analysis of the observed series of events; initial estimates of the remaining trigonometric (respectively, polynomial) parameters are computed from a standard maximum likelihood (respectively, moment-matching) procedure

for an exponential-trigonometric (respectively, exponential-polynomial) rate function. To simulate the fitted process by the method of thinning, we present (a) a procedure for constructing an optimal piecewise linear majorizing rate function; and (b) a "piecewise thinning" simulation procedure based on the inverse transform method for generating events from a piecewise linear rate function. These procedures are applied to the storm-arrival process observed at an off-shore drilling site.

## 1. INTRODUCTION

This paper describes techniques for identification, estimation, and simulation of a nonhomogeneous Poisson process whose rate function contains a cyclic component as well as a long-term evolutionary trend. These techniques can be applied whether the oscillation frequency of the cyclic rate component is known or unknown. The original motivation for this work arose in a simulation study (Hoffman, 1982; Hoffman, Crawford, and Wilson, 1983; Lee 1985) that was designed to evaluate the effects of sea conditions and supply-ship availability on petroleum exploration operations at an off-shore drilling site. For many off-shore exploration activities, work efficiency depends on the response of the drilling vessel to prevailing environmental conditions. In particular, stormy weather at sea causes an increase in the heave of the drilling vessel as a function of wave height; this in turn causes a corresponding reduction in the amount of work completed per unit of time. Sea conditions also affect (a) the movement of supply ships between the port and the site, and (b) the off-loading of supplies from the supply ships to the drilling vessel.

A simulation model was developed to estimate the time required to complete each exploration activity as well as the total time to complete the entire drilling project. It was essential to provide not only an adequate representation of the wave process at the drilling site but also an efficient method for simulating realizations from this representation. A simulated realization of the wave process was obtained by superimposing (a) characteristics of storms that occur over time according to a strongly seasonal arrival pattern, and (b) the characteristics of calm waves that are described by a stationary time series. Crawford (1985) presented conditional probability models for describing the duration of storms and the height of the associated waves, and she

developed linear models of sea-state variables using Box-Jenkins time series analysis techniques.

The arrival pattern for storms (and hence for waves that are high enough to affect operational efficiency) can be modeled as a point process. Typically, the observed waves exhibit seasonal variations corresponding to the meteorological characteristics at a site. Because the rate of occurrence of high waves varies between seasons and from year to year, successive high-wave events (storms) are stochastically interdependent. Ashkar and Rousselle (1983) and Smith and Karr (1983) discussed probabilistic models for time-varying weather-related events. We believe that a nonhomogeneous Poisson process (NHPP) with an appropriate continuous rate function is the most plausible, general type of time-varying process for modeling high-wave events. Cox and Lewis (1966) and Lewis and Shedler (1976b) pointed out that the continuous rate function for such a process can be approximated to arbitrary accuracy with an exponential-polynomial function (EPF). For an NHPP whose behavior is locally cyclic with a long-term evolutionary trend, we seek in this paper to model the instantaneous arrival rate using an exponential function whose exponent is the sum of polynomial and trigonometric components—that is, an exponential-polynomial-trigonometric function (EPTF). Previously Lewis (1970) used spectral methods to analyze an NHPP with an exponential-trigonometric rate function. To model the arrival of patients at an intensive-care unit, Lewis (1972) used an NHPP with an EPTF-type rate function that included a quadratic trend.

Methods for simulating an NHPP have been developed by Lewis and Shedler (1976a, 1976b, 1979), Kaminsky and Rumpf (1977), and Klein and Roberts (1984). The exact simulation procedures are generally based on the inverse transform method or on the method of thinning. For joint application with standard variance reduction techniques, the inverse transform method is preferred because this method facilitates (a) synchronized use of random-number inputs across different simulation experiments, and (b) arrangement of a monotonic dependence of the simulation outputs on the random-number inputs (Bratley, Fox, and Schrage, 1987). However, the inverse transform method is limited to special forms of the rate function (such as a piecewise linear rate) for which the distribution function of the next interevent time

can be inverted (Klein and Roberts, 1984). For an arbitrary rate function with a tractable upper bound, the thinning technique is simple to apply and can be relatively efficient (Lewis and Shedler, 1979). Thus a preliminary step in our procedure for simulating an NHPP with a fitted EPTF-type rate function is to approximate the fitted rate function as closely as possible using a piecewise linear majorizing function—that is, a piecewise linear function which provides a tight upper bound for the fitted rate function. At each event epoch of the target process, a series of events is generated from the majorizing function by the inverse transform method; these events are then screened by the thinning procedure so that the next interevent time for the target process is finalized when the first acceptable event is generated for the majorizing process. With this approach, the computation time can be greatly reduced in simulating a process with a complicated cyclic rate function such as an EPTF.

This paper is organized as follows. In Section 2 we discuss our methods for (a) identifying the degree of the polynomial component of an EPTF-type rate function, and (b) estimating the parameters of this rate function. Section 3 describes algorithms for (a) approximating the fitted rate function with an optimal piecewise linear majorizing rate function, and (b) simulating the fitted NHPP by thinning a realization of the majorizing process. In Section 4 we apply these modeling and simulation procedures to oceanic high-wave events that were observed at an off-shore drilling site in the Arctic Sea. Finally in Section 5 we present the conclusions of this research.

## 2. IDENTIFICATION AND ESTIMATION OF NHPPs

An NHPP $\{N(t) : t \geq 0\}$ is a generalization of a Poisson process in which the instantaneous arrival rate $\lambda(t)$ at time $t$ is a nonnegative integrable function of time. The mean value function (or the integrated rate function) of the NHPP is defined by

$$\mu(t) \equiv E[N(t)] = \int_0^t \lambda(z)\,dz \quad \text{for all} \quad t \geq 0.$$

An NHPP with a continuous mean value function can be transformed into a homogeneous Poisson process. For $s \geq 0$, $\mu^{-1}(s)$ is taken to be the smallest

value of $t$ satisfying the condition $\mu(t) \geq s$. The process $\{M(s) : s \geq 0\}$ defined by $M(s) = N[\mu^{-1}(s)]$ is an NHPP with mean value function

$$E[M(s)] = E\{N[\mu^{-1}(s)]\} = \mu[\mu^{-1}(s)] = s \quad \text{for all} \quad s \geq 0;$$

thus $\{M(s)\}$ is a homogeneous Poisson process with unit rate (Parzen, 1962).

In this study, an NHPP displaying cyclic behavior is assumed to have an EPTF-type rate function. An EPTF of degree $m$ has the form

$$\lambda(t) = \exp\{h_\Theta(m,t)\} \quad \text{with} \quad h_\Theta(m,t) = \sum_{i=0}^{m} \alpha_i t^i + \gamma \sin(\omega t + \phi), \qquad (1)$$

where: $\Theta = [\alpha_0, \alpha_1, \ldots, \alpha_m, \gamma, \phi, \omega]$ is the vector of unknown parameters; the first term in $h_\Theta(m,t)$ is an ordinary polynomial function representing the general trend over time; and the second term in $h_\Theta(m,t)$ is a periodic function representing a cyclic effect exhibited by the process. Under the assumption of an NHPP with an EPTF-type rate function, the events of the point process are distributed according to a complicated joint probability density function. Statistical inference is generally difficult for such a stochastic model. For a simple EPF involving a polynomial of degree one or two, maximum likelihood parameter estimates can be derived analytically (Cox, 1972; Lewis, 1972). However, for an exponential rate function whose exponent includes a polynomial of higher degree or a trigonometric component, numerical methods must be used to obtain maximum likelihood estimates of the associated parameters.

It should also be noted that the analysis of an NHPP with a rate function of the form (1) is substantially more difficult when the oscillation frequency $\omega$ (expressed in radians per unit time) is unknown. Although $\omega$ is often known from prior information about the mechanism generating the events of interest, there is a large class of simulation applications for which such prior information is unavailable so that $\omega$ must be estimated from sample data. To develop a completely general technique for modeling and simulation of an NHPP with an EPTF-type rate function, we assume that the oscillation frequency is unknown and must be estimated along with all of the other parameters of the rate function. If $\omega$ is known, then we simply drop the

last component of $\Theta$ before applying the parameter estimation technique described below.

Consider a sequence of $n$ events occurring at the epochs $t_1 < t_2 < \cdots < t_n$ in a fixed time interval $(0, S]$ according to an NHPP with a rate function of the form (1). Then the log-likelihood function of $\Theta$, given $N(S) = n$ and $\mathbf{t} = (t_1, t_2, \ldots, t_n)$, is

$$\mathcal{L}(\Theta|n, \mathbf{t}) = \sum_{i=0}^{m} \alpha_i T_i + \gamma \sum_{j=1}^{n} \sin(\omega t_j + \phi) - \int_0^S \exp\{h_\Theta(m, z)\}\, dz, \qquad (2)$$

where $T_i = \sum_{j=1}^{n} t_j^i$ for $i = 0, 1, \ldots, m$; see Cox and Lewis (1966). Strictly speaking, we observe that the degree $m$ of the polynomial component of $h_\Theta(m, t)$ is also an unknown parameter which could in principle be estimated along with $\Theta$ from the given sequence of events by the method of maximum likelihood. However, since $m$ is constrained to be a nonnegative integer, we cannot estimate $m$ by solving the usual likelihood equations that are applicable to continuous parameters; moreover the usual regularity conditions ensuring the asymptotic efficiency of maximum likelihood estimators do not apply to the estimation of $m$.

In view of the fundamental problems inherent in maximum likelihood estimation of the degree $m$ of the polynomial rate component, we have chosen to condition the estimation of $\Theta$ on a fixed value of $m$ and then to determine the appropriate value of $m$ by a likelihood ratio test to be described later (see equation (7) and the accompanying discussion given below). Thus for a given value of $m$ where $m \geq 0$, we obtain $m+4$ likelihood equations involving the continuous parameter vector $\Theta$

$$\begin{aligned}
\frac{\partial \mathcal{L}(\Theta|n, \mathbf{t})}{\partial \alpha_i} &= T_i - \int_0^S z^i \exp\{h_\Theta(m, z)\}\, dz = 0, \quad i = 0, 1, \ldots, m, \\
\frac{\partial \mathcal{L}(\Theta|n, \mathbf{t})}{\partial \omega} &= \sum_{j=1}^{n} t_j \cos(\omega t_j + \phi) - \int_0^S z \cdot \cos(\omega z + \phi) \exp\{h_\Theta(m, z)\}\, dz = 0, \\
\frac{\partial \mathcal{L}(\Theta|n, \mathbf{t})}{\partial \gamma} &= \sum_{j=1}^{n} \sin(\omega t_j + \phi) - \int_0^S \sin(\omega z + \phi) \exp\{h_\Theta(m, z)\}\, dz = 0, \\
\frac{\partial \mathcal{L}(\Theta|n, \mathbf{t})}{\partial \phi} &= \sum_{j=1}^{n} \cos(\omega t_j + \phi) - \int_0^S \cos(\omega z + \phi) \exp\{h_\Theta(m, z)\}\, dz = 0.
\end{aligned} \qquad (3)$$

The solution to this system of nonlinear equations can be obtained numerically, yielding the maximum likelihood estimates of the parameters. Unfortunately, general numerical techniques such as the Newton-Raphson method have proven to be unstable when they are applied to (3) outside of a fairly small neighborhood of the optimal solution.

If the oscillation frequency $\omega$ is unknown, then it is important to choose the initial value of $\omega$ sufficiently close to the true maximum likelihood estimate because the log-likelihood function (2) has multiple local maxima due to the trigonometric rate component. In this situation, an initial estimate of $\omega$ can be obtained from a preliminary spectral analysis of the observed series of events; see pp. 361–363 of Lewis (1970). The periodogram of a point process having cyclic behavior should display peaks in the vicinity of the corresponding frequency, even when the process also possesses a long-term evolutionary trend. Let $\omega_0$ denote either the known frequency or an initial estimate of the unknown frequency inferred from the periodogram of the observed series of event times $\{t_j : j = 1, 2, \ldots, n\}$. We obtain the corresponding initial estimates of the amplitude $\gamma$ and the phase $\phi$ as follows. If we assume that the long-term evolutionary trend is nearly constant over the observation interval and that $\omega_0$ is the actual frequency of the cyclic component, then the rate function for the process has the exponential-trigonometric form

$$\lambda_0(t) = \exp\{\alpha + \gamma \sin(\omega_0 t + \phi)\} \quad \text{for all} \quad t \in (0, S]. \tag{4}$$

Under the simplifying assumption (4), the initial estimates $\gamma_0$ and $\phi_0$ (for $\gamma$ and $\phi$ respectively) are obtained from

$$\left.\begin{aligned} \phi_0 &= \tan^{-1}\left[\frac{A(\omega_0)}{B(\omega_0)}\right] \\ \gamma_0 \ &\text{is the solution of } \frac{1}{n_0}\sqrt{A^2(\omega_0) + B^2(\omega_0)} = \frac{I_1(\gamma_0)}{I_0(\gamma_0)} \end{aligned}\right\} \tag{5}$$

where: $n_0$ is the number of events in the time interval

$$\left(0, \left\lfloor \frac{\omega_0 S}{2\pi} \right\rfloor \frac{2\pi}{\omega_0}\right]$$

and $\lfloor z \rfloor$ represents the greatest integer $\leq z$ for all real $z$;

$$A(\omega_0) = \sum_{j=1}^{n_0} \cos(\omega_0 t_j); \qquad B(\omega_0) = \sum_{j=1}^{n_0} \sin(\omega_0 t_j);$$

and $I_u(\gamma_0)$ denotes a modified Bessel function of the first kind of order $u$ for $u = 0, 1$. (Note that the time interval corresponding to the event count $n_0$ is generally taken to be somewhat shorter than the original observation interval $(0, S]$ to ensure that $\omega_0$ is one of the frequencies included in the spectral computations.) For most cases in which the evolutionary trend changes slowly over time, the estimates $\gamma_0$ and $\phi_0$ provided by (5) will be good initial values of the corresponding trigonometric parameters. Equations (46) and (48) of Lewis (1970) provide the basis for this approach to the computation of $\gamma_0$ and $\phi_0$.

To determine initial values for the coefficients $\{\alpha_k : k = 0, 1, \ldots, m\}$ of the polynomial rate component, we use a variant of MacLean's (1974) procedure for estimating the parameters of an exponential-polynomial rate function by moment matching. Suppose that $C(m,t) = \sum_{k=0}^{m} c_k t^k$ is an ordinary polynomial function of degree $m$ whose first $m + 1$ moments over the interval $(0, S]$ match those of $\exp\{h_\Theta(m,t)\}$. Then by the first equation of (3), we have

$$T_i = \int_0^S z^i \exp\{h_\Theta(m,z)\}\, dz = \int_0^S z^i C(m,z)\, dz = \sum_{k=0}^{m} \frac{c_k S^{i+k+1}}{i+k+1}$$

for $i = 0, 1, \ldots, m$. The values of the $\{c_k\}$ can be obtained from this linear equation system by matrix inversion. Next, $h_\Theta(m,t)$ can be approximated by matching its first $m + 1$ moments over the interval $(0, S]$ with the corresponding moments of $\log[C(m,t)]$. Thus for $i = 0, 1, \ldots, m$, we have

$$\int_0^S z^i \log\left(\sum_{k=0}^{m} c_k z^k\right) dz - \gamma \mathcal{M}_{\sin}(i, S; \omega, \phi) = \sum_{k=0}^{m} \frac{\alpha_k S^{i+k+1}}{i+k+1} \qquad (6)$$

where $\mathcal{M}_{\sin}(i, S; \omega, \phi) \equiv \int_0^S z^i \sin(\omega z + \phi)\, dz$, the $i$th moment of $\sin(\omega t + \phi)$ over the interval $(0, S]$. Using the initial values of $\gamma$, $\omega$, and $\phi$ based on (4) and (5), we can evaluate the second term on the left-hand side of equation (6) from the computational formulas for $\mathcal{M}_{\sin}(i, S; \omega, \phi)$ given in

Appendix A; and by using numerical integration, we can also evaluate the first term on the left-hand side of (6). A single matrix inversion for equation system (6) yields initial estimates of the $\{\alpha_k\}$. This elaborate procedure for assembling an initial estimate of the complete parameter vector $\Theta$ is designed to ensure that the $(m+4)$-dimensional Newton-Raphson scheme for solving the full likelihood equation system (3) will start reasonably close to the true maximum of the log-likelihood function (2).

Corresponding to each trial value of the degree $m$ for the fitted EPTF-type rate function, the procedure described above is used to obtain an initial estimate of the associated parameter vector $\Theta_m$; then the log-likelihood function $\mathcal{L}_m(\Theta_m|n,\mathbf{t})$ is optimized by the Newton-Raphson procedure to yield the maximum likelihood estimator $\tilde{\Theta}_m$. (Here the subscript "$m$" is used to emphasize the dependence of the subscripted quantities on the degree of the fitted EPTF.) To determine the appropriate value of $m$, we use a maximum likelihood ratio test (Cox and Hinkley, 1974) as follows. Under the null hypothesis that $m$ is the true degree of the underlying EPTF-type rate function, the difference

$$2\left[\mathcal{L}_{m+1}(\tilde{\Theta}_{m+1}|n,\mathbf{t}) - \mathcal{L}_m(\tilde{\Theta}_m|n,\mathbf{t})\right] \tag{7}$$

is asymptotically chi-square with 1 degree of freedom as $n \to \infty$. Successive differences of this form are evaluated, and the smallest value of $m$ yielding a nonsignificant difference (7) is taken as the degree of the fitted EPTF. The corresponding $\tilde{\Theta}_m$ is taken as the final estimator of the parameters of the underlying NHPP.

In principle, the adequacy of the fitted NHPP could be checked by testing the uniformity of the observed point process after it has been "detrended" using the final estimate of the mean value function. Consider the observed series of event epochs $t_1 < t_2 < \cdots < t_n$ for the original process in the interval $(0, S]$. Let $\tilde{\mu}(t)$ denote the mean value function of the fitted NHPP corresponding to the final estimate $\tilde{\Theta}_m$. If the original process $\{N(t) : t \in (0, S]\}$ is in fact an NHPP whose mean value function $\mu(t)$ has been accurately estimated by $\tilde{\mu}(t)$, then the detrended process $\{\tilde{M}(s) \equiv N[\tilde{\mu}^{-1}(s)] : s \in (0, \tilde{\mu}(S)]\}$ with the event epochs $\tilde{s}_1 = \tilde{\mu}(t_1) < \tilde{s}_2 = \tilde{\mu}(t_2) < \cdots < \tilde{s}_n = \tilde{\mu}(t_n)$

should closely approximate a homogeneous Poisson process with unit rate. Given that $n$ events have occurred in the time interval $(0, S]$ for the original process, we seek to use certain nonparametric goodness-of-fit statistics to test the hypothesis that the detrended event epochs $\{\tilde{s}_i : i = 1, \ldots, n\}$ are uniformly distributed on the interval $(0, \tilde{\mu}(S)]$; see Cox and Lewis (1966). In particular, we seek to use the Kolmogorov-Smirnov statistic $D_{\mathrm{KS}}$ and the Anderson-Darling statistic $D_{\mathrm{AD}}$ to gauge the conditional uniformity of the detrended event epochs. Unfortunately, we do not currently have adequate approximations to the null distributions of these test statistics when they are applied to our procedure for fitting an NHPP with an EPTF-type rate function; this issue is the subject of ongoing research. Consequently, we can only use these goodness-of-fit statistics to provide a general indication of the adequacy of the fitted NHPP when they are considered in conjunction with other statistical and graphical evidence of the quality of the fit.

Diagnostic checking of the adequacy of the fitted NHPP will be based primarily on a careful examination of the detrended interevent times $\{\tilde{X}_i \equiv \tilde{s}_i - \tilde{s}_{i-1} : i = 1, \ldots, n;\ \tilde{s}_0 \equiv 0\}$. If the original process $\{N(t) : t \in (0, S]\}$ is in fact an NHPP whose mean value function $\mu(t)$ has been accurately estimated by $\tilde{\mu}(t)$, then the detrended interevent times $\{\tilde{X}_i\}$ should closely resemble a random sample from an exponential distribution with mean $1/\tilde{\lambda} = 1$. Thus we examine some low-order sample moments of the detrended interevent times—in particular, the coefficient of variation $\hat{V}$, the skewness $\sqrt{\hat{\beta}_1}$, and the kurtosis $\hat{\beta}_2$ of the $\{\tilde{X}_i\}$. Moreover, we apply the following randomness tests to the detrended interevent times: (a) the von Neumann ratio $R_{\mathrm{VN}}$ (von Neumann, 1941), whose critical values are calculated using the approximation of Bingham and Nelson (1981); and (b) $\mathcal{Z}$, the standardized Fisher $z$-transformation of the lag-one sample correlation (Stuart and Ord, 1987). In addition to these summary statistics for the $\{\tilde{X}_i\}$, we also obtain a visual assessment of the linearity of a version of the corresponding exponential probability plot (Hahn and Shapiro, 1967, pp. 292–294). If $\{\tilde{X}_i\}$ is a random sample from an exponential distribution with mean $1/\tilde{\lambda}$ and if $\tilde{X}_{(k)}$ is the $k$th order statistic of the sample, then

$$E\left[\tilde{X}_{(k)}\right] = \frac{1}{\tilde{\lambda}} \sum_{i=0}^{k-1} \frac{1}{n-i} \quad \text{for } k = 1, \ldots, n$$

(Feller, 1971; pp. 19–20); and in this case a plot of $\tilde{X}_{(k)}$ versus $\sum_{i=0}^{k} 1/(n-i)$ should be nearly linear. As a final visual check on the adequacy of the fitted NHPP, we will also compare the plot of the fitted mean value function $\tilde{\mu}(t)$ against the plot of the cumulative number of events $N(t)$ observed up to time $t$ for all $t \in (0, S]$.

## 3. SIMULATION OF NHPPs VIA PIECEWISE THINNING

Simulation of an NHPP with an arbitrary rate function $\lambda(t)$ in the interval $(0, S]$ can be accomplished by generating an alternative NHPP with a majorizing rate function $\lambda^*(t)$; some of the generated points are then rejected ("thinned") to yield a realization of the original NHPP with rate function $\lambda(t)$. The following theorem is the basis for this procedure.

**Theorem (Lewis and Shedler, 1979).** Consider an NHPP with rate function $\lambda^*(t)$ for $t \geq 0$. Suppose that $\tau_1, \tau_2, \ldots, \tau_{n^*}$ are the event epochs of this process in the interval $(0, S]$ and that $0 \leq \lambda(t) \leq \lambda^*(t)$ for all $t \in (0, S]$. If the $i$th event epoch $\tau_i$ is independently deleted with probability $1 - \lambda(\tau_i)/\lambda^*(\tau_i)$ for $i = 1, 2, \ldots, n^*$, then the remaining event epochs constitute an NHPP with rate function $\lambda(t)$ in the interval $(0, S]$.

At the $i$th event epoch $\tau_i$, thinning can be performed by sampling a random number $U_i$ and rejecting $\tau_i$ if $U_i > \lambda(\tau_i)/\lambda^*(\tau_i)$. If successive events are generated and thinned in the usual next-event-scheduling approach of discrete simulation, then it is desirable to select $\lambda^*(t)$ as close to $\lambda(t)$ as possible in order to minimize the number of events that are rejected over the time horizon of the simulation.

There are many ways to determine the majorizing rate function $\lambda^*(t)$ to use with the method of thinning. The simplest approach is to use a constant or piecewise constant function $\lambda^*(t)$, but this can be extremely inefficient for simulating an NHPP with a rapidly fluctuating rate function $\lambda(t)$, such as an EPTF that displays high-amplitude oscillations over a long time interval. To simulate an NHPP with an EPTF-type rate function, Lewis and Shedler (1979) suggested the simple majorizing function $\exp(\sum_{i=0}^{m} \alpha_i t^i + \gamma)$; unfor-

tunately this approach is also inefficient when the amplitude $\gamma$ is large. On the other hand, a majorizing process with a complicated rate function may be difficult to simulate. Perhaps the most suitable form for a majorizing rate function on the interval $(0, S]$ is the piecewise linear function

$$\lambda^*(t) = \sum_{j=1}^{r} (a_j t + b_j)\, I_{(L_{j-1}, L_j]}(t) \quad \text{for all} \quad t \in (0, S], \tag{8}$$

where the points $L_0 \equiv 0 < L_1 < L_2 < \cdots < L_r \equiv S$ constitute a partition of $(0, S]$ and $I_{(L_{j-1}, L_j]}(t)$ is the indicator function for the $j$th subinterval $(L_{j-1}, L_j]$.

In the discussion of equations (9)-(14) below, we present an alternative to the approach of Klein and Roberts (1984) for generating an NHPP with the piecewise linear rate function (8) on the interval $(0, S]$. For notational simplicity in this discussion, the increment of the corresponding mean value function over the subinterval $(t_1, t_2]$ will be written

$$\mu^*(t_1, t_2) = \int_{t_1}^{t_2} \lambda^*(z)\, dz, \quad 0 \le t_1 < t_2 \le S. \tag{9}$$

The cumulative distribution function of the $i$th event time $\tau_i$ conditioned on the value of the previous event time $\tau_{i-1}$ can then be expressed as

$$F_{\tau_i}(t|\tau_{i-1}) = \Pr\{\tau_i \le t|\tau_{i-1}\} = \begin{cases} 0, & t < \tau_{i-1} \\ 1 - \exp\{-\mu^*(\tau_{i-1}, t)\}, & t \ge \tau_{i-1} \end{cases}$$

with $\tau_0 \equiv 0$.

Given the $(i-1)$st event time $\tau_{i-1}$ for an NHPP with rate function (8), we generate the next event time $\tau_i$ for this process via the inverse transform method by sampling a random number $U$, solving for $t$ in the equation $F_{\tau_i}(t|\tau_{i-1}) = U$, and setting $\tau_i = t$. This amounts to solving for $\tau_i$ in the equation $\mu^*(\tau_{i-1}, \tau_i) = -\log(1-U)$. If both $\tau_{i-1}$ and $\tau_i$ occur within the $j$th subinterval $(L_{j-1}, L_j]$, then from (8) and (9), we have

$$\mu^*(\tau_{i-1}, \tau_i) = \int_{\tau_{i-1}}^{\tau_i} (a_j z + b_j)\, dz = \frac{1}{2} a_j (\tau_i^2 - \tau_{i-1}^2) + b_j (\tau_i - \tau_{i-1}) \tag{10}$$

$$\text{for} \quad L_{j-1} \le \tau_{i-1} < \tau_i \le L_j.$$

Otherwise, if $\tau_{i-1}$ occurs in the $j$th subinterval and $\tau_i$ is in the $k$th subinterval where $k \geq j+1$, then

$$\begin{aligned}\mu^*(\tau_{i-1}, \tau_i) &= \mu^*(\tau_{i-1}, L_j) + \mu^*(L_j, L_{k-1}) + \mu^*(L_{k-1}, \tau_i) \\ &= \int_{\tau_{i-1}}^{L_j} (a_j z + b_j)\, dz + \delta_k(j+1) \sum_{\ell=j+1}^{k-1} \int_{L_{\ell-1}}^{L_\ell} (a_\ell z + b_\ell)\, dz \\ &\quad + \int_{L_{k-1}}^{\tau_i} (a_k z + b_k)\, dz \\ &\qquad \text{for } L_{j-1} \leq \tau_{i-1} \leq L_j \leq L_{k-1} \leq \tau_i \leq L_k,\end{aligned} \tag{11}$$

where $\delta_k(j+1) = 0$ if $k = j+1$ and $\delta_k(j+1) = 1$ if $k \neq j+1$.

If the $(i-1)$st event time $\tau_{i-1}$ falls in the $j$th subinterval $(L_{j-1}, L_j]$, then we compute the $i$th event time $\tau_i$ from (10) and (11) using the sampled random number $U$ as follows. If $-\log(1-U) \leq \mu^*(\tau_{i-1}, L_j)$ so that both $\tau_{i-1}$ and $\tau_i$ occur in the $j$th subinterval, then from (10) we have

$$a_j \tau_i^2 + 2b_j \tau_i - \left[a_j \tau_{i-1}^2 + 2b_j \tau_{i-1} - 2\log(1-U)\right] = 0;$$

and thus

$$\tau_i = \begin{cases} \dfrac{-b_j + \sqrt{(a_j \tau_{i-1} + b_j)^2 - 2a_j \log(1-U)}}{a_j}, & a_j \neq 0 \\ \tau_{i-1} - \dfrac{\log(1-U)}{b_j}, & a_j = 0 \end{cases}. \tag{12}$$

On the other hand, if $-\log(1-U) > \mu^*(\tau_{i-1}, L_j)$ so that $\tau_i$ occurs in the $k$th subinterval for some $k \geq j+1$, then from (11) we have

$$\tau_i = \begin{cases} \dfrac{-b_k + \sqrt{(a_k L_{k-1} + b_k)^2 - 2a_k[Q + \log(1-U)]}}{a_k}, & a_k \neq 0 \\ L_{k-1} - \dfrac{Q + \log(1-U)}{b_k}, & a_k = 0 \end{cases}, \tag{13}$$

where

$$\left.\begin{aligned} k &\equiv \min\{\ell : \ell \geq j+1 \text{ and } \mu^*(\tau_{i-1}, L_\ell) \geq -\log(1-U)\} \\ Q &\equiv \mu^*(\tau_{i-1}, L_{k-1}) \end{aligned}\right\}. \tag{14}$$

The "piecewise thinning" procedure nhpp($\tau_{\text{next}}$, valid_event) is designed to generate the next event epoch $\tau_{\text{next}}$ from a target NHPP $\{N(t)\}$ whose rate function $\lambda(t)$ is majorized by a piecewise linear function $\lambda^*(t)$ on the interval $(0, S]$. Procedure nhpp() combines (a) the method of thinning, and (b) the formulation given in equations (9)–(14) for the inverse transform method of simulating an NHPP $\{N^*(t)\}$ with the rate function $\lambda^*(t)$. Now we let $\{\tau_k : k = 1, 2, \ldots\}$ denote the successive event epochs of the target process $\{N(t)\}$. If nhpp($\tau_{\text{next}}$, valid_event) is invoked at time $\tau_{k-1}$, then this procedure returns

$$(\tau_{\text{next}}, \text{valid\_event}) = \begin{cases} (\tau_k, \textbf{true}) & \text{if } \tau_k \le S \\ (S, \textbf{false}) & \text{if } \tau_k > S. \end{cases}$$

At the end of this section, a formal statement of procedure nhpp() is presented in a variant of Pidgin ALGOL (Aho, Hopcroft, and Ullman, 1974). In the description of nhpp(), note that $\tau_\ell^*$ represents the next event epoch for the majorizing process $\{N^*(t)\}$ and $\tau_k$ represents the next event of the target process $\{N(t)\}$. It is useful to preset the thinning test such that $\tau_{\text{next}} = \tau_k = \tau_\ell^*$ is always delivered if an independently sampled random number $U_2$ satisfies the condition $U_2 \le \lambda_*/\lambda^*(\tau_\ell^*)$, provided $\lambda_* \equiv \min\{\lambda(t) : 0 \le t \le S\}$ is positive. By avoiding the (usually expensive) evaluation of $\lambda(t)$ in step 7 of nhpp(), we save computation time—especially in cases where the rate function $\lambda(t)$ remains above a relatively high threshold level or is fairly flat over the entire time interval $(0, S]$.

Appendix B contains a formal statement of the procedure maxline($\lambda$, $r$, $\{L_j\}$, $\{a_j, b_j\}$) that forces a majorizing rate function $\lambda^*(t)$ of the form (8) to match the least upper bound of the target rate function $\lambda(t)$ over each subinterval $(L_{j-1}, L_j]$ of the partition of $(0, S]$. Appendix B also contains some guidelines on how to partition the interval $(0, S]$ in order to simulate an NHPP with an EPTF-type rate function on this interval.

## 4. APPLICATION TO OFF-SHORE WEATHER EVENTS

An NHPP with an EPTF-type rate function was fitted to nine years of wave-height data recorded every six hours at a drilling site in the Arctic Sea. The fitted rate function was then applied to the generation of high-wave

**procedure** nhpp($\tau_{\text{next}}$, valid_event):

0. [Initialize.]
   **if** initial call to nhpp() **then**
   **begin**
   **for** $j \leftarrow 1$ **until** $r$ **do**
   $M_j \leftarrow a_j(L_j^2 - L_{j-1}^2)/2 + b_j(L_j - L_{j-1})$;
   $k \leftarrow 1$; $Q \leftarrow 0$; $Q^+ \leftarrow M_k$;
   $\ell \leftarrow 0$; $\tau_\ell^* \leftarrow 0$
   **end**
1. [Increment event index for $\{N^*(t)\}$.] $\ell \leftarrow \ell + 1$
2. [Generate random numbers.] generate $U_1, U_2 \sim \text{IID}\, U(0,1)$
3. [Determine latest increment $\mu^*(\tau_{\ell-1}^*, \tau_\ell^*)$.] $\xi \leftarrow -\log(1 - U_1)$
4. [Determine $k$ and $Q$ in (14) for latest event epoch $\tau_\ell^*$ of $\{N^*(t)\}$.]
   **while** $\xi > Q^+$ **do**
   **begin**
   **if** $k = r$ **then**
   $\tau_{\text{next}} \leftarrow S$; valid_event $\leftarrow$ **false**; **return**
   **else**
   $Q \leftarrow Q^+$; $k \leftarrow k + 1$; $Q^+ \leftarrow Q^+ + M_k$
   **end**
5. [Determine latest event epoch $\tau_\ell^*$ of $\{N^*(t)\}$.]
   **begin**
   $t^+ \leftarrow \max\{\tau_{\ell-1}^*, L_{k-1}\}$;
   **if** $a_k \neq 0$ **then**
   $$\tau_\ell^* \leftarrow \frac{-b_k + \sqrt{(a_k t^+ + b_k)^2 - 2a_k(Q - \xi)}}{a_k}$$
   **else**
   $\tau_\ell^* \leftarrow t^+ - (Q - \xi)/b_k$
   **end**
6. [Update $Q$, $Q^+$.] $Q \leftarrow 0$; $Q^+ \leftarrow Q^+ - \xi$
7. [Perform thinning to get latest event time for $\{N(t)\}$.]
   **if** $U_2 \leq \lambda(\tau_\ell^*)/\lambda^*(\tau_\ell^*)$ **then**
   $\tau_{\text{next}} \leftarrow \tau_\ell^*$; valid_event $\leftarrow$ **true**; **return**
   **else**
   **goto** step 1

events (that is, storms) in a simulation model of offshore drilling operations at the selected site (Lee 1985). We used IMSL subroutines (International Mathematical and Statistical Library, 1987) for random number generation, numerical integration, and numerical solution of linear and nonlinear equations.

The wave-height threshold defining the onset of a high-wave event (simply called a storm in the following discussion) was selected to be 20 feet. Figure 1 depicts the number of storms occurring at the site during each month in a nine-year observation interval. (For simplicity and consistency with the nomenclature used in the previous sections, we label the observation interval $(0, S]$, where $S = 9$ years.) A total of $n = 302$ storms were observed during this time interval. The observed series of storm-arrival times $\{t_j : j = 1, \ldots, n\}$ has pronounced cyclic behavior, and the rate of occurrence of these events appears to vary smoothly from year to year. Table I displays descriptive summary statistics for the corresponding interevent times $\{X_j = t_j - t_{j-1} : j = 1, \ldots, n\}$, where $t_0 \equiv 0$. Table I also displays goodness-of-fit statistics (specifically, the Kolmogorov-Smirnov statistic $D_{\mathrm{KS}}$ and the Anderson-Darling statistic $D_{\mathrm{AD}}$) for testing the uniformity of the event epochs $\{t_j\}$ over the interval $(0, S]$ given that $n = 302$ events were observed in this interval. Now in this context, we may legitimately employ the usual upper 1% critical values of $D_{\mathrm{KS}}$ and $D_{\mathrm{AD}}$ as shown in Table I for the case in which all parameters required for testing conditional uniformity are known; and thus the observed values of $D_{\mathrm{KS}}$ and $D_{\mathrm{AD}}$ strongly indicate that the storm-arrival process is not a homogeneous Poisson process. This conclusion is reinforced by noting that the sample moments of the interevent times (namely, the coefficient of variation $\hat{V}$, the skewness $\sqrt{\hat{\beta}_1}$, and the kurtosis $\hat{\beta}_2$ of the $\{X_j\}$) differ substantially from the corresponding theoretical values for IID exponential interevent times ($V = 1$, $\sqrt{\beta_1} = 2$, and $\beta_2 = 9$). Furthermore, the existence of correlation among the interevent times $\{X_j\}$ is indicated by the von Neumann ratio $R_{\mathrm{VN}}$ and by $Z$, the standardized Fisher $z$-transformation of the lag-one sample correlation.

To model the storm-arrival process as an NHPP, we first fitted a rate function having the exponential-polynomial (EPF) form—that is, the form

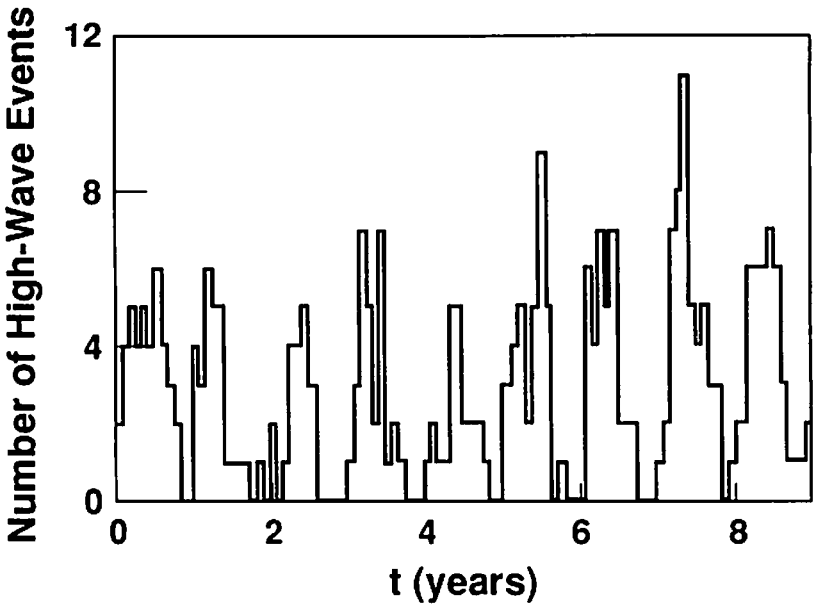


FIG. 1: Number of high-wave events (storms) observed each month over a nine-year observation interval at a site in the Arctic Sea.

TABLE I

Preliminary analysis of the storm-arrival times $\{t_j\}$ with the corresponding interevent times $\{X_j\}$ observed at the Arctic Sea site.

| | | | Critical value at 1% level | |
|---|---|---|---|---|
| Dataset | Statistic | Value | Lower | Upper |
| $\{X_j\}$ | $\hat{V}$ | 1.7340 | | |
| $\{X_j\}$ | $\sqrt{\hat{\beta}_1}$ | 4.8577 | | |
| $\{X_j\}$ | $\hat{\beta}_2$ | 32.2141 | | |
| $\{X_j\}$ | $R_{\mathrm{VN}}$ | 1.5554 | 1.734 | 2.224 |
| $\{X_j\}$ | $Z$ | 3.8354 | −2.575 | 2.575 |
| $\{t_j\}$ | $D_{\mathrm{KS}}$ | 1.8413 | - | 1.628 |
| $\{t_j\}$ | $D_{\mathrm{AD}}$ | 4.5501 | - | 3.857 |

(1) with oscillation amplitude $\gamma = 0$. The Newton-Raphson procedure for estimating the parameters in the rate function failed when the fitted EPF had degree $m > 6$. Let

$$\check{\lambda}(t) = \exp\left(\sum_{i=0}^{6} \check{\alpha}_i t^i\right) \quad \text{for} \quad t \in (0, S] \tag{15}$$

denote the estimate of the EPF-type rate function corresponding to the maximum likelihood estimate $\check{\Theta}_m \equiv [\check{\alpha}_0, \check{\alpha}_1, \ldots, \check{\alpha}_m]$ for $m = 6$, and let $\check{\mu}(t) = \int_0^t \check{\lambda}(z)\,dz$ denote the associated mean value function. We attempted to detrend the event epochs $\{t_j : j = 1, \ldots, n\}$ using $\check{\mu}(t)$ to yield the transformed series $\{\check{s}_j = \check{\mu}(t_j) : j = 1, \ldots, n\}$. Paralleling the goodness-of-fit statistics and sample moments displayed in Table I for the original series of events $\{t_j\}$, analogous results for the transformed series $\{\check{s}_j\}$ are displayed in Table II. In this case, we cannot apply the usual upper 1% critical values for $D_{\mathrm{KS}}$ and $D_{\mathrm{AD}}$ as given in Table I; and thus it is not surprising that $D_{\mathrm{KS}}$ and $D_{\mathrm{AD}}$ do not reveal any marked departures from a uniform distribution on the interval $(0, \check{\mu}(S)]$ when these statistics are computed for the data set $\{\check{s}_j\}$. On the other hand, Figure 2 shows an obvious departure from linearity in the exponential probability plot for the transformed interevent times $\{\check{X}_j = \check{s}_j - \check{s}_{j-1} : j = 1, \ldots, n\}$, where $\check{s}_0 \equiv 0$. This latter discrepancy is also indicated by the tests based on Durbin's transformation (Durbin, 1961). The dispersion of the transformed interevent times $\{\check{X}_j\}$ appears to be greater than would occur for a homogeneous Poisson process; moreover, the $\{\check{X}_j\}$ display stochastic dependencies that are inconsistent with a homogeneous Poisson process. Therefore, we concluded that a computationally feasible EPF-type rate function cannot adequately fit the original series of storm-arrival times $\{t_j\}$. The trend shown in Figure 1 provides strong evidence that an appropriate model for this series of events is an NHPP with an EPTF-type rate function.

There is good reason to believe that the storm-arrival process will exhibit a pronounced time-of-year effect so that the oscillation frequency $\omega = 2\pi$ radians per year. This conclusion is strongly supported by the spectral analysis of the series $\{t_j\}$. As displayed in Figure 3, the periodogram of this series was computed at the frequency $2\pi\ell/S$ radians per year (that is, $\ell/S$ cycles

TABLE II

Statistics for the transformed storm-arrival times $\{\check{s}_j\}$ with the corresponding interevent times $\{\check{X}_j\}$ based on the estimated EPF-type rate function (15).

| Dataset | Statistic | Value |
|---|---|---|
| $\{\check{X}_j\}$ | $\hat{V}$ | 1.5988 |
| $\{\check{X}_j\}$ | $\sqrt{\hat{\beta}_1}$ | 4.2467 |
| $\{\check{X}_j\}$ | $\hat{\beta}_2$ | 24.0325 |
| $\{\check{X}_j\}$ | $R_{VN}$ | 1.5710 |
| $\{\check{X}_j\}$ | $\mathcal{Z}$ | 3.6924 |
| $\{\check{s}_j\}$ | $D_{KS}$ | 0.5754 ( 3.0500) |
| $\{\check{s}_j\}$ | $D_{AD}$ | 0.4653 (19.6756) |

( ) : statistic for Durbin's transformation

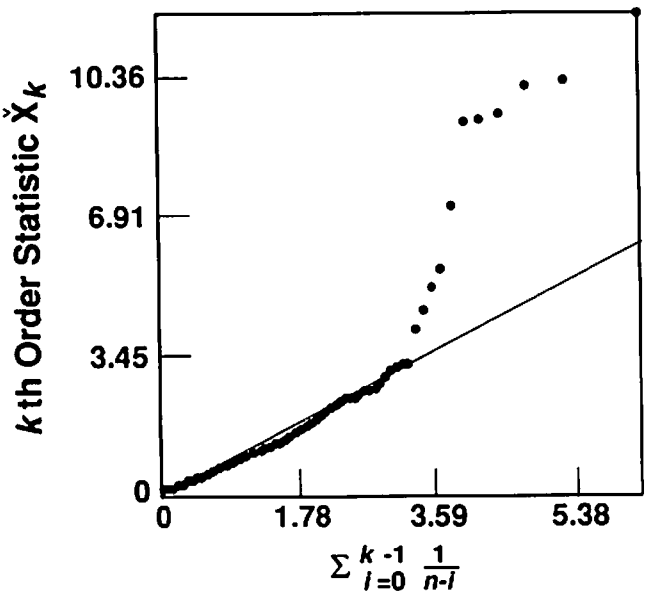


FIG. 2: $\sum_{i=0}^{k-1} 1/(n-i)$ versus the $k$th order statistic $\check{X}_{(k)}$ of the transformed interevent times $\{\check{X}_j = \check{s}_j - \check{s}_{j-1}\}$ based on the estimated EPF-type rate function (15).

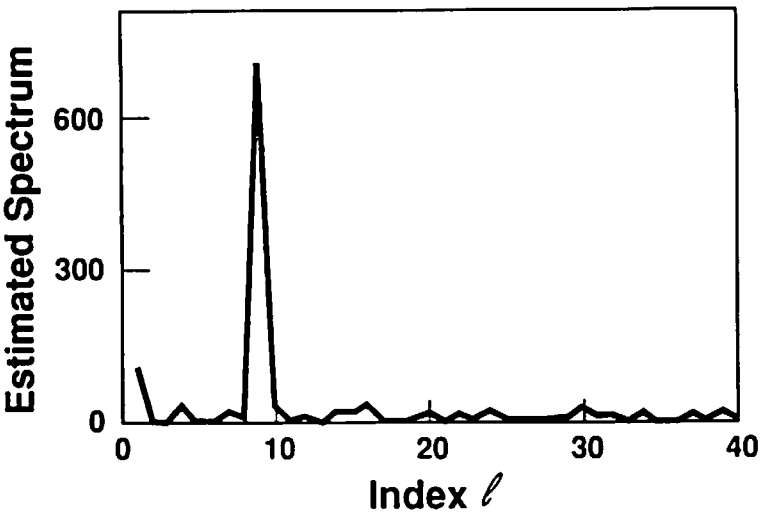


FIG. 3: Estimated spectrum (periodogram) of the observed series of high-wave events $\{t_j\}$.

per year) where the index $\ell$ takes the integer values $\{1, 2, \ldots, 40\}$. The single pronounced spike in the periodogram occurs at the index $\ell = 9$, which corresponds to the frequency $\omega = 2\pi$ radians per year; thus the period of the cyclic behavior in the storm-arrival process is $S/\ell = 1$ year. Note also that the secondary spike in the periodogram near $\ell = 0$ (that is, near zero frequency) indicates the presence of a long-term evolutionary trend in the storm-arrival process. The initial estimates $\gamma_0$ and $\phi_0$ of the amplitude and phase respectively were computed from (5) by using the initial value $\omega_0 = 2\pi$ radians per year. These results are shown in Table III, which also contains the initial estimates $\gamma_0$ and $\phi_0$ corresponding to the initial frequency estimate $\omega_0 = 2\pi\ell/S$ radians per year for $\ell = 7$, 8, 10, and 11. The estimated amplitude $\gamma_0$ is an order of magnitude larger for $\ell = 9$ than for $\ell = 7$, 8, 10, or 11. We concluded that a frequency close to the initial estimate of $\omega_0 = 2\pi$ radians per year (with the corresponding period of one year) can properly represent the cyclic behavior of the process. Our computational experience with the storm-arrival data set indicates that we fail to obtain the optimal

TABLE III

Initial estimates $\gamma_0$ and $\phi_0$ obtained by solving (5) for each frequency of the form $\omega_0 = 2\pi\ell/S$ radians per year, where $\ell = 7, 8, \ldots, 11$.

| $\ell$ | $\gamma_0$ | $\phi_0$ |
|---|---|---|
| 7 | 0.1619 | 0.3520 |
| 8 | 0.1214 | 0.6661 |
| 9 | 1.0592 | −0.7361 |
| 10 | 0.0588 | −1.1193 |
| 11 | 0.1273 | −0.7848 |

parameter estimates for the EPTF if we attempt to solve the system (3) of likelihood equations with other values of $\omega_0$—even with values of $\omega_0$ differing from $2\pi$ radians per year by as little as $\pm 2\pi/S$ radians per year.

The following conclusions can be drawn from the likelihood ratio tests based on the log-likelihood values in Table IV: (a) the purely trigonometric EPTF-type rate function (that is, the EPTF of degree zero) is significant, which indicates pronounced time-of-year effects due to seasonal variations; (b) the EPTF-type rate functions of nonzero degree are significant, which indicates a definite year-to-year trend in the arrival rate of high-wave events over the nine-year period; and (c) the EPTFs of degree greater than three do not provide a significantly better fit to the data than the third-degree EPTF. Note that in Table IV, we use $\hat{\mathcal{L}}_m$ and $\tilde{\mathcal{L}}_m$ respectively to denote the maximum log-likelihood for EPF- and EPTF-type rate functions of degree $m$. The final parameter estimates for the EPTF-type rate function of degree three are contained in Table V. Significance tests and confidence limits for these parameters are discussed in Cox (1972).

Table V shows that the moment-matching technique described by equation (6) can provide good initial values for the parameters of the polynomial rate component. Moreover, the spectral analysis of the event epochs $\{t_j\}$ yielded a good initial estimate of the oscillation frequency; and the corresponding initial estimates of the amplitude and phase based on (4) and (5)

TABLE IV

Maximum log-likelihood (2) for EPF- and EPTF-type rate functions.

| Degree $m$ | EPF-based $\check{\mathcal{L}}_m(\check{\Theta}_m \mid n, t)$ | Difference $\check{\mathcal{L}}_m - \check{\mathcal{L}}_{m-1}$ | EPTF-based $\tilde{\mathcal{L}}_m(\tilde{\Theta}_m \mid n, t)$ | Difference $\tilde{\mathcal{L}}_m - \tilde{\mathcal{L}}_{m-1}$ |
|---|---|---|---|---|
| 0 | 758.9871 | - | 829.6861 | - |
| 1 | 760.8759 | 1.8888 | 833.0218 | 3.3357 |
| 2 | 764.1052 | 3.2293 | 836.6168 | 3.5950 |
| 3 | 769.9283 | 5.8231 | 839.1851 | 2.5683 |
| 4 | 770.1815 | 0.2532 | 839.2283 | 0.0432 |
| 5 | 770.3383 | 0.1568 | 839.4351 | 0.2068 |
| 6 | 772.0798 | 1.7415 | 839.7391 | 0.3040 |

TABLE V

Maximum likelihood parameter estimates for the EPTF-type rate function (16) of degree $m = 3$.

| Parameter | Final estimated value | Initial value |
|---|---|---|
| $\alpha_0$ | 3.6269 | 3.8654 |
| $\alpha_1$ | −0.6324 | −0.6126 |
| $\alpha_2$ | 0.1552 | 0.1486 |
| $\alpha_3$ | −0.0096 | −0.0096 |
| $\gamma$ | 1.0643 | 1.0592 |
| $\omega$ | 6.2581 | 6.2831 |
| $\phi$ | −0.6193 | −0.7361 |

were also quite close to the final estimates obtained via the Newton-Raphson procedure. Let

$$\tilde{\lambda}(t) = \exp\left[\sum_{i=0}^{3} \tilde{\alpha}_i t^i + \tilde{\gamma}\sin(\tilde{\omega}t + \tilde{\phi})\right] \quad \text{for} \quad t \in (0, S] \tag{16}$$

denote the final estimate of the EPTF-type rate function corresponding to the maximum likelihood estimator $\tilde{\Theta}_m$ for $m = 3$, and let $\tilde{\mu}(t) = \int_0^t \tilde{\lambda}(z)\,dz$ denote the associated mean value function. Figure 4 displays a plot of the rate function (16) over the observation interval $(0, S]$.

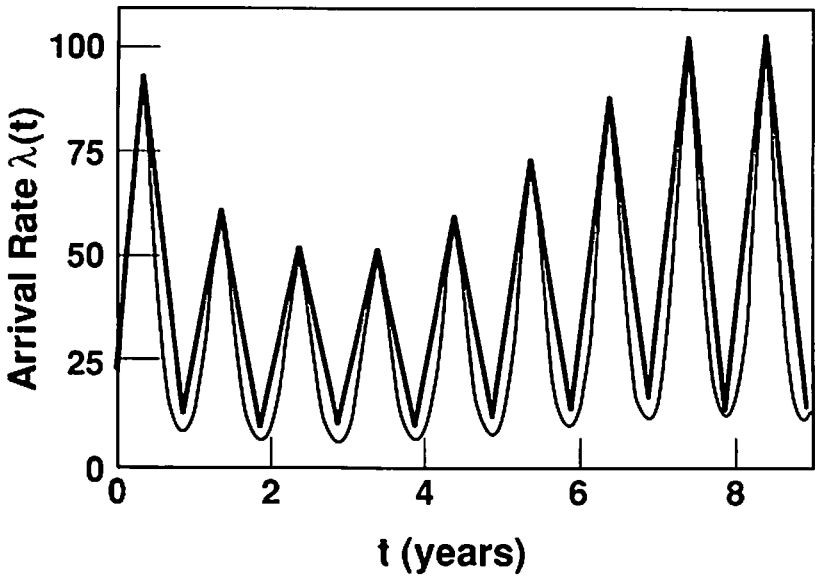


FIG. 4: Estimated EPTF-type rate function (16) (thin curve) and the piecewise linear majorizing function (8) (thick line) constructed by invoking procedure maxline($\tilde{\lambda}, r, \{L_j\}, \{a_j, b_j\}$) with $r = 19$ and a given set of end points $\{L_j\}$ corresponding to the stationary points of (16) in the interval $(0, S]$.

Paralleling the goodness-of-fit statistics and sample moments displayed in Table I for the original series of events $\{t_j\}$ and in Table II for the initially transformed series of events $\{\tilde{s}_j\}$, analogous results for the final detrended series $\{\tilde{s}_j \equiv \tilde{\mu}(t_j) : j = 1, \ldots, n\}$ are displayed in Table VI. As in the case of Table II, we should note that the usual upper 1% critical values for $D_{KS}$ and $D_{AD}$ given in Table I do not apply to the goodness-of-fit statistics reported in Table VI; however, in this case, the observed values of $D_{KS}$ and $D_{AD}$ provide some confirmation of the other graphical and statistical evidence of the adequacy of the fitted NHPP. For example in Figure 5, the exponential probability plot for the detrended interevent times $\{\tilde{X}_j = \tilde{s}_j - \tilde{s}_{j-1} : j = 1, \ldots, n\}$ reveals no marked departures from a straight line with slope one; and this strongly indicates that the detrended process

TABLE VI

Statistics for the detrended storm-arrival times $\{\tilde{s}_j\}$ with the corresponding interevent times $\{\tilde{X}_j\}$ based on the final estimated EPTF-type rate function (16).

| Dataset | Statistic | Value |
|---|---|---|
| $\{\tilde{X}_j\}$ | $\hat{V}$ | 0.9851 |
| $\{\tilde{X}_j\}$ | $\sqrt{\hat{\beta}_1}$ | 1.9225 |
| $\{\tilde{X}_j\}$ | $\hat{\beta}_2$ | 7.2749 |
| $\{\tilde{X}_j\}$ | $R_{\mathrm{VN}}$ | 1.8079 |
| $\{\tilde{X}_j\}$ | $\mathcal{Z}$ | 1.6273 |
| $\{\tilde{s}_j\}$ | $D_{\mathrm{KS}}$ | 0.4603 (1.4693) |
| $\{\tilde{s}_j\}$ | $D_{\mathrm{AD}}$ | 0.1899 (3.0034) |

( ) : statistic for Durbin's transformation

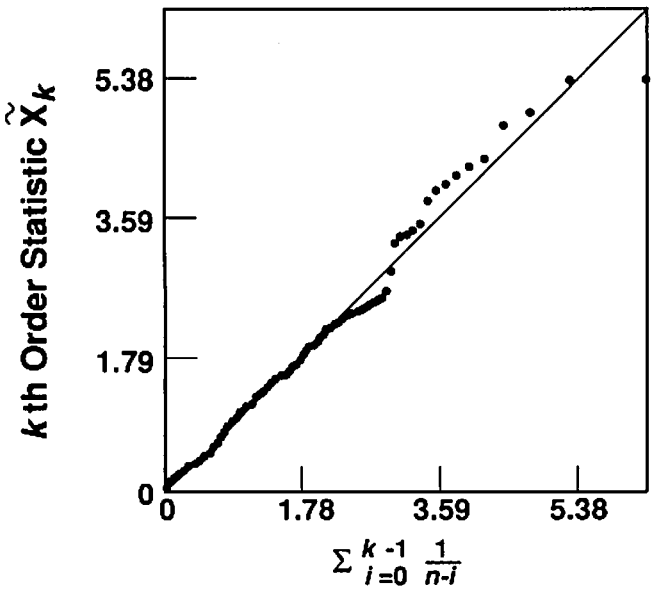


FIG. 5: $\sum_{i=0}^{k-1} 1/(n-i)$ versus the $k$th order statistic $\tilde{X}_{(k)}$ of the detrended interevent times $\{\tilde{X}_j = \tilde{s}_j - \tilde{s}_{j-1}\}$ based on the estimated EPTF-type rate function (16).

$\{\tilde{M}(s) \equiv N[\tilde{\mu}^{-1}(s)] : s \in (0, \tilde{\mu}(S)]\}$ is a homogeneous Poisson process with unit rate. For the detrended interevent times $\{\tilde{X}_j\}$, both the von Neumann ratio $R_{VN}$ and $Z$, the standardized $z$-transformation of the lag-one correlation, show no evidence of nonrandom behavior. Thus, we conclude that the arrival process for high-wave events is consistent with an NHPP having a third degree EPTF-type rate function. Figure 6 contains plots of the cumulative number of observed events $N(t) \equiv \max\{j : t_j \leq t\}$ and the final estimated mean value function $\tilde{\mu}(t)$ for $t \in (0, S]$. We believe that the summary statistics in Table VI and the plots in Figures 5 and 6 collectively constitute good evidence of the adequacy of the fitted NHPP.

The piecewise thinning procedure nhpp() was employed to generate realizations of the fitted NHPP. The nine-year observation interval was divided into 19 subintervals, and a piecewise linear majorizing rate function was constructed using the procedure maxline() described in Appendix B. The thick piecewise linear plot in Figure 4 represents the majorizing rate function, which clearly approximates the fluctuating behavior of the estimated rate function (16) as a close upper bound. A dual CDC CYBER 170/750 computer system was used to execute 100 simulation runs of the fitted process using a FORTRAN 77 implementation of procedure nhpp(). The number of events simulated in each run was distributed between 256 and 342 with standard deviation 15.52 (the expected number of events is 302). To evaluate the efficiency of incorporating the majorizing function depicted in Figure 4 into the thinning procedure, we compared the overall execution time of nhpp() using this majorizing function to the execution time of nhpp() using a constant majorizing rate function $\lambda^*(t) \equiv \max\left\{\tilde{\lambda}(z) : 0 \leq z \leq 9\right\} \approx 100$. The summary results in Table VII indicate that the computation time for piecewise thinning was reduced to almost half that of constant thinning. This improvement resulted from the difference in the total number of events $\{\tau_i^*\}$ generated for the majorizing process by procedure nhpp(), where the generation of each individual event requires at least two random numbers and one evaluation of the target rate function (16). This implies that the cost for inversion of the piecewise linear rate majorizing rate function is relatively small compared to the cost of generating random numbers and evaluating a complicated rate function.

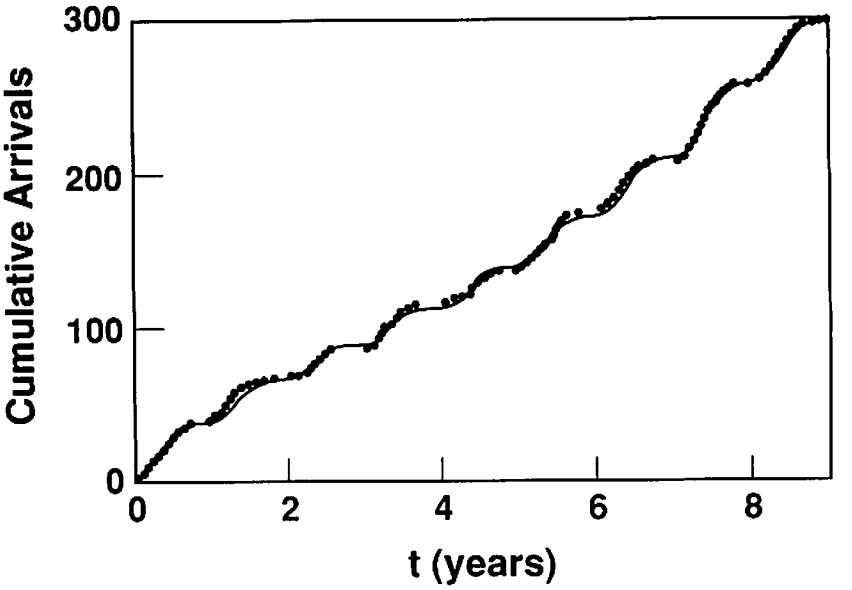


FIG. 6: Cumulative number of events $N(t)$ observed up to time $t$ (dotted curve) versus the mean value function $\tilde{\mu}(t)$ (solid curve) corresponding to the estimated EPTF-type rate function (16).

TABLE VII

Results of 100 independent simulation runs of length $S = 9$ years for the fitted NHPP with rate function (16).

| | Piecewise thinning | Constant thinning |
|---|---|---|
| Total CPU time, CDC CYBER 170/750 | 7.08 sec | 12.36 sec |
| Average number of accepted events | 303.33 | 301.49 |
| Standard deviation of number of accepted events | 15.52 | 17.79 |
| Average number of total generated events | 422.26 | 899.24 |
| Efficiency of thinning | 72 % | 34 % |

## 5. CONCLUSIONS

For modeling and analysis of nonhomogeneous Poisson processes, Lewis (1970, 1972) investigated the EPTF-type rate function with a known oscillation frequency. Lewis and Shedler (1976, 1979) examined NHPPs with an EPF-type rate function, and they developed the thinning technique for simulation of general NHPPs. In this paper, we have extended these results to cover a general EPTF-type rate function with an unknown oscillation frequency, which is an appropriate model for many time-dependent Poisson processes that exhibit cyclic behavior.

It is difficult to characterize the periodic nature of nonstationary point processes. Spectral analysis of the point process is an effective way to study the cyclic features of an NHPP with an EPTF-type rate function. In our experience, an NHPP whose rate function includes a long-term trend or a cyclic rate component can be adequately modeled with an EPTF-type rate function. Moreover, the parameters of an EPTF-type rate function can be effectively estimated by (a) using a maximum likelihood ratio test to identify the degree of the polynomial rate component; and (b) solving the usual likelihood equations to derive the maximum likelihood estimates of the continuous parameters of the rate function. The procedure for obtaining initial parameter estimates using equations (4)–(6) has been found to work reliably in practice.

An NHPP with an EPTF-type rate function can be generated exactly and efficiently by the piecewise thinning algorithm implemented in procedure nhpp(). The supplemental procedure maxline() presented in Appendix B provides a piecewise linear majorizing function which is close enough to the original rate function to achieve high efficiency. The proposed approach to modeling and simulation of NHPPs has been applied to a complex point process with cyclic features as well as a long-term trend. While development of an adequate model required extensive computation and nontrivial numerical techniques for estimating the parameters, the periodic nature of the arrivals of high-wave events was accurately represented by an NHPP with an EPTF-type rate function. The year-to-year trend was also well represented. Simulation applications of the fitted NHPP model have demonstrated the advantages of using our proposed piecewise thinning method.

## APPENDIX A

If $k = 0$, then we have

$$\mathcal{M}_{\sin}(0, S; \omega, \phi) = \omega^{-1}[\cos\phi - \cos(\omega S + \phi)]. \tag{A.1}$$

If $k > 0$, then repeated integration by parts yields

$$\begin{aligned}\mathcal{M}_{\sin}(k, S; \omega, \phi) = \cos(\omega S + \phi) \sum_{r=0}^{\lfloor \frac{k}{2} \rfloor} (-1)^{r+1} \frac{k!}{(k-2r)!} \cdot \frac{S^{k-2r}}{\omega^{2r+2}} \\ + \sin(\omega S + \phi) \sum_{r=0}^{\lfloor \frac{k-1}{2} \rfloor} (-1)^{r} \frac{k!}{(k-2r-1)!} \cdot \frac{S^{k-2r-1}}{\omega^{2r+2}}.\end{aligned} \tag{A.2}$$

where $\lfloor z \rfloor$ represents the greatest integer $\leq z$.

## APPENDIX B

To find a majorizing rate function $\lambda^*(t)$ of the form (8) that most closely approximates the target rate function $\lambda(t)$ on the interval $(0, S]$, we seek to minimize the area

$$\int_0^S [\lambda^*(z) - \lambda(z)]\, dz \tag{B.1}$$

between these two functions. For this purpose, we can use a simple line search—provided we can choose the end points $\{L_j\}$ such that the following conditions hold on the $j$th subinterval $(L_{j-1}, L_j]$ for $j = 1, \ldots, r$: (a) the discrepancy $g_j(t) \equiv \lambda(t) - (a_j t + b_j)$ is strictly quasiconcave (Bazarra and Shetty, 1979); and (b) the equation $g_j(t) = 0$ has at most one root in the open subinterval $(L_{j-1}, L_j)$. Usually, conditions (a) and (b) can be achieved by assigning the end points $\{L_j\}$ to the stationary points of the target rate function $\lambda(t)$. The stationary points of an EPTF-type function $\lambda(t)$ are often close to the stationary points of the associated trigonometric rate component. If local fluctuation of the target rate function $\lambda(t)$ is sufficiently large, then reasonable choices for the number of subintervals $r$ and the corresponding end points $\{L_j\}$ are:

$$r = \left\lfloor \frac{S\omega}{\pi} \right\rfloor + 1 \tag{B.2}$$

and

$$L_j = \frac{1}{\omega}\left[(k+j-1)\pi + \frac{\pi}{2} - \phi\right] \quad \text{for } j = 1, 2, \ldots, r-1, \tag{B.3}$$

where $k$ is the smallest nonnegative integer such that

$$\frac{1}{\omega}\left(k\pi + \frac{\pi}{2} - \phi\right) \geq 0. \tag{B.4}$$

Given the number of subintervals $r$ and the corresponding end points $\{L_j : j = 0, 1, \ldots r\}$ satisfying the conditions (a) and (b) mentioned above, the procedure maxline($\lambda$, $r$, $\{L_j\}$, $\{a_j, b_j\}$) returns the parameters $\{a_j, b_j\}$ of the majorizing function (8) that minimizes (B.1). Two auxiliary procedures are invoked by maxline(): (*i*) procedure bisect($g$, $L_{j-1}$, $L_j$, $\tau$) computes the zero $\tau$ of the function $g(t)$ in the interval $(L_{j-1}, L_j]$ by the bisection method; and (*ii*) procedure goldsect($g$, $p_j$, $q_j$, $\delta$) computes the maximum $\delta$ of the function $g(t)$ in the interval $[p_j, q_j]$ by the golden section method (Bazarra and Shetty, 1979). In the following formal statement of procedure maxline(), note that $g(t) = g(t; c, d) \equiv \lambda(t) - (ct + d)$ for all $t \in (0, S]$.

---

**procedure** maxline($\lambda, r, \{L_j\}, \{a_j, b_j\}$):

0. [Initialize.]
   **begin**
   $y_0 \leftarrow \lambda(L_0)$;
   **for** $j \leftarrow 1$ **until** $r$ **do**
   $y_j \leftarrow \lambda(L_j)$; over_line$_j \leftarrow$ **false**
   **end**
1. [Find the subintervals on which $g(t)$ has constant sign.]
   **for** $j \leftarrow 1$ **until** $r$ **do**
   **begin**
   $c \leftarrow (y_j - y_{j-1})/(L_j - L_{j-1})$; $d \leftarrow y_j - cL_j$;
   $p_j \leftarrow L_{j-1}$; $q_j \leftarrow L_j$;
   [$g(L_{j-1}^+)$ and $g(L_j^-)$ respectively denote right and left limits.]
   **if** sign$[g(L_{j-1}^+)]$ = sign$[g(L_j^-)]$ **then**
   **if** $g(L_{j-1}^+) \leq 0$ **then**
   over_line$_j \leftarrow$ **true**

```
            else;
         else
            begin
               [Find τ, the root of g(t) = 0 in (L_{j-1}, L_j).]
               call bisect(g, L_{j-1}, L_j, τ);
               if g(L_{j-1}^+) ≤ 0 then
                  p_j ← τ
               else
                  q_j ← τ
            end
      end
2. [Set max_line and Δ.] max_line ← true;  Δ ← arbitrary big value;
3. [Find the min_j max{g(t) : t ∈ (L_{j-1}, L_j] and g(t) > 0}.]
   for j ← 1 until r do
      begin
         if over_line_j = false then
            begin
               c ← (y_j − y_{j-1})/(L_j − L_{j-1});  d ← y_j − cL_j;
               [Find δ, the maximum value of g(t) for t ∈ [p_j, q_j].]
               call goldsect(g, p_j, q_j, δ);
               if δ > 0 and δ < Δ then
                  max_line ← false;  k ← j;  Δ ← δ
            end
      end
4. [Update λ*(t) on subinterval (L_{k-1}, L_k] found in step 3.]
   if max_line = true then
      begin
         for j ← 1 until r do
            a_j ← (y_j − y_{j-1})/(L_j − L_{j-1});  b_j ← y_j + a_jL_j;
         return
      end
   else
      y_{k-1} ← y_{k-1} + Δ;  y_k ← y_k + Δ;  over_line_k ← true;

5. [Repeat steps 2–5 until g(t) ≤ 0 for all t ∈ (0, S].]
   goto step 2.
```

## ACKNOWLEDGMENTS

The authors thank Professor David Kelton of the University of Minnesota and Mr. Robert Klein of the Regenstrief Institute for their insightful comments on this paper. The authors also thank Professor J. P. C. Kleijnen and the referees for numerous suggestions that substantially improved the exposition. This research was sponsored by ARCO Oil and Gas Company under Contract No. DE 83-157. It was also partially based upon work supported by the National Science Foundation under Grant No. DMS-8717799.

## BIBLIOGRAPHY

Aho, A. V., Hopcroft, J. E. and Ullman, J. D. (1974). *The Design and Analysis of Computer Algorithms*. Reading, Massachusetts: Addison-Wesley.

Ashkar, F. and Rousselle, J. (1983). 'The effect of certain restrictions imposed on the inter-arrival time of flood events on the Poisson distribution used for modeling flood counts,' *Water Resources Research, 19*, 481–485.

Bartlett, M. S. (1955). *An Introduction to Stochastic Processes*. Cambridge: Cambridge University Press.

Bazarra, M. S. and Shetty, C. M. (1979). *Nonlinear Programming: Theory and Algorithms*. New York: John Wiley & Sons.

Bingham, C. and Nelson, L. S. (1981). 'An approximation for the distribution of the von Neumann ratio,' *Technometrics, 23*, 285–288.

Bratley, P., Fox, B. L., and Schrage, L. E. (1987). *A Guide to Simulation*, Second Edition. New York: Springer-Verlag.

Cox, D. R. (1972). 'The statistical analysis of dependencies in point processes,' *Stochastic Point Processes* (P. A. W. Lewis, Ed.). New York: John Wiley & Sons, 55–66.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman and Hall.

Cox, D. R. and Lewis, P. A. W. (1966). *The Statistical Analysis of Series of Events*. London: Chapman and Hall.

Crawford, M. M. (1985). Time series modeling of sea state variables. Center for Space Research Technical Report No. 85-9. The University of Texas, Austin, TX.

Durbin, J. (1961). 'Some methods of constructing exact tests,' *Biometrika, 48*, 41–55.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications, Volume II*, Second Edition. New York: John Wiley & Sons.

Hahn, G. J. and Shapiro, S. S. (1967). *Statistical Models in Engineering.* New York: John Wiley & Sons.

Hoffman, S. E. (1982). An integrated model of drilling vessel operations. Department of Mechanical Engineering M.S. Thesis. The University of Texas, Austin, TX.

Hoffman, S. E., Crawford, M. M., and Wilson, J. R. (1983). 'An integrated model of drilling vessel operations,' *Proceedings of the 1983 Winter Simulation Conference* (S. D. Roberts, J. Banks, and B. W. Schmeiser, Eds.). Piscataway, NJ: Institute of Electrical and Electronics Engineers, 45–53.

International Mathematical and Statistical Library (1987). *IMSL User's Manual*, IMSL, Inc., Houston, TX.

Kaminsky, F. C. and Rumpf, D. L. (1977). 'Simulating nonstationary Poisson processes: a comparison of alternatives including the correct approach,' *Simulation, 29*, 17–20.

Klein, R. W. and Roberts, S. D. (1984). 'A time-varying Poisson arrival process generator,' *Simulation, 43*, 193–195.

Law, A. M. and Kelton, W. D. (1982). *Simulation Modeling and Analysis.* New York: McGraw-Hill.

Lee, S. (1985). Modeling and simulation of sea-state variables as nonhomogeneous Poisson processes and bilinear time series. Department of Mechanical Engineering M.S. Thesis. The University of Texas, Austin, TX.

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

Lewis, P. A. W. (1972). 'Recent results in the statistical analysis of univariate point processes,' *Stochastic Point Processes* (P. A. W. Lewis, Ed.). New York: John Wiley & Sons, 1–54.

Lewis, P. A. W. and Shedler, G. S. (1976a). 'Simulation of nonhomogeneous Poisson processes with log linear rate function,' *Biometrika, 63*, 501–505.

Lewis, P. A. W. and Shedler, G. S. (1976b). 'Statistical analysis of nonstationary series events in a data base system,' *IBM Journal of Research and Development, 20*, 465–482.

Lewis, P. A. W. and Shedler, G. S. (1979). 'Simulation of nonhomogeneous Poisson processes by thinning,' *Naval Research Logistics Quarterly, 26*, 403–413.

MacLean, C. J. (1974). 'Estimation and testing of an exponential polynomial rate function within the nonstationary Poisson process,' *Biometrika*, *61*, 81–85.

Parzen, E. (1962). *Stochastic Processes*. San Francisco: Holden Day.

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

Stuart, A. and Ord, J. K. (1987). *Kendall's Advanced Theory of Statistics, Volume 1: Distribution Theory*, Fifth Edition, New York: Oxford University Press.

von Neumann, J. (1941). 'Distribution of the ratio of the mean square successive difference to the variance,' *Annals of Mathematical Statistics*, *12*, 367–395.