Geophys. J. Int. (2001) **144,** 517–531

Remarks on the accelerated moment release model: problems of model formulation, simulation and estimation

David Vere-Jones,[1] Russell Robinson[2] and Wenzheng Yang[3]

[1] *Victoria University, Wellington, New Zealand*
[2] *Institute of Geological and Nuclear Sciences,* PO Box 30-368, *Lower Hutt,* 6315 *New Zealand. E-mail: r.robinson@gns.cri.nz*
[3] *Centre for Analysis and Prediction, China Seismological Bureau, Beijing, China*

Accepted 2000 September 25. Received 2000 September 14; in original form 2000 April 26

SUMMARY

This report summarizes a variety of issues concerning the development of statistical versions of the so-called 'accelerated moment release model' (AMR model). Until such statistical versions are developed, it is not possible to develop satisfactory procedures for simulating, fitting or forecasting the model. We propose a hierarchy of simulation models, in which the increase in moment is apportioned in varying degrees between an increase in the average size of events and an increase in their frequency. To control the size distribution, we propose a version of the Gutenberg–Richter power law with exponential fall-off, as suggested in recent papers by Kagan. The mean size is controlled by the location of the fall-off, which in turn may be related to the closeness to criticality of the underlying seismic region. Other points touched on concern the logical structure of the model, in particular the identifiability of the parameter assumed to control the size of the main shock, and appropriate procedures to use for simulation and estimation. An appendix summarizes properties of the Kagan distribution. The simulations highlight the difficulty in identifying an AMR episode with only limited data.

Key words: accelerated moment release, frequency–magnitude distribution, Kagan distribution, simulated catalogues.

1 INTRODUCTION

The accelerated moment release model (AMR model for short) has been explored in a number of recent papers (see e.g. Bufe & Varnes 1993; Sornette & Sammis 1995; Bowman *et al.* 1998; Jaumé & Sykes 1999; Robinson 2000). It is defined in its simplest form by the equation

$$\varepsilon(t) = A - B(t_f - t)^m . \tag{1}$$

Here, the left-hand side, $\varepsilon(t)$, has been variously interpreted as the accumulated seismic moment, the energy release or the Benioff strain release within a specified seismogenic region, from some origin time t_0, say, up to time t. In what follows we shall adopt the interpretation in terms of strain release, although a parallel development could be carried through for the other interpretations. Our reasons for this choice are essentially pragmatic: it follows the majority of previous papers; it represents a compromise between the extremes of total energy release and pure numbers, which seems to work as well as any other choice in practice; and it leads to some interesting but perhaps purely coincidental identifications of parameter values. Whichever interpretation is taken, the idea that the approach to a critical point is associated with power-law behaviour of this general type is supported by deductions from a variety of physical models (see e.g. the review in Main 1999 and further references therein).

The remaining quantities, A, B, t_f and m, are parameters characterizing the seismic episode under study. In empirical studies, if we adopt the strain release interpretation,

$$\varepsilon(t) = \sum_{1}^{N(t)} E_i^{1/2} , \tag{2}$$

where $N(t)$ is the number of events in the region between t_0 and t, and the $\{E_i;\, i \geq 1\}$, denote the successive energy releases from earthquakes in the region. In what follows, we shall write $E_i^{1/2} = S_i$. For most regions, the strain releases have to be roughly estimated from the magnitudes using the relation

$$S_i = E_i^{1/2} = 10^{2.4 + 0.75 M_i} \tag{3}$$

(Kanamori & Anderson 1975), although in principle the seismic moment might offer a better starting point for such studies.

The parameter t_f is interpreted as the origin time of the final event in the sequence, representing the next event whose source region is commensurate with the size of the region being studied, and the exponent m is assumed to lie in the range

$0<m<1$; indeed, there are some theoretical problems if m lies outside this range (see e.g. Main *et al.* 2000). The deficit

$$S_f = A - \sum_{t_i<t_f} S_i \tag{4}$$

is sometimes interpreted as predicting the size (i.e. the Benioff strain release) of the final event, but as we shall see there are difficulties with this interpretation.

Clearly, eq. (1) can describe only in an approximate, average sense the actual behaviour of a particular seismic cycle, if only because the left side is an irregularly increasing step function, while the right side is continuous. The main purpose of this paper is to examine how this equation might be given a more precise sense within a fully specified (if still approximate) stochastic model. Such a model is needed in order to simulate, and hence to forecast, systems exhibiting the AMR type of behaviour, and as a guide to determining the most effective ways of estimating the parameters of the model. At the same time, the exercise of trying to think through the structure of a stochastic model for eq. (1) shows up a number of logical issues that require further clarification.

2 MODEL DEFINITION

The usual approach in developing a stochastic model for a situation described in broad terms by a deterministic equation (as, for example, in developing models for population growth or the spread of epidemics) is to treat the approximating deterministic equation as relating to the expected value (ensemble average) of the random quantity it purports to represent. If we take this point of view, then eq. (1) should be replaced by

$$\mathscr{E}\left[\sum_{1}^{N(t)} S_i\right] = A - B(t_f - t)^m , \tag{5}$$

where $\mathscr{E}$ denotes the expectation (ensemble average). However, there are two notes of caution that should be sounded in regard to this equation. First, it is not clear whether, in the present context of long-tailed or fractal distributions, the expectation necessarily exists. If it does not, then some alternative method of describing the approximate average behaviour of the system is needed—perhaps the median of the distribution of $\varepsilon(t)$ might be taken instead of the mean. The mean has special advantages, however, and to preserve them we shall restrict the discussion in this paper to situations where the expected strain release is finite. In particular, we shall use a form of the frequency/strain release distribution that has an exponential fall-off at infinity and so has finite mean and indeed finite moments of all orders. The second point is that, even if eq. (5) is meaningful, it is far from sufficient in itself to specify a full stochastic model. Thus the task remains of finding fully specified models for which eq. (5) is valid, and of selecting between them.

As a first step in approaching the latter problem, we may observe that the quantity $\varepsilon(t)$ in the original equation contains two irregular components: the *number* of events, namely $N(t)$, and the *sizes* of the individual events, namely $S_i = E_i^{1/2}$. We can clarify the interpretation of the expectation in eq. (5) by introducing an instantaneous rate (intensity function),

$$\lambda(t) = \mathscr{E}\left[\frac{dN(t)}{dt}\right], \tag{6}$$

and a time-varying mean event size,

$$\mu(t) = \mathscr{E}[S \mid S \text{ occurs at } t] . \tag{7}$$

Then the expected increment over a short time interval dt can be written as

$$\mathscr{E}[d\varepsilon(t)] = \lambda(t)\mu(t)dt , \tag{8}$$

so that eq. (1) can be rewritten in differential form as

$$\lambda(t)\mu(t) = Bm(t_f - t)^{-(1-m)} . \tag{9}$$

This seems to be the most plausible way of capturing the intention of eq. (1) within the extended framework of a stochastic model. See also the discussions in Varnes (1989) and Bufe & Varnes (1993) and the more extended review in Main *et al.* (2000).

In the simple simulation models considered below, $\lambda(t)$ will be interpreted as the intensity of a Poisson process with time-varying intensity function, and the successive events will be assumed independently distributed, even though the distributions may vary with time. However, considerably more complex dependence structures could be envisaged without altering the validity of eq. (9). In particular, both the rate at time t and the size of the next event in the sequence could depend on the times and sizes of earlier events in the sequence; all that eq. (9) describes is the average to be expected over many such sequences.

Another important point to note about eq. (9) is that the constant A has dropped out of the equation entirely. This is not a consequence of the change to a stochastic model, but of the change to a description in terms of rates. It implies that, if the model equation eq. (9) is accepted, then *nothing about the value of A (or about the magnitude of the final event, if that can be related to A) can be learned from the expected increments of* $\varepsilon(t)$ *between* t_0 *and* t_f. If any information is to be gleaned about the final magnitude, it must come from aspects of the process not captured in that model, possibly from the conditions in place at the start of observations, the size of the region to which the model is applied, or other factors lying outside the growth of the cumulative strain release. This conclusion impacts on the interpretation that can be given to estimates of A obtained from a least-squares fitting procedure applied to successive values of $\varepsilon(t)$. It implies that such estimates are basically noise values, derived from incautiously applying an estimation technique that was not designed for fitting cumulative data curves (see further the comments in Section 4 below).

Finally, in connection with eq. (9), the problem arises of apportioning the increase in expected strain release between the two factors on the right. Is the increase caused by an increase in the frequency of events, or an increase in their average size, or do both factors play some role? We have found little discussion of this issue in the earlier papers on the topic. Analogies with the approach to criticality in a phase-change context suggest that the size of events—perhaps more accurately the potentiality for larger events to occur—increases as criticality is approached, and correlation radii increase. Other evidence, for example, the increase of acoustic emissions as the failure point is reached, point to an increase in the frequency. The balance is rarely discussed, however. Real examples are equally hard to pin down in terms of general tendency, as can be seen from the summary in Jaumé & Sykes (1999); each example tends to have particular features that dominate the appearance of the sequence and are peculiar to that sequence.

One possible approach, outlined in Vere-Jones (in preparation), is to apply the maximum-entropy formalism to the problem, much along the lines of Main & Burton (1984). In this approach, one seeks the model giving maximum entropy to an observed sequence of events, subject only to constraints on the mean values such as that implied by eq. (9). As in the original statistical mechanical problems studied by Boltzmann and Gibbs, the detailed behaviour of successive events is not known, but if the underlying assumptions made on the distributions via the mean equation (9) are at all correct, observation sequences giving empirical distributions close to the maximum entropy solution should be overwhelmingly more common than observation sequences following some other type of pattern.

The most important conclusions to follow from this approach are ones that might have appeared as natural first assumptions in any case. They can be summarized as follows:

(i) the event times should be modelled as a non-stationary Poisson process, with time-varying intensity, say $\lambda(t)$;
(ii) the event sizes should be treated as independent random variables;
(iii) the distribution of the event sizes should follow a distribution with density function with power-law tail and exponential fall-off;
(iv) conditional on the time t at which it occurs, the mean $\mu(t)$ of this distribution and the rate $\lambda(t)$ of the Poisson process of occurrence times should each be proportional to the square root of their product (eq. 9).

In the absence of any other guidance, and without claiming any great authority for them, we shall use these assumptions in the next section as the basis for developing a series of simulation models. The purpose is to examine the behaviour of the models and to provide visual comparisons with real AMR sequences. The third assumption plays a particularly important role in our models and warrants further comment.

The need to use a distribution for the strain release somewhat more flexible than that corresponding to the usual Gutenberg–Richter (G–R) form was impressed on us by considerations quite apart from the link with maximum entropy. Initially we assumed that the required changes in mean value could be accomplished simply by changes in the b-value, but a number of difficulties with this approach revealed themselves. In particular, the empirical magnitude distribution resulting from sequences of events with changing b-values has the form of a G–R plot with reverse curvature to that usually seen in real G–R plots—for large magnitudes, the slope *decreases* with increasing magnitude. The more usual appearance is a linear plot or a plot with increasing steepness at high magnitudes or, in some cases at least, a plot with a 'kink' at high magnitudes similar to that described by Wesnousky (1994) in connection with the characteristic earthquake model. Thus, if the observed changes in mean strain release were to be accomplished by a change in the magnitude distribution with time, then something other than the simple G–R law seemed to be required.

The candidate that then suggested itself was a power-law form (corresponding to the 'pure' G–R relation) modified by an exponential fall-off in the far right tail. The version that we use in this paper, and call the *Kagan distribution*, has the explicit parametric form

$$\mathrm{Prob}(S-S_0>x)=1-F_S(x)=\left(1+\frac{x}{L+S_0}\right)^{-\alpha}\mathrm{e}^{-\frac{x}{U}}\quad(x>0)\,. \tag{10}$$

This corresponds to the density (obtained by differentiation)

$$f_{S|S_0}(x)=\frac{1}{(S_0+L)}\left(\frac{x+L}{S_0+L}\right)^{-(\alpha+1)}\mathrm{e}^{-\frac{x-S_0}{U}}\left\{\alpha+\frac{L+x}{U}\right\}. \tag{11}$$

Plots of the log survivor function and the density function for this distribution are shown in Figs 1 and 2; see the Appendix for further discussion and a development of basic properties.

In these expressions, S_0 is the threshold strain release, corresponding to the threshold magnitude M_0 for the catalogue in

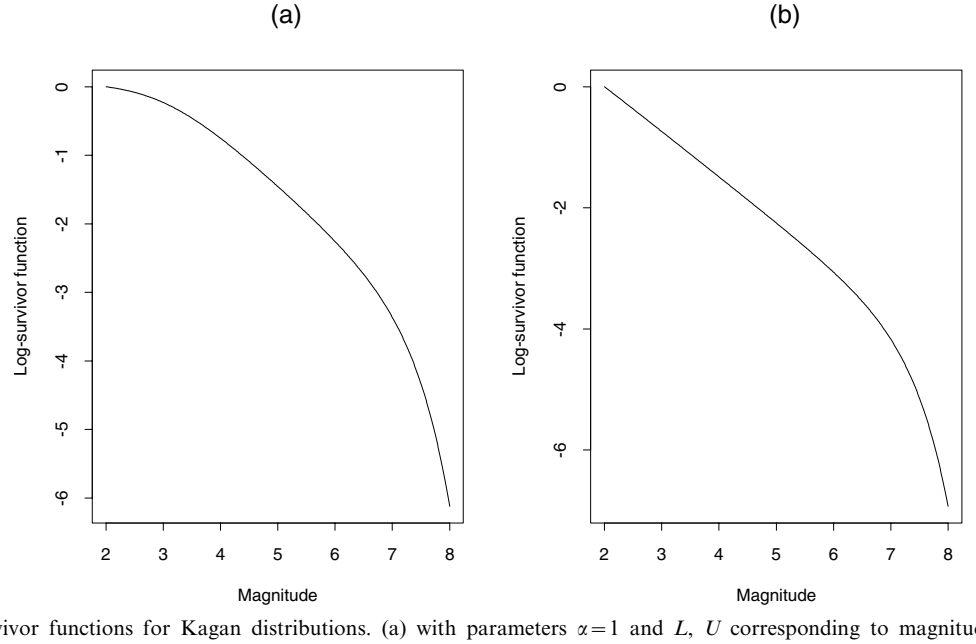

Figure 1. Log survivor functions for Kagan distributions. (a) with parameters $\alpha=1$ and L, U corresponding to magnitudes $\beta=3$ and $\gamma=7$, respectively. (b) two-parameter model; as (a) but with $\beta=0$.

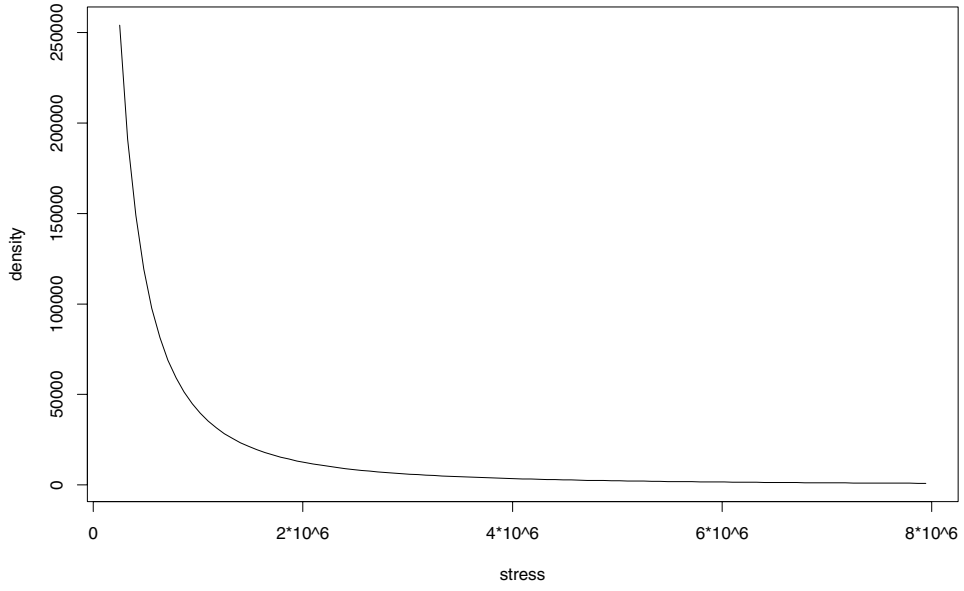

Figure 2. Density for Kagan distribution; parameters as in the three-parameter version of the Fig. 1.

use, $S = x + S_0$ is the strain release, while α is proportional to the b-value in the usual Gutenberg–Richter plot, and determines the slope in the linear portion of the log–log plot. The two parameters L and U both have the dimensions of strain (that is, the same dimensions as the variable S itself) and control the positions of the lower and upper extremes of the linear portion of the log–log plot (see Fig. 1). U corresponds roughly to the *maximum moment* in Kagan's papers, or to the correlation length in percolation studies. L corresponds roughly to a minimum moment, or minimum resolvable crack length. Frequently, we assume $L \approx 0$, so that the lower turning point in the log survivor curve is effectively absent, and the original *three-parameter model* is replaced by a *two-parameter model*. In the diagrams, where the strain axis is frequently replaced by a magnitude axis, we represent L and U by their corresponding equivalent magnitudes, β and γ respectively.

It is of interest to note that mixtures of the Kagan distribution, with the mixing taking place over the parameter U, show not the reverse curvature that moved us away from the straight G–R form, but a range of behaviours that includes straight-line curves with slope greater than 1 (higher b-values), and curves with a kink in the log frequency–magnitude plot similar to that sometimes used as evidence for the characteristic earthquake model. Examples of log survivor function plots (analogous to the usual Gutenberg–Richter plots) for both the basic distribution and for mixtures are shown in Figs 3 and 4 and discussed further in the Appendix. The mixture plots suggest that variations in apparent b-value, as well as more extreme features, could arise naturally if observations were taken over varying stress conditions in the seismogenic zone.

3 SIMULATION MODELS

We proceed to outline a selection of simple simulation models, all satisfying the basic requirement expressed by eq. (9), but corresponding to different apportionments of the increase in mean strain release between frequency and mean event size. For the sake of definiteness we have taken all models to correspond to AMR sequences with $m = 0.5$ in eq. (1). The increases in rate and mean event size are taken as being proportional to $[Bm(t_f - t)^{-(1-m_1)}]$ and $[Bm(t_f - t)^{-(1-m_2)}]$ respectively, where m_1, m_2 are adjustable, subject to $0 \le m_1 \le 1$, $0 \le m_2 \le 1$, $m_1 + m_2 = 1.5$.

The simulation methods outlined are all extensions of the basic methods for simulating Poisson processes with time-varying intensities. The suite of programmes used for this purpose is available from the first author, and will be incorporated into the next edition of the Statistical Seismological Library (SSLib; see Harte 1999).

3.1 Poisson process with increasing rate function and independent magnitudes

Probably the simplest candidate for a stochastic AMR process is a Poisson process with mean rate following the right-hand side of eq. (9), and magnitudes independently selected either from a Gutenberg–Richter law with fixed b-value, or from a Kagan distribution with fixed parameters L, α and U.

The simulation of such a process involves the following components.

(i) In the range $0 < t < t_f$, generate a sequence of event times $\{t_i\}$ for a Poisson process with time-varying intensity,

$$\lambda(t) = \lambda_0 + d\left(1 - \frac{t}{t_f}\right)^{-\kappa}, \tag{12}$$

where t_f, λ_0, d and κ are adjustable parameters that can be selected by the user. In terms of the previous notation, if $E(S)$ is the expected mean strain release for the sequence and $\lambda_0 = 0$, then $d = Bmt_f^{1-m}/E(S)$ and $\kappa = 1 - m$. The option of adding in a constant background rate λ_0 represents a minor extension of the model that seems physically sensible, and leads to a natural comparison of fit between models with $d = 0$ and $d > 0$. It follows Gross & Rundle (1998) and Main (1999), who pointed out that in real-data applications, the background rate may have a greater weight than the AMR component. This stage

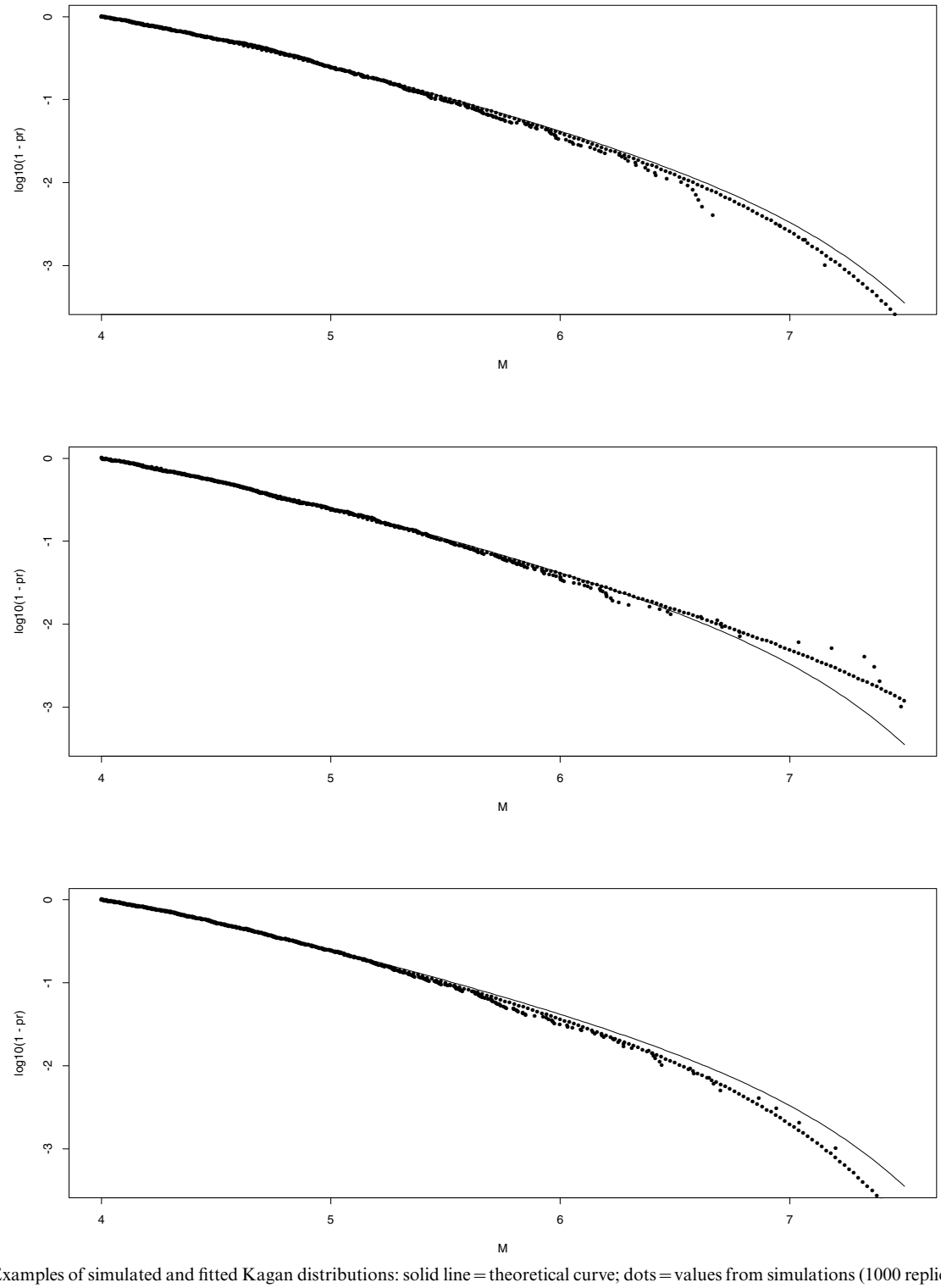

Figure 3. Examples of simulated and fitted Kagan distributions: solid line = theoretical curve; dots = values from simulations (1000 replicates); dotted line = curve fitted to simulated data.

of the simulation can take advantage of standard methods for point processes, using variants of either the Shedler–Lewis thinning method (Lewis & Shedler 1976; Ogata 1981) or the inversion method.

(ii) Generate an associated sequence of strain releases S_i. This can be done either for the simple G–R law or for the Kagan distribution. Since the former is a special case of the latter (set $U = \infty$), it will be enough to describe a procedure for simulating events from the Kagan distribution, assuming α, L, U and the threshold strain release S_0 given. In practice, it is usually more convenient to specify the quantities L, U and S_0 in terms of the equivalent magnitudes (converted via eq. 3 or something similar to it), which we denote by β, γ and M_0 respectively.

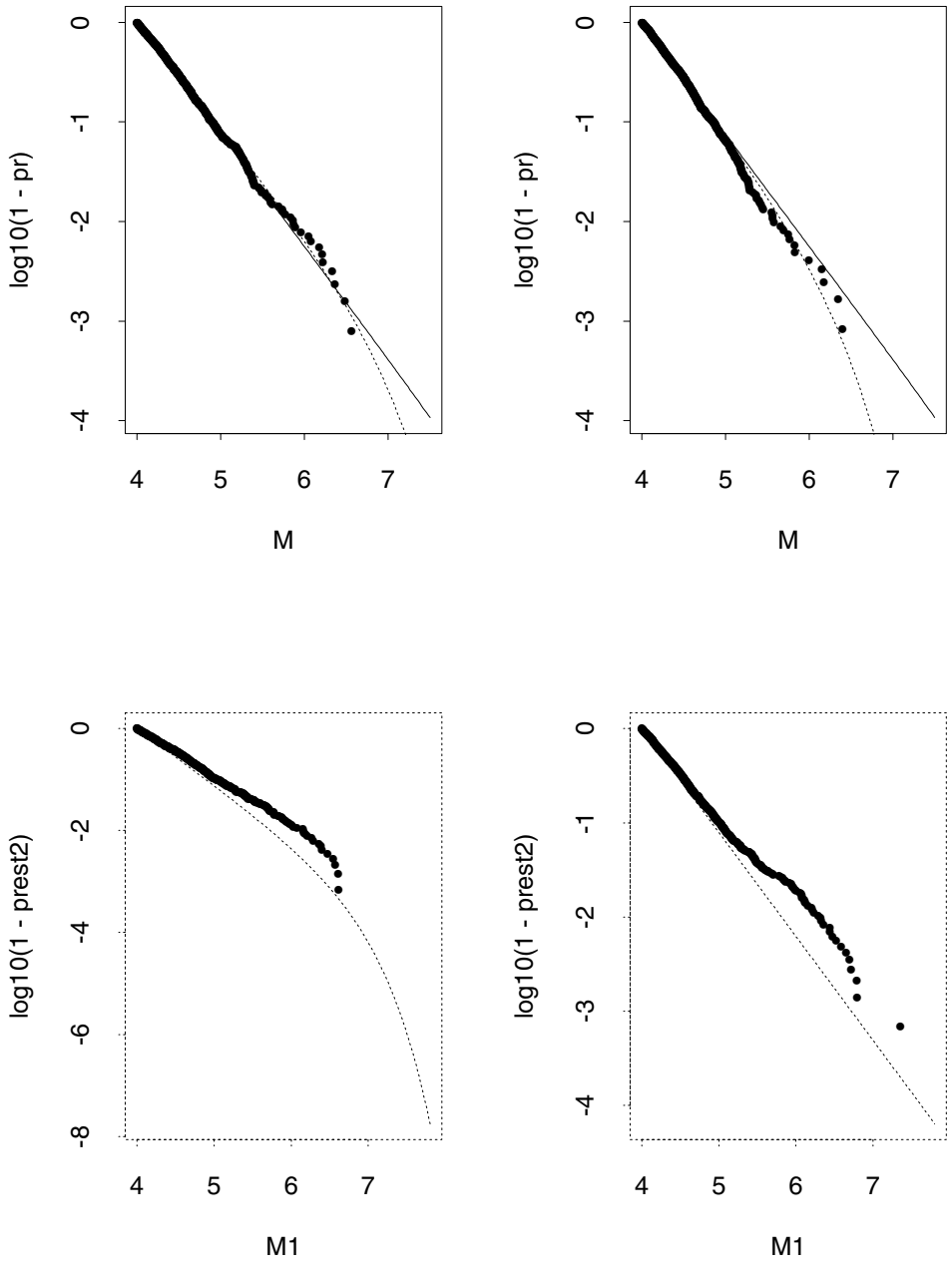

Figure 4. Examples of mixtures of Kagan distributions. Upper graphs: pure gamma mixture showing steepened slope; lower graphs: truncated gamma mixture showing a 'Wesnousky kink'; dotted lines represent the best-fitting Kagan distribution (see Appendix for details).

A simple way of accomplishing this simulation, based on the interpretation of the product form in the survivor function as the distribution of a minimum, is outlined in the Appendix. The procedures (i) and (ii) together result in a sequence of pairs (t_i, S_i) with increasing frequency and strain releases S_i having a fixed distribution.

Two examples of AMR sequences simulated in this manner are illustrated in Fig. 5. The parameters for the first pair of graphs in Fig. 5 have been adjusted to give a rather high occurrence rate (about 250 events over a 20 yr period with threshold magnitude $M_0 = 4$; of these, about 60 per cent come from the AMR signal, the rest from background) in order to illustrate the nature of the relationship. Even for a more realistic number of events (about 40, as in Fig. 5b), the increase in frequency is close to being detectable. Moreover, with this version of the model, the change in frequency should persist at all magnitudes, so the effect could be confirmed by checking the behaviour for smaller events. We discuss methods of estimating the parameters from the data in Section 4.

3.2 Poisson process with constant occurrence rate but increasing mean strain release

Here we assume that the mean rate is constant and the mean strain release varies with time according to the AMR relation. More precisely, in eq. (8) we assume

$$\lambda(t) = \lambda = \text{const.}, \tag{13}$$

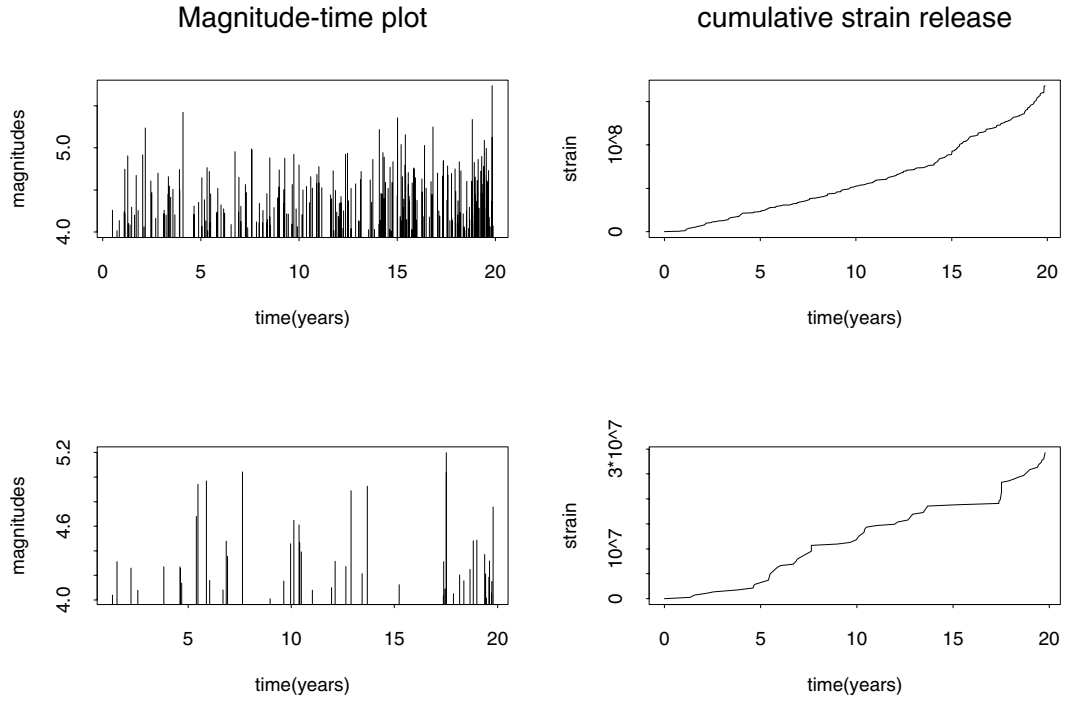

Figure 5. Simulations from the AMR model: AMR signal in frequency only.

and that the instantaneous value of the expected strain release changes according to an equation of the form

$$\mu(t) = \mu_0 + h\left(1 - \frac{t}{t_f}\right)^{-\kappa}. \tag{14}$$

To implement this result we must find an expression for the mean of the Kagan distribution in terms of its parameters, and then solve for the parameters given a particular value of the mean $\mu(t)$. Expressions for the mean (and higher moments) can be obtained in terms of the incomplete gamma function ($\alpha > 1$), or the exponential integral if $\alpha = 1$. Details are given in the Appendix.

Since the distribution involves three parameters, the problem is indeterminate unless we incorporate some additional assumptions. The assumptions we shall adopt are that $\alpha = 1$ and that L corresponds to an event with magnitude 0 (so $\beta = 0$). The latter effectively implies that we are using a high enough magnitude threshold to ensure that the G–R relation shows no lower-end curvature due to catalogue incompleteness. The mean (expected strain release) is then a function $m(U)$ of the upper parameter U, and we have to solve the equation $m(U) = \mu(t)$ to find the U-value corresponding to the value of the mean at any given time t.

If we again express L and U in terms of equivalent magnitudes β and γ, a reasonable range of variation will be obtained if γ varies from around 5 or 6 for the background events up to 7 or 8 as the AMR region approaches a critical state. In principle, γ approaches infinity as the failure time t_f is approached. This does not mean an infinite b-value, however, just an approach to the pure G–R form with b-value around $b = 0.75$, corresponding approximately to the value $\alpha = 1$, for which the mean strain release is already infinite. Of course, the assumptions incorporated here represent only one of many variations that may have been adopted.

The simulation steps implementing these ideas can be summarized as follows.

(i) Simulate the time points t_i using any standard method for a constant-rate Poisson process.

(ii) For each i, set $m(U) = \mu(t_i)$ and solve for U. This creates a sequence of values of U_i that, with the assumptions $\alpha = 1$, $\beta = 0$ and threshold magnitude $M_0 = 4$, uniquely specifies the distribution to be used at time t_i. The explicit relationship between U_i and the equivalent magnitude γ_i is here given by

$$U_i = S_0 \exp\left[\frac{\mu_0 + h\left(1 - \dfrac{t_i}{t_0}\right)^{-\alpha}}{S_0} + 1 - \gamma_i\right]. \tag{15}$$

(iii) For each i, simulate a stress release S_i according to the Kagan distribution with parameters $\alpha = 1$, $\beta = 0$, $\gamma = \gamma_i$, $M_0 = 4$ using the procedure outlined in the preceding example.

Examples of simulations from this version of the model are shown in Fig. 6. For small to moderate data runs, the changes are harder to pick out, and the effects are confined to magnitudes commensurate with the values of γ occurring within the sequence. Thus, the changes would not be any more noticeable (perhaps even less so) if the magnitude threshold were to be lowered.

One feature of models making use of the Kagan distributions in this way is the very small variation in mean magnitude (and hence in apparent b-value) caused by the approach to criticality. Let us take the conventional value $\alpha = 1$, assume that L is negligible relative to the threshold S_0 corresponding to the magnitude threshold M_0, and examine the effects of changing the upper parameter U on the mean strain release and the mean magnitude. For this purpose it is convenient first to represent U in terms of an equivalent magnitude γ via the approximate

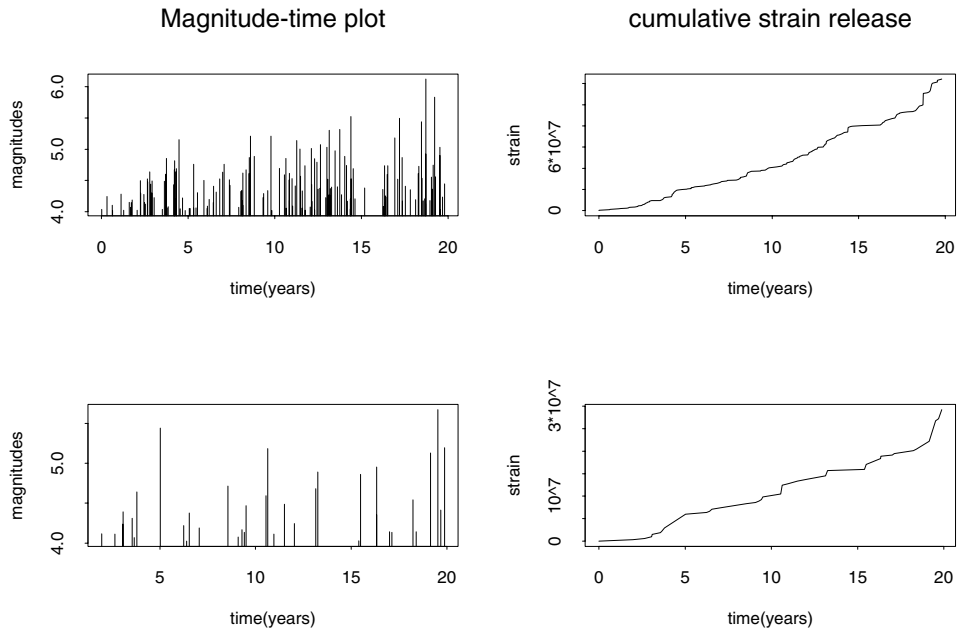

Figure 6. Simulations from the AMR model: AMR signal in mean strain release only.

conversion formula eq. (3). For any given value of γ we may then compute the mean strain release $E(S)$ and its equivalent magnitude. For example, if $\gamma = 5.5$, corresponding to $U = 3.35 \times 10^6$, we find the mean strain release is approximately 8.18×10^5, with equivalent magnitude 4.68. By the time γ has increased to 7.5, the corresponding figures are 1.63×10^6 and 5.08 respectively. This behaviour contrasts sharply with that of the mean magnitudes $E(M)$ themselves, which increase from $E(M) = 4.48$ when $\gamma = 5.5$ to $E(M) = 4.58$ when $\gamma = \infty$, corresponding to a change in the equivalent b-values from 0.90 to 0.75. Moreover, the changes in b-value will be almost imperceptible unless the sample is large enough to contain enough data points in the far tail to pick up the changes in the distribution—the exception rather than the rule in practical applications of the AMR methodology. The moral is that even if the stress field affects the upper tail of the distribution quite markedly, the changes will be hard to pick up from the study of short data runs.

3.3 Models in which both rate and mean strain release change with time

At this point there is no particular difficulty in simulating from models in which both $\lambda(t)$ and $\mu(t)$ are changing with time. The only constraint implied by eq. (1) is that their products should combine as in eq. (8). This can be done in very many ways; one simple way of cutting the Gordian knot is to take

$$\lambda(t) = \exp\left[C + m_1 \log\left(1 - \frac{t}{t_f}\right)\right],$$

$$\mu(t) = \exp\left[D + m_2 \log\left(1 - \frac{t}{t_f}\right)\right],$$

where the ratio $p = m_1/(m_1 + m_2)$ is either fixed or treated as an additional parameter to be estimated. Once the parameters are determined, the simulations can proceed as described in the previous items. Fig. 7 gives examples with $p = 1/2$. Again, the changes are quite hard to pick up from short data runs. They would be reinforced to some extent by lowering the threshold magnitude.

No particular features of interest are observed if the magnitudes from sequences such as those in Fig. 6 are plotted in standard Gutenberg–Richter format. Indeed, the data are often well fitted by a regular Kagan distribution (Fig. 8).

Closer analysis shows that the values of α are larger than in the base model (typically the observed values give $\alpha \sim 1.15$–1.25, whereas the base model has $\alpha = 1$. The situation is similar to that illustrated in Fig. 3, where the mixture has the effect of altering the value of α without otherwise altering the character of the distribution. In fact, there is more curvature in the graphs plotted in Fig. 8 than would be expected from the Kagan distribution itself. This is shown up in the sensitivity of the estimates to the magnitude threshold, but is not readily evident by visual inspection. It is also notable that the parameter estimates are relatively unstable, even with the unrealistically large numbers of events (about 700) used in these two simulations.

4 TESTING AND ESTIMATION PROBLEMS FOR THE AMR

The underlying difficulty in approaching these issues is that because the initial model is not fully defined, it is not possible to state categorically whether one method of estimation is likely to be better (produce smaller estimation errors, or provide more sensitive tests) than another. In general, the best procedures depend on the model. Optimal procedures for a well-defined model are generally available through maximum likelihood techniques, which are then to be preferred; when the underlying model is uncertain, however, robust techniques, which give reasonable estimates under a wider range of model assumptions, may be preferable. In this section we outline and comment on several approaches to estimation and detection problems, both

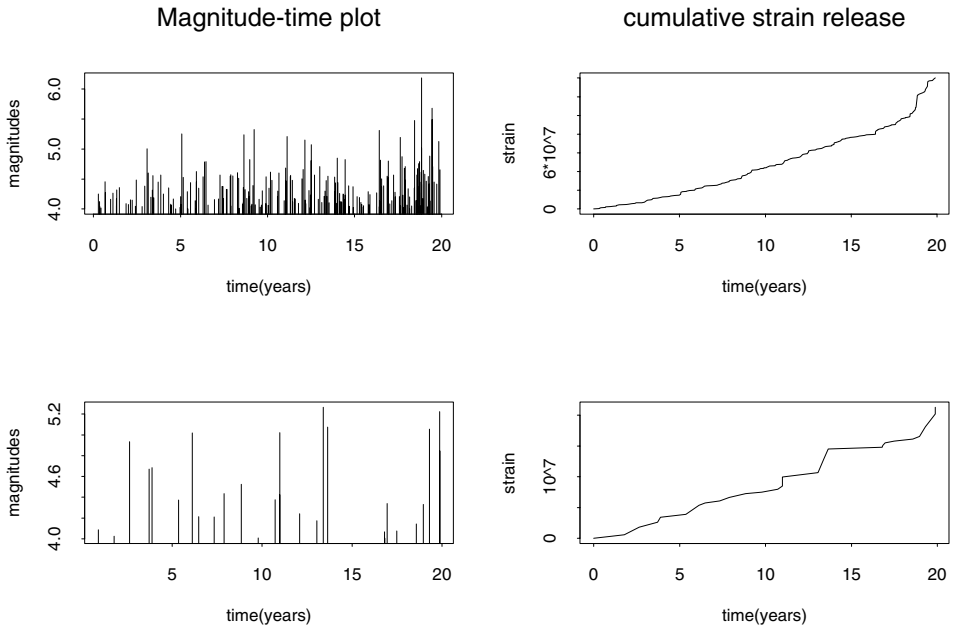

Figure 7. Simulations from the AMR model: AMR signal divided between frequency and mean strain release.

for the initial model (eq. 1) and for the more explicit models that were used in the simulations. We restrict ourselves here to the discussion of general issues and principles; we hope to pursue empirical studies, including in particular parameter estimation and hypothesis testing for the various simulation models considered earlier, in a later study.

4.1 Least-squares fitting to the cumulative moment release curve

In the notation of eq. (1) this means choosing the parameters A, B, t_f and m to minimize the sum of squares

$$\sum (\varepsilon(t_k) - (A - B(t_f - t_k)^m)^2 . \quad (16)$$

The time points (t_k) here may either be equispaced or represent the event times. Despite being the most commonly used technique in the literature so far, it has several major disadvantages.

(i) The unweighted least-squares procedure gives equal weight to all points along the curve, even though the later points clearly depend (from the cumulative nature of the curve) on the values that precede them. Even if weighted least squares were used, the dependence between successive observations would remain, and would invalidate the usual tests and estimates. Fitting the incremental form by weighted least squares should be better from this point of view.

(ii) In general, least-squares estimation procedures are known to be optimal only when the observations are normally distributed, which is not the case for the moment release data.

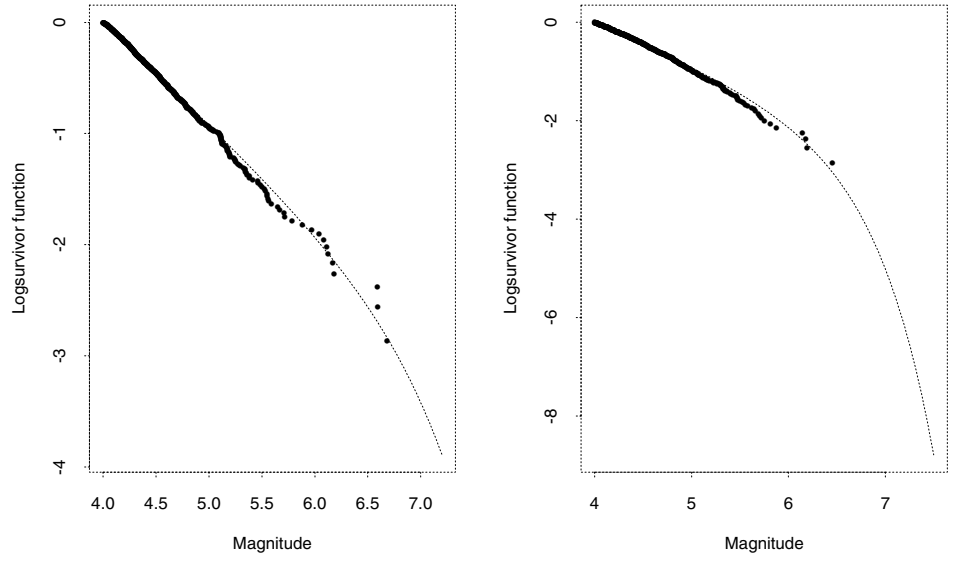

Figure 8. Two examples of Gutenberg–Richter plots from the AMR model with increasing mean strain release. Dotted line: fitted Kagan distribution; dots: empirical log survivor function. The second set had fewer large events and smaller gamma.

In particular, estimates based on least squares are extremely sensitive to occasional large departures from the underlying curve, such as occur when a large event occurs during the sequence. The use of least absolute errors, or some method based on fitting by medians, might produce more stable results.

(iii) The residual sum of squares in this context does not give a meaningful basis for determining estimation errors. In particular, a perfect fit of the cumulative curve to some functional form would not imply that the parameters in the curve were estimated with zero error.

(iv) Inclusion of the constant A in the formula is a logical error, for it represents an initial value and should only be included in the model if the sum of event sizes also starts from a non-zero initial value. Non-zero values of A that arise in the estimates, even when $\varepsilon(0)=0$, represent noise values only.

Despite these criticisms, the least-squares method will usually produce a curve that gives a reasonable visual fit to the data, and for this reason is unlikely to be widely discrepant from those obtained by other means. Another use for the residual sum of squares eq. (16) is as a test for the presence of an AMR signal. Typically, the residual sum of squares is computed as in eq. (16) and then compared with the corresponding sum of squares for a straight-line fit (uniform rate of increase). The ratio provides a test statistic for which significance levels can be computed by simulation, provided there is a well-defined null hypothesis (*cf.* Bowman *et al.* 1998). A general problem is the instability of the estimates, caused by random fluctuations in the numbers of events in the far right tails.

Although the ratio of sums of squares seems a reasonable test statistic to use in this context, the high variability introduced by occasional large events will still tend to diminish its effectiveness. The most obvious alternative would seem to be a likelihood ratio test based on a comparison of the likelihoods from the null hypothesis model and a fully specified AMR model, as outlined further below.

4.2 Least-squares techniques applied to incremental data

Here the data are first binned into intervals along the time axis and the total moment release Q_k is calculated for each bin. These totals are then fitted to the derivative of the AMR form by unweighted least squares. The main disadvantages with this procedure are the usual ones associated with binning procedures, that the choices of the numbers and widths of the bins are somewhat arbitrary and yet can significantly affect the estimates obtained. On the other hand, the contributions to each bin are likely to be close to independent (removing one of the prime objections to the previous case), and if not constant, at least of the same order of magnitude apart from random fluctuations. Because of the long tails of the moment distribution, the contributions to each bin are still likely to be highly variable, and unlikely to be even roughly normally distributed. This means that the estimates are likely to be heavily influenced by any bins that happen to contain one or more large events. To this extent they will be unstable and unreliable.

There are some fairly straightforward techniques that can be used to induce normality, such as first transforming the bin totals by taking square roots or even logarithms and then fitting a correspondingly transformed version of the model. Moreover, the normality assumptions are much more plausible in this context and should allow meaningful standard estimates to be obtained for the regression parameters and their standard errors.

4.3 Fitting a model with changing rate by maximum likelihood

If the assumption can be made that the sequence of time points (t_i) can be modelled as a Poisson process with varying intensity (rate) $\lambda(t|\boldsymbol{\phi})$, where $\boldsymbol{\phi}$ is the parameter vector, there are straightforward point-process techniques for estimating the parameters in $\boldsymbol{\phi}$. In the example considered in Section 2.1, the parameters to be estimated are those appearing in eq. (12), namely $\boldsymbol{\phi}=(\lambda_0, d, t_f, \kappa)$. The log likelihood for the point process takes the form

$$\log L_1 = \sum \log \lambda(t_i|\phi_1) - \int_0^{t_f} \lambda(t|\phi_1)dt\,, \tag{17}$$

which has then to be maximized with respect to $\boldsymbol{\phi}$. Routines for using maximum likelihood methods to fit a wide variety of point-process models defined by their intensities are available within SSLib (Harte 1999).

In this version of the model, the strain releases, s_i, are assumed to be independent of the t_i, and to have a fixed distribution, say $f(s|\boldsymbol{\theta})$. The vector $\boldsymbol{\theta}$ of parameters occurring in this distribution [$\boldsymbol{\theta}=(\alpha, \beta, \gamma)$ if the Kagan distribution is used] can therefore be independently estimated by maximizing the log likelihood

$$\log L_2 = \sum \log f(s_i|\boldsymbol{\theta})\,. \tag{18}$$

4.4 Fitting a model with changing mean energy release by maximum likelihood

Suppose that the moment releases are governed by a family of distributions with density $f(x|\boldsymbol{\theta})$. The evolution of the distribution in time will then be achieved by allowing the parameter vector $\boldsymbol{\theta}$ itself to evolve with time, as in the forms for the G–R and Kagan distributions considered earlier. One specific option is that used in model 2 in Section 2.2, namely the Kagan distribution with $\alpha=1$, $\beta=0$ and $\gamma=\gamma_i$, where γ_i is obtained from solving the equation $m(U)=\mu(t_i)$ for U_i and expressing U_i in terms of its equivalent magnitude γ_i. The parameters to be evaluated are no longer the parameters $\boldsymbol{\theta}$ arising directly in the definition of the magnitude distribution, but rather the parameters arising in the expressions for the components of $\boldsymbol{\theta}$ as functions of time. In the specific example, this would mean the parameters appearing in eq. (14), namely $\phi_2=(\mu_0, D, t_f, \kappa)$. In general, ϕ_2 could contain several sets of parameters, depending on how the various components of $\boldsymbol{\theta}$ were expressed as functions of time. The log likelihood for the event sizes, given their times, is then given by

$$\log L_2 = \sum \log f(x_i|t_i, \phi_2)\,. \tag{19}$$

This expression has then to be maximized with respect to the parameters in ϕ_2.

Since the event times here are assumed to have a constant rate λ, then if we also assume that they follow a Poisson process, the only additional parameter to be estimated is λ itself, which can be obtained by maximizing the corresponding special case of $\log L_1$.

4.5 Fitting a model in which both the rate and mean strain release change with time

If both the rate and the magnitude distributions change with time, the joint log likelihood can still be written in the form

$$\log L = \log L_1 + \log L_2 , \tag{20}$$

with log L_1 and log L_2 defined as in eqs (17)–(19). The new feature is that the parameter vectors $\boldsymbol{\phi}$, $\boldsymbol{\theta}$ appearing in the two parts of the likelihood may have some components in common. When this is the case, the two parts cannot be maximized separately, but must be maximized jointly with respect to the full set of parameters.

A suitable joint model for use in Section 3.3 might involve the forms

$$\lambda(t) = \lambda_0 + d\left(1 - \frac{t}{t_f}\right)^{-\kappa_1},$$

$$\mu(t) = \mu_0 + g\left(1 - \frac{t}{t_f}\right)^{-\kappa_2},$$

involving the seven parameters (λ_0, d, t_f, κ_1, μ_0, g, K_2). Some constraints might be introduced among these, such as $\kappa_1 + \kappa_2 = 1$, to try and avoid redundant parametrization. In either case, the joint likelihood has to be maximized for the parameters that remain.

5 CONCLUSIONS

This paper has addressed some questions of model formulation, simulation and assessment with respect to the AMR model. In the process a number of issues have emerged that were not treated in earlier discussions of the model, but may have some interest in both this and wider contexts. These issues may be summarized as follows.

(i) the non-identifiability of the constant A in eq. (1);

(ii) the difficulties of using the ordinary Gutenberg–Richter distribution in simulation versions of the AMR model;

(iii) the use of the Kagan distribution in the form (1) and the development of its elementary properties;

(iv) the hypothesis that changes in the parameter U (or γ) in the Kagan distribution can be related to the approach to criticality;

(v) the related possibility that variations in observed b-values, and other features of observed Gutenberg–Richter plots, may in fact be interpreted in terms of variations in U, including the occurrence of mixture distributions;

(vi) the difficulty in the AMR model of apportioning the increase in mean strain release between frequency and mean event size;

(vii) the problems, evident in simulations with realistically small numbers of events, of reliably detecting an AMR signal, particularly when it is expressed in terms of an increase in mean moment;

(viii) the need for caution in using least-squares methods to estimate coefficients and their standard errors in the AMR model.

Many features require further discussion and explanation, including the apportionment problem, the factors governing the size of the main event, and the embedding of AMR sequences within wider space–time patterns of earthquake occurrence.

ACKNOWLEDGMENTS

This research was partially supported by grants from the Marsden Fund (DVJ), the New Zealand Foundation for Science, Research and Technology (RR) and the Asia 2000 Foundation (WY). We are grateful to the referees for comments and corrections, which led to numerous improvements, and for additional references.

REFERENCES

Bowman, D.D., Ouillon, G., Sammis, C.G., Sornette, D. & Sornette, A., 1998. An observational test of the critical earthquake concept, *J. geophys. Res.,* **103,** 24 359–24 372.

Bufe, C.G. & Varnes, D.J., 1993. Predictive modelling of the seismic cycle of the greater San Francisco Bay region, *J. geophys. Res.,* **98,** 9871–9883.

Dahmen, K., Ertaş, D. & Ben-Zion, Y., 1998. Gutenberg-Richter and characteristic earthquake behaviour in simple mean field models of heterogeneous faults, *Phys. Rev. E,* **58,** 1494–1501.

Feller, W., 1966. *Introduction to Probability Theory and its Applications,* Vol. 2, Wiley, New York.

Gross, S. & Rundle, J., 1998. A systematic test of time-to-failure analysis, *Geophys. J. Int.,* **133,** 57–64.

Harte, D., 1999. Documentation for the Statistical Seismology Library, *Research Rept,* **98-10,** revised edn, School of Mathematical and Computing Sciences, Victoria University of Wellington, New Zealand.

Jaumé, S.C. & Sykes, L.R., 1999. Evolving towards a critical point: a review of accelerating moment/energy release prior to large and great earthquakes, *Pure appl. Geophys.,* **155,** 279–306.

Kagan, Y.Y., 1991. Seismic moment distribution, *Geophys. J. Int.,* **106,** 123–134.

Kagan, Y.Y., 1997. Seismic moment-frequency relationship for shallow earthquakes; Regional comparisons, *J. geophys. Res.,* **102,** 2835–2852.

Kanamori, H. & Anderson, D.L., 1975. Theoretical basis of some empirical relations in seismology, *Bull. seism. Soc. Am.,* **65,** 1073–1095.

Lewis, P.A.W. & Shedler, G.S., 1976. Simulation of non-homogeneous Poisson processes with loglinear rate function, *Biometrika,* **63,** 501–506.

Main, I.G., 1999. Applicability of time-to-failure analysis to accelerated strain before earthquakes and volcanic eruptions, *Geophys. J. Int.,* **139,** F1–F6.

Main, I.G. & Burton, P.W., 1984. Information theory and the earthquake frequency-magnitude distribution, *Bull. seism. Soc. Am.,* **74,** 1409–1426.

Main, I.G., O'Brien, G. & Henderson, J.R., 2000. Statistical physics of earthquakes: comparison of distribution exponents for source area and potential energy and the dynamic emergence of log periodic energy quanta, *J. geophys. Res.,* **105,** 6105–6126.

Ogata, Y., 1981. On Lewis' simulation method for point processes, *IEEE Trans. Inform. Theory,* **IT-30,** 23–31.

Robinson, R., 2000. A test of the precursory accelerating moment release model on some recent New Zealand earthquakes, *Geophys. J. Int.,* **140,** 568–576.

Saito, M., Kikushi, M. & Kudo, K., 1973. Analytical solution to 'Go-game' model of earthquakes, *Zishin,* **26,** 19–250.

Sornette, D. & Sammis, C.G., 1995. Complex critical exponents from renormalization group theory of earthquakes, *J. Phys. I. France,* **5,** 607–619.

Stauffer, D. & Aharoney, A., 1994. *Introduction to Percolation Theory,* 2nd edn, Taylor & Francis, Philadelphia, PA.

Varnes, D.J., 1989. Predicting earthquakes by analysing accelerating precursory seismic activity, *Pure appl. Geophys.,* **130,** 661–686.

Vere-Jones, D., 1976. A branching model for crack propagation, *Pageoph,* **114,** 711–726.

Vere-Jones, D., 1977. Statistical theories for crack propagation, *Math. Geol.,* **9,** 455–481.

Wesnousky, S., 1994. The Gutenberg-Richter or characteristic earthquake distribution, which is it?, *Bull. seism. Soc. Am.,* **84,** 1940–1959.

APPENDIX A: THE KAGAN DISTRIBUTION

A1 Background and general properties

There are many models leading to distributions that combine a power-law region with an exponential tail. To our knowledge, their first appearances in the earthquake context were in the papers by Saito *et al.* (1973) and Vere-Jones (1976), where they arose in connection with the 'Go-Game' and branching process models for the development of a fracture; see Vere-Jones (1977) for a more extended discussion. The exponential fall-off is related to the closeness of the model to criticality. In the branching model, this corresponds to the point at which the mean number of offspring per ancestor equals unity: the assumption is that as stress increases, so does the number of 'offspring' cracks per 'ancestor' crack, until criticality is reached and the pure power-law form is recovered. Ian Main has pointed out to us that a similar recovery of the pure power-law form occurs also in percolation theory as the correlation length approaches infinity (Stauffer & Aharoney 1994; Main *et al.* 2000), and once again the process approaches criticality.

In most of the derivations, the distribution that emerges has a density that can be written as the product of a power-law term and an exponential decay term that becomes important only in the far tail of the distribution. This version was later derived by Main & Burton (1984) as the maximum entropy distribution when mean magnitude and mean moment are both fixed. It has been used extensively in recent papers by Kagan (e.g. Kagan 1991, 1997) on the distribution of seismic moments; Kagan calls it a gamma distribution, but its behaviour is very different to that of a true gamma distribution, and corresponds rather to a Pareto or power-law distribution with a modified right-hand tail. Distributions of the same general type also arise in the first-passage problem for simple random walks, which has lead to further derivations based on first-passage arguments (see e.g. Dahmen *et al.* 1998). The special case used as an example in Vere-Jones (1977) has a Bessel function representation well-known in queuing theory (see e.g. Feller 1966, Section XIV.6).

The particular version of this general class of distributions that we propose, and have called the *Kagan distribution* in this paper, differs slightly from those described above in that the product form appears in the expression for the *survivor function* (complement of the cumulative distribution function), rather than the density function. The resulting distribution is possibly the simplest from the point of view of analytical manipulation, simulation and interpretation, and leads to a distribution with a slightly thicker tail for otherwise identical parameter values. It has the appealing interpretation that it represents a competition between two mechanisms for controlling the size of the event: one mechanism leads to the pure power-law form, perhaps representing parts of the region where the rock is already in a critical stress state; the other leads to the exponential tail, and perhaps represents gross geological or structural features that intervene and prematurely terminate the growth of the crack. One can imagine that as the general stress level increases, the latter are more easily overcome and the rock condition approaches that of an infinite region in a critical state.

The basic form we take for the Kagan distribution is

$$\Pr(X > x) = 1 - F(x;\, \alpha,\, L,\, U) = \left(1 + \frac{x}{L}\right)^{-\alpha} \mathrm{e}^{-\frac{x}{U}}, \quad x \geq 0\,. \tag{A1}$$

In the seismological context, the random variable X described by this distribution could refer to seismic moment, energy, Benioff strain release, fault length or fault area. As the default option we shall assume that it is related to the magnitude by the equation

$$X = 10^{0.75M + 2.4}\,, \tag{A2}$$

corresponding roughly to the interpretation of X as Benioff strain release. This corresponds to the practice in the AMR context (see e.g. Bowman *et al.* 1998), whereas Kagan preferred to work directly with the scalar seismic moment, as provided by the Harvard catalogue.

The parameters L and U represent lower and upper turning points, while α is the basic index in the power-law range of the distribution. Kagan referred to the parameter corresponding to U as the 'corner moment' or 'maximum moment'. All three parameters are constrained to be non-negative.

The role of the parameters is best illustrated by the plot of the log survivor function against log x (see Fig. 1). For $L \ll U$ (the usual situation in seismological applications) this curve has three portions. For small x ($\log x \to -\infty$), it is asymptotically flat, with 0 as its asymptote. There is then a linear portion, the power-law range, with slope $-\alpha$, where the power-law term dominates. The curve then starts to decrease more rapidly. The parameters L and U mark approximately the turning points between these three sections. A typical example is shown in Fig. 1, where the default parameters from the PKAGAN function in SSLib (Harte 1999) are used, namely $\alpha = 1$, with L and U corresponding approximately to magnitudes 0 and 7 respectively.

In its basic form (eq. A1), the model is assumed to apply from $x = 0$ (magnitude $-\infty$) onwards. This is unrealistic, since in practice there will be a lower bound on the sizes of the events that can be detected. It is therefore necessary to examine the distribution only above a certain threshold value, say x_0, corresponding to a lower cut-off magnitude M_0. In other words, our observations correspond to looking at the conditional distribution of X, given $X > x_0$. A simple calculation shows that the conditional distribution is given by

$$\begin{aligned} 1 - F(x;\, \alpha,\, L,\, U | x_0) &= \frac{1 - F(x;\, \alpha,\, L,\, U)}{1 - F(x_0;\, \alpha,\, L,\, U)} \\ &= \left(1 + \frac{x - x_0}{L + x_0}\right)^{-\alpha} \mathrm{e}^{-\frac{x - x_0}{U}}, \quad x \geq x_0\,. \end{aligned} \tag{A3}$$

As with the simpler Pareto distribution, this truncated and renormalized version of the Kagan distribution turns out to have the same form as the original version; it is enough to replace x by $x - x_0$ and L by $L + x_0$. Indeed, eq. (A3) can be interpreted as giving the distribution of the random variable with the representation

$$X = (x_0 + L)V - L\,, \tag{A4}$$

where the dimensionless variable $V = (X + L)/(x0 + L)$ has the 'standard' Kagan distribution,

$$1 - F(v;\, \alpha,\, \rho) = v^{-\alpha}\, \mathrm{e}^{-\rho(v-1)}, \quad v \geq 1\,, \quad \rho = (x_0 + L)/U\,. \tag{A5}$$

Note that this corresponds to the special case of eq. (A3) with $x_0 = 0$, $x = v - 1$ and $L = 1$. We find the standard version useful in what follows as a means of simplifying computations.

The density functions corresponding to the distributions (A1) and (A3) can also be written explicitly. Thus, the general form (A3) has density function

$$f_{X|x_0}(x) = \frac{1}{x_0+L}\left(\frac{x+L}{x_0+L}\right)^{-(\alpha+1)} \mathrm{e}^{-\frac{x-x_0}{U}}\left\{\alpha + \frac{L+x}{U}\right\}. \qquad \text{(A6)}$$

The basic form (A1) corresponds to the choice $x_0 = 0$, while the standard version has density

$$f_V(x) = v^{-(\alpha+1)}\mathrm{e}^{-\rho(v-1)}[\alpha+\rho v], \qquad \text{(A7)}$$

so that

$$f_{X|x_0}(x) = (x_0+L)^{-1} f_V\left(\frac{x+L}{x_0+L}\right), \qquad \text{(A8)}$$

the initial term corresponding to the change of scale from x to $(x+L)/(x_0+L)$. Functions to compute the distribution and density functions of the Kagan distribution with general choices of the parameters α, L, U, and x_0 are available in SSLib (Harte 1999).

For reasons associated with catalogue completeness, the magnitude threshold will often have to be taken quite large, so that it may be assumed $x_0 \gg L$. Then L can be neglected relative to both x and x_0, and the log frequency/magnitude graph is linear from the threshold magnitude M_0 until magnitudes corresponding to the upper turning point U. Taking L close to M_0 produces a flattening of the frequency/magnitude plot at low magnitudes similar to that caused by catalogue incompleteness. However, modelling incompleteness in this way is not recommended, since it can bias the estimates of α and U. Examples of simulated data sets with fixed M_0 and U but varying L are shown in Fig. 3.

A2 Expression in terms of magnitudes

If the survivor function is rephrased in terms of magnitudes, its logarithm takes the form

$$\log_{10}[1-F_M](\Delta M; \beta, \gamma|x_0)] = -b\Delta M - d(10^{0.75\Delta M}-1), \quad \Delta M \geq 0, \qquad \text{(A9)}$$

where $b = 0.75\alpha$, $d = (x_0/U)\log_{10}\mathrm{e}$ and $\Delta M = 0.75[\log_{10}(x+L) - \log_{10}(x_0+L)]$. This represents a special case of the Gompertz–Makeham distribution, originally used in graduating actuarial tables for human life distribution. This distribution is often characterized in terms of the hazard function (ratio of the density to the survivor function, or derivative of the log survivor function), which here has the general form

$$h(y) = c + d\mathrm{e}^{fy}.$$

In the life distribution context, the parameter c is interpreted as an age-independent background risk (corresponding to an exponential distribution) and the additive exponential term represents an additional, rapidly increasing, but independent source of risk that only cuts in after a certain age but then becomes rapidly dominant. Such a 'competing risk' interpretation could be of interest in the seismological context also, with the exponential decreasing term in eq. (A1) representing an alternative mechanism causing termination of the fracture, possibly related to the size of the seismogenic region or other large-scale boundary features, separate from the features normally controlling the fracture size.

We make one other point concerning the representation of the Kagan distribution in terms of magnitudes. As we have mentioned, in many situations, the physical variable x is inferred from an approximate regression relationship of the type

$$x = 10^{\boldsymbol{\theta} M + \boldsymbol{\phi}}.$$

The natural interpretation of $\boldsymbol{\theta}$ is that it determines the relation between α and the b-value; indeed, it is not difficult to see that $b = \alpha\boldsymbol{\theta}$, so that in the AMR context $\alpha = 1$ corresponds to $b = 0.75$, and $b = 1$ to $\alpha = 1.33$.

If, however, only the magnitudes can be measured directly, then a more subtle interpretation should in principle be considered. Suppose that it is not our original choice X, but rather $X^{1/\delta}$ that actually follows the law (A1), so that X follows the generalized Kagan distribution

$$1-F(x;\alpha,L,U) = \left(1+\frac{x^\delta}{L}\right)^{-\alpha}\mathrm{e}^{-\frac{x^\delta}{U}} \approx x^{-\alpha\delta}\,\mathrm{e}^{-x^\delta/U}, \quad x \geq 0. \qquad \text{(A10)}$$

In terms of magnitudes, this takes the approximate form

$$\log_{10}[1-F_M] \approx -\alpha\delta\boldsymbol{\theta}\Delta M - d(10^{\delta\boldsymbol{\theta}\Delta M}-1), \quad \Delta M \geq 0. \qquad \text{(A11)}$$

In this situation, assuming that only observations on magnitudes are available, only the quantities α and $\delta\boldsymbol{\theta}$ could be separately estimated. In other words, if it were known which function of magnitude followed exactly a Kagan distribution, then the coefficient $\boldsymbol{\theta}$ could be estimated since it would then be assumed that $\delta = 1$; if, on the other hand, the function relating magnitude to energy were known (or if the energies or seismic moments were known directly), then it should be possible to estimate which power of the energy most closely followed a distribution of Kagan type.

A3 Quantiles and moments

The quantiles corresponding to the forms (A1) and (A3) are readily computed by fixing a value for the probability on the right and solving for x. In practice it is convenient to find the quantiles of the normalized version (eq. A5) and then convert back to moments or magnitudes. A simple Newton–Raphson procedure is implemented in the QKAGAN function in SSLib.

Evaluating the moments requires a little more attention. To correspond with averages of observed values, the moments required are those of the conditional distribution eq. (A3). From the representation (A4) we have

$$E(X^k) = E\{[(x_0+L)V - L]^k\} = \sum_{r=0}^{k} C_r^n (x_0+L)^r (-L)^{k-r} E(V^r), \qquad \text{(A12)}$$

where C_r^k is the binomial coefficient. For the standardized variable V we find

$$E(V^k) = \int_1^\infty v^{k-\alpha-1}\,\mathrm{e}^{-\rho(v-1)}[\alpha+\rho v]dv.$$

Integrating by parts leads to the reduction

$$E(V^k) = 1 + k\,\rho^{\alpha-k}\,\mathrm{e}^{\rho}\,I(k-\alpha,\rho), \qquad \text{(A13)}$$

where

$$I(\beta, \rho) = \int_{\rho}^{\infty} y^{\beta-1} \mathrm{e}^{-y} dy$$

is a form of incomplete Γ-function. Essentially three cases arise.

(i) $\beta > 0$ ($k > \alpha$). In this case the integral remains bounded as $\rho \to 0$; indeed, it approaches $\Gamma(\beta)$. We can write

$$I(\beta, \rho) = \Gamma(\beta)\left[1 - \frac{1}{\Gamma(\beta)} \int_0^{\rho} y^{(k-1)} \mathrm{e}^{-y} dy\right], \tag{A14}$$

or in terms of SPlus functions, `I <- gamma(beta)*(1-pgamma(beta,rho))`. An alternative procedure is to expand the final integral as a power series in ρ, namely

$$I(\beta, \rho) = \Gamma(\beta) - \frac{\rho^{\beta}}{\beta} + \frac{\rho^{\beta+1}}{(\beta+1)1!} - \frac{\rho^{\beta+2}}{(\beta+2)2!} - \ldots .$$

(ii) $\beta = 0$ ($k = \alpha$). The integral here reduces to the exponential integral

$$I(0, \rho) = \int_{\rho}^{\infty} y^{-1} \mathrm{e}^{-y} dy = E_1(\rho) .$$

There is no SPlus function for the exponential integral, but since ρ is small, we can again expand the integral as a series:

$$I(0, \rho) = |\log(\rho)| - \gamma + \frac{\rho}{1.1!} - \frac{\rho^2}{2.2!} + \frac{\rho^3}{3.3!} - \ldots , \tag{A15}$$

where γ is Euler's constant, $\gamma \approx 0.5772$. In most situations arising in practice, ρ is small enough for the first three terms to give adequate accuracy.

(iii) $\beta < 0 (k < \alpha)$. In this case the integral diverges as $\rho \to 0$, but a further integration by parts gives a reduction formula for $I(\beta, \rho)$, which can be used to reduce the integral to case (i):

$$I(\beta, \rho) = \frac{1}{-\beta}\left\{\rho^{-\beta} \mathrm{e}^{-\rho} + I(\beta+1, \rho)\right\}, \tag{A16}$$

leading to

$$E[(X+L)^k] = (x_0 + L)^k \left\{1 + \frac{k}{\alpha - k}\left[1 - \rho^{\alpha-k} \mathrm{e}^{\rho} I(k+1-\alpha, \rho)\right]\right\}. \tag{A17}$$

In most applications, $1 < \alpha < 2$, so that at most one iteration of this formula should suffice.

In the applications to the AMR model, the mean is the quantity of greatest interest, and for this we have

$$\begin{aligned} E(X) &= (x_0 + L)E(V) - L \\ &= x_0 + (x_0 + L)\rho^{\alpha-1} \mathrm{e}^{\rho} I(1-\alpha, \rho) \\ &= x_0 + U\rho^{\alpha} \mathrm{e}^{\rho} I(1-\alpha, \rho) . \end{aligned} \tag{A18}$$

An inverse formula is also needed, giving the value of U (or ρ in the reduced form), when values for α, L, x_0 and the mean $E(X|X > x_0)$ are given. For this purpose we used an iteration formula based on the series expansion for $I(\beta, \rho)$.

A4 Parameter estimation

Estimation is considerably simplified by making use of the standardized form eq. (A7) for the density. We initially assume that L, as well as x_0, is fixed and known, and obtain likelihood estimates of α and ρ, from which we can obtain U as a function of L, x_0 and ρ. If it is needed, an estimate of L can be obtained by a further stage of estimation; for example, the procedures implemented in SSLib allow the user to scan the likelihood over a grid of values of L. However, estimating all three parameters from the data often leads to significant increases in the standard errors of the estimates. In particular, estimates of U become quite unstable if there is the option of explaining a slight curvature in the log survivor (G–R) plot as due either to L or to U or to a combination of both. Our experience is that better estimates of U are obtained by increasing the magnitude threshold to a point where there is no lower-end curvature from catalogue incompleteness, despite the loss of data involved in doing so. In fact, the data lost contains little information concerning U, but can have a significant nuisance value if it introduces additional curvature into the log survivor plot.

We assume, therefore, that the data have been transformed into a vector of standardized values $\{v_i, 1 = 1, 2, \ldots, n\}$. Using the density eq. (A7) (but omitting the initial factor), the log likelihood for an i.i.d. sample is given by

$$\log L = \sum \log(\alpha + \rho v_i) - (\alpha + 1) \sum \log v_i - \rho \sum (v_i - 1) . \tag{A19}$$

For brevity write $A = (1/n) \sum \log v_i$, $B = (1/n) \sum (v_i - 1)$. Then the likelihood equations, obtained by equating the first derivatives to zero, can be written in the form

$$\frac{1}{n}\frac{\partial \log L}{\partial \alpha} = -A + \frac{1}{n} \sum \frac{1}{\alpha + \rho v_i} = 0 ,$$

$$\frac{1}{n}\frac{\partial \log L}{\partial \rho} = -B + \frac{1}{n} \sum \frac{v_i}{\alpha + \rho v_i} = 0 .$$

These yield

$$A = \frac{1}{n} \sum \frac{1}{\alpha + \rho v_i} ,$$

$$B = \frac{1}{n} \sum \frac{v_i}{\alpha + \rho v_i} ,$$

whence

$$\alpha A + \rho B = 1 .$$

Substituting for α in terms of ρ gives us the single equation for ρ

$$1 = \frac{1}{n} \sum \frac{1}{1 - \rho(B - Av_i)} , \tag{A20}$$

which may be solved by Newton–Raphson or similar procedures. Care is needed in selecting the right root to this equation, since roots occur between every pair of successive zeros in the denominator. In fact, the relevant root is that between 0 and the smallest value of $1/(B - Av_i)$. The procedure converges successfully provided the initial value and all successive Newton–Raphson approximations are constrained to remain within this interval.

A5 Simulation

The competing risks interpretation of the Kagan distribution, mentioned in Section A2 above, leads to a particularly simple algorithm for simulating observations from the distribution. It implies that the standard variable V can be written

$$V = \min(Y, 1+Z), \tag{A21}$$

where the distribution of Y has the modified Pareto form

$$1-F_Y(x) = x^{-\alpha}, \quad x \geq 1,$$

and Z has an exponential distribution with parameter ρ. To generate a sequence of values X_i from the Kagan distribution with parameters (α, L, U, x_0), we may therefore proceed as follows:

(i) generate a sequence of values Y_i from the modified Pareto distribution with parameter α, for example by setting

$$Y_i = W_i^{-1/\alpha},$$

where the W_i are independent, uniform (0, 1);

(ii) generate a sequence Z_i from an exponential distribution with parameter $\lambda = (L+x_0)/U$;

(iii) set $X_i = (L+x_0)\min(Y_i, 1+Z_i) - L$.

The simulation algorithms will appear in the next edition of SSLib; the output is illustrated in Figs 3 and 4.

A6 Mixtures of Kagan distributions

As mentioned in the main text, there are some reasons for assuming that the upper turning point, or corner moment, U, may vary in both time and space according to the stress environment. If this is the case, then data from an extended region or time period would be more likely to follow a mixture of Kagan distributions than the simple distribution itself. We briefly discuss some of the forms such mixtures can take, limiting ourselves for simplicity to a discussion of mixtures of the standard form eq. (A5) with respect to the parameter ρ. The survivor function for such a mixture, assuming the parameter α remains fixed, has the form

$$1-G(v) = v^{-\alpha}\int e^{-\rho(v-1)}dH(\rho) = v^{-\alpha}h^*(v-1), \tag{A22}$$

where $h^*(s)$ is the Laplace–Stieltjes transform $\int e^{-sx}dH(x)$ of the mixing distribution for ρ.

The most important special case arises when H has a gamma $\Gamma(\beta, \lambda)$ distribution, in which case $h^*(s) = (1+s/\lambda)^{-\beta}$, yielding

$$\begin{aligned} 1-G(v) &= v^{-\alpha}\left(1+\frac{v-1}{\lambda}\right)^{-\beta} \\ &= v^{-(\alpha+\beta)}\left(\frac{1}{\lambda}+\frac{\lambda-1)}{\lambda v}\right)^{-\beta}, \quad v \geq 1. \end{aligned} \tag{A23}$$

Asymptotically, this is just another power-law distribution, with a straight-line form for the corresponding Gutenberg–Richter plot. A similar result will be obtained whenever the mixing distribution satisfies $dH(\rho) \approx \rho^{\beta}d\rho$ for ρ small (U large).

Mixtures of the Kagan distribution thus can produce power-law tails with indices higher than the original α, thereby providing a potential explanation of variations of b-value with stress.

Other remarkable forms can be produced either by truncating the gamma distribution (so that the values of ρ used in the mixture are either equal to a small value ρ_0 with probability $F(\rho_0)$, or distributed according to the original gamma distribution in the region above ρ_0), or by taking a mixture of some values with ρ fixed at ρ_0 and others with ρ distributed according to a gamma distribution with a mean above ρ_0. Such examples produce G–R type plots with a variety of different kinks, resembling for example the kinks ascribed to characteristic earthquake patterns. This is not altogether surprising, since the characteristic earthquake model is itself a mixture (of characteristic and background events); the point here is that such plots can be produced by models requiring much less drastic contrasts in the generating mechanisms, perhaps merely a response to changing stress conditions. Examples are shown in Fig. 4.