Ann. Inst. Statist. Math.
Vol. 56, No. 4, 599–609 (2004)

A BAYESIAN ANALYSIS FOR THE SEISMIC DATA ON TAIWAN

TSAI-HUNG FAN AND ENG-NAN KUO

Graduate Institute of Statistics, National Central University, Chungli, Taiwan 32054, R.O.C.

(Received October 28, 2002; revised November 18, 2003)

Abstract. A Bayesian approach is used to analyze the seismic events with magnitudes at least 4.7 on Taiwan. Following the idea proposed by Ogata (1988, *Journal of the American Statistical Association*, **83**, 9–27), an epidemic model for the process of occurrence times given the observed magnitude values is considered, incorporated with gamma prior distributions for the parameters in the model, while the hyperparameters of the prior are essentially determined by the seismic data in an earlier period. Bayesian inference is made on the conditional intensity function via Markov chain Monte Carlo method. The results yield acceptable accuracies in predicting large earthquake events within short time periods.

Key words and phrases: Epidemic model, prior distribution, hyperparameter, conditional intensity function, MCMC method.

1. Introduction

Statistical research using point process models to analyze the earthquake activity have been considered since Utsu (1961) in which only the temporal component was focused in small areas that are believed to have homogeneous physical characteristics. More sophisticated approaches incorporating the magnitude of the shocks where the relationships between the main shocks and the minor shocks are described can be seen in Ogata (1988, 1989). Ogata and Katsura (1988) and Musmeci and Vere-Jones (1992) have also included the spatial location of the shocks for modeling events that occur in heterogeneous regions. Betrò and Ladelli (1996) address the issue of model selection in different areas of seismic activities in Italy.

Let $\{t_i, m_i\}_{i=1}^N$ denote the occurrence times and the corresponding magnitude of the earthquake events over the time interval $[0, T]$ of interest, and M_r be a prescribed minimum magnitude considered. Ogata (1988) constructed the seismic model based on an epidemic model with conditional intensity function

$$\lambda(t \mid \underline{\theta}) = \mu + \sum_{t_i<t} e^{\beta(m_i - M_r)} \frac{a}{(t - t_i + c)^p}, \tag{1.1}$$

where $\underline{\theta} = (\mu, a, \beta, c, p)$ denotes the vector of unknown but positive parameters and μ relates to the sequence of main shocks, β, a, c and p refer to the sequence of the aftershocks. He used the maximized likelihood estimator (MLE) to analyze and predict the seismic activity along the northeast coast in Japan. The choice of the conditional intensity function has been described extensively in Ogata (1988, 1989) and the references there in. Peruggia and Santner (1996) applied a Bayesian approach to the seismic data at the area of Sannio Matese in Italy based on the intensity function

$$\lambda(t \mid \underline{\theta}) = \mu + \sum_{t_i<t} e^{\beta(m_i - M_r)} a e^{-\alpha(t-t_i)}, \tag{1.2}$$

where $\underline{\theta} = (\mu, \alpha, \beta, a)$ are all positive but unknown parameters. A similar method has been used to the seismic data in Hualien area of the northeastern Taiwan by Fan and Lin (2002).

In this paper, we use a Bayesian approach to analyze the seismic data on the entire island of Taiwan which includes the area between the east longitudes of 119° and 122°, and the north latitudes of 21° and 26°. The data are obtained from the Central Weather Bureau of Taiwan. By restricting attention to earthquakes occurring between January 1984 and December 1997 and their magnitudes, we model the likelihood function using a nonhomogeneous Poisson process with conditional intensity function given by (1.1). Both the AIC (Akaike (1974)) and BIC (Schwarz (1978)) criteria suggest (1.1) is superior to (1.2). At the beginning of 1991, the Central Weather Bureau in Taiwan introduced a new measuring system to record the seismic activity. Therefore, we use it as a cut-off point to collect the prior information. The data before 1991, measured by the old system, are used to determine the prior distribution and those after 1991, measured by the new system, are used in the likelihood. However, according to the Central Weather Bureau, there exists a measurement difference between the two systems. Their rough belief is that the magnitude of 4.5 measured by the old system is equivalent to that of 4.7 by the new one. We have also observed that the average daily rates in the early period with magnitude at least 4.5 and in the later period with magnitude at least 4.7 are about the same. Thus in our study only those data with magnitudes at least 4.5 measured by the old system are considered and adjusted to construct the prior and the minimum magnitude M_r of the earthquake events used in (1.1) is 4.7. In Section 2, we will select between models (1.1) and (1.2). Both the AIC and BIC criteria support (1.1) over (1.2). Then we will describe the prior distributions of the parameters which more or less are adopted from Peruggia and Santner (1996). Section 3 gives the Bayesian inference of the conditional intensity function which can be used to estimate the probability of event occurrence within a short period. Section 4 is the conclusion and discussion.

2. The model and the prior

2.1 *Model selection*

The conditional likelihood for an epidemic model is uniquely determined by its conditional intensity function (Daley and Vere-Jones (1988)). Let N be the number of events occurred over the time interval considered $[0,T]$, $\{t_i\}_{i=1}^N$ and $\{m_i\}_{i=1}^N$ be the observed times and magnitudes of the events, then the conditional likelihood of $\underline{\theta}$ based on the intensity function $\lambda(t;\underline{\theta})$ is

$$\log L_T(\underline{\theta}) = \sum_{i=1}^N \log \lambda(t_i;\underline{\theta}) - \int_0^T \lambda(t;\underline{\theta})dt. \tag{2.1}$$

Thus, using $\lambda(t;\underline{\theta})$ in (1.1), one can deduce the conditional log likelihood of $\underline{\theta} = (\mu, a, \beta, c, p)$ as

$$(2.2)\qquad \log L_T(\underline{\theta}) = \begin{cases} \log\mu + \sum_{i=2}^{N}\log(\mu + a\sum_{t_j<t_i}\frac{e^{\beta(m_j-M_r)}}{(t_i-t_j+c)^p}) - \mu T \\ \quad - \frac{a}{p-1}\sum_{i=1}^{N}[e^{\beta(m_i-M_r)}(c^{-(p-1)} - (T-t_i+c)^{-(p-1)})] \\ \qquad\qquad\qquad\qquad \text{for } \; p>1 \\ \log\mu + \sum_{i=2}^{N}\log(\mu + a\sum_{t_j<t_i}\frac{e^{\alpha(m_j-M_r)}}{(t_i-t_j+c)^p}) - \mu T \\ \quad - a\sum_{i=1}^{N} e^{\beta(m_i-M_r)}\log(\frac{T-t_i+c}{c}) \quad \text{for } \; p=1; \end{cases}$$

and while for $\lambda(t;\underline{\theta})$ is given by (1.2)

$$(2.3)\qquad \log L_T(\underline{\theta}) = \log\mu + \sum_{i=2}^{N}\log\left(\mu + a\sum_{t_j<t_i} e^{-\alpha(t_i-t_j)}e^{\beta(m_i-M_r)}\right) - \mu T + \frac{a}{\alpha}\sum_{i=1}^{N} e^{\beta(m_i-M_r)}[1 - e^{-\alpha(T-t_i)}]$$

with $\underline{\theta} = (\mu, a, \alpha, \beta)$ and all the parameters in $\underline{\theta}$ are positive. Ogata (1988) uses (2.2) to analyze the seismic activity in the northeast coast of Japan by the MLE approach and Perrugia and Santner (1996) perform a Bayesian analysis on the earthquake data in Italy using (2.3). We will first make a model comparison between (2.2) and (2.3) based on the data to be analyzed which include all the earthquake events of magnitudes at least 4.7 ($M_r = 4.7$) on Taiwan over the east longitudes of 119° and 122°, and the north latitudes of 21° and 26° from 1991 to 1997 with $N = 110$. Figure 1 shows the magnitudes versus the event times of all such events.

Two of the most widely used model selection criteria are the AIC (Akaike (1974)) and the BIC (Schwarz (1978)) criteria. For given data of N observations, let $L(\underline{\theta})$ be the likelihood of $\underline{\theta}$ which has k unknown parameters, and $\hat{\underline{\theta}}_{\mathrm{MLE}}$ be its MLE, then the AIC

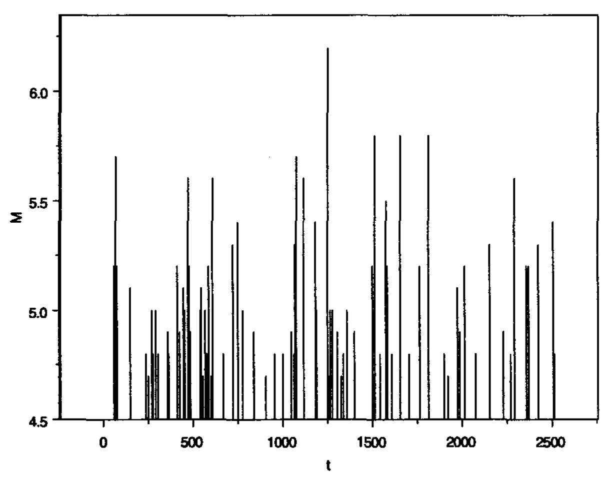

Fig. 1. The magnitudes versus the occurrence times of the earthquakes from 1991 to 1997 in Taiwan.

Table 1. Comparison between Models 1 and 2.

	k	N	$\log L(\hat{\underline{\theta}}_{\rm MLE})$	AIC	BIC
Model 1	5	110	−403.18	816.36	829.87
Model 2	4	110	−419.85	847.70	858.50

value of the model is defined by

$$\text{AIC} = -2\log L(\hat{\underline{\theta}}_{\rm MLE}) + 2k;$$

and the BIC value is

$$\text{BIC} = -2\log L(\hat{\underline{\theta}}_{\rm MLE}) + k\log N.$$

In other words, both AIC and BIC examine the maximized likelihood values plus a penalty term. The best model to be selected is the one with the smallest value. In addition to being penalized by the number of unknown parameters, BIC also takes into account of the sample size effect so that it would correct the bias toward the more complex models. Table 1 gives the corresponding AIC and BIC values based on the data shown in Fig. 1. We see that (2.2) indeed yields bigger maximized log likelihood and it also has smaller AIC and BIC values even though it contains one more parameter. Thus, we conclude that the data support Model 1 (with corresponding log-likelihood given by (2.2)) over Model 2 (with log-likelihood given by (2.3)). Hence, we will only focus on Model 1 with the likelihood given by (2.2) from now on.

2.2 *Prior determination*

In this paper, a Bayesian analysis will be conducted. The prior distribution on $\underline{\theta}$ is more or less adopted from that used by Perrugia and Santner (1996). That is, we assume that the prior distribution of μ is independent of that of (a, β, c, p) by the fact that μ relates to the main shocks and the rests relate to the aftershocks. Moreover, β describes the magnitude effect and (c, p) are about the time effect of the events, hence we assume independence between β and (c, p). Since the value of p is known not to be too big (cf., Ogata (1988)), a uniform prior over $(0, 5)$ is considered here for simplicity. However, the prior distribution of a depends on β, c and p. Furthermore, all μ, β and c are assumed to have gamma priors with corresponding hyperparameters $(\gamma_\mu, \lambda_\mu)$, $(\gamma_\beta, \lambda_\beta)$, and (γ_c, λ_c) respectively. Given β, c, and p, it is also assumed that the conditional prior of a is of gamma distribution with hyperparameters (γ_a, λ_a), but both γ_a and λ_a depend on the values of β, c and p. Based on constraints on the behavior of the magnitude effect (depending on β) and the time effect (depending on c and p) of a shock, one can identify the area where a is concentrated most likely. Detailed discussion can be seen in Perrugia and Santner (1996). Therefore, the joint prior density of $\underline{\theta}$ is

$$\pi(\underline{\theta}) = [\underline{\theta}] = [\mu][a \mid \beta, c, p][\beta][c][p], \tag{2.4}$$

where $[\cdot]$ or $[\cdot \mid \cdot]$ denotes the density or the conditional density of the prior for the corresponding parameter(s).

The values of the hyperparameters of the priors can be determined by specification of the means and percentiles of the prior distributions. Such information here is obtained from an approximate sample of $\underline{\theta}$ based on the early seismic data from January 1984 to December 1990 in the area considered. Beginning in 1991, the Central Weather Bureau in Taiwan changed the measuring system for recording the seismic activity. The data measured by the old system before 1991 are used to construct the prior distribution and those measured by the new system after 1991 are used in the likelihood. As mentioned in Section 1, a measurement difference exists between the two systems. We have investigated that the average occurrence rate of earthquakes with magnitude at least 4.5 (out of 106 occurrences) in the early period was 0.042 per day while that with magnitude above 4.7 (with 110 cases) in the later period was 0.043. Therefore, our prior information is based on the early data with (original) minimum magnitude 4.5. There are 106 such events in the early period. Adjusted values of their magnitudes are made by the Central Weather Bureau which are basically derived by a simple linear regression. Figure 2 shows the cumulative numbers of shocks of the early data with magnitudes at least 4.5 and 4.7 as well as those of the recent data with magnitudes at least 4.7. It indicates that the adjustments look reasonable. To obtain the prior distribution, consider the likelihood function

$$L_0(\underline{\theta}) = \exp(\log L_{T_0}(\underline{\theta})),$$

where $\log L_{T_0}(\underline{\theta})$ is given by (2.2), with $\{t_i^0, m_i^0\}_{i=1}^{N_0}$ being the adjusted data in the early period time interval $[0, T_0]$ with $N_0 = 106$. Letting $\pi_0(\underline{\theta}) = 1$ be the noninformative prior of $\underline{\theta}$, it yields the posterior density of $\underline{\theta}$ as

$$\pi_0(\underline{\theta} \mid \{t_i^0, m_i^0\}) \propto L_0(\underline{\theta})\pi_0(\underline{\theta}) \propto L_0(\underline{\theta}),$$

subject to the dependence of a on β, c, and p as described previously. Note that $\pi_0(\underline{\theta} \mid \{t_i^0, m_i^0\})$ contains all the information about $\underline{\theta}$ carried by the early data. Using Metropolis-Hastings algorithm with a multivariate log-normal distribution for (μ, a, β, c) and a uniform distribution for p as the proposal distributions to generate 500 parallel (independent) chains, each of length 200, an approximate posterior sample of $\underline{\theta}$ of size 500 with respect to $\pi_0(\underline{\theta} \mid \{t_i^0, m_i^0\})$ is drawn from the last observation generated in each chain. Then we find the corresponding hyperparameters by matching the gamma priors with the sample means and quantiles of the generated distributions. The estimated hyperparameters are listed in Table 2.

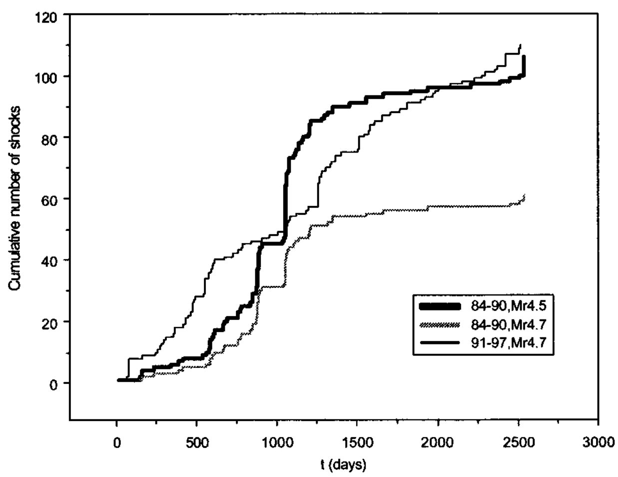

Fig. 2. Cumulative number of shocks of the early and recent data.

Table 2. Estimated hyperparameters of the priors (Unit: day).

Parameter	Gamma Hyperparameters	
$\underline{\theta}$	γ	λ
μ	11.86	1101.76
β	33.57	11.80
c	1.35	34.30

3. Bayesian inference

Given the observed data $\{t_i, m_i\}_{i=1}^N$ over $[0, T]$ with $m_i \geq M_r$ for all $i = 1, \ldots, N$, the posterior density of $\underline{\theta}$ is

$$\pi(\underline{\theta} \mid \{t_i, m_i\}) \propto L_T(\underline{\theta})\pi(\underline{\theta}),$$

where $\log L_T(\underline{\theta})$ and $\pi(\underline{\theta})$ are of the forms in (2.2) and (2.4), respectively. The observed data analyzed here include all the earthquake events of magnitudes at least 4.7 ($M_r =$ 4.7) on Taiwan from 1991 to 1997 with $N = 110$. The prior $\pi(\underline{\theta})$ is constructed based on an early period data from 1984 to 1990 in the same area after adjustments as discussed in Subsection 2.2. It is obvious but not unusual that the posterior $\pi(\underline{\theta} \mid \{t_i, m_i\})$ is not of a closed form. An approximate posterior sample, $\underline{\theta}_1, \underline{\theta}_2, \ldots, \underline{\theta}_M$, of $\underline{\theta}$ of size $M = 10,000$ are generated via the Metropolis MCMC algorithm similar to that described previously. The resulting sample means, denoted by $\hat{\underline{\theta}} = (\hat{\mu}, \hat{a}, \hat{\beta}, \hat{c}, \hat{p})$, which are the commonly used Bayesian estimates of the parameters, and the sample variances are listed in Table 3. Note that, for each t, $\lambda(t; \underline{\theta}_i)$, $i = 1, \ldots, M$, also form an approximate posterior sample of $\lambda(t; \underline{\theta})$ and its sample mean

$$\hat{\lambda}(t; \underline{\theta}) = \sum_{i=1}^{M} \lambda(t; \underline{\theta}_i)/M$$

can be considered as a Bayesian estimate of $\lambda(t; \underline{\theta})$ in contrast to $\lambda(t; \hat{\underline{\theta}})$ used by Perrugia and Santner (1996). One may view $\hat{\lambda}(t; \underline{\theta})$ as the estimation of the posterior intensity mean, or the Bayesian predictive intensity (cf., Rhoades *et al.* (1994) and Ogata (2002)). Furthermore, any Bayesian inference of $\lambda(t; \underline{\theta})$ can be made for each t via the posterior sample. For example, one can obtain a corresponding $(1-\alpha)100\%$ credible set of $\lambda(t; \underline{\theta})$ for each t. Figure 3 plots $\log \hat{\lambda}(t; \underline{\theta})$ and the (estimated) 90% credible bands of $\log \lambda(t; \underline{\theta})$ for each $t \in [0, T]$ pointwisely. We have also examined that $\hat{\lambda}(t; \underline{\theta})$, $\lambda(t; \hat{\underline{\theta}})$ and $\lambda(t; \hat{\underline{\theta}}_{\text{MLE}})$

Table 3. Estimated posterior means (posterior standard deviations) and the MLE of the parameters. (The last line lists the MLE's based on the adjusted early data only.)

Parameter	μ	a	c	β	p
Posterior mean	0.0298	0.0081	0.0420	2.9831	1.1095
(Posterior s.d.)	(0.0014)	(0.0010)	(0.0141)	(0.2917)	(0.0412)
MLE	0.0320	0.0043	0.0491	3.5999	1.2718
Early-MLE	0.019	0.0120	0.0041	2.8167	0.9411

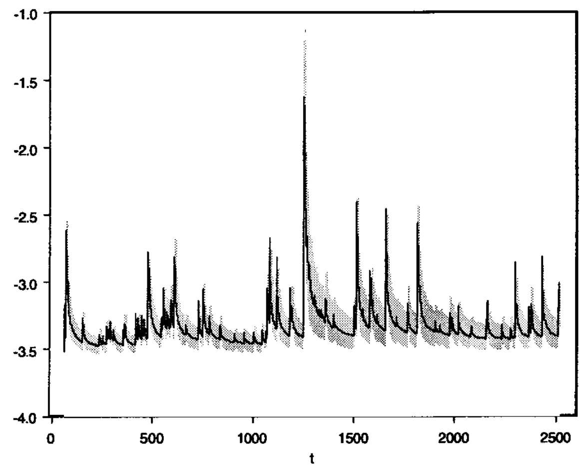

Fig. 3. Graphs of $\hat{\lambda}(t;\underline{\theta})$ (solid line) for $0 < t < T$ and the 90% credible bands of $\lambda(t;\underline{\theta})$ (shaded area). The vertical axis is given in logarithmic scale.

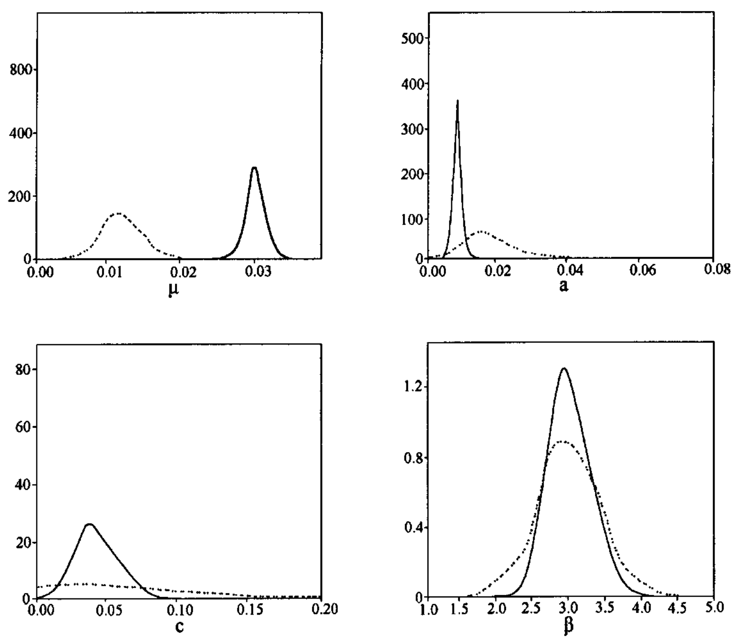

Fig. 4. The prior density (dotted) and the posterior density (solid) of each parameter.

(where $\hat{\underline{\theta}}_{\mathrm{MLE}}$ is the MLE of $\underline{\theta}$ based on the recent data) are not of much difference pointwise, especially the two Bayesian point estimates. The posterior and prior densities of each parameter are also graphed in Fig. 4, from which we see that the prior densities are considerably flatter compared with the posteriors and hence in general are satisfactory by robustness consideration (cf., Berger (1985)).

4. Conclusion and discussion

Given the conditional intensity function $\lambda(t;\underline{\theta})$ as the occurrence rate of a Poisson process, one can calculate the probability of event occurrence within a short time period $(t, t+\Delta t)$ for any $\Delta t > 0$. Let $X(t, t+\Delta t)$ denote the number of such events occurring in $(t, t+\Delta t)$. The probability of at least one occurrence within the short interval is

$$\phi_{\Delta t}(t) = \Pr(X(t, t+\Delta t) > 0) = 1 - \exp\left(-\int_t^{t+\Delta t} \lambda(s;\underline{\theta})ds\right). \tag{4.1}$$

For fixed $\Delta t > 0$ and for each $t > 0$, $\phi_{\Delta t}(t)$ can be estimated via (4.1) by replacing $\lambda(t;\underline{\theta})$ with its estimate. If the estimated probability is high, we may suspect it as a signal of a possible future earthquake (with magnitude at least M_r) within the next Δt time interval. However, $\phi_{\Delta t}(t)$ should be estimated with the data observed only up to time t for prediction purpose. Figure 5 gives the predicted $\log \lambda(t;\underline{\theta})$, only using data up to t for $0 < t < T$ by MLE and the predictive Bayesian methods. The thin line is $\lambda(t;\hat{\underline{\theta}}_{\mathrm{MLE}})$, without taking account of the early period data which results in higher conditional intensity, and the grey line, say $\lambda(t;\hat{\underline{\theta}}_{\mathrm{all}})$, produced by the MLE on all data up to t including the adjusted data from the early period, yields lower conditional intensity; while the thick one, $\hat{\lambda}(t;\underline{\theta})$ (the Bayesian result), lies in between except in the very beginning period. Note that the adjusted data in the early period indeed have much smaller MLE than those in the later period (see Table 3) which might pull down the intensity. On the other hand, the Bayesian approach, incorporating the prior information from previous observations, provides a compromising result and seems to be less sensitive compared with $\lambda(t;\hat{\underline{\theta}}_{\mathrm{all}})$ against the early information. At time goes on with more data collected, the prior plays less role and the Bayesian result turns out to be similar to the MLE with current data only.

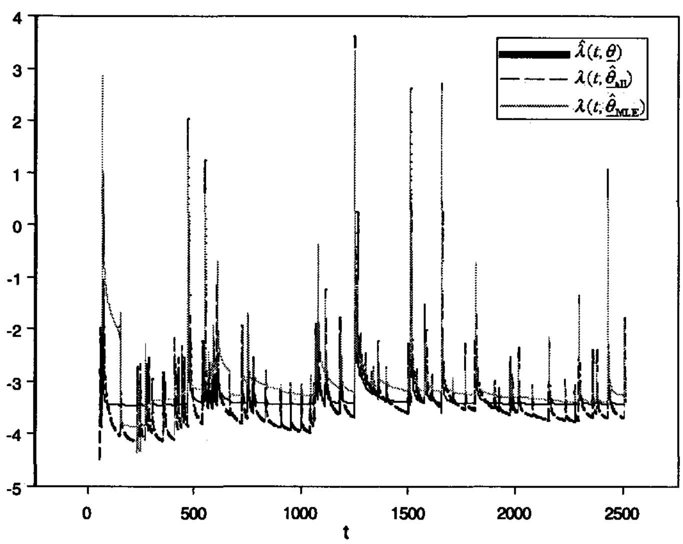

Fig. 5. Graphs of (predicted) conditional intensity functions.

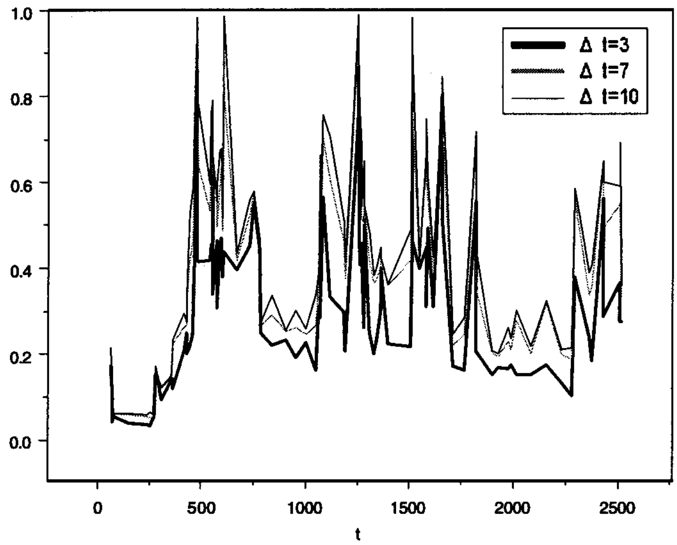

Fig. 6. Graphs of the predictive probabilities $\hat{\phi}_{\Delta t}(t)$ for $0 < t < T$.

Hereafter, we will only look at the estimated $\phi_{\Delta t}(t)$, denoted by $\hat{\phi}_{\Delta t}(t)$, in which $\lambda(t; \underline{\theta})$ is estimated by $\hat{\lambda}(t; \underline{\theta})$ using data up to t. Figure 6 plots $\hat{\phi}_{\Delta t}(t)$ for $0 < t < T$ with $\Delta t = 3, 7$ and 10 (days), respectively. It is then of desired to choose a cutoff value for $\hat{\phi}_{\Delta t}(t)$ in order to make predictions. If $\hat{\phi}_{\Delta t}(t)$ is higher than the cutoff, we conclude it as a signal for a possible earthquake within $(t, t + \Delta t)$. Appropriate cutoff values can be determined based on the prediction results. Let a, b, c, and d represent the numbers of successful forecasts of occurrence, failures to predict, false alarms and successful forecasts of non-occurrence in the prediction, respectively. For example, considering $\Delta t = 7$, among all 109 events studied, there are 45 of them occurred within 7 days. In Fig. 7 we show the prediction error rates of $\hat{\phi}_7(t)$ against different cutoff values, $\underline{\phi}_7$, in which the dotted line (a) is the false alarm (false positive) rate ($= c/(a + c)$), the dashed line (b) is the undetected (false negative) rate ($= b/(a + b)$) and the overall error rate ($= (b + c)/(a + b + c + d)$) is plotted by the solid line (c). Obviously, more signals are detected using smaller cutoff values but it produces higher false positive rates. On the other hand, if bigger cutoff values are used, it results in fewer signals and thus higher false negative rates. In terms of the overall error rates, the best choice for the cutoff is about 0.5. In addition, we also consider the Hanssen-Kuiper skill score, or so called R-score, namely

$$R = \frac{ad - bc}{(a + b)(c + d)}$$

for various cutoffs, and the results are shown in Fig. 8. Again the maximum R-score is achieved at $\underline{\phi}_7 = 0.5$. If $\hat{\phi}_7(t) \geq 0.5$ is used as detection of a signal, there are overall 49 signals detected including 15 false alarms. The correct prediction rate is 34/45 (75.56%) and the correct detection rate is 34/49 (69.39%) with R-score 0.521.

It is also interesting to note that the trend of $\hat{\phi}_{\Delta t}(t)$ is similar to that of the magnitudes (see Fig. 1). It is worth to mention that if the cutoff used is 0.5, the events detected are indeed of magnitudes over 5.4 for $\hat{\phi}_7(t)$, 5.8 for $\hat{\phi}_3(t)$ and 5.3 for $\hat{\phi}_{10}(t)$.

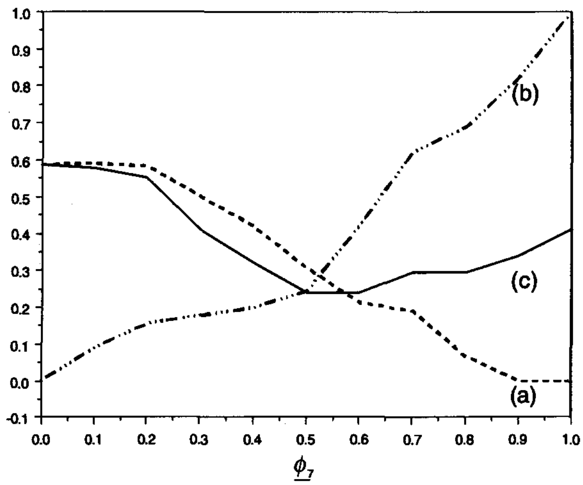

Fig. 7. The prediction error rates for various cutoffs.

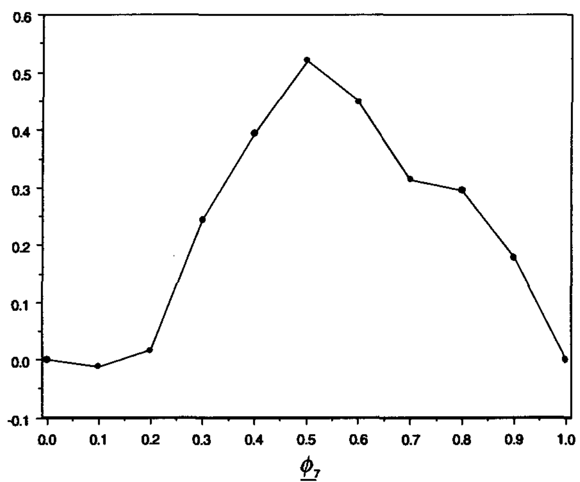

Fig. 8. The R-score for various cutoffs.

The predictive probabilities are relatively lower for smaller Δt and thus indicate less signals, it detects bigger events however. On the contrary, longer period prediction yields higher probabilities and more signals but with more false alarms as well. Higher values of the predictive probabilities correspond to larger earthquake events. How to find the correspondence between $\hat{\phi}_{\Delta t}(t)$ and detected magnitudes within various time intervals should be an important issue in further study.

Acknowledgements

This work was supported by the MOE Program for Promoting Academic Excellent of Universities in Taiwan under the grant numbers 91-N-FA07-7-4 and 91-H-FA07-1-4 and the data were provided by the Central Weather Bureau in Taiwan. The authors would like to thank Prof. Y. Ogata for the FORTRAN codes in computing the MLE's and for his valuable suggestions at a personal contact. The authors also express their gratitude to Prof. Y. I. Chen for her encouragement and discussion on this research project.

References

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, **19**, 716–723.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, New York.

Betrò, B. and Ladelli, L. (1996). Point process analysis for Italian seismic activity, *Applied Stochastic Models and Data Analysis*, **12**, 75–105.

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

Fan, T. H. and Lin, J. S. (2002). A Bayesian analysis for the seismic data in Hualien area, *Journal of Chinese Statistical Association*, **40**, 229–247.

Musmeci, F. and Vere-Jones, D. (1992). A space-time clustering model for historical earthquakes, *Annals of the Institute of Statistical Mathematics*, **44**, 1–11.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes, *Journal of the American Statistical Association*, **83**, 9–27.

Ogata, Y. (1989). Statistical model for standard seismicity and detection of anomalies by residual analysis, *Tectonophysics*, **169**, 159–174.

Ogata, Y. (2002). Slip-size-dependent renewal processes and Bayesian inferences for uncertainties, *Journal of Geophysical Research*, **107**(B11), 2268, doi:10.1029/2001JB000668.

Ogata, Y. and Katsura, K. (1988). Likelihood analysis of spatial inhomogeneity for marked point patterns, *Annals of the Institute of Statistical Mathematics*, **40**, 29–39.

Peruggia, M. and Santner, T. (1996). Bayesian analysis of time evolution of earthquakes, *Journal of the American Statistical Association*, **91**, 1209–1218.

Rhoades, D. A., Van Dissen, R. J. and Dowrik, D. J. (1994). On the handling of uncertainties in estimating the hazard of rupture on a fault segment, *Journal of Geophysical Research*, **99**, 13701–13712.

Schwarz, G. (1978). Estimating the dimension of a model, *The Annals of Statistics*, **6**, 461–464.

Utsu, T. (1961). A statistical study on the occurrence of aftershocks, *Geophysical Magazine*, **30**, 521–605.