Statistical Science
2013, Vol. 28, No. 4, 521–541
DOI: 10.1214/13-STS439

A Prospect of Earthquake Prediction Research

Yosihiko Ogata

Abstract. Earthquakes occur because of abrupt slips on faults due to accumulated stress in the Earth's crust. Because most of these faults and their mechanisms are not readily apparent, deterministic earthquake prediction is difficult. For effective prediction, complex conditions and uncertain elements must be considered, which necessitates stochastic prediction. In particular, a large amount of uncertainty lies in identifying whether abnormal phenomena are precursors to large earthquakes, as well as in assigning urgency to the earthquake. Any discovery of potentially useful information for earthquake prediction is incomplete unless quantitative modeling of risk is considered. Therefore, this manuscript describes the prospect of earthquake predictability research to realize practical operational forecasting in the near future.

Key words and phrases: Abnormal phenomena, aseismic slip, Bayesian constraints, epidemic-type aftershock sequence (ETAS) models, hierarchical space–time ETAS models, probability forecasts, probability gains, stress changes.

1. INTRODUCTION

Through remarkable developments in solid Earth science since the late 1960s, our understanding of earthquakes has increased significantly. The availability of relevant data has steadily increased as the study of earthquakes has progressed remarkably in geophysics. After every major earthquake, researchers have elucidated important seismic mechanisms associated with it. However, even though detailed analysis and discussions have been conducted, large uncertainties remain because of diversity and complexity of the earthquake phenomenon. This leads to unachievable challenges in deterministic earthquake prediction because all diverse and complex scenarios must faithfully reflect the processes of earthquakes to be considered for effective earthquake prediction.

On the other hand, several techniques for predicting earthquakes have been proposed on the basis of anomalies of various types; however, the effectiveness of these techniques is controversial (Jordan et al., 2011). Therefore, objectivity is required for such evaluation; otherwise, arguments presented may lack merit. New prediction models that claim to incorporate potentially useful information over those used in standard seismicity models should be evaluated to determine whether predictive power is improved. Earthquake forecasting models should evolve in this manner.

Recently, there has been growing momentum for seismologists to develop an organized research program on earthquake predictability. An international cooperative study known as *Collaboratory for the Study of Earthquake Predictability* (CSEP; http://www.cseptesting.org/) is currently under way among countries prone to major earthquakes for exploring possibilities in earthquake prediction (e.g., Jordan, 2006). An immediate objective of the project is to encourage the development of statistical models of seismicity, such as those subsequently discussed in Section 2, and to evaluate their predictive performances in terms of probability.

In addition, the CSEP study aims to develop a scientific infrastructure to evaluate statistical significance and *probability gain* (Aki, 1981) of various methods

Yosihiko Ogata is Professor Emeritus, Institute of Statistical Mathematics, Information and System Research Organization, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562 and Institute of Industrial Science, University of Tokyo, Visiting Professor, 4-6-1 Komaba, Megro-Ku, Tokyo 153-8505 (e-mail: ogata@ism.ac.jp).

used to predict large earthquakes by using observed abnormalities such as seismicity anomaly, transient crustal movements and electromagnetic anomaly. Here *probability gain* is defined as the ratio of probability of a large earthquake estimated based on an anomaly to the underlying probability without anomaly. Section 3 describes this important concept, and then discusses statistical point-process models to examine the significance of causality of anomalies and also to evaluate the probability gains conditional on the anomalous events.

For prediction of large earthquakes with a higher probability gain, comprehensive studies of anomalous phenomena and observations of earthquake mechanisms are essential. Several such studies are summarized in Sections 4–6. Particularly, I have been interested in elucidating abnormal seismic activities and geodetic anomalies to apply them for promoting forecasting abilities, as described in these sections.

2. PROBABILITY FORECASTING OF BASELINE SEISMICITY

2.1 Log-Likelihood for the Evaluation Score of Probability Forecast

Through repeated revisions, CSEP attempts to establish standard models to predict probability that conform to various parts of the world. Here I mean the prediction/estimation of probability as predicting/estimating conditional probabilities given the past history of earthquakes and other possible precursors. The likelihood is used as a reasonable measure of prediction performance (cf. Boltzmann, 1878; Akaike, 1985). The evaluation method for probabilistic forecasts of earthquakes by the log-likelihood function has been proposed, discussed and implemented (e.g., Kagan and Jackson, 1995; Ogata, 1995; Ogata, Utsu and Katsura, 1996; Vere-Jones, 1999; Harte and Vere-Jones, 2005; Schorlemmer et al., 2007; Zechar, Gerstenberger and Rhoades, 2010; Nanjo et al., 2012; Ogata et al., 2013). In some such studies, the evaluation score has been referred to as (relative) entropy, which is essentially similar to the log-likelihood.

2.2 Space–Time-Magnitude Forecasting of Earthquakes

Baseline models should be set to compare with and evaluate all predictability models. Based on empirical laws, we can predict standard reference probability of earthquakes in a space–time-magnitude range on the basis of the time series of present and past earthquakes. The framework of CSEP, which has evaluated performances of submitted forecasts of respective regions (Jordan, 2006; Zechar, Gerstenberger and Rhoades, 2010; Nanjo et al., 2011), is similar to that of the California Regional Earthquake Likelihood Models (RELM) project for spatial forecast (Field, 2007; Schorlemmer et al., 2010). Different space–time models were submitted to the CSEP Japan Testing Center at the Earthquake Research Institute, University of Tokyo, for the one-day forecast applied to the testing region in Japan (Nanjo et al., 2012). This means that the model forecasts the probability of an earthquake at each space–time-magnitude bin. However, the CSEP protocols have to be improved to those in terms of point-processes on a continuous time axis for the evaluation including a real-time forecast (Ogata et al., 2013).

Almost all models incorporated the Gutenberg–Richter (G–R) law for forecasting the magnitude factor, and take different variants of the space–time epidemic-type aftershock sequence (ETAS) model (Nanjo et al., 2012, and Ogata et al., 2013). In the following sections, I will outline these models.

2.2.1 *Magnitude frequency distribution*. Gutenberg and Richter (1944) determined that the number of earthquakes increased (decreased) exponentially as their magnitude decreased (increased). Describing this theory in terms of point processes, the intensity of magnitude M is

$$\lambda_0(M) = \lim_{\Delta\to 0} \frac{1}{\Delta} \mathrm{Prob}(M < \mathit{Magnitude} \le M + \Delta) = 10^{a-bM} = Ae^{-\beta M} \tag{1}$$

for constants a and b. In other words, the magnitude of each earthquake will obey an exponential distribution such that $f(M) = \beta\, \mathrm{e}^{-\beta(M-Mc)}$, $M \ge M_c$, where $\beta = b \ln 10$, and M_c is a cutoff magnitude value above which all earthquakes are detected. Traditionally, the b-value had been estimated graphically, however, more efficient estimation is performed by the maximum likelihood method. Utsu (1965) derived it by the moment method. Later, Aki (1965) demonstrated that this is a maximum-likelihood estimate (MLE) and provided the error estimate. It should be noted that the magnitudes in most catalogs are given in the interval of 0.1 (discrete magnitude values), hence, care should be taken for avoiding the bias of the b estimate in likelihood-based estimation procedures (Utsu, 1969).

Although the coefficient b in a wide region is generally slightly smaller than 1.0, Gutenberg and Richter (1944) further determined that the b-value

varies according to location in smaller seismic regions. The b-value varies even within Japan and further varies with time. Temporal and spatial b-value changes have attracted the attention of many researchers ever since Suyehiro (1966) reported a difference between b-values of foreshocks and aftershocks in a sequence.

Here, we consider that β can vary with time, space and space–time according to a function such as $\beta(t)$, $\beta(x, y, z)$ or $\beta(t, x, y)$. Various nonparametric smoothing algorithms such as kernel methods have been proposed (Wiemer and Wyss, 1997). Alternatively, the β value can be parameterized by smooth cubic splines (Ogata, Imoto and Katsura, 1991; Ogata and Katsura, 1993) or piece-wise linear expansions on Delaunay tessellated space (Ogata, 2011c; see also Figure 1, e.g.). In such a case, a penalized log-likelihood (Good and Gaskins, 1971) is used whereby the log-likelihood function is associated with penalty functions in which the coefficients are constrained for smoothness of the β function. For the optimal estimation of the β function, the weights in the penalty function are objectively adjusted in a Bayesian framework, as suggested by Akaike (1980a).

2.2.2 *Aftershock analysis and probabilistic forecasting*. Typical aftershock frequency decays according to the reverse power function with time (Omori, 1894; Utsu, 1961, 1969). First, let $N(s, t)$ be the number of aftershocks in an interval (s, t). Then, the occurrence rate of aftershocks at the elapsed time t since the main

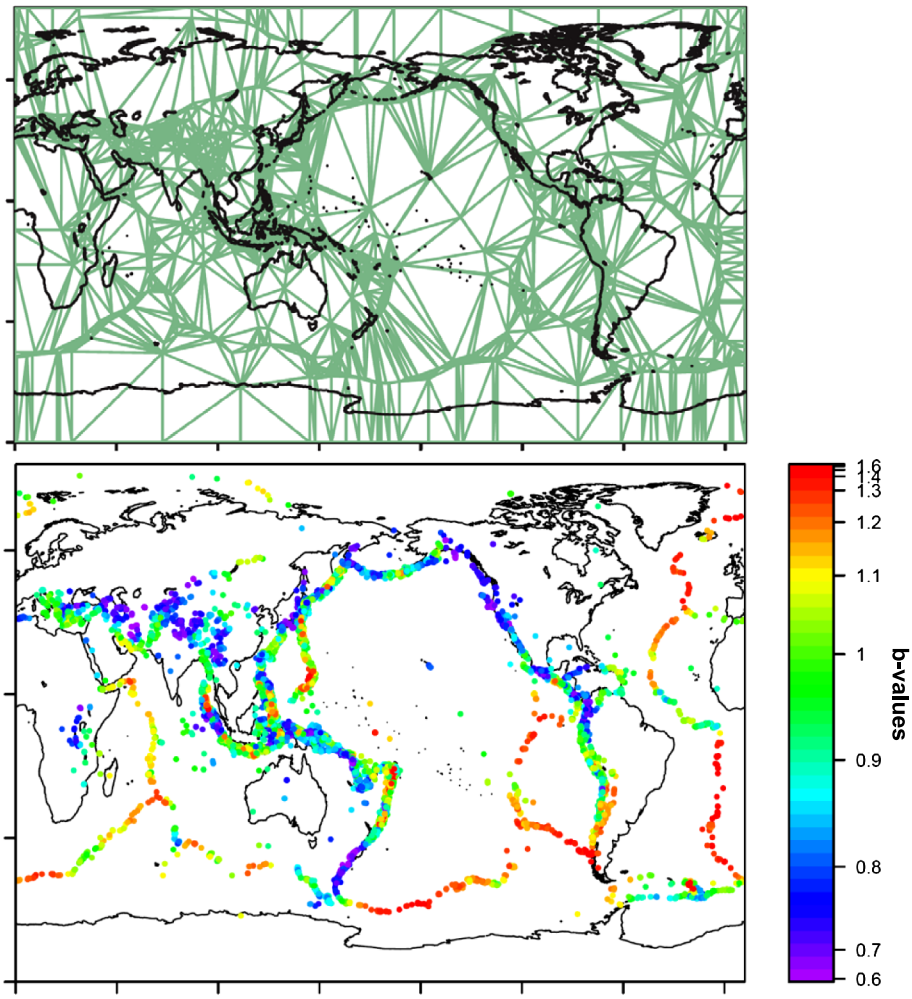

FIG. 1. *Top panel shows Delaunay tessellation upon which the piecewise linear function is defined. The smoothness constraint is posed in the sum of squares of integrated slopes. Bottom panel shows b-value estimates of the G–R formula in equation (1) estimated from the data for earthquakes of* $M \geq 5.4$ *from the Harvard University global CMT catalog (http://www.globalcmt.org/CMTsearch.html). One conspicuous feature is that b-values are large in oceanic ridge zones, but small in plate subduction zones.*

shock is

$$\nu(t) = \lim_{\Delta\to 0}\frac{1}{\Delta}P\{N(t,t+d\Delta)\geq 1| \text{ Mainshock at time } 0\} = \frac{K}{(t+c)^p} \tag{2}$$

for constants K, c and p. This is known as the Omori–Utsu (O–U) law.

Traditionally, estimates of the parameter p have been obtained since the study of Utsu (1961) in the following manner. The numbers of aftershocks in a unit time interval $n(t)$ are first plotted against elapsed time on doubly logarithmic axes, and then are fit to an asymptotic straight line. The slope of this line is an estimate for p. The values of c can be determined by measuring the bending curve immediately after the main shock. Such an analysis is based on the time series of counted numbers of aftershocks. By such a plot, we can find aftershock sequences for which the formula (2) lasts a long period, more than 120 years, for example (Utsu, 1969; Ogata, 1989; Utsu, Ogata and Matsu'ura, 1995).

To efficiently estimate the three parameters directly on the basis of occurrence time records of aftershocks, assuming nonstationary Poisson process with intensity function (2), Ogata (1983) suggested the maximum-likelihood method, which enabled the practical aftershock forecasting. Reasenberg and Jones (1989) proposed a procedure based on the joint intensity rate of time and magnitude of aftershocks (Utsu, 1970) according to the G–R law (1),

$$\lambda(t,M) = \lambda_0(M)\nu(t) = \frac{10^{a+b(M_0-M)}}{(t+c)^p} \quad (a,b,c,p;\ \text{constant}), \tag{3}$$

where M is the magnitude of an aftershock and t is the time following a main shock of magnitude M_0; the parameters are independently estimated by the maximum-likelihood method for respective empirical laws.

After a large earthquake occurs, the Japan Meteorological Agency (JMA) and the United States Geological Survey (USGS) have undertaken operational probability forecast of the aftershocks. However, the forecast is announced after the elapse of 24 h or more. This is due to the deficiency of aftershock data due to overlapping of seismograms after the main shock. In particular, the parameter a is crucial for the early forecast, but difficult to estimate in an early period, whereas the other parameters can be default values for the early forecast [Reasenberg and Jones, 1994; Earthquake Research Committee (ERC), 1998]. The difficulty is because the parameter a can substantially differ even if the magnitudes of the main shocks are almost the same: for example, the numbers of the aftershocks of M $\geq$ 4.0 of two nearby main shocks of the same M6.8 differ by 6–7 times (JMA, 2009).

It is notable that the strongest aftershocks occurred within 24 h in most sequences (JMA, 2008). Therefore, despite adverse conditions during data collection, probabilistic aftershock forecasts should be delivered as soon as possible within 24 h after the main shock to mitigate secondary disasters in affected areas.

For this purpose, it is necessary to estimate time-dependent missing rates, or detection rates, of aftershocks (Ogata and Katsura, 1993, 2006; Ogata, 2005c) because they enable probabilistic forecasting immediately after the main shock (Ogata, 2005c; Ogata and Katsura, 2006). The detection rate of earthquakes is described by a probability function $q(M)$ of magnitude M such that $0 \leq q(M) \leq 1$. The intensity $\lambda(M)$ for actually observed magnitude frequency is described by $\lambda(M) = \lambda_0(M)q(M)$, corresponding to thinning or random deletion. An example of the detection rate function is the cumulative of Gaussian distribution or the so-called error function $q(M) = \text{erf}\{M|\mu,\sigma\}$. The parameter μ represents the magnitude at which earthquakes are detected at a rate of 50%, and σ represents a range of magnitudes in which earthquakes are partially detected. Let a data set of magnitudes $\{(t_i, M_i); i = 1,\ldots,N\}$ be given at a period immediately after the main shock. Assume that the parameters are time-dependent during the period such that

$$\lambda(t,M) = \frac{10^{a+b(M_0-M)}}{(t+c)^p}q\{M|\mu(t),\sigma\} \tag{4}$$

with an improving detection rate $\mu(t)$. An additional parametric approach proposed by Omi et al. (2013) uses the state–space representation method for real-time forecasting within the 24 h period.

2.2.3 *Epidemic-type aftershock sequence (ETAS) model*. The epidemic-type aftershock sequence (ETAS) model describes earthquake activity as a point process (Ogata, 1986, 1988) and includes the O–U law for aftershocks as a descendant process. This model assumes that the background seismicity is a stationary Poisson process with a constant occurrence rate or number of earthquakes per day, μ. The conditional intensity function of the process is described by

$$\lambda_\theta(t|H_t) = \mu + \sum_{\{i:t_i<t\}}\frac{K}{(t-t_i+c)^p}e^{\alpha(M_i-M_0)}, \tag{5}$$

where $H_t = \{(t_i, M_i); t_i < t\}$ is the history of the occurrence times and magnitudes of earthquakes before time t, and M_0 is a reference magnitude throughout the data; it can be a threshold magnitude in case of general seismic activity or a main shock magnitude in case of a single aftershock sequence. The parameters K, α, c and p are constants, and their detailed features are summarized and discussed in Utsu, Ogata and Matsu'ura (1995), for example. Here, in simulations and forecasting, magnitude sequence is usually assumed to be independent and identically distributed according to the G–R law (Section 2.2.1) unless otherwise modeled like in Ogata (1989).

We estimate the ETAS parameters by using the maximum-likelihood estimation where the log-likelihood function, or rigorously partial log-likelihood (Cox, 1975),

$$\log L(\theta; S, T) = \sum_{\{i; S<t_i<T\}} \log \lambda_\theta(t_i|H_{t_i}) - \int_S^T \lambda_\theta(t|H_t)\,dt \tag{6}$$

is maximized with respect to the parameters $\theta = (\mu, K, c, \alpha, p)$. Here, $\{(t_i, M_i), M_i \geq M_c; i = 1, 2, \ldots\}$ are data from the period $[0, T]$ consisting of occurrence times and magnitudes of earthquakes above a thereshold M_c. Here, note that the magnitudes are exogenous variables. The ETAS model is applied to data from the target time interval $[S, T]$. The occurrence history H_S during the precursor period $[0, S]$ is used for sustaining stationarity of the process after the time S.

Then, the model's effectiveness in fitting an earthquake sequence can be evaluated by comparing the cumulative number $N(S, t)$ of earthquakes with the rate predicted by the model

$$\Lambda(S, t) = \int_S^t \lambda(u|H_u)\,du \tag{7}$$

in the time interval $S < t < T$. If earthquakes in the catalog are described effectively by the ETAS model, the transformed time τ_i defined as $\tau_i = \Lambda(t_i)$, which include correction for the O–U law decay, will be distributed according to the stationary Poisson process, and the plot of the actual cumulative number of events versus transformed time should be close to linear (Ogata, 1988). The transformed time τ_i is useful for judging goodness of fit of the ETAS model because it assigns a visual check of the fit to a stationary Poisson process. Anomalous seismicity, not explained by the stationary ETAS model, will appear as systematic deviations from this trend. An example of such an analysis will be presented in Section 4.3 and Figure 4.

To predict in real time, the probability of occurrence of future earthquakes using the data of earthquakes in the past, the ETAS model has been used. For example, the ETAS model and its space–time extensions (see Section 2.2.4) are reviewed in the next version on operational earthquake forecast in California (Working Group on California Earthquake Probabilities, WGCEP, 2012).

2.2.4 *Space–time ETAS model*. The space–time ETAS model considers space–time occurrence rate at the time t and location (x, y), conditional on the occurrence history up to time t, such that

$$\begin{aligned}\lambda(t, x, y|H_t) &= \mu(x, y) \\ &\quad + \sum_j^{t_j<t} \frac{K}{(t - t_j + c)^p} \\ &\quad \times \left[\frac{(x - \bar{x}_j, y - \bar{y}_j) S_j \binom{x-\bar{x}_j}{y-\bar{y}_j}}{e^{\alpha(M_j - M_c)}} + d\right]^{-q},\end{aligned} \tag{8}$$

where S_j is a normalized positive definite symmetric matrix for anisotropic clusters such that

$$\begin{aligned}&(x, y) S (x, y)^t \\ &\quad = \frac{1}{\sqrt{1-\rho^2}} \left\{\left(\frac{\sigma_2}{\sigma_1}\right) x^2 + 2\rho xy + \left(\frac{\sigma_1}{\sigma_2}\right) y^2\right\}.\end{aligned} \tag{9}$$

Here, $(\bar{x}_j, \bar{y}_j)$ is an average location of earthquakes that are placed in the same cluster as (x_j, y_j). Both $(\bar{x}_j, \bar{y}_j)$ and coefficients of S_j for a selected set of large earthquakes j are identified by fitting a bivariate normal distribution to spatial coordinates of the cluster occurring within a square of $3.33 \times 10^{0.5M_j - 2}$ km side-length and within $10^{0.5M_j - 1}$ days after the large event of magnitude M_j, according to Utsu (1969); but I use 1 h in prediction stage. The locations $(\bar{x}_j, \bar{y}_j)$ of all other events, including cluster members, remain the same as the epicenter coordinates of the original catalog; and they are associated with the identity matrix for S_j, namely, $\sigma_1 = \sigma_2 = 1$ and $\rho = 0$. See Figure 2 for an illustrative view of the conditional intensity (8). Further details of the algorithm can be found in studies of Ogata (1998, 2011a, 2011b).

Although several alternative versions to the spatial factor given by the bracket of (8), as described by Ogata (1998), are available, the form in (8) fits best

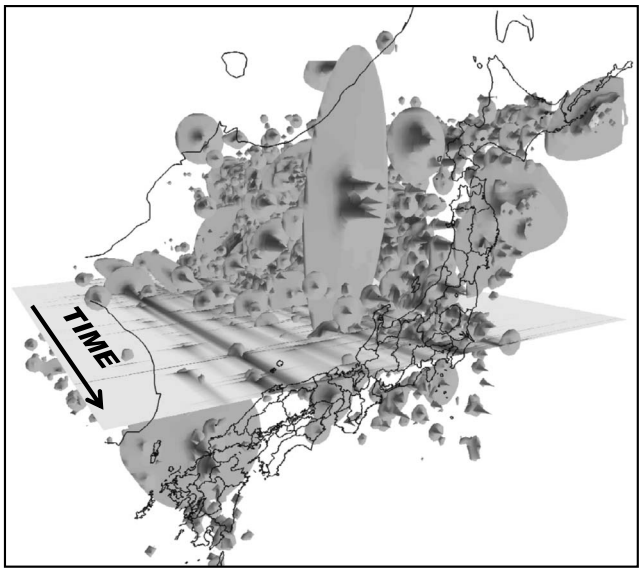

FIG. 2. *Iso-surface plot of the estimated conditional intensity function* (8) *of the space–time ETAS model* (*Ogata*, 1998) *to the JMA hypocenter data of shallow earthquakes* (*depth* $\leq$ 100 *km*) *of magnitude* 5.0 *or larger from the period* 1926–1995; *in addition, Utsu's earthquake catalog for the* 40 *years period before* 1926 (1885–1925) *was used as the preceding occurrence history of the space–time ETAS model. Space means longitude and latitude, whereas the depth data are neglected.*

in terms of the Akaike information criterion (AIC; Akaike, 1974) for Japanese earthquake data sets. All extensions of the temporal ETAS model are referred to as space–time ETAS models (e.g., Nanjo et al., 2012).

2.2.5 *Hierarchical space–time ETAS model.* When a region becomes wide or the number of earthquakes becomes sufficiently large, spatial heterogeneity of seismicity becomes conspicuous. For example, many studies have been conducted on regional variation of seismicity-related parameters such as the b-value of the G–R law and p-values of the O–U law (Utsu, 1961, 1969; Mogi, 1967).

Regarding space–time ETAS models, the aftershock productivity K may differ significantly among locations, even if magnitudes of triggering earthquakes are similar (see Section 2.2.2). Moreover, the main shock–aftershock and swarm-type clusters exhibit significantly different activity patterns. Therefore, we applied an extension to the above space–time model to earthquakes in the entire region for developing a hierarchical space–time ETAS (HIST–ETAS) model, which is a space–time ETAS model in which parameter values μ, K, α, p and q can vary depending on location, such as $\mu(x, y)$, $K(\bar{x}_j, \bar{y}_j)$, $\alpha(\bar{x}_j, \bar{y}_j)$, $p(\bar{x}_j, \bar{y}_j)$ and $q(\bar{x}_j, \bar{y}_j)$.

Thus, coefficients of parameter functions of the space–time ETAS model in equation (8) must be evaluated. Coefficients of each parameter function are defined by values at epicenter locations of earthquakes and a number of points on the region boundary. Hence, each function is uniquely defined by linear interpolation of values at three nearest points (earthquakes) determined by Delaunay tessellation that is constructed by all the earthquake locations and additional points on the boundary of the region (see Figure 1).

For a stable optimal estimation, the freedom of coefficients of parameter functions needs to be constrained to assign penalties against roughness of the functions. The coefficients that maximize the penalized log-likelihood are then sought, which is the equivalent of attaining the maximum posterior distribution. Here, we adjusted the optimal prior function for the parameter constraints in terms of the penalty function by an empirical Bayesian method (Akaike, 1980a). Further details can be found in the studies of Ogata, Katsura and Tanemura (2003) and Ogata (2004a, 2011b). Figure 3 shows the optimal solution of background seismic activities $\mu(x, y)$, which appear useful for long-term prediction of large earthquakes in and near Japan.

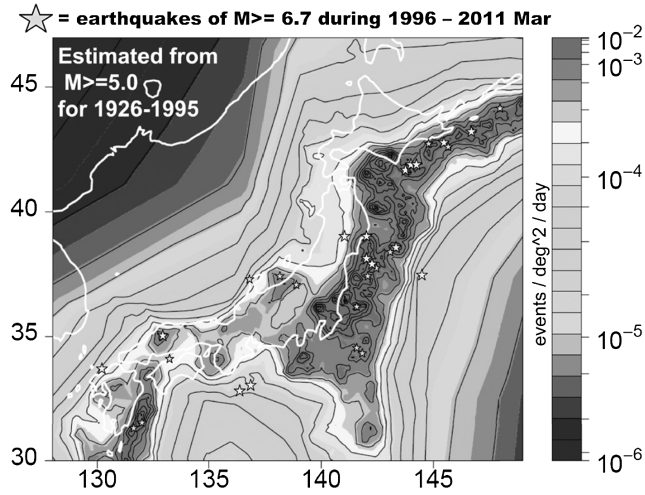

FIG. 3. *The optimal solution of the* μ*-values for background seismicity of the space time ETAS model* (8) *in terms of minimum ABIC priors. The model is estimated from the JMA data with earthquakes of M*5.0 *or larger for the target period* 1926–1995. *In addition, Utsu's earthquake catalog for the* 40 *years period before* 1926 (1885–1925) *was used as the precursory occurrence history of the space–time ETAS model. Contours are equidistant in the logarithmic scale. Stars indicate locations of earthquakes of magnitude* 6.7 *or larger that occurred during* 1996–2009, *which mostly occurred in high background rates. Note however that there are several large earthquakes occurred at very low seismicity rates, the issue of which lead to Section* 2.3.

Moreover, stochastic declustering using the space–time ETAS model can make realizations of background seismicity (Zhuang, Ogata and Vere-Jones, 2002; Zhuang et al., 2005b; Bansal and Ogata, 2013).

2.3 Long-Term Probability Forecasts of Characteristic Earthquakes

A characteristic earthquake is a repeating large earthquake that is traditionally defined from paleoseismology observations. The estimation is made by using recurrence times of a large earthquake on an active fault or a particular seismogenic region on a plate boundary. The Earthquake Research Committee of Japan (ERC, 2001) adopted the Brownian Passage Time (BPT; Matthews, Ellsworth and Reasenberg, 2002) renewal process, in which the inter-event probability density function is given by

$$f(x|\mu,\alpha)=\sqrt{\frac{\mu}{2\pi\alpha^2x^3}}\exp\left\{-\frac{(x-\mu)^2}{2\mu\alpha^2x}\right\}. \tag{10}$$

This equation considers the potential of further model extensions by useful physical concepts in the elastic rebound theory, such as stress interaction from neighboring earthquake ruptures. This physical concept will be subsequently described in Sections 4 and 5. BPT renewal process is based on the following Brownian perturbation process:

$$S(t)=\lambda t+\sigma W(t),\qquad t\geq 0, \tag{11}$$

which includes linearly increasing drift for stress accumulation and diffusion rate σ. An earthquake occurs when the path $S(t)$ attains the critical stress level s_f, and the accumulated stress is released down to the ground state s_0 based on elastic rebound theory of earthquakes (Reid, 1910). Random fluctuations represent the transient stress changes due to the effect of other earthquakes in close proximity (see Section 4). This model includes four parameters: the stress accumulation rate λ, perturbation rate σ, failure state s_f, and ground state s_0. If we assume that failure and ground states s_f and s_0, respectively, are constant, the interval of earthquakes is independent and identically distributed with the BPT distribution, in which parameters are related by $\mu=(s_f-s_0)/\lambda$ and $\alpha=\sigma/\sqrt{\lambda(s_f-s_0)}$.

Because of very small sample size available from each fault, the mean parameter μ has been estimated using methods other than the MLE. The ERC (2001) uses a common α value of 0.24 throughout Japan. This is because better fit of the same α value was shown by the AIC comparison than the different α estimates for respective active faults, for a set of occurrence data with moderate sample sizes from four active faults (ERC, 2001). Also, the ERC has estimated μ in two ways: as the mean of past recurrence intervals and as expected intervals estimated from the slip data of the fault plane. The latter estimate is expressed by $T=U/V$, where U is the slip size per earthquake and V is the deformation rate per year, observed from the escarpment of the fault. The ERC selects and applies one of these estimates for μ of each active fault according to reliability of the data.

Alternatively, Nomura et al. (2011) propose following Bayesian estimation procedure assuming a common prior distribution for fault segments throughout Japan. Consider historical occurrence data $\mathbf{X}_j=\{X_i^j;i=1,2,\ldots,n\}$ in the jth segment of m fault segments. Consider a posterior density

$$\begin{aligned}&\text{posterior}(\mu_j,\alpha_j|T_j,\phi_\mu,\phi_\alpha)\\&\quad=L(\mu_j,\alpha_j|\mathbf{X}_j)\pi_1(\mu_j|T_j,\phi_\mu)\pi_2(\alpha_j|\phi_\alpha),\end{aligned} \tag{12}$$

where likelihood L is based on the renewal process taking account of the forward and backward recurrence times (Daley and Vere-Jones, 2003) and T_j is the above-mentioned geologically estimated slip deformation ratio from slip data. Furthermore, the values of the hyperparameters ϕ_μ and ϕ_α characterizing the prior densities of μ and α are common to all considered fault segments. We obtain their estimates by maximizing the integrated posterior distribution

$$\begin{aligned}\Lambda(\phi)=\prod_{j=1}^{m}\int_0^\infty\int_0^\infty&\text{posterior}(\mu_j,\alpha_j|\\&T_j,\phi_\mu,\phi_\alpha)\,d\mu_j\,d\alpha_j,\end{aligned} \tag{13}$$

where the subscript j represents the jth segment of m fault segments. This maximizing procedure is called the Type II maximum likelihood method (Good, 1965). The selection of the best combination of the prior distribution factors and the optimal values of the hyperparameters in (12) are carried out to attain the smallest value of the Akaike Bayesian information criterion (*ABIC*; Akaike, 1980a) that is defined by $ABIC=-2\max_\phi\log\Lambda(\phi)+2\dim(\phi)$, where $\dim\{\phi\}$ denotes the number of hyperparameters.

The most common forecast technique is a plug-in method, which is a probability forecast that uses a distribution or conditional intensity function with a parameter set to its estimated value, such as MLE. This

method works well if the estimation error is sufficiently small. However, its predictive performance can be inadequate when the sample size is small. Hence, ERC adopts the plug-in method only for μ and uses a common α value of 0.24 throughout Japan. Alternatively, Rhoades, Van Dissen and Dowrick (1994), Ogata (1999b, 2002) and Nomura et al. (2011) propose the Bayesian prediction (Akaike, 1985)

$$\tilde{h}(y|\mathbf{X}) = \int_0^\infty \int_0^\infty h(y|\mu,\alpha)\{1 - F(y|\mu,\alpha)\} \times \prod_{i=1}^{n} f(X_i|\mu,\alpha)\, d\mu\, d\alpha, \tag{14}$$

where $F(y|\mu,\alpha)$ is the cumulative distribution of the density $f(y|\mu,\alpha)$ in (10), and $h(y|\mu,\alpha)$ is the hazard rate function. This prediction is shown to provide a systematically better performance in the sense of expected entropy criterion (Akaike, 1985) than the plug-in predictor in case of very small sample sizes of data (Nomura et al., 2011).

Bayesian framework can also be used when the occurrence times are uncertain, especially when we are dealing with geological data (Ogata, 1999b; Nomura et al., 2011) in addition to the magnitude dependent model (Ogata, 2002) based on the time-predictable model (Shimazaki and Nakata, 1980).

3. PRACTICAL EARTHQUAKE FORECASTING

Probability gain refers to the ratio of predicted conditional probability relative to baseline earthquake probability. As far as I know, most probability gains of predictions are not very high, even relative to the stationary Poisson process model. Therefore, predictions based on a single anomaly data set alone are not satisfactory for disaster prevention because baseline probability of a large earthquake itself is very small according to the G–R law. Also, the BPT renewal process model has been applied to active fault segments to estimate time-dependent probability on the basis of the last earthquake and stress accumulation rate. In Californian, probability gain showed an improvement of approximately 1.7 times over Poisson process model predictions (Jordan et al., 2011).

The key for research progress in practical probability earthquake forecasting is to use a multiple prediction formula (Utsu, 1979) such that total probability gain is approximately equal to the product of individual probability gains (Aki, 1981). The rate of probability gain that an individual anomaly was actually a precursor to an earthquake may be calculated as its success rate of the anomaly divided by precursor time (Utsu, 1979). Success rate can only be determined from accumulation of actual earthquake occurrences; and precursor time can be studied experimentally and theoretically (Aki, 1981). In this section I review important suggestions by Utsu (1979) and Aki (1981) and provide some examples of causality modeling toward improved accuracy for probability gain.

3.1 Abnormal and Precursor Phenomena

The continuing pursuit of possible algorithms used to predict large earthquakes should consider specific developmental patterns listed in the seismic catalog. So far, an alarm-type method of earthquake prediction (Keilis-Borok et al., 1988; Keilis-Borok and Malinovskaya, 1964; Rundle et al., 2002; Shebalin et al., 2006; Sobolev, 2001; Tiampo, et al., 2002) based on seismicity patterns has been operationally implemented, in which predictions are conveyed to seismologists through e-mail. Some predictions in each year are published in an official document such as the Center for Analysis and Prediction, State Seismological Bureau, China (1990–2003). In addition, many papers have been published on earthquake predictions. Some of these predictions may be statistically significant against the stationary Poisson process that is assumed for normal seismicity. Such alarm-type predictions have been further evaluated by Zechar and Zhuang (2010), Jordan et al. (2011), Zhuang and Ogata (2011) and Zhuang and Jiang (2012).

Comprehensive studies of anomalous phenomena and observations of earthquake mechanisms are essential for predicting large earthquakes with high probability gains. However, it is difficult to determine whether detected abnormalities are precursors to large earthquakes. Nevertheless, we aspire to become able to say that the probability of occurrence of a large earthquake, in a certain period and a certain region, has increased a certain extent as compared with the reference probability. Therefore, it is necessary to estimate uncertainty of the nature and urgency of abnormal phenomena relative to their roles as precursors to major earthquakes. For this purpose, it is necessary to study a large number of anomalous cases for potential precursory links to large earthquakes. Thus, incorporation of this information in the design of a prediction model for probability that exceeds the underlying probability is important.

3.2 Conditional Probability of an Earthquake for Multiple Independent Precursors

As previously described, although an individual precursory anomaly is insufficient for providing a fore-

cast of an earthquake with a high probability, forecasting probability can be enhanced if several anomalies are simultaneously observed (Utsu, 1977, 1979, 1982; Aki, 1981). The probability of an anomaly being a precursor of a large earthquake should be estimated through comprehensive observations. Then, it provides medium- or short-term probability forecasts depending on the time scale of enhanced earthquake probability following the anomaly. For example, identification of foreshocks (Section 3.4.2) and seismicity quiescence (Section 5.1) belongs to short- and medium-term forecasting, respectively.

Let us find the probability $P(E_M|A, B, C, \ldots, S)$ of occurrence of an earthquake, with a magnitude greater than M in a specified area, under the condition that N anomalies $A, B, C, \ldots, S$ appeared simultaneously. Assuming that anomalies are conditionally independent on E_M and the complement of E_M, Aki (1981) derived the following equation of Utsu (1977, 1979) using Bayes' theorem:

$$\begin{aligned} &P(E_M|A, B, C, \ldots, S) \\ &= \left[1 + \left(\frac{1}{P_A} - 1\right)\left(\frac{1}{P_B} - 1\right)\left(\frac{1}{P_C} - 1\right)\cdots \right. \\ &\qquad \left. \cdot \left(\frac{1}{P_S} - 1\right) \Big/ \left(\frac{1}{P_0} - 1\right)^{N-1}\right]^{-1}, \end{aligned} \tag{15}$$

where $P_0 = P(E_M)$, $P_A = P(E_M|A)$, $P_B = P(E_M|B), \ldots, P_S = P(E_M|S)$. Note that this formula can be written as a linear relation of logit functions of probabilities [see equation (23) in Section 3.4.2]. These probabilities for a short time interval become very small so that (15) can be approximated by

$$P(E_M|A, B, C, \ldots, S) \approx P_0 \frac{P_A}{P_0}\frac{P_B}{P_0}\frac{P_C}{P_0}\cdots\frac{P_S}{P_0}. \tag{16}$$

The above relation shows that for multiple independent precursors, the conditional rate of earthquake occurrence can be obtained by multiplying the unconditional rate P_0 with ratios of conditional probability to unconditional probability P_0. Each ratio is defined as the probability gain of a precursor. Utsu (1979) retrospectively reported a high probability forecast of the 1978 Izu–Oshima–Kinkai earthquake of M7.0 using the multiple independent precursor formula. This is based on each probability assessment of the anomalous phenomena consisting of uplift in the Izu Peninsula, swarm, and a composite of a radon anomaly, anomalous water table change and volumetric strain anomaly. Each of such probabilities was not very high. Aki (1981) summarized the Utsu report and further explained similar possible calculations for successful prediction of the 1975 Haicheng earthquake of M7.3 in China by considering long-term, intermediate-term, short-term and imminent precursory phenomena.

3.3 Improving Probability Gains by Seeking Statistically Significant Phenomenon

Here I would like to describe several point process models which can enhance the probability gains. To examine whether certain abnormal phenomena affect changes in the baseline rate of earthquake occurrences, Ogata and Akaike (1982), Ogata, Akaike and Katsura (1982) and Ogata and Katsura (1986) analyzed causal relationships between earthquake series from two different seismogenic regions, A and B. Let N_t^A and N_t^B be the number of earthquakes above a certain magnitude thereshold in the time interval $(0, t)$ in regions A and B, respectively. Then, consider a model of the intensity function of point process N_t^A for earthquake occurrences in the region A, conditional on the history of earthquake occurrences H_t^A and H_t^B in both regions:

$$\begin{aligned} \lambda_A(t|H_t^A, H_t^B) &= \mu + \sum_{j=1}^{J} a_j t^j + \int_0^t g(t-s)\, dN_s^A \\ &\quad + \int_0^t h(t-s)\, dN_s^B, \end{aligned} \tag{17}$$

where the first two terms on the right-hand side of the equation represent the Poisson process of a trend, the third term represents the cluster component within region A including aftershocks and swarms, and the last term represents the effect of earthquake occurrences in region B.

Here, it must be noted that even if a significant correlation is observed between the two series of events, it is insufficient from the standpoint of prediction, and it is necessary to identify causality. Thus, we must examine the opposite causality by interchanging A and B in equation (17). If both direction models hold, this process is mutually exciting (Hawkes, 1971). Furthermore, the correlation between A and B regions may be indirect such that activities in both regions may be affected by additional factors, for which the trend term may be useful if the polynomial can efficiently capture such an effect. According to our analysis of seismicity in two seismogenic regions along the subducting Pacific plate interface beneath the central Honshu, Japan, seismicity causality was found as a one-way effect from the deeper to the shallower. The maximum probability gain of the causal effect was several times the average occurrence rate.

Similarly, we can examine the causal relationship from some observed geophysical time series of ξ_s (Ogata and Akaike, 1982; Ogata, Akaike and Katsura, 1982) as follows:

$$\lambda_A(t|H_t^A, H_t^\xi) = \mu + \sum_{j=1}^{J} a_j t^j + \int_0^t g(t-s)\,dN_s^A + \int_0^t h(t-s) f(\xi_s)\,ds. \tag{18}$$

An example is the data of unusual intensities of ground electric potential, which were observed in the vicinity of Beijing, China, during a 16-year period beginning in 1982. The issue was whether or not these factors were useful as precursors to strong earthquakes of $M \geq 4.0$. Electricity anomalies could have been aftereffects of strong earthquakes. However, by comparing the goodness of fit of models (17) by AIC, anomalies were deemed statistically significant as precursors to earthquakes (Zhuang et al., 2005a). Moreover, the conditional intensity rate of declustered earthquakes $M \geq 4.0$ or larger within a radius of 300 km from the Huailai ground-electricity station was given by

$$\lambda(t|H_t) = \mu + \int_S^t h(t-s)\xi(s)^a\,ds = 0.00702 + \sum_{j=S}^{t} 0.000117 e^{-0.142(t-j)} \xi_j^{0.69} \tag{19}$$

(event/day) in the study of Ogata and Zhuang (2001), in which successively occurring $M \geq 4$ earthquakes within five days and 30 km distance were removed from the data to account for the self-exciting effect in equation (18). According to this model, the rate of $M \geq 4$ earthquakes varies from a half to 10 times the average occurrence of 0.0126 event/day.

Furthermore, the time series of electric anomaly records were available from three other stations near Beijing. If we assume that the four sets of the time series are approximately independent, we may consider the following conditional intensity rate by extending the multiple precursor in equation (16):

$$\lambda_A\left(t\Big|\bigcap_{m=1}^{4} H_t^m\right) \approx \hat{\lambda}_A \prod_{m=1}^{4} \frac{\lambda_{A_m}(t|H_t^m)}{\hat{\lambda}_{A_m}} \tag{20}$$

for the common region $A = \bigcap_{m=1}^{4} A_m$ among four circular regions A_m of 300 km radii from the four stations. Retrospective total probability gain varies in the range 1/10–100 times of the average occurrence rate λ_A in the common region (Ogata and Zhuang, 2001).

Here, we considered declustered earthquakes near Beijing, but if we can consider the original data and model that take the triggered clustering effect into account (cf. Zhuang et al., 2005a, 2013), the corresponding model would become

$$\lambda\left(t\Big|\bigcap_{m=1}^{4} H_t^m\right) \approx \lambda_0(t|H_t) \prod_{m=1}^{4} \frac{\lambda(t|H_t^m)}{\lambda_0(t|H_t)}. \tag{21}$$

A similar but more general model is applied in foreshock forecasting discussed in the following section (Section 3.4.2).

Moreover, we examined periodicity, or seasonality, of earthquake occurrences (Ogata and Katsura, 1986; Ma and Vere-Jones, 1997). Although such issues have been frequently discussed in statistical seismology, it was difficult to analyze correlations in the conventional method because the clustering feature of earthquakes frequently leads to incorrect results (Aki, 1956). On the other hand, we found it effective to apply statistical models of stochastic point processes that incorporate a clustering component; further details can be found in the study of Ogata (1999a), and the references therein. These models can also be applied to examine whether or not various geophysical anomalies are statistically significant as precursors of an approaching large earthquake:

$$\lambda_\theta(t|H_t) = \mu + \sum_{j=1}^{J} a_j t^j + \sum_{k=1}^{K} \left\{ c_{2k-1} \cos\frac{2\pi k t}{T_0} + c_{2k} \sin\frac{2\pi k t}{T_0} \right\} + \int_0^t g(t-s)\,dN_s. \tag{22}$$

From estimated amplitudes of the one-year periodic term with $T_0 = 365.24$ days, it is evident that probability gains vary around the average occurrence rate of corresponding $M \geq 4.0$ and $M \geq 5.0$ earthquakes. More extensive studies were reported by Matsumura (1986) by using the above model, in which the trend-term (first two terms) was used for artificial nonstationarity due to an increasing number of observed earthquakes in the long-term global catalog. He detected periodic effects for many mid-latitude seismic inland regions, whereas the seasonality was rarely observed in tropical seismic regions and ocean seismogenic zones. Correlations with precipitation variations were common in these results and are most probably due to pore fluid pressure changes in faults (see Section 4.1 for the

physical mechanism). An extension of the above periodicity model, reported by Iwata and Katao (2006), includes a combination of (lunar) synodic and semi-synodic periods to examine whether or not and how certain seismicity is affected by Earth tides. Statistical models, applications to validate data from the earthquake-induced phenomena and their references were reviewed by Ogata (1999a).

3.4 Probabilistic Identification of Foreshocks

The study of foreshocks should lead to a short-term forecasting. Although a considerable number of foreshocks are observed, most are recognized after occurrence of a large earthquake. Nevertheless, when earthquakes begin to occur in a local region, its residents should determine whether or not such movement is a precursor of a significantly larger earthquake. The probability of foreshock type can be determined statistically from the data of ongoing earthquakes in a particular region. Moreover, by using composite identification data of magnitude sequence and degree of hypocenter concentration, the probability gain of prediction is heightened.

3.4.1 *Working definitions for foreshock discrimination*. When an earthquake of M4.0 or larger occurs, it must first be determined whether or not the movement is a continuation of nearby earthquakes. Precisely, the connection to past earthquakes is determined by the single-link clustering (SLC) algorithm of Frohlich and Davis (1990).

The largest earthquake in a cluster is designated as the main shock. Pre-shocks refer to all earthquakes preceding the main shock of a cluster. All pre-shocks in a cluster become foreshocks when the magnitude gap or magnitude difference between the largest pre-shock and main shock is 0.45 or greater. If the magnitude gap is smaller than 0.45, the cluster with pre-shocks is defined as a swarm. An additional type of cluster is the main shock–aftershock type, in which the main shock occurs first in the cluster. A magnitude gap of 0.45 or larger between the main shock and largest pre-shock occurs in less than approximately 20% of pre-shock clusters in Japan. This 0.45 borderline of foreshock- and the swarm-types has been determined by a trade-off between achievement of a larger magnitude gap, which results in better discrimination of the foreshock, and a greater number of foreshock clusters, which results in better statistics. Here, to characterize these features, we note that clusters of foreshock-type exclude main shock and subsequent aftershocks, whereas other cluster types include all events in each cluster. This designation is made because real-time recognition of the main shock, which is preceded by foreshocks, is easy owing to the large magnitude gap, whereas main shocks of other cluster types are difficult to recognize until the end of the cluster.

3.4.2 *Probability forecast by discrimination of foreshocks*. Using the location (x, y) of the first earthquakes from clusters or isolated single earthquakes, by the empirical Bayesian logit model, Ogata, Utsu and Katsura (1996) obtained a map $\mu(x, y)$ of a probability that the earthquake will be a foreshock of a forthcoming main shock. Such probability varies from 1% to 10% with an average of 3.8% throughout Japan. Probability forecasts using this map have been conducted from January 1994 to April 2011, and their performances have been demonstrated by Ogata and Katsura (2012).

Multiple earthquakes occurring in a cluster provide more effective forecast updates, and certain statistics within the cluster are useful for discriminating foreshocks. Ogata, Utsu and Katsura (1996) revealed that distances between foreshocks in time and space are statistically shorter than those between earthquakes in clusters of other types. Moreover, increasing magnitudes enhance the probability of foreshocks. In the following model, we devise foreshock probability by using such statistics for prospective forecasting of main shocks.

Suppose that multiple earthquakes occur in a cluster c. Then, we consider time differences $t_{i,j} = t_j - t_i$ (days), epicenter separations

$$r_{i,j} = \sqrt{(x_j - x_i)^2 \cos^2 \theta_{i,j} + (y_j - y_i)^2}\ \text{km},$$

where θ_{ij} represents the mean latitude of earthquakes i and j, and magnitude differences $g_{ij} = M_j - M_i$ between earthquakes i and j $(i < j)$. On the basis of a comparative study of these statistics (Ogata, Utsu and Katsura, 1995, 1996), we standardized them into a unit interval. Specifically, time difference was transformed by $\tau = \log(100t)/\log(3000)$ for $0.1 \le t \le 30$ days; otherwise, 0 and 1 were set for $t \le 0.1$ and $t \le 30$, respectively. Epicenter separation was transformed by $\rho = 1 - \exp\{-\min\{(r, 50)/20\}$ km. Finally, magnitude difference was transformed by $\gamma = (2/3)\exp\{g/\sigma_1\}$ and $\gamma = 2/3 + (1/3)\{1 - \exp(g/\sigma_2)\}$ for $g \le 0$ and $g > 0$, respectively, where $\sigma_1 = 0.6709$ and $\sigma_2 = 0.4456$ (km).

Suppose that, at the current time, $c|n$ shows the stage where the nth earthquake $(n = 2, 3, 4, \ldots, \#c)$

has occurred in a cluster c, where $\#c$ is the number of all earthquakes in the cluster c. We propose the forecasting probability $p_{c|n}$ by using the following logistic transformation: Set $f = \operatorname{logit} p = (1-p)/p$, or $p = 1/(1+e^f)$; then,

$$\begin{aligned}\operatorname{logit} p_{c|n} &= \operatorname{logit} \mu(x_1, y_1) \\ &\quad + \frac{1}{\#(i < j \le n)} \\ &\quad \times \sum_{i<j\le n} \Bigg(\mu_0 + \sum_{k=1}^{3} b_k \gamma_{i,j}^k \\ &\qquad + \sum_{k=1}^{3} c_k \rho_{i,j}^k + \sum_{k=1}^{3} d_k \tau_{i,j}^k \Bigg).\end{aligned} \tag{23}$$

Here, the first term $\mu(x_1, y_1)$ indicates the probability that the first earthquake in the cluster is a foreshock, and the second term indicates the sample mean of weighted polynomials of transformed variables defined among all cluster members up to the time of forecasting. Here, the factor $\#(i < j \le n)$ is the number of pairs in the first n members of the cluster c.

It must be noted that interactions between the normalized statistics were not selected in equation (23) by the AIC comparison; namely, independency for the formula (15) is shown between the statistics of time intervals, epicenter separations and magnitude differences. Such conditions can be extended for a case in which the factors are dependent by considering higher order of polynomials; however, the linear factor in equation (23) represented the best fit in this case according to AIC (Ogata, Utsu and Katsura, 1996).

Probability forecasts that use prediction equation (23) have been conducted from January 1994 to April 2011, and their performances have been evaluated by Ogata (2011a) and Ogata and Katsura (2012). Therefore, these forecasts are expected to be applied for practical use in real time in the near future.

4. INCORPORATING PHYSICAL MECHANISMS OF EARTHQUAKES

4.1 Earthquake Dynamics and Interactions

The earth crust and upper mantle lithosphere can be approximately considered as an elastic body. These become distorted under stress, which increases steadily in a particular direction. Fault planes are cracks within the lithosphere or plate boundary interfaces. Earthquakes occur through distortion of subsurface rocks and both sides of the fault plane moving out of alignment. The earthquake location listed in hypocenter catalogs is location at which a fault displacement started, and earthquake magnitude represents eventual size of the displacement. Moreover, some catalogs record the orientations and slips of fault planes of relatively large earthquakes.

For each fault plane, stress tensor in lithosphere is decomposed into two perpendicular components. The shear stress acts in a direction parallel to fault shifting, and the normal stress acts perpendicular to the fault plane. The orientation of each fault plane determines shear and normal stress, which define Coulomb failure stress (CFS):

$$\begin{aligned}\text{CFS} &= (\text{Shear Stress}) - (\text{friction coefficient}) \\ &\quad \times (\text{Normal Stress} - \text{pore fluid pressure}).\end{aligned} \tag{24}$$

CFS increases at a constant rate over time. When CFS exceeds a particular threshold, the fault slips dramatically (an earthquake). Then, the stress reduces to a certain value and accumulates again over decades to result in large earthquakes on plate boundaries, and over thousands of years to result in slips on inland active faults (see Section 2.3).

When an earthquake occurs, the displacement of the source fault causes sudden Coulomb stress changes (ΔCFS) on the peripheral receiver faults. The ΔCFS of each receiver fault plane either decreases or increases depending on its orientations relative to the slip angles of the source fault. On the faults with increased ΔCFS, earthquakes occur earlier than expected, whereas on those with decreased ΔCFS, forthcoming earthquakes are delayed. When faults of similar orientations dominate a region, either seismic activation or quiescence is expected in the region.

In equation (24), the pore fluid pressure of the gap fault related to CFS is generally a constant. However, its changes may be important. For example, pressure changes in the fluid magma gap affect swarm activity in volcanic areas (Dieterich, Cayol and Okubo, 2000; Toda, Stein and Sagiya, 2002). In addition, earthquakes can be induced through increased pore fluid pressure in a fault system (Hainzl and Ogata, 2005; Terakawa, Hashimoto and Matsu'ura, 2013), which is occasionally due to heavy rainfall or shaking of the earth crust owing to propagated strong seismic waves; the latter causes dynamic triggering (Steacy, Gomberg and Cocco, 2005, and papers included in the same volume). Relevantly, the seasonal nature of seismicity or annual periodicity has been discussed in Section 3.3 [cf. equation (22)]. Moreover, the ETAS model applications to

seismicity changes that were induced by dynamic triggering or injection of water were reported (Lei et al., 2008, Lei, Xie and Fu, 2011).

4.2 Predicting Seismicity in the Peripheral Area by Abrupt Stress Changes

To explain earthquake induction or suppression of seismicity, it is useful to determine whether or not ΔCFS was due to a rapid faulting event that caused an earthquake. When a large earthquake occurs, low-frequency seismic waves and global positioning system (GPS) crustal displacement are observed. From such observations, source fault mechanisms of the earthquake can be solved, such as size, orientation and vector of the fault slip. Such source parameters are input into a computer program designed by Okada (1992) to calculate ΔCFS in a receiver fault system on the basis of source fault data. Thus, studies on induction of earthquakes, based on ΔCFS, have become popular. Special issue volumes on this subject have been edited by Harris (1998) and Steacy, Gomberg and Cocco (2005).

For example, Ogata (2004b) examined regional ΔCFS in southwestern Japan by analyzing M7.9 Tonankai and M8.1 Nankai earthquakes in 1944 and 1946, respectively. Conventionally, some seismic quiescence in this period was either considered as a genuine precursor or suspected as an artifact because of incomplete detection of earthquakes during the Second World War. Positive and negative ΔCFS correlated strongly with seismic activation and quiescence, respectively. In particular, this study classified seismicity anomalies into pre-seismic, coseismic and post-seismic before, during and after massive earthquakes, respectively. These scenarios may be helpful in interpreting seismicity in western Japan prior to occurrences of expected subsequent large earthquakes along the Nankai Trough.

4.3 Physical Implication of the ETAS Model and Seismicity

In general, interactions among earthquakes are fairly complex. Once an earthquake occurs in a particular location, CFS of the fault system adjacent to the source fault is considerably increased, and many earthquakes are induced. Traditionally, these earthquakes are called as aftershocks. Some of them are induced outside the aftershock region; these are also called as off-fault aftershocks, or aftershocks in a broad sense. Significant changes in stress result in many aftershocks; even small changes can induce aftershocks to some extent. Furthermore, any aftershock can change stress, too, causing their aftershocks. Because of such complex interactions among invisible fault segments in the crust, detailed calculations of such stress changes are difficult and impractical.

Therefore, statistical models designed to describe the actual macroscopic outcome of these stress interactions are required. For example, the ETAS model in equation (5), which consists of empirical laws of aftershocks, quantifies dynamic forecasting of induced effects. By fitting to the selected data from the catalog earthquake, this ETAS model determines the parameters by the maximum-likelihood method. Thus, prediction of earthquakes conforming to regional diversity is possible.

On the other hand, the friction law of Dieterich (1994), which was developed on the basis of rock fracture experiments with controlled stress, can be linked to statistical laws of earthquake occurrences. In particular, this law reproduces temporal and spatial distribution of the attenuation rate of aftershocks, such as that determined by the O–U law in equation (2). However, because of seismicity diversity, predictions adapting well to development of seismicity appear to be difficult.

Seismicity anomalies can hardly be detected by observing conventional plots of earthquake series because they show a complex generation due to successive occurrences of earthquakes or clustering. The clustering nature is also the main difficulty for traditional statistical test analysis. Complexity due to the clustering feature creates difficulties in revealing anomalies of seismicity caused by slight stress changes, hence, various anomalous signals are missed.

Therefore, some seismologists have devised various declustering methods that include only isolated and largest earthquakes in a clustering group or the main shock, and other earthquakes are excluded. On the basis of declustered data, statistical significance of seismic quiescence was tested against the Poisson process. Occasionally, however, analysis results depend on the choice of criteria of the adopted declustering algorithm (Van Stiphout, Zhuang and Marsan, 2012). Hence, results could be due to artificial treatment. In addition, declustering methods result in a significant loss of information because they discard a large amount of data from the original catalog.

The ETAS model uses original earthquake data without declustering. As mentioned in Section 2.2.3, the

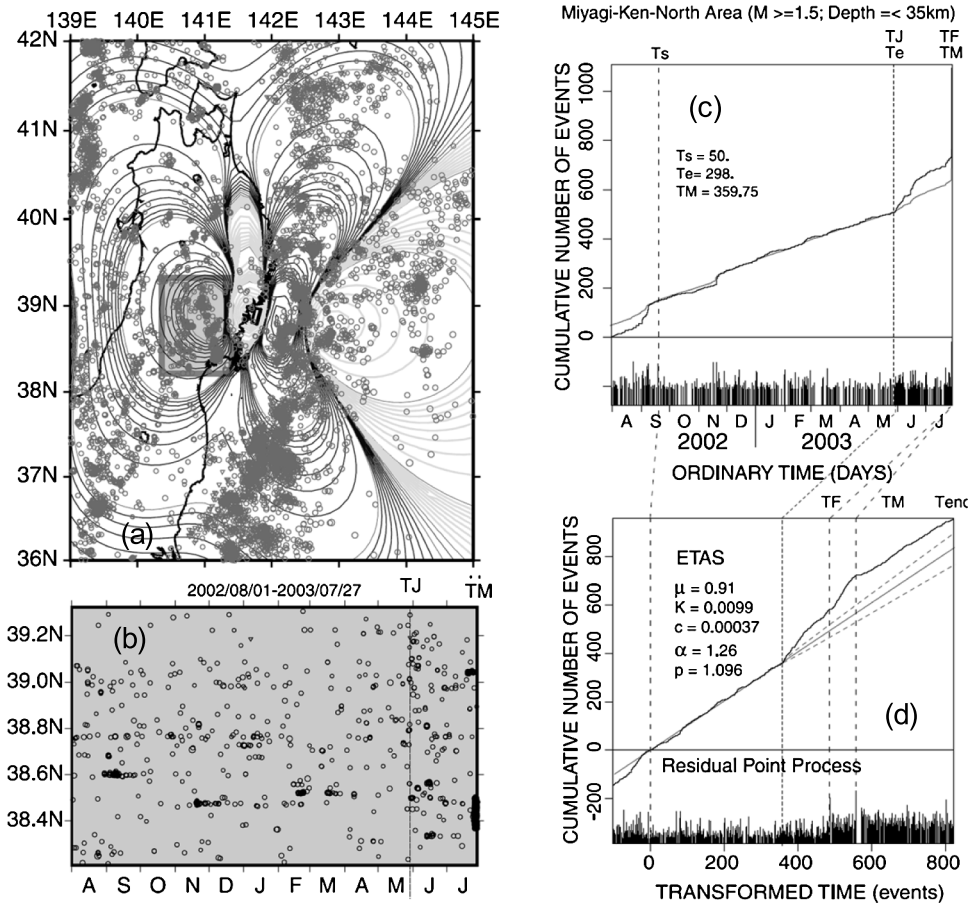

FIG. 4. *Shallow earthquakes in the inland rectangular region in* (a) *were analyzed from August* 2002–*July* 2003 *to investigate the effect of the M*7.0 *earthquake in* 2003. *CFS increments of this region took the largest values transferred from the M*7.0 *earthquake that is shown by the small rectangle fault located at* (141.7°E, 38.8°N) *at a depth of* 71 *km. Epicenters and latitude versus time are shown in* (a) *and* (b), *respectively. The occurrence time of the M*7.0 *earthquake is shown in each panel as the vertical dotted line indicated by TJ. The ETAS model was fitted to the data from the target period* (*TS*, *TJ*). *The panel* (c) *shows cumulative numbers and magnitude versus ordinary time*; *and the panel* (d) *shows these values against the transformed time determined in equation* (7) *by the estimated ETAS model. The black and grey cumulative function in panels* (c) *and* (d) *show the empirical cumulative function* (*step function*) *and the theoretical one* (*curve and straight line*) *estimated and then predicted by the ETAS model*, *respectively. At the southeastern corner of the inland rectangular region*, *a large M*6.2 *earthquake* (*occurrence time indicated by TM*) *and its largest M*5.5 *foreshock* (*indicated by TF*) *occurred on July* 26, 2003. *The panel* (d) *shows that the foreshock activity was more active than was expected*, *which is seen from the steepest slope of the cumulative function in* (d). *In contrast*, *the aftershock activity of the M*6.2 *earthquake*, *during the period TM-Tend*, *appears to be similar to predicted rates in* (d). *Dotted parabola-like envelope curves show twofold standard deviations* (95% *error bands*) *of cumulative numbers of the transformed time.*

ETAS model is a point process model configured to conform to empirical laws derived from various studies such as aftershocks in Japan and the time evolution of seismicity rate. Because regional characteristics of earthquake occurrences can be captured and considered as typical seismicity in this model, it has been accepted by seismologists as a standard model of ordinary seismicity. The ETAS model is used as a "barometer" for detecting significant deviations from normal activities as demonstrated in Figure 4 in Ogata (2005a).

5. SEISMICITY ANOMALIES FOR INTERMEDIATE-TERM FORECASTS

5.1 Seismicity Quiescence Relative to the ETAS Model

The deviation of actual cumulative number of earthquakes is measured relative to the theoretical cumulative function of the earthquake that serves as an indefinite integral in time for the predicted rate function (7) of the ETAS model. Relative quiescence occurs when actual earthquake occurrence rates are systematically lowered in comparison with the predicted inci-

dence that is determined by the ETAS model (Ogata, 1992). Relative quiescence lasting for many years was observed in a broad region before great earthquakes of M8 class and larger occurred in and near Japan (Ogata, 1992, 2006b). Similar phenomena were observed before M9-class large earthquakes in other regions of the world.

Since 2001, I have reported 25 agendas of various seismicity anomalies and forecasting proposals in Japan at the Coordinating Committee for Earthquake Prediction of Japan (CCEP). Except the agenda that reported seismic quiescence of aftershock activity before the largest aftershock (see Section 6.2 for detail), all were ex-post analysis report; the agendas were summarized in Ogata (2009). In addition, among 76 aftershock cases in Japan that I have investigated, relative quiescence was observed in 34 (see Ogata, 2001b, and its appendix for details of the case studies). Moreover, Section 5.4 includes a discussion on the manner in which results of this aftershock study will be used for space–time probability prediction of a neighboring large earthquake with a size similar to that of the main shock.

Here, I will note the results on the aftershock research of inland earthquakes of M6.0 or larger in southwestern Japan that occurred 30 years before and after the M8.1 Nankai earthquake in 1946. Among six earthquakes which occurred before 1946, relative quiescence was observed in five aftershock sequences. In contrast, among seven earthquakes after 1946, relative quiescence was not observed in six aftershock sequences, and these aftershock sequences were on track as expected. Since the ERC forecasts the next large earthquake for the next 30 years 60–70% (see Section 2.3, also see Ogata, 2001b and 2002), it would be worthy to monitor recent and future aftershock activity of similar large inland earthquakes.

5.2 Aseismic Slip, Stress Change and Seismicity Anomalies

Since a dense GPS observation network was established in Japan, aseismic fault motions or slow slips that could not be detected by seismometers have been successively identified in the plate boundary regions. We can take occurrence of such motion into account in discussing the relationship between seismicity anomalies (quiescence or activation) and stress changes.

Specifically, it can be assumed that slow slips on a focal fault or its adjacent part have occurred during a particular period. Then, depending on dominating orientations of receiver faults neighboring the focal fault, CFS could decrease or increase. Accordingly, we expect that seismicity there decrease or increase relative to the expected occurrence rate by the ETAS model. Such seismicity anomalies are revealed before some recent large earthquakes (Ogata, 2005b, 2007, 2010a, 2011c; and Kumazawa, Ogata and Toda, 2010). See Figure 5 for an example. By assuming slow slip on the source fault, the peripheral regions were classified as either of the increasing or decreasing CFS. Then, each region can theoretically correspond to an area that either promoted or suppressed seismicity. Such anomaly patterns of seismicity relative to the ETAS model (5) are in good agreement with those of CFS increment.

5.3 Variation of Local Stress Deduced from Spatio-Temporal Variation of Aftershocks

Local anomalies occur in space–time locations of most of the aftershocks. To elucidate these anomalies, we firstly apply the O–U formula (2) for aftershock decay to data of occurrence times and convert these times by the estimated theoretical cumulative function (7). We then examine whether space–time coordinates on a projected line, such as longitude and latitude, against the converted time remained temporally uniform or not. If nonuniformity in a certain portion of space–time conversion is observed, this implies discrepancies between theoretical and actual aftershock occurrences in such a place. Several possible scenarios for such discrepancies are offered: Secondary aftershocks that follow a large aftershock are obvious once seen as a cluster. Such a cluster shows traces of a new local rupture to extend the peripheral portion of the fault of the main shock. Moreover, when a nonuniform portion other than the secondary aftershocks is observed, it is crucial for us to explore the reasons.

Based on recent accurate aftershock location data, Ogata (2010b) revealed local relative quiescence and activations; these can occur associated with post- or pre-slips of a large aftershock. These anomalies were systematically investigated assuming that they were related to changes in the CFS rate. In addition, assuming several scenarios of stress changes due to slow slips, Ogata and Toda (2010) and Ogata (2010b) performed simulations to reproduce seismicity anomalies of relative activation and quiescence within aftershocks on the basis of the rate/state friction law of Dieterich (1994).

5.4 Space–Time Probability Gain of a Large Earthquake Under Relative Quiescence of Aftershocks

The probability of relative quiescence being precursor to a large earthquake must be evaluated with their

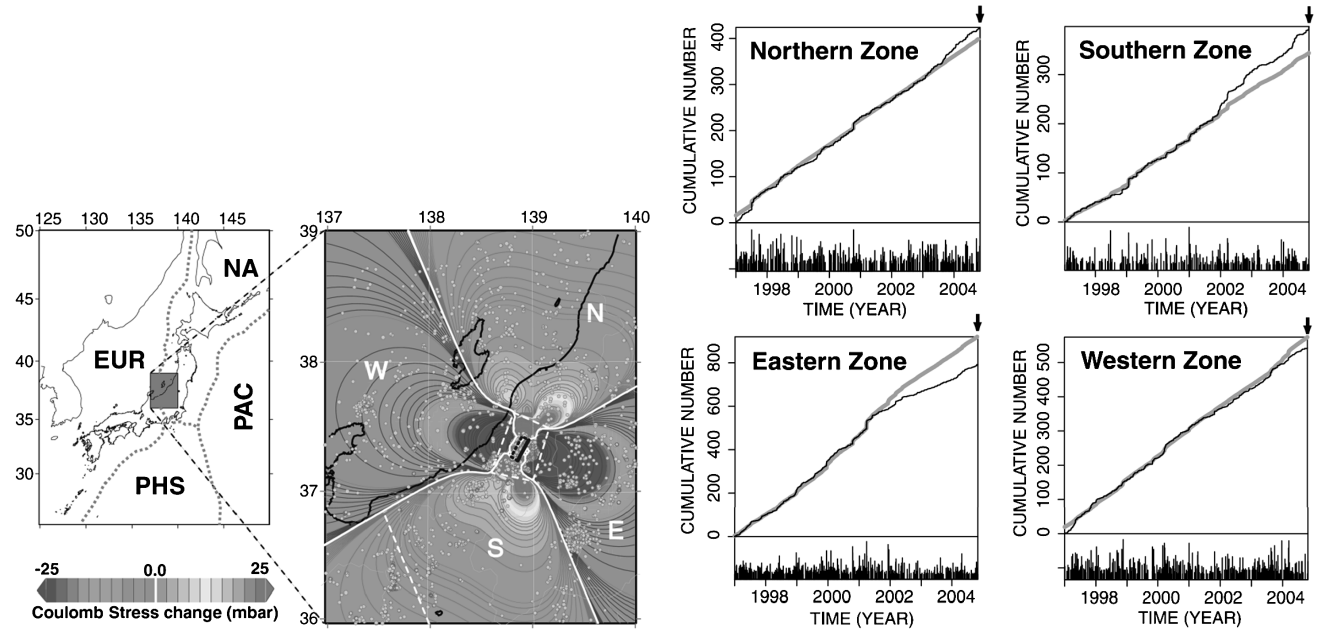

FIG. 5. (a) *Seismic activity for the eight years period before the M*6.8 *earthquake of* 2004. *Dominating fault slip orientations of earthquakes in this area are similar to those of the main shock and its aftershocks. The small black rectangle in the center of the magnified geographical map shows the main shock fault model determined by the GPS observations. The regions of thin and thick contours show positive and negative Coulomb failure stress* (*CFS*) *increments, respectively, assuming that slow slips in the source have occurred for some time. These define four subregions N, S, W and E used for the following ETAS analysis.* (b) *The four panels show the empirical cumulative curve* (*thin black*) *of the sequence of earthquakes of magnitude* 2 *or larger in each of the four divided subregions from* 1997 *until the M*6.8 *earthquake* (*downward arrows*). *Thick gray curves show estimated and predicted cumulative functions before and after each change-point time, respectively. Activation and quiescence relative to those predicted by the ETAS model agrees with increase and decrease in CFS, respectively.*

likely time and location. Because these involve many conditions and a number of parameters, they cannot be easily stated. However, by statistical studies of aftershock sequences in Japan (Ogata, 2001a), what I can say about a probability gain that a large earthquake will occur is as follows: First, if a large earthquake occurred in a particular location, the probability per unit area that another earthquake of similar magnitude will occur in the vicinity is greater than that which will occur in a distant area. This is the result of simple statistics regarding the self-similarity feature (inverse-power law correlations), and also physically suggests that the neighboring earthquake will be more probably induced by a sudden stress change on the periphery because of the abrupt slip of the earthquake. Moreover, if aftershock activity becomes relatively quiet, it becomes more likely that large aftershocks will occur around the boundary of the aftershock area. Furthermore, if relative quiescence lasts for a sufficiently long time more than a few months, the probability that another earthquake of similar magnitude will increase within six years in the vicinity of the aftershock area within 200 km distance.

6. SEISMICITY AND GEODETIC ANOMALIES

6.1 Aseismic Slip and Crustal Deformation

The Geological Survey Institute (GSI) of Japan compiles daily geodetic locations of global positioning system (GPS) stations throughout Japan, and baseline distances between GPS stations can be calculated from data in the GPS catalog. The geodetic time series show that contraction or extension of the distance between stations is basically linear with time because the subducting plate converges with constant speed. However, several years prior to large inland earthquakes of M7 class, the time series of the baseline distance variation around the fault was observed with systematic deviation from a linear trend (Ogata, 2007, 2010a, 2011c; Kumazawa, Ogata and Toda, 2010; GSI, 2009). Each deviation of these baselines was consistently explained by slow slips on the earthquake source fault or on its

down-dip extension. These results were due to post-hoc analysis based on knowledge of the source fault obtained by coseismic displacement.

From a predictive perspective, it is highly desirable to estimate such a fault slip in near real-time to that of occurrence. So far, several estimates of sufficiently large slips on plate boundaries have been obtained from GPS records by inversion analysis. GSI has regularly reported such estimates of coseismic, post-seismic and large-size habitual slips, at the CCEP meeting. However, it is difficult to obtain fine images of small slips, particularly in inland, even though inland GPS stations are arranged closely. This is attributed to high seismicity rather than GPS observation errors. Because strong earthquakes occur frequently, various effects of slow slips in GPS records are mixed up with such stronger changes. Hence, development of statistical models and methods to separate such signals is crucial. To estimate slow slips more precisely, combined modeling and analysis of seismicity and geodetic anomalies will be useful. Analyzing both seismicity and transient geodetic movements in a number of areas and locating the area of aseismic slip is very important for increasing the probability gain of a large earthquake.

6.2 Considering the Scenario of an Earthquake from Aseismic Slip

Observed anomalies of crustal movement and seismicity assume fault mechanisms and locations of slip precursors for prediction probability; therefore, their uncertainty must be estimated. In addition, probabilities of considered scenarios must be estimated. Such tasks are difficult. A possible method is to consider the logic tree of various scenarios regarding destruction of the fault system by attaching appropriate subjective or objective probabilities to tree components, as was performed for long-term predictions in California and Japan. Hence, such a scenario ensemble gives a forecast probability. Similarly, medium- and short-term prediction logic trees of various scenarios must be considered.

At the CCEP meeting on April 6, 2005, I reported relative quiescence of aftershocks of the Fukuoka–Oki earthquake (Ogata, 2005d). In addition, I examined potential slow slip areas on nearby active faults that may have created a stress shadow (region of decreased CFS) to cause relative quiescence in the aftershock sequence. Among these areas, the Kego fault, traversing Fukuoka City, had a large positive ΔCFS because of the main shock rupture, which showed evidence of possible slow-slip induction. Furthermore, the seismogenic zone along the Kego fault had already activated before the Fukuoka–Oki earthquake occurred (Ogata, 2010a). Therefore, because a slow-slip scenario on this fault was possible, I examined the pattern causing stress variation in the aftershock area. However, no stress shadow in the aftershock area was found. Therefore, I determined that probability of slow slip on the Kego fault was quite low. I also examined whether other possible slow slips in neighboring active faults could create a stress shadow that covered the aftershock region. However, no large earthquake has occurred in those faults thus far.

Approximately one month later, however, the largest aftershock occurred at the southeast end of the aftershock zone. Post-mortem examination based on information of the fault mechanism of this aftershock and detailed aftershock data revealed a detailed scenario. This means that by assuming a slow slip into the gap between the fault of the largest aftershock and main shock, relative quiescence of activity in the deeper part of the aftershock zone can be clearly explained (Ogata, 2006a). Moreover, the slip can explain relative quiescence in the induced swarm activity that occurred away from the aftershock area (Ogata, 2006a).

This setting as a prediction of future scenarios is much more vague and difficult to explain, even if it includes an ex-post scenario. Moreover, the time of occurrence must be predicted in addition to location, which is more difficult. Occurrence of slow slip does not always indicate a proximate precursor of fault corruption. Nevertheless, it is desired to keep observing GPS data to form scenarios of forthcoming large earthquake. For example, using Bayesian inversion by using GPS records, Hashimoto et al. (2009) estimated the locked zones on the plate boundary, where the next great earthquakes are expected.

7. CONCLUSIONS

To predict the future of a complex and diverse earthquake generation process, probability forecasting cannot be avoided. The likelihood (log-likelihood) is rational to measure the performance of the prediction. To provide a standard stochastic prediction of seismic activity in long term and short term, it is necessary to construct proper point process models and revise those that conform to each region. By the appearance of the anomaly, we need to evaluate the probability that it will be a precursor to a large earthquake. Namely, we need to forecast that the probability in a space–time zone

will increase to an extent, relative to those of the reference probability. For this, we make use of a point process model for the causality relationship.

It is desired to search any anomaly phenomena that enhance the probability gains. Having such anomalies, application of the multiple element prediction formula increases a precursory probability. A comprehensive physical study between precursory phenomena and earthquake mechanisms is essential for composing useful point process models. These elements must be incorporated to achieve predicted probability exceeding predictions of typical statistical models.

Furthermore, to determine urgency and uncertainty of major earthquakes against abnormal phenomena, numerous research examples must be accumulated. On the basis of these examples, possible prediction scenarios must be presented. Furthermore, to adapt well to diversity of earthquake generation, it is useful to adopt Bayesian predictions (Akaike, 1980b; Nomura et al., 2011) and consider region-specific models.

My experiences thus far confirm that the method of statistical science is essential to elucidate movement leading to prediction of a complex system. There is a need for development of a forecasting model that reflects diversity of the vast amount of information on seismicity and various covariate data. I believe that these will be developed by inventing an appropriate hierarchical Bayesian model. Space–time models for seismicity have become increasingly complicated (Ogata, 1998, 2004a, 2011b; Ogata, Katsura and Tanemura, 2003; Ogata and Zhuang, 2006).

A similar evolution is required for statistical models of geodetic GPS data.

ACKNOWLEDGMENTS

I am grateful to the Japan Meteorological Agency (JMA), the National Research Institute for Earth Science and Disaster Prevention, and universities for the providing hypocenter data. I am also grateful to the anonymous referee, Associate Editor and the Editor for their careful reviews and suggestions, which led to a significant revision of the present manuscript. This work was supported by JSPS KAKENHI Grant Number 23240039, and by the Aihara Innovative Mathematical Modelling Project, the "Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program)," initiated by the Council for Science and Technology Policy.

REFERENCES

AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Control*. **AC-19** 716–723. MR0423716

AKAIKE, H. (1980a). Likelihood and the Bayes procedure. In *Bayesian Statistics* (*Valencia*, 1979) (J. Bernard, M. De Groot, D. Lindley and A. Smith, eds.) 143–166. Univ. Press, Valencia. MR0638876

AKAIKE, H. (1980b). On the use of the predictive likelihood of a Gaussian model. *Ann. Inst. Statist. Math.* **32** 311–324. MR0609025

AKAIKE, H. (1985). Prediction and entropy. In *A Celebration of Statistics* (A. Atkinson and S. Fienberg, eds.) 1–24. Springer, New York. MR0816143

AKI, K. (1956). A review on statistical seismology. *Zisin II* (*J. Seismol. Soc. Japan*) **8** 205–228.

AKI, K. (1965). Maximum likelihood estimate of b in the formula $\log \mathrm{N} = a - bM$ and its confidence limits. *Bull. Earthq. Res. Inst.* **43** 237–238.

AKI, K. (1981). A probabilistic synthesis of precursory phenomena. In *Earthquake Prediction* (*Maurice Ewing Series*, 4) (D. Simpson and P. Richards, eds.) 566–574. American Geophysical Union, Washington, DC.

BANSAL, A. and OGATA, Y. (2013). Non-stationary epidemic type aftershock sequence model for seismicity prior to the 26 December 2004, M9.1 Sumatra-Andaman Islands mega-earthquake. *J. Geophys. Res.* **118** 1–14.

BOLTZMANN, L. (1878). Weitere Bemerkungen uber einige Plobleme der mechanischen Warmetheorie. *Wiener Berichte* **78** 7–46.

CENTER FOR ANALYSIS AND PREDICTION (1990–2003). Study on the seismic tendency in China (for the year 1989, . . . , 2002 and 2003) (in Chines). Beijing, Seismological Press.

COX, D. R. (1975). Partial likelihood. *Biometrika* **62** 269–276. MR0400509

DALEY, D. J. and VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes. Vol. I: Elementary Theory and Methods*, 2nd ed. Springer, New York. MR1950431

DIETERICH, J. (1994). A constitutive law for rate of earthquake production and its application to earthquake clustering. *J. Geophys. Res.* **99** 2601–2618.

DIETERICH, J., CAYOL, V. and OKUBO, P. G. (2000). The use of earthquake rate changes as a stress meter at Kilauea volcano. *Nature* **408** 457–460.

EARTHQUAKE RESEARCH COMMITTEE (1998). Regarding methods for evaluating probility of aftershocks. Available at http://www.jishin.go.jp/main/yoshin2/yoshin2.htm.

EARTHQUAKE RESEARCH COMMITTEE (2001). Regarding methods for evaluating long-term probility of earthquake occurrence. Available at http://www.jishin.go.jp/main/choukihyoka/01b/chouki020326.pdf.

FIELD, E. (2007). Overview of the working group for the development of regional earthquake likelihood models. *Seismology Research Letters* **78** 7–16.

FROHLICH, C. and DAVIS, S. D. (1990). Single-link cluster analysis as a method to evaluate spatial and temporal properties of earthquake catalogues. *Geophys. J. Int.* **100** 19–32.

GOOD, I. J. (1965). *The Estimation of Probabilities*. MIT Press, Cambridge, MA. MR0185724

GSI (2009). Crustal movements in the Tohoku district. *Rep. Coord. Comm. Earthq. Predict.* **83** 59–81.

GOOD, I. J. and GASKINS, R. A. (1971). Nonparametric roughness penalties for probability densities. *Biometrika* **58** 255–277. MR0319314

GUTENBERG, B. and RICHTER, C. (1944). Frequency of earthquakes in California. *Bull. Seismol. Soc. Amer.* **34** 185–188.

HAINZL, S. and OGATA, Y. (2005). Detecting fluid signals in seismicity data through statistical earthquake modeling. *J. Geophys. Res.* **110** B05S07.

HARRIS, R. A. (1998). Introduction to special section: Stress triggers, stress shadows, and implications for seismic hazard. *J. Geophys. Res.* **103** 24347–24358.

HARTE, D. and VERE-JONES, D. (2005). The entropy score and its uses in earthquake forecasting. *Pure Appl. Geophys.* **162** 1229–1253.

HASHIMOTO, C., NODA, A., SAGIYA, T. and MATSU'URA, M. (2009). Interplate seismogenic zones along the Kuril–Japan trench inferred from GPS data inversion. *Nature Geoscience* **2** 14–144.

HAWKES, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58** 83–90. MR0278410

IWATA, T. and KATAO, H. (2006). Correlation between the phase of the moon and the occurrences of microearthquakes in the Tamba region through point-process modeling. *Geophysical Research Letters* **33** L07302.

JMA (2008). Japan Meteorological Agency, Seismic activity in and around Kanto and Chubu Districts (May 2007–April 2008). *Rep. Coord. Comm. Earthq. Predict.* **80** 80–99.

JMA (2009). The Iwate-Miyagi Nairiku earthquake in 2008. *Rep. Coord. Comm. Earthq. Predict.* **81** 101–131.

JORDAN, T. H. (2006). Earthquake predictability, brick by brick. *Seismol. Res. Lett.* **77** 3–6.

JORDAN, T. H., CHEN, Y. T., GASPARINI, P., MADARIAGA, R., MAIN, I., MARZOCCHI, W., PAPADOPOULOS, G., SOBOLEV, G., YAMAOKA, K. and ZSCHAU, J. (2011). Operational earthquake forecasting. State of knowledge and guidelines for utilization. *Ann. Geophys.* DOI:10.4401/ag-5350. Available at http://www.annalsofgeophysics.eu/index.php/annals/article/view/5350/5371.

KAGAN, Y. and JACKSON, D. (1995). New seismic gap hypothesis–5 years after. *J. Geophys. Res. B: Solid Earth* **100** 3943–3959.

KEILIS-BOROK, V. and MALINOVSKAYA, L. (1964). One regularity in the occurrence of strong earthquakes. *J. Geophys. Res.* **70** 3019–3024.

KEILIS-BOROK, V., KNOPOFF, L., ROTWAIN, I. and ALLEN, C. (1988). Intermediate-term prediction of occurrence times of strong earthquakes. *Nature* **335** 690–694.

KUMAZAWA, T., OGATA, Y. and TODA, S. (2010). Precursory seismic anomalies and transient crustal deformation prior to the 2008 Mw = 6.9 Iwate-Miyagi Nairiku, Japan, earthquake. *J. Geophys. Res.* **115** B10312.

LEI, X., XIE, C. and FU, B. (2011). Remotely triggered seismicity in Yunnan, southwestern China, following the 2004 Mw9.3 Sumatra earthquake. *J. Geophys. Res.* **116** B08303.

LEI, X., YU, G., MA, S., WEN, X. and WANG, Q. (2008). Earthquakes induced by water injection at ~ 3 km depth within the Rongchang gas field, Chongqing, China. *J. Geophys. Res.* **113** B10310.

MA, L. and VERE-JONES, D. (1997). Application of M8 and Lin–Lin algorithms to New Zealand earthquake data. *New Zealand Journal of Geology and Geophysics* **40** 77–89.

MATSUMURA, K. (1986). On regional characteristics of seasonal variation of shallow earthquake activities in the World. *Bull. Disas. Prey. Res. Inst., Kyoto Univ.* **36** 43–98.

MATTHEWS, M. V., ELLSWORTH, W. and REASENBERG, P. A. (2002). A Brownian model for recurrent earthquakes. *Bull. Seismol. Soc. Amer.* **92** 2233–2250.

MOGI, K. (1967). Regional variation of aftershock activity. *Bull. Earthquake Res. Inst. Univ. Tokyo* **45** 125–173.

NANJO, K., TSURUOKA, H., HIRATA, N. and JORDAN, T. (2011). Overview of the first earthquake forecast testing experiment in Japan. *Earth Planets Space* **63** 159–169.

NANJO, K., TSURUOKA, H., YOKOI, S., OGATA, Y., FALCONE, G., HIRATA, N., ISHIGAKI, Y., JORDAN, T., KASAHARA, K., OBARA, K., SCHORLEMMER, D., SHIOMI, K. and ZHUANG, J. (2012). Predictability study on the aftershock sequence following the 2011 Tohoku-Oki, Japan, Earthquake: First results. *Geophysical Journal International* **191** 653–658.

NOMURA, S., OGATA, Y., KOMAKI, F. and TODA, S. (2011). Bayesian forecasting of the recurrent earthquakes and its predictive performance for a small sample size. *J. Geophys. Res.* **116** B04315.

OGATA, Y. (1983). Estimation of parameters in the modified Omori formula for aftershock frequencies by the maximum likelihood procedure. *J. Phys. Earth.* **31** 115–124.

OGATA, Y. (1986). Statistical models for earthquake occurrences and residual analysis for point processes. Mathematical Seismology (I) (M. Saito, ed.). Institute of Statistical Mathematics, Tokyo.

OGATA, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.* **83** 9–27.

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

OGATA, Y. (1992). Detection of precursory relative quiescence before great earthquakes through a statistical model. *J. Geophys. Res.* **97** 845–919.

OGATA, Y. (1995). Evaluation of probability forecasts of events; invited discussion as a commentary on "Forecasting Earthquakes and Earthquake Risk" by Prof. D. Vere-Jones. *Int. J. Forecasting* **11** 539–541.

OGATA, Y. (1998). Space–time point-process models for earthquake occurrences. *Ann. Inst. Statist. Math.* **50** 379–402.

OGATA, Y. (1999a). Seismicity analysis through point-process modeling: A review. *Pure Appl. Geophys.* **155** 471–507.

OGATA, Y. (1999b). Estimating the hazard of rupture using uncertain occurrence times of paleoearthquakes. *J. Geophys. Res.* **104** 17995–18014.

OGATA, Y. (2001a). Biases and uncertainties when estimating the hazard of the next Nankai earthquake. *Chigaku Zasshi* (*J. Geography*) **110** 602–614.

OGATA, Y. (2001b). Increased probability of large earthquakes near aftershock regions with relative quiescence. *J. Geophys. Res.* **106** 8729–8744.

OGATA, Y. (2002). Slip-size-dependent renewal processes and Bayesian inferences for uncertainties. *J. Geophys. Res.* **107** 2268.

OGATA, Y. (2004a). Space–time model for regional seismicity and detection of crustal stress changes. *J. Geophys. Res.* **109** B03308.

OGATA, Y. (2004b). Seismicity quiescence and activation in western Japan associated with the 1944 and 1946 great earthquakes near the Nankai trough. *J. Geophys. Res* **109** B04305.

OGATA, Y. (2005a). Detection of anomalous seismicity as a stress change sensor. *J. Geophys. Res.* **110** B05S06.

OGATA, Y. (2005b). Synchronous seismicity changes in and around the northern Japan preceding the 2003 Tokachi-oki earthquake of M8.0. *J. Geophys. Res.* **110** B08305.

OGATA, Y. (2005c). Simultaneous estimation of b-values and detection rates of earthquakes for the application to aftershock probability forecasting (in Japanese). *Rep. Coord. Comm. Earthq. Predict.* **73** 666–669.

OGATA, Y. (2005d). Seismicity changes in and around Kyushu District before the 2005 earthquake of M7.0 in the western offshore of Fukuoka Prefecture (in Japanese). *Rep. Coord. Comm. Earthq. Predict.* **74** 523–528.

OGATA, Y. (2006a). Monitoring of anomaly in the aftershock sequence of the 2005 earthquake of M7.0 off coast of the western Fukuoka, Japan, by the ETAS model. *Geophys. Res. Let.* **33** L01303.

OGATA, Y. (2006b). Seismicity anomaly scenario prior to the major recurrent earthquakes off the east coast of Miyagi Prefecture, northern Japan. *Tectonophysics* **424** 291–306.

OGATA, Y. (2007). Seismicity and geodetic anomalies in a wide preceding the Niigata-Ken-Chuetsu earthquake of 23 October 2004, central Japan. *J. Geophys. Res.* **112** B10301.

OGATA, Y. (2009). 40 Years Activities of the Coordinating Committee for Earthquake Prediction; (4) The Institute of Statistical mathematics, Research Organization of Information and Systems, Inter-University Research Inst Corporation, Edited by the Coordinating Committee for Earthquake Prediction, Geographical Institute of Japan, pp. 374. Available at http://www.ism.ac.jp/~ogata/yotiren/Yotiren2008ISM.pdf.

OGATA, Y. (2010a). Anomalies of seismic activity and transient crustal deformations preceding the 2005 M7.0 earthquake west of Fukuoka. *Pure and Applied Geophysics* **167** 1115–1127.

OGATA, Y. (2010b). Space–time heterogeneity in aftershock activity. *Geophys. J. Int.* **181** 1575–1592.

OGATA, Y. (2011a). Operational probability foreshock forecasts up until Tohoku-Oki earthquake (in Japanese). *Rep. Coord. Comm. Earthq. Predict.* **86** 123–125.

OGATA, Y. (2011b). Significant improvements of the space–time ETAS model for forecasting of accurate baseline seismicity. *Earth, Planets and Space* **63** 217–229.

OGATA, Y. (2011c). Pre-seismic anomalies in seismicity and crustal deformation: Case studies of the 2007 Noto Hanto earthquake of M6.9 and the 2007 Chuetsu-oki earthquake of M6.8 after the 2004 Chuetsu earthquake of M6.8. *Geophys. J. Int.* **186** 331–348.

OGATA, Y. and AKAIKE, H. (1982). On linear intensity models for mixed doubly stochastic Poisson and self-exciting point processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **44** 102–107. MR0655379

OGATA, Y., AKAIKE, H. and KATSURA, K. (1982). The application of linear intensity models to the investigation of causal relations between a point process and another stochastic process. *Ann. Inst. Statist. Math.* **34** 373–387.

OGATA, Y., IMOTO, M. and KATSURA, K. (1991). 3-D spatial variation of b-values of magnitude-frequency distribution beneath the Kanto District, Japan. *Geophys. J. Int.* **104** 135–146.

OGATA, Y., JONES, L. and TODA, S. (2003). When and where the aftershock activity was depressed: Contrasting decay patterns of the proximate large earthquakes in southern California. *J. Geophys. Res.* **108** 2318.

OGATA, Y. and KATSURA, K. (1986). Point-process models with linearly parametrized intensity for application to earthquake data. *J. Appl. Probab.* **23A** 291–310. MR0803179

OGATA, Y. and KATSURA, K. (1993). Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogues. *Geophys. J. Int.* **113** 727–738.

OGATA, Y., KATSURA, K. and TANEMURA, M. (2003). Modelling heterogeneous space–time occurrences of earthquakes and its residual analysis. *J. Roy. Statist. Soc. Ser. C* **52** 499–509. MR2012973

OGATA, Y. and KATSURA, K. (2006). Immediate and updated forecasting of aftershock hazard. *Geophys. Res. Lett.* **33** L10305.

OGATA, Y. and KATSURA, K. (2012). Prospective foreshock forecast experiment during the last 17 year. *Geophys. J. Int.* **191** 1237–1244.

OGATA, Y. and TODA, S. (2010). Bridging great earthquake doublets through silent slip: On- and off-fault aftershocks of the 2006 Kuril Island subduction earthquake toggled by a slow slip on the outer-rise normal fault the 2007 great earthquake. *J. Geophys. Res.* **115** B06318.

OGATA, Y., UTSU, T. and KATSURA, K. (1995). Statistical features of foreshocks in comparison with other earthquake clusters. *Geophys. J. Int.* **121** 233–254.

OGATA, Y., UTSU, T. and KATSURA, K. (1996). Statistical discrimination of foreshocks from other earthquake clusters. *Geophys. J. Int.* **127** 17–30.

OGATA, Y. and ZHUANG, J. (2001). Statistical examination of anomalies for the precursor to earthquakes, and the multi-element prediction formula: Hazard rate changes of strong earthquakes ($M \geqq 4$) around Beijing area based on the ultra-low frequency ground electric observation (1982–1997). *Rep. Coord. Comm. Earthq. Predict.* **66** 562–570.

OGATA, Y. and ZHUANG, J. (2006). Space–time ETAS model and an improved extension. *Tectonophysics* **413** 13–23.

OGATA, Y., KATSURA, K., FALCONE, G., NANJO, K. and ZHUANG, J. (2013). Comprehensive and topical evaluations of earthquake forecasts in terms of number, time, space, and magnitude. *Bull. Seismol. Soc. Amer.* **103** 3.

OKADA, Y. (1992). Internal deformation due to shear and tensile faults in a half-space. *Bull. Seismol. Soc. Amer.* **82** 1018–1040.

OMI, T., OGATA, Y., HIRATA, Y. and AIHARA, K. (2013). Forecasting large aftershocks within one day after the main shock. *Scientific Reports* **3** Article No. 2218.

OMORI, F. (1894). On the aftershocks of earthquake. *J. Coll. Sci. Imp. Univ. Tokyo.* **7** 111–200.

REASENBERG, P. and JONES, L. (1989). Earthquake hazard after a mainshock in California. *Science* **243** 1173–1176.

REASENBERG, P. and JONES, L. (1994). Earthquake aftershocks: Update. *Science* **265** 1251–1252.

REID, H. F. (1910). *The Mechanics of the Earthquake. The California Earthquake of April* 18, 1906, *Report of the State Investigation Commission* **2**. Carnegie Institution, Washington.

RHOADES, D. A., VAN DISSEN, R. J. and DOWRICK, D. J. (1994). On the handling of uncertainties in estimating the hazard of rupture on a fault segment. *J. Geophys. Res.* **99** 13701–13712.

RUNDLE, J., TIAMPO, K., KLEIN, W. and MARTINS, J. (2002). Self-organization in leaky threshold systems: The influence of near-mean field dynamics and its implications for earthquakes, neurobiology, and forecasting. *Proc. Natl. Acad. Sci. USA* **99** 2514–2521.

SCHORLEMMER, D., GESTENBERGER, M. C., WIEMER, S., JACKSON, D. and RHOADES, D. (2007). Earthquake likelihood model testing. *Seism. Res. Lett.* **78** 17–29.

SCHORLEMMER, D., ZECHAR, J., WERNER, M., FIELD, E., JACKSON, D., JORDAN, T. and RELM WORKING GROUP (2010). First results of the regional earthquake likelihood models experiment. *Pure Appl. Geophys.* **167** 859–876.

SHEBALIN, P., KELLIS-BOROK, V., GABRIELOV, A., ZALIAPIN, I. and TURCOTTE, D. (2006). Short-term earthquake prediction by reverse analysis of lithosphere dynamics. *Tectonophysics* **413** 63–75.

SHIMAZAKI, K. and NAKATA, T. (1980). Time predictable recurrence model for large earthquakes. *Geophys. Res. Lett.* **7** 279–282.

SOBOLEV, G. (2001). The examples of earthquake preparation in Kamchatka and Japan. *Tectonophysics* **338** 269–279.

STEACY, S., GOMBERG, J. and COCCO, M. (2005). Introduction to special section: Stress transfer, earthquake triggering, and time-dependent seismic hazard. *J. Geophys. Res.* **110** B05S01.

SUYEHIRO, S. (1966). Difference between aftershocks and foreshocks in the relationship of magnitude to frequency of occurrence for the great Chilean earthquake of 1960. *Bull. Seismol. Soc. Amer.* **56** 185–200.

TIAMPO, K., RUNDLE, J., MCGINNIS, S., GROSS, S. and KLEIN, W. (2002). Mean-field threshold systems and phase dynamics: An application to earthquake fault systems. *Europhys. Lett.* **60** 481–487.

TERAKAWA, T., HASHIMOTO, C. and MATSU'URA, M (2013). Changes in seismic activity following the 2011 Tohoku-oki earthquake: Effects of pore fluid pressure. *Earth Planet. Sci. Lett.* **365** 17–24.

TODA, S., STEIN, R. and SAGIYA, T. (2002). Evidence from the A.D. 2000 Izu Islands swarm that seismicity is governed by stressing rate. *Nature* **419** 58–61.

UTSU, T. (1961). A statistical study on the occurrence of aftershocks. *Geophys. Mag.* **30** 521–605.

UTSU, T. (1965). A method for determining the value of b in a formula $\log n = a - bM$ showing the magnitude-frequency' relation for earthquakes. *Geophys. Bull. Hokkaido Univ.* **13** 99–103.

UTSU, T. (1969). Aftershocks and earthquake statistics (1). *J. Fac. Sci. Hokkaido Univ.* **2** 129–195.

UTSU, T. (1970). Aftershocks and earthquake statistics (II)—Further investigation of aftershocks and other earthquake sequences based on a new classification of earthquake sequences. *J. Fac. Sci. Hokkaido Univ., Ser.* 7 **3** 197–266.

UTSU, T. (1977). Probability in earthquake prediction. *Zisin* (*J. Seismol. Soc. Japan*, 2*nd Ser.*) **30** 179–185.

UTSU, T. (1979). Calculation of the probability of success of an earthquake prediction (In the case of Izu-Oshima-Kinkai earthquake of 1978). *Rep. Coord. Comm. Earthq. Predict.* **21** 164–166.

UTSU, T. (1982). The probability in earthquake prediction (The second paper). *Bull. Earthq. Res. Inst., Univ. Tokyo* **57** 499–524.

UTSU, T., OGATA, Y. and MATSU'URA, R. (1995). The centenary of the Omori formula for a decay law of aftershock activity. *J. Phys. Earth* **43** 1–33.

VAN STIPHOUT, T., ZHUANG, J. and MARSAN, D. (2012). Declustering; The community online resource for statistical seismicity analysis (CORSSA). Available at http://www.corssa.org/articles/themev/van_stiphout_et_al/vanstiphoutetal2012.pdf.

VERE-JONES, D. (1999). Probabilities and in formation gain for earthquake forecasting. *Computational Seismology* **30** 248–263.

WIEMER, S. and WYSS, M. (1997). Mapping the frequency-magnitude distribution in asperities: An improved technique to calculate recurrence times? *J. Geophys. Res.* **102** 15.

WORKING GROUP ON CALIFORNIA EARTHQUAKE PROBABILITIES (2012). The uniform California earthquake rupture forecast, Version 3 (UCERF3) project plan. Available at http://www.wgcep.org/.

ZECHAR, J., GERSTENBERGER, M. and RHOADES, D. (2010). Likelihood-based tests for evaluating space–rate-magnitude earthquake forecasts. *Bull. Seismol. Soc. Amer.* **100** 1184–1195.

ZECHAR, J. and ZHUANG, J. (2010). Risk and return: Evaluating reverse tracing of precursors earthquake predictions. *Geophys. J. Int.* **182** 1319–1326.

ZHUANG, J. C. and JIANG, C. S. (2012). Evaluation of the prediction performance of the Annual Cosultation Meeting on Earthquake Tendency by using the gambling score. *Chinese Journal of Geophysics* (*in Chinese with English Abstract*) **55** 1695–1709.

ZHUANG, J., OGATA, Y. and VERE-JONES, D. (2002). Stochastic declustering of space–time earthquake occurrences. *J. Amer. Statist. Assoc.* **97** 369–380. MR1941459

ZHUANG, J. and OGATA, Y. (2011). Evaluation of warning forecasts by a gambling score. *Rep. Coord. Comm. Earthq. Predict.* **85** 451–452.

ZHUANG, J., VERE-JONES, D., GUAN, H., OGATA, Y. and MA, L. (2005a). Preliminary analysis of observations on the ultra-low frequency electric field in a region around Beijing. *Pure and Applied Geophysics* **162** 1367–1396.

ZHUANG, J., CHANG, C., OGATA, Y. and CHEN, Y. (2005b). A study on the background and clustering seismicity in the Taiwan region by using point process models. *J. Geophys. Res.* **110** B5.

ZHUANG, J., OGATA, Y., VERE-JONES, D., MA, L. and GUAN, H. (2013). Statistical modeling of earthquake occurrences based on external geophysical observations: With an illustrative application to the ultra-low frequency ground electric signals observed in the Beijing region. In *Imaging, Modeling and Assimilation in Seismology, Vol. II* (Y. Li, ed.) De Gruyter, Berlin.