

Statistics of Earthquake Activity: Models and Methods for Earthquake Predictability Studies

Yosihiko Ogata

Institute of Statistical Mathematics, Tokyo 190-8562, Japan; email: ogata@ism.ac.jp

Annu. Rev. Earth Planet. Sci. 2017. 45:497–527

First published as a Review in Advance on June 28, 2017

The *Annual Review of Earth and Planetary Sciences* is online at earth.annualreviews.org

<https://doi.org/10.1146/annurev-earth-063016-015918>

Copyright © 2017 by Annual Reviews.
All rights reserved



ANNUAL
REVIEWS **Further**

Click here to view this article's online features:

- Download figures as PPT slides
- Navigate linked references
- Download citations
- Explore related articles
- Search keywords

Keywords

earthquake predictability research, point process models, probability forecast, statistical seismology

Abstract

Statistical methods and various models in time-space-magnitude parameter space of earthquakes are being developed to analyze seismic activity based on earthquake hypocenter catalogs that are routinely accumulated. Considering complex geophysical environments and uncertainties, we seek proper stochastic modeling that depends on the history of earthquake occurrences and relevant geophysical information for describing and forecasting earthquake activity. Also, we need empirical Bayesian models with many parameters in order to describe nonstationary or nonhomogeneous seismic activity. This review is concerned with earthquake predictability research aimed at realizing practical operational forecasting. In particular, uncertainty lies in identifying whether abnormal phenomena are precursors to large earthquakes. The predictability of such models can be examined by certain statistical criteria.

1. INTRODUCTION

Conditional intensity function: used for short-term prediction of probability of an event occurrence under the condition of past history variables; in Equation 1 and elsewhere, the conditional variables are separated by a vertical bar (|) that can be read as “under the condition that”

Epidemic-type aftershock sequence (ETAS) model: model that uses aftershock productivity combined with other statistical relationships to forecast earthquake rate (see Sections 4.1–4.4); its parameters are entirely empirically determined from history of earthquake occurrence times associated with magnitudes

Akaike information criterion: a statistical indicator that compares the predictive performance of parametric models when the future data set is not available (see Section A.2 of Appendix A); a model with a smaller Akaike information criterion value is expected to be a better predictor

Statistical studies of earthquake occurrences have been carried out since the early years of seismology (e.g., Omori 1894). Although the causes of earthquakes were not clear until the last half century, statistical seismology (e.g., Aki 1956) had already played a major role in earthquake research by that time. In most statistical studies before 1970, the analysis of earthquakes was mainly based on the number of their occurrences or their interval lengths. However, after the 1970s, applications of stochastic point process theory to earthquake occurrences have drastically transformed the analysis of seismic activity. The point process is a mathematical abstraction of the stochastic occurrences of events. Among the concepts that characterize a point process, the conditional intensity function is important and represents the imminent likelihood of an event occurrence; that is, the conditional intensity function predicts a short-term occurrence rate that changes depending on the relevant information, including the history of past event occurrences (for a pioneering sketch of earthquake forecasting by point processes, see Vere-Jones 1978).

Earthquakes are represented by point events in a five-dimensional space-time-magnitude continuum. Thus, statistical models have been developed not only for time series of events but also for space-time data. Empirical Bayesian models with a large number of parameters are also rapidly being developed for location-dependent seismic activities.

In the following sections, such models and applications are reviewed, with an emphasis on the work of the author’s group. For example, epidemic-type aftershock sequence (ETAS) models are developed in terms of the conditional intensity functions built upon preceding empirical studies in statistical seismology, including the Omori–Utsu law for aftershock frequency (Omori 1894, Utsu 1961). These models can reproduce many of the statistical features observed in seismicity catalogs, and are used to construct forecasts that indicate how earthquake probabilities change over the short term. There are other types of point-process models for seismicity periodicities, causal relationships between earthquake occurrences, and the influence of exogenous geophysical data on focal earthquake occurrences.

These models are expected to be further developed to allow for practical probability forecasting of large earthquakes, and to discover transient geophysical phenomena such as changes in stress field, slow slip events, and the like. The methods of estimation, simulation, and diagnostic analysis can be applied straightforwardly to the data through the conditional intensity function. To examine the significance of such influences, the Akaike information criterion (Akaike 1974) and Akaike Bayesian information criterion (Akaike 1980) are useful (see the sidebar titled Glossary of Terms and Sections A.2 and A.3 of Appendix A).

I use general setups of space-time point processes to suggest that there is considerable room for developing new models of seismicity. One of the main applications of statistical seismology is implementing probability forecasts for large earthquakes in the presence of various anomaly data. Jordan et al. (2011) reviewed research on earthquake predictability to provide guidelines for the implementation of operational earthquake forecasting. I describe the prospects of relevant statistical studies to effectively predict earthquakes. The statistical methodologies and their principles are briefly summarized in Appendix A, and readers are referred to the **Supplemental Material** for citations of the traditional methods in statistical seismology and the references therein. The Community Online Resource for Statistical Seismicity Analysis (<http://www.corssa.org>) is also useful. In addition, Utsu (1999b) extensively reviews seismicity studies. The relevant statistical principles are summarized in an anthology of papers by Akaike (Parzen et al. 1998).

2. POINT PROCESS MODELING

Earthquake hypocenter catalogs include the five coordinates of earthquake-generating positions, namely, occurrence time, longitude and latitude of epicenter, focal depth, and magnitude. In this

GLOSSARY OF TERMS

Akaike Bayesian information criterion: a statistical indicator that compares the predictive performance of empirical Bayesian models when the future data set is not available (see Section A.3 of Appendix A)

Akaike information criterion: a statistical indicator that compares the predictive performance of parametric models when the future data set is not available (see Section A.2 of Appendix A); a model with a smaller Akaike information criterion value is expected to be a better predictor

Brownian passage time model: a stochastic model for rupture times on a recurrent earthquake source, adding Brownian perturbations to steady tectonic loading; rupture is assumed to occur when this process reaches a critical-failure threshold

Conditional intensity function: used for short-term prediction of probability of an event occurrence under the condition of past history variables; in Equation 1 and elsewhere, the conditional variables are separated by a vertical bar (|) that can be read as “under the condition that”

Epidemic-type aftershock sequence (ETAS) model: model that uses aftershock productivity combined with other statistical relationships to forecast earthquake rate (see Sections 4.1–4.4); its parameters are entirely empirically determined from history of earthquake occurrence times associated with magnitudes

Maximum a posteriori solution: parameters that maximize a posterior function that represents a trade-off between a likelihood function of parameters for the fit to data and prior probability for constraints among parameters; equivalent to the solution of parameters θ that minimize the penalized log-likelihood function in Equation 20 given the weights τ

Maximum likelihood estimate: the optimal estimate of parameters that characterize a set of stochastic models are obtained by maximizing their likelihood function that measures prediction performance (see Section A.2 of Appendix A)

Residual point process (RPP): a series of the transformed times from the occurrence time data using the integration of the conditional intensity in time; it becomes the standard stationary Poisson process if the conditional intensity of the integrand generated the data (see Sections A.4 and A.6 of Appendix A)

review, analysis of earthquake activity is discussed in relation to probability forecasting based on the four-dimensional time-space-size continuum (t_n, x_n, y_n, M_n) , $n = 1, 2, \dots, N$, ignoring the depth coordinate unless otherwise mentioned, due to low accuracy in comparison to the other coordinates. The hypocenters are regarded as samples from a stochastic point process (e.g., Daley & Vere-Jones 2003, Karr 1986).

Consider the probability of an earthquake occurring at a time t , a location (x, y) , and a magnitude M , conditional on the history of occurrence records $H_t = \{(t_j, x_j, y_j, M_j); t_j < t\}$ and relevant information F_t , including relevant exogenous records, such that

$$\frac{P \{ \text{an event in } [t, t + \Delta t] \times [x, x + \Delta x] \times [y, y + \Delta y] \times [M, M + \Delta M] \mid H_t, F_t \}}{\Delta t \Delta x \Delta y \Delta M} \\ \approx \lambda(t, x, y, M \mid H_t, F_t), \quad (1)$$

where the differential function, λ , is called the conditional intensity function representing the imminent likelihood of an event occurrence (see the sidebar titled Glossary of Terms). An illustrative use of the formula (Equation 1) is found in daily forecasting experiments of the Collaboratory for the Study of Earthquake Predictability (<http://www.cseptesting.org>), which aims to develop a scientific infrastructure of probability forecasts and their objective evaluations (Figure 1).

Akaike Bayesian information criterion: a statistical indicator that compares the predictive performance of empirical Bayesian models when the future data set is not available (see Section A.3 of Appendix A)

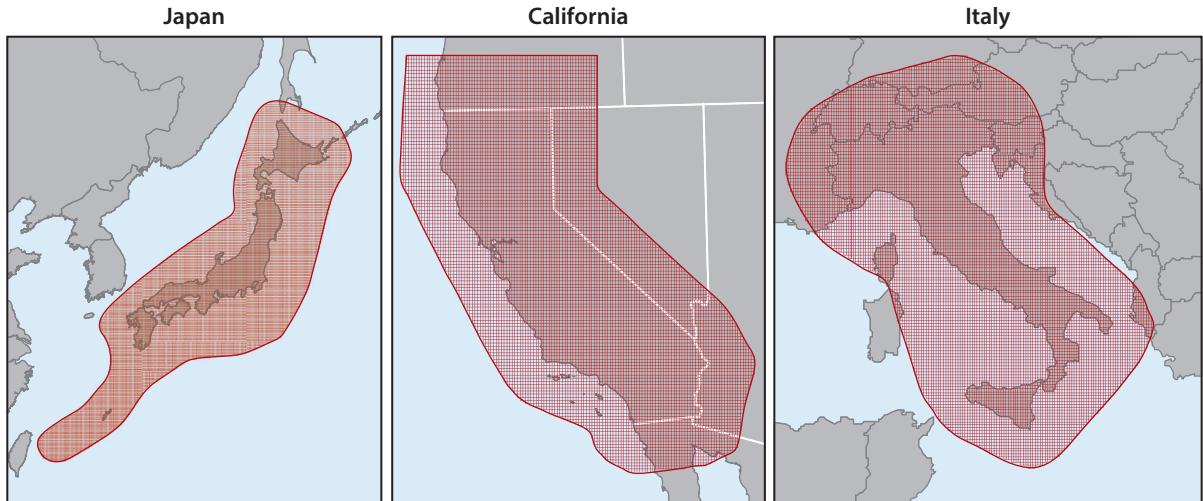


Figure 1

Collaboratory for the Study of Earthquake Predictability (CSEP) testing regions in Japan, California, and Italy. Daily forecasts of probabilities in the time-space-magnitude bins are submitted to the CSEP testing centers before the occurrence of earthquakes. According to the CSEP protocol, forecasters should provide conditional probabilities (see Equation 1) for each bin of $\Delta t \approx 1$ day, $\Delta x \Delta y \approx 0.1^\circ \times 0.1^\circ$ cell, and cells are divided into magnitude bins of $\Delta M \approx 0.1$ unit. The forecasters in the other submission categories can ignore the time axis and put a probability $\lambda(x, y, M) \Delta x \Delta y \Delta M$ in each space-magnitude bin for a given future time interval, such as 3 months, 1 year, and 5 years.

Hereafter, for simplicity, I assume that space-time locations and magnitudes of earthquakes are conditionally independent in such a way that

$$\lambda(t, x, y, M | H_t, F_t) = \lambda(t, x, y | H_t, F_t) \lambda(M | t, x, y, H_t, F_t), \quad (2)$$

which is referred to as the separability (Rathbun 1993, Schoenberg 2004). The modeling of non-separable seismicity phenomena will be a future task in predicting large earthquakes.

3. STATISTICS OF EARTHQUAKE MAGNITUDES

3.1. Magnitude Frequency Distributions

Generally, the theoretical earthquake size at time t and location (x, y) is defined in terms of a conditional intensity function as

$$\lambda(M | t, x, y, H_t, F_t) dM = P(M < \text{Magnitude} \leq M + dM | t, x, y, H_t, F_t), \quad (3)$$

which also depends on the history of occurrences and/or other relevant information. This approach is needed to model a probability forecast based on a potential precursory phenomenon, if it exists, as discussed in the following sections.

If we assume that the magnitude frequency is independent of the conditional variables in Equation 3 and has the parameters of the Gutenberg–Richter formula (Gutenberg & Richter 1944), then we have

$$\lambda(M) = 10^{a-bM} = Ae^{-\beta M}, \quad (4)$$

with constants a , b , and $\beta = b \ln 10$. Restricting the range of earthquake sizes as $M \geq M_c$, we can derive the probability density distribution $f(M) = \lambda(M)/\Lambda(M_c) = \beta e^{-\beta(M-M_c)}$, where $\Lambda(M_c) = \int_{M_c}^{\infty} \lambda(M)dM$ is the expected total number of earthquakes with $M \geq M_c$.

At first, Utsu (1965) derived the b -value estimator $\hat{\beta} = (\bar{M} - M_c)^{-1}$, or $\hat{b} = \log_{10}e/(\bar{M} - M_c)$, where \bar{M} is the mean magnitude of earthquakes with $M \geq M_c$, and Aki (1965) demonstrated that this is the maximum likelihood estimate obtained by maximizing the likelihood function

$$L(\beta) = \prod_{i=1}^n f_{\beta}(M_i) = \beta^n \prod_{i=1}^n \exp\{-\beta(M_i - M_c)\}, \quad (5)$$

which also provides the error estimate. Here, it should be noted that magnitudes in most earthquake catalogs are given in intervals of 0.1 (discrete magnitude values); hence care should be taken for avoiding bias in the b estimate. Therefore, the cutoff magnitude should be calibrated $M_c = M_{\min} - 0.5 \Delta M$, where ΔM is the discretization unit of the magnitude scale (Utsu 1965). Maximum likelihood estimates are obtained for many modified Gutenberg–Richter magnitude frequency distributions (see Utsu 1999a).

Maximum likelihood estimate: the optimal estimate of parameters that characterize a set of stochastic models are obtained by maximizing their likelihood function that measures prediction performance (see Section A.2 of Appendix A)

3.2. Location-Dependent b -Values

Here, consider that the β value in Equation 4 depends on location, such that $\beta(z)$ is a function of $z = (x, y)$ or $z = (x, y, b)$ and b represents depth. Because the maximum likelihood estimate of the b -value is given by the reciprocal of the magnitude average as stated above, local changes of the b -values are obtained by various kernel methods or moving weighted averages (e.g., Smith 1986, Wiemer & Wyss 1997, Wyss et al. 1997).

I consider that β in Equations 4 and 5 is represented by a flexible function of time and/or location using cubic B-spline expansions (Ogata & Katsura 1993, Ogata et al. 1991). Alternatively, in the case of highly concentrated earthquake clusters, I use a piecewise linear function defined on a space tessellated by triangles (Delaunay 1934) (**Figure 2**); coefficient values of the function are given at locations of the epicenters and additional points in the region (Ogata 2011b, Ogata et al. 2003b). Hence, the function is uniquely defined by linear interpolation of values at the three nearest points (earthquakes) determined by a Delaunay triangle.

Thus, the characterization of such flexible functions for β needs a set of high-dimensional coefficients θ . In such a case, a penalized log-likelihood (Good & Gaskins 1971),

$$Q(\theta|\tau) = \sum_{i=1}^n \ln \beta_{\theta}(z_i) e^{-\beta_{\theta}(z_i)(M_i - M_c)} - \text{penalty}(\theta|\tau), \quad (6)$$

describes the trade-off between the log-likelihood function that is an extension of Equation 5 for the goodness of fit to the data, and a penalty function that penalizes fluctuations (roughness) of the β function with constraining weights τ called hyperparameters in a Bayesian framework. Then, the best selection among the distinctive parameterizations for the penalty and tuning of hyperparameter values is objectively carried out as summarized in Section A.3 of Appendix A.

For example, Ogata et al. (1991) considered the magnitude frequency distribution beneath Kwanto, Japan, to a depth of 100 km using a function $\phi(x, y, b) = \ln \beta_{\theta}(x, y, b)$ of longitude x , latitude y , and depth b . Here, the penalty for the roughness of the β function is described by an integral of the sum of squares of the first and second partial derivative functions of ϕ , but the hyperparameters τ for constraining strengths are not always the same in the space axes, particularly in the horizontal and depth directions. Hence, the crucial point is the optimal determination of suitable τ for a given data set. A similar but important selection takes place in the case of space-time changes, replacing depth b by time t .

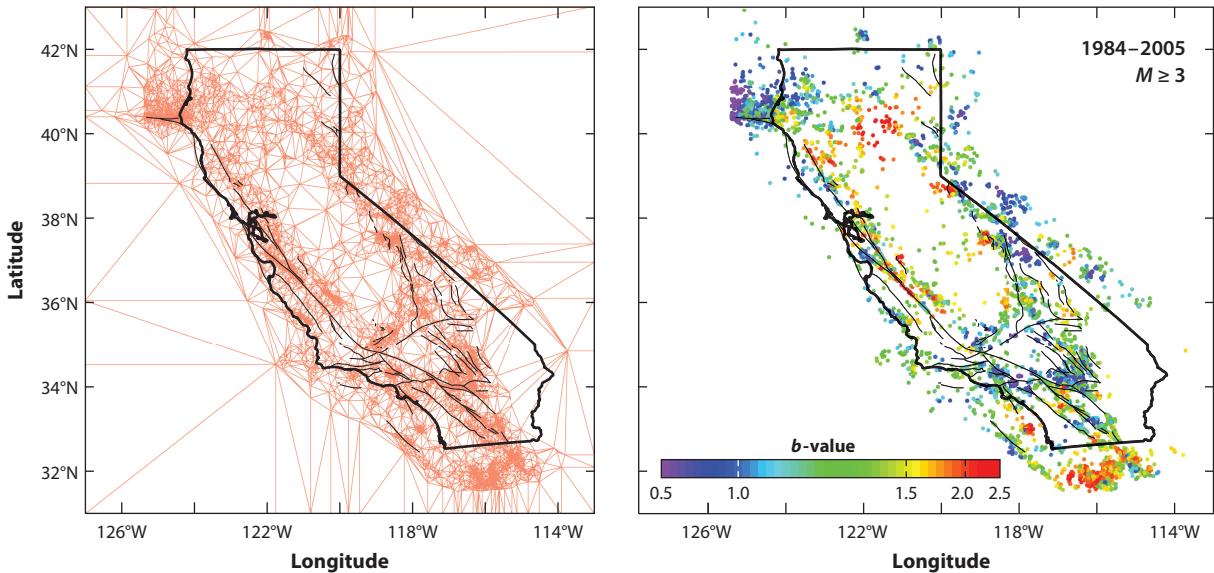


Figure 2

The two-dimensional, piecewise linear function on the Delaunay triangles for California seismicity, where earthquake data are taken from the Advanced National Seismic System catalog. The triangle vertices are earthquake locations, except those on the rectangle boundaries; colors represent the optimal maximum a posteriori (MAP) b -values at the vertices (see also Section A.3 of Appendix A).

3.3. Time- and History-Dependent Evolution of b -Values

Given a sequence of aftershocks $\{(t_i, M_i); i = 1, 2, \dots, N\}$ on the time interval $(0, T)$, the logarithm of $\beta(t)$ can be parameterized by flexible functions, such as a step function (Omi et al. 2013, 2014a,b), broken line function (Kumazawa & Ogata 2013; Kumazawa et al. 2016a,b), and cubic spline function (Ogata 1989). Then, consider a penalty function, replacing the variable z in Equation 6 by time t , that is, either the sum of the squares of the slopes of the piecewise line or the sum of the squares of the differences between consecutive function values. The optimal weights on the constraining strengths put on the penalties are objectively determined, and the goodness of fit of different models can also be compared, as explained in Section A.3 of Appendix A. In general, the optimal b -value changes in time, showing rapid changes in clustered periods, with otherwise slow changes (Ogata 1989).

It may be the case that a magnitude sequence is history dependent on past earthquakes or some other data of potential precursory phenomena as in Equation 3. In fact, empirical statistical analysis shows that magnitude sequences of even large earthquakes depend on long time periods (Ogata & Abe 1991). As an illustrative example, Ogata (1989) considered a statistical model of a history-dependent magnitude sequence in such a way that for $\sigma(t) = \beta(t)^{-1}$,

$$\sigma(\tau) = \sigma(\tau | H_\tau) = \mu + \sum_{\tau_i < \tau} g_J(\tau - \tau_i) + \sum_{\tau_i < \tau} h_K(\tau - \tau_i) \cdot M_i$$

for the transformed time τ (see Equation 24) by the epidemic-type aftershock sequence (ETAS) model (Equation 11), where $g_J(\cdot)$ and $h_K(\cdot)$ are parameterized by the Laguerre-type polynomials $g_J(t) = \sum_{j=1}^J a_j t^{j-1} e^{-ct}$ and $h_K(t) = \sum_{k=1}^K b_k t^{k-1} e^{-dt}$. Then, the optimal orders J and K with the maximum likelihood estimates of the coefficients are searched by minimizing the

Akaike information criterion (see Section A.2 of Appendix A), where the likelihood function is given by

$$L(\mu, a_1, \dots, a_J, c; b_1, \dots, b_K, d) = \prod_{i=1}^N \sigma(\tau_i)^{-1} e^{-(M_i - M_c)/\sigma(\tau_i)},$$

and M_c is the calibrated cutoff magnitude. The model with the minimum Akaike information criterion for Ogata's (1989) data showed a similar b -value evolution to that obtained by the empirical Bayesian method explained above; here, we have to be careful that seemingly low b -values may be due to smaller aftershocks missing from the catalog (i.e., being poorly recorded) immediately after the mainshock. We need to apply such analysis in conjunction with models for incompletely detected earthquakes (Ogata & Katsura 2006; Omi et al. 2013, 2014a,b).

Regarding history-dependent magnitude, Rhoades & Evison (2004) proposed a forecasting model for a strong earthquake based on the precursory swarm hypothesis (Evison 1977, Sekiya 1976); Smyth & Mori (2011), a medium-term forecast using an autoregressive b -value sequence; and Ogata & Katsura (2014), magnitude forecasts based on space-time-magnitude features of earthquake clusters as discussed in the next section.

Physically, a temporal b -value change in the earthquake sequence is interpreted as being due to the material experiencing a stress change (Scholz 1968). Statistically, a decrease in b -values implies a higher probability of a large earthquake, as shown in b -value studies of foreshock sequences reported by Suyehiro (1966). However, we must be careful as to whether the b -value change is due to migration of earthquake activity in the region in which the b -values are significantly location dependent, which can be explained by heterogeneity in strength of the material (Mogi 1962).

3.4. Foreshocks

Unlike aftershocks, foreshocks in an individual earthquake cluster are usually rare or small in number, and the pattern is varied even when many foreshocks occur (e.g., Mogi 1963). However, we can see some robust features by data stacking, where each mainshock occurrence time is taken to be the origin of the time coordinate. Their frequencies versus reverse time from the mainshock obey inverse power decays (e.g., Agnew & Jones 1991, Jones & Molnar 1979, Ogata et al. 1995, Papazachos 1974, von Seggern et al. 1981). There are some reports that individual or stacked foreshocks converge to the mainshock in distance (Ogata et al. 1995, Smith 1986, Wong & Wyss 1985). ETAS models (see Section 4) can reproduce these statistical aspects of retrospectively identified foreshock sequences (Helmstetter & Sornette 2003, Ogata & Katsura 2014).

However, foreshocks are quite uncertain to be recognized a priori. Utsu (1978) searched for magnitude patterns that would provide the highest foreshock probability. Studies following a priori foreshocks include those by Yamashina (1981a,b), Jones (1985, 1994), Agnew & Jones (1991), Maeda (1993), Savage & de Polo (1993), and Console et al. (1993). However, the working definitions are different among such foreshock studies. Hence, we have to take substantial care in comparing the results owing to varying definitions and algorithms for foreshock probability.

Ogata et al. (1996) implemented a grouping of earthquakes of $M \geq 4$ such that each cluster contains all connected earthquakes within a threshold distance (the single-link method; Frohlich & Davis 1990). They then defined a cluster as a foreshock-mainshock type if the largest earthquake called a mainshock is preceded by a significantly smaller earthquake with a magnitude difference 0.45 or more. At first, the location-dependent foreshock probability of the first earthquake in the cluster (that may eventually be an isolated earthquake) is determined for each region in Japan, followed by forecasts of the foreshock probability p_c of the cluster c ; this consists of variables of time intervals t , distances r , and magnitude differences g between the early earthquakes before the

mainshock in the cluster. Specifically, such a set of variables t , r , and g is nonlinearly converted to τ , ρ , and γ , uniformly distributed in the unit cube $[0, 1]^3$, and the negative logit function of p_c , $\text{Logit}(p_c) = \log\{(1-p_c)/p_c\}$ (cf. Section A.7 of Appendix A), is represented by

$$\text{Logit}(p_c) = v(x, y) + E[f(\tau, p, \gamma)], \text{ where } f(\tau, p, \gamma) = \sum_{l=0}^3 b_l \tau^l + \sum_{m=0}^3 c_m \rho^m + \sum_{n=0}^3 d_n \gamma^n,$$

where $v(x, y)$ is the foreshock probability determined only by the position (x, y) of the first earthquake in a cluster, $E[\cdot]$ indicates a certain average, and the maximum likelihood estimate is obtained using the 1926–1993 $M \geq 4.0$ data of the Japan Meteorological Agency data. Ogata et al. (1996) also applied general multivariate polynomials for $f(\tau, p, \gamma)$, but the above simple form is selected according to the Akaike information criterion (see Section A.2 of Appendix A). Hence, the variables (τ, p, γ) appear to be independent of each other, and p_c is simply the multielement probability formula with respect to (τ, p, γ) (see Section A.7 of Appendix A). In general, a flexible function such as a three-dimensional spline function or a Delaunay-based, three-dimensional piecewise linear function $f_\theta(\tau, p, \gamma)$ can be used to minimize the Akaike Bayesian information criterion (see Section A.3 of Appendix A; Nomura 2015; Nomura & Ogata 2015a,b).

Furthermore, this foreshock forecasting experiment (see Section A.2 of Appendix A; Ogata & Katsura 2012) was evaluated using data in the period from 1994 until March 10, 2011. The probabilities that used only the information of the location of the first earthquake vary in the range of approximately 1–10% per month throughout Japan. The use of the above predictor p_c for clusters with multiple earthquakes yields predicted probabilities that span the range from 0% to 30%. We confirmed that the forecasts were performed appropriately by testing the frequencies of actual foreshock-type clusters against the predicted probabilities. Moreover, the log-likelihood ratio (relative entropy) score in Equation 16 indicated a significantly better forecast than the average foreshock probability (7.9%). Incidentally, under the restriction that the mainshock is no less than M6.5, the predicted discrimination between the foreshock type and the others appears more clearly.

3.5. Earthquake Detection Rates in Time and Space

Most of the long-term earthquake catalogs are heterogeneous in terms of time and space in that the detection rate of smaller earthquakes is dependent on time and location due to the configuration of seismic networks changing in time. We want to separate the real seismicity from such variations in detection rates. The traditional method that restricts earthquakes above a completely detected magnitude loses a large amount of data.

The magnitude frequency of all detected earthquakes can be considered using the detection rate $q(M)$ of earthquakes such that $0 \leq q(M) \leq 1$. An example of the detection rate function is the cumulative function of the normal distribution,

$$q(M | \mu, \sigma) = (\sqrt{2\pi}\sigma)^{-1} \int_{-\infty}^M \exp(-(x - \mu)^2 / 2\sigma^2) dx, \quad (7)$$

suggested by Ringdal (1975). Here, the parameter μ represents the magnitude value at which 50% of earthquakes are detected, and σ relates to a range of magnitudes at which earthquakes are partially detected. For all observed magnitude frequencies, Ogata (1991) and Ogata & Katsura (1993) consider an intensity $\lambda(M | b, \mu, \sigma) = \lambda(M | b)q(M | \mu, \sigma)$ of magnitudes to obtain the maximum likelihood estimate $(\hat{\beta}, \hat{\mu}, \hat{\sigma})$.

To examine spatial inhomogeneity of the frequency distributions, we assume that the parameters (β, μ, σ) are parameterized by flexible functions of the location (x, y) , such as cubic-spline or piecewise-linear surfaces in a Delaunay tessellated space (Ogata 2011b, Ogata & Katsura 1993, Ogata et al. 1991). Then, given a data set of earthquake epicenters and magnitudes (x_i, y_i, M_i) , we

consider the log-likelihood with a penalty against the roughness of $\ln \beta$, $\ln \mu$, and $\ln \sigma$, respectively. The optimal determination of the hyperparameters for the penalties is therefore crucial to the method as discussed in Section A.3 of Appendix A.

Similarly, the temporal changes of b -value and detection rates in a narrow region can be modeled, expressing the parameters $b(t)$, $\mu(t)$, and $\sigma(t)$ by using the same flexible functions of time as those cited in Section 3.3. In particular, a temporal change analysis is indispensable for the aftershock forecast in the several hours immediately after a mainshock and its large aftershocks (see Section 4.3).

3.6. Bias Visualization and Corrections

There are two kinds of errors in estimating earthquake magnitude, random errors and systematic errors. The random errors can be reduced by increasing the amount of data, but the systematic errors are a serious subject of research. Utsu (2002) investigated the systematic differences in magnitude scales between different catalogs, for example.

It is useful to compare the magnitude of the same earthquake in other catalogs to examine whether transient magnitude changes exist in a catalog. Consider the time-dependent systematic difference $\psi(t, M^{(1)})$ in the magnitude $M^{(1)}$ of an earthquake in catalog 1 compared to its magnitude $M^{(2)}$ in catalog 2. Then, consider the following penalized sum of squares,

$$\sum_{i=1}^N \left\{ M_i^{(2)} - M_i^{(1)} - \psi(t, M_i^{(1)}) \right\}^2 + \text{penalty}(\theta|\tau),$$

which relates to the penalized log-likelihood function of the normal distribution that is objectively optimized by the Bayesian method described in Section A.3 of Appendix A.

For example, Ogata (1998a) applied this procedure to examine the differences between the old Japan Meteorological Agency magnitudes and body-wave magnitudes of the United States Geological Society (USGS) Preliminary Determination of Epicenters catalog in and around Japan during the period of 1963–1989. The optimally estimated function $\hat{\psi}$ for the bias indicates that the old Japan Meteorological Agency magnitudes below 5.0 are substantially underestimated in the period after 1976, which was due to application of the traditional magnitude calibration to velocity-type seismographs. This transient magnitude change (magnitude shift) affected the analysis of seismic quiescence. Also, transient magnitude shifts of the International Seismological Centre catalog provided different results on seismicity from those using data from the National Earthquake Information Center catalog (Bansal & Ogata 2013).

Similar analysis is applied for real-time correction of the epicenter location. For example, Ogata et al. (1998) utilized this model to calibrate the global earthquake location estimates from a small seismic array at the Japan Meteorological Agency Matsushiro Seismological Observatory, and this has been adopted for operational determinations for the Matsushiro catalog.

4. STATISTICS OF EARTHQUAKE OCCURRENCES

4.1. Space-Time Epidemic-Type Aftershock Sequence Models

The term epidemic in the ETAS model comes from quantifying births and deaths (Hawkes & Oakes 1974, Kendall 1949) in epidemiology. Hawkes (1971) modeled the birth process, allowing immigration at a constant rate μ per unit time, in terms of the conditional intensity rate

$$\lambda(t|H_t) = \mu + \sum_{\{i: t_i < t\}} g(t - t_i)$$

depending on the history of occurrence times t_i . This is naturally extended to the space-time model

$$\lambda(t, x, y | H_t) = \mu(x, y) + \sum_{\{j : t_j < t\}} g(t - t_j, x - x_j, y - y_j | M_j)$$

for a time period $[0, T]$ and locations (x, y) in a region A ; note here that the magnitude series $M_j \geq M_c$, $j = 1, 2, \dots$, consists of exogenous variables. The term $\mu(x, y)$ represents background seismicity rates, and $g(\cdot)$ is the impulse response function of triggered earthquakes, as used by Kagan (1991), Musmeci & Vere-Jones (1992), and Rathbun (1993).

Candidates for the space-time ETAS models (Ogata 1998b, Ogata & Zhuang 2006) were considered based on aftershock studies (Utsu 1969, 1971) and remote triggering phenomena, by way of the following impulse response functions:

$$g(t - t_j, x - x_j, y - y_j | M_j) = \frac{K_0}{(t - t_j + c)^p} \left[\frac{(x - \bar{x}_j, y - \bar{y}_j) S_j^{-1}(x - \bar{x}_j, y - \bar{y}_j)^t}{e^{\alpha(M_j - M_c)}} + d \right]^{-q}, \quad (8)$$

$$g(t - t_j, x - x_j, y - y_j | M_j) = \frac{K_0 e^{\alpha(M_j - M_c)}}{(t - t_j + c)^p} [(x - \bar{x}_j, y - \bar{y}_j) S_j^{-1}(x - \bar{x}_j, y - \bar{y}_j)^t + d]^{-q}, \quad (9)$$

$$g(t - t_j, x - x_j, y - y_j | M_j) = \frac{K_0 e^{\alpha(M_j - M_c)}}{(t - t_j + c)^p} \left[\frac{(x - \bar{x}_j, y - \bar{y}_j) S_j^{-1}(x - \bar{x}_j, y - \bar{y}_j)^t}{10^{q(M_j - M_c)}} + d \right]^{-q}, \quad (10)$$

where S_j is either a unit matrix or a covariance matrix, and (\bar{x}_j, \bar{y}_j) are the centroid epicenter coordinates of a cluster replaced by the original epicenter coordinates (x_j, y_j) in a catalog. A practical estimation procedure of the covariance matrix and centroid epicenter coordinates, using log-likelihood (Equation 21) and the Akaike information criterion (see the sidebar titled Glossary of Terms and Section A.2 of Appendix A), is given by Ogata (1998b, 2011b).

4.2. Hierarchical Space-Time Epidemic-Type Aftershock Sequence Model

When a seismogenic region becomes wide or the number of earthquakes large, spatial heterogeneity of seismicity becomes conspicuous. The aftershock productivity K_0 may differ significantly at different locations, even if the triggering earthquakes are of the same magnitude. Indeed, this has been the main difficulty in real-time aftershock forecasting. Moreover, mainshock-aftershock and swarm activities exhibit significantly different patterns, which can be related to location-dependent α -values.

Therefore, we develop a hierarchical space-time ETAS model in which the parameter values K_0 , α , p , and q can also be location dependent, such that each parameter function is defined by a linear interpolation of values at the three nearest earthquakes or additional sampling points determined by the Delaunay tessellation (cf. Section 3.2).

For a stable optimal estimation, we need to put penalties on the gradients of the piecewise linear function so as to minimize its roughness. The coefficients that maximize the penalized log-likelihood are then sought by an empirical Bayesian method (see Section A.3 of Appendix A). Further details can be found in work by Ogata et al. (2003b) and Ogata (2004, 2011b). In particular, background seismic activities $\mu(x, y)$ of the model in Section 4.1 using Equation 8 appear useful for long-term prediction of large earthquakes (**Figure 3**).

It is worth examining whether forecasting performance of the model in Equation 2 by $\lambda_{\text{HIST-ETAS}}(t, x, y | H_t) g(M | x, y)$ (Ogata 2011b) that was submitted to the Collaboratory for the Study of Earthquake Predictability can be improved if $g(M | x, y)$ is suitably replaced by

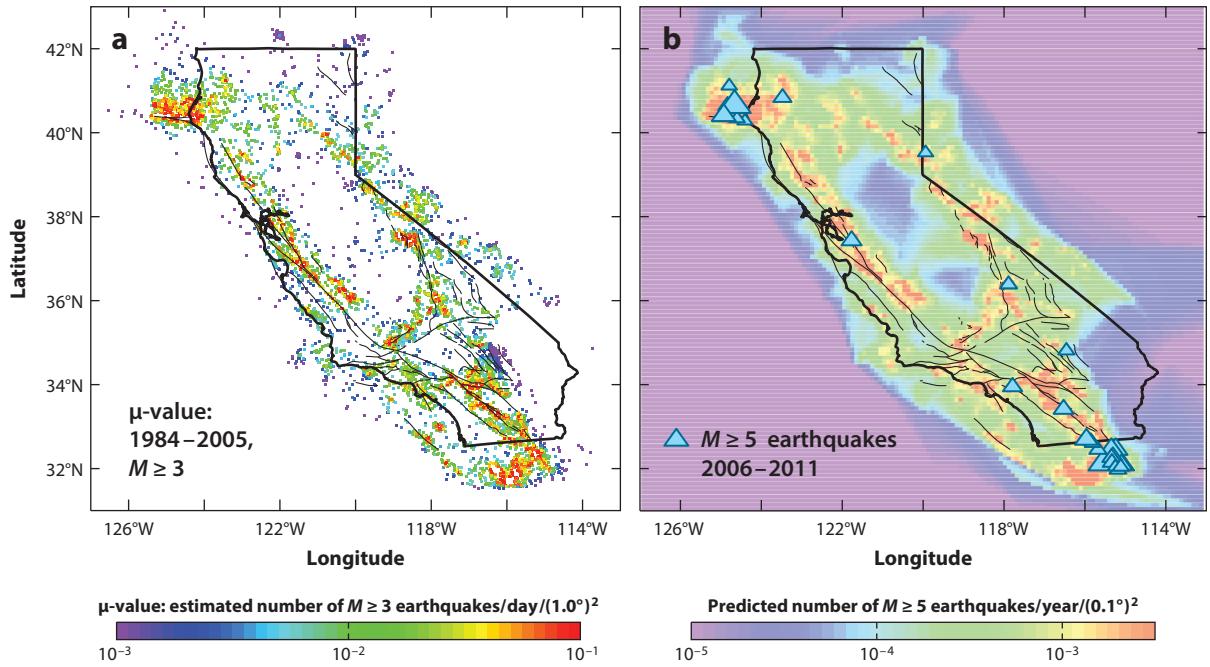


Figure 3

The (a) background seismicity μ -values and (b) forecasted spontaneous $M \geq 5$ earthquakes for the next 5 years in California. The hierarchical space-time-epidemic-type aftershock sequence model is fitted to Advanced National Seismic System data of earthquakes of $M \geq 3$ in the period of 1984–2005. The colored interpolated rates in panel b are calculated using μ -values from panel a and the location-dependent Gutenberg–Richter law from Figure 2. The blue triangles in panel b are $M \geq 5$ earthquakes that occurred during the period 2006–2011. Most of them occurred in the highest-rate zones (red areas).

$\gamma(M|t, x, y, H_t)$ of history-dependent space-time configuration as in the case of the foreshock probability forecast (Section 3.4; Ogata & Katsura 2014).

4.3. Temporal Epidemic-Type Aftershock Sequence Model and the Trigger Models

Here, we restrict ourselves to considering seismic activity, ignoring earthquake locations. The data set consists of the occurrence times of earthquakes associated with their magnitudes. Thus, the temporal ETAS model (Ogata 1988, 1989, 1992, 2006a,b) is

$$\lambda(t|H_t) = \mu + \sum_{\{j, t_j < t\}} K_0 e^{\alpha(M_j - M_c)} (t - t_j + c)^{-p}, \quad (11)$$

which is reduced by spatial integration of the conditional intensity of the space-time ETAS model (Equations 8–10). A similar model but with a different form was given by Kagan & Knopoff (1987) as the critical branching model (Kagan 2014).

Kumazawa & Ogata (2013) and Kumazawa et al. (2016a,b) consider a nonstationary ETAS model such that μ and K_0 in Equation 11 are time dependent to examine induced effects from exogenous factors, including slow stress changes due to fault weakening by abrupt pore pressure changes; $\mu(t)$ and $K_0(t)$ are inverted by the empirical Bayesian method (see Section A.3 of Appendix A).

If the conditional intensity function depends only on time t , this characterizes a nonstationary Poisson process. Well-known examples include the Omori–Utsu formula (Utsu 1961, Utsu et al. 1995)

$$\lambda(t) = K/(t + c)^p$$

for aftershock sequences. A comprehensive review of the past century of aftershock studies is given by Utsu et al. (1995). Ogata (1983a,b) suggested the maximum likelihood method for estimating parameters, which enabled practical aftershock forecasting. Special care was placed on objective selection of the target interval $[S, T]$ in Equation 22 to avoid bias in the maximum likelihood estimate due to the substantial lack of smaller aftershocks in the earthquake catalog during the period immediately after the mainshock.

Reasenberg & Jones (1989) implemented aftershock forecasting based on the joint intensity rate (Equation 2) of the time and magnitude of aftershocks (Utsu 1970). To overcome the missing aftershock problem, Ogata & Katsura (2006) use the detection rate function $q\{\cdot\}$ in Equation 7 in such a way that

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

for forecasting early aftershock activity, and Omi et al. (2013) implement this by an automatic sequential procedure. Furthermore, Omi et al. (2014a,b) use the ETAS model, which enables medium-term aftershock forecasting based on early aftershock observations.

Before the study of secondary aftershocks by Utsu (1970), earthquakes were implicitly categorized into primary events (mainshocks) and induced events (aftershocks). The trigger model of Vere-Jones (1970; Vere-Jones & Davies 1966) generates stationary Poisson events as primary events, which characterize aftershocks by the intensity function $g(x)$ of a nonstationary Poisson process. This model was extended in various ways by Lomnitz & Hax (1966), Lomnitz & Nava (1983), and Utsu (1972).

When the category of each earthquake in the data was not determined, the parameters were estimated using the spectral likelihood (Whittle 1962) by which Hawkes & Adamopoulos (1973) showed a better fit of the intensity function $g(x) = \alpha\rho_1 e^{-\rho_1 x} + (1-\alpha)\rho_2 e^{-\rho_2 x}$ than of $g(x) = \rho e^{-\rho x}$; the former is a simple approximation of the Omori–Utsu function (cf. Ogata et al. 1993).

If the category of each event of a data set is identified beforehand, then for a series of occurrence times $\{t_m; m \in P\}$ of the primary events, the conditional intensity is given by

$$\lambda(t | H_t) = \mu + \sum_{t_j < t} K_j (t - t_j + c)^{-p},$$

where K_j is a positive constant for the primary events $j \in P$, otherwise $K_j = 0$, so that we can compute the log-likelihood (Equation 22). For example, by setting $K_j = K_0 e^{\alpha(M_j - M_0)}$ for $j \in P$, otherwise $K_j = 0$, this model (the restricted trigger model; Ogata 1988) was applied to the same data set in which each earthquake was empirically classified by Utsu (personal communication; see also table 1 in Ogata 1988), but the results provide a significantly better fit to the ETAS model.

For descriptive clustering analysis, the restricted trigger model can be used to select some large earthquakes as the primary events (Ogata 2001a). Specifically, setting the parameters $\mu = v^2$ and $K_j = \vartheta_j^2$ for all $j \in P$ to maximize the log-likelihood function of v and ϑ starting from small, positive initial values, such maximum likelihood estimates $\hat{\theta}$ can be stably obtained because of the unimodality and concavity of the log-likelihood function due to the everywhere positive definite of the Hessian matrix, or the matrix of the second derivative components (see Ogata 1978). The

short-term earthquake probability model is the same as the restricted trigger model, and its space-time extension (Gerstenberger et al. 2005) was used for seismicity forecasting in California by the USGS.

4.4. Seismic Anomalies Relative to the Epidemic-Type Aftershock Sequence Model

Since the publication of Inouye's (1965) paper, seismic quiescence has been studied by many investigators for earthquake predictability research. For statistical studies of quiescence, a number of researchers have made declustered earthquake catalogs to test the quiescence against a stationary Poisson process; however, the declustering methods are not universal (see Ogata 1999a, Zhuang et al. 2002, and references therein).

Relative quiescence is found using the ETAS model as follows. First, differences in seismicity before and after a possible change-point are examined using the modified Akaike information criterion, because the maximum likelihood estimate of the change-point does not satisfy ordinary large-sample theory (Ogata 1978). Ogata (1992) and Kumazawa et al. (2010) provide calibration of the Akaike information criterion. For a significant change-point, the residual point process (RPP; see Section A.6 of Appendix A) is made based on the ETAS model that is fitted to the first part of the period. If the seismicity is well predicted for the extended period by the ETAS model, the predicted cumulative curve overlaps with the actual empirical cumulative curve, or the extended cumulative curve of the RPP data defines nearly a straight line. By contrast, a significant decrease in slope for the extended part shows general seismicity or aftershock activity decreasing from the expected occurrence rates.

Since 2001, my colleagues and I have reported on 25 agendas of seismicity anomalies in Japan (<http://www.ism.ac.jp/~ogata/yoti.html>) at the Coordinating Committee for Earthquake Prediction (<http://cais.gsi.go.jp/YOCHIREN/report.html>) of Japan. All agendas except that of Ogata (2006a) were ex-post analysis reports. In addition, among 76 aftershock cases in and around Japan that I have investigated, relative quiescence was observed in 34 cases (for details of the case studies, see Ogata 2001b and its appendix); in Section 5.2, I use the results of this aftershock study for space-time probability prediction of a neighboring earthquake with a size similar to that of the mainshock or larger.

4.5. Periodicity in Seismicity

The analysis of periodicity in seismicity has been attracting the interest of many investigators, who usually use conventional tests of declustered earthquake catalogs against the null hypothesis that the azimuth angles distribute uniformly. However, the declustered method considerably reduces data due to the removal of clustered as well as smaller events to guarantee homogeneity of the data.

In order to overcome these limitations, Ogata (1983b) looked for periodicity by using an extended version of the self-exciting model (Hawkes 1971),

$$\lambda(t|H_i) = \mu + P_J(t) + C_K(t) + \sum_{t_i < t} g_L(t - t_i), \quad (12)$$

where $P_J(t)$ is a polynomial of order J representing the long-term trend in detection capability for earthquakes, $C_K(t)$ is a Fourier series of order K for the periodic component of period T_0 , and the impulse response function $g_L(\cdot)$ of the self-exciting component is represented by a Laguerre-type

Residual point process (RPP):

a series of the transformed times from the occurrence time data using the integration of the conditional intensity in time; it becomes the standard stationary Poisson process if the conditional intensity of the integrand generated the data (see Sections A.4 and A.6 of Appendix A)

function of the order L (see Section 3.3), standing for the effect of clustering. The minimum Akaike information criterion procedure searches for the best combinations of J , K , and L .

The model was applied to shallow seismicity in western Japan (Ogata 1983b) and elsewhere in the world (Li & Vere-Jones 1997, Matsumura 1986, Ogata & Katsura 1986). All the cases are associated with seasonality in precipitation, which likely causes pore-pressure changes to induce seismicity (for more details, see Ogata 1999a). Besides the additive model in Equation 12, certain multiplicative models in the presence of the ETAS model were used in testing the effects on earthquake rates of triggering clusters and earth tides (Iwata 2015; Iwata & Katao 2006a,b).

4.6. Causality Analysis of Seismicity

Tectonically, some seismic causality can be understood as being due to conveyed stress changes. For example, Mogi (1973) discusses the relation between shallow and deep earthquakes along subducting slabs. Utsu (1975) tested the correlation between intermediate-depth earthquakes beneath central Japan and the shallower earthquakes in the Kanto region around Tokyo, along the subducting Pacific plates, and concluded that there was a significant correlation.

Cross-correlation between two point-process data sets $\{t_i\}$ and $\{u_j\}$ is quantified by the cross-Palm intensity (Daley & Vere-Jones 2003), which is realized by the superposition of point configurations $\{t_i - u_j\}$ for all i and j (Cox & Lewis 1966). To test the independence between two stationary series of events from this statistic, the significance level of the cross-covariance density should be calculated (e.g., Mantovani et al. 1986, Ogata 1988). However, the analysis of causality using the cross-correlation is again affected by clustering of events. Declustering of the data sets may avoid this difficulty, but the results depend on the adopted declustering method. Furthermore, a significant cross-correlation does not discriminate among the following four cases: (a) $\{t_i\}$ causes $\{u_j\}$; (b) $\{u_j\}$ causes $\{t_i\}$; (c) $\{u_j\}$ and $\{t_i\}$ cause each other; and (d) some hidden process causes both $\{u_j\}$ and $\{t_i\}$.

To make the analysis in the presence of clustering events and also to discriminate among the four cases, Ogata (1983b) suggested a parametric model,

$$\lambda(t|H_t, F_t) = \mu + P_J(t) + \sum_{t_i < t} g_L(t - t_i) + \sum_{s_i < t} b_M(t - u_i), \quad (13)$$

where H_t is the history of $\{t_i\}$, and F_t is the history of $\{u_i\}$. $P_J(t)$ is a polynomial of order J representing the long-term trend in detection capability for earthquakes. The impulse response function $g_L(x)$ and $b_M(x)$ are the Laguerre-type polynomials of order L and M , respectively. The response functions in the last two terms of Equation 13 can be extended to the ETAS type in cases where magnitude effect is very significant. Thus, the maximum likelihood estimates of the minimum Akaike information criterion model lead to the best combination of orders J , L , and M to provide the timescale and shape of response functions. The conjugate model of $\{u_j\}$ versus $\{t_i\}$ is also considered.

Ogata et al. (1982) applied this model to Utsu's (1975) data mentioned above. The results of these analyses showed that earthquakes in the Kanto region were subject to one-way causality from occurrences in the deeper zone. Moreover, Ogata & Katsura (1986) applied Equation 13 to wider regions with lower magnitude thresholds for a longer period, from 1926 to 1984. Consequently, besides the same one-way causality, a decreasing seismicity trend is revealed in the Kanto region: The trend was caused by the M7.9 Kanto earthquake of 1923. Equation 13 is also applied to shallow and deep earthquakes in other regions of the world (De Natale et al. 1988, Ogata 1983b).

Until now, many earthquake prediction techniques have been proposed on the basis of geophysical anomalies of various types. However, the effectiveness of these techniques is controversial

(Jordan et al. 2011). Therefore, objective evaluation methods are required for such techniques. In particular, we should provide statistical significance (see Section A.2 of Appendix A) of a causal mechanism and assessment of probability gain (see Section A.7 of Appendix A) against the standard seismicity model in the target area. In this respect, Equation 13 can be used to investigate the statistical relation between geophysical anomalies as suspected precursor signals and large earthquakes. For example, Nishizawa et al. (1994) examined the one-way causality of seismoelectric signals (Varotsos et al. 1986) at a station in Greece to earthquake events in a region around the station listed in USGS catalogs, but the risk enhancement factor (see Section A.7 of Appendix A) relative to the self-exciting model was not high.

Furthermore, in cases where suspected precursor signals are given by analog records $\{\xi(u)\}$, or by series of anomaly occurrence times together with their sizes, we can apply a modified version of the model (Ogata & Akaike 1982, Ogata et al. 1982):

$$\lambda_\theta(t|F_t) = a_0 + \sum_{t_i < t} g(t - t_i) + \int_0^t b(t - s)\xi(s)ds. \quad (14)$$

An example is provided by the unusual intensities $\xi(t)$ of ground electric potential observed by day in the Beijing area of China in 1982–1998. By comparing the Akaike information criterion, anomalies were deemed statistically significant as precursors to earthquakes (Zhuang et al. 2005). The conditional intensity rate (events/day) of earthquakes of $M \geq 4$ within a radius of 300 km from a ground-electricity station was given by

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

in the study of Zhuang et al. (2014), in which successively occurring $M \geq 4$ earthquakes within five days and a 30-km distance were removed from the data to negate the self-exciting effect in Equation 14. According to this model, the rate of $M \geq 4$ earthquakes varies from one-half to at most ten times the average occurrence of 0.0126 event/day.

Furthermore, three sets of time series of electric anomaly records were available from three other stations near Beijing. If the four sets of time series were independent of one another, though this is difficult to confirm, we could calculate the conditional intensity rate using Equation 27 in the common region of circular areas within a 300-km radius centered at each of the four stations; such probability gain (except aftershock occurrence rates) varies in the range of approximately one-tenth to one hundred times the average occurrence rate (Ogata & Zhuang 2001).

Also, Kumazawa et al. (2016a) consider a model in which the first two terms of Equation 14 are replaced by the ETAS model (Equation 11), $b(t)$ is an exponential function, and $\xi(s)$ are the hourly changes of volumetric strains at a Japan Meteorological Agency station in a volcanic region. Then the last term of Equation 14 turned out to agree with the background rate changes of the nonstationary ETAS model that is inverted from the corresponding swarm activities (see Section 4.3).

4.7. Models for Recurrent Earthquakes

Here, I consider models for stress-renewal earthquake occurrences, including large interplate events, paleoearthquakes in an active fault segment, and small, repeating earthquakes on a plate boundary. Consider a time series of such events $t_1 < t_2 < \dots < t_i < \dots$. A stationary renewal process is defined if the interval length $Y_i = t_{i+1} - t_i$; $i = 1, 2, \dots$, is independent of the history of occurrences and identically distributed according to a distribution, say, $F_\theta(y)$ of $y \geq 0$. Utsu

(1984) applied various distributions to data sets of interplate earthquakes around Japan, but no single distribution outperformed the others according to the Akaike information criterion.

The conditional intensity function of the renewal process is given by

$$\lambda_\theta(t|H_t) = v_\theta(t - t_{\text{last}}) = f_\theta(t - t_{\text{last}}) / \{1 - F_\theta(t - t_{\text{last}})\},$$

where $v_\theta(\cdot)$ is any nonnegative function of the elapsed time since the last occurrence at time t_{last} , and $f_\theta(\cdot)$ is the density of $F_\theta(\cdot)$. Then, the probability that at least one event will take place in the future interval $(t, t + x)$ is

$$F_\theta(t + x) - F_\theta(t) = 1 - \exp \left\{ - \int_t^{t+x} v(s - t_{\text{last}}) ds \right\}.$$

The logarithm of the exact log-likelihood function given in Equation 22 in the observed interval $[S, T]$ then becomes

$$\ln L_{[S,T]}(\theta) = -\ln v_\theta + \ln \{1 - F_\theta(t_1 - S)\} + \sum_{i=2}^n \ln f_\theta(t_i - t_{i-1}) + \ln \{1 - F_\theta(T - t_n)\},$$

assuming that the process is stationary (cf. Ogata 1978), where $v_\theta = \int_0^\infty t f_\theta(t) dt$. This exact log-likelihood becomes important as the observed number of earthquakes becomes small, including only one observed earthquake; in fact, most of the inland active faults in Japan have only a few recorded events. As the log-likelihood function is ill posed due to sample paucity, Bayesian predictive forecasting (Equation 18) is more useful than the plug-in maximum likelihood estimate predictor (see Sections A.2 and A.3 of Appendix A; see also Nomura et al. 2011, Ogata 2002).

Ogata (1999b) and Nomura et al. (2011) were concerned with analysis of paleoearthquake occurrence dates obtained by trenching of active fault segments, where the uncertain occurrence times of events are given by intervals or probability densities on the time axis.

The Brownian passage time model (Matthews et al. 2002) is a relatively new renewal process model based on a diffusion process $dX_t = vt + \sigma dW_t$, with a loading rate v and a perturbation rate σ . When the perturbed loading reaches the level of critical failure rate of earthquake occurrence, it drops to the background rate, and the Brownian oscillation repeats between the two levels. Thus, the Brownian passage time model has a mean occurrence time μ representing the reciprocal stress-accumulating rate, and the dispersion parameter α value indicates how frequently an active fault earthquake interacts with earthquakes on the neighboring faults.

The Brownian passage time model has been adopted for long-term probability forecasting by the Earthquake Research Committee (ERC 2001), and the observed stress-loading rate and slip sizes are also utilized for estimation purposes. A current issue is the adequacy of $\alpha = 0.24$ commonly adopted for time series of ruptures on major active fault segments by the Earthquake Research Committee. Nomura & Ogata (2015a,b) obtained the maximum likelihood estimate $\alpha = 0.44$ using occurrence data from 61 active faults that have had more than two large earthquakes in inland Japan; then the Akaike information criterion becomes smaller by a difference of 25.6; namely, the model with maximum likelihood estimate $\alpha = 0.44$ is $\exp\{-\Delta\text{AIC}/2\} \approx 1.2 \times 10^{11}$ times more likely compared to the case with $\alpha = 0.24$.

In recent years, small earthquakes repeating on plate boundaries have also attracted much attention (Nadeau & McEvilly 1999). Their recurrence cycles are short, and they are useful for monitoring interplate slip rates around their fault patches. To investigate transient slip-rate changes, Nomura et al. (2014, 2016) considered a nonstationary version of the Brownian passage time model where the loading rate is time dependent, and further extended this to a space-time version. The space-time loading rates are optimally inverted assuming certain proper constraints (see Section A.3 of Appendix A).

Brownian passage time model:

a stochastic model for rupture times on a recurrent earthquake source, adding Brownian perturbations to steady tectonic loading; rupture is assumed to occur when this process reaches a critical-failure threshold

Finally, the stress-release model of Vere-Jones (1978) has a recurrence feature similar to that of the Brownian passage time model, and its extensions have been applied and discussed, for example, by Lu et al. (1999) and Bebbington & Harte (2003).

5. TOWARD OPERATIONAL PROBABILITY FORECASTING OF LARGE EARTHQUAKES

5.1. Forecasts and Probability Gains

Although the probability of a major earthquake under normal circumstances is very small, the probability is increased in the presence of anomalies as potential precursors. This urges us to perform quantitative studies on the statistics of anomalies against earthquakes using various relevant data sets. We describe probability forecasts of a major earthquake based on observed anomaly data or other prediction elements, and consider probability gain defined as the ratio of probability of a large earthquake estimated based on an anomaly to the underlying probability without the anomaly. Section A.7 of Appendix A deals with this concept, and the conditional intensity is calculated using statistical point process models to examine the significance of causality of anomalies and also to evaluate the probability gains, or enhancement factors, conditional on the anomalous events (Section 4.6).

Therefore, we have to search anomalies for potential precursors to apply the multielement probability formula (Equation 25) of Section A.7 of Appendix A, even if the hit rate of a large earthquake by each respective anomaly is low. This demands extensive quantitative studies on the statistics of anomalies using various relevant data sets. There are possibly many delicate anomalies that can be revealed after a diagnostic analysis. In fact, the author and his colleagues have made efforts to discuss seismicity anomalies by diagnostic analysis using the ETAS model fitted to their seismicity data (Ogata 1988, 1989, 1992, 1998a, 2001b, 2006a, 2007, 2010a, 2011a; Ogata et al. 2003a,b; Kumazawa & Ogata 2013; Kumazawa et al. 2010, 2016b). The geodetic anomalies would also be searched for the time series of baseline distances between GPS stations (e.g., Ogata 2007, 2010a, 2011a; Wang et al. 2013), or various strain time series within triangular regions of GPS stations, after developing the proper standard space-time geodetic models. Furthermore, other anomalies including magnetic or electric fields should also be statistically studied in relation to seismicity.

5.2. Applications of the Multielement Probability Formula

In the second half of the 1970s, there were medium-term anomalies in and around the Izu Peninsula, Japan, including crustal uplift, volumetric anomalies, water-level changes, and chemical radon changes; these were reported by the abovementioned Coordinating Committee for Earthquake Prediction. Then, when conspicuous seismic activity started near Izu-Oshima island, which was suspected to be foreshocks, the Japan Meteorological Agency announced a warning a few hours before the 1978 Izu-Oshima-Kinkai M7 earthquake occurred.

Later, to review the experiments, Utsu (1979b) retrospectively calculated the probability of occurrence of a target earthquake of M6.5 or larger by the multielement probability formula (see Section A.7 of Appendix A) assessing probabilities from the anomalies reported by the Coordinating Committee for Earthquake Prediction. In fact, he illustrated that the focal event could be predicted with a probability in the range of approximately 4–89% per day, despite the fact that the probability of each anomaly was extremely low. This illustrative assessment was reviewed by Aki (1981), who gave similar retrospective predictions of the 1975 M7.3 Haicheng China earthquake and three other earthquakes in China (Cao & Aki 1983).

On April 16, 2016, an M7.3 earthquake occurred on the Futagawa fault in the Kumamoto District of Japan. Two days prior to this event, an M6.5 earthquake had occurred and was followed

Table 1 Conditional probability assessments for earthquakes of $M \geq 7^a$

Unit time	Seismicity indicators and anomalies	30 years	1 year	1 month	1 day
Secular^b					
S1	ETAS μ and Gutenberg–Richter	3.0	0.099	0.0081	0.00027
S2	#($M \geq 4$) and Gutenberg–Richter	6.4	0.21	0.017	0.00058
Long-term^c					
A1	Futagawa fault	0.9	0.03	0.0025	8.2E-05
A2	Kumamoto District	10.5	0.35	0.029	0.00096
A3	Central Kyushu	21.0	0.70	0.058	0.0019
Medium-term^d					
B1	Triggering	NA	0.5	0.042	0.0014
B2	Quiescence	NA	2.0	0.17	0.0055
Short-term^e					
C	Foreshock	NA	NA	5.0	0.17

^aAll the probability values in the table are given as percentages.

^bThe secular probabilities of a major earthquake ($M \geq 7.0$ per $1.0^\circ \times 1.0^\circ$ area) in Kumamoto District (Item S) were calculated by the Gutenberg–Richter law of magnitude frequency with either the background rate of the epidemic-type aftershock sequence (ETAS) model (Item S1) or the average number of $M \geq 4.0$ earthquakes per year of the region (Item S2).

^cLong-term probability (Items A1–3) is the 30-year probability of a rupture in one of the active faults in three regions, including the Futagawa fault.

^dMedium-term probability (Items B1,2) is based on the empirical triggering probability of proximate large and induced large earthquakes, respectively, in the case of relative quiescence of the aftershock activity.

^eShort-term probability (Item C) is a foreshock probability per month. Various estimates are considered to evaluate the range of dispersion using multielement probabilities. The probability per day values are all small, but the resulting probabilities evaluated by the multielement prediction formula are not negligible in some cases (see Table 2).

by aftershocks that were actually foreshocks of the M7.3 earthquake. In this case, independent anomalies and time-dependent long-term prediction, for each of which the probability gain can be evaluated, are classified in Table 1 as follows.

- (A) Long-term probabilities of 30 years' rupture incidence are assessed, on the nearest fault (Futagawa fault), or a broad fault system including the central Kyushu, based on the Earthquake Research Committee report; in addition, the half probability of rupture in the central Kyushu can be assessed for the wider Kumamoto region.
- (B) Medium-term probability gain can be assessed such that the preceding M6.5 and M6.4 earthquakes could induce a proximate large earthquake with a probability of approximately 0.5% per year in the near field within about 100 km according to Ogata (2001b). The other probability gain can be assessed by the anomaly phenomena that some aftershock quiescence will induce a large earthquake. Specifically, the aftershock activity preceding the M7.3 earthquake lowered relative to the ETAS model after five hours. Also, aftershock sequences of several past large earthquakes in and around Kyushu indicate relative quiescence (Kumazawa et al. 2016b). Hence, the probability of a large earthquake increases to approximately 2% per year within the 6 years in the near field within about 100 km by such relative quiescence in an aftershock sequence (Ogata 2001b).
- (C) Short-term probability gain indicates that the M6.5 aftershock sequence will be foreshocks of a larger earthquake; the foreshock probability increased from 2% to approximately 5% until the M7.3 earthquake, although this probability is still below the average among such foreshock probabilities in Japan (see Section 3.4).

Table 2 Multielement probabilities of an earthquake of $M \geq 7^a$

Unit time	1 day	3 days	1 week	1 month
$P(S1 A1 \cap B1 \cap C)$	0.25	0.74	1.7	7.3
$P(S1 A1 \cap B2 \cap C)$	0.99	2.9	6.6	23.9
$P(S1 A2 \cap B1 \cap C)$	2.8	8.0	17.0	47.8
$P(S1 A2 \cap B2 \cap C)$	10.4	25.9	45.1	78.6
$P(S1 A3 \cap B1 \cap C)$	5.5	14.9	29.1	64.7
$P(S1 A3 \cap B2 \cap C)$	18.8	41.1	62.2	88.0
$P(S2 A1 \cap B1 \cap C)$	0.043	0.13	0.30	1.3
$P(S2 A1 \cap B2 \cap C)$	0.17	0.51	1.2	5.1
$P(S2 A2 \cap B1 \cap C)$	0.50	1.5	3.4	13.6
$P(S2 A2 \cap B2 \cap C)$	2.0	5.7	12.4	38.8
$P(S2 A3 \cap B1 \cap C)$	0.99	2.9	6.6	24.0
$P(S2 A3 \cap B2 \cap C)$	3.9	10.8	22.1	55.9

^aAll the probability values in the table are given as percentages.

- (S) Finally, we need to evaluate the secular probability of an earthquake of $M \geq 7$ occurring in the Kumamoto region that is calculated by the seismicity rate (number of earthquakes per unit area and unit time) together with the Gutenberg–Richter law. Alternatively, the secular seismicity rate may be obtained from the background seismicity rate (μ -value) of the ETAS model together with the Gutenberg–Richter law.

Then, after transforming each probability mentioned above to the probability per day as summarized in **Table 1**, I get the probability of a $M \geq 7.0$ earthquake occurrence in Kumamoto District based on the multielement probability formula (see Section A.7 of Appendix A), which varies from 0.04% to 19% per day, 0.3% to 62% per week, and 1.3% to 88% per month (see **Table 2**). Readers are referred to the recent article by Ogata (2017) for details and discussions.

6. CONCLUSION

Earthquake occurrences under complex conditions and uncertain elements are incorporated into point-process modeling that leads to stochastic predictions. In particular, a large uncertainty lies in determining whether abnormal phenomena are precursors to large earthquakes, and also in assigning urgency to a given earthquake prediction. This review describes the prospects of earthquake predictability research to realize operational forecasting in the near future.

APPENDIX A: OVERVIEW OF STATISTICAL CONCEPTS AND METHODOLOGIES

A.1. Rationale of Log-Likelihood as the Evaluation Score of Probability Forecast

Evaluation of the forecast performance is essential to researching the predictability of earthquakes. The Collaboratory for the Study of Earthquake Predictability (see Section 2) requires forecasting probabilities $p_k, k = 1, 2, 3, \dots, K$, that are sorted from time-space-magnitude bins (l, i, j, m) . Suppose that n_k events are observed in each bin k and the total number of earthquakes is N . The evaluation score measures how likely the forecast probabilities $\underline{p} = (p_1, p_2, p_3, \dots, p_K)$ can realize

observed relative frequencies $\underline{v} = (v_1, v_2, v_3, \dots, v_K)$, where $v_k = n_k/N$, which is given by

$$\Lambda(\underline{p} | \underline{v}) = \frac{n!}{n_1! n_2! n_3! \cdots n_K!} p_1^{n_1} p_2^{n_2} p_3^{n_3} \cdots p_K^{n_K} \approx \text{const.} \times \exp \left\{ N \sum_k v_k \ln \frac{p_k}{v_k} \right\},$$

and its logarithm,

$$\ln \Lambda(\underline{p} | \underline{v}) = \sum_k v_k \ln p_k - \sum_k v_k \ln v_k + \text{const.}, \quad (15)$$

is the relative entropy (Boltzmann 1878); only the first term depends on the forecasting model \underline{p} , which is the log-likelihood as a natural measure of prediction performance (cf. Akaike 1985). Therefore, evaluation of probabilistic forecasts of earthquakes has been implemented using the log-likelihood score (e.g., Kagan & Jackson 1995; Nanjo et al. 2012; Ogata 1995; Ogata & Katsura 2014; Ogata et al. 1996, 2013; Schorlemmer et al. 2010; Vere-Jones 1999; Woessner et al. 2011).

For example, suppose that we have a series of (conditional) probability forecasts $\{p_i; i = 1, \dots, I\}$. Then, to test whether the forecast is consistent with the outcomes $\{\xi_i; i = 1, \dots, I\}$ such that $\xi_i = 1$ when the earthquake actually occurs and such that $\xi_i = 0$ when it does not, the performance of such forecasts can be measured by using the log-likelihood score as follows. The forecasted probabilities are compared with the unconditional probability forecast p_0 , which can be the average probability. Then, assuming the binomial series for $i = 1, 2, \dots, I$, the log-likelihood ratio (relative entropy) $\ln(L_1/L_0) = \sum_i Q_i$, where

$$Q_i(p | p_0) = \eta_i \log \frac{p_i}{p_0} + (1 - \eta_i) \log \frac{1 - p_i}{1 - p_0} \quad (16)$$

indicates the gain or loss (depending on whether a positive or negative value is taken, respectively) of the forecast p_i against the unconditional probability forecast p_0 at each event i .

Also, the performance of two forecasts $\{p_i; i = 1, \dots, I\}$ and $\{q_i; i = 1, \dots, I\}$ can be compared by the score of the sum of gains or losses:

$$Q_i(p | q) = \eta_i \log \frac{p_i}{q_i} + (1 - \eta_i) \log \frac{1 - p_i}{1 - q_i}.$$

A.2. Log-Likelihood for Parameter Estimation and Model Selection

The relative entropy in Equation 15 can be extended to the case of continuous variables as

$$\ln \Lambda(\underline{p} | \underline{v}) = \int v(\underline{x}) \ln\{p(\underline{x})/v(\underline{x})\} d\underline{x} = E_{\mathbf{X}} \{ \ln p(\mathbf{X}) \} - E_{\mathbf{X}} \{ \ln v(\mathbf{X}) \}, \quad (17)$$

where \mathbf{X} represents a random vector for the stochastic collection of the data sets X during a “learning period” defined by the forecast developer with the probability density distribution $v(x)$, and $E_{\mathbf{X}}\{\cdot\}$ represents the expectation with respect to $v(x)$. Consider that models contain a parameter θ such that $p_{\theta}(x)$. Suppose that the data X from $v(x)$ are observed during the learning period. Then we seek the optimal model among a set of parameterized forecasting models $p_{\theta}(y|X)$ to perform the best forecast of a future data set $Y \in \mathbb{Y}$; namely, the model that most closely simulates the future events with the underlying distribution $v(x)$. The unbiased estimate of the first integral in Equation 17 is $\ln p_{\theta}(X)$, which is the log-likelihood function of θ , and the maximum likelihood estimate, $\hat{\theta} = \hat{\theta}(X)$, that satisfies $\ln p_{\hat{\theta}}(X) = \max_{\theta \in \Theta} \ln p_{\theta}(X)$ is eligible according to the entropy maximization principle. Thus, the predictor becomes $p_{\hat{\theta}(X)}(y|X)$, which we call the maximum likelihood estimate plug-in predictor.

Here, note that both data sets X and Y are generated from the same though unknown underlying distribution $v(\cdot)$. To estimate the forecasting score without using future data Y , Akaike (1974)

introduced the concept of the expected relative entropy $E_{\mathbf{Y}}[E_{\mathbf{Y}}\{\ln p_{\theta}(\mathbf{Y}|\mathbf{X})\}]$, and he found that the maximum log-likelihood score $\ln p_{\hat{\theta}(X)}(X|X)$ has a biased value from the expected relative entropy. Specifically, $E_{\mathbf{Y}}[\max_{\theta} \ln p_{\theta}(\mathbf{X}|\mathbf{X})] \approx E_{\mathbf{Y}}[E_{\mathbf{Y}}\{\ln p_{\hat{\theta}(X)}(\mathbf{Y}|\mathbf{X})\}] + \dim\{\theta\}$, and the statistics $-\text{AIC}/2 = \ln p_{\hat{\theta}(X)}(X|X) - \dim\{\theta\}$ is the unbiased estimate of the expected relative entropy, where

$$\text{AIC} = -2 \max \ln \text{likelihood} + 2 (\text{number of parameters})$$

is the Akaike information criterion (Akaike 1974). A model with a smaller Akaike information criterion value is expected to be a better predictor; for the difference of Akaike information criterion values between two competing models, $\exp\{-\Delta\text{AIC}/2\}$ can be interpreted as the relative probability of how much the better model is superior to the other (Akaike 1978b) in view of the rationale presented in Section A.1.

Maximum a posteriori solution:
 parameters that maximize a posterior function that represents a trade-off between a likelihood function of parameters for the fit to data and prior probability for constraints among parameters; equivalent to the solution of parameters θ that minimize the penalized log-likelihood function in Equation 20 given the weights τ

A.3. Empirical Bayesian Methods

In the case where the log-likelihood function is ill-conditioned, with nonuniqueness of the maximizing parameter, the plug-in predictor does not work well. Instead, we need to consider a suitable prior probability $\pi(\theta|\tau)$ that provides constraints on the parameter values θ with adjusting hyperparameters τ ; such examples are exponential to the negative penalty function given in Sections 3 and 4. Then, together with the likelihood function $p_{\theta}(X)$, we obtain the posterior function posterior $(\theta|X, \tau) = p_{\theta}(X) \cdot \pi(\theta|\tau)$ and the Bayesian predictor

$$p(y|X, \hat{\tau}) \propto \int p_{\theta}(y) \cdot \text{posterior}(\theta|X, \hat{\tau}) d\theta. \quad (18)$$

The Bayesian predictor is more robust than the plug-in predictor for the case of a small number of data against the number of parameters, as seen in Section 4.7. Here, the optimal selection of hyperparameters $\hat{\tau}$ is derived by minimizing the Akaike Bayesian information criterion (Akaike 1978a, Parzen et al. 1998):

$$\text{ABIC} = -2 \max_{\tau} \ln \int \text{posterior}(\theta|X, \tau) d\theta + 2\dim(\tau). \quad (19)$$

To get a robust and sensible estimate of a model with many parameters, we have to assume constraints among the parameters, as defined by a penalty function $\text{penalty}(\theta|\tau)$; the constraining strengths are adjusted by a set of hyperparameters τ . For fixed values of τ , the penalized log-likelihood function (Good & Gaskins 1971)

$$Q(\theta|\tau) = \ln L(\theta|X) - \text{penalty}(\theta|\tau) \quad (20)$$

is minimized to get the estimate of the parameters θ . Thus, the crucial point becomes how to determine the optimal weights τ . The exponential of the penalized log-likelihood provides a posterior distribution, but note here that $\exp\{-\text{penalty}(\theta|\tau)\}$ is normalized for the prior probability $\pi(\theta|\tau)$ in calculating the Akaike Bayesian information criterion (Equation 19). If the penalized log-likelihood function (Equation 20) has a quadratic form with respect to the parameters θ , or the posterior is proportional to the normal distribution, its integration is reduced to the determinant of the variance-covariance matrix; otherwise, it needs the Laplace (1774) approximation by using the Taylor's expansion of log posterior function around the maximum a posteriori solution, or maximum penalized log-likelihood. To get the optimal hyperparameters and to calculate the Akaike Bayesian information criterion value, one has to repeat optimizations until they eventually converge (for the details, see Ogata & Katsura 1988, 1993; Ogata et al. 1998, 2003b).

A.4. Simulation of Point Processes by Thinning

Lewis & Shedler (1979) introduced the thinning simulation method. Given an intensity function $\lambda(t)$ of a nonstationary Poisson process, consider a constant Λ such that $\lambda(t) \leq \Lambda$ on a time interval $S \leq t \leq T$. First, simulate a stationary Poisson process of a constant intensity Λ ; namely, generate uniform random number U_n to take $t_{n+1} = t_n - \Lambda^{-1} \ln U_n$; then accept t_n with probability $\lambda(t)/\Lambda$, or otherwise reject it. The accepted times constitute a series of events from the nonstationary Poisson process. This thinning method can be extended for general point processes of the conditional intensity $\lambda(t|H_t)$ by recursively setting stepwise conditional intensity $\Lambda(t|H_t)$ such that $\lambda(t|H_t) \leq \Lambda(t|H_t)$. The thinning simulation method can be further extended to multivariate, multichannel, and marked point processes with a general state, including space-time point processes (for the details, see Ogata 1981, 1998b).

A.5. Log-Likelihood Function of Point Processes and the Maximum Likelihood Estimate

Suppose that we have observed earthquakes at time-space-magnitude coordinates (t_n, x_n, y_n, M_n) , $M_n \geq M_c$, $n = 1, 2, \dots, N$, during a period $[S, T]$ and in a region A . Consider the space-time conditional intensity function that is parameterized as $\lambda_\theta(t, x, y, M|H_t)$. Then, the log-likelihood function (see Daley & Vere-Jones 2003) of θ is given by

$$\ln L(\theta) = \sum_{i=1}^N \ln \lambda_\theta(t_i, x_i, y_i, M_i|H_{t_i}) - \int_S^T \iint_A \int_{M_c}^\infty \lambda_\theta(t, x, y, M|H_t) dt dx dy dM. \quad (21)$$

Also, consider a parameterized conditional intensity function $\lambda_\theta(t|H_t)$ for the temporal sequence of occurrence times (t_n, M_n) , $n = 1, 2, \dots, N$, in a seismogenic region; the log-likelihood is

$$\ln L(\theta) = \sum_{i=1}^N \ln \lambda_\theta(t_i|H_{t_i}) - \int_S^T \lambda_\theta(t|H_t) dt. \quad (22)$$

The maximum likelihood method involves maximization of the log-likelihood function with respect to θ , and the maximizing parameter value $\hat{\theta}$ is called the maximum likelihood estimate. The maximum likelihood estimate has some optimal features in terms of estimation according to the statistical theory of large samples (Karr 1986, Kutoyants 1984, Ogata 1978). Numerically, the maximization is implemented by sequentially calculating the log-likelihood value and its gradients for a given parameter vector by using a standard quasi-Newton optimization technique.

A.6. Diagnostic Analysis for Point Processes

For example, suppose that the maximum likelihood estimate $\hat{\theta} = (\hat{\mu}, \widehat{K}_0, \hat{\epsilon}, \hat{\alpha}, \hat{p})$ of the epidemic-type aftershock sequence (ETAS) model in Equation 11 is given. Then the integral of the conditional intensity function,

$$\Lambda_{\hat{\theta}}(t|H_t) = \int_S^t \lambda_{\hat{\theta}}(u|H_u) du, \quad (23)$$

provides the expected cumulative number of earthquakes on the time interval $[S, t]$. If the model of the maximum likelihood estimate represents the real seismicity, Equation 23 closely agrees with the empirical cumulative counts $N(t)$ of the observed earthquakes (**Figure 4**).

Furthermore, consider

$$\tau = \Lambda_{\hat{\theta}}(t|H_t), \quad (24)$$

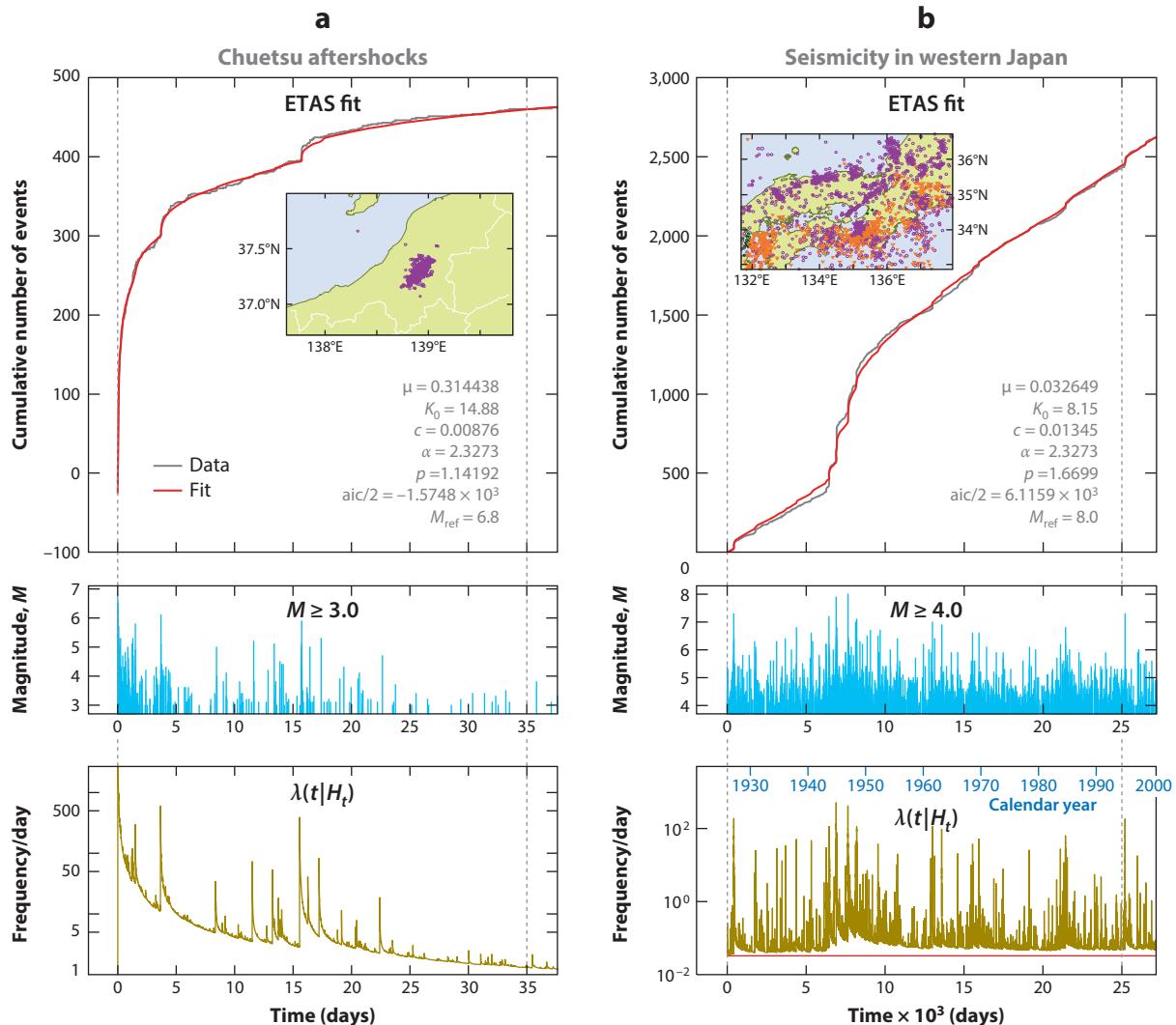


Figure 4

Application of the epidemic-type aftershock sequence (ETAS) model to (a) aftershocks of the 2004 M6.8 Niigata-Ken Chuetsu earthquakes and (b) seismic activity in western Japan for the period of approximately 1926–2000. Shown in both panels are the empirical (gray curve) and theoretical (red curve) cumulative numbers against time (top rows), magnitude versus time (middle rows), and conditional intensity rates against time (bottom rows). The horizontal red line at the bottom of panel b indicates the background rate in the seismicity of western Japan.

that transforms the ordinary occurrence time of aftershocks (t_1, t_2, \dots, t_N) into the sequence $(\tau_1, \tau_2, \dots, \tau_N)$, called a residual point process (RPP) (Ogata 1988). This also implies that the RPP appears to be a stationary Poisson process with unit intensity (Papangelou 1972). If, on the contrary, we find a misfit of the RPP $(\tau_1, \tau_2, \dots, \tau_N)$ to a stationary Poisson process, this suggests either the model is incorrect or the data include some anomalies.

To examine whether or not the RPP obeys the standard stationary Poisson process, we first examine the cumulative curve of the RPP in comparison with that of the uniform distribution.

In addition, various other tests against the stationary Poisson assumption are carried out (Cox & Lewis 1966, Ogata 1988). In particular, to find local unusual characteristics in RPP that might invalidate the stationary Poisson assumption, we consider the number of points $\xi_b(\tau) = N(\tau - b, \tau)$ in the interval $(\tau - b, \tau)$. Under the null hypothesis, $\xi_b(\tau)$ should behave as a Poisson random variable with mean b and variance b for each τ . Setting a not very small b , the time series of $\xi_b(\tau)$ of τ should behave as a stationary normal process with a short-range correlation. For example, having applied the ETAS model to wide-area seismicity, a span of outlying values above a significant high level corresponds to a period of swarm activity (Ogata 1988).

For an estimated space-time model $\lambda_{\hat{\theta}}(t, x, y | H_t)$, a space-time diagnostic analysis can be carried out by estimating a flexible function $\xi(t, x, y; \phi)$ that is a factor in the new intensity function

$$\eta_{\phi}(t, x, y) = \lambda_{\hat{\theta}}(t, x, y | H_t) \exp \{ \xi(t, x, y; \phi) \}.$$

This is applied to the same data as the model from which $\lambda_{\hat{\theta}}$ was estimated, under a constraint against roughness of ξ using the penalized log-likelihood (Equation 20). The optimal maximum a posteriori solution of the ξ -function provides an image of the residual, showing where and when a systematic deviation (enhanced activity or quiescence) takes place relative to the originally estimated model $\lambda_{\hat{\theta}}(t, x, y | H_t)$. Examples can be seen in the work of Ogata et al. (2003b) and Ogata (2004, 2010b).

A.7. Multielement Probability Formula for Plural Precursory Phenomena

Let the event M be an earthquake of magnitude M or more occurring within a time interval Δ . The occurrence probability is $P(M)$, which is calculated by assuming a stationary Poisson process. Let the events $\{A_n, n = 1, 2, \dots, N\}$ be observed anomalous phenomena, and any other information that is suspected to have some precursory relation to the earthquake. An anomalous event A is useful if the conditional probability $P(M|A)$ is larger than $P(M)$, and $P(M|A)/P(M)$ is called the probability gain, which measures the effectiveness of the anomaly as a predictive precursor. However, in many cases, suspected anomalous phenomena include record outliers or noise, and the probability gain is not very substantial, so that use of a single anomaly is rarely practical.

Assuming that anomalies $\{A_n, n = 1, 2, \dots, N\}$ are independent of each other (conditional on the impending earthquake M), Utsu (1977) derived the following conditional probability by using the Bayes rule (see Aki 1981):

$$P \left(M \mid \bigcap_{n=1}^N A_n \right) = \left[1 + \prod_{n=1}^N \left\{ P(M|A_n)^{-1} - 1 \right\} \middle/ \left\{ P(M)^{-1} - 1 \right\}^{N-1} \right]^{-1}, \quad (25)$$

called the multielement probability formula. Aki (1981) approximated this by

$$P \left(M \mid \bigcap_{n=1}^N A_n \right) \approx P(M) \prod_{n=1}^N \frac{P(M|A_n)}{P(M)}, \quad (26)$$

when the time interval Δ is small.

If we consider conditional intensity function models $\lambda(t, x, y, M | H_t, F_t)$ that depend on occurrence history $\{H_t\}$ and data sets of various anomalies $F_t = \{F_t^k; k = 1, 2, \dots, K\}$ that are mutually independent among k conditional on their histories, then we can consider Equation 26 in such a way that

$$\lambda(t, x, y, M | H_t, F_t^1, \dots, F_t^K) = \lambda_0(t, x, y, M | H_t) \prod_{k=1}^K \frac{\lambda_k(t, x, y, M | H_t, F_t^k)}{\lambda_0(t, x, y, M | H_t)}, \quad (27)$$

where $\lambda_0(t, x, y, M | H_t)$ represents the conditional intensity for reference seismic activity. Vere-Jones (1978) called the ratios in Equation 27 risk enhancement factors. When a large earthquake is forecast for a long period in a given seismogenic region, one may use an approximate reference seismicity model such as the stationary Poisson process $\lambda_0(M)$ in time (Utsu 1979) or nonhomogeneous spatial Poisson processes $\lambda_0(x, y, M)$ (Zhuang & Jiang 2012).

The multielement prediction formula is based on the strong assumption of independence between anomalous events. Among long-, medium-, and short-term anomalous phenomena, we may well assume their mutual independence. However, when we have many kinds of anomalies, independence is not necessarily justified, and it is necessary to consider the generalized version. For a probability value p ($0 \leq p \leq 1$), consider the logit transformation (e.g., Cox & Snell 1989)

$$f = \text{logit } p = \ln \eta(p), \text{ where } \eta(p) = p/(1-p), \text{ or } p = 1/(1 + e^{-f}),$$

and then the formula in Equation 25 is represented by

$$\text{logit } P\left(M | \bigcap_{n=1}^N A_n\right) = \sum_{n=1}^N \text{logit } P(M | A_n) - (N - 1) \text{logit } P(M). \quad (28)$$

In general, this relation can be extended by logistic regression to the nonlinear form,

$$\text{logit } P\left(M | \bigcap_{n=1}^N A_n\right) = \varphi(f_1, f_2, \dots, f_N; f_0), \quad (29)$$

in terms of $f_n = \text{logit } P(M | A_n)$ for $n = 1, 2, \dots, N$ and $f_0 = \text{logit } P(M)$. The Akaike information criterion is useful for the selection of specific competing models of Equation 29, including Equation 28 (e.g., Ogata et al. 1996). Also, the log conditional intensity $\ln \lambda(t, x, y, M | H_t, F_t^1, \dots, F_t^N)$ is represented in terms of $f_0 = \ln \lambda_0(t, x, y, M | H_t)$ and $f_n = \ln \lambda_n(t, x, y, M | H_t, F_t^n)$.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The reviewers' comments provided useful clarification. This manuscript was partly supported by grants (26240004 and 17H00727) from the Japan Society for the Promotion of Science Grants-in Aid for Scientific Research (KAKENHI).

LITERATURE CITED

- Agnew DC, Jones LM. 1991. Prediction probabilities from foreshocks. *J. Geophys. Res.* 96(B7):11959–71
- Akaike H. 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Control* 19:716–23
- Akaike H. 1978a. A new look at the Bayes procedure. *Biometrika* 65:53–59
- Akaike H. 1978b. On the likelihood of a time series model. *Statistician* 27:217–35
- Akaike H. 1980. Likelihood and Bayes procedure. In *Bayesian Statistics*, ed. JM Bernard, MH De Groot, DU Lindley, AFM Smith, pp. 143–66 (discussion 185–203). Valencia, Spain: Univ. Press
- Akaike H. 1985. Prediction and entropy. In *A Celebration of Statistics: The ISI Centenary Volume*, ed. AC Atkinson, SE Fienberg, pp. 1–24. New York: Springer-Verlag
- Aki K. 1956. A review on statistical seismology. *Zisin II* [J. Seismol. Soc. Jpn.] 8:205–28
- Aki K. 1965. Maximum likelihood estimate of b in the formula $\log N = a - bM$ and its confidence limits. *Bull. Earthq. Res. Inst. Univ. Tokyo* 43:237–39

- Aki K. 1981. A probabilistic synthesis of precursory phenomena. In *Earthquake Prediction*, ed. DW Simpson, PG Richards, pp. 566–74. Washington, DC: Am. Geophys. Union
- Bansal AR, Ogata Y. 2013. A non-stationary epidemic type aftershock sequence model for seismicity prior to the December 26, 2004 M9.1 Sumatra-Andaman Islands mega-earthquake. *J. Geophys. Res. Solid Earth* 118:616–29
- Bebbington M, Harte D. 2003. The linked stress release model for spatio-temporal seismicity: formulations, procedure and applications. *Geophys. J. Int.* 154:925–46
- Boltzmann L. 1878. Weitere Bemerkungen über einige Probleme der mechanischen Warmetheorie. *Wiener Ber.* 78:7–46
- Cao T, Aki K. 1983. Assigning probability gain for precursors of four large Chinese earthquakes. *J. Geophys. Res.* 88(B3):2185–90
- Console R, Murru M, Alessandrini B. 1993. Foreshock statistics and their possible relationship to earthquake prediction in the Italian region. *Bull. Seismol. Soc. Am.* 83:1248–63
- Cox DR, Lewis PAW. 1966. *The Statistical Analysis of Series of Events*. London: Methuen
- Cox DR, Snell EJ. 1989. *Analysis of Binary Data*. London: Chapman & Hall/CRC. 2nd ed.
- Daley DJ, Vere-Jones D. 2003. *An Introduction to the Theory of Point Processes*. Vol 1: *Elementary Theory and Methods*. New York: Springer. 2nd ed.
- De Natale G, Musmeci F, Zollo A. 1988. A linear intensity model to investigate the causal relation between Calabrian and North-Aegean earthquake sequences. *Geophys. J.* 95:285–93
- Delaunay B. 1934. Sur la sphère vide. *Bull. Acad. Sci. URSS* 6:793–800
- ERC (Earthq. Res. Comm.). 2001. *Regarding methods for evaluating long-term probability of earthquake occurrence*. Rep., Hqrs. Earthq. Res. Promot., Tokyo. <http://www.jishin.go.jp/main/choukihyoka/01b/chouki020326.pdf>
- Evison FF. 1977. The precursory earthquake swarm. *Phys. Earth Planet. Inter.* 15:19–23
- Frohlich C, Davis SD. 1990. Single-link cluster analysis as a method to evaluate spatial and temporal properties of earthquake catalogs. *Geophys. J. Int.* 100:19–32
- Gerstenberger M, Wiemer S, Jones R. 2005. Real-time forecasts of tomorrow's earthquakes in California. *Nature* 435:328–31
- Good IJ, Gaskins RA. 1971. Nonparametric roughness penalties for probability densities. *Biometrika* 58:255–77
- Gutenberg R, Richter CF. 1944. Frequency of earthquakes in California. *Bull. Seismol. Soc. Am.* 34:185–88
- Hawkes AG. 1971. Point spectra of some mutually exciting point processes. *J. R. Stat. Soc. B* 33:438–43
- Hawkes AG, Adamopoulos L. 1973. Cluster models for earthquakes—regional comparisons. *Bull. Int. Stat. Inst.* 45(Book 3):454–61
- Hawkes AG, Oakes DA. 1974. A cluster process representation of a self-exciting process. *J. Appl. Probab.* 11:493–503
- Helmstetter A, Sornette D. 2003. Foreshocks explained by cascades of triggered seismicity. *J. Geophys. Res.* 108(B10):2457
- Inouye W. 1965. On the seismicity in the epicentral region and its neighborhood before the Niigata earthquake. *Kenshin-jibo* [Q. J. Seismol.] 29:139–144
- Iwata T. 2015. Earthquake forecasting based on the correlation between earth tides and earthquake occurrences. *Proc. Inst. Stat. Math.* 63(1):129–44
- Iwata T, Katao H. 2006a. *Correlation between the phase of the moon and the occurrences of microearthquakes in the Tamba region by a modified conditional intensity model*. Presented at Seismol. Soc. Jpn. Fall Meet., Oct. 31–Nov. 2, Nagoya, Jpn., Abstr. P212
- Iwata T, Katao H. 2006b. Correlation between the phase of the moon and the occurrences of microearthquakes in the Tamba region through point-process modeling. *Geophys. Res. Lett.* 33:L07302
- Jones LM. 1985. Foreshocks and time-dependent earthquake hazard assessment in southern California. *Bull. Seismol. Soc. Am.* 75:1669–79
- Jones LM. 1994. Foreshocks, aftershocks and earthquake probabilities: accounting for Landers earthquakes. *Bull. Seismol. Soc. Am.* 85:892–99
- Jones LM, Molnar P. 1979. Some characteristics of foreshocks and their possible relationship to earthquake prediction and premonitory slip on faults. *J. Geophys. Res.* 84(B7):3596–608

- Jordan TH, Chen YT, Gasparini P, Madariaga R, Marzocchi W, et al. 2011. Operational earthquake forecasting: state of knowledge and guidelines for utilization. *Ann. Geophys.* 54(4):315–91
- Kagan YY. 1991. Likelihood analysis of earthquake catalogues. *Geophys. J. Int.* 106:135–48
- Kagan YY. 2014. *Earthquakes: Models, Statistics, Testable Forecasts*. Chichester, UK: Wiley
- Kagan YY, Jackson D. 1995. New seismic gap hypothesis—5 years after. *J. Geophys. Res.* 100(B3):3943–59
- Kagan YY, Knopoff L. 1987. Random stress and earthquake statistics: time dependence. *Geophys. J. R. Astron. Soc.* 88:723–31
- Karr AN. 1986. *Point Processes and Their Statistical Inference*. New York: Marcel Dekker
- Kendall DG. 1949. Stochastic processes and population growth. *J. R. Stat. Soc. B* 11:230–64
- Kumazawa T, Ogata Y. 2013. Quantitative description of induced seismic activity before and after the 2011 Tohoku-Oki earthquake by non-stationary ETAS models. *J. Geophys. Res. Solid Earth* 118:6165–82
- Kumazawa T, Ogata Y, Kimura K, Maeda K, Kobayashi A. 2016a. Background rates of swarm earthquakes that are synchronized with volumetric strain changes. *Earth Planet. Sci. Lett.* 442:51–60
- Kumazawa T, Ogata Y, Toda S. 2010. Precursory seismic anomalies and transient crustal deformation prior to the 2008 $M_w = 6.9$ Iwate-Miyagi Nairiku, Japan, earthquake. *J. Geophys. Res.* 115:B10312
- Kumazawa T, Ogata Y, Tsuruoka H. 2016b. Statistical monitoring of seismicity in Kyushu District before the occurrence of the 2016 Kumamoto earthquakes of M6.5 and M7.3. *Rep. Coord. Comm. Earthq. Predic.* 96(12–8):412–23
- Kutoyants YU. 1984. *Parameter Estimation for Stochastic Processes*. Berlin: Heldermann
- Laplace PS. 1774 (1986). Memoir on the probability of causes of events, Vol. 6, transl. SM Stigler. *Stat. Sci.* 1(19):364–78
- Lewis PAW, Shedler GS. 1979. Simulation of nonhomogeneous Poisson processes by thinning. *Nav. Res. Logist. Q.* 26(3):403–13
- Li M, Vere-Jones D. 1997. Application of M8 and Lin-Lin algorithms to New Zealand earthquake data. *N. Z. J. Geol. Geophys.* 40:77–89
- Lomnitz C, Hax A. 1966. Clustering in aftershock sequences. In *The Earth Beneath the Continents*, ed. JS Steinhart, TJ Smith, pp. 502–8. Washington, DC: Am. Geophys. Union
- Lomnitz C, Nava FA. 1983. The predictive value of seismic gaps. *Bull. Seismol. Soc. Am.* 73:1815–24
- Lu C, Harte D, Bebbington M. 1999. A linked stress release for historical Japan earthquakes: coupling among major seismic regions. *Earth Planets Space* 51:907–16
- Maeda K. 1993. An empirical alarm criterion based on immediate foreshocks: a case study for the Izu region. *Zisin II* [J. Seismol. Soc. Jpn.] 45:373–83
- Mantovani E, Albarello D, Mucciarelli M. 1986. Seismic activity in North Aegean region as middle-term precursor of Calabrian earthquakes. *Phys. Earth Planet. Int.* 44:264–73
- Matsumura K. 1986. On regional characteristics of seasonal variation of shallow earthquake activities in the world. *Bull. Disaster Prev. Res. Inst. Kyoto Univ.* 36:43–98
- Matthews MV, Ellsworth WL, Reasenberg PA. 2002. A Brownian model for recurrent earthquakes. *Bull. Seismol. Soc. Am.* 92:2233–50
- Mogi K. 1962. Magnitude frequency relations for elastic shocks accompanying fractures of various materials and some related problems in earthquakes. *Bull. Earthq. Res. Inst. Univ. Tokyo* 40:831–53
- Mogi K. 1963. Some discussions on aftershocks, foreshocks, and earthquake swarms—the fracture of a semi-infinite body caused by an inner stress origin and its relation to the earthquake phenomena. *Bull. Earthq. Res. Inst. Univ. Tokyo* 41:615–58
- Mogi K. 1973. Relationship between shallow and deep seismicity in the western Pacific region. *Tectonophysics* 17:1–22
- Musmeci F, Vere-Jones D. 1992. A space-time clustering model for historical earthquakes. *Ann. Inst. Stat. Math.* 44:1–11
- Nadeau RM, McEvilly TV. 1999. Fault slip rates at depth from recurrence intervals of repeating microearthquakes. *Science* 285:718–21
- Nanjo KZ, Tsuruoka H, Yokoi S, Ogata Y, Falcone G, et al. 2012. Predictability study on the aftershock sequence following the 2011 Tohoku-Oki, Japan, earthquake: first results. *Geophys. J. Int.* 191(2):653–58

- Nishizawa O, Lei X, Nagao T. 1994. Hazard function analysis of seismo-electric signals in Greece. In *Electromagnetic Phenomena Related to Earthquake Prediction*, ed. M Hayakawa, Y Fujinawa, pp. 459–74. Tokyo: Terra
- Nomura S. 2015. *Real-time classifier of foreshocks using probability density ratio estimation*. Presented at Jpn. Geosci. Union Meet., May 24, Chiba City, Jpn., Abstr. SSS01-01
- Nomura S, Ogata Y. 2015a. *Foreshock discrimination with probability density ratio estimation and its implementation for short-term forecast*. Presented at Seismol. Soc. Jpn. Fall Meet., Oct. 27, Kobe City, Jpn.
- Nomura S, Ogata Y. 2015b. *Spatial distribution of the coefficient of variation for the paleo-earthquakes in Japan*. Presented at AGU Fall Meet., Dec. 15, San Francisco, Abstr. S21D-05
- Nomura S, Ogata Y, Komaki F, Toda S. 2011. Bayesian forecasting of the recurrent earthquakes and its predictive performance for a small sample size. *J. Geophys. Res.* 116:B04315
- Nomura S, Ogata Y, Nadeau RM. 2014. Space-time model for repeating earthquakes and analysis of recurrence intervals on the San Andreas Fault near Parkfield, California. *J. Geophys. Res. Solid Earth* 119(9):7092–122
- Nomura S, Ogata Y, Uchida N, Matu'ura M. 2016. Spatiotemporal variations of interplate slip rates in northeast Japan inverted from recurrence intervals of repeating earthquakes. *Geophys. J. Int.* 208:468–81
- Ogata Y. 1978. The asymptotic behavior of maximum likelihood estimator of stationary point processes. *Ann. Inst. Stat. Math.* 30:243–61
- Ogata Y. 1981. On Lewis' simulation method for point processes. *IEEE Trans. Inf. Theory* 27:23–31
- Ogata Y. 1983a. Estimation of the parameters in the modified Omori formula for aftershock frequencies by the maximum likelihood procedure. *J. Phys. Earth.* 31:115–24
- Ogata Y. 1983b. Likelihood analysis of point processes and its applications to seismological data. *Bull. Int. Stat. Inst.* 50(2):943–61
- Ogata Y. 1988. Statistical models for earthquake occurrences and residual analysis for point processes. *J. Am. Stat. Assoc.* 83:9–27
- Ogata Y. 1989. Statistical model for standard seismicity and detection of anomalies by residual analysis. *Tectonophysics* 169:159–74
- Ogata Y. 1991. Space and time variations of detection rate and the magnitude frequency of earthquakes. *Tokei Suri [Proc. Inst. Stat. Math.]* 39(2):245–56
- Ogata Y. 1992. Detection of precursory relative quiescence before great earthquakes through a statistical model. *J. Geophys. Res.* 97(B13):19845–71
- Ogata Y. 1995. Evaluation of probability forecasts of events. *Int. J. Forecast.* 11:539–41
- Ogata Y. 1998a. Quiescence relative to the ETAS model. *Zisin II [J. Seismol. Soc. Jpn.]* 50:115–27
- Ogata Y. 1998b. Space-time point-process models for earthquake occurrences. *Ann. Inst. Stat. 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(B8):17995–8014
- Ogata Y. 2001a. Exploratory analysis of earthquake clusters by likelihood-based trigger models. *J. Appl. Probab.* 38(A):202–12
- Ogata Y. 2001b. Increased probability of large earthquakes near aftershock regions with relative quiescence. *J. Geophys. Res.* 106(B5):8729–44
- Ogata Y. 2002. Slip-size-dependent renewal processes and Bayesian inferences for uncertainties. *J. Geophys. Res.* 107(B11):2268
- Ogata Y. 2004. Space-time model for regional seismicity and detection of crustal stress changes. *J. Geophys. Res.* 109(B3):B03308
- 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. Lett.* 33:L01303
- Ogata Y. 2006b. *Statistical Analysis of Seismicity—Updated Version (SASeis2006)*. Comput. Sci. Monogr. 33. Tokyo: Inst. Stat. Math. http://www.ism.ac.jp/~ogata/Ssg/ssg_softwaresE.html
- Ogata Y. 2007. Seismicity and geodetic anomalies in a wide area preceding the Niigata-Ken-Chuetsu earthquake of 23 October 2004, central Japan. *J. Geophys. Res.* 112:B10301

- Ogata Y. 2010a. Anomalies of seismic activity and transient crustal deformations preceding the 2005 M7.0 earthquake west of Fukuoka. *Pure Appl. Geophys.* 167(8–9):1115–27
- Ogata Y. 2010b. Space-time heterogeneity in aftershock activity. *Geophys. J. Int.* 181(3):1575–92
- Ogata Y. 2011a. 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.* 186:331–48
- Ogata Y. 2011b. Significant improvements of the space-time ETAS model for forecasting of accurate baseline seismicity. *Earth Planets Space* 63:217–29
- Ogata Y. 2017. Forecasting of a large earthquake: an outlook of the research. *Seismol. Res. Lett.* 88(4):1117–26
- Ogata Y, Abe K. 1991. Some statistical features of the long term variation of the global and regional seismic activity. *Int. Stat. Rev.* 59:139–61
- Ogata Y, Akaike H. 1982. On linear intensity models for mixed doubly stochastic Poisson and self-exciting point process. *J. R. Stat. Soc. B* 44:102–7
- Ogata Y, Akaike H, 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. Stat. Math.* 34:373–87
- Ogata Y, Imoto M, Katsura K. 1991. 3-D spatial variation of b -values of magnitude-frequency distribution beneath the Kanto District, Japan. *Geophys. J. Int.* 104:135–46
- Ogata Y, Jones LM, Toda S. 2003a. When and where the aftershock activity was depressed: contrasting decay patterns of the proximate large earthquakes in southern California. *J. Geophys. Res.* 108(B6):2318
- Ogata Y, Katsura K. 1986. Point process model with linearly parameterized intensity for the application to earthquake data. *J. Appl. Probab.* 23(A):291–310
- Ogata Y, Katsura K. 1988. Likelihood analysis of spatial inhomogeneity for marked point patterns. *Ann. Inst. Stat. Math.* 40:29–39
- Ogata Y, Katsura K. 1993. Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogues. *Geophys. J. Int.* 113:727–38
- Ogata Y, Katsura K. 2006. Immediate and updated forecasting of aftershock hazard. *Geophys. Res. Lett.* 33(10):L10305
- Ogata Y, Katsura K. 2012. Prospective foreshock forecast experiment during the last 17 years. *Geophys. J. Int.* 191(3):1237–44
- Ogata Y, Katsura K. 2014. Comparing foreshock characteristics and foreshock forecasting in observed and simulated earthquake catalogs. *J. Geophys. Res. Solid Earth* 119(11):8457–77
- Ogata Y, Katsura K, Falcone G, Nanjo KZ, Zhuang J. 2013. Comprehensive and topical evaluations of earthquake forecasts in terms of number, time, space, and magnitude. *Bull. Seismol. Soc. Am.* 103:1692–708
- Ogata Y, Katsura K, Tanemura M. 2003b. Modelling heterogeneous space-time occurrences of earthquakes and its residual analysis. *J. R. Stat. Soc. C* 52(4):499–509
- Ogata Y, Kobayashi A, Mikami N, Murata Y, Katsura K. 1998. Correction of earthquake location estimation in a small-seismic-array system. *Bernoulli* 4:167–84
- Ogata Y, Matsu'ura RS, Katsura K. 1993. Fast likelihood computation of epidemic type aftershock-sequence model. *Geophys. Res. Lett.* 20:2143–46
- Ogata Y, Utsu T, Katsura K. 1995. Statistical features of foreshocks in comparison with other earthquake clusters. *Geophys. J. Int.* 121:233–54
- Ogata Y, Utsu T, Katsura K. 1996. Statistical discrimination of foreshocks from other earthquake clusters. *Geophys. J. Int.* 127:17–30
- Ogata Y, Zhuang J. 2001. Statistical examination of anomalies for the precursor to earthquakes, and the multi-element prediction formula: hazard rate changes of strong earthquakes ($M \geq 4$) around Beijing area based on the ultra-low frequency ground electric observation (1982–1997). *Rep. Coord. Comm. Earthq. Predict.* 66:562–70
- Ogata Y, Zhuang J. 2006. Space-time ETAS models and an improved extension. *Tectonophysics* 413:13–23
- Omi T, Ogata Y, Hirata Y, Aihara K. 2013. Forecasting large aftershocks within one day after the main shock. *Sci. Rep.* 3:2218
- Omi T, Ogata Y, Hirata Y, Aihara K. 2014a. Estimating the ETAS model from an early aftershock sequence. *Geophys. Res. Lett.* 41:850–57

- Omi T, Ogata Y, Hirata Y, Aihara K. 2014b. Intermediate-term forecasting of aftershocks from an early aftershock sequence: Bayesian and ensemble forecasting approaches. *J. Geophys. Res. Solid Earth* 120:2561–78
- Omori F. 1894. On the aftershocks of earthquake. *J. Coll. Sci. Imp. Univ. Tokyo* 7:111–200
- Papangelou F. 1972. Integrability of expected increments of point processes and related random change of scale. *Trans. Am. Math. Soc.* 165:483–506
- Papazachos BC. 1974. On certain aftershock and foreshock parameters in the area of Greece. *Ann. Geophys.* 27:497–515
- Parzen E, Tanabe K, Kitagawa G, eds. 1998. *Selected Papers of Hirotugu Akaike*. Tokyo: Springer-Verlag
- Rathbun SL. 1993. Modeling marked spatio-temporal point patterns. *Bull. Int. Stat. Inst.* 55:379–96
- Reasenberg PA, Jones LM. 1989. Earthquake hazard after a main shock in California. *Science* 243:1173–76
- Rhoades DA, Evison FF. 2004. Long-range earthquake forecasting with every earthquake a precursor according to scale. *Pure Appl. Geophys.* 161(1):47–72
- Ringdal F. 1975. On the estimation of seismic detection thresholds. *Bull. Seismol. Soc. Am.* 65:1631–42
- Savage MK, de Polo DM. 1993. Foreshock probabilities in the western Great-Basin eastern Sierra Nevada. *Bull. Seismol. Soc. Am.* 83:1910–38
- Schoenberg FP. 2004. Testing separability in spatial-temporal marked point processes. *Biometrics* 60:471–81
- Scholz C. 1968. The frequency-magnitude relation of microfracturing in rock and its relation to earthquakes. *Bull. Seismol. Soc. Am.* 58(1):399–415
- Schorlemmer D, Zechar J, Werner M, Field E, Jackson D, et al. 2010. First results of the Regional Earthquake Likelihood Models experiment. *Pure Appl. Geophys.* 167(8–9):859–76
- Sekiya H. 1976. The seismicity preceding earthquakes and its significance to earthquake prediction. *Zisin II* [J. Seismol. Soc. Jpn.] 29:299–312
- Smith WD. 1986. Evidence for precursory changes in the frequency magnitude b -value. *Geophys. J. Int.* 86:815–38
- Smyth C, Mori J. 2011. Statistical models for temporal variations of seismicity parameters to forecast seismicity rates in Japan. *Earth Planets Space* 63:231–38
- 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. Am.* 56:185–200
- 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 (I): some parameters which characterize an aftershock sequence and their interaction. *J. Fac. Sci. Hokkaido Univ., Ser. 7*, 3:129–95
- 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. 1971. Aftershock and earthquake statistic (III): analyses of the distribution of earthquakes in magnitude, time and space with special consideration to clustering characteristics of earthquake occurrence (1). *J. Fac. Sci. Hokkaido Univ., Ser. 7*, 3:379–441
- Utsu T. 1972. Aftershocks and earthquake statistics (IV): analyses of the distribution of earthquakes in magnitude, time, and space with special consideration to clustering characteristics to earthquake occurrence (2). *J. Fac. Sci. Hokkaido Univ., Ser. 7*, 3:379–441
- Utsu T. 1975. Correlation between shallow earthquakes in Kanto region and intermediate earthquakes in Hida region, central Japan. *Zisin II* [J. Seismol. Soc. Jpn.] 28:303–11
- Utsu T. 1977. Probability in earthquake prediction. *Zisin II* [J. Seismol. Soc. Jpn.] 30:179–85
- Utsu T. 1978. An investigation into the discrimination of foreshock sequences from earthquake swarms. *Zisin II* [J. Seismol. Soc. Jpn.] 31:129–35
- 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.* 21:164–66
- Utsu T. 1984. Estimation of parameters for recurrence models of earthquakes. *Bull. Earthq. Res. Inst. Univ. Tokyo* 59:53–66

- Utsu T. 1999a. Representation and analysis of the earthquake size distribution: a historical review and some new approaches. *Pure Appl. Geophys.* 155(2):509–35
- Utsu T. 1999b. *Seismicity Studies: A Comprehensive Review*. Tokyo: Univ. Tokyo Press
- Utsu T. 2002. Relationships between magnitude scales. In *International Handbook of Earthquake and Engineering Seismology*, Vol. 81, Part A, ed. WHK Lee, H Kanamori, PC Jennings, C Kisslinger, pp. 733–46. London/San Diego: Academic/Elsevier
- Utsu T, Ogata Y, Matsu'ura RS. 1995. The centenary of the Omori formula for a decay law of aftershock activity. *J. Phys. Earth* 43:1–33
- Varotsos P, Alexopoulos K, Nomicos K, Lazaridou M. 1986. Earthquake predictions and electric signals. *Nature* 322:120
- Vere-Jones D. 1970. Stochastic models for earthquake occurrence. *J. R. Stat. Soc. B* 32:1–62
- Vere-Jones D. 1978. Earthquake prediction—a statistician's view. *J. Phys. Earth* 26:129–46
- Vere-Jones D. 1999. Probabilities and information gain for earthquake forecasting. *Comput. Seismol.* 30:248–63
- Vere-Jones D, Davies RB. 1966. A statistical survey of earthquakes in the main seismic region of New Zealand: Part 2—Time series analyses. *N. Zealand J. Geol. Geophys.* 9:251–84
- von Seggern D, Alexander SS, Baag CE. 1981. Seismicity parameters preceding moderate to major earthquakes. *J. Geophys. Res.* 86(B10):9325–51
- Wang T, Zhuang J, Kato T, Bebbington M. 2013. Assessing the potential improvement in short-term earthquake forecasts from incorporation of GPS data. *Geophys. Res. Lett.* 40:2631–35
- Whittle P. 1962. Gaussian estimation in stationary time series. *Bull. Int. Stat. Inst.* 39:105–29
- Wiemer S, Wyss M. 1997. Mapping the frequency-magnitude distribution in asperities: an improved technique to calculate recurrence times? *J. Geophys. Res.* 102(B7):15115–28
- Woessner J, Hainzl S, Marzocchi W, Werner MJ, Lombardi AM, et al. 2011. A retrospective comparative forecast test on the 1992 Landers sequence. *J. Geophys. Res.* 116:B05305
- Wong KC, Wyss M. 1985. Clustering of foreshocks and preshocks in the Circum-Aegean region. *Earthq. Predict. Res.* 1:121–40
- Wyss M, Shimazaki K, Wiemer S. 1997. Mapping active magma chambers by b values beneath the off-Ito volcano, Japan. *J. Geophys. Res.* 102(B9):20413–22
- Yamashina K. 1981a. A method of probability prediction for earthquakes in Japan. *J. Phys. Earth* 29:9–22
- Yamashina K. 1981b. Some empirical rules on foreshocks and earthquake prediction. In *Earthquake Prediction: An International Review*, ed. DW Simpson, PG Richards, pp. 517–26. Washington, DC: Am. Geophys. Union
- Zhuang J, Jiang C. 2012. Scoring annual earthquake predictions in China. *Tectonophysics* 524:155–64
- Zhuang J, Ogata Y, Vere-Jones D. 2002. Stochastic declustering of space-time earthquake occurrences. *J. Am. Stat. Assoc.* 97:369–80
- Zhuang J, Ogata Y, Vere-Jones D, Ma L, Guan H. 2014. 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 *Seismic Imaging, Fault Rock Damage and Healing*, ed. Y Li, pp. 351–76. Berlin/Boston: de Gruyter/High. Educ. Press
- Zhuang J, Vere-Jones D, Guan H, Ogata Y, Ma L. 2005. Preliminary analysis of observations on the ultra-low frequency electric field in a region around Beijing. *Pure Appl. Geophys.* 162:1367–96

Contents

Researching the Earth—and a Few of Its Neighbors <i>Susan Werner Kieffer</i>	1
The Fascinating and Complex Dynamics of Geyser Eruptions <i>Shaul Hurwitz and Michael Manga</i>	31
Plant Evolution and Climate Over Geological Timescales <i>C. Kevin Boyce and Jung-Eun Lee</i>	61
Origin and Evolution of Water in the Moon's Interior <i>Erik H. Hauri, Alberto E. Saal, Miki Nakajima, Mahesh Anand, Malcolm J. Rutherford, James A. Van Orman, and Marion Le Voyer</i>	89
Major Questions in the Study of Primate Origins <i>Mary T. Silcox and Sergi López-Torres</i>	113
Seismic and Electrical Signatures of the Lithosphere–Asthenosphere System of the Normal Oceanic Mantle <i>Hitoshi Kawakatsu and Hisashi Utada</i>	139
Earth's Continental Lithosphere Through Time <i>Chris J. Hawkesworth, Peter A. Cawood, Bruno Dhuime, and Tony I.S. Kemp</i>	169
Aerosol Effects on Climate via Mixed-Phase and Ice Clouds <i>T. Storelvmo</i>	199
Hydrogeomorphic Ecosystem Responses to Natural and Anthropogenic Changes in the Loess Plateau of China <i>Bojie Fu, Shuai Wang, Yu Liu, Jianbo Liu, Wei Liang, and Chiyuan Miao</i>	223
Interface Kinetics, Grain-Scale Deformation, and Polymorphism <i>S.J.S. Morris</i>	245
Back-Projection Imaging of Earthquakes <i>Eric Kiser and Miaki Ishii</i>	271
Photochemistry of Sulfur Dioxide and the Origin of Mass-Independent Isotope Fractionation in Earth's Atmosphere <i>Shubei Ono</i>	301
Southeast Asia: New Views of the Geology of the Malay Archipelago <i>Robert Hall</i>	331

Forming Planets via Pebble Accretion <i>Anders Johansen and Michiel Lambrechts</i>	359
Tungsten Isotopes in Planets <i>Thorsten Kleine and Richard J. Walker</i>	389
Shape, Internal Structure, Zonal Winds, and Gravitational Field of Rapidly Rotating Jupiter-Like Planets <i>Keke Zhang, Dali Kong, and Gerald Schubert</i>	419
Effects of Partial Melting on Seismic Velocity and Attenuation: A New Insight from Experiments <i>Yasuko Takei</i>	447
Origin and Evolution of Regional Biotas: A Deep-Time Perspective <i>Mark E. Patzkowsky</i>	471
Statistics of Earthquake Activity: Models and Methods for Earthquake Predictability Studies <i>Yosihiko Ogata</i>	497
Tectonic Evolution of the Central Andean Plateau and Implications for the Growth of Plateaus <i>Carmala N. Garzio, Nadine McQuarrie, Nicholas D. Perez, Todd A. Ehlers, Susan L. Beck, Nandini Kar, Nathan Eichelberger, Alan D. Chapman, Kevin M. Ward, Mihai N. Ducea, Richard O. Lease, Christopher J. Poulsen, Lara S. Wagner, Joel E. Saylor, George Zandt, and Brian K. Horton</i>	529
Climate and the Pace of Erosional Landscape Evolution <i>J. Taylor Perron</i>	561
The Rise of Animals in a Changing Environment: Global Ecological Innovation in the Late Ediacaran <i>Mary L. Droser, Lidya G. Tarhan, and James G. Gehling</i>	593
The Late Heavy Bombardment <i>William F. Bottke and Marc D. Norman</i>	619
Reconstructing Climate from Glaciers <i>Andrew N. Mackintosh, Brian M. Anderson, and Raymond T. Pierrehumbert</i>	649
Autogenic Sedimentation in Clastic Stratigraphy <i>Elizabeth A. Hajek and Kyle M. Straub</i>	681

Errata

An online log of corrections to *Annual Review of Earth and Planetary Sciences* articles
may be found at <http://www.annualreviews.org/errata/earth>