STATISTICAL ANALYSIS ON MARKOV-MODULATED POISSON PROCESSES

N. I. RAMESH

School of Mathematics, Statistics and Computing, The University of Greenwich, Wellington Street, Woolwich, London SE18 6PF, U.K.

SUMMARY

A class of doubly stochastic Poisson processes, which is termed a Markov-modulated Poisson process, is studied. The maximum likelihood method is used to make inferences about the Markov-modulated Poisson process. Expressions are derived for the likelihood function and for second-order properties of both counts and intervals. A simple two-state model is applied to a set of exposure data and to simulated data. Bivariate generalization of this process is also studied.

KEY WORDS   conditional intensity; doubly stochastic Poisson process; maximum likelihood; point process; spectral density

1. INTRODUCTION

Doubly stochastic Poisson processes or Cox processes, in which the rate of occurrence is determined by a stationary non-negative stochastic process, were introduced by Cox[1] and have been studied by many authors, including Bartlett,[2] Kingman,[3] Grandell,[4] Rudemo[5] and Snyder and Miller.[6] Statistical analysis of such processes is usually performed in a rather *ad hoc* manner by calculation of their second-order properties, because the likelihood function is not usually available in a useful form. The purpose of this paper is to describe the properties of a class of doubly stochastic Poisson processes, the so-called Markov-modulated Poisson processes (MMPPs), for which the likelihood can be calculated.

The MMPP is a Cox process in which the arrival rate is modulated or directed by an underlying continuous-time irreducible Markov process $\{X(t)\}$ on a finite state space. In most applications, only the point process $\{N(t)\}$ of occurrences is observed and the underlying Markov process $\{X(t)\}$ is unobserved. In modelling rainfall or the occurrence of storms, $\{X(t)\}$ can be interpreted as an environmental or climatological process.

Smith and Karr[7] used a special MMPP to model summer season rainfall occurrences, Neuts[8] discussed the application of this process to queueing models. Meier-Hellstern[9] developed an iterative statistical procedure for fitting MMPPs having two arrival rates. Our aim is to propose a statistical analysis based on maximum likelihood estimation[10] and second-order properties of the process.

Section 2 describes the process, parameter estimation, and the second-order properties of both the intervals and counts. Section 3 extends the ideas to bivariate MMPPs. A simulation study in Section 4 aims to find how much data are needed before the asymptotic properties of likelihood quantities apply. In Section 5 both univariate and bivariate models are applied to two series of notional exposures to time-integrated air concentration of radionuclides.

2. MARKOV-MODULATED POISSON PROCESS

2.1. Likelihood approach

Suppose $\{N(t)\}$ is an MMPP whose underlying process $\{X(t)\}$ is a stationary irreducible Markov chain with k states, labelled $1, 2, \ldots, k$, and infinitesimal generator $Q_{k \times k}$. We shall assume $\{X(t)\}$ is initially in equilibrium, with equilibrium probabilities $\pi = (\pi_1, \pi_2, \ldots, \pi_k)$. Conditionally on $\{X(t)\}$, the point process $\{N(t)\}$ is a Poisson process with rate $\lambda_{X(t)}$, where $0 \leqslant \lambda_i < \infty$, $i = 1, 2, \ldots, k$ and at least one of the λ_i is positive. That is, whenever the Markov chain $\{X(t)\}$ is in state i, arrivals occur according to a Poisson process of rate λ_i. Let $L_{k \times k} = \text{diag}(\lambda_1, \ldots, \lambda_k)$. The parameters in L and Q are to be estimated from the observations of the process $\{N(t)\}$, as $\{X(t)\}$ is unobserved.

To obtain an expression for the likelihood function, we first define the conditional probabilities[5,10]

$$\psi_{ij}(t) = P\{X(t) = j, N(t) = 0 | X(0) = i, N(0) = 0\}, \quad i, j = 1, \ldots, k.$$

Let $\Psi(t)$ be the matrix function with entries $\psi_{ij}(t)$. Then a forward argument shows that

$$\Psi(t) = \exp\{(Q - L)t\} = \sum_{n=0}^{\infty} \frac{t^n (Q - L)^n}{n!}, \tag{1}$$

where $(Q - L)^0 = I$.

Suppose $N(t)$ is observed on $[0, T]$, during which period points occur at times $t_1 < t_2 < \cdots < t_n$. Let $\pi_{1 \times k}$ be the initial probability vector of $X(t)$. Then the likelihood for Q and L given $t_1, \ldots, t_n$ is[10]

$$f(t_1, \ldots, t_n | Q, L) = \pi \left[\prod_{i=1}^{n} \{\Psi(t_i - t_{i-1})L\} \right] \Psi(T - t_n) \mathbf{1}, \tag{2}$$

where $\mathbf{1}$ is a $k \times 1$ vector of ones and $t_0 = 0$. If the eigenvalues of the matrix $Q - L$ are distinct then $Q - L$ can be expressed as ADA^{-1}, where D is the diagonal matrix of eigenvalues of $Q - L$ and A is a matrix whose columns are the eigenvectors of $Q - L$. Therefore $\Psi(t)$ can be written as $A \exp\{Dt\}A^{-1}$, and (2) reduces to

$$f(t_1, \ldots, t_n | Q, L) = \pi \left\{ \prod_{i=1}^{n} [A \exp\{D(t_i - t_{i-1})\}A^{-1}L] \right\} A \exp\{D(T - t_n)\}A^{-1} \mathbf{1}, \tag{3}$$

and computationally this is a useful expression.

In general, numerical maximization of the likelihood (3) requires a search in k^2 dimensions. The practical difficulties arising in maximization as well as in the spectral analysis are greatly reduced if for values of $k > 2$ we confine the underlying Markov process to possess transitions only between adjacent states, which is a sensible assumption for many applications.

2.2. Stationarity

Associated with a stationary point process is the stationary process of intervals between successive points. The subtle connection between stationarity of point processes and interval stationarity is discussed by Daley and Vere-Jones (reference 11, Chapter 12). The distribution of this process of intervals is the Palm distribution, sometimes referred to as arbitrary event initial conditions. The stationary distribution of the point process corresponds to asynchronous sampling or arbitrary time initial conditions. The arbitrary time initial condition for the point

process is simply the stationary distribution π of $\{X(t)\}$. To see the arbitrary event initial conditions, we proceed as follows.

Consider the epochs of successive arrivals in our MMPP and assume that $t = 0$ is an arrival epoch. Let $Z_i, i = 1, \ldots, n$ be inter-event times. Define the transition probability distribution matrix $F_{k \times k}(\cdot)$ of the times between events of the process

$$F(z) = \int_0^z \exp\{(Q - L)t\}L\,dt, \quad z \geqslant 0,$$

where $F(\cdot)$ has (i, j) element

$$F_{ij}(z) = P\{X(t_n) = j, Z_n \leqslant z | X(t_{n-1}) = i\}, \quad n \geqslant 1,$$

and $X(t)$ is the state of the underlying Markov chain at t. Then it may be verified that[8]

$$F(\infty) = \int_0^\infty \exp\{(Q - L)t\}L\,dt = (L - Q)^{-1}L \tag{4}$$

is a stochastic matrix with stationary probability vector $\pi^* = \pi L(\pi L 1)^{-1}$. The invertibility of $L - Q$ follows from the irreducibility of $\{X(t)\}$ and Perron–Frobenius theory for ML-matrices (Seneta,[12] p. 40). It is clear from the definition of $F(\cdot)$ that $P = F(\infty)$ is the transition probability matrix of the embedded Markov chain at arrivals. When all λ_i are positive, P is also irreducible.

Now since $\{X(t)\}$ is stationary and irreducible and the intensity of the process is finite, it follows that, provided $\sum_{j=1}^k \pi_j(q_j + \lambda_j) < \infty$, the Palm probability measure for $\{X(t)\}$ coincides with the probability measure for $\{X(t)\}$ corresponding to the initial distribution (π_i^*), where $\pi_i^* = \pi_i \lambda_i / \sum_{j=1}^k \pi_j \lambda_j$ and $q_j = -q_{jj} = \sum_{i \neq j} q_{ji}$, (Rudemo,[13] Section 6). That is, the arbitrary event initial distribution is the stationary distribution of the embedded Markov chain at arrivals.

2.3. Second-order properties of intervals

When $t = 0$ is an arrival epoch and the process is interval stationary, it follows from (2) that the joint density of the inter-event times $Z_1, Z_2, \ldots, Z_n$ is

$$f(z_1, \ldots, z_n | Q, L) = \pi^* \left[\prod_{i=1}^n \{\Psi(z_i)L\}\right] 1. \tag{5}$$

Hence the marginal density and the expectation of the inter-event times Z_i are

$$f(z_i | Q, L) = \pi^*[\exp\{(Q - L)z_i\}]L1,$$

$$E(Z_i) = \pi^* \int_0^\infty z_i \exp\{(Q - L)z_i\}L1\,dz_i = \pi^*(L - Q)^{-2}L1. \tag{6}$$

Note that since π^* is the stationary distribution of P and $P = (L - Q)^{-1}L$ is a stochastic matrix, $\pi^*(L - Q)^{-1}L = \pi^*$ and $(L - Q)^{-1}L1 = 1$. Following the above argument, we can easily show that

$$E(Z_i Z_{h+i}) = \pi^*(L - Q)^{-2}L\{(L - Q)^{-1}L\}^{h-1}(L - Q)^{-2}L1, \quad i = 1, \ldots, n - 1,$$

$$h = 1, \ldots, n - i.$$

Thus

$$\text{cov}(Z_1, Z_{h+1}) = \pi^*(L - Q)^{-1}\{P^h - 1\pi^*\}(L - Q)^{-1}1. \tag{7}$$

Since $\lim_{h \to \infty} P^h = 1\pi^*$ it is clear that $\text{cov}(Z_1, Z_{h+1}) = C_h \to 0$ as $h \to \infty$. It may also be

possible to derive equation (7) using the Laplace–Stieltjes transform of the transition probability distribution matrix $F(\cdot)$.[9] When P is diagonalizable it follows from the spectral representation of a stochastic matrix that P^h can be written as

$$P^h = \beta_1^h B_1 + \cdots + \beta_k^h B_k; \quad h = 0, 1, 2, \ldots,$$

where the eigenvalues of P, $\beta_1, \ldots, \beta_k$, are such that $\beta_1 = 1$ and $|\beta_i| < 1$, $2 \leqslant i \leqslant k$ and $B_1 = 1\pi^*$ (see, for example, Çinlar,[14] p. 368). Then the spectral density function of the inter-event times of the process can be written as

$$f(\omega) = \frac{1}{2\pi}\left[1 + \frac{1}{V}\left\{\pi^*(L-Q)^{-1}\left(\sum_{i=2}^{k}\left(\frac{2\beta_i\cos\omega - 2\beta_i^2}{1 - 2\beta_i\cos\omega + \beta_i^2}\right)B_i\right)(L-Q)^{-1}\mathbf{1}\right\}\right] \tag{8}$$

for $-\pi \leqslant \omega \leqslant \pi$, where $V = C_0$ is the variance of the inter-event times.

2.4. Second-order properties of counts

The second-order properties of the counting process $\{N(t)\}$ are best described via those of the underlying Markov process $\{X(t)\}$ that directs its rate of occurrence. The transition probability matrix of $X(t)$ is given by $P(t) = \exp(Qt)$. Since the rate of occurrence is directed by $\{X(t)\}$, we can interpret the rate process $\{\Lambda(t)\}$ as a continuous-time Markov chain with states $\lambda_{X(t)}$, provided that λ is a one-to-one function. Hence $\{\Lambda(t)\}$ is a stationary irreducible continuous-time Markov chain with state space $S = \{\lambda_1, \lambda_2, \ldots, \lambda_k\}$, transition probability matrix $P(t)$, and stationary probability distribution π.

We derive the counting properties of the rate process $\{\Lambda(t)\}$ first. Suppose the system is in equilibrium at $t = 0$. Then for $t \geqslant 0$,

$$E\{\Lambda(0)\Lambda(t)\} = \sum_{i=1}^{k}\lambda_i E\{\Lambda(t)|\Lambda(0) = \lambda_i\}P\{\Lambda(0) = \lambda_i\}$$

$$= \sum_{i=1}^{k}\lambda_i\sum_{j=1}^{k}P_{ij}(t)\lambda_j\pi_i = \pi L P(t) L \mathbf{1}.$$

Now from stationarity and the equilibrium property it follows that the mean, variance and autocovariance of $\{\Lambda(t)\}$ are

$$\bar{\Lambda} = E\{\Lambda(t)\} = \pi L\mathbf{1}, \quad \sigma_\Lambda^2 = C(0) = \pi L\{I - \mathbf{1}\pi\}L\mathbf{1}, \quad C(t) = \pi L\{P(t) - \mathbf{1}\pi\}L\mathbf{1},$$

where I is the $k \times k$ identity matrix. The counting process $\{N(t)\}$ is also stationary with intensity $m(t) = \bar{\Lambda} = m = \pi L\mathbf{1}$. The covariance density of a stationary doubly stochastic Poisson process equals the autocovariance of its occurrence rate,[2] and so $\{N(t)\}$ has covariance density

$$\gamma_+(t) = C(t) = \pi L\{P(t) - \mathbf{1}\pi\}L\mathbf{1}.$$

For a stationary point process the covariance density is given by (Cox and Lewis,[15] p. 74)

$$\gamma_+(t) = m\{m_f(t) - m\} \tag{9}$$

where $m = m(t)$ is the overall intensity and $m_f(t)$ is the conditional intensity of the process. Therefore the conditional intensity of $\{N(t)\}$ is

$$m_f(t) = \frac{\pi L P(t) L \mathbf{1}}{\pi L\mathbf{1}}, \tag{10}$$

which tends to $\pi L1 = m$ as $P(t) \to 1\pi$, using the stationarity of $\{\Lambda(t)\}$. This expression for the conditional intensity function can be used to obtain various other properties of $\{N(t)\}$. It follows from equations (4.5.16) and (4.5.20) of Cox and Lewis[15] that the spectrum of counts and the Laplace transform of the variance–time curve for the MMPP are

$$g_+(\omega) = \frac{1}{\pi}\{m + \pi L P^*(i\omega)L1 + \pi L P^*(-i\omega)L1\}, \quad \omega \geqslant 0, \tag{11}$$

and

$$V^*(s) = \frac{\pi L1}{s^2} + \frac{2\pi L P^*(s)L1}{s^2} - \frac{2\pi L1 \pi L1}{s^3} \tag{12}$$

where $P^*(s)$ is the Laplace transform of the transition probability matrix $P(t)$. The variance–time curve $V(t)$ is obtained by inversion of (12). The index of dispersion is $I(t) = V(t)/mt$.

3. BIVARIATE MARKOV-MODULATED POISSON PROCESS

Though for clarity we describe only the bivariate case, the ideas of Section 2 are easily extended to the case where several MMPPs are observed. A bivariate Markov-modulated Poisson process comprises two MMPPs $\{N^{(1)}(t)\}, \{N^{(2)}(t)\}$ that are conditionally mutually independent Poisson processes given the underlying process.

3.1. Likelihood approach

Let $\{N(t)\}$ be a stationary orderly bivariate MMPP of which the marginal process $\{N^{(1)}(t)\}, \{N^{(2)}(t)\}$ are univariate MMPPs, assumed to be independent given $\{X(t)\}$. Let $Q_{k \times k}$ be the infinitesimal generator of $X(t)$ and let the rate of occurrences of the marginal processes be given by matrices $L_1 = \text{diag}(\lambda_1, \ldots \lambda_k)$ and $L_2 = \text{diag}(\lambda_1^*, \ldots, \lambda_k^*)$. Suppose that n events are observed in an interval $[0, T]$ at times $t_1 < \cdots < t_n$ and $S_1, \ldots, S_n$ is a binary sequence of the types of events of the bivariate process. Now we define the conditional probabilities, as for the univariate case,

$$\phi_{ij}(t) = P\{X(t) = j, N^{(l)}(t) = 0, l = 1, 2 | X(0) = i, N^{(l)}(0) = 0, l = 1, 2\}.$$

Let $\Phi(t)$ be the $k \times k$ matrix with entries $\phi_{ij}(t)$. Then it is straightforward to see that

$$\Phi(t) = \exp[\{Q - (L_1 + L_2)\}t].$$

Hence the likelihood for Q, L_1 and L_2 given $\{t_i, S_i\}, i = 1, \ldots, n$ can be written as

$$f(t_1, \ldots, t_n; S_1, \ldots, S_n | Q, L_1, L_2) = \pi \left[\prod_{i=1}^n \{\Phi(t_i - t_{i-1})L(i)\} \right] \Phi(T - t_n)1, \tag{13}$$

where π and 1 are defined in Section 2, $t_0 = 0$, and $L(\cdot)_{k \times k}$ is defined by

$$L(i) = \sum_{j=1}^2 \delta_{j S_i} L_j;$$

δ is a Dirac delta function. That is, $L(i)$ is the matrix of rates of occurrence of the type S_i events. Expression (13) can be maximized numerically.

3.2. Second-order properties

The superposed process, points of $N^{(1)}(t)$ and $N^{(2)}(t)$ regardless of their types, is a MMPP with mean arrival rate

$$m_s = m_1 + m_2 = \pi L_1 \mathbf{1} + \pi L_2 \mathbf{1} = \pi L_s \mathbf{1},$$

where m_1 and m_2 are the mean rates of occurrences of type I and type II events respectively, and $L_s = L_1 + L_2$ is the matrix of rates of occurrence of the superposed process. The second-order properties of this process do not need to be described here, as it is a univariate MMPP. For the bivariate process, the cross-covariance densities can be defined as (Cox and Isham,[16] p. 120)

$$\gamma_1^{(2)}(t) = \lim_{\delta_1, \delta_2 \to 0} (\delta_1 \delta_2)^{-1} \operatorname{cov} \{N^{(2)}(t, t + \delta_2), N^{(1)}(-\delta_1, 0)\}, \quad t \neq 0.$$

Then it can be shown, by conditioning on the underlying process, that

$$\gamma_i^{(j)}(t) = \pi L_i \{P(t) - \mathbf{1}\pi\} L_j \mathbf{1}, \quad i, j = 1, 2, \tag{14}$$

with the usual extension when $t = 0$ for $i = j$, where $P(t)$ is the transition probability matrix of $\{X(t)\}$. Now it is fairly straightforward to show, using equation (5.6) of Cox and Isham,[16] that the cross-intensity function of the bivariate MMPP is given by

$$h_i^{(j)}(t) = \lim_{\delta_1, \delta_2 \to 0} \delta_2^{-1} \Pr \{N^{(j)}(t, t + \delta_2) > 0 | N^{(i)}(-\delta_1, 0) > 0\} = \frac{\pi L_i P(t) L_j \mathbf{1}}{\pi L_i \mathbf{1}}, \quad i, j = 1, 2 \tag{15}$$

which converges to m_j as $t \to \infty$. The cross-spectrum of counts of the bivariate process is given by

$$
\begin{aligned}
g_1^{(2)}(\omega) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega t} \gamma_1^{(2)}(t) \, dt \\
&= \frac{1}{2\pi} \int_0^{\infty} \{\pi L_1 \{P(t) - \mathbf{1}\pi\} L_2 \mathbf{1} + \pi L_2 \{P(t) - \mathbf{1}\pi\} L_1 \mathbf{1}\} \cos \omega t \, dt \\
&\quad + i \frac{1}{2\pi} \int_0^{\infty} \{\pi L_2 \{P(t) - \mathbf{1}\pi\} L_1 \mathbf{1} - \pi L_1 \{P(t) - \mathbf{1}\pi\} L_2 \mathbf{1}\} \sin \omega t \, dt \\
&= c(\omega) + iq(\omega), \tag{16}
\end{aligned}
$$

since $\gamma_1^{(2)}(-t) = \gamma_2^{(1)}(t)$. The real and imaginary components of this complex-valued function indicate the relative phases of the fluctuations in the process of events of type I and type II. If the underlying process $\{X(t)\}$ is reversible then $g_1^{(2)}(\omega)$ is real. The covariance-time function of the bivariate MMPP may then be worked out from (3.18) of Cox and Lewis,[17] as

$$V^{(12)}(t) = \int_0^t (t - v) \{\pi L_1 \{P(v) - \mathbf{1}\pi\} L_2 \mathbf{1} + \pi L_2 \{P(v) - \mathbf{1}\pi\} L_1 \mathbf{1}\} dv. \tag{17}$$

4. NON-REGULAR LIKELIHOOD RATIO STATISTICS

There are model comparisons for MMPPs for which the usual asymptotic properties of likelihood ratio statistics fail to hold. One form of non-regularity arises when one or more components of the true parameter lie on the boundary of the parameter space. For example, a renewal Cox process with Markovian intensity (RCM),[7] arises from an MMPP with $k = 2$ when $\lambda_1 = 0$. If we wish to test the hypothesis $\lambda_1 = 0$ the usual asymptotic distribution of the likelihood ratio statistic cannot be applied, as one component of the true parameter lies on a boundary.

Table I. Percentage points of the likelihood ratio statistics for testing $\lambda_1 = 0$ for MMPP(2), and $\lambda_1 = \lambda_3 = 0$ for the MMPP(3) based on 1000 simulations. Expected number of events for M_1, M_2 and M_3 are 100, 200 and 500

Percentage $p \times 100\%$	H_0: $\lambda_1 = 0$ for MMPP(2)				H_0: $\lambda_1 = \lambda_3 = 0$ for MMPP(3)			
	Simulated values				Simulated values			
	M_1	M_2	M_3	$\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$	M_1	M_2	M_3	$\frac{1}{4}\chi_0^2 + \frac{1}{2}\chi_1^2 + \frac{1}{4}\chi_2^2$
70	0·18	0·26	0·44	0·275	0·28	0·79	1·08	1·06
75	0·28	0·37	0·55	0·455	0·45	1·07	1·35	1·35
80	0·46	0·55	0·76	0·708	0·65	1·36	1·79	1·73
85	0·78	0·78	1·04	1·074	1·16	1·81	2·31	2·23
90	1·10	1·22	1·45	1·642	1·78	2·51	3·02	2·94
95	1·86	2·27	2·37	2·706	2·90	3·68	4·25	4·23
99	3·73	3·93	4·40	5·412	5·67	5·76	7·58	7·28

Problems of this type have been studied by Moran,[18] Self and Liang,[19] Smith[20] and others. Self and Liang[19] describe the theoretical asymptotic distributions of the likelihood ratio statistic for such cases. Our aim in this section is to see by simulation how much data are needed before the asymptotic results apply to MMPPs.

We first consider the case when one parameter is on the boundary. Suppose that we wish to test the hypothesis H_0: $\lambda_1 = 0$ for a two state MMPP, denoted MMPP(2). The results of Self and Liang[19] show that the appropriate asymptotic distribution of the likelihood ratio statistic, W, is the mixture $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$, where χ_0^2 is the distribution with mass one at zero and $\frac{1}{2}:\frac{1}{2}$ are mixture probabilities. We shall compare the simulated distribution of the likelihood ratio statistic with its theoretical distribution.

One way of stimulating MMPP is to define an event to be either an arrival from $N(t)$ or a transition of $X(t)$. Given that $X(t)$ is in state i, the next event is an arrival with probability $\lambda_i(\lambda_i + q_i)^{-1}$ and a transition to state j with probability $q_{ij}(\lambda_i + q_i)^{-1}$, where $q_i = \sum_{j \neq i} q_{ij}$. The time to the next event is an exponential variate of parameter $(\lambda_i + q_i)$.

The process is simulated, for some predetermined parameter values, in three intervals in which the average number of points are $M_1 = 100$, $M_2 = 200$ and $M_3 = 500$. We present the results of one such simulation using parameters $(q_1, q_2, \lambda_1, \lambda_2) = (0·10, 0·20, 0·0, 1·5)$. The log-likelihood functions for the two models are maximized, giving W. Results based on 1000 simulations for each of M_1, M_2 and M_3 are summarized in Table I, which shows the simulated percentage points w_p of W together with the corresponding theoretical values, where $P\{W \leqslant w_p\} = p$. The simulated values approach the theoretical ones as the average number of events increases. Although the simulated w_p's for M_3 at $p = 0·9$ and $0·95$ seem to be a bit away from their asymptotic values, the simulated levels of significance at $1·642$ and $2·706$ are 91 per cent and 96 per cent, which is quite reasonable.

Suppose now that we observe a Markov-modulated Poisson process with $k = 3$, denoted MMPP(3), in which the transitions are only possible between neighbouring states. Let q_{12}, q_{21}, q_{23} and q_{32} be the transition rates and λ_1, λ_2 and λ_3 be the arrival rates, where $q_{12} \neq q_{32}$. We call this model 1. When $\lambda_1 = \lambda_3 = 0$ the process reduces to another non-regular case with two parameters on the boundary of the parameter space. In this case, following Self and Liang,[19] the approximate distribution of the likelihood ratio statistic W for testing $\lambda_1 = \lambda_3 = 0$ is $\frac{1}{4}\chi_0^2 + \frac{1}{2}\chi_1^2 + \frac{1}{4}\chi_2^2$. Under the null hypothesis the process has just five parameters (model 2). In order to compare the significance points of W with the corresponding theoretical asymptotic values, we generate data

from model 2. Table I gives percentage points w_p of W based on 1000 simulations for M_1, M_2 and M_3. The table shows that the percentage points are more accurate for M_3 than those for M_1 and M_2. The simulated distributions of W for other parameter configurations show the same and are in good agreement, for M_3, with the asymptotic theoretical distribution of W. This suggests that the theoretical results of Self and Liang[19] apply to MMPP data with at least about 500 events.

Another form of non-regularity arises when we want to test H_0: $\lambda_1 = \lambda_2$ for a MMPP(2). Under the null hypothesis the process is a homogeneous Poisson process and hence some parameters become unidentifiable. The same problem emerges in testing H_0: $\lambda_1 = \lambda_2 = 0$ for the MMPP(3) described earlier. The asymptotic properties of the maximum likelihood estimators for these particular problems are hard to deal with although the general approach given by Davies[21,22] may be of use. For these and other situations Monte Carlo tests seem likely to be valuable tools.

5. DATA ANALYSIS

5.1. Data and models

In this section we apply the models described in Sections 2 and 3 to a set of artificial exposure data for long-range atmospheric transport of radionucleides. The data were generated in the course of using a computer model to examine possible long-range exposures to radioactivity due to accidental releases.[23,24] The computer model, described briefly by Davison and Smith,[25] simulates the atmospheric transport dispersion and deposition of radionucleides at distances 100–1000 km or so away from notional sources in Western Europe, based on 'present weather' observations. The data considered here are the times of releases leading to exposures to time-integrated air concentrations of pollutant at receptors 100 and 300 km north of a notional source at Ispra in northern Italy. Releases are deemed to occur every three hours throughout 1976. The times of releases leading to exposures are available in 3 h units for a period of 8640 hours.

5.2. Univariate analysis

The times of exposures are believed to depend on the weather, *windy* weather being associated with isolated exposures and *calm* weather leading to more clustered exposures. In 'blocking anticyclones', such as can occur over the British Isles in the summer, for example, plumes will move more slowly and hence tend to be wider than in the higher winds associated with the easterly movement of Atlantic depressions. From Figure 1, which displays the cumulative plots of two time series of notational exposures, it seems that the points occur mostly in small clusters. In view of this, we initially interpret the events of exposures as an MMPP(2) whose underlying process is a stationary Markov chain representing the background state of the environment, ignoring seasonality, of which the data show no evidence. The parameters of this model are

$$Q = \begin{bmatrix} -q_1 & q_1 \\ q_2 & -q_2 \end{bmatrix}, \quad L = \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix}.$$

The rate process $\Lambda(t)$ is a stationary continuous-time Markov chain with state space $S = \{\lambda_1, \lambda_2\}$ and generator Q. The transition probability function $P(t) = e^{Qt}$ is easily obtained using the spectral representation of Q. The stationary distribution π of the rate process is $\pi = (q_2, q_1)(q_1 + q_2)^{-1}$.

We begin by fitting an MMPP(2) to the data for the first series, at 100 km, for which $n = 452$. The first column of Table II gives the maximum likelihood estimates and their standard errors when this model is fitted by applying routine AMOEBA of Press *et al.*[26] to (3). Standard errors
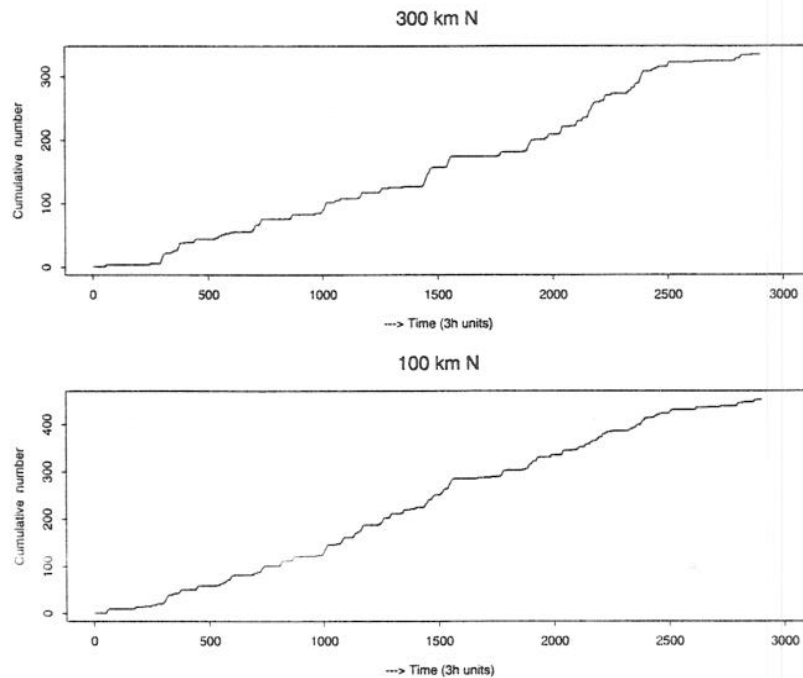
Figure 1. Empirical cumulative plots for the number of exposures to time-integrated air concentration of I131(p) at the two receptors 100,300 km north of Ispra during 1976

are based on the observed information matrix, which was obtained by numerical differentiation. The mean sojourn times of the rate process in state 1 and state 2 are $1/\hat{q}_1 = 24 \cdot 69$ and $1/\hat{q}_2 = 8 \cdot 78$ with respective estimated rates $\hat{\lambda}_1 = 0 \cdot 005$ and $\hat{\lambda}_2 = 0 \cdot 582$. The standard errors suggest that $\lambda_1 = 0$, in which case our model reduces to a special case which is simultaneously a renewal process and a Cox process.[3] This is the RCM model studied by Smith and Karr.[7] The maximum likelihood estimates of the parameters of the RCM model and their standard errors, given in Table II, are very close to those of the MMPP(2).

The null hypothesis H_0: $\lambda_1 = 0$ throws one parameter on the boundary of the parameter space.

Table II. Maximum likelihood estimates and their standard errors in parentheses. The final row gives the values of the log-likelihood $l(\theta)$

	100 km		300 km		100 and 300 km Bivariate models	
	MMPP(2)	RCM	MMPP(2)	RCM	MMPP(2)	RCM
q_1	0·041(0·007)	0·047(0·006)	0·030(0·004)	0·030(0·004)	0·038(0·005)	0·046(0·005)
q_2	0·114(0·017)	0·123(0·017)	0·117(0·014)	0·116(0·019)	0·114(0·016)	0·125(0·015)
λ_1	0·005(0·003)	–	0·001(0·002)	–	0·006(0·004)	–
λ_2	0·582(0·035)	0·575(0·035)	0·575(0·043)	0·576(0·043)	0·593(0·032)	0·584(0·028)
λ_1^*	–	–	–	–	0·001(0·001)	–
λ_2^*	–	–	–	–	0·447(0·031)	0·432(0·025)
$l(\theta)$	−1029·11	−1030·14	−780·58	−780·76	−822·69	−831·24

Table III. Estimated and theoretical values of the second moments of intervals

| | 100 km | | | | | 300 km |
	MMPP(2)	RCM	Empirical	MMPP(2)	RCM	Empirical
$E(Z)$	6·41	6·32	6.36	8·46	8·55	8·59
$V(Z)$	239·75	235·83	226·72	499·05	533·63	569·37
CV	2·41	2·43	2·37	2·64	2·70	2·76

The appropriate asymptotic distribution for the likelihood ratio statistic is $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$,[19] whose 95 per cent significance point is 2·7. The observed value of the likelihood ratio statistic W for testing $\lambda_1 = 0$ is 2·06, which is significant at about the 7 per cent level. Moreover, when a three-state model is fitted the log-likelihood increases only by 0·01, which suggests that a two-state model is adequate.

The expectation of the times between events may be worked out, from (6), as

$$E(Z) = \frac{1}{\alpha^2\beta^2(q_1\lambda_2 + q_2\lambda_1)(\beta - \alpha)}[q_2\lambda_1^2\{\alpha^2(q_2 + \lambda_2 - \beta) - \beta^2(q_2 + \lambda_2 - \alpha)\}$$
$$- q_1\lambda_2^2\{\beta^2(q_2 + \lambda_2 - \beta) - \alpha^2(q_2 + \lambda_2 - \alpha)\}$$
$$+ (\alpha^2 - \beta^2)\lambda_1\lambda_2\{(q_2 + \lambda_2 - \beta)(q_2 + \lambda_2 - \alpha) + q_2\}] \tag{18}$$

where α, β are the roots of the quadratic equation

$$v^2 - (q_1 + q_2 + \lambda_1 + \lambda_2)v + q_1\lambda_2 + q_2\lambda_1 + \lambda_1\lambda_2 = 0.$$

When $\lambda_1 = 0$ (18) simplifies considerably. Similarly $E(Z^2)$ and hence the variance $V(Z)$ can be found. The estimated mean, variance and coefficient of variation $CV = V(Z)^{1/2}/E(Z)$ of the times between events of the process, together with their theoretical counterparts for MMPP(2) and RCM models, are given in Table III. It is clear that the process is over-dispersed relative to the Poisson process, for which the coefficient of variation equals one. Finally, the expression for the spectral density function of the times between events is given in (8). Although we do not show them, the estimated and theoretical spectrum agree closely and depart only slightly from the spectrum for a renewal process.

It follows from (10) that the intensity and the conditional intensity of our MMPP(2) are given by

$$m = \pi L1 = (q_2\lambda_1 + q_1\lambda_2)/(q_1 + q_2), \tag{19}$$

$$m_f(t) = m + \left\{\frac{q_1q_2(\lambda_1 - \lambda_2)^2 e^{-(q_1 + q_2)t}}{(q_1 + q_2)(q_1\lambda_2 + q_2\lambda_1)}\right\}, \tag{20}$$

respectively. The conditional intensity function is greater than the unconditional intensity m for small t and decreases exponentially to m unless $\lambda_1 = \lambda_2$, when $m_f(t) = m$, i.e. we have a Poisson process. Figure 2 shows the theoretical and estimated conditional intensity functions (Cox and Lewis,[15] p. 121) which match very well. Smoothed estimates of $m_f(t)$ are obtained using a Daniell weight function. The confidence bands are pointwise simulation envelope based on 39 simulations of the fitted process. Now from (20) or equivalently from (11) the spectral density function
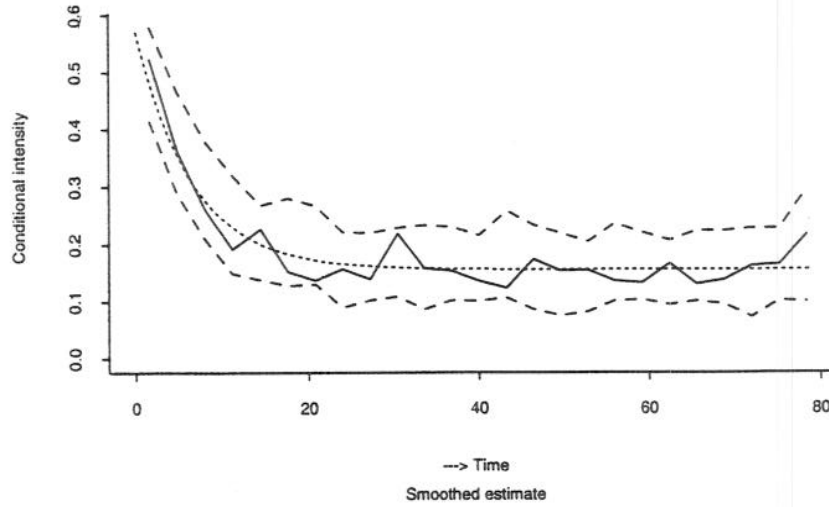
Figure 2. Estimated and theoretical conditional intensity functions for the times of exposures at 100 km north

of counts for an MMPP(2) can be shown to have the form

$$g + (\omega) = \frac{m}{\pi}\left\{1 + \frac{2q_1q_2(\lambda_1 - \lambda_2)^2}{((q_1 + q_2)^2 + \omega^2)(q_1\lambda_2 + q_2\lambda_1)}\right\}, \quad \omega \geqslant 0. \tag{21}$$

When $\lambda_1 = \lambda_2$ this reduces to the constant spectrum of counts of a Poisson process.

Finally, (12) and an inverse Laplace transform gives us the variance–time curve and the index of dispersion of the process as

$$V(t) = mt + \left\{\frac{2mq_1q_2(\lambda_1 - \lambda_2)^2}{(q_1 + q_2)^2(q_1\lambda_2 + q_2\lambda_1)}\right\}\left\{t - \frac{1}{(q_1 + q_2)}(1 - e^{-(q_1 + q_2)t})\right\}, \tag{22}$$

$$I(t) = 1 + \left\{\frac{2q_1q_2(\lambda_1 - \lambda_2)^2}{(q_1 + q_2)^2(q_1\lambda_2 + q_2\lambda_1)}\right\}\left\{1 - \frac{1}{(q_1 + q_2)t}(1 - e^{-(q_1 + q_2)t})\right\}. \tag{23}$$

The spectrum of counts and the variance–time curve give a satisfactory fit for the data. The value of $I(t)$ suggests that the process is over-dispersed relative to the Poisson process. Properties of RCM process can be obtained by setting $\lambda_1 = 0$ in expressions (18) to (23).

Also given in Table II are the parameter estimates of MMPP(2) and RCM models, when fitted to the exposure times at the receptor 300 km away from the source, for which $n = 334$. The estimates suggest that the distance does not reduce either the arrival rate or the mean sojourn times in state 2, but instead that it increases the mean sojourn times in state 1. The second moment results of the times between events for this series are shown in Table III.

5.3. Bivariate analysis

In order to describe the joint properties of the two exposure time series, here we consider a special bivariate MMPP whose underlying process $\{X(t)\}$ is an irreducible two-state continuous time
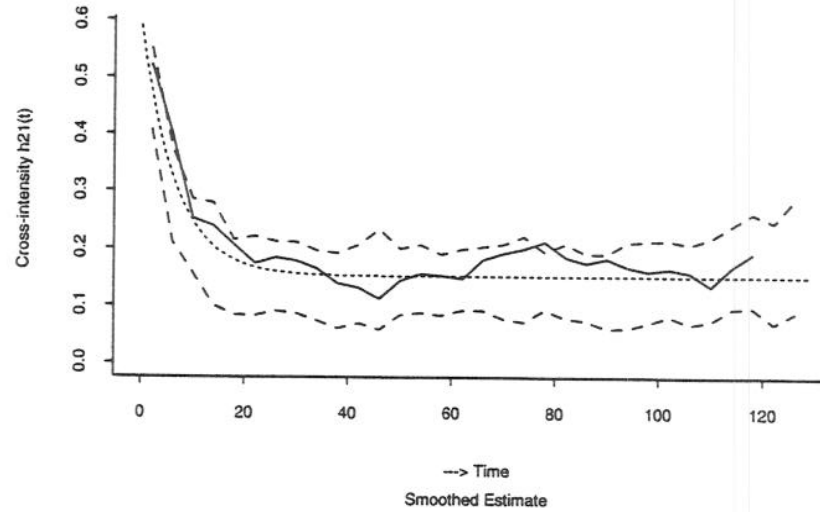
Figure 3. Estimated and theoretical cross-intensity functions for exposure data sets

Markov chain with transition rates q_1, q_2. Let $\{N_1(t)\}$ and $\{N_2(t)\}$ be the marginal processes of the bivariate MMPP $\{N(t)\}$ with respective rate matrices

$$L_1 = \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} \quad L_2 = \begin{bmatrix} \lambda_1^* & \\ & \lambda_2^* \end{bmatrix}.$$

The expressions for the second-order functions: cross-covariance density for $t \neq 0$, cross-intensity function, covariance-time function and cross-spectrum of counts for this bivariate MMPP, are (from equations (14) to (17))

$$\gamma_1^{(2)}(t) = \frac{q_1 q_2 (\lambda_2 - \lambda_1)(\lambda_2^* - \lambda_1^*) \, e^{-(\lambda + \mu)t}}{(q_1 + q_2)^2}$$

$$h_1^{(2)}(t) = \frac{(q_2 \lambda_1^* + q_1 \lambda_2^*)}{(q_1 + q_2)} + \frac{q_1 q_2 (\lambda_2 - \lambda_1)(\lambda_2^* - \lambda_1^*) \, e^{-(q_1 + q_2)t}}{(q_1 + q_2)(q_1 \lambda_2 + q_2 \lambda_1)},$$

$$V^{(12)}(t) = \frac{2 q_1 q_2 (\lambda_2 - \lambda_1)(\lambda_2^* - \lambda_1^*)}{(q_1 + q_2)^3} \left[t - \frac{1}{(q_1 + q_2)} \{1 - e^{-(q_1 + q_2)t}\} \right],$$

$$g_1^{(2)}(\omega) = \frac{m_1}{2\pi} \left\{ \frac{2 q_1 q_2 (\lambda_2 - \lambda_1)(\lambda_2^* - \lambda_1^*)}{((q_1 + q_2)^2 + \omega^2)(q_1 \lambda_2 + q_2 \lambda_1)} \right\}.$$

The transition probability function $P(t)$ and the stationary distribution π show explicitly that $\{X(t)\}$ satisfies the detailed balanced conditions and hence is reversible, which leads to a real cross-spectrum.

This bivariate model is fitted to data on exposures at the receptors 100 and 300 km north of the notional source at Ispra, Italy. The likelihood (13) is maximized to obtain estimates of Q, L_1 and L_2, and the result, given in Table II, is consistent with the univariate case.

To assess goodness of fit we use the second-order functions as for the univariate case. The estimated and the theoretical cross-intensity functions are shown in Figure 3. The confidence bands
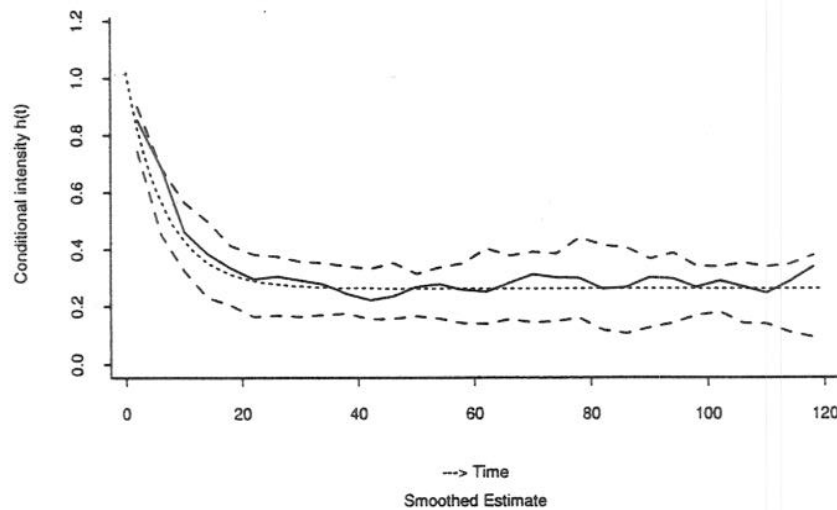
Figure 4. Estimated and theoretical conditional intensity functions for the pooled process of exposures

are simulation envelopes generated as the pointwise maximum and minimum cross-intensities for 19 simulations from the fitted process. The estimated function $\tilde{h}_2^{(1)}(t)$ is in good agreement with its theoretical counterpart. Figure 4 displays the conditional intensity $h^{(\cdot)}(t)$ of the superposed process, together with a pointwise confidence band based on 19 simulations, which gives an even better fit. Estimation of $h_2^{(1)}(t)$ is as proposed by Cox and Lewis (reference 17, Section 6.3). Smoothed estimates with a Daniell weight function have been used in all cases. The normalized cross-spectrum and the covariance-time functions give reasonably good fit for the data. The second-order functions demonstrate the adequacy of fit of the model described here to the bivariate point process of exposure times at the two receptors.

The null hypothesis H_0: $\lambda_1 = \lambda_1^* = 0$ puts two parameters on the boundary of the parameter space. The test statistic, whose asymptotic null distribution is $\frac{1}{4} + \frac{1}{2}\chi_1^2 + \frac{1}{4}\chi_2^2$, takes the value 17·1, which is strong evidence of exposures in state 1. However, the likelihood ratio test statistic for testing H_0: $\lambda_1 = 0$ takes the value 16·7 and rejects the null hypothesis at 5 per cent significance level whereas for testing H_0: $\lambda_1^* = 0$ it takes 0·2 which is barely significant. This leads us to the conclusion that the exposures take place in both states at 100 km north of the source but only in one state at the receptor 300 km away from the source. This is plausible as the frequency of exposures drops with the distance.

One aspect of the data which is not addressed in our analysis is the discreteness, as the exposure data are recorded in three-hour intervals. Perhaps a better approach is to model exposure times as a binary time series generated by an unobserved point process. Modelling exposure times at number of receptors using discrete-time MMPPs, for which the likelihood and second-order properties can be calculated in a manner similar to that described in this paper, is currently under investigation.[27]

REFERENCES

1. Cox, D. R. 'Some statistical methods connected with series of events (with Discussion)', *Journal of the Royal Statistical Society, Series B*, **27**, 129–164 (1955).

2. Bartlett, M. S. 'The spectral analysis of point processes (with Discussion), *Journal of the Royal Statistical Society, Series B*, **25**, 264–296 (1963).

3. Kingman, J. F. C. 'On doubly stochastic Poisson processes, *Proceedings of the Royal Cambridge Philosophical Society*, **60**, 923–960 (1964).

4. Grandell, J. *Doubly Stochastic Poisson Processes*, Lecture notes in Mathematics No. 529, Springer-Verlag, Berlin, 1976.

5. Rudemo, M. 'Doubly stochastic Poisson processes and process control', *Advances in Applied Probability*, **4**, 318–338 (1972).

6. Snyder, D. L. and Miller, M. I. *Random Point Processes in Time and Space*, 2nd edn, Springer-Verlag, New York, 1991.

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

8. Neuts, M. F. *Structured Stochastic Matrices of M/G/1 Type and Their Applications*, Marcel Dekker, New York, 1989.

9. Meier-Hellstern, K. S. 'A fitting algorithm for Markov-modulated Poisson processes having two arrival rates', *Euro. J. Oper. Res.*, **29**, 370–377 (1987).

10. Smith, R. L. 'Contribution to the discussion of Stern, R. D. and Coe, R. (1984). A model-fitting analysis of daily rainfall data (with Discussion), *Journal of the Royal Statistical Society, Series A*, **147**, 24–25 (1984).

11. Daley, D. J. and Vere-Jones, D. *An Introduction to the Theory of Point Processes*, Springer-Verlag, New York, 1988.

12. Seneta, E. *Non-Negative Matrices*, George Allen & Unwin, London, 1973.

13. Rudemo, M. 'Point processes generated by transitions of Markov Chains', *Advances is Applied Probability*, **5**, 262–286 (1973).

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

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

16. Cox, D. R. and Isham, V. *Point Processes*, Chapman and Hall, London, 1980.

17. Cox, D. R. and Lewis, P. A. W. 'Multivariate point processes', *Proceedings of the 6th Berkeley Symposium*, 401–448 (1972).

18. Moran, P. A. P. 'Maximum likelihood estimation in non-standard problems', *Proceedings of the Cambridge Philosophical Society*, **70**, 441–450 (1971).

19. Self, S. G. and Liang, K.-Y. 'Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions', *Journal of the Americal Statistical Association*, **82**, 605–610 (1987).

20. Smith, R. L. 'A survey of nonregular problems', *Bulletin of the 47th Session of the International Statistical Institute, Paris*, **3**, 353–372 (1989).

21. Davies, R. B. 'Hypothesis testing when a nuisance parameter is present only under the alternative', *Biometrika*, **64**, 274–254 (1977).

22. Davies, R. B. 'Hypothesis testing when a nuisance parameter is present only under the alternative', *Biometrika*, **74**, 33–43 (1987).

23. ApSimon, H. M., Goddard, A. J. H. and Wrigley, J. 'Long-range transport of radioisotopes: I, The MESOS model', *Atmosp. Environ.*, **19**, 99–111 (1985).

24. ApSimon, H. M. and Davison, A. C. 'A statistical model for deriving probability distributions of contamination for accidental releases', *Atmosp. Environ.*, **20**, 1249–1259 (1986).

25. Davison, A. C. and Smith, R. L. 'Models for exceedances over high thresholds (with Discussion), *Journal of the Royal Statistical Society, Series B*, **52**, 393–442 (1990).

26. Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vettering, W. T. *Numerical Recipes. The Art of Scientific Computing*, Cambridge University Press, 1986.

27. Davison, A. C. and Ramesh, N. I. 'Some models for discretized series of events', in preparation. 1993.