Pure appl. geophys. 162 (2005) 2299–2319
0033–4553/05/122299–23
DOI 10.1007/s00024-005-2777-5

Pure and Applied Geophysics

Information Gains for Stress Release Models

MARK S. BEBBINGTON[1]

Abstract— The information gain of a point process model quantifies its predictability, relative to a reference model such as the Poisson process. This is bounded above by the entropy gain, or difference between the point process entropy rates. This provides a bound on the utility of the model as a forecasting tool, separate from the usual "goodness of fit" assessment criteria. The stress release model is a point process with an underlying state variable increasing linearly with time, and decreased by events. Assuming the intensity to be an exponential function of this state, we derive an analytic expression for the entropy gain. This is illustrated, using various magnitude distributions, for earthquake data from north China, and extensions to a multivariate linked model outlined. The results measure the effectiveness of the stress release process as a predictive tool. Comparisons are made with a scale derived from the Gamma renewal process and using Molchan's $\nu - \tau$ diagram.

Key words: Point processes, entropy rates, self-correcting point process, earthquake forecasting.

1. Introduction

In attempting to forecast some phenomena using a point process (or indeed any statistical model), evaluating the model presents two facets. The first of these is the usual question of how well the model fits the data, which is quantified through the likelihood, and perhaps some residual analysis (OGATA, 1988). The second aspect is one of "assessment": How accurate or precise is the forecast? This can be quantified by the information gain (VERE-JONES, 1998; DALEY and VERE-JONES, 2003, 2004). In this case a further question begs answering: Is the model of any use? In some cases, it may be possible to place an upper bound on the information gain, and hence the maximum predictability for forecasting purposes, of the model. A case for study is provided by point process models with history dependent conditional intensities for earthquake occurrence (see, for example, OGATA, 1999, and references therein). In view of the considerable recent interest, there is a need for assessing and comparing the performance of these models in a forecasting role. Similar concerns arise in weather forecasting, see for example KLEEMAN *et al.* (2002).

[1]IIST, Massey University, Private Bag 11222, Palmerston North, New Zealand.
E-mail: m.bebbington@massey.ac.nz

We will next define the information gain, and the entropy gain which provides an upper bound. In Section 2 we will outline two point processes for which we can calculate the entropy gain. Section 3 contains our main result, explicit formulae for the entropy gain of the stress release process, including three alternative jump (magnitude) distributions. The calculations are illustrated in Section 4 using data from north China. An extension to a multivariate model is briefly considered in Section 5. The results are discussed in Section 6, and compared to other descriptive assessment techniques.

1.1 The Information Gain

Let us consider a stationary, ergodic, point process $N(t) \equiv N(0,T]$ with conditional intensity $\lambda(t)$ such that

$$\lambda(t)dt = \mathbb{E}(dN(t) \mid \mathscr{H}_t),$$

where the conditioning is taken with respect to the σ-algebra $\mathscr{H}_t$ of events defined on the history of the process up to time t. We wish to generate estimates of the probability of at least one event in each of a sequence of non-overlapping intervals $(t_i, t_i + \delta_i)$, covering a realization of the process. In principle, the forecasts can be updated between intervals (VERE-JONES, 1998; HARTE and VERE-JONES, 2005), allowing the conditional intensity to reflect the activity in the preceding intervals. Using the conditional intensity we can generate, through simulation if necessary, forecast probabilities $p_i = \mathbb{P}(Y_i = 1)$, where the outcomes are denoted by

$$Y_i = \begin{cases} 1, & \text{if } N(t_i + \delta_i) - N(t_i) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

The log-likelihood for the series of intervals is then the binomial score

$$B = \sum_i (Y_i \ln p_i + (1 - Y_i) \ln(1 - p_i)).$$

This rewards high forecast probabilities for intervals where an event occurs, and low forecast probabilities where events do not occur, and penalizes false alarms and missed events. If the intervals cover a total period $T = \sum_i \delta_i$, an estimate, over the finite period $(0, T)$, of the information per unit time generated by the process is B/T. We would prefer to have the information gain relative to a reference process, and in accordance with the maximum entropy principle, the natural reference process is the Poisson process with the same mean rate $\bar{\lambda} = \mathbb{E}[\lambda(t)]$ ($= \mathbb{E}[\lambda(0)]$ by stationarity). This choice also facilitates calculations. If $\bar{p}_i = 1 - \exp(-\bar{\lambda}\delta_i)$ denotes the forecast probabilities for this reference process, the average difference between the binomial scores,

$$\bar{\rho}_T = \frac{B - \bar{B}}{T} = \frac{1}{T} \sum_i \left(Y_i \ln\left(\frac{p_i}{\bar{p}_i}\right) + (1 - Y_i) \ln\left(\frac{1 - p_i}{1 - \bar{p}_i}\right) \right),$$

is the *mean information gain per unit time* associated with the forecasting procedure using $\lambda(t)$. The ratio $p_i/\bar{p}_i$ is the *probability gain*, and thus the information gain is the expected value (the random element is the process history, $\mathscr{H}_t$) of the log probability gain. If the test data $\{Y_i\}$ is different to the data from which the model parameters have been estimated, bias due to the differing numbers of parameters between models should not arise, so that the models can be ranked according to the value of their gains against the reference model.

As $B - \bar{B}$ is simply the difference in the log-likelihoods of the fitted and reference models, when the information gain is calculated using observed data, it can be interpreted as a measure of goodness of fit. But, in the "assessment" role outlined above, if data are simulated from the model and refitted by maximum likelihood, the information gain then quantifies the predictability of the process relative to the reference model. A larger information gain indicates a more predictable (or powerful) model and this enables models to be compared. It is possible to go still one step further, and determine a theoretical bound on the information gain of a model, requiring neither fitting nor simulated forecasts.

1.2 The Entropy Gain

Recall that the *entropy rate* for a stationary, ergodic, point process is

$$H = -\mathbb{E}[\lambda(0)(\ln \lambda(0) - 1)].$$

We can then define the generalized entropy rate, relative to the corresponding Poisson process, as

$$G = \mathbb{E}\left[\lambda(0)\ln\lambda(0) - \bar{\lambda}\ln\bar{\lambda}\right] = \mathbb{E}\left[\lambda(0)\ln\left(\lambda(0)/\bar{\lambda}\right)\right], \tag{1}$$

which we call the *entropy gain per unit time*. This is effectively a number characterizing the inherent predictability of the model.

The following two results then relate the information gain to the entropy gain.

Proposition 1 (Vere-Jones, 1998). For all subdivisions of $(t, t+T)$, $\mathbb{E}[\bar{\rho}_T] \leq G$. *Furthermore,* $\mathbb{E}[\bar{\rho}_T] \to G$ *as* $\max_i \delta_i \to 0$, *and* $\bar{\rho}_T \to G$ *as* $T \to \infty$.

Noting that forecasts are inherently made in discrete time, this implies that the performance of such forecasts (i) is bounded above by the entropy gain, calculated in continuous-time, and (ii) tends to the bound as the intervals become smaller. Thus the performance of any probability forecasting scheme based on a point process model is bounded by an intrinsic property of the model itself.

The average information gain over a long series of trials could easily be well below the entropy gain, particularly if there are considerable differences between the actual data and that produced by the model. In this case the mean information gain (or equivalently the average log-likelihood per observation) forms the basis of a goodness of fit test (DALEY and VERE-JONES, 2003, Section 7.6; HARTE and VERE-JONES, 2005).

Instead of dividing $B - \bar{B}$ by the total time T to obtain a mean score per unit time, we can divide by the total number of events $N(T)$ to obtain an average score per event. We can then compare this, as above, to the Poisson reference process to form an average information gain per event, in which case the corresponding upper bound and limit is the *entropy gain per event* $G/\bar{\lambda}$. In this case Proposition 1 has the following corollary.

Proposition 2 (VERE-JONES, 1998).

$$\mathbb{E}\left[\frac{1}{N(T)}\sum_{i=1}^{N(T)} \ln\left(\frac{p_i}{\bar{p}_i}\right)\right] \approx \mathbb{E}\left[\frac{T}{N(T)}\bar{\rho}_T\right] \leq G/\bar{\lambda}.$$

KAGAN and KNOPOFF (1977) suggested using the averaged log-likelihood as a measure of the information rate for the process, which yields (VERE-JONES, 1998)

$$\frac{1}{T}\ln L(0,T) \sim \mathbb{E}[\lambda(0)(\ln\lambda(0) - 1)],$$

where $\ln L(0,T) = \sum_i \ln\lambda(t_i) - \int_0^T \lambda(t)\,\mathrm{d}t$ is the log-likelihood when points occur at times $0 < t_1 < t_2 < \cdots < T$. Combining this with

$$G = \mathbb{E}\big[\lambda(0)\ln(\lambda(0)/\bar{\lambda})\big] = \mathbb{E}[\lambda(0)(\ln\lambda(0) - 1)] + \bar{\lambda} - \bar{\lambda}\ln\bar{\lambda},$$

we obtain

$$G \simeq \frac{1}{T}\ln L(0,T) + \bar{\lambda} - \frac{N(T)}{T}\ln\bar{\lambda}. \qquad (2)$$

This implies that choosing the model that maximizes the likelihood roughly corresponds to choosing the model with the best prediction performance. However, in practice allowance has to be made for the number of unknown parameters estimated from the data, which leads to criteria such as the Akaike Information Criterion (AKAIKE, 1977).

2. Some Point Process Models for Earthquake Occurrence

Besides the Poisson process, the most common point process used in earthquake modelling is the renewal process. For a stationary renewal process with inter-event distribution $F(\cdot)$, survivor function $\bar{F}(\cdot)$, density $f(\cdot)$ and mean $1/m = \int_0^\infty \bar{F}(y)\mathrm{d}y$, the conditional intensity is $\lambda(t) = f(B_t)/\bar{F}(B_t)$, where B_t denotes a backward recurrence time, the time since the last event. It can then be shown (DALEY and VERE-JONES, 2003, Section 7.6) that

$$G = m\left(1 - \ln m + \int_0^\infty f(y) \ln f(y) \mathrm{d}y\right). \quad (3)$$

The usual renewal density considered in earthquake modelling is the Lognormal

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(\frac{1}{2\sigma^2}(\ln y - \mu)^2\right),$$

as this provides a conditional intensity

$$\lambda(t) = f(t)/\bar{\Phi}((\ln t - \mu)/\sigma), \quad (4)$$

where $\bar{\Phi}(z) = \int_z^\infty (2\pi)^{-1/2} \exp(-z^2/2) dz$, that first increases, and then decreases, with t. It also readily allows for occasional very long inter-event times. We find that

$$\int_0^\infty f(y) \ln f(y) dy = -\ln(\sqrt{2\pi}\sigma) - \mathbb{E}[\ln Y] - \frac{1}{2\sigma^2} \mathbb{E}\left[(\ln Y - \mu)^2\right]$$

and so

$$G = \exp(-\mu - \sigma^2/2)\left(\frac{1 + \sigma^2}{2} - \ln(\sqrt{2\pi}\sigma)\right).$$

Since $\bar{\lambda} = m = \exp(-\mu - \sigma^2/2)$, we see that the entropy gain per event is

$$G/\bar{\lambda} = \frac{1 + \sigma^2}{2} - \ln(\sqrt{2\pi}\sigma),$$

a function of σ alone. The conditional intensity and the entropy gains are shown in Figure 1. We see that the entropy gain per event can become arbitrarily large, as the events become increasingly infrequent. This derives from the very long tail of the distribution, and the consequent ability to forecast an absence of events with considerable confidence.

DALEY and VERE-JONES (2004) calculate the entropy gain for the Gamma renewal process, while IMOTO (2004) determines the probability gains from renewal models with Lognormal, Gamma,Weibull and Brownian distributions.

The renewal process corresponds, in earthquake modelling, to a slip predictable model, where the size of the event depends on the interval before it, but the interval after an event is independent of its size. However, a causal relationship between the size of an earthquake and the following interval is desirable when modelling large earthquakes (REID, 1910) which leads in the point process setting to the stress release process (VERE-JONES, 1978). This models the gradual build-up of stress by tectonic movements and its release in the form of earthquakes. The stress level $X(t)$ increases deterministically between earthquakes and is reduced stochastically during earthquakes, as

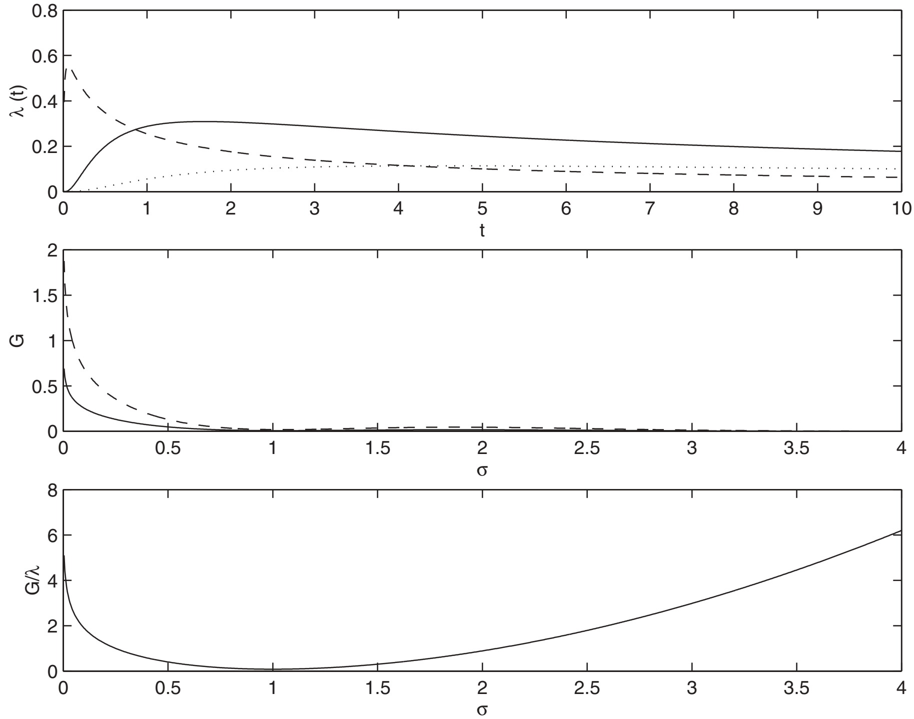

Figure 1
Top: Conditional intensity of the log-normal renewal process with $\mu = \sigma = 1$ (solid), $\mu = 2, \sigma = 1$ (dashed) and $\mu = 1, \sigma = 2$ (dotted). Middle: Entropy gain per unit time with $\mu = 1$ (solid) and $\mu = 2$ (dashed) for varying σ. Bottom: Entropy gain per event as a function of σ.

$$X(t) = X(0) + \rho t - S(t)$$

where $X(0)$ is the initial value, ρ is a constant loading rate from external tectonic forces, and $S(t)$ is the accumulated stress release from earthquakes within the region over the period $(0, t)$. That is, $S(t) = \sum_{t_i < t} S_i$, where t_i and S_i are the origin time and the stress release associated with the i-th earthquake. This release is usually inferred from the magnitude of the event M_i using

$$S_i = 10^{\eta(M_i - M_0)}, \tag{5}$$

where M_0 is a magnitude cutoff. The two commonly used values of η are $\eta = 0.75$ in which case the 'stress release' corresponds to Benioff strain, and $\eta = 1.5$, where the 'stress release' is actually the seismic moment.

The conditional intensity of an earthquake occurrence is a function $\psi(X)$ which is monotonically increasing. It is usually chosen as

$$\psi(X) = e^{\beta(X - x_0)}, \tag{6}$$

where the constant x_0 is the unknown initial value of stress,while the constant $\beta > 0$ is an amalgam of the strength and heterogeneity of the crust in the region. The result is a point process in time-stress space with conditional intensity function

$$\lambda(t) = \psi(X(t)) = \exp(\beta(X(0) + \rho t - S(t) - x_0)). \tag{7}$$

The amount of stress released in an event initiated by a stress level $X(t) = x$ has distribution $J(dy \mid x)$. In practice this is usually assumed to be independent of the stress level,

$$J(dy \mid x) = J(dy), \tag{8}$$

and so the stress drops are independent identically distributed random variables with distribution J.

Borovkov and Vere-Jones (2000) show that, given (6) and (8), X has characteristic function

$$\varphi(s) = \mathbb{E}\left[e^{isX}\right] = e^{isR}\gamma(s)\prod_{k=1}^{\infty}\exp\left(\frac{is}{\beta k}\right)\frac{\gamma(s - ik\beta)}{\gamma(-ik\beta)}, \tag{9}$$

where $R = x_0 + \beta^{-1}(\ln(\beta\rho) - \Gamma), \Gamma = 0.5772\ldots$ is Euler's constant,

$$\gamma(s) = \frac{1 - \chi(s)}{is\mathbb{E}[S]}, \tag{10}$$

and $\chi(s) = \int_0^{\infty} e^{-isx} J\ (dx)$.

3. *Entropy Gain of the Stress Release Process*

If we assume (see Zheng, 1991; Borovkov and Vere-Jones, 2000) the existence of a stationary distribution $f(x)$ for the stress $X(t)$, we have the forward equation

$$\rho f(x) = \int_x^{\infty} \psi(y) f(y) \mathbb{P}(S > y - x)\, \mathrm{d}y.$$

where S is the random variable for the stress released in an event, which yields (Vere-Jones, 1998)

$$\frac{\rho}{\mathbb{E}[S]} = \int_{-\infty}^{\infty} \psi(y) f(y) dy = \mathbb{E}[\psi(X(0))] = \bar{\lambda}$$

and

$$\mathbb{E}[X\psi(X)] = \frac{\rho}{\mathbb{E}[S]}\left(\mathbb{E}[X] + \frac{\mathbb{E}[S^2]}{2\mathbb{E}[S]}\right). \tag{11}$$

So, from (1), the entropy gain per unit time is

$$G = \mathbb{E}\left[\psi(X)\left(\ln\psi(X) - \ln\left(\frac{\rho}{\mathbb{E}[S]}\right)\right)\right] = \mathbb{E}[\psi(X)\ln\psi(X)] - \frac{\rho}{\mathbb{E}[S]}\ln\left(\frac{\rho}{\mathbb{E}[S]}\right). \tag{12}$$

Theorem 1. *If the stress release process has a hazard function* (6), *and a jump distribution* $J(dx)$ *such that*

$$\int_0^\infty x^2 J(\mathrm{d}x) < \infty, \tag{13}$$

then the entropy gain per unit time is

$$G = \frac{\rho}{\mathbb{E}[S]}\left(\ln(\beta\mathbb{E}[S]) - \Gamma + \beta\sum_{k=1}^{\infty}\frac{\int_0^\infty xe^{-k\beta x}J(dx)}{1 - \int_0^\infty e^{-k\beta x}J(dx)}\right). \tag{14}$$

The proof is deferred to the Appendix. We note that G is proportional to ρ, and thus to the expected number of events. Hence the following corollary follows trivially.

Corollary 1. *Under the conditions of Theorem 1, the entropy gain per event is*

$$G/\bar{\lambda} = \ln(\beta\mathbb{E}[S]) - \Gamma + \beta\sum_{k=1}^{\infty}\frac{\int_0^\infty xe^{-k\beta x}J(\mathrm{d}x)}{1 - \int_0^\infty e^{-k\beta x}J(\mathrm{d}x)}. \tag{15}$$

We observe from (15) that the entropy gain per event is essentially unaffected by changes in ρ.

Theorem 1 and its corollary provide expressions for the entropy gain in terms of the distribution of the stress drops $J(dx)$. We will now consider possibilities for the latter.

3.1 Truncated Gutenberg-Richter Distribution

The truncated Gutenberg-Richter distribution has density

$$g(m) = \frac{\theta e^{\theta(m_0 - m)}}{1 - e^{\theta(m_0 - m_{\max})}}, \qquad m \in (m_0, m_{\max}),$$

where the "b-value" $\theta\log_{10}e$ describes the fall-off in frequency with magnitude. Since b-values less than one are common, the distribution must be truncated at some maximum magnitude $m_{\max}$. Using (5), we obtain

$$\mathbb{P}(S \le s) = \frac{1 - s^c}{1 - e^{\theta(m_0 - m_{\max})}}, \qquad s \in (1, s_{\max}), \tag{16}$$

where $s_{\max} = 10^{\eta(m_{\max} - m_0)}$ and $c = -(\theta/\eta)\log_{10} e$. Thus $J(dx) = c_0 x^{c-1}\mathrm{d}x$, for $x \in (1, s_{\max})$, with $c_0 = -c/(1 - e^{\theta(m_0 - m_{\max})})$. Then

$$\mathbb{E}[S] = \int_1^{s_{\max}} c_0 x^c \mathrm{d}x = \frac{c_0}{c+1}(s_{\max}^{c+1} - 1)$$

and

$$\int_0^\infty x e^{-k\beta x} J(\mathrm{d}x) = c_0 \int_1^{s_{\max}} e^{-k\beta x} x^c \mathrm{d}x.$$

The latter integral can be expressed in terms of the incomplete Gamma function $P(a,x) = \int_0^x e^{-t} t^{a-1} \mathrm{d}t / \Gamma(a)$ yielding

$$\int_0^\infty x e^{-k\beta x} J(\mathrm{d}x) = \frac{c_0 \Gamma(c+1)}{(k\beta)^{c+1}} (P(c+1, k\beta s_{\max}) - P(c+1, k\beta)).$$

Using the recurrence formula $P(a,x) = P(a+1,x) + x^a e^{-x}/\Gamma(a+1)$ (ABRAMOWITZ and STEGUN, 1964) we eventually get

$$\int_0^\infty x e^{-k\beta x} J(\mathrm{d}x) = \frac{c_0 \Gamma(c+1)}{(k\beta)^{c+1}} (P(c+2, k\beta s_{\max}) - P(c+2, k\beta)) + \frac{c_0}{c+1}\left(s_{\max}^{c+1} e^{-k\beta s_{\max}} - e^{-k\beta}\right),$$

where $c + 2 > 0$ for b-values less than 1.5, facilitating numerical calculation. Similarly,

$$\int_0^\infty e^{-k\beta x} J(\mathrm{d}x) = \frac{c_0 \Gamma(c)}{(k\beta)^c} (P(c+2, k\beta s_{\max}) - P(c+2, k\beta)) + \frac{c_0}{c}\left(\left(\frac{k\beta s_{\max}}{c+1} + 1\right) s_{\max}^c e^{-k\beta s_{\max}} - \left(\frac{k\beta}{c+1} + 1\right) e^{-k\beta}\right).$$

3.2 Tapered Pareto Distribution

VERE-JONES *et al.* (2001) examine a distribution for the stress itself,

$$\mathbb{P}(S > s) = \left(\frac{s}{s_0}\right)^{-\alpha} \exp\left(\frac{s_0 - s}{U}\right), \qquad s > s_0, \tag{17}$$

where the 'upper turning point' U is usually specified through the equivalent magnitude $\gamma = m_0 + \eta^{-1}\log_{10} U$ obtained via inverting (5). This is sometimes known as the Modified Gutenberg-Richter distribution (see, e.g., SORNETTE and SORNETTE, 1999). Note that in order to obtain a sensible magnitude distribution, α must roughly be inversely proportional to η.

We find that

$$\mathbb{E}[S] = s_0 + \int_{s_0}^{\infty} \mathbb{P}(S > s)\,\mathrm{d}s = s_0 + d_0 U^{1-\alpha} I(1-\alpha, s_0/U)$$

where $d_0 = s_0^{\alpha} e^{s_0/U}$ and $I(a,x) = \int_x^{\infty} t^{a-1} e^{-t} \mathrm{d}t = \Gamma(a)(1 - P(a,x))$. Similarly,

$$\mathbb{E}[S^2] = s_0^2 + 2\int_{s_0}^{\infty} s\mathbb{P}(S > s)\,\mathrm{d}s = s_0^2 + 2d_0 U^{2-\alpha} I(2-\alpha, s_0/U).$$

Differentiating (17) we find that $J(\mathrm{d}x) = (\alpha/x + 1/U)\mathbb{P}(S > x)dx$, and so if $d_k = k\beta + 1/U$, for $k > 0$, then

$$\int_0^{\infty} x e^{-k\beta x} J(dx) = d_0 \int_{s_0}^{\infty} \left(\frac{\alpha}{x} + \frac{1}{U}\right) x^{1-\alpha} e^{-d_k x} dx$$
$$= d_0 \alpha d_k^{\alpha-1} I(1-\alpha, d_k s_0) + (d_0/U) d_k^{\alpha-2} I(2-\alpha, d_k s_0).$$

Similarly,

$$\int_0^{\infty} e^{-k\beta x} J(dx) = d_0 \alpha d_k^{\alpha} I(-\alpha, d_k s_0) + (d_0/U) d_k^{\alpha-1} I(1-\alpha, d_k s_0).$$

3.3 Empirical Magnitude Distribution

Bebbington and Harte (2001) argue for using the observed empirical magnitude distribution when forecasting through simulation, as this removes a level of approximation. Suppose that we observe N earthquakes of magnitude $m_i, i = 1, \ldots, N$. Converting these magnitudes to stresses s_i using (5), the entropy gain per unit time is

$$G = \frac{N\rho}{\sum_{i=1}^{N} s_i} \left(\ln\left(\frac{\beta}{N} \sum_{i=1}^{N} s_i \right) - \Gamma + \beta \sum_{k=1}^{\infty} \frac{\sum_{i=1}^{N} s_i e^{-k\beta s_i}}{N - \sum_{i=1}^{N} e^{-k\beta s_i}} \right).$$

In the case that the stress distribution is degenerate, i.e., a characteristic earthquake, $J(\mathrm{d}x) = \delta(x - s)$ say,

$$G = \frac{\rho}{s} \left(\ln(\beta s) - \Gamma + \beta s \sum_{k=1}^{\infty} \frac{e^{-k\beta s}}{1 - e^{-k\beta s}} \right).$$

Table 1

Fitted stress release parameters in Eq. (7) using north China data

η	$\hat{\beta}$	$\hat{\rho}$	$\hat{x_0}$	$\ln L$
0.75	0.010	1.176	246.2	−195.87
1.5	0.000134	47.3	18193	−196.68

4. Illustration Using Data from North China

We will now use the results of the previous section to calculate the entropy gains for the stress release model fitted to an historical earthquake catalogue from north China. The maximum likelihood estimates (see, for example, DALEY and VERE-JONES, 2003, Section 7.1) of the parameters in the conditional intensity (7) are given in Table 1. For our analyses we will need the distribution of the recorded magnitudes, as given in Table 2. The MLE of the b-value is $\hat{b} = \log_{10} e/(\bar{M} - m_0) = 0.61$, using a cut-off magnitude $m_0 = 6.0$. Figure 2 shows the catalogue of 65 observed events, and the fitted conditional intensity in the case $\eta = 0.75$.

Examples of the entropy gain for the stress release process with the magnitude distributions considered above are presented in Table 3. The magnitude parameters were chosen according to two criteria. Firstly, to provide a reasonable maximum magnitude, hence $m_{\max} = 9.0$, also used by VERE-JONES (1998), and $\gamma = 8.5$, which is the MLE to the one decimal place in the data. The remaining parameters θ and α were then chosen to preserve the mean jump size $\mathbb{E}[S]$ and hence the event rate $\bar{\lambda}$, meaning that the reference process is the same in all cases, allowing for direct comparison.

We see that, even leaving aside the degenerate distribution which appears to provide a minimum of entropy gain, a range of quite respectable magnitude distributions leads to order of magnitude differences in the entropy gain. The dependence of G on the magnitude distribution arises from the fact that although the magnitudes are independent of the history, the history and hence the conditional intensity is dependent on the magnitudes through the stress drops. The larger entropy gains from using $\eta = 1.5$ point to the gain resulting largely from the depressive effect of large events on the intensity (7). In other words, we are likely to successfully forecast an absence of activity subsequent to a large event. This factor also goes far

Table 2

Tally of observed magnitudes from north China

M	6.0	6.2	6.3	6.5	6.7	7.0	7.2	7.3	7.4	7.5	7.8	8.0	8.5	8.6
No.	20	7	1	8	5	7	2	1	1	4	1	6	1	1

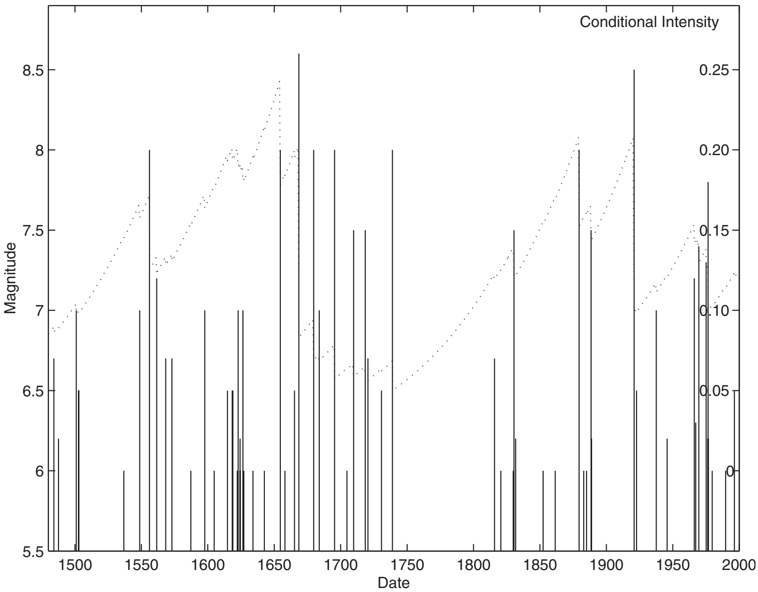

Figure 2
Magnitudes of observed events (solid lines) and fitted stress release conditional intensity (dotted line) for North China catalog.

towards explaining the variation between the distributions, with those more likely to produce larger stress drops providing greater entropy gain. A secondary source of entropy gain in the stress release model derives from the fact that small events are likely to both result from high stress levels, and not depress these levels, and so the likelihood of successfully forecasting an event subsequently is raised. Using $\eta = 1.5$ does better here too, as the smaller events have less relative effect on the stress level $X(t)$. We note that the approximation (2) yields an entropy gain $G = 0.0078$ ($\eta = 0.75$) or $G = 0.0071$ ($\eta = 1.5$), of the right order, although smaller than the

Table 3

Calculated entropy gains for north China using selected magnitude distributions

Stress Distribution, $J(dx)$	η	$E[S]$	$\bar{\lambda}$	G	$G/\bar{\lambda}$
Empirical, Table 2	0.75	8.81	0.1335	0.0120	0.0902
Degenerate, $M = 7.26$	0.75	8.81	0.1335	0.0029	0.0220
Truncated G-R, $\theta = 1.2225$, $m_{\max} = 9.0$	0.75	8.81	0.1335	0.0160	0.1200
Tapered Pareto, $\alpha = 0.6949$, $\gamma = 8.5$	0.75	8.81	0.1335	0.0171	0.1284
Empirical, Table 2	1.5	330.5	0.1431	0.0213	0.1486
Degenerate, $M = 7.68$	1.5	330.5	0.1431	0.0016	0.0111
Truncated G-R, $\theta = 1.3950$, $m_{\max} = 9.0$	1.5	330.5	0.1431	0.0480	0.3353
Tapered Pareto, $\alpha = 0.3688$, $\gamma = 8.5$	1.5	330.5	0.1431	0.0300	0.2097

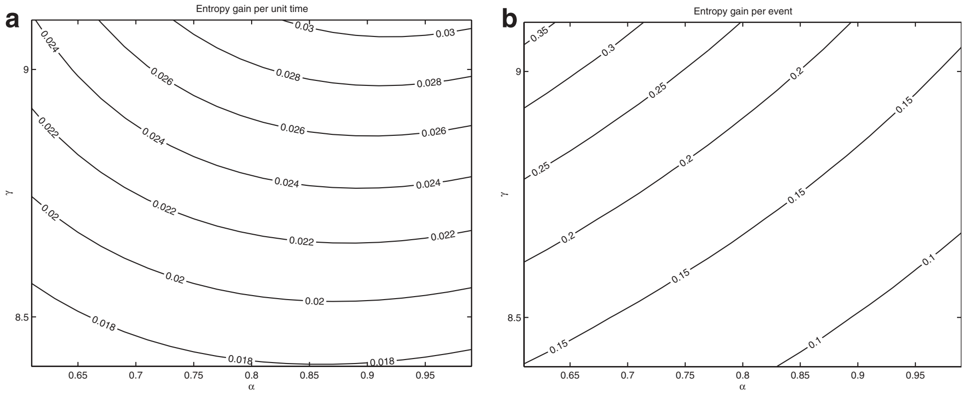

Figure 3
Entropy gain per time unit (left) and per event (right) for the stress release model with $\eta = 0.75, \beta = 0.010, \rho = 1.176$, and the Pareto stress drop distribution (17) for varying α, γ.

values derived from the empirical distributions in Table 3. HARTE and VERE-JONES (2005) show that the differences are not significant, using Monte Carlo inference based on repeated simulation of the model. In other words, the data produced by the model is not substantially different from the observed data, according to this measure.

Using the results for the tapered Pareto distribution, or indeed the truncated Gutenberg-Richter distribution, we can also investigate the dependence of the entropy gains on the magnitude distribution parameters (Fig. 3), and on the process parameters (Fig. 4). We will illustrate this only for the tapered Pareto distribution with $\eta = 0.75$. The other cases show similar behavior. The entropy gain for the tapered Pareto distribution indicates that for a given γ, there is a value of α, within

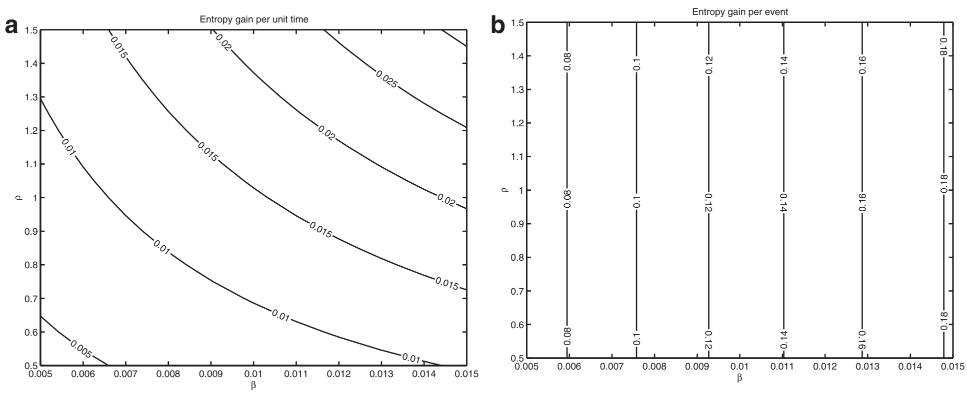

Figure 4
Entropy gain per time unit (left) and per event (right) for the stress release model with the tapered Pareto stress drop distribution (17) using $\eta = 0.75, \alpha = 0.6949, \gamma = 8.5$ and varying β, ρ.

the feasible region, that maximizes the entropy gain. Conversely, we observe that the entropy gain increases with γ, regardless of α. Figure 4 shows that, although the entropy gain increases with ρ, this is solely due to having more frequent events, as we knew from Corollary 1. On the other hand, we see that the entropy gain increases with β, which results from the risk function growing faster with stress, and decreasing the uncertainty about the time of the next event. This accords with the observation of LU and VERE-JONES (2001) who, in fitting the stress release model to a variety of synthetic catalogs with quite different properties, found that the "predictability" increased with β. Further investigation of the truncated Gutenberg-Richter distribution and the tapered Pareto distribution shows that, although the average rate (and hence the entropy gain per event) varys considerably with the 'b-value'parameters θ and α, the entropy gain is relatively insensitive. On the other hand, the entropy gain appears to increase considerably with an increase in the 'upper magnitude cutoff' parameters $m_{\max}$ and γ.

5. *Extension to the Linked Stress Release Model*

An extension of the stress release model to multiple "regions" was proposed by LIU *et al.* (1998). The evolution of stress $X_i(t)$ in the ith region is supposed to follow

$$X_i(t) = X_i(0) + \rho_i t - \sum_j \theta_{ij} S^{(j)}(t),$$

where $S^{(j)}(t)$ is the accumulated stress release in region j over the period $(0, t)$, and the coefficient θ_{ij} measures the fixed proportion of the stress drop, as initiated in region j, which is transferred to region i. Conventionally, $\theta_{jj} = 1$ for all j. This *linked* (or *coupled*) *stress release model* has a point process conditional intensity function

$$\lambda_i(t) = \psi(X_i(t)) = \exp\left(\beta_i\left(X_i(0) + \rho_i t - \sum_j \theta_{ij} S^{(j)}(t) - x_0\right)\right)$$

for each region i. If $\theta_{ij} = \delta_{ij}$, the Kronecker delta, then the process is a simple aggregate of independent stress release models.

For such a multivariate process, we can define a binomial score as before, for each region, and sum them to obtain an overall score

$$B = \sum_k \sum_i (Y_{ik} \ln p_{ik} + (1 - Y_{ik}) \ln(1 - p_{ik})),$$

where k denotes the k-th region. The information gain is thus

$$\bar{\rho}_T = \frac{B - \bar{B}}{T} = \frac{1}{T} \sum_k \sum_i \left(Y_{ik} \ln\left(\frac{p_{ik}}{\bar{p}_{ik}}\right) + (1 - Y_{ik}) \ln\left(\frac{1 - p_{ik}}{1 - \bar{p}_{ik}}\right)\right),$$

where $\bar{B}$ is again the score from a reference process, and the entropy gain can be extended similarly as $\sum_k G_k$, where G_k is the entropy gain in the k-th region taken in isolation.

Proposition 3. *Suppose that events in a process with intensity $\lambda = \lambda(t)$ and entropy gain G are now divided by region, with the result being a multivariate process with intensity $\boldsymbol{\lambda} = \boldsymbol{\lambda}(t) = \{\lambda_k(t)\}$. If $\lambda = \sum_k \lambda_k$ and $\bar{\lambda} = \sum_k \bar{\lambda}_k$ then the entropy gain of the multivariate process $\sum_k G_k > G$.*

The proof is deferred to the Appendix.

Breaking down the data from north China in the previous section into four regions from West to East (see ZHENG and VERE-JONES, 1994) and using the individual empirical magnitude distributions for the four regions, we obtain the results in Table 4. Hence, taking $\theta_{ij} = \delta_{ij}$, and assuming that the reference model is an aggregate of Poisson processes of rate $\{\bar{\lambda}_k\}$, we have aggregate figures for the process of

$$\sum_i \bar{\lambda}_k = 0.1278, \qquad \sum_k G_k = 0.0457, \qquad \text{and} \qquad \frac{\sum_k G_k}{\sum_k \bar{\lambda}_k} = 0.3576.$$

We can compare (using the empirical magnitude distribution in both cases) this with values of $G = 0.0120$ and $G/\bar{\lambda} = 0.0902$ for the univariate case. We should not be surprised at the increase in the entropy gain. Basically, there are four times asmany spatio-temporal ‘‘cells’’, and no more earthquakes, thus we can improve our predictions by forecasting that no event will occur. However, this increase in the entropy gain is predicated on correct model fitting and identification, which the demands of the additional parameters render problematical.

Moreover, if the regions are interacting by transferring stress between themselves, a question arises as to the correct reference model. Assuming the reference model to be a multivariate Poisson process, we still need to calculate appropriate rates $\bar{\lambda}_k$. One solution is to take the rate as $\bar{\lambda}_k = \upsilon_k/\mathbb{E}[S_k]$, where the stress ‘input’ rates $\{\upsilon_k\}$ are found as the solutions to the simultaneous equations

$$\upsilon_k = \rho_k - \sum_{j \neq k} \theta_{kj} \upsilon_j,$$

Table 4

Calculated entropy gains for North China (by region)

k	$\hat{x_{0_k}}$	$E[S_k]$	$\bar{\lambda}_k$	G_k	$G_k/\bar{\lambda}_k$
1	143.3	11.3	0.0430	0.0128	0.2979
2	66.6	7.6	0.0182	0.0042	0.2285
3	47.5	5.4	0.0371	0.0112	0.3023
4	98.1	11.8	0.0295	0.0175	0.5939

for all k (BEBBINGTON and HARTE, 2003). If we consider the linked model, estimating $\{\mathbb{E}[S_k]\}$ by means of the empirical magnitude distributions, and with parameters in the obvious matrix notation: $\boldsymbol{\rho} = 0.204 \times \mathbf{1}$,

$$\mathbf{x}_0 = \begin{pmatrix} 265 \\ 60.2 \\ 45.8 \\ 78.8 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} 0.017 \\ 0.048 \\ 0.088 \\ 0.061 \end{pmatrix}, \quad \boldsymbol{\Theta} = \begin{pmatrix} 1 & -1.53 & 0 & 0 \\ 0.14 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -0.63 & 1 \end{pmatrix}$$

we find that $\boldsymbol{\upsilon} = (0.426\ 0.144\ 0.204\ 0.333)'$. This produces a reference model with $\bar{\boldsymbol{\lambda}} = (0.0377\ 0.0189\ 0.0378\ 0.0282)'$, against which the aggregate of 4 independent stress release models, as in Table 4, has an entropy gain per event of $\sum_k G_k/(\sum_k \bar{\lambda}_k) = 0.3728$. Simulating the actual *linked* model and calculating the average information gain, we obtain an information gain per event of 0.375. This somewhat small (if actually positive) improvement on the aggregate of independent stress release models is in line with the small log-likelihood improvement resulting from the fitted linked model (BEBBINGTON and HARTE, 2003).

6. *Discussion*

We have derived an expression (14) for the entropy gain per unit time that, while not in closed form, is readily calculated to any desired degree of accuracy in the case of several suitable stress drop distributions. It also shows how G increases with β, in a nonlinear fashion illustrated in Figure 4. The dependence on the stress drop distribution J, and on its mean $\mathbb{E}[S]$ is more complex, although the factor $\rho/\mathbb{E}[S]$ is effectively the rate of events, which thus disappears in the entropy gain per event (15).

The numerical results show that the choice of magnitude distribution, and the values of any parameters therein, are a major factor in the entropy gain of the process. LU and VERE-JONES (2001) also noted that the selection of the magnitude distribution is the factor in the stress release model which has the greatest effect on simulated activity, affecting the periodicity and hence predictability of the results. In general, it appears that the entropy gain increases with the likelihood of larger events, which probably indicates that we are gaining in the forecasting of quiescence after very large events, and possibly from forecasting activity following smaller events.

For purposes of comparison, DALEY and VERE-JONES (2004) give a scale, based on the Gamma renewal process with unit mean, for the entropy gain per event (cf. Eq. 3) as a function of the shape parameter κ. We find that the stress release process possesses approximately the degree of predictability of a Gamma renewal process with shape parameter between 2 ($G/\bar{\lambda} = 0.12$) and 3 ($G/\bar{\lambda} = 0.25$) or between 0.6

($G/\bar{\lambda} = 0.11$) and 0.5 ($G/\bar{\lambda} = 0.22$). The former corresponds to more regular event times, and the latter to clustering, which appears the more likely from Figure 2. Fitting the inter-event times by a Gamma distribution we find that the estimate of the shape parameter is $\hat{\kappa} = 0.516$, for which the formula in DALEY and VERE-JONES (2004) predicts $G/\bar{\lambda} \approx 0.20$. This is corroborated by fitting the Lognormal renewal intensity (4), which produces an entropy gain $G = 0.0225$ ($G/\bar{\lambda} = 0.18$). Hence we see that the inclusion of magnitudes in the conditional intensity for the stress release model appears to lead to little information gain, unless $\eta = 1.5$ (seismic moments) is used rather than Benioff strains.

MOLCHAN (1990, 1991) considered the problem of comparing prediction strategies for stationary point process models of earthquake sequences. The idea is to prescribe a prediction threshold and consider the rate of false alarms and missed events as this threshold varies. In the present setting, this is equivalent to setting a hazard threshold λ_0. We can then consider the rate of missed events to be

$$\nu = \mathbb{P}(\lambda(t_i^-) < \lambda_0) \approx \frac{1}{N} \sum_{i=1}^{N} \mathbf{I}_{(0,\lambda_0]}\left(\lambda\left(t_i^-\right)\right)$$

where events occur at times $t_1, \ldots, t_N$. Similarly, we can approximate a false alarm rate as

$$\begin{aligned}\tau &= \mathbb{P}(\lambda(t) > \lambda_0) \\ &\approx \frac{1}{t_N - t_1} \sum_{i=1}^{N-1} (t_{i+1} - t_i) \min\left(1, \max\left(0, \frac{\ln \lambda(t_{i+1}^-) - \ln \lambda_0}{\ln \lambda(t_{i+1}^-) - \ln \lambda(t_i^+)}\right)\right),\end{aligned}$$

where the approximation follows from the form of the intensity (7). By varying the threshold λ_0, we produce Molchan's $\nu - \tau$ diagram. All curves have end-points $(\nu, \tau) = (1, 0)$ and $(\nu, \tau) = (0, 1)$. Random guessing produces a curve $\nu = 1 - \tau$. Predictions improving on this are in the region $\nu < 1 - \tau$, improving to perfect prediction as the curve gets closer to $(\nu, \tau) = (0, 0)$. Figure 5 shows prediction curves for some of the models in Table 3, obtained by simulating the models forward for 10,000 events.

The diagram can be related to the entropy score by expressing the latter in terms of $d\nu/d\tau$ (HARTE and VERE-JONES, 2005). We see that the truncated Gutenberg-Richter distribution provides uniformly poor predictive performance relative to the tapered Pareto and empirical distributions, while the tapered Pareto distribution is slightly better than the empirical. Of particular interest is that the gamma renewal process provides predictive power roughly equivalent to the stress release model, as was indicated by the entropy gains. In absolute terms, we find that the stress release model with $\eta = 0.75$ is likely to be ineffective as a forecasting tool,

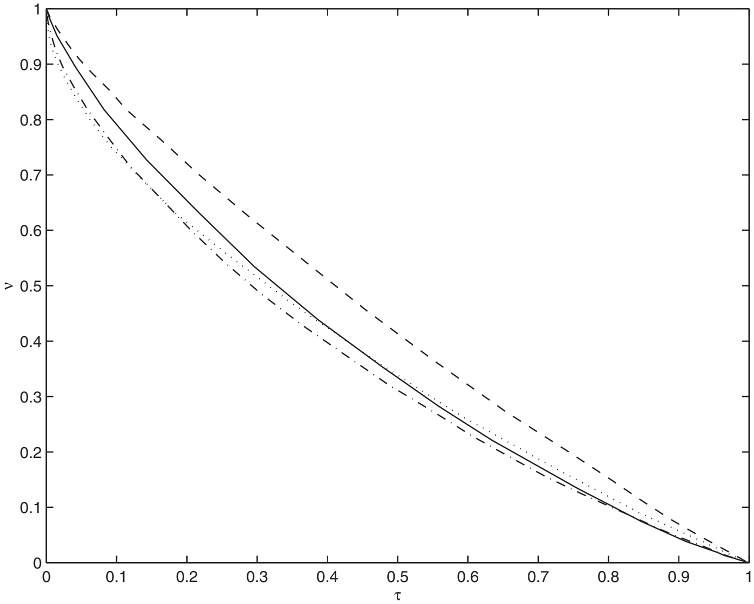

Figure 5
Molchan's $\nu - \tau$ diagram for selected North China models. Stress release models ($\beta = 0.010, \rho = 1.176, \eta = 0.75$): solid line = empirical magnitude distribution, dashed line = truncated G-R with $\theta = 1.22, m_{\max} = 9.0$, dash-dot line = tapered Pareto with $\alpha = 0.695, \gamma = 8.5$. Renewal model: dotted line = gamma renewal process with $\kappa = 0.516$.

as even blocking into 5-yearly intervals (cf., VERE-JONES, 1998) leads to an entropy gain of at best 0.25, or a probability gain of approximately 30% over the Poisson process.

Acknowledgements

This work was supported by the Marsden Fund, administered by the Royal Society of New Zealand. Valuable discussion was provided by David Vere-Jones and David Harte. The author is grateful to the Institute for Mathematics and its Applications, University of Minnesota, for its hospitality.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. (Eds.), *Handbook of Mathematical Functions* (U.S. Govt. Print. Office, Washington 1964).

AKAIKE, H., *On entropy maximization priciple*. In *Applications of Statistics* (ed. Krishnaiah, P.R.) (North-Holland, Amsterdam, 1977) pp. 27–41.

BEBBINGTON, M. and HARTE, D. (2001), *On the Statistics of the Linked Stress Release Process,* J. Appl. Probab. *38A,* 176–187.

BEBBINGTON, M. and HARTE, D. (2003), *The Linked Stress Release Model for Spatio-temporal Seismicity: Formulations, Procedures and Applications,* Geophys. J. Int. *154,* 925–946.

BOROVKOV, K. and VERE-JONES, D. (2000), *Explicit Formulae for Stationary Distributions of Stress Release Processes,* J. Appl. Probab. *37,* 315–321.

DALEY, D.J. and VERE-JONES, D. *An Introduction to the Theory of Point Processes*, Vol. 1, second edition (Springer, New York, 2003).

DALEY, D.J. and VERE-JONES, D. (2004), *Scoring Probability Forecasts for Point Processes: The Entropy Score and Information Gain,* J. Appl. Probab. *41A,* 297–312.

HARTE, D.S. and VERE-JONES, D. (2005), *The Entropy Score and its Uses in Earthquake Forecasting,* Pure Appl. Geophys., 162, 1229–1253.

IMOTO, M. (2004), *Probability Gains Expected for Renewal Process Models,* Earth Planets Space *56,* 563–571.

KAGAN, Y.Y. and KNOPOFF, L. (1977), *Earthquake Risk Prediction as a Stochastic Process,* Phys. Earth Plan. Int. *14,* 97–108.

KLEEMAN, R., MAJDA, A.J., and TIMOFEYEV, I. (2002), *Quantifying Predictability in a Model with Statistical Features of the Atmosphere,* Proc. Nat. Acad. Sci. *99,* 15291–15296.

LIU, J., VERE-JONES, D., MA, L., SHI, Y., and ZHUANG, J.C. (1998), *The Principal of Coupled Stress Release Model and its Application,* Acta Seismologica Sinica *11*, 273–281.

LU, C. and VERE-JONES, D. (2001), *Statistical Analysis of Synthetic Earthquake Catalogs Generated by Models with Various Levels of Fault Zone Disorder,* J. Geophys. Res. *106,* 11115–11125.

MOLCHAN, G.M. (1990), *Strategies in Strong Earthquake Prediction,* Phys. Earth Plan. Int. *61,* 84–98.

MOLCHAN, G.M. (1991), *Structure of Optimal Strategies of Earthquake Prediction,* Tectonophysics *193,* 267–276.

OGATA, Y. (1988), *Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes,* J. Amer. Statist. Assoc. *83,* 9–27.

OGATA, Y. (1999), *Seismicity Analysis through Point-Process Modeling: A Review,* Pure Appl. Geophys. *155,* 471–507.

REID, H.F., *The mechanism of the earthquake*. In *The California Earthquake of April 18, 1906, Report of the State Investigation Commission,* Vol. 2 (Carnegie Institute of Washington, Washington D.C. 1910) pp. 16–28.

SORNETTE, D. and SORNETTE, A. (1999), *General Theory of the Modified Gutenberg-Richter Law for Large Seismic Moments,* Bull. Seismol. Soc. Amer. *89,* 1121–1130.

VERE-JONES, D. (1978), *Earthquake Prediction: A Statistician's View,* J. Phys. Earth *26,* 129–146.

VERE-JONES, D. (1998), *Probabilities and Information Gain for Earthquake Forecasting,* Comput. Seismol. *30,* 248–263.

VERE-JONES, D., ROBINSON, R. and YANG, W. (2001), *Remarks on the Accelerated Moment Release Model: Problems of Model Formulation, Simulation and Estimation,* Geophys. J. Int. *144,* 517–531.

ZHENG, X. (1991), *Ergodic Theorems for Stress Release Processes,* Stoch. Proc. Appl. *37,* 239–258.

ZHENG, X. and VERE-JONES, D. (1994), *Further Applications of the Stochastic Stress Release Model to Historical Earthquake Data,* Tectonophysics *229,* 101–121.

Appendix

Proof of Theorem 1. From (6) and (12) we have

$$G = \beta \mathbb{E}[X\psi(X)] - \beta x_0 \mathbb{E}[\psi(X)] - \frac{\rho}{\mathbb{E}[S]} \ln\left(\frac{\rho}{\mathbb{E}[S]}\right). \tag{18}$$

Differentiating (9) we get

$$\varphi'(s) = iR\varphi(s) + \frac{\gamma'(s)}{\gamma(s)}\varphi(s) + \sum_{k=1}^{\infty}\left(\frac{i}{\beta k} + \frac{\gamma'(s-ik\beta)}{\gamma(s-ik\beta)}\right)\varphi(s). \tag{19}$$

The condition (13) and Lebesgue's Dominated Convergence Theorem allow us to write

$$\chi'(s) = \frac{d}{ds}\int_0^{\infty} e^{-isx}J(dx) = -i\int_0^{\infty} xe^{-isx}J(dx),$$

and

$$\chi''(s) = \frac{d^2}{ds^2}\int_0^{\infty} e^{-isx}J(\mathrm{d}x) = -\int_0^{\infty} x^2e^{-isx}J(\mathrm{d}x),$$

hence $\chi'(0) = -i\mathbb{E}[S]$ and $\chi''(0) = -\mathbb{E}[S^2]$. From (10),

$$\gamma'(s) = \frac{-is\chi'(s)\mathbb{E}[S] - (1-\chi(s))i\mathbb{E}[S]}{-s^2\mathbb{E}[S]^2} = \frac{i(s\chi'(s) + 1 - \chi(s))}{s^2\mathbb{E}[S]},$$

and thus

$$\frac{\gamma'(s)}{\gamma(s)} = -\frac{s\chi'(s) + 1 - \chi(s)}{s(1-\chi(s))} = -\frac{\chi'(s)}{1-\chi(s)} - \frac{1}{s}.$$

Also, $\chi(0) = 1 = \gamma(0)$ and so appealing to L'Hôpital's rule we find that

$$\gamma'(0) = \lim_{s\to 0}\frac{i(s\chi''(s) + \chi'(s) - \chi'(s))}{2s\mathbb{E}[S]} = \frac{i\chi''(0)}{2\mathbb{E}[S]} = \frac{-i\mathbb{E}[S^2]}{2\mathbb{E}[S]}.$$

Substituting into (19), and noting that $\varphi(0) = 1$, we have

$$\begin{aligned}\mathbb{E}[X] &= -i\varphi'(0)\\ &= R - i\left(\frac{-i\mathbb{E}[S^2]}{2\mathbb{E}[S]}\right) - i\sum_{k=1}^{\infty}\left(\frac{i}{\beta k} + \frac{\gamma'(-ik\beta)}{\gamma(ik\beta)}\right)\\ &= x_0 + \beta^{-1}(\ln(\beta\rho) - \Gamma) - \frac{\mathbb{E}[S^2]}{2\mathbb{E}[S]} + \sum_{k=1}^{\infty}\frac{\int_0^{\infty} xe^{-k\beta x}J(dx)}{1 - \int_0^{\infty} e^{-k\beta x}J(dx)}.\end{aligned} \tag{20}$$

Substituting (20) into (11) yields

$$\mathbb{E}[X\psi(X)] = \frac{\rho}{\mathbb{E}[S]}\left(x_0 + \beta^{-1}(\ln(\beta\rho) - \Gamma) + \sum_{k=1}^{\infty}\frac{\int_0^{\infty} xe^{-k\beta x}J(dx)}{1 - \int_0^{\infty} e^{-k\beta x}J(dx)}\right).$$

The result follows by substituting this into (18), and observing that $\mathbb{E}[\psi(X)] = \rho/\mathbb{E}[S]$ and $\ln(\beta\rho) - \ln(\rho/\mathbb{E}[S]) = \ln(\beta\mathbb{E}[S])$. □

Proof of Proposition 3. By definition, the entropy gain of the multivariate process is

$$\begin{aligned}\sum_k G_k &= \sum_k \mathbb{E}\left[\lambda_k \ln\left(\frac{\lambda_k}{\bar{\lambda}_k}\right)\right] \\ &= \sum_k \mathbb{E}\left[\lambda_k\left(\ln\left(\frac{\lambda_k}{\sum_j \lambda_j}\right) + \ln\sum_j \lambda_j - \ln\sum_j \bar{\lambda}_j - \ln\left(\frac{\bar{\lambda}_k}{\sum_j \bar{\lambda}_j}\right)\right)\right] \\ &= \mathbb{E}\left[\sum_k \lambda_k\left(\ln\sum_j \lambda_j - \ln\sum_j \bar{\lambda}_j\right)\right] \\ &\quad + \mathbb{E}\left[\sum_k \lambda_k\left(\ln\left(\frac{\lambda_k}{\sum_j \lambda_j}\right) - \ln\left(\frac{\bar{\lambda}_k}{\sum_j \bar{\lambda}_j}\right)\right)\right].\end{aligned}$$

The first term is the entropy gain of the univariate process, and the second term is always positive by Jensen's inequality. □

(Received August 11, 2004, revised January 11, 2005; accepted February 10, 2005)
Published Online First: July 29, 2005