Space-Time Point Process Modeling of Seismicity

Yosihiko Ogata, Koichi Katsura and Masaharu Tanemura
The Institute of Statistical Mathematics
4-6-7 Minami-Azabu, Minato, Tokyo 106-8569, Japan
e-mail ogata@ism.ac.jp

1. Motivation

The seismic gap of the 2nd kind is the region in which seismic activity become quiet before a large earthquake takes place. Kanamori (1981) reviewed the studies of the seismic gaps showing some examples of space-time patterns illustrating the gaps, and he hypothesized their physical mechanism. However, Lomnitz and Nava (1983) criticize that the gaps are merely owing to the reduction of aftershocks of the previous large earthquakes. They simulated space-time Poisson cluster process to indicate seeming seismic gaps, and claimed that there is no information in seismic gaps about the time of the occurrence nor the magnitude of the next event.

We have been interested in these papers and the followed debates. If we restrict ourselves to the case of temporal changes of seismic activity, we can well predict the expected occurrence rate of earthquakes in normal sequences by using the Epidemic Type Aftershock-Sequences (ETAS) model (Ogata, 1988). Comparing the predicted rate with that of observed occurrence data, periods of relatively decreased or increased seismic activity can be recognized (Ogata, 1988, 1992), which can be clearly seen by the time-change based on the integration of the estimated intensity function. In this paper we are concerned with the space-time extension of this serious problem relating to earthquake prediction.

2. Space-time point processes and the dataset

The *conditional intensity function* of a space-time point process is defined as the occurrence rate at time t and the location (x, y) conditional on the past history of the occurrences such that

$$\lambda(t, x, y|\, \mathcal{H}_t) \equiv \frac{E\{N(dt, dx, dy)|\, \mathcal{H}_t)\}}{dt \cdot dx \cdot dy} = \frac{P\{an\ event\ in\ (dt, dx, dy)|\ History\}}{dt \cdot dx \cdot dy}$$

where $\mathcal{H}_t = \{(t_i, x_i, y_i, M_i); t_i < t\}$ is the history of occurrence times $\{t_i\}$ up to time t, with the corresponding epicenters $\{(x_i, y_i)\}$ and magnitudes $\{M_i\}$. Given the origin times and space coordinates of earthquakes with their magnitudes $\{(t_i, x_i, y_i, M_i);\ M_i \geq M_0,\ i = 1, \cdots, n\}$ during a period $[0, T]$ and in a region A, the log likelihood of the model is given by

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log \lambda_{\boldsymbol{\theta}}(t_i, x_i, y_i|\ H_t) - \int_0^T \int \int_A \lambda_{\boldsymbol{\theta}}(t, x, y|\ H_t) dt dx dy. \quad (1)$$

We use the Hypocenter Data File of Japan Meteorological Agency (JMA) as the source of the data in this study. We select the data of earthquakes of magnitude (M) 5.0 or larger with depths shallower than 100 km throughout whole Japan (within the rectangular region bounded by 128°E and 149°E meridians, and 26°N and 47°N parallels) for the period from 1926 through 1995.

3. Extension of the ETAS model to the space-time seismicity

We assume that tectonic seismic activity is given by a superposition of earthquake clusters triggered by relatively large earthquakes, so that the conditional intensity function is expressed by

$$\lambda(t, x, y|\, \mathcal{H}_t) = \mu(x, y) + \sum_{\{j:\ t_j < t\}} \nu(t, x, y|\ t_j, x_j, y_j; M_j),$$

where $\mu(x,y)$ represents background seismicity and $\nu(t,x,y|\ t_j,x_j,y_j;M_j)$ is the rate of aftershocks ($M \geq M_c$) at space-time coordinate (t,x,y) following j-th earthquake of (t_j,x_j,y_j,M_j). We further assume that the activity is stationary and the space-time transfer function ν is factorized to be the products of functions of size, time and space. Furthermore, we would like to obtain the ETAS model (Ogata, 1988) by integrating the space-time conditional intensity function with respect to the location variable (x,y). These requirements are satisfied by the following form:

$$\begin{aligned}
\nu(t,x,y \mid t_j,x_j,y_j;M_j) &= \nu(t-t_j,\ x-x_j,\ y-y_j|\ M_j) \\
&= \kappa(M_j)\cdot g(t-t_j|\ M_j)\cdot f(x-x_j,\ y-y_j|\ M_j) \\
&= \frac{K_0 e^{\alpha(M_j-M_c)}}{(t-t_j+c)^p}\cdot f(x-x_j,\ y-y_j|\ M_j)
\end{aligned}$$

Thus the remaining aspect for modeling is the parametric form of the response function f. Owing to the empirical studies by Utsu (1969) the function should take the following common form

$$f(x-x_j,\ y-y_j|\ M_j) = \frac{|S_j|}{2\pi\sigma(M_j)}\ f_0\left\{(x-x_j,\ y-y_j)\ \frac{S_j}{\sigma(M_j)}\ \begin{pmatrix} x-x_j \\ y-y_j \end{pmatrix}\right\},$$

where S is a positive-definite symmetric matrix (actually dependent on the event j) for anisotropic clusters such that

$$(x,y)S\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{\sqrt{1-\rho^2}}\left(\frac{\sigma_2}{\sigma_1}x^2 - 2\rho xy + \frac{\sigma_1}{\sigma_2}y^2\right).$$

Here, Ogata (1998) compared the three possible function forms for f_0 which arise from the questions of physical interest including the selection of either short or long range decay of the spatial distributions of aftershock clusters and the question whether or not the cluster regions scale with magnitudes. Consequently, the AIC (Akaike, 1974) indicated the clear preference of the following function form

$$f(x,y|M) = \frac{(q-1)d^{q-1}|S|}{\pi\sigma(M)}\left\{\frac{(x,y)S\begin{pmatrix} x \\ y \end{pmatrix}}{\sigma(M)} + d\right\}^{-q}.$$

Here we have parameters $\boldsymbol{\theta} = (\mu, K_0, c, \alpha, p, d, q)$ to maximize the log likelihood function in (1).

4. Modeling of spatially heterogeneous seismicity

However, as the data size increases by lowering the threshold magnitude, or as the geophysical region get wider, each characteristic parameter in $\boldsymbol{\theta} = (\mu, K_0, c, \alpha, p, d, q)$ tends to take significantly different value from place to place. Therefore we have to further consider that each parameter should be a function of location. Thus we extend the previous space-time ETAS model to a hierarchical model which may be called as the *Hierarchical Space-Time ETAS model*; in short *HIST-ETAS model*.

Specifically, we assume that the parameters except for c and d are the functions of location (x,y) as follows. Given all locations of earthquakes (x_i,y_i), $i=1,2,\cdots,N$, consider the Delaunay triangulation of the region A tessellated based on the epicenter locations and some additional points (say, M points) in the boundaries including the corners. Then, consider the parameter function $\mu(x,y)$ to be 2-dimensional piecewise linear, namely, the function consisting of facets defined on the tessellated triangles where, for each vertex of a triangular (x_i,y_i), it takes the value $\mu_i = \mu(x_i,y_i)$ for each $i = 1,2,\cdots,N{+}M$. In the same manner the other parameter functions $K_0(x,y)$, $\alpha(x,y)$, $p(x,y)$ and $q(x,y)$ are piecewise linearly expanded by the parameters $K_{0,i}$, α_i, p_i and q_i for $i = 1,2,\cdots,N{+}M$. This modeling using Delaunay tessellation is suited for the observations on clustered locations regardless the dimemsion of the space: later we consider the three dimensional version. That is to say, we can see detailed changes in the region where the observations are densely populated while smoother changes are expected in the region of sparsely populated.

The number of unknown parameters is actually about five times as many as the number of data events. Therefore we need to formulate the penalized log likelihood (Good and Gaskins, 1971)

$$Q(\theta \mid w_\mu, w_K, w_\alpha, w_p, w_q) = \log L(\boldsymbol{\theta}) - penalty(\boldsymbol{\theta} \mid w_\mu, w_K, w_\alpha, w_p, w_q)$$

to get the unique solution for the parameters $\boldsymbol{\theta} = \{(\mu_i, K_{0,i}, \alpha_i, p_i, q_i);\ i = 1, 2, \cdots, N{+}M\}$, or the maximum a posterior (MAP) solution, where $\log L(\boldsymbol{\theta})$ stands the log likelihood in (1) and the penalty function is equal to

$$\int\int_A dxdy \Big\{ w_\mu(\mu_x^2 + \mu_y^2) + w_K(K_x^2 + K_y^2) + w_\alpha(\alpha_x^2 + \alpha_y^2) + w_p(p_x^2 + p_y^2) + w_q(q_x^2 + q_y^2) \Big\},$$

and μ_x and μ_y, for example, denote partial derivative of the function $\mu(x, y)$ with respect to the variables x and y, respectively. The penalized log-likelihood takes the trade-off between the good fit to the data and the uniformity of the parameter functions. The optimal tuning of the weight is objectively carried out by the Bayesian method suggested by Good (1965) and Akaike (1979). Since the above defined penalty function is quadratic with respect to the coefficients, the prior distribution ($\propto \exp\{-penalty\}$) is a multivariate Normal distribution. Therefore, the Normal approximation of the posterior function is available to calculate the integrated likelihood, or likelihood of the Bayesian model, in order to find the optimal hyper-parameters (the weights, $\hat{w}_1, \cdots, \hat{w}_5$) which maximize the likelihood (e.g., Ogata and Katsura, 1988, 1993; and Ogata et al., 1991, 2000).

Before carrying out the estimation of the HIST-ETAS model that needs the optimization of the five weights (hyperparameters), we fitted the simpler model of the non-homogeneous spatial Poisson intensity to the point pattern of epicenters, where we need to tune only one hyperparameter. Then the estimated spatial intensity function with the optimal MAP is obtained.

Eventually, we obtained the optimal weights for the HIST-ETAS model, and then the corresponding MAP estimates of each parameter function using some numerical iteration procedure for nonlinear optimization using Incomplete Cholesky Conjugate Gradient method. Thus, by applying the HIST-ETAS model, we have succeeded in discriminating the background activity and the clustering effect which can be seen by comparing the estimates $\hat{\mu}(x, y)$ and $\hat{K}_0(x, y)$ with the MAP intensity estimate for spatial Poisson process. The range of the MAP $\hat{p}(x, y)$ is $0.99 \sim 1.25$, which is consistent with our experience in p-value estimates of aftershock sequences.

5. Space-time residual analysis of point processes

Let us denote the above estimated conditional intensity function of the HIST-ETAS model with MAP coefficients by $\hat{\lambda}(t, x, y \mid H_t)$. In order to detect the temporal change of the seismicity relative to the estimated model, we consider the *residual function* $\xi(t, x, y)$ in such a way that the new space-time intensity function

$$\eta(t, x, y) = \hat{\lambda}(t, x, y \mid H_t)\ \exp\{\xi(t, x, y)\}$$

is defined to apply the same data as in the previous sections. The estimated model $\hat{\lambda}(t, x, y \mid H_t)$ is shown to be a good-fit in space-time volumes where $\xi(t, x, y) \approx 0$ hold. However, we are particularly interested in the volumes where $\xi(t, x, y) < 0$ hold significantly, which may be called as the *relative seismic gaps*.

In order to estimate $\xi(t, x, y)$, the whole three-dimensional volume $[0, T] \times A$ is divided into the Delaunay tetrahedra (e.g., Tanemura et al. 1983) whose vertices consist of the coordinates of earthquake data $\{(t_i, x_i, y_i)\}$ and some additional points in the boundaries, which are associated with unknown parameter ξ_i to be estimated. Then, consider 3-dimensional piecewise linear function $\xi(t, x, y; \boldsymbol{\theta})$ defined on the tessellated volume where, for each vertex of a tetrahedron (t_i, x_i, y_i), the value ξ_i is attached, namely $\xi_{\boldsymbol{\theta}}(t_i, x_i, y_i) = \xi_i$. If the conditional intensity $\lambda_{\boldsymbol{\theta}}$ in the log likelihood (1) is replaced by the intensity $\eta_{\boldsymbol{\theta}}$ with $\boldsymbol{\theta} = (\xi_i)$, we can define the log likelihood function of $\boldsymbol{\theta}$.

Thus, for the flatness constraint of parameters, we consider the penalty function

$$penalty(\boldsymbol{\theta}) = \int_0^T \int \int_A \left\{ w_1 \xi_t^2 + w_2 \left(\xi_x^2 + \xi_y^2 \right) \right\} dt dx dy,$$

where ξ_t, ξ_x and ξ_y are partial derivative of the function $\xi(t, x, y; \boldsymbol{\theta})$ with respect to the variables t, x and y, respectively. Since the penalty function is quadratic with respect to the parameters $\boldsymbol{\theta} = (\xi_i)$, the prior distribution ($\propto \exp\{-penalty\}$) is multivariate Normal distribution. Therefore, the Normal approximation of the posterior function is available to calculate the integrated likelihood in order to find the optimal hyperparameters (the weights, w_1, w_2) which maximize the integrated likelihood.

Thus obtained three dimensional image $\xi(t, x, y; \hat{\boldsymbol{\theta}})$ is displayed by means of the AVS, from which we can see some space-time subvolumes of $\xi(t, x, y; \hat{\boldsymbol{\theta}}) < 0$ before some great earthquakes, which includes the source region of the event. In particular, we are interested in the most active seismic area, around off the east coast of Tohoku District. The cross-sectional residual image at a certain particular longitude shows some lowering of the rate before the great events in the latitude range with enough number of earthquakes in which enough precision of the image is expected. Those hollows show space-time gaps relative to the predicted seismic activity by the estimated HIST-ETAS model. This is encouraging result in that we may hope to forecast the region and time of forthcoming great earthquakes using abundant space-time data of the smaller earthquakes.

In summary, the HIST-ETAS model and the proposed residual analysis are useful for measuring characteristic of seismic activity and also for the space-time forecasting of the large earthquakes. The interested readers can referred to the full paper submitted elsewhere.

REFERENCES

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

Akaike, H. (1979) Likelihood and Bayes procedure, in *Bayesian Statistics* (Bernard et al., eds.), University Press, Valencia, Spain.

Good, I.J. (1965) *The Estimation of Probabilities*, M.I.T. Press, Cambridge, Massachusetts.

Good, I.J. and R.A. Gaskins (1971) Nonparametric roughness penalties for probability densities, *Biometrika* **58**, 255-277.

Kanamori, H. (1981) The nature of seismicity patterns before large earthquakes, in *Earthquake Prediction, Maurice Ewing Series*, **4**, D. Simpson and P. Richards, Editors, Am. Geophys. Union, Washington D.C., pp. 1-19.

Lomnits, C. and F.A. Nava (1983) The predictive value of seismic gaps, *Bull. Seism. Soc. Am.*, **73**, pp. 1815-1824.

Ogata, Y. (1988) Statistical models for earthquake occurrences and residual analysis for point processes, *J. Amer. Statist. Assoc.*, **83**, No. 401, 9-27.

Ogata, Y. (1992) Detection of precursory seismic quiescence before major earthquakes through a statistical model, *J. Geophys. Res.*, **97**, 19845-19871.

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

Ogata, Y. and K. Katsura (1988) Likelihood analysis of spatial inhomogeneity for marked point patterns, *Ann. Inst. Statist. Math.* **40**, pp. 29-39.

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

Ogata, Y. and K. Katsura (1993) Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogues, *Geophys. J. Int.* **113**, pp. 727-738.

Ogata, Y., Koichi Katsura, Niels Keiding, Claus Holst and Anders Green (2000), Empirical age-period-cohort analysis of retrospective incidence data, *Scand. J. Statist.*, **27**, pp. 415-432.

Tanemura, M., T. Ogawa, and N. Ogita (1983), A new algorithm for three-dimensional Voronoi tessellation, *J. Comput. Phys.* **51**, pp. 191-207.

Utsu, T. (1969) Aftershocks and earthquake statistics (I): some parameters which characterize an aftershock sequence and their interaction, *J. Faculty Sci., Hokkaido Univ., Ser. VII (geophysics)* **3**, 129-195.