Ann Inst Stat Math (2012) 64:87–106
DOI 10.1007/s10463-010-0302-9

Markov modulated Poisson process associated with state-dependent marks and its applications to the deep earthquakes

Shaochuan Lu

Received: 25 April 2008 / Revised: 13 January 2010 / Published online: 27 May 2010

Abstract In this paper, we introduce one type of Markov-Modulated Poisson Process (MMPP) whose arrival times are associated with state-dependent marks. Statistical inference problems including the derivation of the likelihood, parameter estimation through EM algorithm and statistical inference on the state process and the observed point process are addressed. A goodness-of-fit test is proposed for MMPP with state-dependent marks by utilizing the theories of rescaling marked point process. We also perform some numerical simulations to indicate the effects of different marks on the efficiencies and accuracies of MLE. The effects of the attached marks on the estimation tend to be weakened for increasing data sizes. Then we apply these methods to characterize the occurrence patterns of New Zealand deep earthquakes through a second-order MMPP with state-dependent marks. In this model, the occurrence times and magnitudes of the deep earthquakes are associated with two levels of seismicity which evolves in terms of an unobservable two-state Markov chain.

Keywords MMPP · Hidden Markov model · EM algorithm · Exponential family · New Zealand deep earthquakes · Marked point process

1 Introduction

The topic of the current paper is a type of Markov-modulated Poisson process (MMPP), in which the occurrence times of the point process are attached to marks. More

S. Lu
School of Mathematical Science,
Beijing Normal University, Beijing, China

S. Lu (✉)
School of Mathematics, Statistics and Operation Research,
Victoria University of Wellington, Wellington, New Zealand
e-mail: shaochuan.lv@gmail.com

accurately, it is a marked doubly stochastic point process for which the stochastic intensity of the ground process, and the mark distribution are determined by an underlying irreducible finite Markov chain. Such an extension of MMPP has potential applications in spatial–temporal point patterns, multivariate point processes or other point processes with attached mark.

MMPP is particularly useful in modelling time-varying intensity rate processes such as traffic flows of communication networks, Internet traffic flows and queuing systems. A collection of properties of ordinary MMPP is given in Fischer and Meier-Hellstern (1993). The parameter estimation for MMPP through EM algorithm and its comparison with downhill simplex algorithm are addressed by Ryden (1996a). Jensen (2005) discusses the likelihood process of MMPP with discrete type marks.

In this paper, we obtain an extension for the likelihood of MMPP with marks and outline a procedure for parameter estimation by the EM algorithm. We pay especial attention to the case when the marks come from an exponential family distribution. Then we discuss the inference of the state process and the observed point process. A method of assessing the goodness-of-fit for MMPP with marks based on the rescaling theory of the marked point process (Vere-Jones and Schoenberg 2004) is also suggested. Then we present some simulations to show the effects of the additional marks on the estimation efficiencies and accuracies. The effects are demonstrated by comparing the variances of the estimates for series of events with fixed number of observations simulated from a MMPP without mark and MMPPs attached by several pairs of marks, in which the infinitesimal generator and the intensity rates are assumed identical and the state-dependent distributions of the marks are separated in different degrees, so that the effects posed by the additional marks are comparable.

In the last section, we apply these methods to the deep earthquakes around the North Island of New Zealand. Catalogue completeness and homogeneity are demonstrated by cusum (cumulative sum) charts and coverage information of monitoring networks in the early period. We suggest the main occurrence patterns of New Zealand deep earthquakes, in a relatively large time scale, is the time-varying seismic behaviour. A switching Poisson model, which is a Poisson process with the Poisson rates switching between two levels according to an unobserved two-state Markov chain, is proposed to characterize the time-varying seismic activities. In this model, the magnitudes of the earthquakes are treated as marks whose distributions are also dependent on the underlying Markov chain.

2 The likelihood

We denote the infinitesimal generator matrix of the underlying irreducible Markov chain $X(t)$ by $\mathbf{Q} = (q_{ij})_{r\times r}$ with its (i,i)-th element $-q_i \hat{=} q_{ii}$ satisfying $q_i = \sum_{j:j\neq i} q_{ij}, i = 1,\ldots,r$ and $q_{ij} > 0$ for $i \neq j$. The observed marked point process is specified by the conditional intensity rate $\lambda_{X(t)} f_{X(t)}(z)$, where $\lambda_{X(t)}$ is the conditional intensity rate of the ground process $N(t)$ and $f_{X(t)}(z)$ is the probability density of the mark with respect to a reference measure μ on the mark space $\mathcal{Z}$ conditional on the current state of $X(t)$. The initial distribution $\pi' = (\pi_1,\ldots,\pi_r)$ of the Markov chain is chosen according to the stationary vector which satisfies $\pi'\mathbf{Q} = 0$ and

$\pi'\mathbf{1} = 1$, where $\mathbf{1}$ is a column vector with all entries being unity. The internal history generated by this process is denoted by $\mathcal{F} = \{\mathcal{F}_t\}_{t\geq 0}$. The order r of the Markov chain is also called the order of the MMPP. Given the occurrence times and associated marks (t_i, z_i), $i = 0, 1, \ldots, n$ over $[0, T]$ and assuming $t_0 = 0$, the likelihood is obtained as following:

Let $X_k = X(t_k)$ and $Y_k = t_k - t_{k-1}$. Note that the sequence $(X_i, Y_i, Z_i)_{i=1}^n$ which is equivalent to the sequence $(X_i, t_i, Z_i)_{i=1}^n$ forms a Markov sequence. For any $y > 0$ and Borel set $B \in \mathcal{B}(\mathcal{Z})$, the transition probability for this sequence is given by

$$\begin{aligned} &P\{X_k = j, Y_k \leq y, Z_k \in B | X_{k-1} = i\} \\ &= P\{X_k = j, Y_k \leq y | X_{k-1} = i\} P\{Z_k \in B | X_k = j\} \\ &= \int_0^y \int_B \mathbf{e}_i' \exp\{(\mathbf{Q} - \Lambda)t\} \Lambda \Upsilon(z) \mathbf{e}_j \mu(\mathrm{d}z)\mathrm{d}t, \end{aligned} \tag{1}$$

where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$, $\Upsilon(z) = \mathrm{diag}(f_1(z), \ldots, f_r(z))$ and $\mathbf{e}_i$ is a unit column vector with the ith entry being unity. See Meier-Hellstern (1987) for the derivation of the second equality in the above equations. We characterize a MMPP by $(\pi, \mathbf{Q}, \Lambda, \theta)$, where θ are parameters in $\Upsilon(z)$. Given that the initial distribution is stationary vector π, the likelihood for the observations $(Y_i, z_i)_{i=1}^n$ is written by

$$L(\pi, \mathbf{Q}, \Lambda, \theta) = \pi' \exp\{(\mathbf{Q} - \Lambda)Y_1\} \Lambda \Upsilon(z_1) \cdots \exp\{(\mathbf{Q} - \Lambda)Y_n\} \Lambda \Upsilon(z_n)\mathbf{1}. \tag{2}$$

To evaluate the likelihood and other statistics involved in MMPP with marks efficiently, we introduce the forward and backward probabilities as in the context of the discrete-time hidden Markov models. The forward and backward probability densities are defined by $\alpha_t(i) \hat{=} Pr\{$ points occur at $t_1, \ldots, t_{N(t)}$ with marks $z_1, \ldots, z_{N(t)}$ in $(0, t]$ and $X(t) = i\}$ and $\beta_t(j) \hat{=} Pr\{$ points occur at $t_{N(t)+1}, \ldots, t_{N(T)}$ with marks $z_{N(t)+1}, \ldots, z_{N(T)}$ in $(t, T]$ given $X(t) = j\}$, respectively. Denote $L_k = \exp\{(\mathbf{Q} - \Lambda) Y_k\} \Lambda \Upsilon(z_k)$. For $0 < t < T$, the forward and backward probabilities are written by $\alpha_t(i) = \pi' L_1 \cdots L_{N(t)} \exp\{(\mathbf{Q} - \Lambda)(t - t_{N(t)})\}\mathbf{e}_i$ and $\beta_t(j) = \mathbf{e}_j' \exp\{(\mathbf{Q} - \Lambda)(t_{N(t)+1} - t)\} \Lambda \Upsilon(z_{N(t)+1}) L_{N(t)+2} \cdots L_n \mathbf{1}$ respectively. The likelihood in terms of this device is obviously $L(\pi, \mathbf{Q}, \Lambda, \theta) = \sum_{i=1}^r \alpha_t(i)\beta_t(i)$ for all $t \in (0, T)$.

The likelihood is invariant under the permutation of states. It is still possible that seemingly different MMPPs may have the same laws. For the MMPP with marks, two point processes $N(t, z)$ and $\overline{N}(t, z)$ are equivalent if and only if their conditional intensity rates $\lambda_{X(t)} f_{X(t)}(z)$ and $\overline{\lambda}_{X(t)} \overline{f}_{X(t)}(z)$ have the same laws. The question of under what conditions two MMPPs are equivalent can be translated into the same question of discrete-time hidden Markov models through Poisson randomization of the continuous-time Markov chain $X(t)$ and utilizing results of Ito et al. (1992) on the identifiability of discrete-time hidden Markov model (see also Ryden (1996b)).

3 Parameter estimation through EM algorithm

The EM algorithm (Dempster et al. 1977) is an iterative method for MLE in incomplete data problem which bears several appealing properties. It is easily implemented

since it only involves two steps in each iteration: taking expectations of the complete data likelihood with respect to missing data and maximizing the conditional expectation which in many cases is in explicit form, particularly for the exponential family. Furthermore, it is numerically stable for that each iteration returns with an increased likelihood. More exactly, let $\log L^c(\phi|\mathbf{X}, \mathbf{Y})$ denote the complete data log-likelihood, in which $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{Y} = (Y_1, \ldots, Y_m)$ are observations and missing data, respectively. In E-step, the expectation of the complete data log-likelihood given the observations, namely $Q(\phi) = \mathrm{E}(\log L^c(\phi|\mathbf{X}, \mathbf{Y})|\mathbf{X})$, is evaluated. In M-step, $Q(\phi)$ is maximized with respect to ϕ. Then the iteration steps are repeated until some convergence criterion is met; see McLachlan and Krishnan (1997) for a comprehensive introduction of the EM algorithm and its extensions.

The key point of using the EM algorithm, as Asmussen et al. (1996) and Ryden (1996a) point out, is taking the whole trajectory of the Markov chain $X(t)$ as missing data. Here, we omit the detailed derivation of the likelihood of the Markov chain and the complete likelihood of MMPP. Readers are suggested to refer to Ryden (1996a). Note that the complete likelihood is partitioned into the likelihood of the underlying Markov chain $X(t)$ and the conditional likelihood of MMPP arrivals associated with marks conditioned on $X(t)$. The complete data log-likelihood of this extended MMPP can be summarized by a group of sufficient statistics in the form:

$$\begin{aligned}
&\log L^{\mathcal{C}}(\pi, \mathbf{Q}, \Lambda, \theta) \\
&= \sum_{i=1}^{r} \left\{ \log \pi_i \mathbf{1}\{X(0) = i\} - (q_i + \lambda_i) T_i + \sum_{j:j\neq i} N_{ij} \log q_{ij} + N_i \log \lambda_i \right\} \\
&\quad + \sum_{i=1}^{n} \log f_{X_i}(z_i) \\
&= \log L_1^{\mathcal{C}}(\pi, \mathbf{Q}, \Lambda) + \log L_2^{\mathcal{C}}(\theta), \qquad (3)
\end{aligned}$$

in which $T_i \hat{=} \int_0^T \mathbf{1}\{X(t) = i\}\,\mathrm{d}t$ is the sojourn time of $X(t)$ in state i, $N_{i,j} \hat{=} \#\{t : 0 < t \leq T, X(t-) = i, X(t) = j\}$ is the number of transition times of $X(t)$ from i to j and $N_i \hat{=} \int_0^T \mathbf{1}\{X(t) = i\}\,\mathrm{d}N(t)$ is the number of arrivals occurred in the state i. We limit our attention only on those marks whose distributions have no common parameters with $\log L_1^{\mathcal{C}}(\pi, \mathbf{Q}, \Lambda)$. Such a simplification is sufficient for most practically interested problems at this stage. Therefore, $\log L_1^{\mathcal{C}}(\pi, \mathbf{Q}, \Lambda)$ and $\log L_2^{\mathcal{C}}(\theta)$ can be maximized separately within M-step after taking expectations conditioned on the observations, i.e. the arrival times and associated marks.

In general, we suppose the mark distribution comes from an exponential family distribution in its canonical form $f(z) = h(z) \exp\{\theta^T t(z)\}/c(\theta)$, where the sufficient statistic $t(z)$ and parameter vector θ are d dimensional vectors, $h(z)$ and $c(\theta)$ are scalar functions. The parameter space Ω is a d-dimensional convex set such that $\Omega = \{\theta : \int h(z) \exp\{\theta^T t(z)\}\mathrm{d}z < \infty\}$. Note that in the E-step the conditional expectation $Q_1(\pi, \mathbf{Q}, \Lambda) \hat{=} \mathrm{E}\{\log L_1^{\mathcal{C}}(\pi, \mathbf{Q}, \Lambda)|\mathcal{F}_T\}$ depends only on the expectations of the sufficient statistics (T_i, N_{ij}, N_i) for parameters $\mathbf{Q}$ and Λ. They can be

written as $\mathrm{E}\{N_i|\mathcal{F}_T\} = \sum_{k=1}^{n} \alpha_{t_k}(i)\beta_{t_k}(i)/L$, $\mathrm{E}\{T_i|\mathcal{F}_T\} = \int_0^T \alpha_t(i)\beta_t(i)\,\mathrm{d}t/L$ and $\mathrm{E}\{N_{ij}|\mathcal{F}_T\} = \int_0^T \alpha_t(i)q_{ij}\beta_t(j)\,\mathrm{d}t/L$, see Asmussen et al. (1996). We denote them by N_i^*, T_i^* and N_{ij}^*, respectively, throughout later discussions. After taking conditional expectation of $\log L_2^{\mathcal{C}}(\theta)$ given the history of the marked point process and ignoring terms without θ, we have

$$\begin{aligned} Q_2(\theta) &\hat{=} \mathrm{E}\left\{\log L_2^{\mathcal{C}}(\theta)|\mathcal{F}_T\right\} \\ &= \sum_{k=1}^{r} \mathrm{E}\left\{\theta_k^T\left(\sum_{i=1}^{n} t(z_i)\mathbf{1}(X_i = k)\right) - N_k \log c(\theta_k)\middle|\mathcal{F}_T\right\} + \sum_{i=1}^{n}\log h(z_i) \\ &= \sum_{k=1}^{r}\left\{\theta_k^T\left(\sum_{i=1}^{n}\frac{\alpha_{t_i}(k)\beta_{t_i}(k)}{L}t(z_i)\right) - N_k^*\log c(\theta_k)\right\} + \sum_{i=1}^{n}\log h(z_i). \end{aligned} \tag{4}$$

M-step is followed by maximizing $Q_1(\pi, \mathbf{Q}, \Lambda)$ and $Q_2(\theta)$, respectively, since they have no parameters in common. After an explicit M-step in $Q_1(\pi, \mathbf{Q}, \Lambda)$, the parameters in $\log L_1^{\mathcal{C}}(\pi, \mathbf{Q}, \Lambda)$ are updated by

$$\hat{q}_{ij} = \frac{N_{ij}^*}{T_i^*}, \quad \hat{\lambda}_i = \frac{N_i^*}{T_i^*}, \quad i \neq j,\ 1 \le i,\ j \le r. \tag{5}$$

The resulting equation is same as that given by Ryden (1996a) except for additional marks incorporated into the forward, backward equations to update T_i^*, N_{ij}^* and N_i^*. On maximization of $Q_2(\theta)$ with respect to θ in Ω, it follows that

$$\frac{\partial Q_2(\theta)}{\partial\theta_k} = \sum_{i=1}^{n}\frac{\alpha_{t_i}(k)\beta_{t_i}(k)}{L}t(z_i) - \frac{N_k^*}{c(\theta_k)}\frac{\partial c(\theta_k)}{\partial\theta_k}. \tag{6}$$

Note that $\mathrm{E}_\theta\{t(z)\} = \frac{\partial \log c(\theta)}{\partial\theta}$, the M-step requires θ to be chosen by solving the equation

$$\mathrm{E}_{\theta_k}(t(z)) = \frac{1}{N_k^*}\sum_{i=1}^{n}\frac{\alpha_{t_i}(k)\beta_{t_i}(k)}{L}t(z_i). \tag{7}$$

If the above equation can be solved in Ω, then it is uniquely solvable due to the convexity property of minus the log likelihood of the regular exponential family.

We list several examples to demonstrate applications of the method.

Example 1 Assume each attached mark is an indicator of the class to which the point belongs. Then such a MMPP with discrete marks forms a multivariate MMPP with intensity rate $\lambda_{X(t)}\prod_j (p_{X(t)}^{(j)})^{\mathbf{1}(z=j)}$,where $\sum_j p_k^{(j)} = 1$, $p_k^{(j)} > 0$, $1 \le j \le M$, $1 \le k \le r$. Rewrite the mark distribution in its canonical form $f_k(z) = \exp\{\sum_j \mathbf{1}(z = j)\log p_k^{(j)}\}$ and note that the sufficient statistic $\mathbf{1}(z = j)$ satisfies $\mathrm{E}_{\theta_k}\{\mathbf{1}(z = j)\} = p_k^{(j)}$, One obtains

$$\hat{p}_k^{(j)} = \frac{1}{N_k^*} \sum_{i=1}^{n} \frac{\alpha_{t_i}(k)\beta_{t_i}(k)}{L} \mathbf{1} \quad (z_i = j),$$

see also Jensen (2005).

Example 2 When the mark of the MMPP is exponential distributed variable with probability density $f(z) = \theta \exp\{-\theta z\}\mathbf{1}\{z > 0\}$, the expectation value of the sufficient statistic z given $\theta = \theta_k$ is $\mathrm{E}_{\theta_k}(z) = \frac{1}{\theta_k}$. The explicit EM estimation is given by

$$\hat{\theta}_k = N_k^* \left(\sum_{i=1}^{n} \frac{\alpha_{t_i}(k)\beta_{t_i}(k)}{L} z_i \right)^{-1}. \tag{8}$$

This example is used for modelling the deep earthquakes (see Sect. 6.2).

Example 3 Suppose the mark variable is p-variate normal distribution $N_p(\mu, \Sigma)$ with canonical probability density $\exp\{-\frac{1}{2}(\mathrm{V}(zz')\mathrm{V}(\Sigma^{-1}) - 2\mu'\Sigma^{-1}z)\}\frac{1}{c(\mu,\Sigma)}$, where V is the vectorization of a matrix and $c(\mu, \Sigma) = (2\pi)^{p/2}|\Sigma|^{1/2}\exp\{\frac{1}{2}\mu'\Sigma^{-1}\mu\}$. The expectation of the sufficient statistics (z, zz') for parameter (μ, Σ) is

$$\begin{cases} \mathrm{E}_{\theta_k}(z) = \mu_k; \\ \mathrm{E}_{\theta_k}(zz') = \Sigma_k + \mu_k\mu_k'. \end{cases}$$

Thus, in terms of Eq. (7), the closed form of EM iteration step is given by

$$\begin{cases} \hat{\mu}_k = \dfrac{1}{N_k^*} \displaystyle\sum_{i=1}^{n} \frac{\alpha_{t_i}(k)\beta_{t_i}(k)}{L} z_i; \\ \hat{\Sigma}_k = \displaystyle\sum_{i=1}^{n} \frac{\alpha_{t_i}(k)\beta_{t_i}(k)}{LN_k^*} z_i z_i' - \left(\sum_{i=1}^{n} \frac{\alpha_{t_i}(k)\beta_{t_i}(k)}{LN_k^*} z_i \right)\left(\sum_{i=1}^{n} \frac{\alpha_{t_i}(k)\beta_{t_i}(k)}{LN_k^*} z_i' \right). \end{cases}$$

Example 4 The above procedure is not automatically applicable for those marks from non-exponential family distribution. For instance, the t-distribution does not belong to the exponential family, but it has many applications in applied statistics. We assume the mark variable Z comes from the multivariate t-distribution $t_p(\mu, \Sigma, \nu)$ with location parameter μ, positive definite inner matrix Σ and ν degrees of freedom. Given weight u, $Z|u \sim N(\mu, \Sigma/u)$, where the random variable U corresponding weight u is distributed as $\Gamma(\frac{1}{2}\nu, \frac{1}{2}\nu)$. For known ν, a closed form EM step can be obtained for the T-type mark since $Z|u$ and u are from exponential family distribution. We treat not only the whole trajectory of the Markov chain but also the mark Z as missing data.

Then the second part $\log L_2^{\mathcal{C}}$ of the complete-data log likelihood is written by

$$\log L_2^{\mathcal{C}}(\theta) = c - \sum_{k=1}^{r} \frac{N_k}{2} \log\left|\Sigma_k\right| - \sum_{i=1}^{n}\sum_{k=1}^{r} \frac{u_i}{2} \mu_k' \Sigma_k^{-1} \mu_k \mathbf{1}\{X_i = k\} - \sum_{i=1}^{n}\sum_{k=1}^{r} \frac{u_i}{2} \left(Z_i' \Sigma_k^{-1} Z_i - 2\mu_k' \Sigma_k^{-1} Z_i\right) \mathbf{1}\{X_i = k\} + \sum_{i=1}^{n} g(u_i), \tag{9}$$

where c and $g(u_i)$ are terms without unknown parameters. In the E-step, the conditional expectation will be taken both over $(Z_i Z_i', Z_i)\mathbf{1}(X_i = k)$ and u_i given the observed marked point process and current values of parameters. It turns out that the conditional distribution of U_i is distributed as $U_i \big| Z_i \sim \Gamma(m_1, m_2)$, where $m_1 = \frac{1}{2}(\nu + p)$ and $m_2 = \frac{1}{2}(\nu + (Z_i - \mu)'\Sigma^{-1}(Z_i - \mu))$. So we have $U_i^* \hat{=} \mathrm{E}(U_i \big| Z_i) = \frac{\nu+p}{\nu+(Z_i-\mu)'\Sigma^{-1}(Z_i-\mu)}$. After the execution of an explicit E-step, each U_i is replaced by U_i^* and if we denote $\frac{\alpha_{t_i}(k)\beta_{t_i}(k)}{L}$ by $p_{t_i}(k)$, the parameters are updated by

$$\begin{cases} \hat{\mu}_k = \sum_{i=1}^{n} \dfrac{U_i^* Z_i\, p_{t_i}(k)}{\sum_{i=1}^{n} U_i^* p_{t_i}(k)}; \\ \hat{\Sigma}_k = \sum_{i=1}^{n} \dfrac{U_i^* (Z_i Z_i' - 2\hat{\mu}_k Z_i' + \hat{\mu}_k \hat{\mu}_k{}')\, p_{t_i}(k)}{N_k^*}. \end{cases}$$

4 Statistical inference on the state process and observed point process, model evaluation

There exist state smoothers for MMPP such as those given by Elliott and Malcolm (2005). Much the same as in the context of discrete-time hidden Markov models, a computationally efficient algorithm, fixed point smoothing algorithm, is available to estimate the probability of the underlying Markov chain in a given state at a specific time conditioned on all available observations, which only involves the forward and backward probabilities (see MacDonald and Zucchini (1997, pp 85)). Let $\hat{p}_t(i) = P\{X(t) = i | \mathcal{F}_T\}$. After taking conditional expectation of $\mathbf{1}\{X(t) = i\}$ conditioned on observations, one obtains

$$\hat{p}_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^{r} \alpha_t(j)\beta_t(j)}, \tag{10}$$

which is obviously continuous with respect to t.

For the purpose of simulation, model evaluation and prediction based on the intensity rate of MMPP, we estimate the intensity rate of MMPP by the conditional expectation of $\lambda_{X(t)} f_{X(t)}(z) = \sum_{i=1}^{r} \lambda_i f_i(z)\mathbf{1}\{X(t) = i\}$ conditioned on the observations, namely,

$$\hat{\lambda}(t,z)\hat{=}\sum_{i=1}^{r}\hat{\lambda}_i\hat{f}_i(z)p_t(i)=\sum_{i=1}^{r}\hat{\lambda}_i\hat{f}_i(z)\frac{\alpha_t(i)\beta_t(i)}{\sum\alpha_t(i)\beta_t(i)}. \tag{11}$$

It is a weighted sum of all intensity rates $\hat{\lambda}_i\hat{f}_i(z)$ which suggests MMPP is a mixing process with Markov dependence. From Eq. (11), the estimated ground intensity is given by $\hat{\lambda}(t)=\sum_{i=1}^{r}\hat{\lambda}_i\frac{\alpha_t(i)\beta_t(i)}{L}$ and the conditional mark density is given by $\hat{f}(t,z)=\sum_{i=1}^{r}\hat{f}_i(z)\frac{\alpha_t(i)\beta_t(i)\hat{\lambda}_i}{\sum\alpha_t(k)\beta_t(k)\hat{\lambda}_k}$.

Rescaling a marked point process (Vere-Jones and Schoenberg 2004) technique is very useful for point process model evaluation (Baddeley et al. 2005). The detection of deviations of the residual point process (Ogata 1988) from a standard process, the homogeneous unit rate Poisson process with stationary mark, forms a key tool to suggest deficiencies and even illuminate possible improvements or alternatives of current models. Vere-Jones and Schoenberg (2004) show that under quite general conditions by changing each point from (t_k,z_k) to $(\int_0^{t_k}\lambda(t,z_k)\mathrm{d}t,z_k)$, the collection of these rescaled ones forms a stationary compound Poisson process with unit ground intensity and stationary mark distribution on the rescaled space. Especially when the mark admits a univariate continuous cumulative distribution function $F(z)$ with respect to a reference probability measure $\mu(dz)$, the doubly transformed points $(\int_{t_0}^{t_k}\lambda(t,z_k)\mathrm{d}t,F(z_k))$ form a unit rate planar Poisson process over $\mathcal{R}\times[0,1]$ (see also Daley and Vere-Jones (2008, chapter 14.6)). On numerically evaluating

$$\Lambda(t_k,z_k)=\int_{t_0}^{t_k}\frac{1}{L}\sum_{i=1}^{r}\lambda_i f_i(z_k)\alpha_t(i)\beta_t(i)\mathrm{d}t, \tag{12}$$

we are computing the integral of the matrix exponential involved in MMPP with marks, a numerical technique which can be implemented through matrix eigenvalue decomposition (Ryden 1996a), Poisson randomization (Klemm et al. 2003) or a certain matrix exponential in higher order (Van Loan 1978). Then the usual tests for Poisson processes such as the K-functions, L-functions and nearest neighbor tests etc. can be applied to evaluate the goodness-of-fit of the proposed hidden Markov models (see Cressie (1991)). For instance, the estimated K-functions $K(d)$ which indicates the proportion of paired points per unit area within a specified distance d is a powerful test to detect clustering or regularity appearing in the spatial point pattern.

5 Implementation and simulation study

5.1 Implementation

One step of EM iterations is carried out as follows:

1) In E-step, update the forward, backward probabilities: let $\alpha(0) = \pi$, $\alpha(k) = \alpha(k-1)\exp\{(\mathbf{Q} - \Lambda)Y_k\}\Lambda\Upsilon(z_k)$, $\beta(n+1) = \mathbf{1}$ and $\beta(k) = \exp\{(\mathbf{Q} - \Lambda)Y_k\}\Lambda\Upsilon(z_k)\beta(k+1)$, for $k = 1, \ldots, n$.
2) Calculate $A_{ij} = \sum_{k=1}^{n} \alpha(k-1) \int_{t_{k-1}}^{t_k} \exp\{(\mathbf{Q} - \Lambda)(t - t_{k-1})\}\mathbf{e}_i\mathbf{e}_j' \exp\{(\mathbf{Q} - \Lambda)(t_k - t)\}\Lambda\Upsilon(z_k)\,\mathrm{d}t\beta(k+1)$, $B_i = \sum_{k=1}^{n} \alpha(k)\mathbf{e}_i\mathbf{e}_i'\beta(k+1)$ and $C_i = \sum_{k=1}^{n} \alpha(k)\mathbf{e}_i\mathbf{e}_i'\beta(k+1)t(z_k)$.
3) In M-step, update the parameters by: $\hat{q}_{ij} = q_{ij}^0 \frac{A_{ij}}{A_{ii}}$, $\hat{\lambda}_i = \frac{B_i}{A_{ii}}$ and solving $\mathrm{E}_{\theta_i}(t(z)) = \frac{C_i}{B_i}$ to obtain θ_i, where q_{ij}^0 is obtained from the previous EM steps.

To avoid the overflow or underflow problem in computation, it is necessary to employ scale procedures in the EM iteration steps (see Roberts et al. (2006)).

5.2 Simulation study

For MMPP without marks, it is noted by Lu (2009) that the better the intensity rates are separated, the better the estimates are close to the true values for given number of observations when $\mathbf{Q}$ fixed. Intuitively, due to an increase of the mutual information between the observed point process and the underlying Markov chain brought by additional marks, the estimates of MMPP attached by state-dependent marks should be better than that of MMPP without marks when the infinitesimal generator $\mathbf{Q}$ and the intensity rates λ_i remain unchanged. To confirm this, we illustrate several simulations to compare the efficiencies and accuracies of the estimates for several types of MMPPs: one of them is MMPP without marks and the others are MMPP associated with several different types of state-dependent marks whose $\mathbf{Q}$ matrix and intensity rates Λ are identical. We also demonstrate that among MMPP with marks whose $\mathbf{Q}$ matrix and intensity rates Λ are identical, the better the mark distributions are separated , the better the estimates are close to the true values of the model parameters even for different types of marks. How well the mark distributions are separated is measured by the Kullback–Leibler divergence.

One approach to appraise how the accuracies and efficiencies of the MLEs vary according to different types of marks attached to MMPP is by evaluating the observed Fisher information, i.e. minus the second-order derivative of the log likelihood function given the observations. However, directly evaluating the Fisher information or the observed Fisher information requires the derivatives of matrix exponential and its products, which is unfortunately numerically complicated. Existing approximation methods are not accurate enough to the level to enable delicate effects posed from different types of marks to be compared and demonstrated in detail. We turn to Monte Carlo simulations.

Assume all types of MMPP with or without marks in the simulation have identical infinitesimal generator $\mathbf{Q} = \begin{pmatrix} -1 & 1 \\ 0.5 & -0.5 \end{pmatrix}$ and the intensity rates $\Lambda = \mathrm{diag}(5, 1)$. We attach four pairs of marks to MMPP observations, e.g. $(N(0,1), N(1,1))$, $(N(0,1), N(1, 0.5^2))$, $(\exp(2.5), \exp(3))$ and $(\exp(1), \exp(10))$ which are denoted by MMPP(1), MMPP(2), MMPP(3) and MMPP(4), respectively, throughout later discussions. MMPP without marks is denoted by MMPP(0). The Kullback–Leibler divergences $d_{KL}(\cdot, \cdot)$ of these four pairs of marks comparatively satisfy

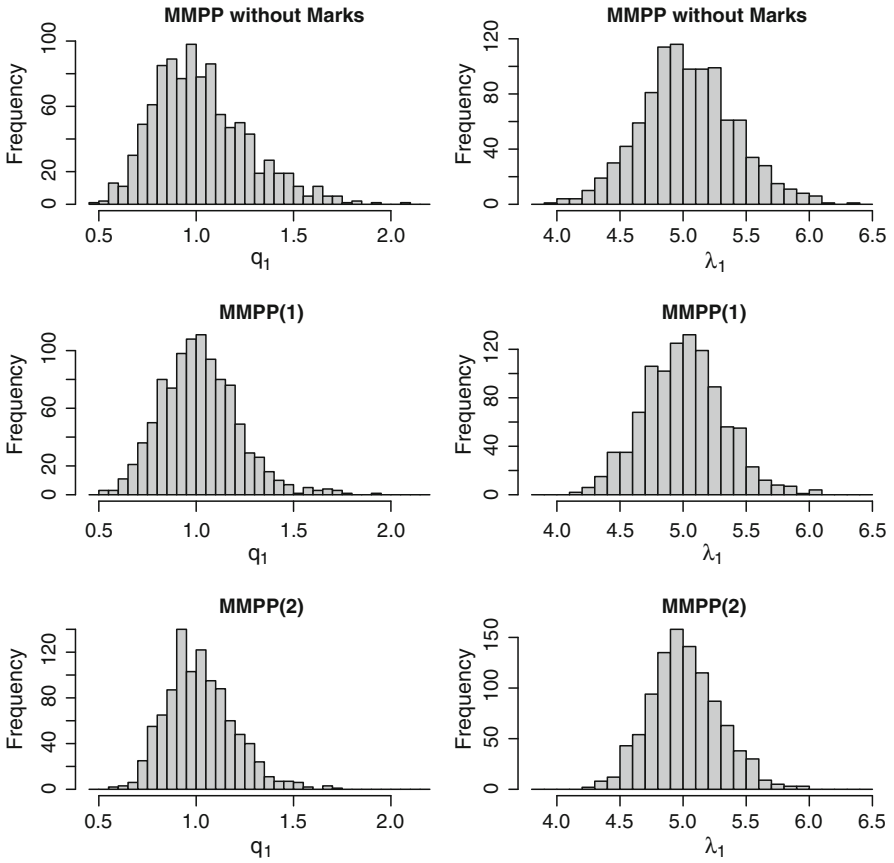

Fig. 1 Histograms of estimated q_1 and λ_1 for 1000 replicates obtained from simulated series of events with 1000 observations

$d_{KL}(\exp(2.5), \exp(3)) < d_{KL}(N(0, 1), N(1, 1)) < d_{KL}(N(0, 1), N(1, 0.5^2)) < d_{KL}(\exp(1), \exp(10))$. Then we generate many series of events with 1000 observations according to the parameters of MMPP and MMPP with the four pairs of marks listed above, each one of them repeats 1000 times. For each series of MMPP observations, the parameter estimation is implemented through the EM algorithm.

The histograms of the estimates for MMPPs attached by marks and MMPP without marks clearly indicate the effects of different marks on the estimation errors (see Figs. 1 and 2). From the histograms, it is evident that the estimates are more and more centralized about the true values of the parameters when the paired marks are more and more separated in terms of the Kullback–Leibler divergence (see also the standard deviations of the estimates in Table 1). The standard deviations of the estimates in Table 1 confirm that the accuracies and efficiencies of MLEs are in the order of the Kullback–Leibler divergences of the paired marks.

It is worth noting that the difference of the standard deviations of the estimates between MMPP(0) and MMPP(3) are very small, suggesting very weak leverage effect

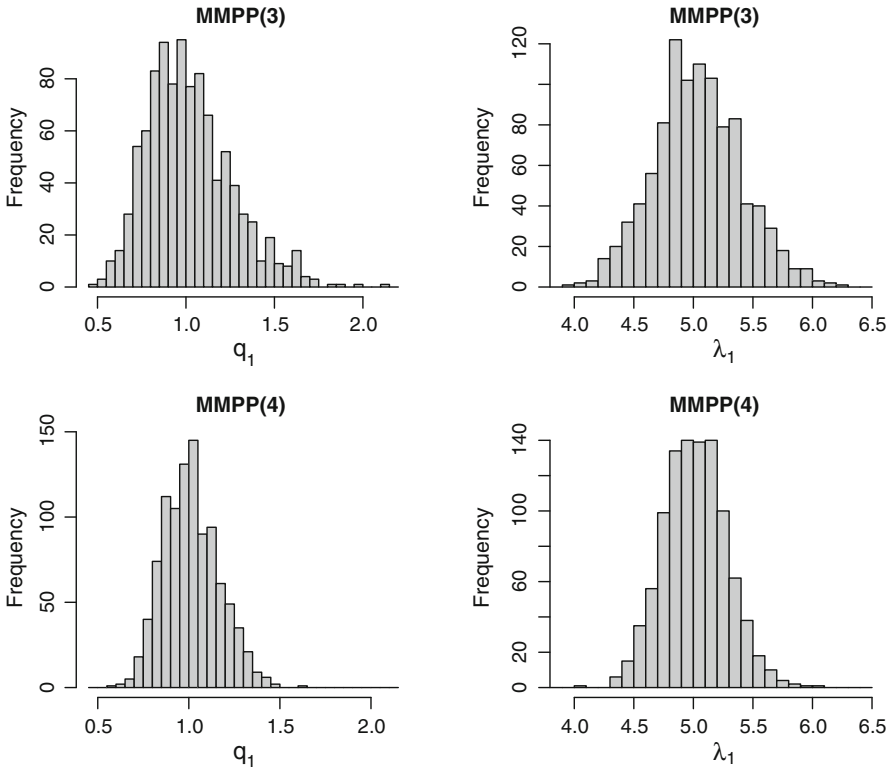

Fig. 2 Histograms of estimated q_1 and λ_1 for 1000 replicates obtained from simulated series of events with 1000 observations

Table 1 Standard deviations for Q and Λ according to 1000 replicates obtained from simulated series of events with 1000 observations

SD	MMPP(0)	MMPP(1)	MMPP(2)	MMPP(3)	MMPP(4)
q_1	0.242	0.198	0.174	0.240	0.152
q_2	0.123	0.102	0.076	0.122	0.067
λ_1	0.368	0.314	0.276	0.366	0.265
λ_2	0.123	0.096	0.075	0.122	0.068

posed by the paired marks (exp(2.5), exp(3)). The simulation study for MMPP(3) is an analogy to the real application in the next section.

Similar comparisons for MMPP(0) and MMPP(3) with larger data sizes are demonstrated. When the data sizes n are increased to 5000, the histograms and standard deviations of estimates for MMPP(0) and MMPP(3) are given in Fig. 3 and Table 2, respectively. From Table 2, it suggests that the estimates are roughly convergent to the true values of the parameters at the rate of $1/\sqrt{n}$. When comparing MMPP(0) and MMPP(3), It is more clear that the leverage effects posed by the paired marks (exp(2.5), exp(3)) on the estimation is even weaker.

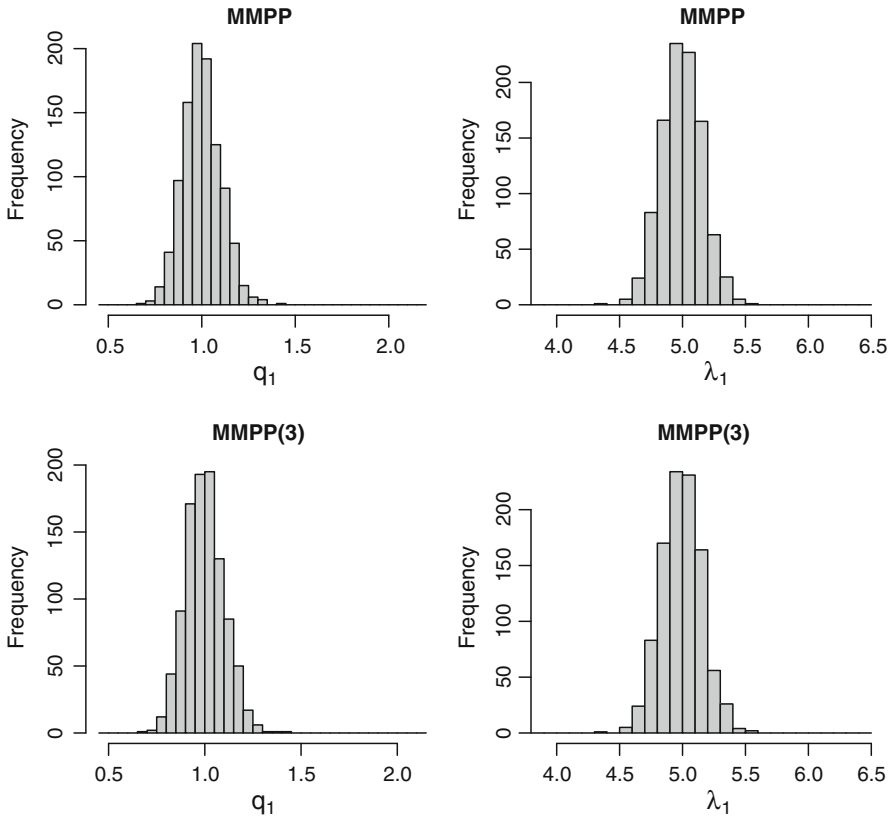

Fig. 3 Histograms of estimated q_1 and λ_1 for 1000 replicates obtained from simulated series of events with 5000 observations

Table 2 Standard deviations for Q and Λ according to 1000 replicates obtained from simulated series of events with 5000 observations

SD	q_1	q_2	λ_1	λ_2
MMPP(0)	0.1004	0.0525	0.1612	0.0558
MMPP(3)	0.0997	0.0521	0.1605	0.0554

6 Applications to the deep earthquakes

6.1 Catalogue completeness and homogeneity

The data used in this study are selected from New Zealand catalogue within the confine defined in Fig. 4 starting from 1945 to 2007 at depth greater than 45 km with magnitude greater than 5 in Richter scale. Generally speaking, the selected deep events are well located, either under the land area or near the shore. However, due to the sparse monitoring network coverage and limited detectability of the instruments in

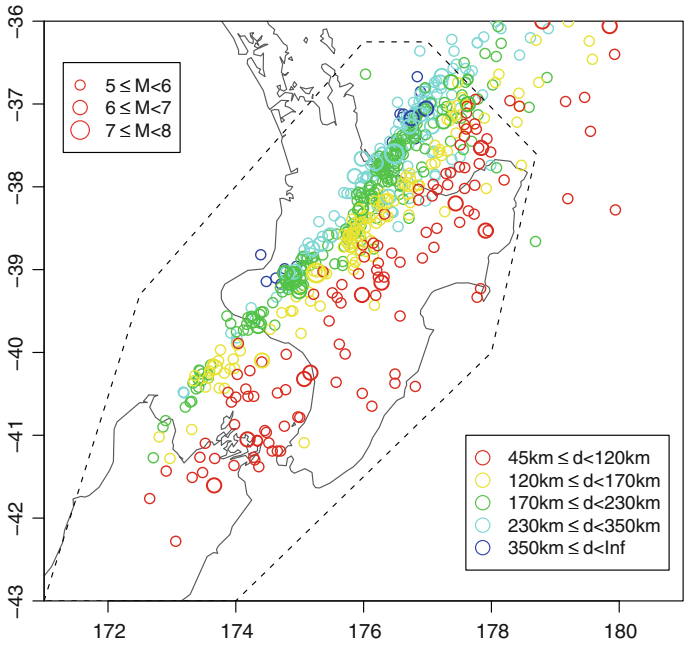

Fig. 4 Epicentral distribution of the deep earthquakes. The legends indicate the depths d and magnitudes M of the deep earthquakes. The deep earthquakes with their epicentre circled by *dash lines* and the lower boundary of the map are used in this study

early periods, the epicentres might be in significant error. So, we limit our analysis only on the temporal patterns rather than both spatial and temporal patterns together. Before doing so, the completeness and homogeneity of the catalogue data are evaluated according to the coverage information of monitoring networks in the early period and cusum (cumulative summation) statistics.

From the study of Vere-Jones et al. (1964), it is concluded that the catalogue coverage would be incomplete below surface 50 km for events with magnitude less than five between 1942 and 1961. The events we select are still within the confines of reasonable coverage in this early period. Afterwards, the catalogue coverage is improved for smaller events because of the upgradations of the monitoring networks in 1960s and late 1980s. We further confirm this by using magnitude frequency relation and reverse Cusum statistics.

The left top plot in Fig. 5 indicates the selected events are magnitude complete since the logarithms of the frequencies of magnitudes nearly form a straight line as supposed for a complete catalogue according to Gutenberg–Richter magnitude-frequency law. The right-top plot gives the cumulative magnitude of events at three levels with magnitude greater than 5, 5.5 and 6, respectively. The cumulative magnitude curves show steeper slopes at late 1980s when the monitoring networks were under upgradations. It is mainly ascribed to a period of active seismicity starting from late 1980s rather than improved instrumental detectability. The bottom row plots are cumulative frequency and(or) magnitude deducted by a reference mean value $\overline{x}$ obtained by averaging over

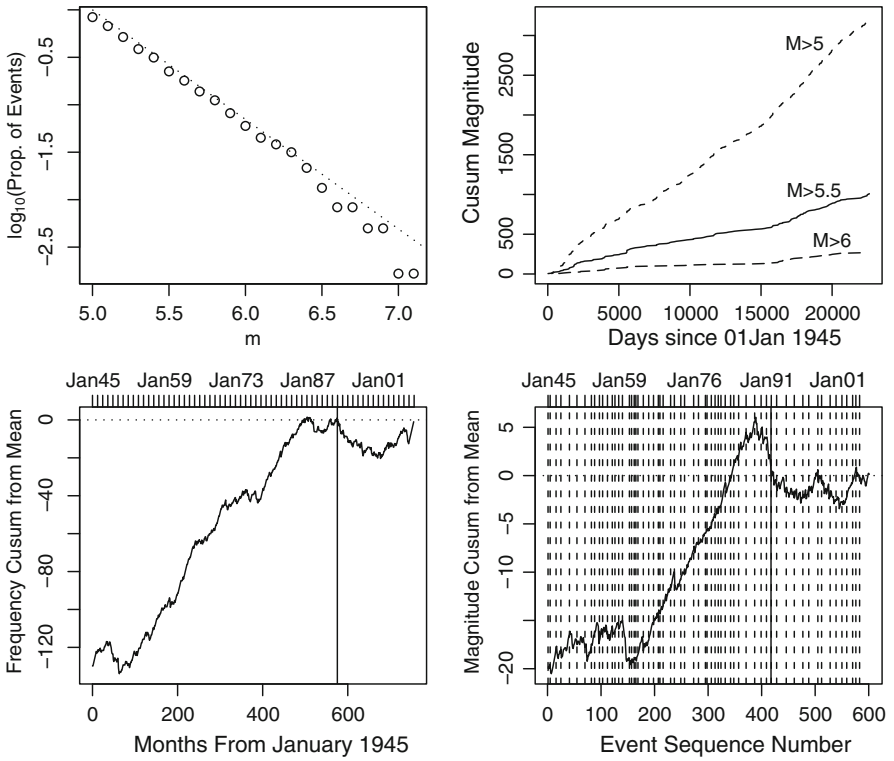

Fig. 5 Frequency versus magnitude, frequency, magnitude Cusum statistics. Events are selected from New Zealand catalogue within the confine defined in Fig. 4 and depth greater than 45 km from January 1, 1945 up to January 1, 2007. *Vertical solid lines* in the bottom mark the beginning of training period for obtaining $\bar{x}$. *Vertical dash lines* in the right bottom delimit the calendar years

most recent period when the catalogue is believed to be homogeneous and complete. These reverse cumulative plots should be read in a reverse direction, from the present to the back, since they are obtained by calculating the corresponding quantities $\sum_i (x_i - \bar{x})$ from the most recent period to the past. The quantities that we are interested in should not be the absolute values, but the slopes which give indications of when and how the earthquake activities are changed relative to the most recent period. Generally, a fast dropping slope in cumulative frequency combined with a fast ascending slope in cumulative magnitude at early periods may suggest catalogue incompleteness due to limited instrumental detectability and sparse monitoring networks at early periods.

Here, both the cumulative monthly frequency $\sum(f_i - \overline{f})$ and cumulative magnitude $\sum(m_i - \overline{m})$ based on event by event scale show dropping slopes between early 1960s and late 1980s. This should be interpreted as seismic quiescent period as pointed out by Reyners (1989) rather than catalogue incompleteness since both plots show nearly horizontal slopes in the early period.

In conclusion, the catalogue might still miss some events, but the effect of the missing tends to be limited and should not affect general conclusions.

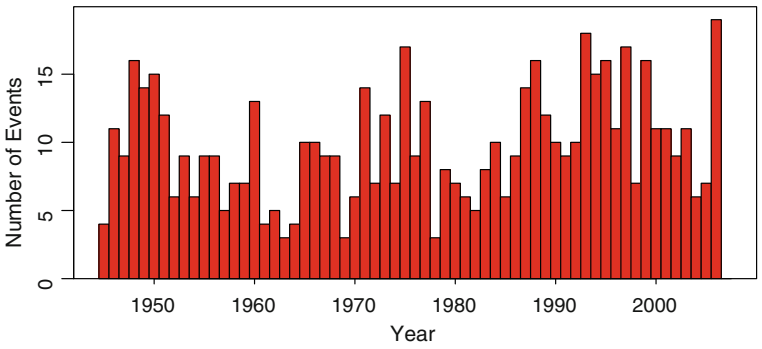

Fig. 6 Yearly counts of the deep events

6.2 Model and model fitting

Generally speaking, the deep earthquakes rarely have following sequences with numerous small aftershocks which decay in terms of Omori's law. Some deep earthquakes even do not have aftershocks at all (see Frohlich (2006)). Partly because of these factors, those models usually used for the shallow earthquakes fit the deep earthquakes less satisfactory. Instead, the main evolution feature of the deep earthquakes around the North Island of New Zealand is that the deep earthquakes vary from time to time, active in one period, relatively quiescent in another (see the yearly occurrence numbers of events in Fig. 6). The mechanism and reason behind this time-varying behaviour are still not well-understood. It is this lack of interpretability forming the main motivation of applying the "hidden" Markov model. We suggest a switching Poisson model, a second-order MMPP, to characterize the occurrence patterns of New Zealand deep earthquakes.

The occurrence times and magnitudes of the deep earthquakes are fitted through a second-order MMPP with marks; two levels of seismicity are associated with seismic active state (state 1) and seismic quiescent state (state 2), respectively. In this model, the magnitudes of the earthquakes are treated as marks which are exponentially distributed variables according to Gutenberg-Richter law (see the left top of Fig. 5). The magnitude distributions are also dependent on the underlying Markov process. Parameter estimation of this model is implemented according to equation (8) in Example 2. Whether the magnitude distribution is also varying simultaneously with the occurrence rate is demonstrated by comparing the model performance of MMPP with state-dependent marks , MMPP with state-independent marks and Poisson model with stationary marks according to some information theoretical criterion.

The initial values used in the EM iteration procedure should be selected carefully. A natural approach is to approximate the continuous time process by time discretization as set out in Deng and Mark (1993). In this procedure, the time interval is divided into many small bins and the number of arrivals in each bin is counted. By assuming that the state transitions occur only at bin boundaries, the model can be treated as discrete-time hidden Markov model with Poisson observations. After performing fixed point smoothing procedure for discrete-time hidden Markov models, the distri-

Table 3 Estimates of second-order MMPP with state-dependent marks (model (1)) and MMPP with state-independent marks (model (2))

	q_1	q_2	λ_1	λ_2	θ_1	θ_2	$\log L$
model(1)	0.1824	0.2299	7.491	12.928	3.06	2.38	760.455
model(2)	0.1637	0.1483	7.225	12.423	2.65	$\theta_1 = \theta_2$	758.329

butions of marks can be specified accordingly. Therefore, the initial values in the EM iteration steps for parameters of the marks are obtainable. Then, the transition rate q_{ij} and the intensity rates λ_i can be straight-forwardly converted from the discrete-time hidden Markov model, i.e. $q_{ij} = p_{ij}/h$ and $\lambda_i = b_i/h$, where p_{ij} is the state transition probability of the discrete-time Markov chain, b_i is the Poisson parameter of the discrete-time hidden Markov model and h is the bin width. We also use different initial values in the EM iteration steps to validate the convergence results. Generally speaking, the initial values have no significant influence on the convergence results, and the iteration solutions begin to converge within hundreds steps. The estimated parameters $\mathbf{Q}$, Λ and the decaying parameters θs in the magnitude distributions are listed in Table 3. The time scale in the estimation is about 1 year (365 days).

The probabilities of the underlying Markov chain in the second state (the seismic active state) are obtained by the fixed point smoothing algorithm according to equation (10) in Sect. 4 which gives the probabilities of the underlying Markov chain in the second state. We evaluate this probability at many pre-selected points and connect them by straight lines. Since these quantities are continuous with respect to t, this approach is sufficient to indicate the evolution of the underlying process. The evolution of the observed ground process are similarly demonstrated according to Eq. (11). The magnitude versus time, the estimated probabilities of the underlying Markov chain in the second state (seismic active state) and the estimated ground intensity are given in Fig. 7.

The AIC values for the three models, i.e. MMPP with state-dependent marks, MMPP with state-independent marks and Poisson model with stationary marks are listed in Table 4. From Table 4, it is suggested that the MMPP with state-independent marks is much better than Poisson model with stationary marks. The MMPP with state-dependent marks outperforms both MMPP with state-independent marks and Poisson model with stationary marks. Considering the marks can only exert very limited leverage effects on the estimation as suggested in Figs. 1, 2 and 3 and Tables 1 and 2, the difference of AIC between MMPP with state-dependent marks and MMPP with state-independent marks is still not negligible.

By 10000 Monte-Carlo simulations, the likelihood ratio of MMPP with state-dependent marks and MMPP with state-independent marks for an event with magnitude greater than 6 is approximately 1.36 when assuming the seismicity is in active state. But when assuming the seismicity being relatively quiescent, the likelihood ratio for an event with magnitude greater than 6 is decreased to about 0.698. These simulations suggest that in the second state, it is more likely for a medium size or large earthquake occurring.

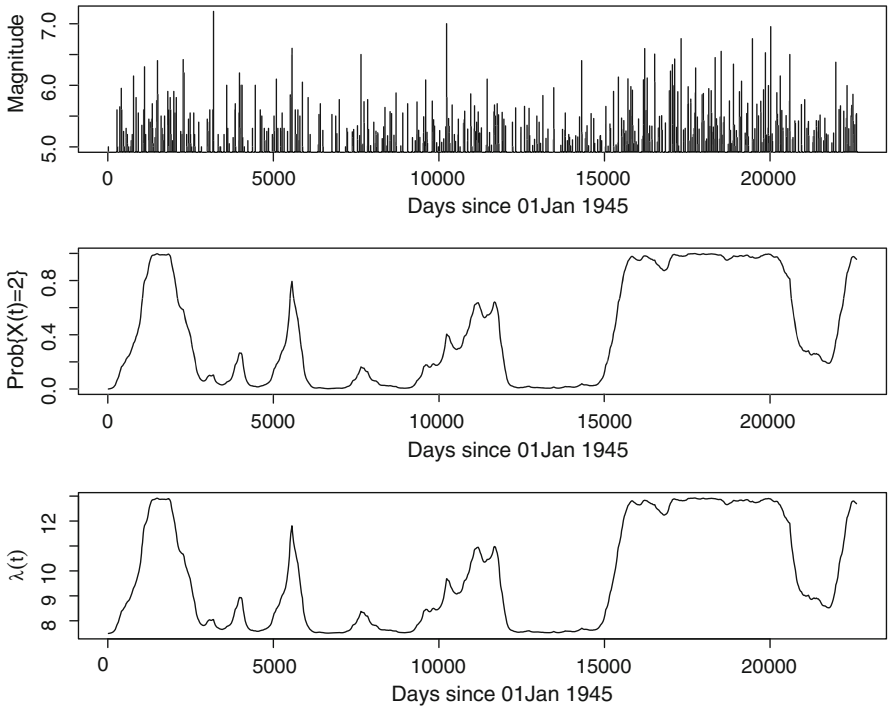

Fig. 7 Magnitude versus time plot, estimated probabilities of $X(t)$ in state 2 and estimated intensity rate of the ground process

Table 4 The AIC values of the three nested models: MMPP with state-dependent marks (model (1)), MMPP with state-independent marks (model (2)) and Poisson model with stationary marks (model (3)) for the deep earthquakes

Model	Model (1)	Model (2)	Model (3)
AIC	−1508.9	−1506.6	−1489.9

The residual point process obtained by rescaling the arrival times and marks according to Eq. (12) is demonstrated in Fig. 8. We use the estimated K-function $K(d)$ which indicates the proportion of paired points per unit area within a specified distance d to detect deviations of the residual point process from a unit rate planar Poisson process. After performing K-test via Ripley's correction for the boundary effect over the doubly rescaled marked point process, no obvious clustering or regularity pattern can be detected in the transformed process (see Fig. 9).

A full third-order MMPP are also considered in this analysis. The estimated parameters and the likelihood are listed in Table 5. From Table 5, it is clearly indicated that even for a reduced model when two of intensity rates and (or) two of the mark distributions are identical, the third-order MMPP will not outperform the second-order MMPP in terms of AIC since the gain of the likelihood is very limited even for the full model, and a reduced model with parameters defined in a restricted parameter space

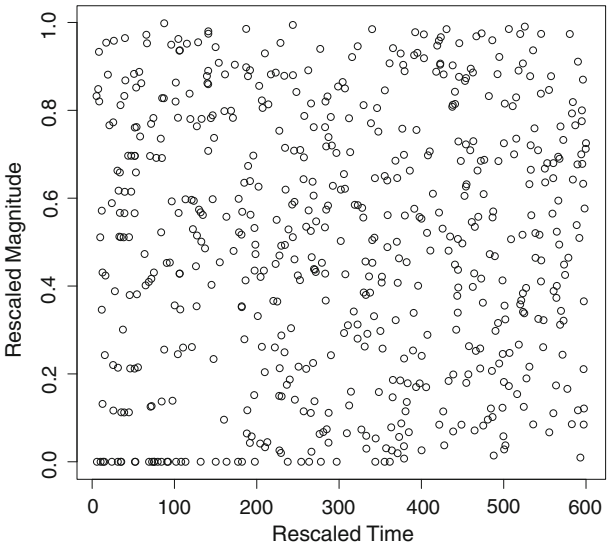

Fig. 8 Doubly rescaled planar point process

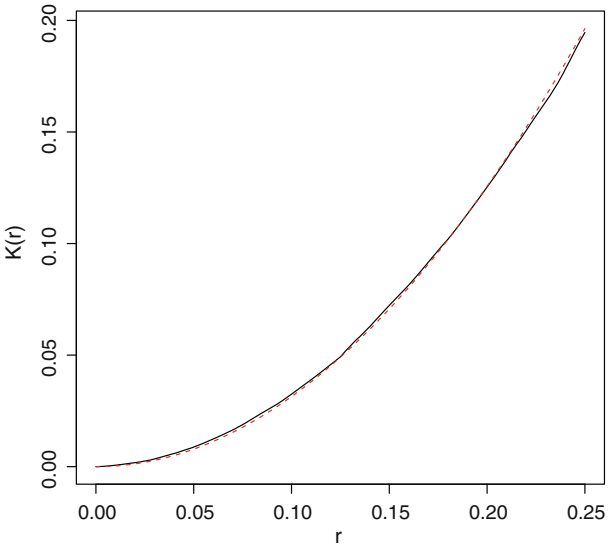

Fig. 9 K-test via Ripley's correction for the boundary effects over the doubly rescaled marked point process. The *solid line* gives the estimated K-function for the doubly rescaled process and the *dash line* indicates the theoretical values of the K-function for a Poisson process

has at least ten parameters, six parameters for the transition rates, four parameters for the occurrence rates and the decaying parameters of the magnitude distributions. Also, in Table 5, two transition rates are very close to zero, which strongly suggests lack of information about the type of state transitions due to limited available observations. At this stage, we will not assume particular parametrization for the Q matrix to allow

Table 5 Estimated parameters of the third order MMPP with state-dependent marks

q_{12}	q_{13}	q_{21}	q_{23}	q_{31}	q_{32}	λ_1
0.742	4e–10	1e–07	0.178	0.136	0.067	3.89
λ_2	λ_3	θ_1	θ_2	θ_3	LL	
8.32	12.967	2.57	3.16	2.33	763.43	

Table 6 95% confidence intervals and standard errors for parameters of model (1) listed in Table 3 obtained from 1000 bootstrap replications

	q_1	q_2	λ_1	λ_2	θ_1	θ_2
C.I.	[0.04, 0.81]	[0.04, 0.86]	[5.8, 8.6]	[10.7, 15.3]	[2.54, 3.79]	[2.34, 3.01]
SE	0.271	0.261	0.757	1.19	0.313	0.172

some q_{ij} being zeros, since we have no prior information of the underlying process about which state transitions are actually avoided. Generally speaking, a third-order MMPP with or without restrictions in the parameter space is over-fit for this relatively small data set (600 observations).

We also evaluate the estimation errors of parameters in MMPP with state-dependent marks (model (1)) appearing in Table 3. The estimation errors are estimated by parametric bootstrap methods. We simulate 1000 series of events within the same observation period as the real data according to the estimated parameters of MMPP (1) in Table 3. Then the parameters of the model are estimated via the EM algorithm for each series of simulated data. From the 1000 bootstrap replicates, we obtain the 95% bootstrap percentile confidence intervals and the standard errors given in Table 6. From Table 6, it is observed that the estimates are slightly unstable for this small data set, particularly for the transition rates in Q matrix.

6.3 Conclusions

From this analysis, we conclude that the main pattern of New Zealand deep earthquakes is its time-varying behaviour of occurrence rates, which is well characterized by a switching Poisson model. The magnitude distribution of the deep earthquakes tends to vary correspondingly with the occurrence rates of events. The associations between the occurrence rates and magnitude distributions suggest that it is more likely for a medium size or large deep earthquake occurring around the North Island in New Zealand when the deep seismicity is in seismic active state.

Acknowledgments This research is motivated by earthquakes modelling. I am grateful to David Vere-jones and David Harte for their many helpful ideas and comments. The suggestions of two referees are acknowledged. This study is supported by New Zealand Institute of Mathematics and its Applications.

References

Asmussen, S., Nerman, O., Olsson, M. (1996). Fitting phase-type distribution via the EM algorithm. *Scandinavian Journal of Statistics, 23*, 419–441.

Baddeley, A. J., Turner, R., Møller, J., Hazelton, M. (2005). Residual analysis for spatial point processes (with discussion). *Journal of the Royal Statistical Society, Series B, 67*, 617–666.

Cressie, N. A. C. (1991). *Statistics for spatial data*. New York: Wiley.

Daley, D. J., Vere-Jones, D. (2008). *An introduction to the theory of point processes* (Vol. II). Berlin: Springer.

Dempster, A. P., Laird, N. M., Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journals of the Royal Statistical Society, Series B, 39*, 1–38.

Deng, L., Mark, J. W. (1993). Parameter estimation for Markov modulated Poisson processes via the EM algorithm with time discretization. *Telecommunication Systems, 1*, 321–338.

Elliott, R. J., Malcolm, W. P. (2005). General smoothing formulas for Markov-modulated Poisson observations. *IEEE Transactions on Automatic Control, 50*, 1123–1134.

Fischer, W., Meier-Hellstern, K. (1993). The Markov-modulated Poisson process (MMPP) cookbook. *Performance Evaluation, 18*, 149–171.

Frohlich, C. (2006). *Deep earthquakes*. UK: Cambridge University Press.

Ito, H., Amari, S. I., Kobayashi, K. (1992). Identifiability of hidden Markov information sources and their minimum degrees of freedom. *IEEE Transactions on Information Theory, 38*(2), 324–333.

Jensen, A. T. (2005). *Statistical inference for doubly stochastic Poisson processes*. PhD Thesis, University of Copenhagen.

Klemm, A., Lindemann, C., Lohmann, M. (2003). Traffic modeling of IP networks using the batch Markovian arrival process. *Performance Evaluation, 54*(2), 149–173.

Lu, S. (2009). *Extensions of Markov modulated Poisson processes and their applications to deep earthquakes*. PhD Thesis, Victoria University of Wellington.

MacDonald, I. L., Zucchini, W. (1997). *Hidden Markov and other models for discrete-valued time series*. London: Chapman & Hall.

McLachlan, G. J., Krishnan, T. (1997). *EM algorithm and extensions*. New York: Wiley.

Meier-Hellstern, K. S. (1987). A fitting algorithm for Markov-modulated Poisson processes having two arrival rates. *European Journal of Operational Research, 29*, 370–377.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association, 83*(401), 9–27.

Reyners, M. (1989). New Zealand seismicity 1964–87: An interpretation. *New Zealand Journal of Geology and Geophysics, 32*, 307–315.

Roberts, W. J. J., Ephraim, Y., Dieguez, E. (2006). On Ryden's EM algorithm for estimating MMPPs. *IEEE in Signal Processing Letters, 13*(6), 373–376.

Ryden, T. (1996a). An EM algorithm for estimation in Markov-modulated Poisson processes. *Computational Statistics and Data Analysis, 21*, 431–447.

Ryden, T. (1996b). On identifiability and order of continuous-time aggregated Markov chains, Markov-modulated Poisson processes, and phase-type distributions. *Journal of Applied Probability, 33*, 640–653.

Van Loan, C. F. (1978). Computing integrals involving the matrix exponential. *IEEE Transactions on Automatic Control, AC-23*(3), 395–404.

Vere-Jones, D., Schoenberg, F. (2004) Rescaling marked point processes. *Australian & New Zealand Journal of Statistics, 46*(1), 133–143.

Vere-Jones, D., Turnovsky, S., Eiby, G. A. (1964). A statistical survey of earthquakes in the main seismic region of New Zealand. *New Zealand Journal of Geology and Geophysics, 7*(4), 722–744.