# How robust is the skill score of probabilistic earthquake forecasts?

## *Sulla robustezza della stima delle performance predittive di modelli sismologici probabilistici*

Alessia Caponera and Maximilian J. Werner

**Abstract** Earthquake scientists continue to improve models of the spatio–temporal evolution of seismicity, including complex aftershock sequences. The Collaboratory for the Study of Earthquake Predictability (CSEP) prospectively evaluates the predictive skill of probabilistic forecasts by such models. Here, we assess the robustness of one popular skill score, the information gain per earthquake, with respect to temporal fluctuations of the seismicity rate. We conduct a numerical experiment with a widely-used temporal stochastic seismicity model, a special case of Hawkes process. Our simulations reveal that the information gain fluctuates substantially with time, because a central limit theorem does not hold in a realistic parameter regime. Our results may eventually contribute to more robust inferences.

**Abstract** *Gli scienziati della terra propongono modelli probabilistici sempre più sofisticati per descrivere l'evoluzione spazio–tempo dei terremoti. Il CSEP (Collaboratory for the Study of Earthquake Predictability) stima prospettivamente le performance predittive di tali modelli. Qui, viene valutata l'incertezza relativa a uno stimatore utilizzato da CSEP, l'information gain per earthquake. Viene condotto un esperimento numerico con un noto modello temporale di sismicità, caso particolare di processo di Hawkes. Le simulazioni effettuate rivelano che, per valori realistici dei parametri, l'information gain mantiene una variabilità elevata nel tempo. I risultati possono contribuire a rendere le conclusioni inferenziali più robuste.*

**Key words:** Earthquake forecasting, temporal Hawkes processes, information gain per earthquake

---

Alessia Caponera
Department of Statistical Sciences, Sapienza University, Rome, Italy
London Mathematical Laboratory, London, UK
e-mail: alessia.caponera@uniroma1.it

Maximilian J. Werner
School of Earth Sciences and Cabot Institute, University of Bristol, Bristol, UK
London Mathematical Laboratory, London, UK
e-mail: max.werner@bristol.ac.uk

# 1 Introduction

Over the last decade, stochastic and physics-based models of seismicity have matured to sophisticated system-specific forecast models that can reliably forecast the evolution of seismicity, including complex aftershock sequences. The international Collaboratory for the Study of Earthquake Predictability (CSEP) provides independent and prospective evaluations of model forecasts, and thereby aims to support robust inferences about the performances of models and guide model development [8]. One challenge is the lack of data: large earthquakes are rare, especially at the regional scales. In addition, seismicity fluctuates over orders of magnitude because earthquakes cluster in space on faults and in time during aftershock sequences.

Because short-term earthquake forecasts are now starting to inform decision-making of societal relevance, there is an urgent need to understand quantitatively the robustness of performance metrics. Here, we focus on a popular measure of the relative predictive skill of two models: the information gain per earthquake.

Despite the information gain's growing importance, its robustness has not been studied in detail. Using simulations from a popular model of clustered seismicity, we show that the information gain suffers from substantial fluctuations, because a central limit theorem does not hold under realistic parameters. Our ultimate goal is to make CSEP model inferences more robust by providing guidelines for the uncertainty in the information gain.

# 2 The Information Gain per Earthquake

The information gain per earthquake is defined as the log likelihood ratio between two models, say $A$ and $B$, divided by the total number of earthquakes $N$, observed in a given time window, i.e.

$$\mathscr{I}_N(A,B) = \frac{\log L_A/L_B}{N} = \frac{\log L_A - \log L_B}{N}. \tag{1}$$

Popular statistical models for earthquake occurrences are based on marked point processes [1, Chap. 6], with magnitudes and locations as marks. We will refer here only to the time–magnitude analysis, discarding the spatial component.

Consider a marked point process on $\mathbb{R}^+ \times \mathbb{R}^+$, adapted to the filtration $\{\mathscr{F}_t\}_{t\geq 0}$, with conditional intensity function $\lambda(t,m|\mathscr{F}_{t-})$, given the history up to but not including time $t$. Suppose, in addition, that the process is such that the likelihood, given a realization $\{(t_i,m_i)\}_{i=1}^n$ over the interval $[0,T]$, for some positive finite $T$, is well defined (for further details, see [1, Chap. 7]). Then, the likelihood $L$ of such a realization is expressible in the form

$$L = \left[\prod_{i=1}^{n} \lambda(t_i,m_i|\mathscr{F}_{t_i-})\right] \exp\left(-\int_0^T \int_{\mathbb{R}^+} \lambda(t,m|\mathscr{F}_{t-}) \mathrm{d}m \mathrm{d}t\right). \tag{2}$$

Its log likelihood ratio on $[0,T]$ relative to a compound Poisson process [1, Chap. 6] with constant intensity $\nu$ and mark distribution $\pi(m)$, that is independent of time $t$, is given by

$$\log\frac{L}{L_0} = \sum_{i=1}^{n}\log\frac{\lambda(t_i,m_i|\mathscr{F}_{t_i-})}{\nu\pi(m_i)} - \int_0^T\int_{\mathbb{R}^+}\lambda(t,m|\mathscr{F}_{t-})\mathrm{d}m\mathrm{d}t + \nu T. \tag{3}$$

The basic idea behind the information gain is that a forecast with a higher joint log likelihood is "better". However, before observing an earthquake sequence, $\mathscr{I}_N(A,B)$ is a random variable and has its own uncertainty. Additionally, it depends on the duration of the time interval in which we observe the events. We use simulations from a popular model of seismicity to estimate the uncertainty of this statistic and to explore its robustness to the addition of new sequences.

## 3 The Epidemic Type Aftershock Sequence (ETAS) Model

The Epidemic Type Aftershok Sequence (ETAS) model was introduced by Ogata [6]. Belonging to the class of marked Hawkes processes [4, 1], the model approximates seismicity by an epidemic process: any earthquake increases the rate of future events for some period of time (Hawkes' self-exciting property), and large quakes induce more aftershocks (higher infection rate).

Formally, the ETAS model corresponds to a marked point process [1, Chap. 6] on $\mathbb{R}^+\times\mathbb{R}^+$, adapted to the filtration $\{\mathscr{F}_t\}_{t\geq 0}$, with conditional intensity function

$$\lambda(t,m|\mathscr{F}_{t-}) = \left[\mu + \sum_{i:t_i<t}k(m_i)g(t-t_i)\right]p(m), \tag{4}$$

where $\mu > 0$ represents the background seismicity rate; the term $k(m_i)g(t-t_i)$ is the contribution to seismicity rate by the $i$-th event $(t_i,m_i)$, specifically

$$k(m) = Ae^{\alpha(m-m_0)}, \qquad m\geq m_0, \tag{5}$$

is the mean number of direct offspring from an event sized $m$, $m_0$ being the magnitude threshold, and

$$g(t) = \frac{c^{p-1}(p-1)}{(t+c)^p}, \qquad t\geq 0, \tag{6}$$

is the modified Omori law [7] for the occurrence times of direct offspring. Magnitudes are distributed independently according to the Gutenberg–Richter law [3]

$$p(m) = \beta\, e^{-\beta\,(m-m_0)}, \qquad m\geq m_0, \tag{7}$$

which is the probability density function of a translated exponential distribution with rate parameter $\beta = b\log 10,\ b>0$. Magnitudes are independent of past seismicity. $A,\ \alpha,\ c,\ p$ are constant positive parameters. For the sake of simplicity, we fix $m_0 = 0$.

The evolution of the information gain is tied to the evolution of the underlying process, as we will show in Section 4 (see also [1, 2]). Thus, we are interested in studying the behaviour of the process itself over time.

Stability properties of the ETAS model are simpler to derive in its branching process representation; we can interpret the mark $m_i$ as the "type" of an individual in a multi-type Galton–Watson process with a modified time dimension. In this context, stability is closely related to the concepts of criticality and a branching ratio. The branching ratio is defined as the number of descendants for one immigrant over the size of their entire family (all descendants plus the original immigrant); that is

$$\rho = \frac{A\beta}{\beta - \alpha}. \tag{8}$$

Sufficient conditions for the existence of a stationary version [1, 2] are

$$p > 1, \qquad \beta > \alpha, \qquad \rho < 1, \tag{9}$$

which implies a *subcritical* process. When $\rho > 1$, the process is *supercritical*: there is a finite probability of an infinite number of events in a unit time interval.

Now, let $N(T)$ denote the number of events in the interval $(0,T]$[1]. If in addition $\beta > 2\alpha$, it can be shown that, for every sequence $\{T_n\}_{n\in\mathbb{N}}$ such that $T_n \to \infty$, a central limit theorem holds [5], namely

$$\sqrt{T_n}\left(\frac{N(T_n)}{T_n} - \frac{\mu}{1-\rho}\right) \xrightarrow{d} \mathrm{N}\left(0, \frac{\mu(1+\sigma^2)}{(1-\rho)^3}\right), \qquad T_n \to \infty, \tag{10}$$

where

$$\sigma^2 = \frac{A^2\beta}{\beta - 2\alpha} - \rho^2.$$

The previous condition $\beta > 2\alpha$ is necessary and sufficient for the existence of $\sigma^2$, and hence the asymptotic variance in (10). These results suggest that

$$\mathbb{E}\left[\frac{N(T)}{T}\right] \approx \frac{\mu}{1-\rho}, \qquad \mathbb{V}\left[\frac{N(T)}{T}\right] \approx \frac{1}{T} \times \text{const} \tag{11}$$

for sufficiently large $T$. However, the condition $\beta > 2\alpha$ does not hold for quakes.

## 4 Simulation Study

For each model in Table 1, we simulate ten thousand catalogs within the time window $[0, T_{\max}]$, with $T_{\max} = 10\,000$ days, and compute the information gain per earthquake over a finite grid of time $T_1 < T_2 < \cdots < T_k = T_{\max}$, based on the log likelihood

[1] $N(0) = 0$ almost surely.

ratio in (3). The compound Poisson process used in (3) provides a benchmark. We set $\nu = 5$ and $\pi(\cdot) = p(\cdot)$ as the mark distribution.

| Experiment | | $\mu$ | $\beta$ | $\alpha$ | $n$ | $p$ | $c$ |
|---|---|---|---|---|---|---|---|
| 1 | no clustering | 1 | 2.3 | 0 | 0 | – | – |
| 2 | short memory cl. | 1 | 2.3 | 0 | .5 | 5 | .1 |
| 3 | long memory cl. | 1 | 2.3 | $\beta/3$ | .5 | 1.2 | .1 |
| 4 | no CLT | 1 | 2.3 | 2.2 | .5 | 1.2 | .1 |

**Table 1** Simulation scheme.

Remarkable results are displayed in Fig. 4. Experiment 3 shows well behaved trajectories of the number of events per unit time $N(T)/T$. The sample variance is bounded from above by $1/T$ times the asymptotic variance $\mu(1+\sigma^2)/(1-\rho)^3$ from (10). As a result, the information gain stabilizes around a single value. On the other hand, in experiment 4, for which the central limit theorem does not hold, trajectories and sample variance have a completely different behaviour. The information gain that does not converge to a stable value but continues to fluctuate. This is a result of large seismicity variations caused by the high aftershock rates of great earthquakes.

## 5 Discussion and Conclusions

The lack of an obvious convergence of the information gain per earthquake to a stable value is a warning flag: a gain measured at a moment in time, even if supported by a large data set, may change substantially in the future. The fluctuations result from the empirically-supported near-equality between the Gutenberg–Richter exponent $\beta$ and the productivity exponent $\alpha$. Under these conditions, a central limit theorem, which otherwise ensures convergence, does not hold.

A next step is to investigate the importance of a physically required maximum magnitude that truncates the Gutenberg–Richter law. This will theoretically restore the central limit theorem. However, observed magnitudes near the upper limit are extremely rare, and therefore the finite variance may not reign in the fluctuations for decades. Our ultimate goal is to provide CSEP with guidelines for inferring relative model performance on the basis of the information gain.

## References

1. Daley, D.J., Vere-Jones, D.: An Introduction to the Theory of Point Processes, vol. I, 2nd edn. Springer-Verlag New York (2003)
2. Daley, D.J., Vere-Jones, D.: An Introduction to the Theory of Point Processes, vol. II, 2nd edn. Springer-Verlag New York (2008)
3. Gutenberg, B., Richter, C.F.: Frequency of earthquakes in California. Bulletin of the Seismological Society of America **34**(4), 185–188 (1944)

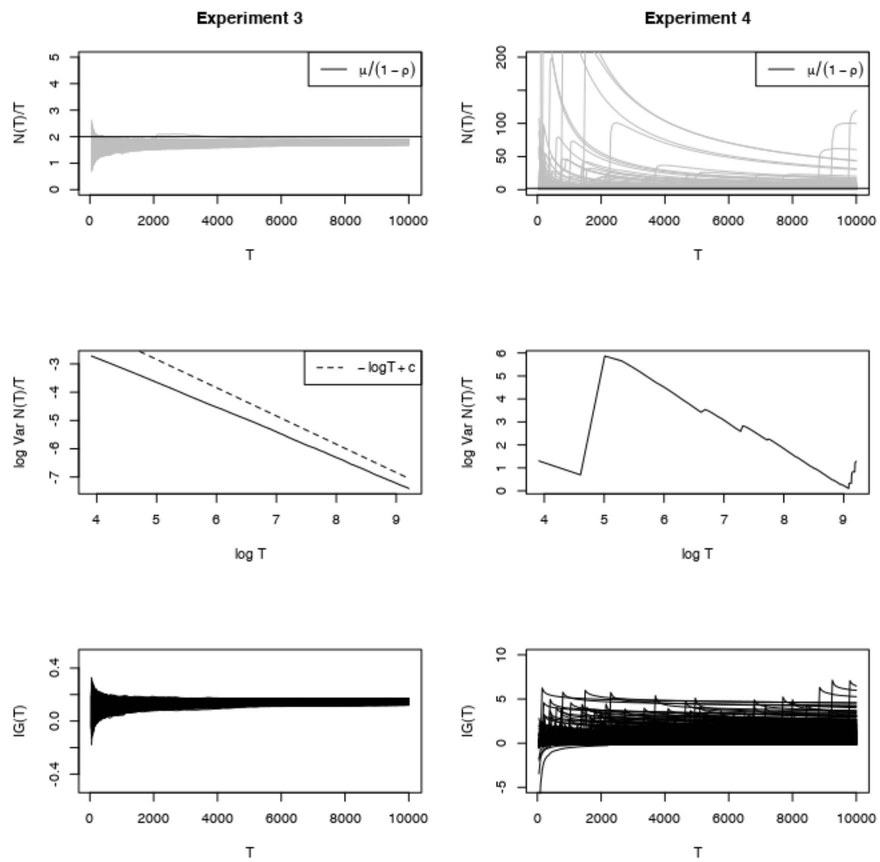


**Fig. 1** From top to bottom: ETAS trajectories (number of events per unit time); sample variance (solid line) and asymptotic variance (dashed line) with time, $c = \log\left(\mu(1+\sigma^2)/(1-\rho)^3\right)$; and information gain per earthquake. Note the different y-axis scales between the left and right panels.


4. Hawkes, A.G., Oakes, D.: A cluster process representation of a self-exciting process. Journal of Applied Probability **11**(3), 493–503 (1974)
5. Karabash, D., Zhu, L.: Limit theorems for marked hawkes processes with application to a risk model. Stochastic Models **31**(3), 433–451 (2015)
6. Ogata, Y.: Statistical models for earthquake occurrences and residual analysis for point processes. Journal of the American Statistical association **83**(401), 9–27 (1988)
7. Utsu, T.: A statistical study on the occurrence of aftershocks. Geophysical Magazine **30**(4) (1961)
8. Zechar, J.D., Schorlemmer, D., Liukis, M., Yu, J., Euchner, F., Maechling, P.J., Jordan, T.H.: The Collaboratory for the Study of Earthquake Predictability perspective on computational earthquake science. Concurrency and Computation: Practice and Experience **22**(12), 1836–1847 (2010)