Slip-size-dependent renewal processes and Bayesian inferences for uncertainties

Yosihiko Ogata
Institute of Statistical Mathematics, Tokyo, Japan

Received 15 June 2001; revised 8 April 2002; accepted 17 April 2002; published 1 November 2002.

[1] This paper is initially concerned with uncertainty of hazard estimates using renewal process models where the parameters are poorly constrained because of the scarcity of large earthquakes occurring along the same fault segment. A Bayesian inference is adopted, taking account of the whole likelihood function of the parameters. Also, new statistical models are considered which make use of knowledge on the slip associated with earthquake ruptures, may help to improve the hazard estimate in a positive way. Including these models, the predictive efficiency is compared by using the Akaike's Bayesian information criterion (ABIC). Three data sets are analyzed for the illustrations. The Brownian Passage Time (BPT) model is selected to fit the first data set, consisting of 10 historical great earthquakes from Nanaki trough in Japan. However, its predictive hazard function shows a large uncertainty (>100 years) of likely occurrence time around 2070. The second data set consists of the last three events of the first data set but associated with record of slip sizes, for which the ABIC selected the extended lognormal renewal process model where the time intervals between successive events are normalized by the corresponding slip sizes of the starting events of the intervals. The estimated predictive hazard function implies that the next event is likely to occur around 2040 ± 10. The last data set, consisting of 4 events with estimated occurrence times and slip sizes from a submarine fault. The ABIC selected the slip-size-dependent BPT model for this data. This model indicates that the likelihood of occurrence time of the next event is decreasing from now, and the period of its half decay is more than several hundred years. For the data sets of the last two examples, it was also shown that the slip size records are useful for better prediction of the next event. *INDEX TERMS:* 7223 Seismology: Seismic hazard assessment and prediction; 7221 Seismology: Paleoseismology; 3210 Mathematical Geophysics: Modeling; *KEYWORDS:* Brownian Passage Time process, lognormal distribution, Weibull distribution, posterior distribution, time-predictable model, renewal processes

Citation: Ogata, Y., Slip-size-dependent renewal processes and Bayesian inferences for uncertainties, *J. Geophys. Res.*, *107*(B11), 2268, doi:10.1029/2001JB000668, 2002.

1. Introduction

[2] In order to assess the probability of rupture on an active fault, the coefficients of the hazard function are substituted by the maximum likelihood estimates (MLE) that are obtained by fitting a renewal process to historical or paleoearthquake occurrence data. Since the available records of such events are usually small in number, it is recommended to reduce the number of parameters to be adjusted after the comparison of goodness of fit of the model, using the Akaike information criterion (AIC) [*Akaike*, 1974, 1998], with the restricted models where a parameter is fixed with a standard value for each renewal process such as the shape factor of lognormal distribution [*Nishenko and Buland*, 1987; *Savage*, 1991; *Working Group on Assessment Methods of Long Term Earthquake Probability*, 2001; *Ogata*, 1999].

[3] However, when the data set is too small size or has a particular configuration, the MLE has not only large errors but also tends to provide a biased hazard function, namely, a very different outcome from our statistical intuition. For example, for the data where several events have occurred in such a way that the successive time intervals are of almost equal lengths, the hazard rate with the MLE coefficients leads to an unusually high probability of occurrence of the next event in a very short time span, so as to keep the same length from the last event. The bias and large uncertainty of the MLE are due to a spiky peak and extremely asymmetric, heavy-tailed shape of the log likelihood function, which often takes place when the data set is small.

[4] To overcome such an inherent difficulty, we adopt a Bayesian inference which takes all the information regarding the entire log likelihood function into consideration in estimating the current and future hazard and its uncertainty,

0148-0227/02/2001JB000668$09.00

Table 1. Historical Earthquakes at Nankai Trough

Occurrence Time	Date, years
t_1	684
t_2	887
t_3	1099
t_4	1233
t_5	1361
t_6	1498
t_7	1605
t_8	1707
t_9	1854
t_{10}	1946

rather than looking at a representative parameter value for the likelihood function. Thus we can consider the uncertainty of the hazard rate functions based on the posterior distribution and their expectation to evaluate the probability of the occurrence of the next event. This Bayesian method includes selection of models by using the Akaike's Bayesian information criterion (ABIC), which is useful for the selection of Bayesian models, namely, not only the prior but also the likelihood function. For example, it will be worth examining the predictive utility of the slip-size records for each data set by making use of the ABIC, since some authors doubt the reliability of the slip-dependent model [e.g., *Mulargia and Gasperini*, 1995].

2. Data

[5] In this paper we will analyze three sets of occurrence data. The first data set in Table 1 consists of 10 great earthquakes of M8 class at Nankai trough, taken from *Utsu* [1999, Table 10.5], where the fourth event of early thirteenth century [*Sangawa*, 1992] is provisionally set to be 1233 in this paper.

[6] The second data set listed in Table 2 consists of the last three events in the above mentioned Nankai earthquakes where the coseismic rise at Muroto Peninsula (the water level changes at Murotsu Port) is assumed to be proportional to the slip sizes [*Shimazaki and Nakata*, 1980]. In addition to these records, we make note on the fact that the great event at Nankai trough has never occurred since 1946, using this information as appropriate.

[7] The third data set is listed in Table 3, which is from Off Toyooka Fault, one of the submarine faults in Beppu Bay, Khushu, Japan. This data set is based on the inverse image of submarine normal fault structure using sonar reflection, boring data of strata, and their calibration of ^{14}C ages estimated by radiocarbon dating [e.g., *Nakata and Shimazaki*, 1993]. Here, they identified the last four events and estimated the occurrence times and the associated slip sizes. This active fault is estimated to create an earthquake of magnitude 6 to 7 class. In addition to the above data we note that the event has never occurred since 300 B.C., making use this information as appropriate.

Table 2. Last Three Earthquakes at Nankai Trough

Occurrence Time	Date, years	Slip Size ξ, m
t_1	1707	1.8
t_2	1854	1.2
t_3	1946	1.15

Table 3. Fault Off Coast of Toyooka, Beppu Bay

Occurrence Time	Date, years B.P.[a]	Slip Size ξ, m
t_1	5900	1.0
t_2	4500	0.6
t_3	3600	0.7
t_4	2250	1.3

[a] B.P. indicates years before 1950 A.D.

3. Models and Hazards

[8] Consider a random series of events, $t_1 < t_2 < \cdots < t_i < \cdots$, and the interval lengths $X_i = t_i - t_{i-1}$ ($i = 1, 2, $) between the consecutive events. If the sequence $\{X_i\}$ is independently and identically distributed, the original series of events $\{t_i\}$ is called a renewal process. Hereinafter the density distribution of the interval is denoted by $f(x|\theta)$, and the cumulative distribution function is denoted by $F(x|\theta)$, which are characterized by parameter vector θ. In particular, if the exponential distribution

$$f(x|\lambda) = \lambda e^{-\lambda x} \tag{1}$$

$$F(x|\lambda) = 1 - e^{-\lambda x} \tag{2}$$

hold, then the series is a stationary Poisson process with intensity λ, where λ is the expected number of events in the unit time interval, or reciprocal of the mean interval length between the successive events. In addition, we later consider a slip-size-dependent renewal process using this distribution.

[9] We further consider two distributions for the renewal processes and their extensions throughout this paper. The first one is the lognormal distribution

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma x}\exp\left\{\frac{(\ln x - \mu)^2}{2\sigma^2}\right\} \tag{3}$$

$$F(x|\mu,\sigma) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right) \tag{4}$$

where "ln" denotes natural logarithm, $\Phi(x)$ denotes the standard normal distribution, and the parameters μ and σ^2 represent the mean and variance of the variable ln x. The second distribution is the Weibull distribution

$$f(x|b,\sigma) = \frac{b}{\sigma}\left(\frac{x}{\sigma}\right)^{b-1}\exp\left\{-\left(\frac{x}{\sigma}\right)^{b}\right\} \tag{5}$$

$$F(x|b,\sigma) = 1 - \exp\left\{-\left(\frac{x}{\sigma}\right)^{b}\right\}, \tag{6}$$

where b represents the power of the polynomial in its hazard function and σ is the scaling parameter.

[10] Let T be current time and consider records of occurrence times $\{t_i;\ i = 1, 2, \cdots, n\}$ of events in the

observed period $[t_1, T]$. If we can assume renewal processes for the data, then we have the following log likelihood function

$$\ln L(\theta) = \sum_{i=2}^{n} \ln f(t_i - t_{i-1}|\theta) + \ln \mathcal{F}(T - t_n|\theta) \tag{7}$$

[e.g., *Brillinger*, 1982; *Davis et al.*, 1989; *Ogata*, 1999], where

$$\mathcal{F}(x|\theta) = \text{Prob}\{\text{Time} > x\} = \int_x^\infty f(t|\theta)dt = 1 - F(x|\theta) \tag{8}$$

is called as the survivor function, which means the probability of no event in the time span of length x from the last event. This is derived by the stationarity of renewal processes for an arbitrary time at random after the last event [*Daley and Vere-Jones*, 1988, section 4.2]. The last term in the log likelihood in equation (7) stands for the fact that no events occurred from the last event up until the current time T, and the contribution of this term cannot be neglected especially when the number of events n is small or the length of $T - t_n$ is larger than the mean recurrence time $(T - t_1)/n$ [cf. *Davis et al.*, 1989; *Ogata*, 1999].

[11] Suppose that estimates $\hat{\theta}$ of the parameters are obtained somehow. For example, the maximum likelihood estimates (MLE) are the values of parameters that maximize the log likelihood function (7). Also, there are a few other good estimates, as will be discussed in section 4. The hazard function of the occurrence of the next event, at the time when x years has elapsed since the last event, is derived by the calculation of conditional probability such that

$$\begin{aligned}\lambda(x|\theta) &= \lim_{\Delta\to 0} \frac{\text{Prob}\{x < \text{Time} < x + \Delta | \text{Time} > x\}}{\Delta} \\ &= \lim_{\Delta\to 0} \frac{\text{Prob}\{x < \text{Time} < x + \Delta\}}{\Delta\ \text{Prob}\{\text{Time} > x\}} \\ &= \frac{f(x|\theta)}{\mathcal{F}(x|\theta)} = -\frac{d}{dx}\ln\mathcal{F}(x|\theta).\end{aligned} \tag{9}$$

Therefore, assuming no event has occurred up to the current time t_c since the last event, the probability that at least one event takes place in the future interval $(t_c,\ t_c + y)$ is calculated by

$$\mathcal{P}(t_c; y|\theta) = 1 - \exp\left\{-\int_{t_c}^{t_c+y} \lambda(t|\theta)dt\right\} = 1 - \frac{\mathcal{F}(t_c + y|\theta)}{\mathcal{F}(t_c|\theta)}. \tag{10}$$

By taking the derivative of the function in equation (10) with respect to y, we have the density function of the occurrence time of the next event

$$\begin{aligned}p(t_c; y|\theta) &= \lambda(t_c + y|\theta)\exp\left\{-\int_{t_c}^{t_c+y} \lambda(x|\theta)dx\right\} \\ &= \frac{\mathcal{F}(t_c + y|\theta)}{\mathcal{F}(t_c|\theta)} f(t_c + y|\theta).\end{aligned} \tag{11}$$

[12] Now, extending the renewal processes, we consider models to investigate the statistical uncertainty of the time-predictable model [*Shimazaki and Nakata*, 1980] for the last two data sets listed in section 2. Given a data set of occurrence times of events $\{t_i: i = 1, 2, \cdots, n\}$ associated with slip sizes $\{\xi_i: i = 1, 2, \cdots, n\}$, consider the ratio of interval length between consecutive earthquakes to the slip size of the starting event of the interval. Assume that the ratios are independently and identically distributed according to a density function $f(ratio|\theta)$. Then we can state the log likelihood as follows:

$$\begin{aligned}\ln L(\theta) &= \ln L(\theta|t_1, \cdots, t_n, T; \xi_1, \cdots, \xi_n) \\ &= \sum_{i=2}^{n} \ln f\left(\frac{t_i - t_{i-1}}{\xi_{i-1}}|\theta\right) + \ln \mathcal{F}\left(\frac{T - t_n}{\xi_n}|\theta\right)\end{aligned} \tag{12}$$

For such a slip-size-dependent renewal process, the hazard rate at an elapsed time y from the last event is written using the slip size ξ_n of the last event as follows. That is to say,

$$\begin{aligned}\lambda(x|\xi_n, \theta) &= \lim_{\Delta\to 0} \frac{\text{Prob}\{x < \text{Time} < x + \Delta | \text{Time} > x\}}{\Delta} \\ &= \lim_{\Delta/\xi_n\to 0} \frac{\text{Prob}\{x/\xi_n < \text{Time}/\xi_n < x/\xi_n + \Delta/\xi_n\}}{(\Delta/\xi_n)\text{Prob}\{\text{Time}/\xi_n > x/\xi_n\}\,\xi_n} \\ &= \frac{f(x/\xi_n|\theta)}{\mathcal{F}(x/\xi_n|\theta)}\,\xi_n^{-1} = -\frac{d}{dx}\ln\mathcal{F}(x/\xi_n|\theta).\end{aligned} \tag{13}$$

Using this equation, we can calculate the cumulative probability and its density function replacing $\lambda(x|\theta)$ in the first equalities of equations (10) and (11), respectively, by $\lambda(x|\xi_n,\theta)$.

[13] The distributions for this type of model include the exponential, lognormal, and Weibull, as stated above. Note here that the process using the exponential distribution for the ratio is not a Poisson process.

[14] Another type of slip-size-dependent model for the stochastic version of the time-predictable model is available, that is, the original form of the Brownian Passage Time (BPT) distribution [*Schrödinger*, 1915; *Seshadri*, 1998, etc.]. This reduces to the standardized version of the BPT distributions [*Matthews*, 1998; *Ellsworth et al.*, 1998] if we assume the same slip size for each event. The Brownian passage process is compatible with the physical mechanism in which the rupture takes place when tectonic loading with stress accumulation, adding Brownian perturbations, reaches the Coulomb threshold and then the stress drops to repeat the stress accumulation. The Brownian perturbations are assumed for many effects including external stress transfer effects from earthquakes outside the target source, aseismic load variations, pore pressure, dilatancy, healing, and general evolution of spatial heterogeneity.

[15] The density and cumulative distribution function of this model are given by

$$f(x|\xi, v, \sigma) = \frac{\xi}{\sqrt{2\pi\sigma^2 x^3}} \exp\left\{-\frac{(\xi - vx)^2}{2\sigma^2 x}\right\} \tag{14}$$

$$F(x|\xi, v, \sigma) = \Phi\left(\frac{vx - \xi}{\sigma\sqrt{x}}\right) + \exp\left\{\frac{2\xi v}{\sigma^2}\right\}\Phi\left(-\frac{vt + \xi}{\sigma\sqrt{x}}\right), \tag{15}$$

respectively, where $\Phi(\ )$ denotes the cumulative function of the standard normal distribution, v is the accumulation rate

Table 4. Parameter Normalization

Models	Normalized Parameters	Relation to the Original Parameters
Lognormal	μ_1	$\mu - \ln \nu$
	σ_1	σ
Weibull	b_1	b
	σ_1	σ/ν
BPT	ν_1	$\nu\nu$
	σ_1	σ
Exponential	λ_1	$\lambda\nu$

of the stress, σ^2 is the variance of the Brownian fluctuation, and ξ is the variable that is proportional to the stress drop or slip size of the event. Then the log likelihood of this model is given by

$$\ln L(\nu, \sigma) = \sum_{i=2}^{n} \ln f_\theta(t_i - t_{i-1}|\xi_{i-1}, \nu, \sigma) + \ln \mathcal{F}_\theta(T - t_n|\xi_n, \nu, \sigma), \tag{16}$$

and the hazard rate at elapsed time y from the last event is given by

$$\lambda(x|\xi, \nu, \sigma) = \frac{f(x|\xi, \nu, \sigma)}{\mathcal{F}(x|\xi, \nu, \sigma)} = -\frac{d}{dx} \ln \mathcal{F}(x|\xi, \nu, \sigma). \tag{17}$$

Replacing $\lambda(x|\theta)$ in the first equation of equations (10) and (11) by this hazard function, we can derive the cumulative probability and its density functions, respectively.

[16] Here it should be noted that application of the ordinary renewal process models to the events without slip size information, such as the data in Table 1, is equivalent to applying the slip-size-dependent renewal processes assuming same slip size, say $\xi_i = 1.0$.

4. Recurrence Time Normalization

[17] For accurate calculation of log likelihoods at any possible parameter values and their integrals, we normalize the data of occurrence times relative to the mean recurrence time ν of the stationary Poisson process model as discussed by *Nishenko and Buland* [1987]. That is to say, set $\nu = (T - t_1)/n$, and then transform the original occurrence times $\{t_i\}$ in the observation period $[t_1, T]$ to the normalized occurrence times $\{\tau_i\}$ in the normalized time span $[0, n]$ by $\tau_i = (t_i - t_1)/\nu$. Thus, in the normalized times, the mean recurrence time becomes 1.0. Then the corresponding parameters of the normalized renewal processes become as listed in Table 4, in terms of the original parameter values. Thus, we can see that the log likelihood values $\ln L_1$ of the normalized occurrence data with respect to the normalized parameters are respectively related to the original log likelihood values by

$$\ln L(\theta) = \ln L_1(\theta_1) - n \ln \nu.$$

This relation also holds for the integrated likelihood (likelihood of a Bayesian model) that will be introduced in section 5. The normalized parameters are also useful in comparing the shapes of the distributions for data sets of different timescales. From here on we will be concerned with the normalized data and the corresponding parameters unless otherwise mentioned.

5. A Bayesian Inference

[18] The maximum likelihood estimator (MLE) works typically well in the case where the likelihood function is symmetric with respect the parameter. However, if it is asymmetric and heavy tailed, the MLE provides a biased estimate [e.g., *Box and Tiao*, 1973]. This can take place when the sample size is small, and the maximum likelihood procedure is not suitable for such a case.

[19] In this section we denote the likelihood function by $L(\theta|\mathbf{X})$ for the data set that is also denoted by the vector $\mathbf{X}$. For the Bayesian inference [e.g., *O'Hagan*, 1994], we assume a proper prior distribution (i.e., a probability density) $\pi(\theta)$ of the parameters θ. Then the posterior density distribution is given by

$$f(\theta|\mathbf{X}) = \frac{L(\theta|\mathbf{X})\pi(\theta)}{\int_\Theta L(\theta|\mathbf{X})\pi(\theta)d\theta}, \tag{18}$$

which indicates the certainty of the model at each parameter. Here Θ is a parameter space that will be described explicitly in section 6.

[20] On the basis of the posterior distribution, we can define two estimates. The first one is the maximum posterior (MAP) estimate

$$\hat{\theta}_M = \arg\left\{ \max_\theta f(\theta|\mathbf{X}) \right\} = \arg\left\{ \max_\theta L(\theta|\mathbf{X})\pi(\theta) \right\}, \tag{19}$$

which is expected to be approximately equal to the MLE that is obtained when $\pi(\theta) = 1$ in equation (19). The second one is the Bayes estimate, or the posterior mean,

$$\hat{\theta}_B = \int_\Theta \theta f(\theta|\mathbf{X})d\theta = \frac{\int_\Theta \theta L(\theta|\mathbf{X})\pi(\theta)d\theta}{\int_\Theta L(\theta|\mathbf{X})\pi(\theta)d\theta}. \tag{20}$$

[21] Substituting these estimates to the hazard functions in equation (9), (13), or (17) in such a way that

$$\hat{\lambda}_M(x) = \lambda\left(x|\hat{\theta}_M\right) \tag{21}$$

and

$$\hat{\lambda}_B(x) = \lambda\left(x|\hat{\theta}_B\right), \tag{22}$$

we can assess the probabilities of the next event by calculating those in equations (10) and (11). However, when the sample size is too small, the uncertainty for each estimate can be too large to provide a practical probability assessment. The alternative is obtained by taking the average of all possible hazard functions with respect to the posterior distribution in such a way that

$$\hat{\lambda}_P(x) = \int_\Theta \lambda(x|\theta) f(\theta|\mathbf{X})d\theta \tag{23}$$

[e.g., *Rhoades et al.*, 1994; *Ogata*, 1999]. We call this the Bayesian predictive hazard function. The superiority of the prediction performance using this hazard function over

Table 5. Historical Earthquakes at Nankai Trough[a]

Model	Hyperparameters		ABIC	Normalized Timescale				Real Timescale			
	First	Second		MAP1	MAP2	BE 1	BE 2	MAP1	MAP2	BE 1	BE 2
Poisson	1.0	–	121.9	0.898	–	0.871	–	6.82×10^{-3}	–	6.61×10^{-3}	–
Lognormal	0.25	–	110.7	0.030	0.270	0.259	0.296	4.91	0.270	4.91	0.296
Weibull	4.0	1.0	113.7	3.74	1.17	3.38	1.17	3.74	154.	3.38	154.
BPT	1.0	0.25	109.9[b]	0.942	0.266	0.923	0.292	7.15×10^{-3}	0.266	7.01×10^{-3}	0.292

[a] Mean recurrence time is 131.7 years. Hyperparameters show the optimal parameter values of the prior distribution, and the minimum ABIC value for each model is given. The MAP1 and MAP2 are the first and second components of the parameter estimates with maximum a posterior distribution, respectively. The BE 1 and BE 2 are the first and second components of the posterior mean estimates, or the Bayes estimate, respectively (see text for further explanation).

[b] Overall best fitted model (i.e., with the minimum ABIC value) for the data.

those based on a single estimate such as in (21) and (22) can be shown by a similar inequality to that discussed by *Akaike* [1978, 1998].

[22] Now, in order to define our prior distributions for data set of events with the normalized recurrence times, we consider transformations of parameters θ from the original infinite parameter space Θ into the unit square $\Theta_0 = (0, 1) \times (0, 1)$. For the lognormal case, $\theta = (\mu, \sigma) \in \Theta = (-\infty, \infty) \times (0, \infty)$ is transformed to $\theta_0 = (\mu_0, \sigma_0)$ by the relation

$$\mu = \ln \frac{\mu_0}{1 - \mu_0} \tag{24}$$

$$\sigma = -\frac{\ln (1 - \sigma_0)}{\alpha} \tag{25}$$

for some $\alpha > 0$. These transformation lead to the uniform prior density distribution $\pi_0(\theta_0) = 1$ on the parameter space $\Theta_0 = (0, 1) \times (0, 1)$, if we assume the prior

$$\pi(\theta) = \pi(\mu)\, \pi(\sigma) = \frac{d\mu_0}{d\mu} \frac{d\sigma_0}{d\sigma} = \frac{e^{-\mu}}{(1 + e^{-\mu})^2}\, \alpha e^{-\alpha\sigma} \tag{26}$$

in the original parameter space. Furthermore, taking account of the Jacobian matrix that preserves the integral of the posterior function on the corresponding regions in Θ and Θ_0, we obtain the posterior density distribution on the parameter space Θ_0 such that

$$f_0(\mu_0, \sigma_0|\alpha) \propto L\left(\ln \frac{\mu_0}{1 - \mu_0}, -\frac{\ln (1 - \sigma_0)}{\alpha}\right). \tag{27}$$

Namely, the posterior density distribution is given by normalizing the function on the right-hand side of equation (27) by its integral over the region Θ_0 with respect to the parameters μ_0 and σ_0. Later, we will make use of the function in (27) for sampling parameters θ_i, $i = 1, 2, \cdots, n$ from the posterior function.

[23] Similarly, since all the parameters (a, b) for the Weibull distribution and (v, σ) for the BPT model are positive valued, we take the same transformation as in equation (25) so that the prior distributions are

$$\pi(a)\, \pi(b) = \frac{da_0}{da} \frac{db_0}{db} = \alpha e^{-\alpha a}\, \beta e^{-\beta b} \tag{28}$$

and

$$\pi(v)\, \pi(\sigma) = \frac{dv_0}{dv} \frac{d\sigma_0}{d\sigma} = \alpha e^{-\alpha v}\, \beta e^{-\beta\sigma}, \tag{29}$$

respectively, in the original parameter space Θ. Then we have the respective posterior density distributions

$$f_0(a_0, b_0|\alpha, \beta) \propto L\left(-\frac{\ln (1 - a_0)}{\alpha}, -\frac{\ln (1 - b_0)}{\beta}\right) \tag{30}$$

and

$$f_0(v_0, \sigma_0|\alpha, \beta) \propto L\left(-\frac{\ln (1 - v_0)}{\alpha}, -\frac{\ln (1 - \sigma_0)}{\beta}\right). \tag{31}$$

By a similar manner, we can define the prior and posterior distributions of both the stationary Poisson process and the slip-size-dependent exponentially distributed renewal processes.

[24] The optimal values of the hyperparameters α and β in the prior distribution are selected by maximizing the normalizing factor of the posterior distribution in equation (18). For example, in the case of the BPT model, choose α and β such that they maximize

$$\begin{aligned} \mathcal{L}(\alpha, \beta) &= \int_{\Theta} L(v, \sigma|\mathbf{X}) \pi(v, \sigma|\alpha, \beta) dv d\sigma \\ &= \int_{\Theta_0} L\left(-\frac{\ln(1 - v_0)}{\alpha}, -\frac{\ln(1 - \sigma_0)}{\beta}\right) dv_0 d\sigma_0. \end{aligned} \tag{32}$$

This maximizing procedure is called the type II maximum likelihood method of *Good* [1965], where $\mathcal{L}(\vartheta) = \mathcal{L}(\alpha, \beta)$ in equation (32) is called the integrated likelihood or the likelihood of Bayesian model. In a manner similar to the BPT model, we can calculate the integrated likelihood for the exponential, lognormal, and Weibull models. Furthermore, the Akaike's Bayesian Information Criterion [*Akaike*, 1980, 1998],

$$\begin{aligned} \text{ABIC} &= (-2) \max (\ln \text{ integrated likelihood}) \\ &\quad + 2\, (\text{number of adjusted hyperparameters}) \\ &= (-2) \max_{\vartheta} \ln \mathcal{L}(\vartheta) + 2 \dim\{\vartheta\}, \end{aligned} \tag{33}$$

is useful in order to compare the predictive efficiency of the stated Bayesian models of different priors or likelihood functions based on the data, where $\dim\{\vartheta\} = 1$ for the exponential and lognormal models and $\dim\{\vartheta\} = 2$ for Weibull and BPT models. A model with a smaller ABIC value is considered to provide a better prediction. The physical fundamentals of the present procedure from the

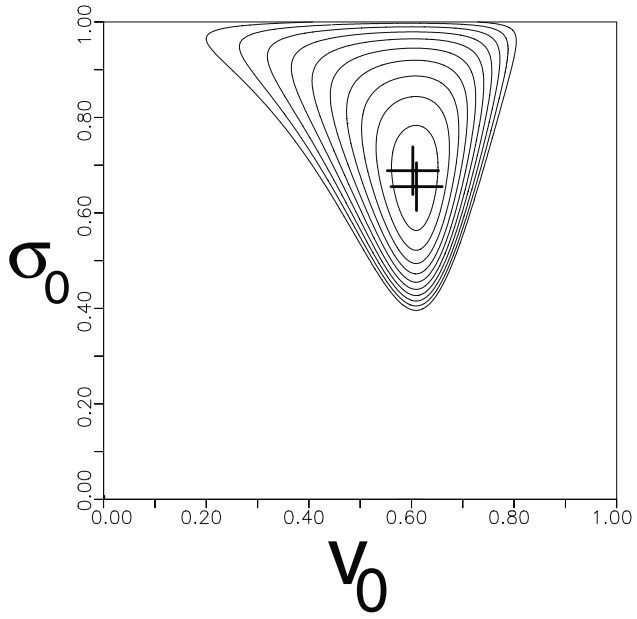

Figure 1. Heights of the posterior density distribution in equation (31) of v_0 (the horizontal axis) and σ_0 (the vertical axis) for the data in Table 1 are contoured at the unit intervals in logarithmic scale for the range from -10 to -1, with no contours being shown below the value -10. The lower and upper pluses indicate the MAP and posterior mean estimate (Bayes estimate), respectively.

prediction viewpoint are effectively presented by *Akaike* [1985, 1998].

[25] We have described only the two types of priors, logistic and exponential distributions of the shift and scaling parameters, respectively, for technical reasons such as the compatibility with the considered transformation of parameters. One can also consider and compare any other proper prior distributions, including those based on physical mechanisms or empirical laws, by means of the ABIC.

6. Results and Uncertainties

6.1. Historical Earthquakes at Nankai Trough

[26] The optimal hyperparameter values, $\hat{\alpha}$ and $\hat{\beta}$, were searched among restricted values on the grids $\{2^i; i = \cdots, -1, 0, 1, \cdots\}$ to maximize the integrated likelihood as in equation (32). Table 5 lists the estimates and ABIC values of the four renewal process models (exponential, lognormal, Weibull, and BPT) that have been applied to the data in Table 1. The ABIC values indicate the BPT model is the best fit.

[27] Figure 1 shows a contour map of the posterior function (31) of the BPT model with respect to the transformed parameters. From Table 5 and Figure 1 we see that the MAP estimate and Bayes estimate (posterior mean estimate) are close for v but differ slightly for σ. Similar features can be seen from the contour map for lognormal model of *Ogata* [2001]. This is due to the fact that the posterior function is symmetric with respect to v while asymmetric and rather heavy tailed with respect to the scaling parameter σ.

[28] Now, in order to see the variability of the estimated hazard functions, we simulate 100 samples of parameters $\theta_0 = (v_0, \sigma_0)$ in Θ_0 from the posterior distribution $f(v_0, \sigma_0)$ in (31), repeating the following rejection algorithm:

1. Simulate a pair of uniform random numbers (v_0, σ_0) in the unit square Θ_0.

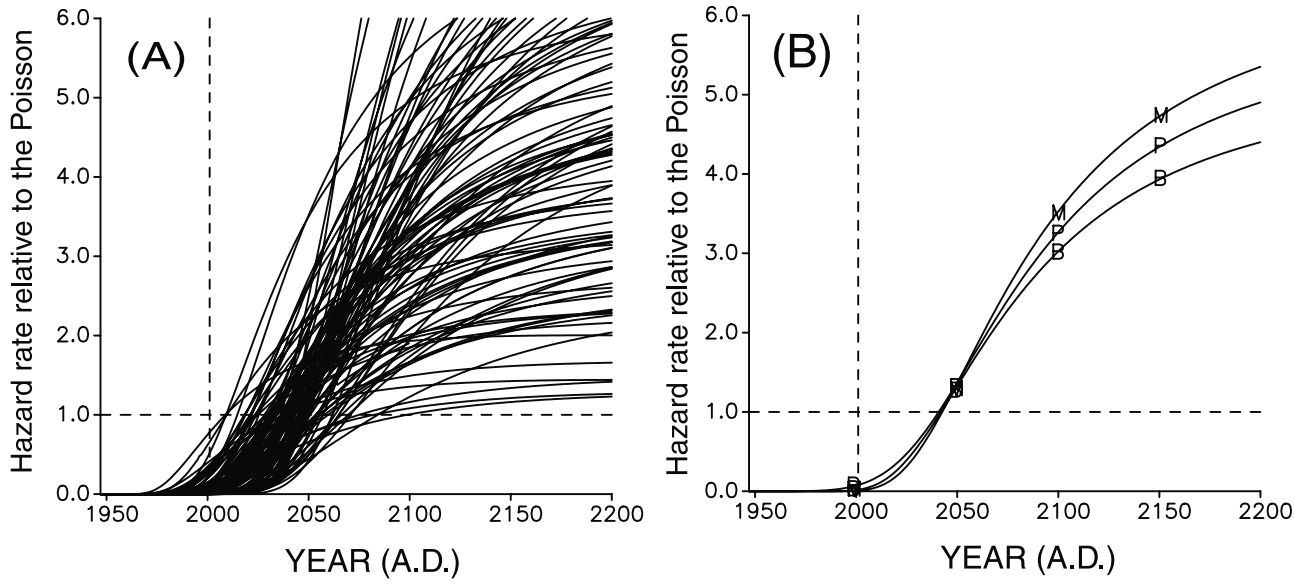

Figure 2. (a) One hundred samples of hazard rate functions of time, representing the variability of the estimated hazard rate function. Their parameter values are simulated from the posterior probability distribution in equation (31) of the BPT model applied to the data in Table 1. The horizontal dashed line indicates the level of the mean hazard rate (0.076 event/yr) throughout the whole time span, with the vertical dotted line indicating the current time (March 2002). (b) The curves with M, B, and P show the hazard function equation (21) with the MAP estimate in equation (19), the hazard function (22) with the Bayes estimate (posterior mean estimate) in equation (20), and the Bayesian predictive hazard function in equation (23), respectively.

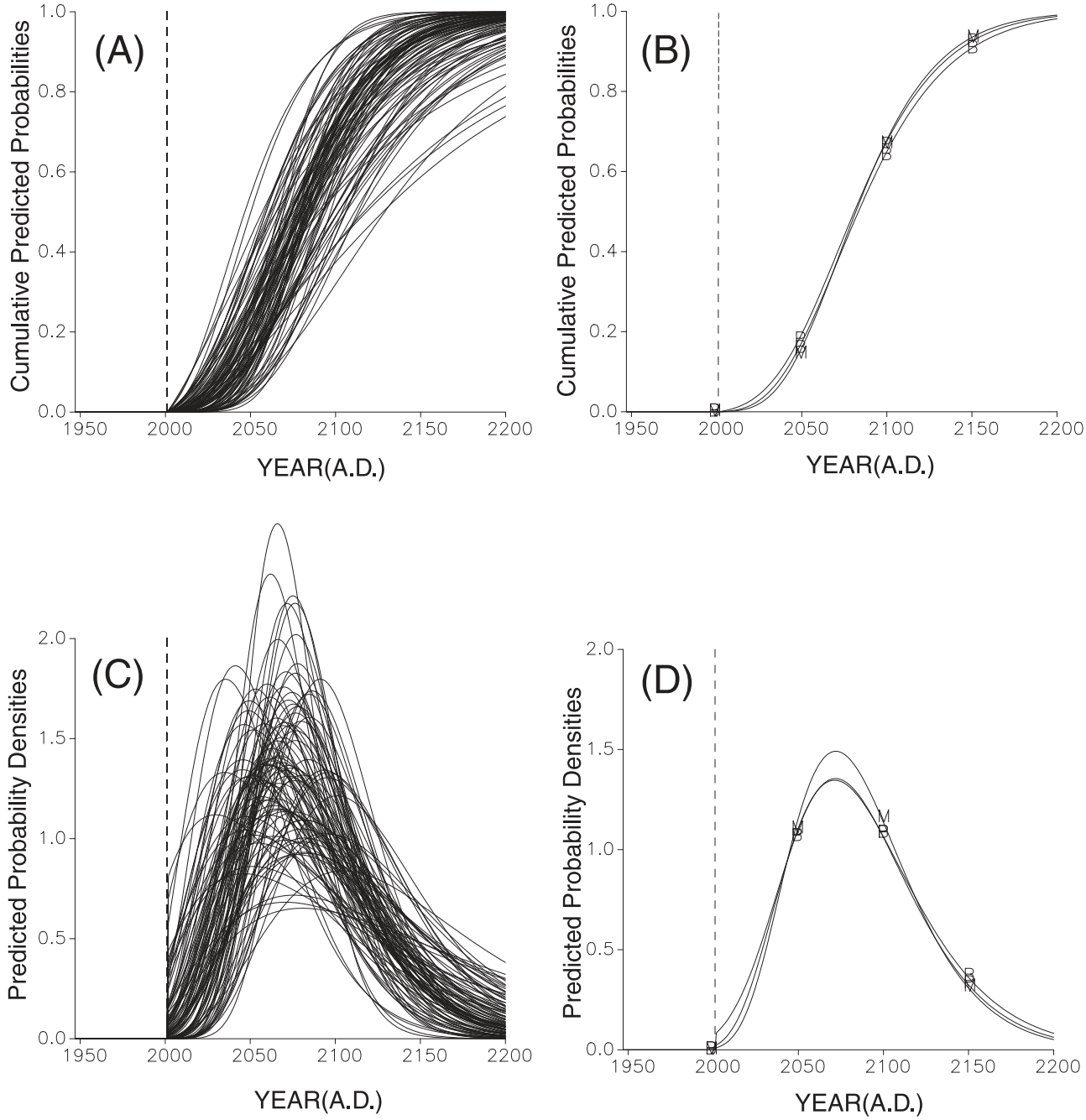

Figure 3. (a) One hundred samples of cumulative probability functions of time from present time (the vertical dashed line), corresponding to each sample path of the hazard rate functions in Figure 2a. These are calculated by the equation in equation (10) and represent the variability of the probability estimate. (b) The curves with M, B, and P show probability functions based on the MLE, Bayes estimate (posterior mean estimate), and the Bayesian predictive hazard function, respectively. (c) and (d) Predicted probability density functions in equation (11) corresponding to Figures 3a and 3b, respectively.

2. For a uniform random number $U \in [0, 1]$, accept (ν_0, σ_0) as a sample if $U \leq f_0(\nu_0, \sigma_0)/\max_{(x,y)\in}\Theta_0 f_0(x, y)$ hold, otherwise reject it.

[29] Note here that the function f_0 need not be the normalized one. That is, the function on the right-hand side of equation (31) can be used since the ratio in the step 2 of the rejection algorithm cancels the normalizing constant. Thus Figure 2a shows the hundred samples of hazard function $\lambda(x|\theta_i)$, which shows the variability of the estimated hazard function inferred from the data in Table 1, assuming the BPT renewal process.

[30] On the other hand, Figure 2b shows the hazard functions with the MAP and posterior mean estimates and also the Bayesian predictive hazard function of the BPT model. Here, we have made two-dimensional numerical integrations for a 400 $\times$ 400 lattice on Θ_0. We also note

Table 6. Last Three Earthquakes at Nankai Trough[a]

Model	Hyperparameters		ABIC	Normalized Timescale				Real Timescale			
	First	Second		MAP1	MAP2	BE 1	BE 2	MAP1	MAP2	BE 1	BE 2
Slip-Size Data Used											
Exponential	1.0	–	34.9	0.955	–	0.838	–	9.77×10^{-3}	–	8.57×10^{-3}	–
Lognormal	0.0625	–	30.0*	−0.211	0.0322	−0.210	0.0577	4.37	0.0322	4.37	0.0577
Weibull	32.0	1.0	30.9	36.0	0.821	22.2	0.831	36.0	80.2	22.2	81.2
BPT	1.0	0.125	30.8	1.229	0.0433	1.213	0.0936	0.0126	0.0433	0.0124	0.0936
Slip-Size Data Not Used											
Poisson	0.5	–	36.4	0.669	–	0.505	–	6.84×10^{-3}	–	5.16×10^{-3}	–
Lognormal	0.5	–	36.8	0.170	0.235	0.199	0.408	4.75	0.235	4.78	0.408
Weibull	4.0	1.0	37.7	4.95	1.31	3.29	1.28	4.95	128.	3.29	125.
BPT	1.0	0.5	34.9	0.815	0.213	0.700	0.388	8.34×10^{-3}	0.213	7.16×10^{-3}	0.388

[a] Mean recurrence time is 97.73 years. See Table 5 for the table notes.

that the posterior mean estimate is expected to be approximately equal to the following arithmetic mean:

$$\hat{\theta}_B \approx \frac{1}{N}\sum_{i=1}^{N}\theta_i$$

of the sampled parameters θ_i from the posterior distribution, while the Bayesian predictive hazard function is expected to be approximately equal to the arithmetic mean,

$$\hat{\lambda}_P(x) \approx \frac{1}{N}\sum_{i=1}^{N}\lambda(x|\theta_i),$$

of the functions shown in Figure 2a. We see that the three hazard functions in Figure 2b differ only slightly from one another.

[31] Figure 3a shows 100 samples of cumulative probability function, where each of the samples is derived from the corresponding sample hazard function in Figure 2a by the relation in equation (10). From these samples of cumulative probability, the error distribution of the predicted probability until time t, say, is obtained by making the histogram of points at which the cumulative curves crossed the vertical line at the time t [cf. *Savage*, 1991, 1992]. Figure 3b shows the three cumulative probability functions corresponding to the three hazard rate in Figure 2b. We see little difference between the estimates. Figures 3c and 3d show the estimated probability density of the next event corresponding to Figures 3a and 3b, respectively. It can be seen that the next Nankai event will most likely to occur around the years of 2070–2080, with a likely occurrence range of more than 100 years, if we applied the best fitted BPT renewal process using the data in Table 1. These results do not change much for the lognormal [*Ogata*, 2001] and Weibull models with the optimal selection of the hyperparameters.

6.2. Last Three Earthquakes at Nankai Trough

[32] As described in section 2, the data in Table 2 are only for the last three Nankai events, for which records of slip sizes are available. Thus the sample size is indeed very small even if we have the implicit information that no event has occurred since 1946. Nevertheless, the posterior distributions itself is well defined for each model.

[33] Again, the optimal hyperparameter values were selected among the restricted values on the grids $\{2^i; i = \cdots, -1, 0, 1, \cdots\}$ in order to maximize the integrated likelihood. Table 6 lists the maximizing hyperparameter estimates, the MAP and posterior mean estimates, and ABIC values of the four models (exponential, lognormal, Weibull, and BPT) which have been applied to the data in Table 2. The same models are also applied, but assuming the same slip size (say, 1.0) for all the three events, which represents the case where slip information is not useful. The results show that the renewal models using slip sizes are better fitted than those models without slip-size information. Overall, the slip-size-dependent (SSD) lognormal model shows the best fit.

[34] From Table 6 we see that the MAP and posterior mean estimates for SSD lognormal and SSD-BPT models are similar for μ and ν but vary greatly for σ, respectively. This is due to the fact that both the posterior distributions are almost symmetric with respect to μ and ν, while they are asymmetric and heavy tailed with respect to the scaling parameter σ. Furthermore, the MAP estimates of σ of SSD lognormal and SSD-BPT models are very small due to the fact that the ratios $\{(t_i - t_{i-1})/\xi_{i-1}; i = 2, 3\}$ are almost equal to one another. The MLE is almost the same as the MAP estimate, both being inadequate as the estimate for σ. Rather, the posterior mean estimate seems to be closer to our intuition. Figure 4 shows a contour map of the posterior function of the SSD lognormal model with respect to the transformed parameters.

[35] By the rejection algorithm described in section 6.1, we simulate 100 samples of parameters $\theta_0 = (\mu_0, \sigma_0)$ in Θ_0 from the posterior distribution $f(\mu_0, \sigma_0)$ of the SSD lognormal model. Figure 5a shows the 100 hazard functions $\lambda(x|\theta_0)$, which show the variability of the estimated hazard function inferred from the data in Table 2, assuming the SSD lognormal model.

[36] In comparison with Figure 2a for the data in Table 1 this shows many sample functions rising steeply around 2030–2040, due to the fact that the ratios of the consecutive interval lengths to the corresponding slip sizes are almost equal to each other, which typically supports the hypothesis of the time-predictable model. On the other hand, it is also indicated that the remainder of samples of hazard functions can also simulate such approximately equal ratios with some probability.

[37] As a result, the hazard functions with the MAP and posterior mean estimates in addition to the Bayesian predictive hazard function are shown in Figure 5b. Among

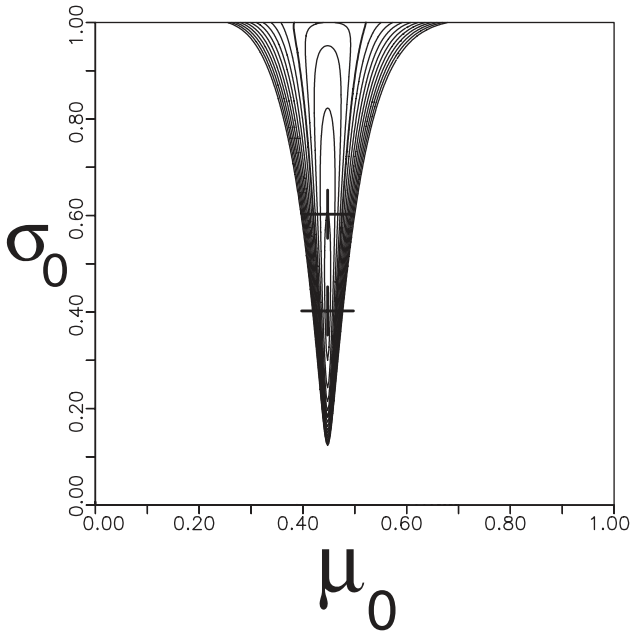

Figure 4. Heights of the posterior density distribution in equation (30) for the data in Table 2 are contoured at the unit intervals in logarithmic scale for the range from −10 to 4, with no contours being shown below the value −10. The lower and upper pluses indicate the MAP and Bayes estimate (posterior mean estimate), respectively.

these we choose the Bayesian predictive hazard function for the same reason as described in section 4.

[38] Figure 6a shows 100 samples of cumulative probability function that is calculated by the first equality in equation (10), where each of the samples is derived from the corresponding sample hazard function in Figure 5a. Figure 6b shows the cumulative probability functions corresponding to the hazard rate in Figure 5b. Figures 6c and 6d show the estimated probability density of the next event corresponding to Figures 6a and 6b, respectively. It is seen that the next Nankai event will most likely to occur around the years of 2040 based on the data in Table 2 associated with the slip-size information. The other models (SSD Weibull and SSD-BPT models), with the optimal hyperparameter values in the prior distributions, also provide the similar shapes to those in Figures 5b, 6b, and 6d, and the probability of the occurrence by 2040 is more than 60%.

6.3. Events of the Fault in Beppu Bay

[39] This data set consisting of four events with slip sizes appears to fit to the time-predictable model very well during the period until the last event [*Nakata and Shimazaki*, 1993]. However, the next event should have occurred at about 300 $\sim$ 400 years ago if the series of events exactly obeyed the time-predictable model.

[40] The optimal hyperparameter values were searched among the restricted values on the grids $\{2^i; i = \cdots, -1, 0, 1, \cdots\}$ in order to maximize the integrated likelihood. Table 7 lists the estimates and ABIC values of the considered models. It can be seen that the renewal models using slip sizes, with the exception of the exponential distribution, are better fitted than those without using such information. Overall, the SSD-BPT model is best fitted for this data. However, the difference between the ABICs (i.e., 1.8) of the SSD-BPT model and the ordinary BPT model without slip-size effect is not so large compared to those between the lognormal models in the previous example (i.e., 6.8 from Table 6). This may be due to the fact that there are larger uncertainties in the estimation of occurrence times and smaller slip sizes than the previous example.

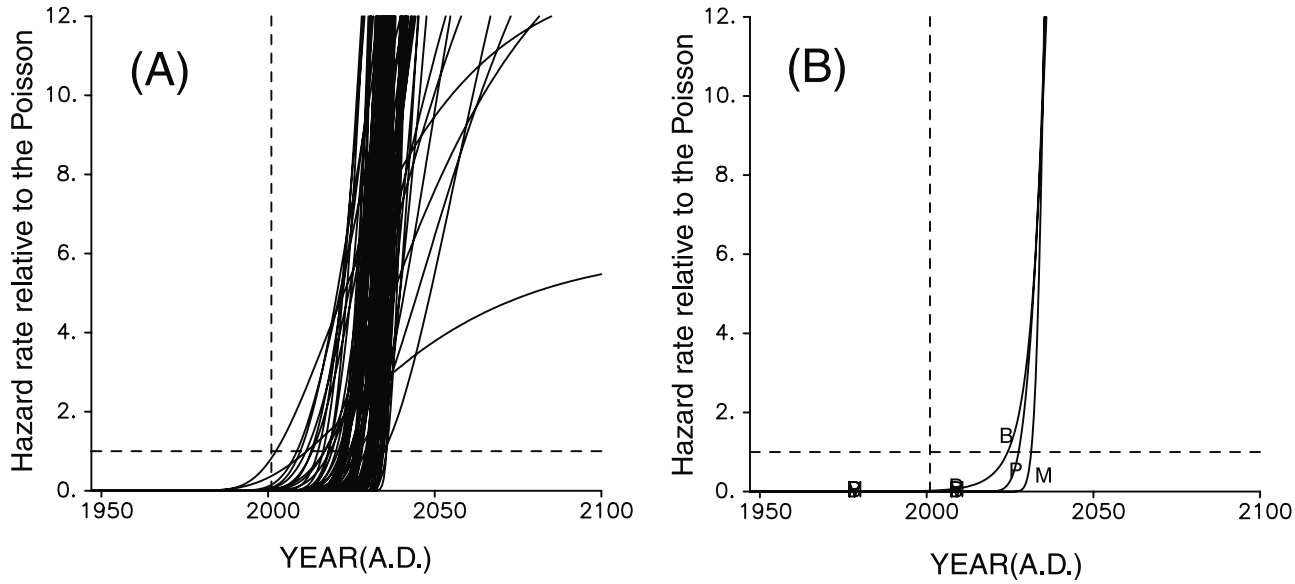

Figure 5. (a) One hundred samples of hazard rate functions of time that start from the occurrence time of the last event, simulated from the posterior probability distribution in equation (30) for the data in Table 2, consisting of only three events but associated with their slip-size records. The horizontal dashed line indicates the mean hazard rate (0.0102 event/yr) with the vertical dashed lines indicating the current time. (b) The estimated hazard rate functions are similarly defined to those in Figures 2a and 2b.

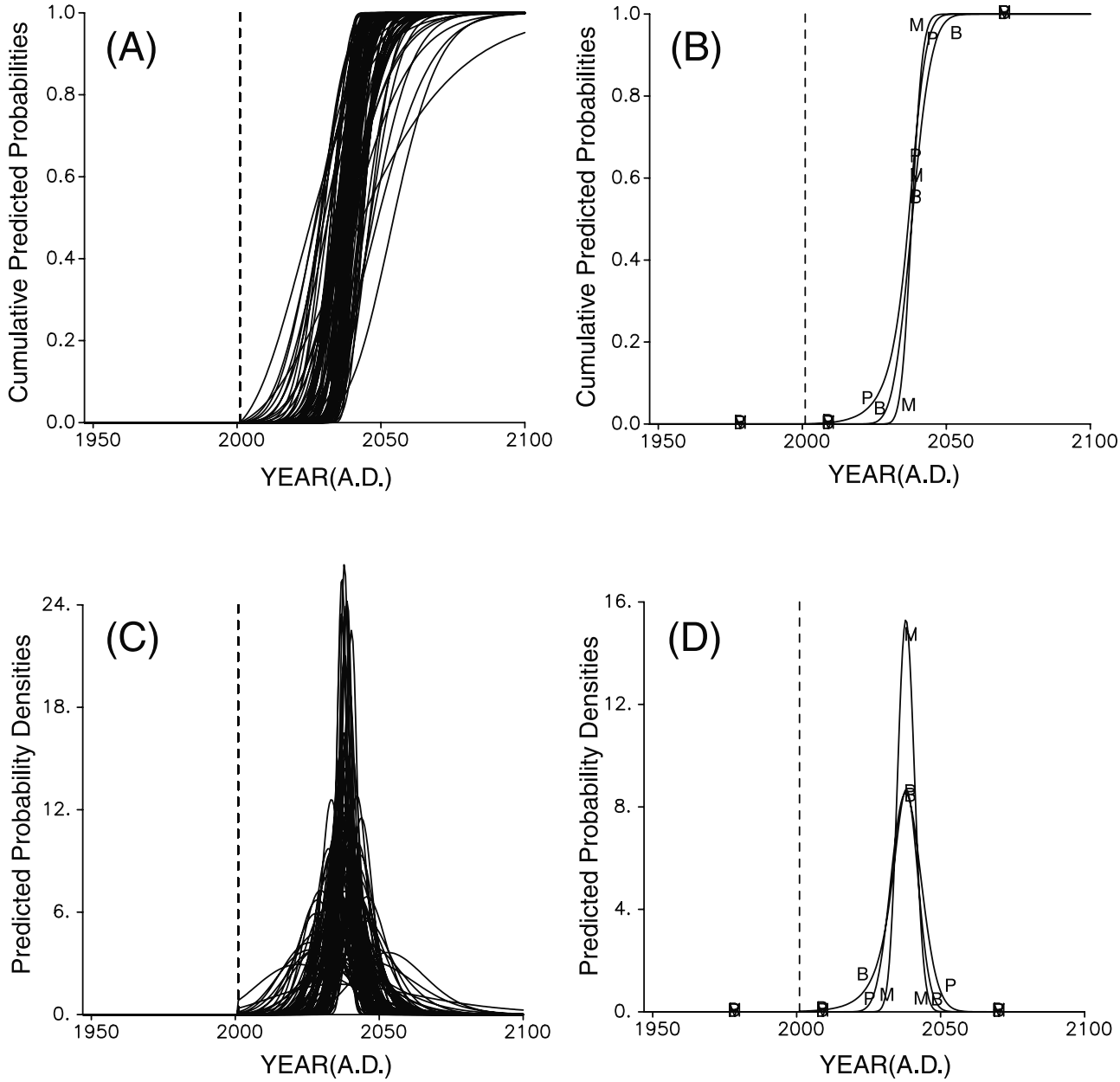

Figure 6. (a) One hundred samples of cumulative probability functions of time, starting from the current time, derived from the hazard rate functions in Figure 5a, using the first part of the equation in equation (10). (b) The similar cumulative probability functions based on the MAP, posterior mean estimate, and the Bayesian predictive hazard function in equation (23). (c) and (d) Predicted probability density functions in equation (11) corresponding to each sample path of the function in Figures 5a and 5b, respectively.

[41] Figure 7 shows a contour map of the posterior function of the SSD-BPT model with respect to the transformed parameters (ν_0, σ_0). In a similar manner to the previous sections, 100 samples of parameters $\theta_0 = (\nu_0, \sigma_0)$ in Θ_0 are simulated from the posterior distribution of the SSD-BPT model. Figure 8a shows the hundred hazard functions $\lambda(x|\theta_0)$, which show the variability of the estimated hazard function inferred from the data in Table 3, assuming the SSD-BPT model.

[42] The steep rise of many simulated hazard functions in Figure 8a is due to the equality of the ratio of the consecutive intervals to the corresponding slip sizes, which supports the hypothesis of the time-predictable model. However, the variability is not only larger than that of the previous models applied to the data in Table 2, but also the differences between the three estimates in Figure 8b are larger than those in Figure 5b. This is due to the fact that the time interval from the last event up until the current time substantially exceeds the one expected by the time-predictable model of *Shimazaki and Nakata* [1980]. This outcome is led by the log likelihood functions that include the log-survivor function as in equations (7), (12), and (16).

Table 7. Events in Fault Off Coast of Toyooka, Beppu Bay[a]

Model	Hyperparameters		ABIC	Normalized Timescale				Real Timescale			
	First	Second		MAP1	MAP2	BE 1	BE 2	MAP1	MAP2	BE 1	BE 2
Slip-Size Data Used											
Exponential	0.5	–	70.4	0.678	–	0.541	–	4.56×10^{-4}	–	3.64×10^{-4}	–
Lognormal	0.25	–	67.3	0.120	0.156	0.144	0.232	7.42	0.156	7.45	0.232
Weibull	8.0	1.0	69.2	7.64	1.20	5.66	1.23	7.64	1790.	5.66	1830.
BPT	1.0	0.25	64.0*	0.923	0.123	0.864	0.198	6.20×10^{-4}	0.123	5.81×10^{-4}	0.198
Slip-Size Data Not Used											
Poisson	1.0	–	69.7	0.750	–	0.729	–	5.04×10^{-4}	–	4.90×10^{-4}	–
Lognormal	1.0	–	70.2	0.0	0.431	0.0614	0.721	7.30	0.431	0.614	0.721
Weibull	2.0	1.0	72.7	2.34	1.15	1.76	1.14	2.34	1710.	1.76	1690.
BPT	1.0	0.5	65.8	1.22	0.223	1.06	0.384	8.21×10^{-4}	0.223	7.10×10^{-4}	0.384

[a] Mean recurrence time is 1475.0 years. See Table 5 for the table notes.

[43] As a result, the hazard functions with the MAP and posterior mean estimates and the Bayesian predictive hazard function shown in Figure 8b differ greatly from each other. Among these we select the Bayesian predictive hazard function for the same reason as described in section 4. It should also be noted from Figures 8a and 8b that the current hazard rate is several times higher than the average hazard rate for the fault.

[44] Figure 9a shows 100 samples of cumulative probability function (10) where each of the samples is derived from a sample hazard function in Figure 8a. Figure 9b shows the cumulative probability functions corresponding to the hazard functions in Figure 8b. We see large differences between the three estimates compared to Figure 3b for the case of the data in Table 1. Figures 9c and 9d show the estimated probability density of the next event corresponding to Figures 6a and 6b, respectively. It can be seen that the next event will most likely to occur in current years, based on the models of the optimal prior distributions fitted to the data in Table 3 with the slip-size information. Namely, the probability of the occurrence within the 21st century is about 30%.

7. Discussion

7.1. Methodological Aspect

[45] Any estimates based on a finite number of events are inevitably associated with some errors. For a large enough number of events, the errors of the hazard function can be approximately calculated using the second derivatives of the log likelihood (Hessian matrix) around the MLE [e.g., *Imoto*, 1999]. However, this does not always work well for data sets with small sample sizes. In contrast, the simulation of parameter values from posterior function is useful to see the variability or uncertainty of the estimated functions for prediction, such as those given in Figures 2a, 3a, 3c, 5a, 6a, 6c, 8a, 9a, and 9c.

[46] The extremely small MAP estimate of the scale parameter σ leads to the steep rise of the hazard rate function in Figure 5. Even in the case where only the occurrence times are available (or all the slip sizes are the same), the similar spiky and extremely skewed posterior function and then the steep hazard rises are conducted if intervals of consecutive events are about the same length such as the historical earthquakes around Odawara in Japan and the Parkfield earthquakes in San Andreas Fault. If we adopt this estimate or the MLE, the hazard function predicts with high probability that the next event will take place to keep the equidistant time interval from the last event. However, this kind of data is actually rare throughout the world, and examples of occurrences with such intervals number $4 \sim 5$ at most. In contrast, it is not unlikely that an ordinary inhibitory renewal process, such as the gamma, Weibull and lognormal, can simulate the occasional realization of events with $4 \sim 5$ consecutively equidistant intervals as discussed by *Utsu* [1994].

[47] In seismic hazard assessment, one of the most critical parameters in renewal models is the earthquake recurrence in time. Usually this parameter is poorly constrained due to the scarcity of large earthquakes occurring along the same fault segments. The only improvement to this unfortunate situation is to include some constraints from other explanatory variables. This paper has presented this issue and provided a way of introducing the slip information through a Bayesian inference. In this sense even if the recurrence

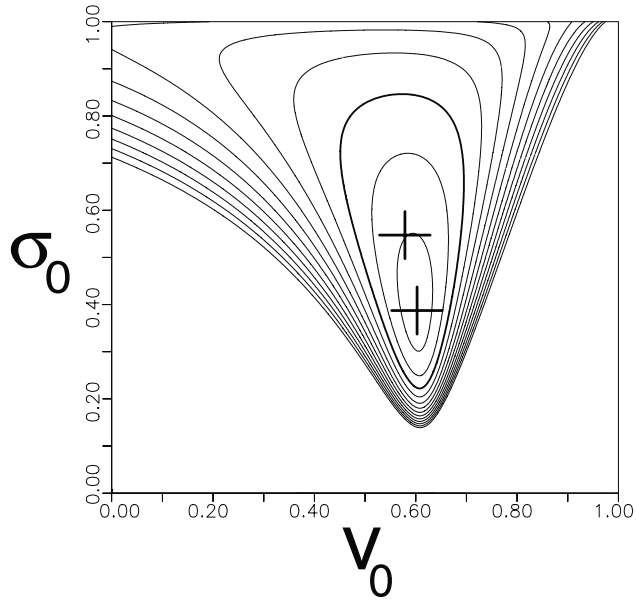

Figure 7. Heights of the posterior density distribution in equation (31) for the data in Table 3 are contoured at the unit intervals in logarithmic scale for the range from -10 to 2, with no contours being shown below the value -10. The lower and upper pluses indicate the MAP and posterior mean estimate (Bayes estimate), respectively.

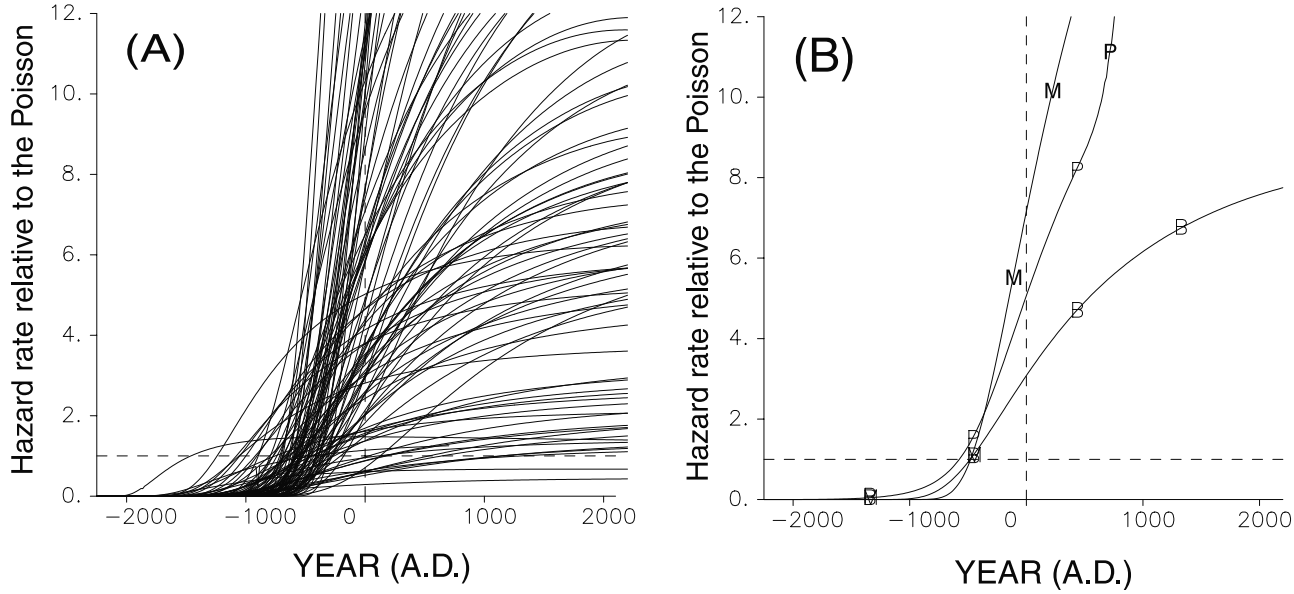

Figure 8. (a) One hundred samples of hazard rate functions of time, starting from the occurrence of the last event, simulated from the posterior probability distribution in equation (31) for the data in Table 3, consisting of four events and their associated estimated slip sizes. The horizontal dashed line indicates the mean hazard rate (6.72×10^{-4} event/year) with the vertical dashed lines indicating the current time. (b) The estimated hazard rate functions are similarly defined to those in Figures 2a and 2b.

data is poor, the knowledge on the slip may help to improve the hazard estimates in a positive way.

[48] For a further improvement of the hazard assessment, extensions of the BPT model will be promising if some records of physical factors that comprise the Brownian perturbations become available. In particular, if we are given information of slip mechanisms of events from neighboring faults together with their estimates of occurrence times and slip sizes, the external stress transfer effects from earthquakes outside the target source could become useful explanatory variables to improve the prediction performance.

7.2. On the Outcomes for the Next Nankai Earthquakes

[49] With regards to the long-term earthquake prediction at Nankai trough, the probabilities calculated based on the data sets in Tables 1 and 2 differ significantly to one another as described in section 6, even taking the uncertainty of their estimates into account. The probabilities in section 6.1 are consistent with those in section 3.1.1.1 of *Working Group on Assessment Methods of Long Term Earthquake Probability* [2001], using similar data to the data in Table 1, whereas the probabilities in section 6.2 are consistent with those in section 3.1.1.2 by the lognormal distribution using the stress accumulation rate and the last slip size with the empirically reasonable variability parameter values $\hat{\sigma} = 0.2$ or 0.3, as suggested by *Working Group on California Earthquake Probabilities* [1988, 1995].

[50] *Hori and Oike* [1996] considered a point-process model for the occurrence rate change of the historical earthquakes in the inland zone around the Nankai trough, forecasting that the next interplate event is most likely to take place around 2040, which is again consistent with the result in section 6.2.

[51] Another data set considered by *Ogata* [2001] includes the same 10 events as listed in Table 1 but those are associated with ranges of magnitudes as estimated by *Usami* [1996]. The slip-size-dependent lognormal model using the estimated seismic moment from the mean magnitudes is applied to obtain the outcome that the likely time of the next event is during this century, with maximum around 2020. Thus the results on the whole considered for the great events at Nankai trough are rather inconsistent. This means that investigation of sizes of the historical events and paleoearthquakes is crucial for the hazard assessment.

8. Conclusion

[52] In order to overcome the difficulty with the MLE under a small sample size, we have adopted the Bayesian inference, using the Bayesian predictive hazard function. The degree of uncertainty of the estimated hazard functions of the parametric model fitted to the data set is obtained by simulating the parameter values from the posterior function.

[53] Generalized models of the renewal processes are considered for the occurrence data associated with their slip sizes. These are stochastic extensions of the time-predictable model, and we have adopted the Bayesian method to objectively measure the predictive efficiency of the models, including the optimal choice of the prior distributions and the selection of models for likelihood function. Thus we have analyzed the three sets of data, using the proposed methods.

[54] The BPT model is selected to fit the first data set, consisting of ten historical great earthquakes from Nanaki trough. Its predictive hazard shows a large uncertainty (>100 years) of likely occurrence time centered around 2070. The second data set consists of the last three events in the first data set but associated with a record of slip sizes,

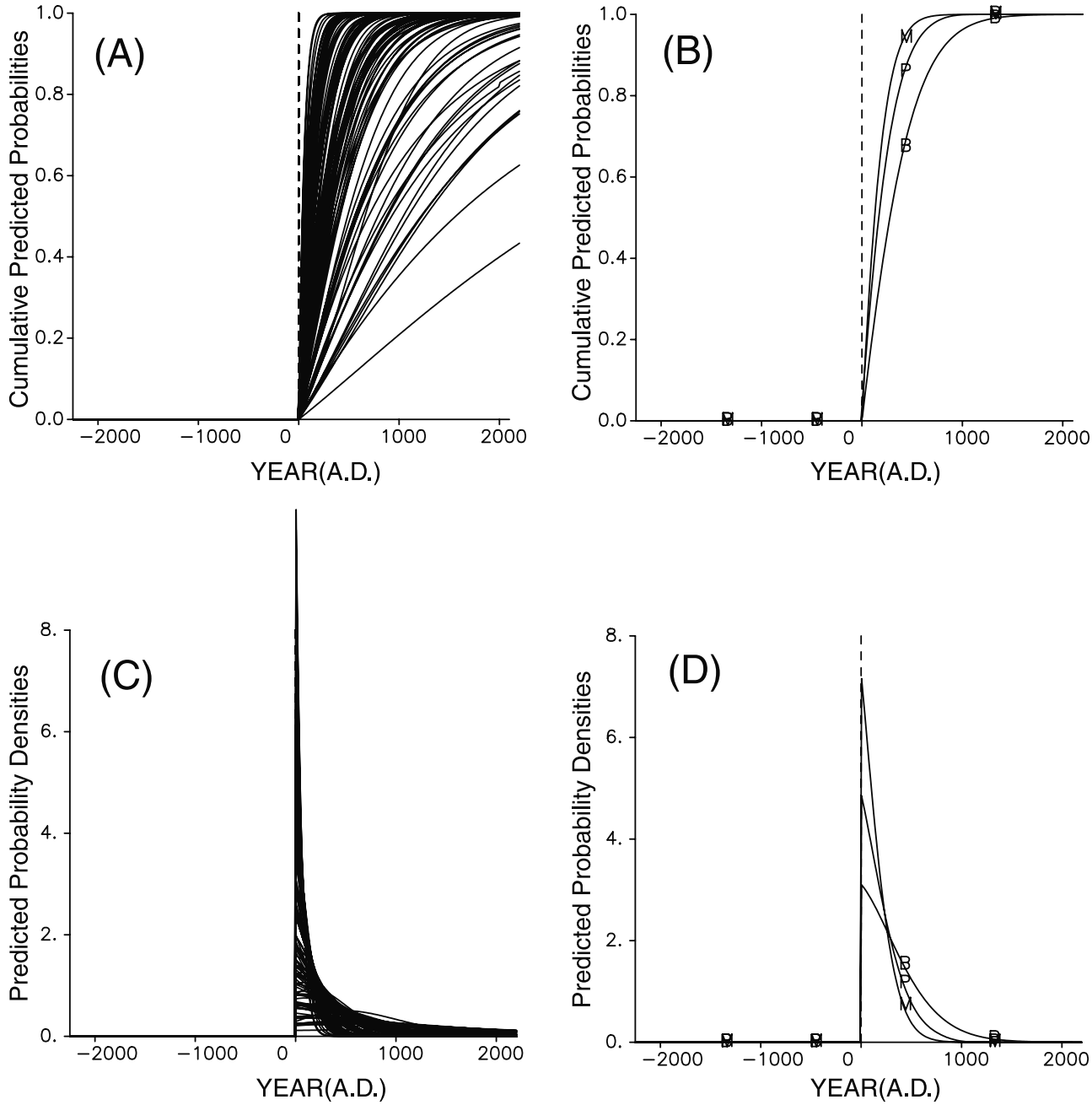

Figure 9. (a) One hundred samples of cumulative probability functions of time, starting from the current time, derived from the hazard rate functions in Figure 7a, using the first part of the equation in equation (10). (b) The similar cumulative probability functions based on the MAP, Bayes estimate (posterior mean estimate), and the Bayesian predictive hazard function. (c) and (d) Predicted probability density functions in equation (11) corresponding to each sample path of the function in Figures 7a and 7b, respectively.

for which the ABIC selected the slip-size-dependent log-normal renewal process model. The estimated predictive hazard function implies that next event is likely to occur around 2040 ± 10. The slip-size-dependent BPT model is selected to fit the last data set, consisting of four events with estimated occurrence times and slip sizes from the submarine fault. This model indicates that the likelihood of the next occurrence time is decreasing from now and that the period of its half decay is more than several hundred years.

[55] For each data set in the last two examples, it was also shown that the slip-size records are useful for better prediction of the next event.

[56] **Acknowledgments.** The theme of this paper is inspired by the discussions in the Working Group on Assessment Methods of Long Term Earthquake Probability (Chairman, K. Shimazaki) in the Committee of Earthquake Research, the Headquarters for Earthquake Research Promotion, the Prime Minister's Office, Japan. The completion of this paper was

facilitated by the generous help of Koiti Katsura, in particular, production of the figures. This study is partly supported by Grant-in-Aid 11680334 for Scientific Research, Ministry of Education, Science, Sports and Culture. Comments by Warner Marzocchi and the anonymous referee were useful clarification.

References

Akaike, H., A new look at the statistical model identification, *IEEE Trans. Automat. Control*, *19*, 716–723, 1974.

Akaike, H., A new look at the Bayes procedure, *Biometrika*, *65*, 53–59, 1978.

Akaike, H., Likelihood and Bayes procedure, in *Bayesian Statistics*, edited by J. M. Bernard et al., pp. 1–13, Univ. Press, Valencia, Spain, 1980.

Akaike, H., Prediction and entropy, in *A Celebration of Statistics, The ISI Centenary Volume*, edited by A. C. Atkinson and S. E. Feinberg, pp. 1–24, Springer-Verlag, New York, 1985.

Akaike, H., *Selected Papers of Horotugu Akaike*, edited by E. Parzen, K. Tanabe, and G. Kitagawa, 434 pp., Springer-Verlag, New York, 1998.

Box, G. P. B., and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, 588 pp., Addison-Wesley-Longman, Reading, Mass., 1973.

Brillinger, D. R., Seismic risk assessment: Some statistical aspect, *Earthquake Predict. Res.*, *1*, 183–195, 1982.

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

Davis, P. M., D. D. Jackson, and Y. Y. Kagan, The longer it has been since the last earthquake, the longer the expected time till the next?, *Bull. Seismol. Soc. Am.*, *79*, 1439–1456, 1989.

Ellsworth, W. L., M. V. Matthews, R. M. Nadeau, S. P. Nishenko, P. A. Reasenberg, and R. W. Simpson, A physically-based earthquake recurrence model for estimation of long-term earthquake probabilities, paper presented at Second Joint Meeting of the UJNR Panel on Earthquake Research, 1998.

Good, I. J, *The Estimation of Probabilities*, MIT Press, Cambridge, Mass., 1965.

Hori, T., and K. Oike, A statistical model of temporal variation of seismicity in the Inner Zone of Southwest Japan related to the interplate earthquakes along the Nankai trough, *J. Phys. Earth.*, *44*, 349–356, 1996.

Imoto, M., Errors of earthquake probabilities derived from renewal models for recurrent earthquakes (in Japanese), *Zisin II*, *52*, 337–340, 1999.

Matthews, M. V., A stochastic model for recurrent earthquakes, preprint, Walden Consult., Wayland, Mass., 1998.

Mulargia, F., and P. Gasperini, Evaluation of the applicability of the time- and slip-predictable earthquake recurrence models to Italian seismicity, *Geophys. J. Int.*, *120*, 453–473, 1995.

Nakata, T., and K. Shimazaki, Investigation of submarine faults (in Japanese), *Kagaku* (*Science*), *63*, 593–599, 1993.

Nishenko, S. P., and R. Buland, A generic recurrence interval for earthquake forecasting, *Bull. Seismol. Soc. Am.*, *77*, 1382–1399, 1987.

Ogata, Y., Statistical analysis of uncertain paleoearthquake occurrence times to forecast the hazard of rupture on an active fault segment, *J. Geophys. Res.*, *104*, 17,995–18,014, 1999.

Ogata, Y., Uncertainties in estimating the hazard of the next Nankai earthquake (in Japanese with English summary and figure captions), *Chigaku Zassi* (*J. Geogr.*), *10*, 602–614, 2001.

O'Hagan, A., *Kendall's Advanced Theory of Statistics*, vol. 2B, *Bayesian Inference*, 332 pp., Arnold, London, 1994.

Rhoades, D. A., R. J. Van Dissen, and D. J. Dowrik, On the handling of uncertainties in estimating the hazard of rupture on a fault segment, *J. Geophys. Res.*, *99*, 13,701–13,712, 1994.

Sangawa, A., *Seismoarchaeology* (in Japanese), 251 pp., Chuokoron-sha, Tokyo, 1992.

Savage, J. C., Criticism of some forecasts of the national earthquake prediction evaluation council, *Bull. Seismol. Soc. Am.*, *81*, 862–881, 1991.

Savage, J. C., The uncertainty in earthquake conditional probabilities, *Geophys. Res. Lett.*, *19*, 709–712, 1992.

Schrödinger, E., Zur theorie der fall und steigversuche an teilchen mit Brownscher bewegung, *Phys. Z.*, *16*, 289–295, 1915.

Seshadri, V., *The Inverse Gaussian Distribution*, *Lect. Notes Stat.*, vol. 137, 347 pp., Springer-Verlag, New York, 1998.

Shimazaki, K., and T. Nakata, Time-predictable recurrence model for large earthquakes, *Geophys. Res. Lett.*, *7*, 279–282, 1980.

Usami, T., *Materials for Comprehensive List of Destructive Earthquakes in Japan, 416-1995* (in Japanese), 494 pp., Univ. of Tokyo Press, Tokyo, 1996.

Utsu, T., Recurrence of several earthquakes at almost equal time intervals (in Japanese), *Zisin II*, *47*, 93–95, 1994.

Utsu, T., *Seismicity Studies: A Comprehensive Review*, 876 pp., Univ. of Tokyo Press, Tokyo, 1999.

Working Group on Assessment Methods of Long Term Earthquake Probability, Report on the assessment methods of long term earthquake probability and some applied examples (in Japanese), 74 pp., Comm. of Earthquake Res., Headquarters for Earthquake Res. Promotion, Prime Minist. Off., Tokyo, 2001.

Working Group on California Earthquake Probabilities, Probability of large earthquakes occurring in California on the San Andreas Fault, *U. S. Geol. Surv. Open File Rep., 88-398*, 1988.

Working Group on California Earthquake Probabilities, Seismic hazards in southern California: Probable earthquakes, 1994 to 2024, *Bull. Seismol. Soc. Am.*, *85*, 379–439, 1995.

Y. Ogata, Institute of Statistical Mathematics, Minani-Azabu 4-6-7, Minato-Ku, Tokyo 106-8569, Japan. (ogata@ism.ac.jp)