Comment on "A Review of Self-Exciting Spatiotemporal Point Process and Their Applications" by Alex Reinhart

Yosihiko Ogata

Abstract. In my discussion, I would like to comment on our early reactions to Hawkes' enlightening paper on the self-exciting model; further, I would like to comment on developments of the extended models with some applications.

Key words and phrases: Akaike Bayesian Information Criterion (ABIC), Akaike information criterion (AIC), causality analysis, conditional intensity function, empirical Bayesian method, epidemic-type aftershock sequence (ETAS) model, hierarchical space-time ETAS (HIST-ETAS) model, maximum-likelihood method, penalized log-likelihood, statistical seismology, study of earthquake predictability, thinning simulation method.

1. INTRODUCTION

This review paper by Mr. Reinhart has provided me with an opportunity to explicitly explore the 30 years of developments attained in the field. On the other hand, I am aware that the Hawkes-type process is also becoming more frequently cited in the fields of financial transactions and insurance. Anyhow, I am pleased to learn that some models and methods developed for analyzing earthquake data in statistical seismology are also useful in other scientific fields. I personally believe that the most appreciable contribution for a statistician would be one that expands a range of applications of statistics in the fields of science and technology. In fact, serious application issues in a particular field could be sources of new models and methods in statistical science.

In my discussion, I would like to comment on our early reactions to Hawkes' enlightening paper on the self-exciting model; further, I would like to comment on developments of the extended models with some applications.

Yosihiko Ogata is Emeritus Professor, The Institute of Statistical Mathematics, Midori-Cho 10-3, Tachikawa city, Tokyo, Japan 190-8562 (e-mail: ogata@ism.ac.jp).

2. LIKELIHOOD-BASED INFERENCE AND FORECASTING

In 1976, at the Institute of Statistical Mathematics (ISM) in Tokyo, David Vere-Jones provided us with a series of special lectures on point processes. The time-series research group in the ISM that was led by Hirotugu Akaike then discovered a new fertile ground of developing likelihood-based estimation of the conditional intensity function that predicted the short-term occurrence rate of events. Therefore, we devoted considerable attention to the Hawkes' self-exciting model because its conditional intensity function takes a linear regressive form of the past occurrence times of events; namely, in a discretized time axis, the corresponding conditional probabilities of binary series of events take a linear regression to the configuration of 0–1 series of the occurrence of past events.

In the estimation of point process models and goodness-of-fit comparisons between different models, Hawkes and his student (Hawkes and Adamopoulos, 1973, Adamopoulos, 1976) calculated the spectral likelihood (Whittle, 1951), which is also applicable to the trigger model of Vere-Jones (1970) that is a special case of the Neyman–Scotts' model. However, such estimates based on the second-order moment are not sufficiently accurate or efficient for point-process data, compared with the maximizing log-likelihood

estimates of regular point processes as discussed by Rubin (1972). Accordingly, we were interested in the maximum-likelihood method to obtain the Akaike information criterion (AIC; see the library edition edited by Parzen, Tanabe and Kitagawa, 1998) to select and estimate the response functions of the Hawkes-type models using the Laguerre polynomials (Ozaki, 1979, Vere-Jones and Ozaki, 1982, Ogata and Akaike, 1982). These were computed by an efficient quasi-Newton optimization method such as the Davidon–Fletcher–Powell method (e.g., Fletcher and Powell, 1963/1964).

3. MODELING FOR THE CAUSALITY ANALYSIS

One of the missing but an important aspect, in Mr. Reinhart's review paper, would be the modeling of point processes for causality analysis from a stochastic process including another point process. In Akaike's research group (see the library edition of Akaike's papers edited by Parzen, Tanabe and Kitagawa, 1998), the causality analysis of multi-variate time series is an important subject in system identification, prediction and optimal control using multivariate autoregressive models. Thus, we have extended the self-exciting and mutually exciting point process models in various ways; namely, in the case of temporal point processes, they take the following general form:

$$(1) \quad \lambda(t|H_t, F_t) = \mu(t|F_t) + \sum_{t_i < t} g(t - t_i|M_i),$$

where H_t is history of occurrence times of the target point process, F_t is the history of outer information inputs and M_i represents a mark associated with the ith event.

Such models include trends, cyclic or seasonal, components in the presence of clustering effects and causal relationships of a target point process from another stochastic process (Ogata and Akaike, 1982, Ogata, Akaike and Katsura, 1982, Ogata and Katsura, 1986, Ogata, 1999 and references therein). Indeed, the models for the causality have become increasingly required in many scientific fields. For example, in seismology, some earthquake causality can be understood due to conveyed stress changes from a nearby fault rupture. We may use the cross-correlation statistic between two stationary point process data sets $\{t_i\}$ and $\{u_j\}$, which is quantified by the cross Palm intensity, realized by the superposition of point configurations $\{t_i - u_j\}$ for all i and j (Cox and Lewis, 1966). However, in the analysis using cross-correlation, it is difficult to identify the causality relationship in the presence of clustering events. Conventional de-clustering of the data

sets may avoid this difficulty, but the results depend on the adopted de-clustering method. Furthermore, a significant cross-correlation does not discriminate among the following four cases: (i) $\{t_i\}$ causes $\{u_j\}$; (ii) $\{u_j\}$ causes $\{t_i\}$; (iii) $\{u_j\}$ and $\{t_i\}$ cause each other and (iv) some hidden process causes the both $\{u_j\}$ and $\{t_i\}$. Therefore, to discriminate among (i)–(iv) in the presence of clustering events, we have used a mixture of a self-exciting process and a doubly stochastic process

$$\lambda(t|H_t, F_t) = \mu + P_J(t)$$
$$+ \sum_{t_i < t} g_L(t - t_i) + \sum_{u_j < t} h_M(t - u_j)\xi_j,$$

where $P_J(t)$ includes a trend and a cyclic function, where g_L and h_M is a Laguerre polynomial of order L and M, respectively.

Now, so far, warning-type large earthquake prediction techniques have been proposed based on various seismicity anomalies, whereas the effectiveness of these techniques is controversial (see Jordan et al., 2011, and references therein). Therefore, objective evaluation methods for the predictions are required for such techniques. Namely, we should provide statistical significance of a causality, and also assessment of probability gains against a standard reference seismicity model in a target area (e.g., Nishizawa, Lei and Nagao, 1994, Zhuang et al., 2005, 2014, Han et al., 2016). In this respect, the causality model can be used to investigate the statistical relation between geophysical anomalies, as suspected precursor signals, and large earthquakes. In future, to explore and evaluate the intensity distribution of the location and time of large earthquakes by monitoring the precursory signals, it will be quite ambitious to develop successful space-time extensions of the models (1), for example, using the space-time geodetic records of the Global Navigation Satellite Systems.

Ogata (1988) explored various extensions (1) of the Hawkes self-exciting models for an earthquake time series associated with earthquake magnitudes; and examined the possible response functions $g(\cdot)$ in equation (1) looking for the standard seismicity model based on statistical empirical laws of aftershock sequences studied and established in statistical seismology (see Utsu, Ogata and Matsu'ura, 1995). In particular, for statistically establishing the epidemic-type aftershock sequence (ETAS) model as the standard model for seismic activity, I adopted the AIC for model selections, and examined various residual analyses for point processes.

Incidentally, Ross (2016), referred to in Section 3.3 in Reinhart review paper, claimed that the log-likelihood of the ETAS model required an $O(n^2)$ operation, but Ogata, Matsu'ura and Katsura (1993) provided an accurate and fast approximation of the $O(n)$ operation by making use of a recursive structure (Ogata and Akaike, 1982) and double exponential integration approximation for numerical integration.

Further extension to a particular nonstationary ETAS model and applications by Kumazawa and Ogata (2014) and Kumazawa et al. (2016) successfully provided accurate quantitative geophysical relationship between changing swarm-type activity and stress change or fault-weakening caused by the migration or intrusion of fluid (e.g., Ogata and Katsura, 1986, Hainzl and Ogata, 2005).

4. SPACE-TIME ETAS MODELS AND BAYESIAN METHOD

The characteristic parameters of the space-time ETAS models (Ogata, 1998) have been extended to be the one with location-dependent coefficients, and thus describe the spatial heterogeneity of seismic activity (Ogata et al., 2003). This model is named the hierarchical space-time ETAS (HIST-ETAS) model. The HIST-ETAS model and other variants of the space-time ETAS models (see Ogata, 1998, 2011 and the respective papers in the same special issue; Ogata et al., 2013) forecast earthquakes in time, space and magnitude in the entire Japanese region, and the Testing Centers of Collaboratory for the Study of Earthquake Predictability (CSEP; http://www.cseptesting.org/) have been testing their performance for more than 10 years.

Technically, to adapt to the heterogeneous seismicity patterns, the models include a few times more coefficients than the data size. For a stable optimal estimation of the models, the variability of the coefficients needs to be constrained by assigning penalties against rough changes and, therefore, use the penalized log-likelihood function (Good and Gaskins, 1971). The coefficients that maximize the penalized log-likelihood are then sought. This procedure is equivalent to attaining the maximum *a posteriori* (MAP) distribution as a tomographic inversion, but we need to adjust the optimal prior function for the strength of smoothness constraints in terms of the penalty functions using an empirical Bayesian method and Akaike Bayesian Information Criterion (ABIC; Akaike, 1980, Parzen, Tanabe and Kitagawa, 1998). I use a Laplace approximation of the posterior function at each repeating step

to attain the MAP, and then to calculate the ABIC to be minimized by a set of the respective weights of the penalties for the constraints. However, so far, Markov chain Monte Carlo (MCMC) methods including the Metropolis method have not yet proved useful in implementing the same procedure. The details of our procedure can be found in the studies of Ogata et al. (2003) and Ogata (2004, 2011).

5. SIMULATION AND INFERENCES OF POINT PROCESSES

The thinning method of Lewis and Shedler (1979) is useful for the simulation of nonstationary Poisson processes regardless of the spatial dimension. Some statisticians regarded this method as essentially the same as the traditional acceptance–rejection simulation method of sample generation from a general probability density, but an advantage of the thinning method is that the total number of the events is automatically obtained without integration of the intensity function.

Using the conditional intensity function, Ozaki (1979) simulated a Hawkes process by numerically solving the hazard equations in equation (7) in the Reinhart review paper. However, the thinning technique can be extended to general regular point processes characterized by any conditional intensity and this can be further extended to multi-variate (multi-channel) and marked point processes with general states, including the Hawkes' mutually exciting process; which can be proved by using martingale theory: refer to Ogata (1981) and Daley and Vere-Jones (2003), Section 7.5.

Incidentally, I am skeptical about the review of Wang et al. (2013) that claimed that there is bias and error in the estimate of the MLE in the inference and simulations of the ETAS models with a smaller sample size. Although the consistency and asymptotic normality hold by the assumptions of stationarity and ergodicity of the data, we have to aware that the response function $g(\cdot)$ of the ETAS model is inverse power decaying in time. Therefore, the accurate simulation data need a long enough burn-in time for stationarity. In addition, a long enough precursory time span is required before the target period for the unbiased estimation. Hence, I always take care of the latter issue when fitting the ETAS models to focal seismic activity (see, e.g., Ogata, 2013).

Finally, I would like to add my recent reviews in statistical seismology (Ogata, 2013 and 2017) written for statisticians and geophysicists, respectively.

ADDITIONAL REFERENCES

ADAMOPOULOS, L. (1976). Cluster models for earthquakes: Regional comparisons. *Math. Geol.* **8** 463–475.

AKAIKE, H. (1980). Likelihood and the Bayes procedure. In *Bayesian Statistics* (*Valencia*, 1979) 143–166. Univ. Press, Valencia. MR0638876

COX, D. R. and LEWIS, P. A. W. (1966). *The Statistical Analysis of Series of Events*. Methuen & Co., Ltd., London. MR0199942

FLETCHER, R. and POWELL, M. J. D. (1963/1964). A rapidly convergent descent method for minimization. *Comput. J.* **6** 163–168. MR0152116

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

HAINZL, S. and OGATA, Y. (2005). Detecting fluid signals in seismicity data through statistical earthquake modeling. *J. Geophys. Res.* **110** B05S07. DOI:10.1029/2004JB003247.

HAN, P., ZHUANG, J., HATTORI, K. and OGATA, Y. (2016). An interdisciplinary approach for earthquake modelling and forecasting. In 2016 *Fall Meeting of the American Geophysical Union* (*AGU*), Moscone Center, San Francisco, CA, 13 December 2016 (Oral).

HAWKES, A. G. and ADAMOPOULOS, L. (1973). Cluster models for earthquakes—Regional comparisons. *Bull. Int. Stat. Inst.* **45** 454–461.

OGATA, Y. (1999). Seismicity analyses through point-process modelling: A review. In *Seismicity Patterns*, *Their Statistical Significance and Physical Meaning* (M. Wyss, K. Shimazaki and A. Ito, eds.). *Pure and Applied Geophysics* **155** 471–507. Birkhauser, Basel.

JORDAN, T. H., CHEN, Y. T., GASPARINI, P., MADARIAGA, R. and MARZOCCHI, W. (2011). Operational earthquake forecasting: State of knowledge and guidelines for utilization. *Ann. Geophys.* **54** 315–391.

KUMAZAWA, T. and OGATA, Y. (2014). Nonstationary ETAS models for nonstandard earthquakes. *Ann. Appl. Stat.* **8** 1825–1852. MR3271355

KUMAZAWA, T., OGATA, Y., KIMURA, K., MAEDA, K. and KOBAYASHI, A. (2016). Background rates of swarm earthquakes that are synchronized with volumetric strain changes. *Earth Planet. Sci. Lett.* **442** 51–60. DOI:10.1016/j.epsl.2016.02.049.

LEWIS, P. A. W. and SHEDLER, G. S. (1979). Simulation of nonhomogeneous Poisson processes by thinning. *Nav. Res. Logist. Q.* **26** 403–413. MR0546120

NISHIZAWA, O., LEI, X. and NAGAO, T. (1994). Hazard function analysis of seismo-electric signals in Greece. In *Electromagnetic Phenomena Related to Earthquake Prediction* (M. Hayakawa and Y. Fujinawa, eds.) 459–474. Terra Publishing Company, Tokyo.

OGATA, Y. (1981). On Lewis' simulation method for point processes. *IEEE Trans. Inform. Theory* **27** 23–31.

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

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

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

OGATA, Y. (2013). A prospect of earthquake prediction research. *Statist. Sci.* **28** 521–541. MR3161586

OGATA, Y. (2017). Statistics of earthquake activity: Models and methods for earthquake predictability studies. *Annu. Rev. Earth Planet. Sci.* **45** 497–527. DOI:10.1146/annurev-earth-063016-015918.

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

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

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

OGATA, Y., MATSU'URA, R. S. and KATSURA, K. (1993). Fast likelihood computation of epidemic type aftershock-sequence model. *Geophys. Res. Lett.* **20** 2143–2146.

OGATA, Y., KATSURA, K., FALCONE, G., NANJO, K. Z. and ZHUANG, J. (2013). Comprehensive and topical evaluations of earthquake forecasts in terms of number, time, space, and magnitude. *Bull. Seismol. Soc. Am.* **103** 1692–1708.

OZAKI, T. (1979). Maximum likelihood estimation of Hawkes' self-exciting point processes. *Ann. Inst. Statist. Math.* **31** 145–155. MR0541960

PARZEN, E., TANABE, K. and KITAGAWA, G. (eds.) (1998). *Selected Papers of Hirotugu Akaike*. Springer, New York. MR1486823

RUBIN, I. (1972). Regular point processes and their detection. *IEEE Trans. Inform. Theory* **18** 547–557. MR0378096

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

VERE-JONES, D. (1970). Stochastic models for earthquake occurrence. *J. Roy. Statist. Soc. Ser. B* **32** 1–62. MR0272087

VERE-JONES, D. and OZAKI, T. (1982). Some examples of statistical inference applied to earthquake data. *Ann. Inst. Statist. Math.* **34** 189–207.

WANG, T., ZHUANG, J., KATO, T. and BEBBINGTON, M. (2013). Assessing the potential improvement in short-term earthquake forecasts from incorporation of GPS data. *Geophys. Res. Lett.* **40** 2631–2635.

WHITTLE, P. (1951). *Hypothesis Testing in Times Series Analysis*. Almqvist & Wiksells Boktryckeri AB, Uppsala.

ZHUANG, J., VERE-JONES, D., GUAN, H., OGATA, Y. and MA, L. (2005). Preliminary analysis of observations on the ultra-low frequency electric field in a region around Beijing. *Pure Appl. Geophys.* **162** 1367–1396. DOI:10.1007/s00024-004-2674-3.

ZHUANG, J., OGATA, Y., VERE-JONES, D., MA, L. and 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* (Y. Li, ed.) 351–376. De Gruyter, Berlin together with Higher Education Press, Beijing.