Ann. Inst. Statist. Math.
34 (1982), Part B, 189-207

SOME EXAMPLES OF STATISTICAL ESTIMATION APPLIED TO EARTHQUAKE DATA

I. CYCLIC POISSON AND SELF-EXCITING MODELS

D. VERE-JONES AND T. OZAKI

(Received Mar. 23, 1981; revised Oct. 1, 1981)

Abstract

Likelihood methods are described for fitting cyclic Poisson and Hawkes' self-exciting models to Kawasumi's historical earthquake series and to more recent data supplied by the Japan Meteorological Agency. Identification of the model is discussed from the standpoint of an entropy maximization principle. The cyclic effect is shown to be not statistically significant after clustering has been allowed for; its physical significance therefore remains questionable.

1. Introduction

This paper describes the fitting of two types of point process model to earthquake data: the cyclic Poisson process; and a general version of the "self-exciting" process introduced by Hawkes [7], [8].

The cyclic Poisson process is perhaps the simplest model to embody a periodic effect of the type postulated by Kawasumi [10] in his analysis of historical earthquake records in the South Kwanto area of Japan. In view of the important implications of this work, we give further analysis of Kawasumi's data in Section 2, in the context of the general problem of detecting and fitting a cyclic effect in point process data when the frequency of the cyclic effect is unknown a priori.

In Section 3 the periodic model is contrasted with a self-exciting model in which the complete intensity function

$$\lambda_\infty(t)dt = \mathrm{E}\,[dN(t)\,|\,t_i:\ -\infty < t_i < t]\ ,$$

representing the instantaneous intensity given the times t_i of all past earthquakes, is parameterized according to the proposal of Akaike (Ozaki and Akaike [17]) to take

$$\lambda_\infty(t)=\mu+\int_{-\infty}^{t} g(t-u)dN(u) \tag{1.1}$$

with

$$g(t)=\sum_{m=0}^{M} \alpha_m t^m e^{-\beta t}\ , \tag{1.2}$$

and suitable restrictions on α_m and β to ensure that

$$\sum_{0}^{M} \alpha_m t^m \geqq 0 \qquad \text{for } t\geqq 0 \tag{1.3}$$

and

$$\int_0^\infty g(t)dt<1\ . \tag{1.4}$$

Loosely speaking, the function $g(t)$ describes the decay of clustering or aftershock activity after an initial event. The particular parameterization chosen is sufficiently flexible to accommodate situations where the counting spectrum may show a peak away from zero frequency, and avoids the numerical difficulty which can arise in maximum likelihood computations with the alternative parameterization

$$g(t)=\sum_{j=1}^{J} \alpha_j e^{-\beta_j t}$$

suggested by Hawkes [7] and Hawkes and Adamopoulos [9] (see Ozaki [16]). In Section 3 the model (1.2) is fitted to Kawasumi's data, to selected earthquake data supplied by the Japan Meteorological Agency, and to some simulated data sets.

Both models are of the general type described by Vere-Jones [20] for which the conditional intensity function (conditional risk) has a form simple enough to allow the method of maximum likelihood to be used for parameter estimation, starting from the formula (Rubin [18]) for the log-likelihood*)

$$\log L=\sum_{1}^{N} \log \lambda(t_i)-\int_0^T \lambda(t)dt\ . \tag{1.5}$$

Note that the conditional intensity $\lambda(t)$ appearing here differs from the complete intensity in that conditioning is taken from the start of the observation period instead of from the infinite past; thus in (1.5)

$$\lambda(t)dt=\mathrm{E}\,[dN(t)\,|\,t_i\colon\ 0\leqq t_i<t]\ .$$

This is an important distinction in principle, but the two intensities coincide for the Poisson process (lack of memory property) while the

*) Logarithms to base e.

numerical difference between them is slight for the self-exciting model because of the exponential term in the function $g(t)$. In the sequel, in dealing with the self-exciting process, we shall approximate $\lambda(t)$ by an expansion of the form (1.1) with the integral taken from 0 to t.

Under wide circumstances, which certainly include the models treated in the present paper, the use of maximum likelihood methods assures the asymptotic efficiency of the resulting estimates (see Ogata [15], Kutoyants [12]). This is in contrast to spectral methods, which will not be fully efficient when, as with many point process models, the structure of the process is not fully expressed through its second order properties. Maximum likelihood methods also allow some comparison between models to be made using the AIC values defined by

$$\text{AIC} = -2\log L + 2k \tag{1.6}$$

where L is maximum likelihood (i.e. the likelihood of the fitted model) and k is the number of fitted parameters. This follows the entropy maximization principle set out by Akaike [1]. Although the likelihood ratios do not always follow the same distribution under the null hypothesis, so that tests based on the likelihood ratios should be approached with caution, the AIC values nevertheless provide a useful heuristic guide to the models most likely to provide effective predictions.

This paper provides the statistical background for fitting two of the models suggested for earthquake data in Vere-Jones [20]. The third model suggested there will be discussed in a forthcoming paper with Y. Ogata.

2. The cyclic Poisson process

2.1. *Maximum likelihood estimation*

Following Lewis [13] we take the time-dependent intensity of the process in the form

$$\lambda(t) = \exp\{\alpha + \rho \sin(\omega_0 t + \theta)\} \tag{2.1}$$

where e^α is a measure of the average rate of occurrence of the process, ρ determines the proportional amplitude of the cyclic fluctuations, and ω_0, θ describe the frequency and phase of the cyclic term. Likelihood equations based on (1.5) take the form (with $\Lambda = \log L$),

$$\frac{\partial \Lambda}{\partial \alpha} = \int_0^t \frac{\partial \log \lambda}{\partial \alpha} \{dN(t) - \lambda(t)dt\} \ . \tag{2.2}$$

For the model (2.1) they have been studied by Lewis [14], who showed that if we are willing to ignore the edge terms introduced by the fact that T may not be an exact multiple of $2\pi/\omega$, the equations for α, ρ and θ can be reduced to the following,

$$N = Te^{\alpha} I_0(\rho) \tag{2.3a}$$

$$S(\omega) \equiv \sum_1^N \sin \omega t_i = Te^{\alpha} \cos \theta I_1(\rho) \tag{2.3b}$$

$$C(\omega) \equiv \sum_1^N \cos \omega t_i = Te^{\alpha} \sin \theta I_1(\rho)\ , \tag{2.3c}$$

where $I_j(\rho)$ denotes the modified Bessel function of order j, viz.

$$I_j(\rho) = \frac{1}{2\pi} \int_0^{2\pi} \cos ju \exp(\rho \cos u) du \qquad (j=0, 1, 2, \cdots)\ .$$

Similarly, if we restrict attention to the terms of leading order in T, the equation for the unknown frequency ω can be written

$$\cos \theta \sum t_i \cos \omega t_i - \sin \theta \sum t_i \sin \omega t_i = 0\ . \tag{2.4}$$

Simplifying, we see from (2.3) that the amplitude ρ is determined by the value of the quantity

$$[S^2(\omega) + C^2(\omega)]/N = (2\pi T/N) I_T(\omega) \tag{2.5}$$

where $I_T(\omega)$ is the value at ω of the Bartlett periodogram (Bartlett [2]; Cox and Lewis [3]). Also (2.4) is just the requirement that $I_T'(\omega)=0$, i.e. that the estimate of frequency corresponds to a stationary value (maximum) of the periodogram. At least for this model, therefore, the periodogram plays much the same role as it does for the more familiar situation of a periodic signal in white noise, $\lambda(t)$ playing the role of the signal, and $dZ(t) = dN(t) - \lambda(t)dt$ playing the role of the white noise.

In seeking a maximum of the periodogram, attention must be restricted to a finite frequency interval of the form (ε, Ω), where $\varepsilon > 0$ and $\Omega < \infty$ are determined a priori. Without such a constraint, the search for a maximum becomes meaningless, as it follows directly from the properties of almost periodic functions that no matter what individual t_i values appear, there will be values of ω for which (2.5) approaches arbitrarily close to its theoretical maximum of N. The need to bound the interval away from 0 arises only because the mean term has not been removed from the periodogram, a point which can easily be corrected in practice. With these two restrictions, an analysis along the lines of Hannan [6] shows that, provided the true frequency itself lies within (ε, Ω), the frequency $\hat{\omega}_T$ corresponding to the maximum of the periodogram over this range is a strongly consistent estimate of the true frequency, with a standard error which is $O(T^{-3/2})$ as $T\to\infty$. Details are given in Vere-Jones [21].

The choice of the upper bound is arbitrary. Often, however, a natural choice will be the frequency $\bar{\Omega}=\pi N/T$, corresponding to one half-cycle over the average time interval between events. This has the character of a Nyquist frequency, with the attractive feature that the range $(0, \bar{\Omega})$ then contains just $[N/2]$ of the special frequencies $\omega_k=2\pi k/T$, for which the periodogram ordinates are asymptotically independent (Bartlett [2]). As Bartlett points out, it is desirable that the number of such frequencies included in the range should roughly match the number of observations.

2.2. *Testing for the significance of the cyclic term*

If the frequency of the cyclic term is known a priori, as in the examples of a daily or weekly cycle described by Lewis, a simple test can be based on the fact that the individual periodogram ordinates are asymptotically exponentially distributed (χ^2 on 2 d.f.). This argument cannot be applied when the frequency is unknown a priori, as we are concerned then with testing the *maximum* of the periodogram over a specified frequency range. Nor is it possible to use standard asymptotic likelihood ratio theory, as the null hypothesis (Poisson with constant rate) is not obtained by fixing values of the parameters in the alternative hypothesis (thus: ω and θ are undefined under the null hypothesis). The exact distribution of the periodogram maximum is not known even in the Gaussian case, still less in the present situation.

In looking for an approximation, a starting point is provided by Bartlett's observation (Bartlett [2]) that the periodogram ordinates at the frequencies ω_k are approximately independent and exponentially distributed with (under the null hypothesis of a constant rate Poisson process) an expected value $\lambda/2\pi$. This suggests that the maximum of the periodogram in this range might have a distribution not greatly different from that of the maximum of $N/2$ such independent exponential variates. A more satisfactory approach to problems of this type has been suggested recently by Davies [4], who uses level crossing theory to find approximations to the distribution of the maximum. In Davies [5] this method is applied to the classical time series problem, and his approach can be extended with only trivial changes to the point process context. A slight extension of the results of Sharpe [19] on the level crossings of a χ^2-process leads to the conclusion that V_α, the $100\alpha\%$ significance level for a test based on $I_T(\hat{\omega}_T)$ is given as the solution to the equation

$$2\pi V_\alpha/\lambda\equiv\theta=\ln(\bar{\Omega}T/\sqrt{48\pi})+(1/2)\ln\theta-\ln\alpha \tag{2.6}$$

which may be compared with the value

$$\theta^* = \ln(\bar{\Omega}T/\pi) - \ln\alpha \tag{2.7}$$

for testing the maximum of the ordinates $2\pi I_T(\omega_j)$ in $(0, \bar{\Omega})$. A comparison of these values for various values of N is shown in Table 1. It should be noted that both values are based on asymptotic theories which ignore edge effects and the departure from normality in the distribution of $C(\omega)$ and $S(\omega)$.

Davies [5] has examined the edge effects in detail and provided an exact upper bound in the Gaussian case for the probability of exceeding a given crossing level; this leads to upper bounds for the significance level which are very close to the values θ^* noted in Table 1. The effect of departures from normality is also likely to be small, insofar as the terms $C(\omega)$ and $S(\omega)$ can both be represented in terms of random sums of bounded random variables, for which the approximation provided by the Central Limit Theorem should be good even for small values of n.

Table 1. Significance levels for the periodogram maximum over the range $(0, \bar{\Omega})$

5% significance levels	$N=$					
	16	33	100	200	500	1000
θ^*	5.52	5.80	6.91	7.60	8.52	9.21
θ	5.72	6.03	7.23	7.97	8.94	9.68
1% significance levels	$N=$					
	16	33	100	200	500	1000
θ^*	7.13	7.41	8.52	9.21	10.13	10.82
θ	7.47	7.76	8.94	9.68	10.64	11.37

The values of θ^* and θ are based on formulae (2.6) and (2.7) in the text.

2.3. *Fit to Kawasumi's Data; adjustments for clustering*

The periodogram for Kawasumi's data (see Appendix) is shown in Fig. 1. A correction for the mean has been made by taking

$$\tilde{S}(\omega) = \sum_1^N \sin\omega t_i - \int_0^T \bar{\lambda}\sin\omega t\,dt = \sum_1^N \sin\omega t_i - N\frac{1-\cos\omega T}{\omega T}$$

$$\tilde{C}(\omega) = \sum_1^N \cos\omega t_i - \int_0^T \bar{\lambda}\cos\omega t\,dt = \sum_1^N \cos\omega t_i - N\frac{\sin\omega T}{\omega T}$$

in place of $S(\omega)$, $C(\omega)$ in (2.5).

As Kawasumi pointed out, there is a sharp peak at $\hat{\omega}_T = 0.91$

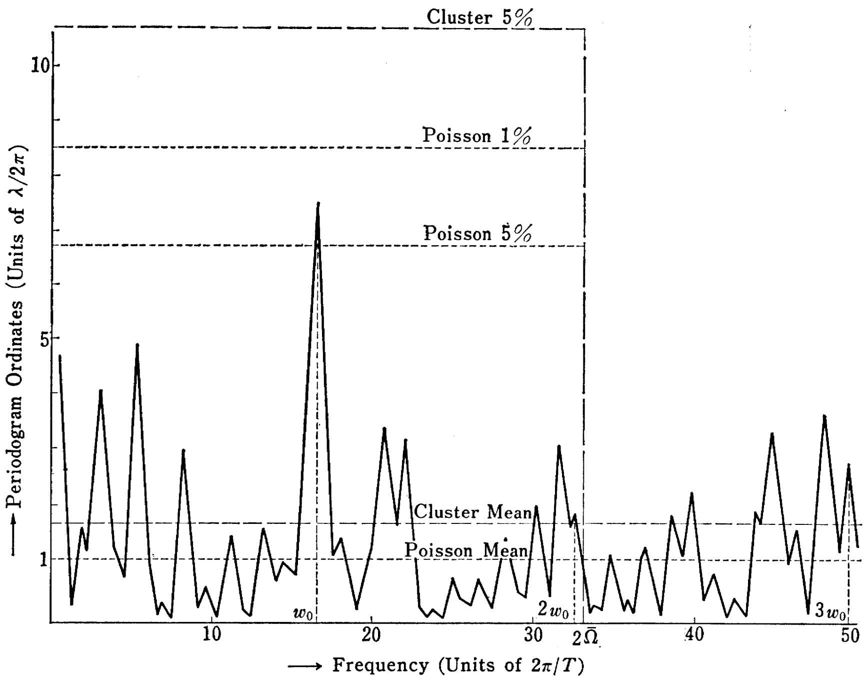

Fig. 1. Periodogram for Kawasumi's Data (full set).

radians/year, corresponding to a period of 69 years. Note also the smaller peaks at the harmonics. Since $\hat{\omega}_T$ falls very close to $\bar{\Omega}$, we have used the frequency range $[0, 2\bar{\Omega}]$ rather than $[0, \bar{\Omega}]$ in tests for the maximum. The peak/mean ratio is about 7.75 which is significant at the 5% level according to the values listed in Table 1. Parameters for the best-fitting cyclic Poisson model, obtained by likelihood maximization, are listed in Table 2; the fitted model for the years 1900–2000 AD is shown in Fig. 2.

Table 2. Parameter values for the Cyclic Poisson Model (Kawasumi's Data)

	$\hat{\omega}_0$	$\hat{\alpha}$	$\hat{\rho}$	$\hat{\theta}$
Full data set	.092	−3.85	1.08	1.32
Reduced data set (clusters removed)	.092	−3.92	0.733	1.46

The parameters are defined in equation (2.1).

Although the high peak leads to a clear rejection of the null hypothesis of a constant rate Poisson process, one should beware of concluding too hastily from this result that the alternative of a cyclic model is thereby accepted. In fact many alternatives are possible. In

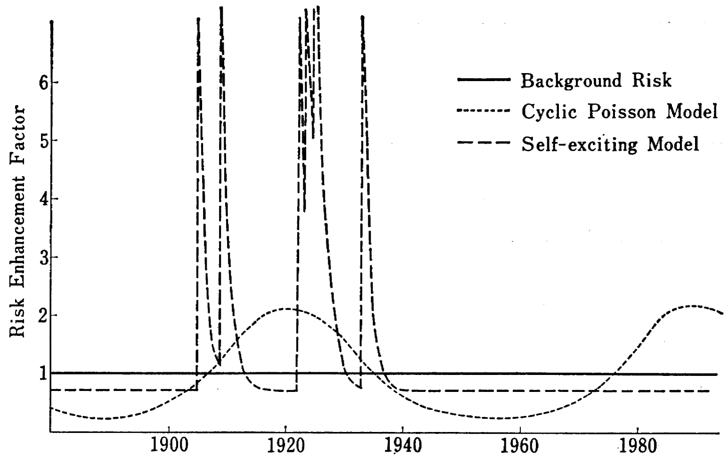

Fig. 2. Cyclic Poisson and Cluster Models for Kawasumi's Data.

particular, the effects of clustering, which turns out to play a critical role in the interpretation of the results, have not yet been taken into account.

In fact, examination of the data shows three triplets (1647–8–9, 1853–4–5, and 1922–3–4) in which earthquakes were recorded in each of three consecutive years, and one doublet (1240–1241). These multiple events occur close to the peaks of the fitted cyclic rate function, and contribute materially to its significance. They are enough by themselves, however, to reject the hypothesis of a Poisson process, whether the rate be constant or cyclically varying. Indeed, since even the cyclic rate function never exceeds a value of 0.07 shocks/year, the probability that earthquakes will occur on each of three consecutive years is dominated by $(.07)^3 = .000343$. This number can also be considered as an upper bound on the rate of occurrence of triplets, so that the probability of obtaining 3 or more triplets in a period of some 1140 years is dominated by the sum of the third and higher terms in a Poisson distribution with parameter $1140 \times .000343 = .38$, i.e. by about .0065. Even this crude bound is enough to show that the occurrence of three such triplets is incompatible with the notion of a Poisson process with bounded rate, so that clustering has to be taken into account before the periodic effect can be assessed adequately.

The simplest way of handling clusters is to ignore the small time differences between the members of each cluster, and to treat the process as having a constant (or cyclically varying) rate of occurrence of clusters, the sizes of individual clusters being independently determined according to some common distribution. If a process of this type were taken as null hypothesis, the periodogram ordinates $I_T(\omega)$ would still be asymptotically exponentially distributed, but with expected values

$(2\pi)^{-1}\lambda m_2/m_1$ in place of $(2\pi)^{-1}\lambda$, where λ in both cases represents the average rate of occurrence of all earthquakes, including those in clusters, and m_k denotes the kth moment of the cluster size. The ratio m_2/m_1, calculated from the occurrence of three triplets and one doublet, is equal to 1.61, reducing the ratio of the periodogram maximum to its expected value from 7.75 for the null hypothesis of a simple Poisson process to 4.8 for the null hypothesis of a compound Poisson process. It can be seen from Table 1 that the peak is then no longer significant at either of the levels listed.

An alternative procedure is to develop directly the likelihood analysis of the compound process, assuming a cyclically varying rate function of the form (2.1) for the occurrence of cluster centres, and a simple parametric form of distribution (the geometric distribution seems most appropriate) for the cluster sizes. Assuming these aspects are independent, the log likelihood breaks up into the sum of two terms, the first of the form (1.5) but with the t_i representing the times of occurrence of cluster centres, and the second of the form

$$\sum_{i=1}^{m} \log p(n_i)$$

where m is the total number of clusters and $p(n)=(1-\rho)\rho^{n-1}$ is the probability of observing just n earthquakes in a given cluster. Estimates for the periodic effect involve only the first of these two terms. In other words, for this simple cluster model the periodic effect should be investigated by treating the clusters as single points and analyzing the resulting reduced process as a simple cyclic Poisson model. This leads both to more powerful tests and to more accurate estimates of the periodic terms than use of the full data set, the additional "noise" introduced by the varying cluster sizes serving merely to increase the variances.

The periodogram for the reduced process is shown in Fig. 3. It is of interest that the dominant feature is still the peak at approximately 69 years. The peak/mean ratio, however, is now only 3.4, which is even less significant than the value obtained earlier for the full data set.

The overall conclusion from this analysis must be that the evidence in favour of a 69-year cycle is not conclusive, although it is certainly suggestive, particularly in view of the AIC values listed in Table 3. Physical considerations tend to reinforce a negative conclusion. The data set is very heterogeneous, containing both large earthquakes in the offshore trench and relatively small earthquakes in the vicinity of Kamakura itself. It is hard to conceive of any physical mechanism which could produce periodic effects in such data, particularly as the

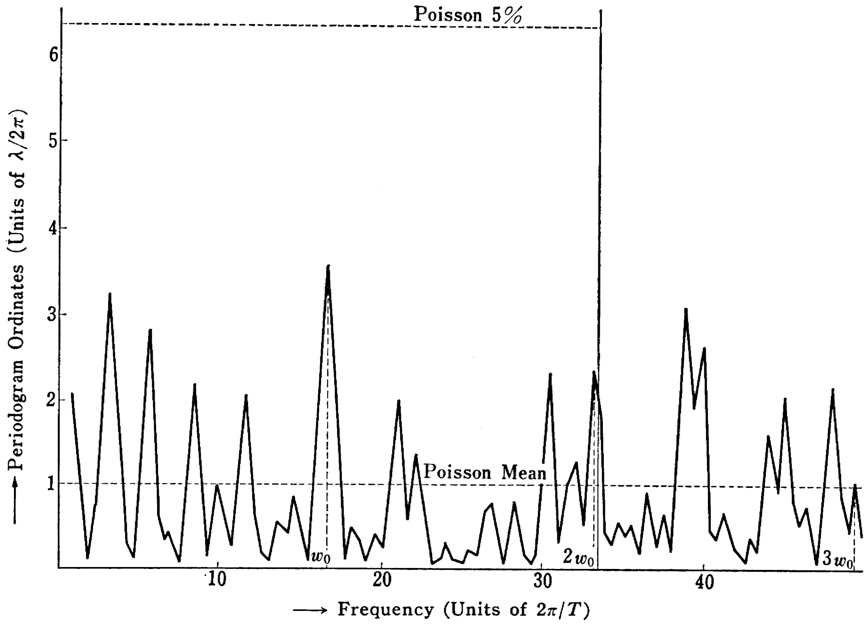

Fig. 3. Periodogram for Kawasumi's Data (reduced set).

proposed period of 69 years does not tally with any other known geophysical effects. The absence of strict periodicities would not rule out the operation of some weaker type of relaxation model, and this is perhaps the most plausible explanation of the sequence of large earthquakes in the set. One model of this type is described in Vere-Jones [20], but the data are not sufficient to allow any clear conclusions to be drawn.

Table 3. AIC values for Cyclic Poisson and Self-Exciting Models (Kawasumi's Data)

	$-\log L$	No. parameters	AIC
Simple Poisson, constant rate	149.2	1	300.3
Compound Poisson, constant rate	140.8	2	285.6
Simple Poisson, cyclic rate	139.5	4	287.0
Compound Poisson, cyclic rate	133.6	5	277.2
Self-exciting, order 0	141.3	3	288.6
Self-exciting, order 4	140.9	7	295.7

3. The self-exciting model

3.1. *Descriptive properties*

We shall refer to the model (1.1) with parameterization (1.2) as Akaike's model of order M. When the process is stationary, it holds that

$$\lambda = \mathrm{E}\,[\lambda(t)] = \mu + \lambda \int_0^\infty g(t)dt .$$

For the Akaike model (1.4) implies that

$$\int_0^\infty g(t)dt = \sum_{m=0}^{M} \frac{\alpha_m \cdot m!}{\beta^{m+1}} < 1 . \tag{3.1}$$

Then

$$\lambda = C \cdot \mu$$

where

$$C = 1\Big/\left(1 - \int_0^\infty g(t)dt\right) \tag{3.2}$$

will be referred to as the *cluster factor* of the model.

The counting spectrum of the self-exciting process is given by Hawkes [8] as

$$\mathrm{P}\,(\omega) = \frac{\lambda}{2\pi}\{1 - G(\omega)\}^{-1}\{1 - G(-\omega)\}^{-1}$$

where $G(\omega)$ is the Fourier transform of the response function $g(t)$. Using this representation, the counting spectrum for Akaike's model becomes

$$\mathrm{P}\,(\omega) = \mu\left(1 - \sum_0^M \alpha_m m!/\beta^{m+1}\right)^{-1}\left|1 - \sum_0^M \alpha_m m!/(\beta + i\omega)^{m+1}\right|^2 . \tag{3.3}$$

3.2. *Maximum likelihood fitting*

Throughout this and the following section we shall use the approximation referred to in the introduction, replacing $\lambda_\infty(t)$ by the approximate form

$$\lambda(t) = \mu + \int_0^t g(t-u)dN(t) .$$

The effect of this approximation is restricted to a transient term at the beginning of the observation period, and it is not difficult to show, as in the discussion of Ogata [15], that estimates based on the approximate form are still asymptotically efficient.

From (1.6) we can write the log-likelihood for Akaike's model of order M in the form

$$\log L = -\mu T - \sum_{m=0}^{M} \alpha_m B(m) + \sum_{i=1}^{N} \log C(i)$$

where

$$B(0) = \frac{1}{\beta} \sum_{i=1}^{n} (e^{-\beta(T-t_i)} - 1) ,$$

and for $m \geqq 1$,

$$B(m) = \sum_{i=1}^{N} \int_0^{T-t_i} x^m e^{-\beta x} dx = \frac{m!}{\beta^{m+1}} \sum_{i=1}^{N} \left[1 - \sum_{j=0}^{m} \frac{\beta^j (T-t_i)^j}{j!} e^{-\beta(T-t_i)} \right] ,$$

$$C(i) = \mu + \sum_{m=0}^{M} \alpha_m A_m(i) ,$$

and

$$A_m(1) = 0 ,$$

$$A_m(i) = \sum_{t_j < t_i} (t_i - t_j)^m e^{-\beta(t_i - t_j)} \qquad \text{for } i \geqq 2 .$$

Gradients of the log-likelihood are given by

$$\frac{\partial \log L}{\partial \alpha_m} = -B(m) + \sum_{i=1}^{N} \frac{A_m(i)}{C(i)} \qquad (m = 0, \cdots, M)$$

$$\frac{\partial \log L}{\partial \beta} = \sum_{m=0}^{M} \alpha_m B(m+1) + \sum_{i=1}^{N} \frac{-D(i)}{C(i)}$$

$$D(i) = \sum_{m=0}^{M} \alpha_m A_{m+1}(i)$$

$$\frac{\partial \log L}{\partial \mu} = -T + \sum_{i=1}^{N} \frac{1}{C(i)} .$$

The Hessian of the log-likelihood is given by

$$\frac{\partial^2 \log L}{\partial \alpha_k \partial \alpha_j} = \sum_{i=1}^{N} \frac{-A_j(i) A_k(i)}{C(i)^2}$$

$$\frac{\partial^2 \log L}{\partial \alpha_j \partial \beta} = B(j+1) + \sum_{i=1}^{N} \frac{-A_{j+1}(i) \cdot C(i) + A_j(i) \cdot D(i)}{C(i)^2}$$

$$\frac{\partial^2 \log L}{\partial \beta^2} = -\sum_{m=0}^{M} \alpha_m B(m+2) + \sum_{i=1}^{M} \frac{\sum_{m=0}^{M} \alpha_m A_{m+2}(i) \cdot C(i) - D(i)^2}{C(i)^2}$$

$$\frac{\partial^2 \log L}{\partial \mu \partial \alpha_j}=\sum_{i=1}^{N}\frac{-A_j(i)}{C(i)^2}$$

$$\frac{\partial^2 \log L}{\partial \beta \partial \mu}=\sum_{i=1}^{N}\frac{D(i)}{C(i)^2}$$

$$\frac{\partial^2 \log L}{\partial \mu^2}=\sum_{i=1}^{N}\left[\frac{-1}{C(i)^2}\right].$$

Maximum likelihood estimates of the parameters of Akaike's models can now be obtained using the above representations and the nonlinear optimization technique of Davidon-Fletcher-Powell (see Kowalik et al. [11]).

3.3. *Application to Kawasumi's Data*

Akaike's models of order 0 and of order 4 were fitted to the data. Parameter values are listed in Table 4, and AIC values compared with those of the cyclic Poisson models in Table 3. Note the large differences in the estimates of the mean rates and cluster sizes for the different order models. The reason for this difference is illustrated in Fig. 4, which shows the response function for the 4th order model.

Table 4. Parameter values for self-exciting models

A

	Mean rate	Cluster size	$\hat{\mu}$	$\hat{\beta}$	AIC value
Kawasumi's Data					
Order 0	0.2783	1.24	0.2241	0.6387	288.6
Order 4	0.1789	18.04	0.9927 $\times 10^{-2}$	0.2993 $\times 10^{-1}$	295.7
JMA Data					
Order 2	0.3470	1.9286	0.1799	1.5809	3997.0
Order 4	0.3470	2.0218	0.1717	1.7400	3974.0
Order 6	0.3664	2.3374	0.1482	0.9479	3959.2

B

	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$	$\hat{\alpha}_6$
Kawasumi's Data							
Order 0	0.1243						
Order 4	0.2535 $\times 10^{-1}$	−0.8627 $\times 10^{-3}$	−0.2627 $\times 10^{-4}$	−0.3857 $\times 10^{-6}$	0.1845 $\times 10^{-8}$		
JMA Data							
Order 2	1.0587	−2.2779	1.4288				
Order 4	1.4178	−5.7732	8.7619	−4.8280	0.9511		
Order 6	1.2009	−4.3545	5.9690	−3.5241	0.9972	−0.1295	−0.6182 $\times 10^{-2}$

It takes several hundred years to die away, whereas the 0th order response function dies away very rapidly. With only 33 earthquakes recorded in a period of about 1000 years, the situation is similar to that of fitting an autoregressive time series model for which the characteristic roots of the fitted model lie close to the unit circle and the number of data points is small. In such a situation the estimates are

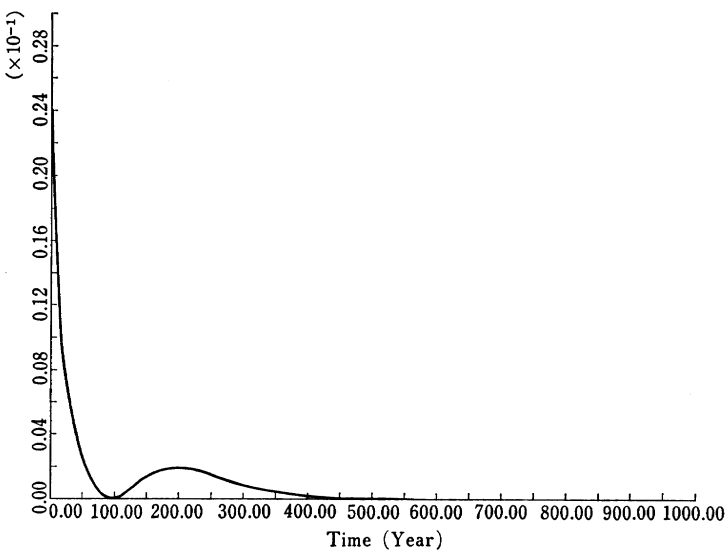

Fig. 4. Response function for 4th order self-exciting (Akaike) model for Kawasumi's Data.

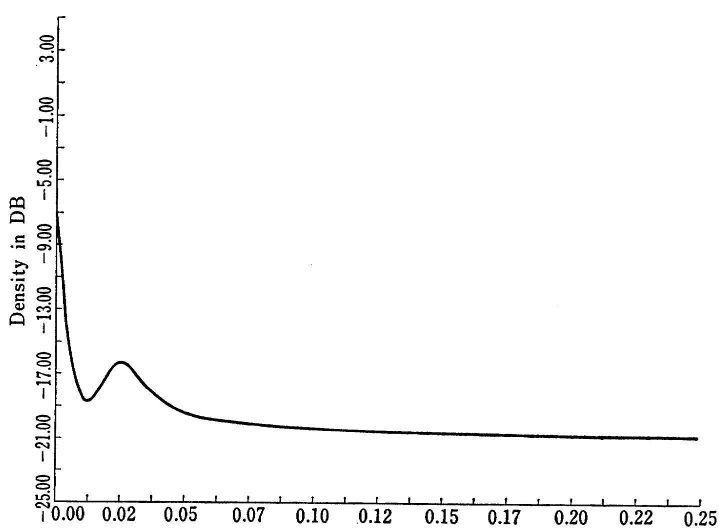

Fig. 5. Spectrum of 4th order self-exciting (Akaike) model for Kawasumi's Data.

unstable and asymptotic theories must be applied with caution.

The spectrum of the 4th order model is shown in Fig. 5, where a dull hill is seen to cover periods from 60 years to 360 years. Fitting higher order models might allow the hill to split into peaks corresponding to the 69 year periodicity and its harmonics, but the sample size and observation period seem too small for this type of treatment.

In summary, the zero order model could be used to give a first approximation to the clustering effect, but the data are unsuitable for the application of higher order models.

3.4. *Application to JMA Data*

The Akaike model was also applied to earthquake occurrence data supplied by the Japan Meteorological Agency for shallow earthquakes in a region roughly corresponding to the Kwanto area, Japan, from 1926 until 1975. The area used was a quadrilateral with vertices at the points (39°N, 143°E), (35°N, 141°E), (36°N, 138°E) and (40°N, 140°E) and the depth <70 km. There were altogether 1268 earthquakes in the list with magnitudes 5 or greater. We assume as before that the initial effect can be ignored, and take $0=t_0<t_1<\cdots<t_{1267}$. Models of several different orders were fitted, with estimated parameters given in Table 4. The figures of the estimated response functions and spectra for some of the fitted models are shown in Figs. 6 and 7.

The results are not very revealing in physical terms. The 6th order model gives the minimum AIC value and should be preferred from the point of view of entropy maximization. The dominant feature is the high peak of the response function near the origin, with a similar peak

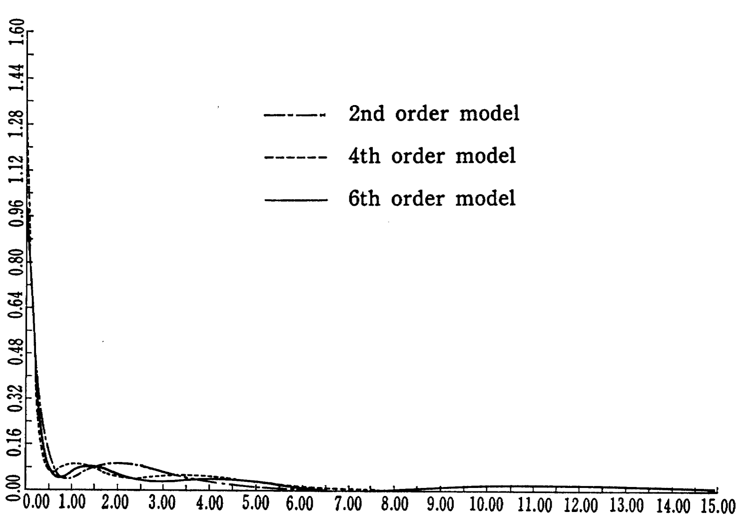

Fig. 6. Response functions for JMA data.

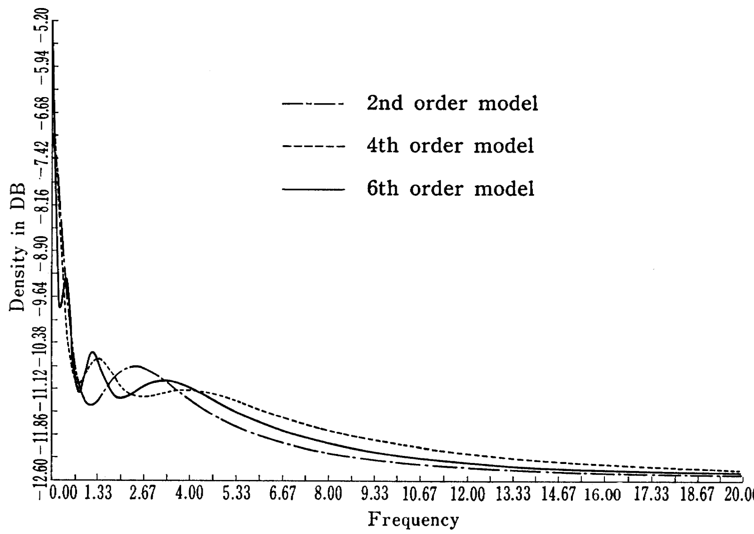

Fig. 7. Spectra for JMA Data.

in the spectrum. This is indicative of the strong clustering effect. The spectrum of the model suggests that there may be weak periodicities with periods at about 4.5, 1.5 and 0.6 days. The physical interpretation of such features is obscure. The second and fourth order models merge these features in dull hills. Some other applications of the model to earthquake data is discussed in Ozaki and Akaike [17].

3.5. *Simulation Data*

It should be noted that it was necessary to use a constrained optimization technique in fitting the model of order 4, in order to get a model which satisfied the non-negativity condition (1.3). Although such constraining is not necessary in most situations, it may be needed when the impulse response function of the true stationary model has a valley with its bottom close to zero. To examine such effects we simulated the model of order 2 with impulse response function (see Fig. 8)

$$g(t)=(0.5-1.15t+0.7t^2)e^{-t},$$

and $\mu=0.025$. The estimated response function obtained by constrained optimization is

$$g(t)=(0.4439-1.0824t+0.6403t^2)e^{-0.9547t}$$

and the impulse response function obtained by free optimization is

$$g(t)=(0.4753-1.1261t+0.6671t^2)e^{-0.9764t}.$$

The two functions are illustrated in Fig. 8; they cannot be distinguished.

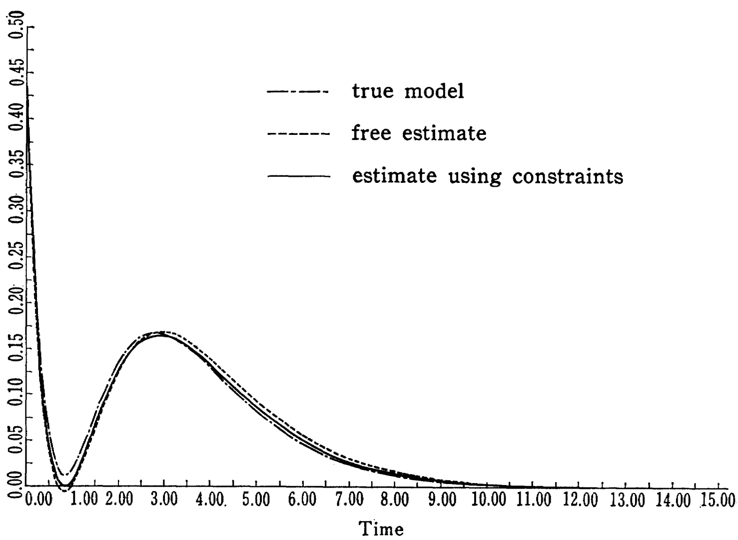

Fig. 8. True and estimated response functions for simulation data.

4. Conclusions

1. The method of maximum likelihood can be used successfully in fitting both the cyclic Poisson process and the self-exciting process with Akaike's form (1.2) for the impulse response function $g(t)$.
2. The cyclic rate compound Poisson process provides the most effective description of Kawasumi's earthquake data from the point of view of entropy maximization. However the periodic effect is not statistically significant when clustering is properly taken into account. In general, the number of data points is too small to distinguish properly between models.

APPENDIX

KAMAKURA EARTHQUAKE DATA (after Kawasumi [10])

Date (AD)	Magnitude	Date (AD)	Magnitude	Date (AD)	Magnitude
818	7.9	1498	8.6	1782	7.3
841	7.0	1525	6.6	1812	6.6
878	7.4	1590	7.2	1853	6.5
1096	8.4	1605	7.9	1854	8.4
1213	6.8	1633	7.1	1855	7.5
1227	6.3	1647	6.8	1905	7.5
1240	6.9	1648	7.1	1909	7.0
1241	7.0	1649	6.5	1922	6.9
1257	7.2	1670	6.4	1923	7.9
1293	7.1	1697	7.2	1924	7.2
1433	7.1	1703	8.2	1933	7.0

Acknowledgements

The authors are grateful to Professor Shibata and to the Japan Meteorological Agency for discussions and for supplying data for the present analyses. The work was partly supported by a Grant from the Ministry of Education, Science and Culture, and was initiated when the first author was a visiting Professor supported by the Japan Society for the Promotion of Science.

VICTORIA UNIVERSITY OF WELLINGTON
THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Akaike, H. (1977). On entropy maximization principle, *Applications of Statistics* (ed. P. R. Krishnaiah), North Holland.

[2] Bartlett, M. S. (1963). The spectral analysis of point processes, *J. Roy. Statist. Soc.*, B, **25**, 264-296 (with discussion).

[3] Cox, D. R. and Lewis, P. A. W. (1966). *The Statistical Analysis of Series of Events*, London (Methuen).

[4] Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika*, **64**, 247-254.

[5] Davies, R. B. (1981). The Chi-squared process with applications to hypothesis testing and time-series analysis, Office of Naval Research Report, No. ONR 80.

[6] Hannan, E. J. (1973). The estimation of frequency, *J. Appl. Prob.*, **10**, 510-519.

[7] Hawkes, A. G. (1971a). Spectra of self-exciting and mutually exciting point processes, *Biometrika*, **58**, 83-90.

[8] Hawkes, A. G. (1971b). Spectra of some mutually exciting point processes, *J. Roy. Statist. Soc.*, B, **33**, 438-443.

[9] Hawkes, A. G. and Adamopoulos (1973). Cluster models for earthquakes—regional comparisons, *Bull. Int. Statist. Inst.*, **45**(3) (Proc. 39th Session of ISI), 454-461.

[10] Kawasumi, H. (1970). Proofs of 69 years periodicity and imminence of destructive earthquakes in Southern Kwanto district and problems in the countermeasures thereof, *Chigaku Zasshi*, **79**(3), 115-138.

[11] Kowalik, J. and Osborne, M. R. (1968). *Methods for Unconstrained Optimization Problems*, American Elsevier, New York.

[12] Kutoyants, Y. A. (1980). *Parameter Estimation for Stochastic Processes*, Academy of Sciences of Armenian SSR, Erevan (in Russian).

[13] Lewis, P. A. W. (1970). Remarks on the theory, computation and application of the spectral analysis of series of events, *J. Sound Vib.*, **12**, 953-975.

[14] Lewis, P. A. W. (1972). Recent results in the statistical analysis of univariate point processes, *Stochastic Point Processes* (ed. P. A. W. Lewis), Wiley, New York, 1-54.

[15] Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes, *Ann. Inst. Statist. Math.*, **30**, A, 145-156.

[16] Ozaki, T. (1979). Maximum likelihood estimation of Hawkes' self-exciting point processes, *Ann. Inst. Statist. Math.*, **31**, B, 145-155.

[17] Ozaki, T. and Akaike, H. (1979). On a generalized model for Hawkes' self-exciting point processes, Research Memo. No. 199, Institute of Statistical Mathematics, Tokyo.

[18] Rubin, I. (1972). Regular point processes and their detection, *IEEE Trans. Inf. Theory*, IT-**18**, 547-557.

[19] Sharpe, K. (1978). Some properties of the crossing process generated by a stationary chi-square process, *Adv. Appl. Prob.*, **10**, 373-391.

[20] Vere-Jones, D. (1978). Earthquake prediction—a statistician's view, *J. Phys. Earth.*, **26**, 129-146.

[21] Vere-Jones, D. (1982). On the estimation of frequency in point process data, to appear in Moran Festschrift.

Ann. Inst. Statist. Math.
39 (1987), Part B, 243

CORRECTIONS TO

"SOME EXAMPLES OF STATISTICAL ESTIMATION APPLIED TO EARTHQUAKE DATA"

D. VERE-JONES AND T. OZAKI

(This Annals Vol. 34, No. 1 (1982), pp. 189–207)

Dr. Mark Berman, Division of Mathematics and Statistics, CSIRO, Sydney, has drawn our attention to the following algebraical and numerical slips in the above paper.

p. 193, equation (2.6): Replace "$\ln(\bar{\Omega}T/\sqrt{48\pi})$" by "$\ln(\bar{\Omega}T/\sqrt{12\pi})$"

p. 194, equation (2.7): Replace "$\ln(\bar{\Omega}T/\pi)$" by "$\ln(\bar{\Omega}T/2\pi)$"

In Table 1, for $N=16$ read $N=25$, and for the θ-values listed read as follows

N =	25	33	100	200	500	1000
$\alpha=.05$	6.48	6.78	7.97	8.71	9.68	10.40
$\alpha=.01$	8.21	8.50	9.68	10.40	11.36	12.09

These changes only reinforce the conclusion in the paper that the cyclic effect in Kawasumi's data cannot be clearly established.

We would like to record our appreciation of Dr. Berman's interest.

VICTORIA UNIVERSITY OF WELLINGTON
THE INSTITUTE OF STATISTICAL MATHEMATICS