DEPARTMENT OF THE INTERIOR
U.S. GEOLOGICAL SURVEY

A Clarification, Correction, and Updating of *Parkfield, California, Earthquake Prediction Scenarios and Response Plans* (USGS Open-File Report 87-192)

Allan Goddard Lindh
and
Marie R. Lim

U.S. Geological Survey
345 Middlefield Road -- MS 977
Menlo Park, CA 94025

17 October 1995

Open-File Report 95-695

ABSTRACT

A careful reexamination of Bakun et al. (1987), the *Parkfield, California, Earthquake Prediction Scenarios and Response Plans* (USGS Open-File Report 87-192), has uncovered two small algebraic errors in the foreshock probability calculations, and one clear mis-statement in the accompanying explanations as to how those probabilities were arrived at. The algebraic errors, when corrected, do not change the three day probability associated with a M4.5 event from 0.37, the threshold for triggering a level A alert, and a public warning by OES. Thus they would have had no practical scientific or societal implications, had they been detected earlier.

However, it is also clear that the lack of detailed documentation of the precise derivation of the foreshock probabilities in the *Parkfield Scenario* document has contributed to the confusion concerning how they were arrived at (Jones and Michael, 1994; Michael and Jones, 1995), and whether or not they are (even approximately) correct. It is also clear that the lack of clarity in the *Parkfield Scenario* document reflected, to some degree, a lack of fundamental understanding of the problem on the part of at least one of the co-authors (AGL).

Fortunately, since 1987, Duncan Agnew has provided a clear derivation of how to compute conditional probabilities for potential foreshocks (Agnew and Jones, 1991), and with the benefit of that analysis we "re-derive" the *Parkfield Scenario* probabilities, show what they do, and do not, include, illustrate the errors, and show that the answers obtained were basically correct. Rather than reflecting any Divine intervention, this good fortune seems rather to reflect the fact that at Parkfield there is a history of recurrent foreshocks that provides empirical underpinning for the calculations; if probabilities had been calculated that were grossly inconsistent with the observations, it would have been noted, and the process continued until "reasonable" results were obtained.

Unfortunately, in "re-deriving" the Parkfield probabilities, it also became clear that the assumed "flat" foreshock-magnitude distribution that underlies the Parkfield calculations, and Agnew and Jones' later work, has difficulties; we show that the foreshock-mainshock data (such as they are) are better fit by a log-normal distribution. Utilizing this new result, and taking into account the lower long-term probabilities appropriate at Parkfield since the failure of the "Parkfield Recurrence Model" of Bakun and Lindh (1985), we compute a new set of Parkfield foreshock probabilities. The new values are somewhat lower than those obtained eight years ago, with a M5 event near the 1966 hypocenter having a probability of 0.18 of being followed within 3 days by a characteristic Parkfield earthquake.

> The remaining phenomenon of the same kind there will be no difficulty in reasoning out by the method of probabilities. A man may sometimes set aside meditations about eternal things, and for recreation turn to consider the truths of generation which are probable only; he will thus gain a pleasure not to be repented of, and secure for himself while he lives a wise and moderate pastime. Let us grant ourselves this indulgence, and go through the probabilities relating the same subjects which follow next in order.
>
> Plato, *Timaeus*

Introduction

Lindh and Jones (1985) and Bakun et al. (1987, hereafter referred to as BEA) outlined an approach to using past data on foreshock occurrence at Parkfield, to estimate the probability that future events at Parkfield might be foreshocks to future characteristic Parkfield earthquakes (Bakun and Lindh, 1985). The basic approach was to estimate the rate at which foreshocks to past Parkfield main-events had occurred, estimate the background rate of potential foreshocks, and combine these into an estimate of the conditional probability.

> "In applying it to the Parkfield region, the limited data available require that the rate of occurrence of foreshocks and the rate of occurrence of all earthquakes be determined separately. These rates are determined as a function of magnitude of the earthquakes, and then the ratio of these rates is an estimate of the probability that a given earthquake will be a foreshock to the next characteristic Parkfield earthquake." (Jones and Lindh, 1986, p. 4-5)

Given that there were four characteristic Parkfield events in this century, (at least two of which had M5 foreshocks), given that approximately half of the M5's at Parkfield since 1934 have been foreshocks, and given that the next Parkfield earthquake was considered imminent, BEA concluded that there was a 37% chance that any M4.5 (or larger) earthquake within the Middle Mountain box would be a foreshock to a characteristic Parkfield event.

The result of these considerations, tempered by a liberal dose of caution that the public not be misled as to the primitive level of our understanding, was a warning scheme in which any event greater than M4.5, located in a box approximately 15 km on a side, centered on the 1966 Parkfield epicenter, would trigger a "Level A alert", which would result in a public warning by the State of California Office of Emergency Services that a M6 earthquake might be imminent. The probability associated with a level A alert was 37% or greater, and it was emphasized that this was an "... approximate (*i.e. order of magnitude*) estimate(s)..." (BEA, p. 21) and that "While these probabilities are quoted to 2 significant figures, they are approximate and somewhat subjective, and are best treated as order of magnitude estimates." (BEA, p. 25). These cautions were not rhetorical; it was clearly understood that the formalism used to compute the probabilities was *ad hoc*, and at best an approximation. It was also appreciated that even with four apparent recurrences of M6 events in this century, that the data was inadequate to constrain the future with perfect confidence.

Since this work was completed, eight years have passed and there have been a number of developments in the field, such that a review of the 1987 BEA calculations is in order. The most important of these developments are:

1. Agnew and Jones (1991, referred to here-after as A&J) have provided an improved formulation of the foreshock probability problem, worked out in terms of conditional probabilities.
2. The Parkfield earthquake has not occurred as anticipated by Bakun and Lindh (1985) necessitating a re-evaluation of the long-term probabilities at Parkfield,
3. The background seismicity rate at Parkfield has increased, particularly in the vicinity of Middle Mountain in the critical M4-5 range, necessitating a re-evaluation of the background seismicity rates used in the probability calculations,
4. Eight years experience has been gained in locating earthquakes, and estimating their magnitudes in real-time, permitting a re-evaluation of the assumptions made in the original calculations to account for real-time observational errors,
5. Meagher (1993) has undertaken a detailed relocation of the moderate seismicity in the Parkfield region since the 1930's, providing a better catalog for estimating background seismicity rates for events of M>4 since the 1930's, and
6. Cole and Ellsworth (1995) have undertaken a detailed study of the waveforms of many of the critical moderate Parkfield events since the 1930s, providing some clarification on precisely how the foreshocks, and non-foreshocks, beneath Middle Mountain were located relative to one another.

In addition, in the process of applying the A&J formulation to the Parkfield data set, we have discovered two errors in the BEA calculations, and have been able to produce a somewhat clearer derivation of how the results in BEA were obtained. The consequences of these errors however, are very small, and in addition they cancel in the critical M4-5 range.

More seriously, we have new evidence on the statistical properties of foreshocks that calls into question the "flat" foreshock distribution that implicitly underlies BEA, (and is explicit in A&J). This distribution is not well fit by existing data sets, and we show that a log-normal function better characterizes the distribution of the difference between the magnitude of the main event and its largest foreshock.

In the work that follows we briefly review the A&J formulation of the foreshock probability problem, and put it into a form that facilitates its application to Parkfield. In the second section we use this formulation to clarify and correct the BEA calculations and presentation. In the third section we present a new foreshock data set that clearly illustrates that the "flat" foreshock distribution used by BEA and A&J is neither a reasonable choice, nor does it fit the existing data; we reformulate the problem with a log-normal distribution. In the fourth section we address the question of whether foreshock rates at Parkfield are sufficiently different from the "generic" data set for strike slip earthquakes throughout the San Andreas province to justify treating them as a special case. And in the final section we present a revised set of foreshock probabilities for Parkfield, based on the log-normal distribution, and the better data sets available for estimating the background rates.

The Agnew and Jones formulation.

A&J provided an improved formulation of the foreshock probability problem, embodying essentially the same assumptions as BEA:

> This probability turns out to depend on the long-term probability of the mainshock, the rate of background seismicity along the fault, and some assumed characteristics of the relations between mainshocks and foreshocks." A&J p. 11,959.

They formulated the conditional probability of a main event, given a potential foreshock, as (their equation (4)):

$$P(C|F \cup B) = \frac{P(F)}{P(F)+P(B)} = \frac{P(C)P(F|C)}{P(C)P(F|C)+P(B)} \qquad (1)$$

where *P(F)*, *P(B)*, and *P(C)* are the unconditioned *a priori* probabilities of a foreshock, background event (i.e. a non-foreshock), and a main event, respectively. *P(F|C)* is the conditional probability of foreshock, given that a main event has occurred; by using this expression they have transformed *P(F)*, which is in general unknown, into the product of two terms, (*P(C)* and *P(F|C)*) which can, at least in principle, be estimated from existing data.

A&J then generalized this expression to a series of integrals in eight dimensions, (two space, one time, one magnitude each for mainshocks and foreshocks), worked through the general case, and then made a long series of simplifying assumptions to obtain:

$$P(C|F \cup B) = \frac{(N_m P(C)/A_C \delta_1)}{(N_m P(C)/A_C \delta_1) + \Lambda_S(x_0)e^{-\beta M}} \qquad (2)$$

where *P(C)* is the probability of the main event occurring within a 3 day window, A_C is the length (in km) of the fault segment for which *P(C)* was computed (which approximates the expected rupture zone of the main event in most cases), δ_1 is the time interval after the occurrence of the foreshock for which the probability is being estimated (3 days in this case), β is 2.3 times the *b* value of the background seismicity, *M* is the magnitude of the candidate foreshock and $\Lambda_S(x_0)$ is the rate density of background earthquake occurrence at x_0 (the location of the candidate earthquake).

Λ_S is obtained by the following formula:

$$\Lambda_S = \frac{2.3 \times 10^a b}{A \times 60^2 \times 24 \times 365.25 \times T} \qquad (3)$$

where *T* is the time period (in years), and *A* the length of the fault segment for which the *a* and *b* values were computed. (Note that *A* and A_C are not always the same; A_C is the segment length for which the long term probability of a main event was computed, *A* is the segment length for which the background seismicity rate was estimated. Since the fault has been divided into smaller segments for estimating the background rates than for estimating the long-term recurrence intervals, A can be, and often is, shorter than A_C).

The crucial term in (2) is N_m, a "normalizing factor," which includes the assumptions, and information, about the actual frequency and distribution of foreshocks; equation (25) from A&J is reproduced below.

$$N_m = \frac{\alpha\beta'}{1+\beta'(M_B - M_D)} \tag{4}$$

where α is the fraction of earthquakes (of magnitude greater than M_B) with foreshocks (of magnitude greater than M_D). M_B and M_D are the lower limits of the magnitudes of the main events and foreshocks characterizing the data set used to determine α, and β' is 2.3 times the b value for the main events; they set $\beta' = 2.3$, which assumes b=1.0. The crucial assumption is that Φ_m (the density function describing the magnitude distribution of foreshocks, see (6) below), is constant; that is, for those events that have foreshocks, all foreshock magnitudes are equally likely. They estimated α to be 0.5 when M_B is 5.0, and M_D is 2.0, (Jones, 1984), and from this compute a value for N_m of 0.15.

This formulation becomes somewhat clearer if we substitute (3) into (2), cancel common terms, and rearrange, obtaining:

$$P(C|F \cup B) = \frac{(A/A_C)P(C)N_m}{(A/A_C)P(C)N_m + (\delta_1/T')\beta \cdot 10^{a-bM}} \tag{5}$$

where T' is equivalent to T in (3), except that it is expressed in seconds rather than years. It is now clear that the numerator corresponds approximately to the numerator in (1), $P(C)P(F|C)$, which is simply $P(F)$, the probability of a foreshock within a 3 day window. The second term in the denominator corresponds approximately to $P(B)$, the probability of a background event of magnitude M within a 3 day window. (These correspondence are obscured, however, by the fact that a Taylor expansion has been used on the background rate term, to cancel out the explicit magnitude dependence in the numerator; see below).

An alternative approach is to work directly from (1), make the same simplifying assumptions, and formulate the problem directly in terms of $P(F|C)$, $P(B)$, and $P(C)$. When we do this we get the following expressions.

$$P(F|C) = \int_{m_1}^{m_2} \Phi_m dM = N_m \int_{m_1}^{m_2} dM = N_m(m_2 - m_1) \tag{6}$$

which is simply the probability of a foreshock of magnitude between m_1 and m_2, given a mainshock of any magnitude greater than m_2. This follows directly from A&J's assuming a flat foreshock distribution, and setting $\Phi_m = N_m$ (see their (15), (16), (24) and (26)).

Similarly, we can write an explicit expression for the rate at which background events are expected:

$$P(B) = 2.3b\left(\frac{\delta_1}{T'}\right)\int_{m_1}^{m_2} 10^{a-bM} dM = \left(\frac{\delta_1}{T'}\right)(10^{a-bm_1} - 10^{a-bm_2}), \tag{7}$$

which follows directly from the Gutenberg-Richter frequency-magnitude relation.

Substituting (6) and (7) into (1), we obtain:

$$P(C|F \cup B) = \frac{(A/A_C)P(C)N_m(m_2 - m_1)}{(A/A_C)P(C)N_m(m_2 - m_1) + (\delta_1/T')(10^{a-bm_1} - 10^{a-bm_2})}. \tag{8}$$

A&J simplified this expression by substituting $M \pm \mu$ for m_2 and m_1, respectively. Then by taking μ small, and using the first term of a Taylor expansion, (7) can be simplified to:

$$P(B) \cong \left(\frac{\delta_1}{T'}\right)\beta \cdot 10^{a-bM}(m_2 - m_1). \tag{7'}$$

If we substitute (6) and (7') into (1), we obtain

$$P(C|F \cup B) = \frac{(A/A_C)P(C)N_m(m_2 - m_1)}{(A/A_C)P(C)N_m(m_2 - m_1) + (\delta_1/T')\beta \cdot 10^{a-bM}(m_2 - m_1)} \tag{8'}$$

If we cancel common terms, we obtain (5), closing the loop.

The numerator of (8) is *P(F)*, the expectation of the number of foreshocks between m_2 and m_1 that will occur during the time δ_1. The product of the first two terms, $(A/A_C)P(C)$ simply represents the probability of a main shock within any three day window; the extra A/A_C term is required because the *P(C)* values in A&J are for segments of length A_C, while the *A* values are the (sometimes shorter) lengths of the segments for which the background earthquake rates were computed. Thus $(A/A_C)P(C)$ is the probability of a main event initiating within a subsection of the segment corresponding to *P(C)*, and essentially involves an assumption that the epicenter is equally likely anywhere within the segment. The third and fourth terms in the numerator are just *P(F|C)* (from (6)), the fraction of mainshocks preceded by foreshocks between m_1 and m_2 within a three day window.

The second half of the denominator in (8) is *P(B)*, which can now be clearly seen as simply the expectation of the number of events between m_1 and m_2, from the Gutenberg-Richter relation for a catalog characterized by *a* and *b*, for the time period δ_1. The number of events expected within time δ_1 is obtained by multiplying by δ_1/T', the ratio of the time considered to the time period covered by the catalog.

Thus in essence, the foreshock probability problem has been reduced to one of drawing a single ball from an urn containing a mixture of black and white balls. *P(F)* is an estimate of the number of white balls, *P(B)* an estimate of the number of black balls, and *P(F)/(P(F)+P(B))* is an estimate of the probability that any single ball drawn will be white.

The *Parkfield Scenario* document formulation

"It is very tempting to impose clarity and order upon events that had neither one nor the other. It is very tempting and very dangerous. That is how you become a philosopher before your time. I shall therefore try to relate what happened, and what was thought as it happened. If this at first seems to you to be chaotic and hazy, take heart; subsequently things will be only too orderly, too clear."

from *A Night of Serious Drinking*,
by René Daumal (Boulder: Shambala, 1979)

BEA used an expression for the probability of a characteristic Parkfield event, following a foreshock of magnitude M,

$$P_F = 3.1 \times 10^{-4+0.62M} \tag{9}$$

which followed directly from the work of Lindh and Jones (1985) and Jones and Lindh (1986) in which the probability was estimated from the ratio of the expectation of the number of foreshocks per year, to the annual expectation of the number of background events:

$$P_F = \frac{N_f/yr}{N/yr} = \frac{1/43.4}{10^{1.86-.62M}} = 3.1 \times 10^{-4+.62M} \tag{10}$$

The numerator reflects the observation of one M5 foreshock per two recurrence cycles at Parkfield, and the denominator is the expectation of the number of events of M or greater per year; this equation anticipated (5) above, the main result from A&J.

We should note, however, that there is one fundamental difference between the approach of BEA and A&J. Given the observational basis of repeated characteristic events at Parkfield, at least some of which had large foreshocks, BEA estimated *P(F)*, the expectation of the number of foreshocks with a given time window, directly from the observed (minimum) rate for Parkfield foreshocks. A&J had no such luxury -- for most of the fault segments they had no observed characteristic earthquake, let alone foreshocks -- and therefore had to estimate the expected number of foreshocks "indirectly", by estimating *P(C)* and *P(F|C)*, and then using their product as an estimate of *P(F)* (see (1) above). Otherwise the two approaches are, in principle, identical.

Unfortunately if we compare (1) and (5) with (10), there clearly are errors in (10); the first is that the denominator is based on a declustered catalog from which foreshocks and aftershocks have been removed, and thus should contain a second term as an estimate of the number of foreshocks; the denominator should more closely follow the form of (5) or (8). (In terms of the urn analogy, (10) is the ratio of the number of white balls to the number of black balls; the denominator should be the total number of balls.) Thus (10) should have included an estimate of the expectation of the number of foreshocks per year in the denominator as well:

$$P_F = \frac{N_f/yr}{N_f/yr + N/yr} = \frac{1/43.4}{1/43.4 + 10^{1.86-.62M}} \tag{11}$$

This mistake is small, negligible below M4, and by itself would have changed the Level A probability from 0.37 to about 0.33 at M4.5 (compare curves 1&2 in Figure 1b). In addition, as we show below, it was canceled by a second small error of opposite sign.

The second error in BEA becomes clear if we write out (11) in a fuller form, as if BEA had benefited from the above analysis, and followed the form of (8) above. (Because BEA estimated P(F) directly from the data, and the rates are per year, the A/A_C, $P(C)$ and δ_1/T' terms do not appear at this stage, they are incorporated, and discussed, below.)

$$P_F = \frac{N_f/yr}{N_f/yr + N/yr} = \frac{\frac{1}{43.4}(m_2 - m_1)}{\frac{1}{43.4}(m_2 - m_1) + (10^{1.86-.62m_1} - 10^{1.86-.62m_2})} \tag{12}$$

Clearly what BEA did, in effect, was set $m_2 - m_1 = 1$, and $m_1 = M$ (compare (12) and (11)), so that a more correct version of (11) would have been:

$$P_F = \frac{N_f/yr}{N_f/yr + N/yr} = \frac{\frac{1}{43.4}}{\frac{1}{43.4} + (10^{1.86-.62M} - 10^{1.86-.62(M+1)})} = \frac{\frac{1}{43.4}}{\frac{1}{43.4} + .76 \cdot 10^{1.86-.62M}} \tag{13}$$

Now it is clear that BEA omitted the second exponential term in the denominator (compare (11) & (13)), which in effect omitted a factor of 0.76 in the final term of the denominator. However, this error was partially canceled by the first error -- omitting the first term of the denominator -- so that the total net effect was very small; compare curves 1 & 3 in Figures 1a&b. By extreme good fortune these two curves cross at about M4.5, so that the net effect of the two algebraic errors in BEA on the setting of A level alerts was nil. The curves diverge slightly at lower values, so that the threshold for a B level alert might have been set at 0.13 instead of 0.11 for instance, had BEA had the benefit of this analysis in 1987.

This formulation also makes it easy to see what effect it would have had, if BEA had formulated the problem the same way A&J did. The main difference is that BEA integrated over a one magnitude unit range **above** the magnitude of interest (M), while A&J integrated over a range of magnitudes **centered** on M, which is equivalent to setting $m_2 = M + \frac{1}{2}$ and $m_1 = M - \frac{1}{2}$ in (12):

$$P_F = \frac{N_f/yr}{N_f/yr + N/yr} = \frac{\frac{1}{43.4}}{\frac{1}{43.4} + (10^{1.86-.62(M-\frac{1}{2})} - 10^{1.86-.62(M+\frac{1}{2})})} = \frac{\frac{1}{43.4}}{\frac{1}{43.4} + 1.55 \cdot 10^{1.86-.62M}} \tag{14}$$

which is essentially the same form as (5), with $\beta \cong 1.5$. As is clear in Figure 1 (curve 4), the net effect on the probabilities is very small, if one compares values for $M \geq 4.5$ for curves 1,2 & 3, with values for M=5 for curve 4. (Curves 1,2 & 3 are for a one magnitude unit range above M4.5; curve 4 is for a one magnitude unit range centered on M5.)

It is now clear, with the benefit of hindsight, that not only did BEA and A&J make essentially the same assumptions, but they got essentially the same result; (11) through (14) have precisely the same *functional form* as (5). And as we will show below, BEA got essentially the same answers they would have obtained had they used the A&J formulation, but that first requires clarifying the differences that do exist between the two approaches.

As we described above, A&J estimated *P(F)* "indirectly" by taking the product of *P(C)* and *P(F|C)*; they converted the observed frequency of foreshocks before main-events, and the assumption of a flat foreshock distribution, into an estimate of the frequency of foreshocks of any magnitude within the time window for which *P(C)* is estimated -- three days in their case. Thus this formulation explicitly includes the estimate of the probability of the main event within the specified time window.

BEA on the other hand estimated *P(F)* directly from the observed frequency of M5 foreshocks at Parkfield, and then made an implicit assumption of a flat foreshock distribution to generalize to other magnitudes. (Admittedly the documentation of this implicit assumption in Jones and Lindh (1986) and BEA leaves something to be desired, but the formula, and the results are clear.) This resulted in an estimate of the foreshock probability, (what BEA called P_F) which includes foreshocks at any time before the main event (A&J assumed three days), and is independent of any consideration of where one might be in the recurrence cycle of the main event (by including *P(C)* in their foreshock probability calculation, A&J directly incorporated the "probability gain" associated with time-varying long-term probability.)

Because of these differences in the way in which BEA and A&J formulated the problem, two additional factors must be considered before final results can be compared. The expression shown above, (11), is an estimate of the probability that a given event is a foreshock (and therefore that a characteristic event is about to occur), given that either a foreshock or background event has occurred, but it does not explicitly include a specification of how soon the main event will occur. (This differs from the formulation of A&J, where the foreshock probabilities were explicitly for 3-day windows before the main event.) BEA assumed that the likelihood of a main event decayed exponentially, following the time of the potential foreshock, with a characteristic time of 48 hours. This results in an estimate that there is a 0.79 chance that the main event will occur within 72 hours,

$$P_{F,72} = -P_F \cdot \frac{1}{48}\int_0^{72} e^{-t/48}dt = 0.79P_F \tag{15}$$

In addition, rather than including $P(C)$ within the foreshock formulation (as did A&J), BEA treated the current estimate of the probability of the main event as a probability gain versus the Poisson probability, and utilized the formulation of Utsu (1979, reproduced in Aki, 1981) to allow for the fact that some of the probabilities are no longer infintesimals. (Note: *P(C)* in A&J is entirely equivalent to P_C in BEA, and they are sometimes used interchangeably here. Both represent an estimate of the probability within a three day window of a characteristic event.)

$$P = \frac{1}{1 + r_1 r_2 / r_0} \tag{16}$$

where

$$r_0 = \frac{1}{P_0} - 1, P_0 = 3.8 \cdot 10^{-4}, r_1 = \frac{1}{P_C} - 1, P_C = 1.23 \times 10^{-3}, \text{and } r_2 = \frac{1}{P_{F,72}} - 1$$

These formulas have been applied to formulas (10), (11), (13) & (14), and P is plotted versus M as curves 1-4 in Figure 1b. Just as for the corresponding curves in Figure 1a, curves 1&3 intersect near M4.5 ($P \sim 0.37 - 0.39$), implying that BEA got essentially the "right answer" for M4.5 (that is the answer that their data and assumptions implied, given our current level of understanding of the problem, and no algebraic error. In addition, the same answer is obtained for curve 4 for an event of M5 ($P \sim 0.38$), again showing that it makes no difference whether the computations are made for a magnitude range above the target magnitude, or centered on it, assuming that the algebra has been done correctly.

This expression for the total probability now makes it possible to directly compare the results from BEA with A&J's formulation of the problem. If we use the same values for *a*, *b*, and P_C used above, disregard A/A_C, set δ_1/T' to 3/365, and set $N_m = 0.5$ in (5), we obtain curve 5 in Figure 1b. (Remember, $N_m = 0.15$ was obtained by A&J from the observation that one-half of the strike slip events in California had foreshocks within three magnitude units (Jones, 1984); $N_m = 0.5$ corresponds to BEA's observation that one-half the Parkfield main events had foreshocks within one magnitude unit.) It is clear in Figure 1b that comparable results are obtained with either approach. Curve 5 is somewhat higher than curve 4, to which it can most directly be compared, because it does not include the factor of 0.79 from expression (15).

One additional problem with BEA is in the documentation of how the foreshock probabilities were obtained. In particular the statement on page 23 that:

1) The next characteristic Parkfield earthquake is assumed to have a 0.5 chance of having some foreshocks, magnitude unspecified, within the Middle Mtn. alert zone;

is vague, to put it kindly. In light of current understanding of the problem, it is clear that it should have said:

1) Any earthquake of magnitude 4.5 to 5.5, within the Middle Mtn. alert zone, is assumed to have a 0.5 chance of being a foreshock to the next characteristic Parkfield earthquake;

This problem with documentation, the two algebraic errors corrected above, and the fact that the Jones and Lindh (1986) manuscript was never completed, clearly have contributed to the confusion that exists as to how the BEA probabilities were obtained, and whether or not they were even approximately correct (Jones and Michael, 1994; Michael and Jones, 1995). It is clear that at least one of the authors in 1987 (AGL) did not have a firm grasp of the problem, and was in fact "winging it" on an intuitive hunch as to how the problem "should" be formulated. That BEA in fact got a reasonable "first approximation" to the correct algebraic formulation, and almost exactly the "right answer", can only be attributed to a remarkable bit of good luck.

The Problem with *P(F)*

However, one fundamental problem with all of these foreshock probability calculations is the assumption of a "flat" foreshock distribution; this assumption underlies the formulations of both BEA and A&J, as has been illustrated above. The first problem with this assumption is that it implies a density function for foreshock magnitude distribution that is not a proper density function; as pointed out by A&J, they predict that for δM's above about 6, more than 100% of all earthquakes are estimated to have foreshocks.

The second problem is that the foreshock probability relations derived by both BEA and A&J predict results that are clearly not reasonable as the magnitude of the potential foreshock increases. In Figure 1a, for instance, BEA's original expression (curve 1) asserts that a M5.5 (or larger) event at Parkfield would have an 86% chance of being followed by a characteristic event, and even the corrected version (curve 3) provides an estimate of 70%. BEA avoided this problem by truncating the probabilities at 0.37, corresponding to a M4.5 or larger event.

A similar problem exists with the A&J work, where for the Palm Springs segment, their formula produced an estimate that a M7 event in the Coachella Valley had a probability of 70% of being followed by an M7.5 characteristic event (curve 1 in Figure 2.) Nor is this simply an academic problem, as the A&J results were used as the basis for public warning issued after the Joshua Tree (Jones, 1994) and Lander's earthquakes. The assertion that Joshua Tree had a 21% chance of being followed by a M7.5 on the Coachella Valley directly flies in the face of the facts; there have been two M6, and one M5.6, events near the Palm Springs segment in the last 50 years, the recurrence interval on this segment is believed to be in excess of 250 years. When many such events are expected over a 250 year period, any one of which might be a foreshock, it cannot possibly be correct that any one of them has a probability of 21% of being a foreshock; something has gone wrong.

We believe it is clear that one fundamental problem with this work is that a flat foreshock distribution does not fit the data; that is, Φ_m is not constant with magnitude as explicitly assumed by A&J, and implicitly by BEA. A&J based their estimate of the frequency of foreshocks on a set of 15 strike slip events that occurred in California in the period 1966-1980 (Jones, 1984), and concluded that about half of the events had foreshocks within three magnitude units of the main event. The main shocks in this data set were all of M5 or greater, and all but one of the foreshock sequences included at least one foreshock of M3 and greater, so the data may be relatively complete, even given the relatively sparse seismic networks operating in California in the 1960's and 70's.

However, they did not base their assumption that foreshocks have a flat magnitude distribution on this small data set. They instead used the entire Caltech catalog for southern California for the period 1932 to 1987, apparently including all main events of M3 and greater, and all identified foreshocks of M2 and greater. From a bivariate plot of all foreshocks (their Figure 3) against main shock magnitude, they concluded that:

> "....(this data) suggests that for any narrow range of mainshock magnitude, foreshock magnitudes close to that of the mainshock are more common; however, for the larger mainshock magnitudes of interest here, the (admittedly sparse) data suggest that all foreshock magnitudes are equally likely for given mainshock magnitude." (p. 11,964).

However, if we simply plot their data (from their Figure 3) as a function of δM (the difference between the magnitude of the main event and the magnitude of the largest foreshock) it is clear that their assumption is unwarranted (Figure 3a), and that the data have a strong central tendency around a δM of 1-2 magnitude units.

This result is not entirely surprising. Papazachos (1975) had earlier obtained a similar result for the seismicity catalog in Greece; the mean of his data set is 1.9 magnitude units, with a standard deviation of 0.89 (Figure 3b). There are difficulties with both of these data sets however, because they are drawn from catalogs that span a period of time during which significant changes occurred in the recording and analysis of earthquakes, so it is unclear precisely what lower magnitude threshold applies. Thus it is unclear what confidence to attach to the results in Figure 3, particularly on the right hand side of the plots ($\delta M > 2$), where the apparent decrease could be due to catalog incompleteness. Moreover, both data sets almost certainly contain some swarm sequences (Mogi's (1963) Type 3), associated with the extensional environments of the Imperial Valley and southern Basin & Range (Figure 3a) and Gulf of Corinth (Figure 3b) which complicates the interpretation.

We should note that the problem with the "flat foreshock distribution" is not the only problem with the A&J results. The more serious problem is that they used microearthquake catalogs spanning only 10-15 years to estimate *a* and *b*, and thereby systematically underestimated the rate at which moderate events occur (see Wesnousky, 1994, for illustrations of this general problem), sometimes by at least an order-of-magnitude (Figure 2). For instance, the *a* and *b* values used by A& J predict an event of M6 or greater an average of once every 726 years on the Palm Springs segment, but two such events have occurred on, or near, this segment in the last 50 years. The implications of this problem for estimating Parkfield foreshock probabilities are discussed below.

Foreshock Data Set

In an attempt to improve the situation with the foreshock distribution model, we have assembled a new foreshock data set for the period 1970-1995, for (predominantly) strike-slip events of M5 and greater in the San Andreas tectonic province (Table 1). We followed Jones (1984) in calling anything a foreshock that was located within 10 km and 3 days before a main event. The time period and magnitude range was chosen to ensure that conclusions could be drawn for foreshocks as small as M2, so that thte data may be essentially complete for $\delta M \leq 3$. We excluded:

1. Thrust events, because they appear to have foreshocks less frequently than strike slip events (Jones, 1984),
2. All Basin and Range, Coso, and Mammoth events, and those south of the Imperial Valley, because swarm behavior is so prevalent in those areas, and
3. All events at or north of the Mendocino Triple junction because the completeness threshold is uncertain for offshore events, and because the wide variety of focal mechanisms in the area make it problematic identifying strike-slip events during at least the 1970's. In addition, routine location uncertainties for the offshore events would make applying the distance criterion difficult.

We left in the Homestead Valley sequence, even though it was clearly a swarm with four "main events" of about M5, because to remove it would have raised the possibility that we were "trimming" the data to favor our eventual conclusion. Given that the data set now includes 30 main events, one event more or less makes little difference anyway.

This resulted in a list of 30 events (Table 1), 15 of which had (at least one) foreshocks of magnitude 2.0 or greater. A frequency plot of the "main-shock minus largest foreshock difference" (Figure 4), has a strong central tendency, essentially eliminating the possibility of a uniform distribution of foreshock magnitudes for strike-slip events in California. The data have a mean of 1.89 ($\pm$ 1.04 (1 s.d.)) almost exactly equal to the mean value of 1.9 obtained by Papazachos (1975) for the same calculation on a somewhat larger foreshock data set in Greece.

There are two problems, however, with using a Gaussian distribution to characterize δM. First is that the left hand tail of the distribution extends to negative values, yet by the definition of a foreshock, it cannot be larger than the main event. The second problem is that a Gaussian distribution predicts very small probabilities for foreshocks more than 4 magnitude units smaller than the main event ($\delta M > 2\sigma$), and while we have very little data in this part of the domain, there are a few very small foreshocks to large events, and it clearly makes little sense to assume that they never occur.

A log-normal distribution provides a better over-all fit to the data (dashed line in Figure 4) since it goes smoothly to zero as δM goes to zero, and has a long asymmetrical tail to the right, solving both of the above problems. The best-fit Gaussian on $\ln(\delta M)$ has a mean value ($\overline{\delta M_e}$) of 0.52, and a standard deviation (σ_e) of 0.62 (Figure 4). The expectations for the mean and standard deviation of the log-normal distribution on δM are 2.06 ± 1.10, little different than the values cited above for a Gaussian distribution. This log-normal distribution, with these values, will be used to characterize the foreshock-mainshock relation in the work that follows.

Table 1.
Main Events / Foreshocks
In California, 1970-1995
(Data from the CIT, UCB, and USGS catalogs)

Year	Month	Day	Hour	Min.	Sec.	Latitude	Longitude	Depth	MS Mag	FS Mag
1970	9	12	14	30	52.9	34.269	117.54	8	5.2	4.1
1971	9	30	22	46	11.3	33.033	115.82	8	5	
1972	2	24	15	56	51	36.59	121.2	5.4	5.1	3.4
1974	11	28	23	1	24.6	36.92	121.47	6.1	5.1	
1975	6	1	1	38	48.7	34.512	116.488	0.1	5	3.4
1976	11	4	10	41	37.7	33.115	115.616	4.6	5	4
1979	3	15	21	7	16.5	34.327	116.444	2.4	5.3	4.9
1979	8	6	17	5	22.9	37.1	121.51	8.9	5.6	
1979	10	15	23	16	53.4	32.613	115.318	12.2	6.4	
1980	1	24	19	0	8.7	37.83	121.77	14.5	5.4	2.6
1980	2	25	10	47	38.5	33.501	116.513	13.5	5.5	
1981	4	26	12	9	28.4	33.098	115.632	3.7	5.7	4.1
1981	9	4	15	50	50.1	33.651	119.093	6	5.5	
1984	1	23	5	40	19.9	36.36	121.89	10	5	
1984	4	24	21	15	18.8	37.31	121.68	8.6	6	2.3
1986	1	26	19	20	50.9	36.8	121.28	8.9	5.5	
1986	3	31	11	55	39.8	37.48	121.68	8.9	5.7	2.7
1986	7	8	9	20	44.5	33.999	116.608	10.3	5.6	
1987	11	24	13	15	56.7	33.014	115.852	10.8	6.6	6.2
1988	2	20	8	39	57.2	36.79	121.31	9.7	5.1	
1988	6	13	1	45	36.5	37.39	121.74	9.8	5	
1988	12	3	11	38	26.4	34.151	118.13	14.2	5	
1988	12	16	5	53	4.9	33.979	116.681	8.1	5	2.6
1989	8	8	8	13	27.4	37.14	121.93	13.9	5.4	
1989	10	18	0	4	15.3	37.04	121.88	16.8	7	
1990	2	28	23	43	36.7	34.143	117.697	4.4	5.4	3.7
1992	4	23	4	50	23.2	33.961	116.318	12.3	6.1	4.6
1992	6	28	11	57	34.1	34.2	116.437	0.9	7.3	3.6
1992	7	11	18	14	16.1	35.21	118.066	10.6	5.7	
1995	4	23	8	41		36.627	121.201	7.3	5	3.2

Reformulation

The generality of the Agnew and Jones formulation of the foreshock probability problem allows us to easily incorporate this new result, since the function plotted in Figure 4 can be utilized directly as $\Phi_m(M_C, M_F)$. To this end, equation (8) can easily be written in a more general form.

$$P(C|F \cup B) = \frac{(A/A_C)P(C)\int_{m_1}^{m_2} \Phi_m(M_C, M_F)dM}{(A/A_C)P(C)\int_{m_1}^{m_2} \Phi_m(M_C, M_F)dM + (\delta_1/T')(10^{a-bm_1} - 10^{a-bm_2})} \tag{17}$$

where $P(F|C) = \int_{m_1}^{m_2} \Phi_m(M_C, M_F)dM$, and $\Phi_m(M_C, M_F)$ is the density function that describes the distribution of the largest foreshocks (M_F) before main events (M_C). ($\Phi_m(M_C, M_F) = N_m$ in A&J.)

The log-normal distribution fitted to the data in Figure 4 can be directly incorporated by taking:

$$\Phi_m = \frac{\alpha}{\sigma_e \cdot \delta M\sqrt{2\pi}} e^{-\frac{(\ln(\delta M) - \overline{\delta M_e})^2}{2\sigma_e^2}} \tag{18}$$

where α is the fraction of main events with foreshocks (0.5 for the data in Table 1 and Figure 4), $\delta M = (M_C - M_F)$, and $\overline{\delta M_e}$ and σ_e are the mean and standard deviation of the $\ln(\delta M)$ distribution. Substituting Φ_m into the expression for $P(F|C)$, results in:

$$P(F|C) = \frac{\alpha}{\sigma_e \cdot \delta M\sqrt{2\pi}} \int_{M_F-\mu}^{M_F+\mu} e^{-\frac{(\ln(M_C - M) - \overline{\delta M_e})^2}{2\sigma_e^2}} dM$$

$$= \frac{\alpha}{2}\left[erf\left\{ \frac{\ln(\delta M + \mu) - \overline{\delta M_e}}{\sigma_e\sqrt{2}} \right\} - erf\left\{ \frac{\ln(\delta M - \mu) - \overline{\delta M_e}}{\sigma_e\sqrt{2}} \right\} \right] \tag{19}$$

where μ is the uncertainty in the estimation of the foreshock magnitude. (In this study μ is taken as 0.25 magnitude units. Note that with the use of a log-normal density function for Φ_m, μ cannot be taken small and eliminated with a Taylor expansion, as was done in (8') above, and in A&J. Since μ appears in both the numerator and denominator, however, the results obtained are not strongly dependent on the exact value used. Moreover, in any realistic use of these results, there would in fact be an uncertainty of the order of 0.25 magnitude units in any real-time magnitude estimate, so it would be artificial and misleading if μ did not appear in the final expression.)

Substituting this expression for $P(F|C)$ into (17) results in:

$$P(C|F \cup B) = \frac{\frac{A}{A_C}P(C)\frac{\alpha}{2}\left[erf\left(\frac{\ln(\delta M+\mu) - \overline{\delta M_e}}{\sigma_e\sqrt{2}}\right) - erf\left(\frac{\ln(\delta M-\mu) - \overline{\delta M_e}}{\sigma_e\sqrt{2}}\right) \right]}{\frac{A}{A_C}P(C)\frac{\alpha}{2}\left[erf\left(\frac{\ln(\delta M+\mu) - \overline{\delta M_e}}{\sigma_e\sqrt{2}}\right) - erf\left(\frac{\ln(\delta M-\mu) - \overline{\delta M_e}}{\sigma_e\sqrt{2}}\right) \right] + \frac{\delta_1}{T'}10^{a-bM_F}(10^{b\mu} - 10^{-b\mu})} \tag{20}$$

if one derives a and b from a declustered catalog from which all foreshocks have been removed, or

$$P(C|F \cup B) = \frac{(A/A_C)P(C)\frac{\alpha}{2}\left[erf(\frac{\ln(\delta M+\mu)-\overline{\delta M_e}}{\sigma_e\sqrt{2}}) - erf(\frac{\ln(\delta M-\mu)-\overline{\delta M_e}}{\sigma_e\sqrt{2}}) \right]}{(\delta_1/T')10^{a'-b'M_F}(10^{b'\mu} - 10^{-b'\mu})} \quad (20')$$

if one derives a' and b' from a catalog, which contains background events and foreshocks.

The impact of these changes is clear in Figure 2, where *P(F)*, *P(B)*, and $P(C|F \cup B)$ are plotted for the A&J formulation with a flat foreshock distribution (curve 1), and for the log-normal distribution with two sets of *a & b* values (curves 2 & 3). The peak in the log-normal curve for P(F) at about M6, translates into a peak in the corresponding curves (2&3) $P(C|F \cup B)$ at about M6.5. Curve 2 is based on the unreasonably low a value used by A&J for the Palm Springs segment, curve 3 is based on more realistic values from Wesnousky (1994) for the southern San Andreas, scaled to a segment of 55 km length. We believe that curve 3 is a better estimate of the probability of an earthquake on the Palm Springs segment being followed by an M7.5 characteristic event on the San Andreas in the Coachella Valley, and suggests that the probability following the Joshua Tree earthquake was about 7%, rather than the 21% cited by Jones (1994).

Is Parkfield different?

In their probability calculations BEA used the observation that one half of the M6 events at Parkfield had foreshocks within 1 magnitude unit, while Jones (1984) had found that for strike slip events throughout California, about half the events studied had foreshocks within three magnitude units. Thus BEA effectively used a foreshock occurrence rate at Parkfield three times that observed for the state as a whole.

Michael and Jones (1995) have questioned whether this was justified, given the small sample size of four events, at least two of which had M5 foreshocks. They attempted a statistical test of the question of whether the Parkfield foreshock occurrence rate in this century (50% within one magnitude unit of the main event, sample of four) was significantly different from the prediction of their flat foreshock distribution (15% within one magnitude unit). They used a binomial distribution to conclude that there was an 11% chance of such an occurrence, and therefore concluded that the hypothesis that Parkfield was drawn from the generic population could not be rejected at the 95% level.

There are several things wrong with this test. First is that the flat foreshock distribution results in a hypothesis test whose outcome is arbitrary, and depends entirely on the range of integration. For instance, Parkfield earthquakes have not had a distribution of foreshocks between M5 and M6, but rather the last two events have had essentially identical M5 foreshocks (Bakun and Lindh, 1985). It would have been just as reasonable to test for two foreshocks of exactly M5, or more reasonably, for two foreshocks of $M5 \pm \mu$ where μ is a measure of the magnitude uncertainty. If we repeat Michael and Jones test with a value of $\mu = 0.25$, the same value that was used above, Parkfield foreshocks are different from the generic distribution at the 97% level. But no sensible test can depend so critically on the range of integration; tests must be done against real density distributions with expectations and variances. The correct test is a Student's t-test between the means of the two distributions, but the flat foreshock distribution has a mean that again, depends entirely on the arbitrary choice of a range of integration.

The second problem with Michael and Jones test is that their own data set doesn't fit their generic distribution. The data set of 31 earthquakes they have assembled has 17 events with foreshocks; only three of these sequences have foreshocks within 1 magnitude unit. That is, in their own data set, fewer than 10% of the events have foreshocks within one magnitude unit, even though their assumed generic distribution gives 16.67%. If we were to do a non-parametric significance test against this data set, we would have to remove the '66 Parkfield event from the data set, leaving 2 in 30 sequences with foreshocks within 1 magnitude unit. Using the observed p=2/30=0.067, the significance of rejecting Parkfield, as drawn from the generic distribution, approaches 99%.

Fortunately, most of these difficulties are removed if we use the log-normal distribution derived above. While the expectation of this best fit log-normal distribution is 2.0, and the median is 1.6, the peak of the distribution occurs at value of $\delta M = 1.1$ (Figure 4). Thus if we attempt a maximum likelihood solution of a log-normal distribution to the Parkfield foreshock data, it is clear we will get something indistinguishable from the solution obtained above for the San Andreas system. The Parkfield data has two events with $\delta M \sim 1$, the best fit to the San Andreas data set predicts most likely values of $\delta M = 1.1$, therefore the Parkfield data is entirely consistent with the larger data set. (We remain somewhat concerned that the log-normal distribution shown in Figure 4 may be biased to the left, toward lower values, by the inherent incompleteness of the data set at values of δM greater than 3-4, due to the detection threshold of the networks that record the small foreshocks, but this will require further work, and more data, to clarify. The Parkfield probabilities may have to be revisited after this work, some of which is outlined in the **Discussion and Summary** section, is completed.)

Estimating Parkfield Foreshock Probabilities

Having demonstrated that the log-normal distribution derived above provides a reasonable description of foreshock magnitudes, in this section we use this distribution, in conjunction with the generalization derived above of the A&J foreshock formulation, to obtain estimates of Parkfield foreshock probabilities. However some choices must still be made before we proceed with new estimates of the probability that an earthquake at Parkfield will be a foreshock to a characteristic event. These include:

1. The background seismicity rates (*a* and *b* values) to be used, which follows directly from the choice of geographic coordinates of the fault segment considered, the earthquake catalog used, and the time period considered.
2. The probability of a characteristic event without a foreshock (P_C), which is a slightly more complicated question than it appears in BEA (or A&J for that matter).

To summarize our choices before defending them, we have chosen to:

1. Use new *a* and *b* values for two boxes, the original Parkfield box of BEA, and the small Middle Mountain box of M&J. For both boxes we have used the NCSN catalog for the period 1971-1995 so as to obtain reliable rate estimates for events in the M2-4 range, and the catalog for Meagher (1993) for the period 1934-1969 to help constrain the rate of M4-5 events.
2. We have accepted the consensus estimate of Hager et al. (1994) that there is approximately a 10% change per year of a characteristic Parkfield earthquake, but have partitioned this probability equally between the Parkfield box, and the smaller Middle Mountain box.

To obtain estimates of the background seismicity rates at Parkfield we have combined data from two or more catalogs, covering different time periods and magnitude ranges. We used the NCSN micro-earthquake catalog covering the period since 1971, which has the advantage of precise location and magnitude estimates, and is complete to below the M2 level (diamonds in Figure 6). We obtained estimates of the annual rates of M4-6 events from the Meagher catalog for the period 1934-1970 to ensure that estimates for moderate events are stable (squares in Figure 6). Also shown as a point of reference is the rate for Parkfield main events (M6-6.2) from the Ellsworth (1990) catalog, which includes events from 1881 to the present.

We have followed BEA in partitioning the Parkfield segment into two regions, one beneath Middle Mountain, which contains the 1966 main event and foreshocks, and the other which contains the balance of the San Andreas seismicity at Parkfield. Because the relatively large size of the original Middle Mountain box was determined in part by concerns about the difficulty of obtaining precise real-time locations for potential foreshocks, and because those concerns have proven unwarranted, we have followed Michael and Jones (1995) in reducing the size of this box, to include little more than the 1966 epicenters (Figure 7). We have chosen not to decluster the catalogs, and will use the corresponding equation (20') in the computation of the foreshock probabilities.

For the Parkfield region, excluding the small Middle Mountain box, both sets of frequency magnitude data are reasonably well fit by a single straight line (Figure 6a); the a and b values of 3.06 and 0.87 imply higher rates at the M3-5 level than those obtained by Michael and Jones, which likely reflects our inclusion of Meagher's data for M4-5 events, and their declustering.

The situation is rather worse for the small Middle Mountain box. The frequency-magnitude plot clearly is not well fit by a single straight line; a line through the NCSN catalog, spanning 15 years, clearly implies rates of M4-5 events a factor of three, or more, less than observed in the Meagher catalog. We have chosen to use a single fit to all the data as a compromise; the resulting *a* and *b* values of 1.41 and 0.54 again imply higher rates in the M3-5 range than those obtained by Michael and Jones, most likely for the same reasons. As we will show below, however, these difference result in negligible differences in the final computed probabilities.

The probabilities that result from the application of equation (20'), and the related assumptions discussed above, to these *a* and *b* values, are plotted as a function of magnitude in Figure 5. The values for the Parkfield box top out at about 13% for M5.3 events (small dashes), and the largest probabilities for the Middle Mountain box are about 19% for M5.2 events (solid line). Because of the peak in the log-normal foreshock distribution curves, both curves roll off rapidly at larger magnitudes; both are below 10% for events of M5.6 or greater. For the same reason, both curves also roll off faster at lower magnitudes than the corresponding curves in BEA, or Michael and Jones.

Cholame Segment

Sieh (1978) documented the felt reports for the two moderate foreshocks to the great 1857 earthquake; he speculated that these events represented "...a moderate Parkfield-Cholame sequence similar to those of 1901, ... 1966." Harris and Archuleta (1988) estimated that sufficient strain had accumulated along the 50 km of the San Andreas southeast of Cholame to make an M7.2 event possible along that segment. Lindh (1988) used similar arguments, and the geologic observations of Lienkaemper and Sturm (1989) to estimate that there was a 58% chance of such an event in the next thirty years; Dieterich et al. (1988) used slightly different parameters, and estimated there was an approximately 30% chance of an M7 on the Cholame segment in the next thirty years.

A&J attempted an estimate of the probability that some future event might be a foreshock to an M7 on the Cholame segment (Curve 1 in Figure 8). There are clearly problems with their approach on this segment -- the probability of 52% that an M6 is a foreshock to an M7 is clearly too high -- and they attempted to deal with this by computing a value based solely on the rate of Parkfield main events. This ad hoc adjustment produced estimates that ranged from 3 to 10%, depending on the foreshock scenario assumed.

The log-normal model, and approach presented here, allows a somewhat more straightforward calculation. One problem with A&J's work is that they estimated background seismicity rates for a Cholame box which did not include the Parkfield box; this resulted in a low *a* value, and unreasonably high probabilities; it also is of little use in estimating the likelihood that an M6 at Parkfield is a Cholame foreshock. We have corrected this by expanding the Cholame box to include the Parkfield box, and estimated a set of *a* and *b* values (2.49 and 0.67) (Figure 6c) that differ little from those for the Parkfield box alone. If we use these *a* and *b* values, with Dieterich et al.'s (1988) long term probability estimate ($P(C)=(.3*3)/(30/365)=8.2 \times 10^{-5}$) and the log-normal model, we obtain an estimate of 5% that a M6 Parkfield earthquake will be followed within three days by an M7 event on the Cholame segment (Figure 8). (Remember, these magnitudes have an (1 sigma) uncertainty of 0.25 magnitude units, so M7 (± 0.25) includes Harris and Archuleta's M7.2 event.)

Summary and Discussion

Original Parkfield foreshock probability calculations (OFR 87-192)

Since the expiration in January of 1993 of the Bakun and Lindh (1985) long-term prediction of the next Parkfield earthquake, many questions have been raised about the conduct of the Parkfield Prediction Experiment. While it is true that it is always easier to find fault with a model, and the conduct of a critical experiment, after the model has been experimentally falsified, it is also true that in the ten years that have passed since the Parkfield Prediction Experiment formally began, additional work has been done, and some new understanding gained. Thus, it would be surprising if nothing could be found to criticize in the original arguments and calculations, and it is certain that if we were to begin the experiment today, some things would be done differently, and certainly some arguments would be stated differently. In this regard, we find it somewhat surprising, given the errors that were made, and the limited documentation provided, that the original foreshock probability calculations stand up as well as they do. The value of 0.37 for the probability that a M4.5 (or larger) event within the Middle Mountain box is a foreshock, is remarkably close to the value of 0.39 that we would obtain today, using the same assumptions and input values (compare curves 1&3 in Figure 1a). The result would be stated somewhat differently -- the probability would be for a potential foreshock of $M5.0(\pm 0.5$ magnitude units) -- and hopefully the explanation and documentation would be clearer and more extensive, but the answer would be essentially the same.

The same general conclusion applies to the probability calculations for events of smaller magnitude, although here the differences are a little larger. The probabilities associated with M2.5 and M3.5 events in BEA are 0.028 and 0.11 respectively; the revised formula (equation 13 above) gives values of 0.036 and 0.14 for the same two magnitudes (compare curves 1 and 3 in Figure 1). These differences would have had no practical consequences of any kind.

Log-normal model

The more significant problem with the calculations contained in BEA, and in the subsequent work by A&J, is the nature of the assumption made about the distribution of foreshock magnitudes. Although the "flat foreshock distribution" is simple and convenient, it does not fit the data (Figures 3 and 4). In particular the flat distribution predicts that foreshocks almost as large as the main event would be as likely as any other size, and this does not appear to be the case. (It also causes difficulties with the definition of a foreshock, since an event within a few tenths of a magnitude unit or less than the main event would in effect "use up" the same strain energy required by the main event, and would in effect make the main event less likely, not more so.) It also has unfortunate algebraic properties, since a constant foreshock density function eventually predicts that more than 100% of events would have foreshocks, if the detection threshold extends to about six magnitude units below the main event.

A much better fit to the available foreshock data, for strike-slip main events of $M \geq 5$ within the San Andreas system, is obtained with a log-normal distribution with a peak about one magnitude unit less than the main event, and an area under the curve such that 50% of events have foreshocks. This distribution fits the data, and is easily incorporated into the Agnew formalism for computing foreshock probabilities.

Applying the log-normal model to the Parkfield foreshocks

Before we apply this log-normal foreshock distribution to the Parkfield case, one remaining question must be answered. Based on their understanding of the locations of the last two Parkfield foreshock/mainshock sequences, BEA defined a three-dimensional box centered on the 1966 hypocenter, within which they assumed that the epicenter of the next Parkfield earthquake, and its foreshocks, would occur; is this assumption still valid today? The answer is clearly maybe. Subsequent work has confirmed the detailed locations on which this assumption was based, but the non-occurrence of the next Parkfield event as expected (as of October 1995), and the somewhat greater diversity of moderate Parkfield seismicity since 1992 (Langbein, 1993), clearly raises questions. Should the Parkfield segment be treated as if foreshocks are equally likely everywhere, or should some preference still be given to the area beneath Middle Mountain where the last two foreshock sequences have occurred?

In an attempt to find a compromise solution to these questions, we have followed Michael and Jones (1995) in defining a smaller Middle Mountain box, but have assumed that there is a 50% chance that the hypocenter of the next Parkfield event will occur inside this smaller box. Similarly we have assumed that there is a 50% chance that the next Parkfield event will initiate elsewhere in the Parkfield region. In both cases we have applied the log-normal foreshock model illustrated in Figure 4. This results in an estimate that a M5 (±0.25) event in the small Middle Mountain box has a 17% chance of being a foreshock to a characteristic Parkfield event, while the number drops to 10% for an M5 event elsewhere in the Parkfield box (curves 1&2 in Figure 6.)

Thus we have in effect partitioned Parkfield foreshocks into two alternate scenarios, and assumed that each has an equal likelihood of occurring. In the first scenario, an M5 event located very near the 1966 hypocenter is assigned a relatively high probability of being a foreshock (about 1 in 6), while according to the second scenario, an M5 elsewhere in the Parkfield region has approximately a 1 in 10 chance of being a foreshock.

One implication of this work is that if it were adopted as an update of the BEA scenario document, and if we followed the Southern California scenario document in revising the probability levels that correspond to level A through D alerts, there would no longer be any level A alerts at Parkfield, since the probabilities would never reach 25%. If the 5% threshold were used for level B alerts, this would require an M4.2 event in the small Middle Mountain box, and an M4.6 event elsewhere. Similarly, if the 1% threshold were used for level C alerts, this would require an M3.4 event in the small Middle Mountain box, and a M 4.0 event elsewhere.

Other implications of the log-normal model

The log-normal foreshock model also has potentially interesting implications for some related seismological questions. For instance, the flat foreshock distribution model explicitly asserts that the potential foreshock contains no information about the magnitude of the impending earthquake. Conversely, the log-normal model includes a peaked distribution, with the most likely magnitude for the impending earthquake being just the foreshock magnitude, plus $\overline{\delta M}$. Thus the "information content" of the potential foreshock is increased in some sense. This might have practical implications if further work should confirm that the log-normal model proposed here is a general feature of foreshock behavior.

One implication of this model has to do with the question of whether an earthquake "knows" how large it will grow, in the first second or two after it initiates, when the dimensions and energies involved still appear to correspond to those of a small event. Brune (1979) asked whether we could falsify what he called the "nonpredictable earthquake model." In this model, the stress, and strength, distribution along a fault vary in a manner that makes the earthquake inherently unpredictable. In particular, Brune suggested that the stress level required to produce "premonitory phenomenon" might be less than the ambient stress over significant portions of the fault, even though stresses were high enough that rupture could propagate through the area, even if it could not initiate there. Thus the suggestion was that any M1 had the possibility of becoming an M2, every M2 the possibility of growing into an M3, etc., right up to M7's having the possibility of growing into an M8, and that no "premonitory signals" would be possible, even in principle, to warn us of the size of impending earthquake.

One implication of the log-normal model presented here is that the "nonpredictable earthquake model" may be falsifiable at some level, at least for the one-half of the San Andreas system strike-slip earthquakes that have foreshocks. If over a period of days to hours before a moderate to large earthquake, a foreshock sequence is underway which contains very significant information about the magnitude of the eventual event, it cannot be the case that the eventual magnitude is entirely unknowable until the dynamic rupture has grown to "mainshock dimensions." In particular, for significant main events with dimensions of ten's of km. and greater, the implication seems unavoidable that foreshock source regions, with source dimensions of one or two km. or less, contain significant information about the final dimension of the main event.

Additional work to be done

The difficulty with distinguishing between the flat- and log-normal foreshock models is the limited magnitude range over which microearthquakes are recorded. The two distributions differ significant only at magnitudes very close to the main event, and at small magnitudes ($\delta M > 3$ or 4); in both cases the log-normal model predicts fewer foreshocks than the flat model. At magnitudes close to the main event, the question is somewhat confused by swarms and paired earthquakes, although we believe the San Andreas data clearly favor the log-normal model. At large δM's, the data is limited by the incompleteness of most micro-earthquake catalogs below M1.5-2; catalogs prior to 1970 are often incomplete below M3 or greater. Thus it is very difficult to determine whether the lack of events at $\delta M > 3$ might be an artifact of catalog incompleteness.

One critical test of this question is possible. For the largest events, *without foreshocks*, that have occurred in the last 20-25 years in California, (since dense seismic networks were installed), it should be possible to push the detection threshold much lower for the period immediately before the main event by re-examining the original seismograms for the stations closest to the main event. This has been done for the Loma Prieta earthquake (M7) (White and Ellsworth, 1993), and it is clear that there are no foreshocks, even for $\delta M \sim 6$ or greater. Similarly for the two M6 events along the Calaveras fault in 1979 and 1984, it is clear that there are no foreshocks for $\delta M \sim 5$. Since there are only about 15 strike-slip events in modern times in the San Andreas system without foreshocks, it will not be a great deal of work to check them all. If very few of them have foreshocks at larger $\delta M's$, this would essentially eliminate the flat foreshock distribution model; remember, the A&J distribution implies that all earthquakes should have foreshocks when catalogs are complete to the $\delta M \sim 6$ level. It would also help place some additional constraints on the log-normal model.

Another problem that clearly requires additional work is the characterization of the foreshock sequence itself. Currently most workers follow Jones (1984) in characterizing a foreshock sequence by a single parameter, the difference between the magnitudes of the main event and its largest foreshock. While this is all that can be done for sequences containing only a single foreshock, many sequences (Oroville, 1975; Haichang, 1975; Joshua Tree and Landers, 1992; Kobe, 1995) contain several events within a short time period, at or very near the main shock hypocenter; many of these sequences appear to be non-aftershock like in character. If these sequences could be characterized quantitatively, and proved to be as relatively uncommon as cursory examination suggests, they might offer the potential of computing larger probability gains, based on foreshock clustering, before some moderate to large events with particularly well developed foreshock sequences.

The foreshock probability formulation contained in equation (17) above could be easily generalized to include an estimate of the background rate of clusters, for instance, and then if a $\Phi_c(M_F, n_F)$ could be determined that described the distribution of clusters preceding larger events, it would be straight-forward to compute cluster probabilities just as foreshock probabilities have been computed above. Since clusters appear to be rarer than single events, this has the potential of resulting in larger probability gains.

ACKNOWLEDGMENTS

We wish to thank Lucy Jones, Andy Michael, and Duncan Agnew for their cooperation in this study, Bill Ellsworth for careful reviews of various versions of this work, and Steve Wesnousky for a very helpful conversation on the pitfalls of making robust estimates of *a* and *b* in the Guterberg-Richter relation from small data sets.

REFERENCES

Agnew, D.C., and L.M. Jones (1991). Prediction Probabilities from foreshocks, *J. Geophys. Res., 96*, 11,959-11,971.

Aki, K., A Probabilistic Synthesis of Precursory Phenomenon, in Simpson and Richards (1981), 566-574.

Bakun, W.H. and A.G. Lindh (1985). The Parkfield, California, Earthquake Prediction Experiment, *Science, 229*, pp. 619-624.

Bakun, W.H., K.S. Breckenridge, J. Bredehoeft, R.O. Burford, W.L. Ellsworth, M.J.S. Johnston, L. Jones, A.G. Lindh, C. Mortensen, R.J. Mueller, C.M. Poley, E. Roeloffs, S. Schulz, P. Segall, and W. Thatcher (1987). Parkfield, California, Earthquake Prediction Scenarios and Response Plans: *U.S. Geol. Surv. Open-File Rep. 87-192*, 45 p.

Brune, J.N. (1979). Implications of earthquake triggering and rupture propagation for earthquake prediction based on premonitory phenomena, *Jour. Geophys. Res., 84*, pp. 2195-2198.

Cole, A. and W. Ellsworth (1995). Earthquake Clustering and the Long-term Evolution of Seismicty near Parkfield, California, 1931-1994, *Seis. Res. Let.*, *66*, p. 28.

Dieterich, J.H., L.S. Cluff, D.C. Agnew, C.R. Allen, W.L. Ellsworth, R.L. Keeney, A.G. Lindh, S.P. Nishenko, D.P. Schwartz, K.E. Sieh, W. Thatcher, and R.L. Wesson (1988). Probabilities of Large Earthquakes Occurring in California on the San Andreas fault, *U.S. Geol. Surv. Open-File Report 88-398*, 62 pages.

Ellsworth, W.L. (1990). Earthquake History, 1769-1989, in R.E. Wallace (ed.), *U.S. Geol. Surv. Prof. Paper 1515*.

Hager, B.H., C.A. Cornell, W.M. Medigovich, K. Mogi, R.M. Smith, L.T. Tobin, J. Stock, R. Weldon (AKA NEPEC Working Group to Evaluate the Parkfield Earthquake Prediction Experiment) (1994). Earthquake Research at Parkfield California, 1993 and Beyond: *U.S. Geol. Surv. Circular 1116*, 14 pages.

Harris, R.A. and R.J. Archuleta (1988). Slip Budget and Potential for a M7 Earthquake in Central California, *Geophys. Res. Let.*, *15*, pp. 1215-1218.

Jones, L.M. (1984). Foreshocks (1966-1980) in the San Andreas system, California, *Bull. Seis. Soc. Amer.*, *74*, pp. 1361-1380.

Jones, L.M., and A.G. Lindh (1986). Foreshocks and Earthquake Prediction at Parkfield, California, administrative report, 24 pages, 8 figures.

Jones, L.M. K.E. Sieh, D. Agnew, C. Allen, R. Bilham, M. Ghilarducci, B. Hager, E. Hauksson, K. Hudnut, D. Jackson, A. Sylvester, K. Aki, and F. Wyatt (1991). Short-term earthquake hazard assessment for the San Andreas fault in southern California, *U.S. Geol. Surv., Open-File Report, 91-32*, 24 pp.

Jones, Lucile M. (1994). Foreshocks, Aftershocks, and Earthquake Probabilities: Accounting for the Landers Earthquake, *Bull. Seis. Soc. Amer.*, *84*, pp. 892-899.

Jones, L.M. and A.J. Michael (1994). A Re-evaluation of the Parkfield Alert Level Probabilities, *EOS*, *75*, p. 469.

Langbein, J. (1993). Parkfield: First Short-Term Earthquake Warning, *EOS*, *74*, pp. 152-153.

Lienkaemper, J.J. and T.A. Sturm, Reconstruction of a channel offset in 1857(?) by the San Andreas fault near Cholame, California, *Bull. Seis. Soc. Amer.*, *79*, 901-909.

Lindh, A.G. (1988). Estimates of Long-Term Probabilities for Large Earthquakes Along Selected Fault Segments of the San Andreas Fault System in California, in *Earthquake Prediction -- Present Status*, Univ. of Poona, Pune, India, pp. 189-200.

Lindh, A.G. and L.M. Jones (1985). Parkfield Foreshock Probabilities, *EOS*, *66*, p. 982.

Meagher, K. L. (1993). Relocation of earthquakes in the Parkfield region, Central California, 1932-1969, in *U.S. Geol. Surv. Open-File Report 93-542*, pp. 215-223.

Michael, A.J. and L.M Jones (1995). A Re-Evaluation of the Seismicity Alert Probabilities at Parkfield, California, *U.S. Geol. Surv., Open-File Report, 95-630*, 19 pp.

Mogi, Kiyoo (1963). Some Discussions on Aftershocks, Foreshocks and Earthquake Swarms -- the Fracture of a Semi-infinite Body Caused by an Inner Stress Origin and Its Relation to the Earthquake Phenomenon (Third Paper), *Bull. Earthq. Res. Inst., 41*, 615-658.

Papazachos, B.C. (1975). Foreshocks and Earthquake Prediction, *Tectonophysics, 28*, 213.

Sieh, K. (1978). Central California Foreshocks of the Great 1857 Earthquake, *Bull. Seis. Soc. Amer., 68*, 1731-1750.

Simpson, D.W. and P.R. Richards (1981). *Earthquake Prediction, An International Review*, American Geophysical Union Monograph, Washington D.C.

Utsu, T. (1979). Calculation of the probability of success of an earthquake prediction (in the case of Isu-Oshima-Kinkai earthquake of 1978), *Rep. Coord. Comm. for Earthquake Prediction*, 164-166.

Wesnousky, S.G. (1994). The Gutenberg-Richter or Characteristic Earthquake Distribution, Which is it? *Bull. Seis. Soc. Amer., 84*, pp. 1940-1959.

White, R.A. and W.L. Ellsworth (1993). Near-source short- to intermediate-period ground motions, in *U.S. Geol. Surv. Prof. Paper 1550-C*, (M.J.S. Johnston, editor), pp. C31-C46.

FIGURE CAPTIONS

Figure One

a) Plots of the probability that a given earthquake within the Middle Mountain alert box at Parkfield, is a foreshock (P_F) vs. M, the event's magnitude.
   - Curve 1. From expression (10), the same as the expression for P_F in Bakun et al., 1987.
   - Curve 2. From expression (11); same as curve 1, except that a foreshock expectation term has been added in the denominator.
   - Curve 3. From expression (13); same as curve 2, except that a factor of 0.76 has been added to the second term in the denominator.
   - Curve 4. From expression (14); same as curve 3, except that the calculation is for a one magnitude unit range centered on M, rather than for a one magnitude unit range above M, as in curves 1-3.

b) Curves 1-4, same as a), except that these curve are for P, the "total probability", that results from application of (15) and (16) to the P_F values plotted in a). Curve 5 is $P(C|F \cup B)$ from expression (5), using the same a, b, and P_C values used in curves 1-4, and an N_M of 0.5.

Figure Two

A plot of *P(F)*, *P(B)*, and $P(C|F \cup B)$, vs M_F for the Palm Springs segment, using the data and formulation of A&J. To allow the plotting of real expectations and probabilities, the slightly more general formulation of (8) has been used, with $\mu = 0.25$; that is, all the appropriate density functions have been integrated over a range of one-half magnitude unit centered on M_B and M_F.

Curve $P_1(F)$: Expected frequency of foreshocks, from the "flat" foreshock distribution of A&J. Computed from the numerator of equation (8), with A/AC=50/110.

Curve $P_2(F)$: Expected frequency of foreshocks, from the log-normal foreshock distribution derived above. Computed from the numerator of equation (20).

Curve $P_1(B)$: Expected background rate from A&J. Annual *a&b* values are 2.96 and 0.97, respectively, derived from their Table 1, with T=11. Computed from equation (7).

Curve $P_2(B)$: Expected background rate from Wesnousky (1994). Annual *a&b* values are 3.24 and 0.89, respectively. Computed from equation (7).

Curve $P_1(C|F \cup B)$: Probability that a given event is a foreshock, from the "flat" foreshock distribution of A&J. Computed from equation (7), using $P_1(F)$ and $P_1(B)$ from above.

Curve $P_2(C|F \cup B)$: Probability that a given event is a foreshock, from the log-normal foreshock distribution. Computed from equation (20), using the same data as $P_2(F)$ and $P_1(B)$ from above.

Curve $P_3(C|F \cup B)$: Probability that a given event is a foreshock, from the log-normal foreshock distribution. Computed from equation (20'), using the same data as $P_2(F)$ and $P_2(B)$ from above.

Figure Three

Frequency histogram vs. δM, the difference between the main shock magnitude, and the magnitude of its largest foreshock, in one-half magnitude unit boxes. Please note that the paucity of events for $\delta M > 3$ does not necessarily imply that there are no foreshocks with those values; limitations of the catalogs used to construct the data sets are such that no such inference is possible.

a) Data from Figure 3 of Agnew and Jones (1991). Data are only approximate, given the nature of the figure, and represent sums along the diagonal, for mainshocks of magnitude M4 and larger.

b) Data from Table III of Papazachos (1975). Events with foreshocks occurring more than 3 days prior to the main event have been omitted.

Figure Four

Same reservations as Figure 3 about the paucity of events for $\delta M > 3$. In constructing Table 1 we attempted to select events such that we believe the data is essentially complete for $\delta M < 3$, but no inferences should be made for larger values.

a) Frequency histogram vs. δM, in one-magnitude-unit boxes, for the fifteen earthquakes from Table 2 with one or more foreshocks. The solid curve is a best-fit log-normal distribution with a log-mean of 0.52, and a log-s.d. of 0.62. The dashed curve is a best-fit Gaussian distribution with a mean of 1.89 and a s.d. of 1.04.

b) Cumulative plot of the same data (solid line), with the best fit log-normal distribution superposed (dashed line).

Figure Five

A plot of $P(C|F \cup B)$ vs. M, computed from equation (20'), for events in the small Middle Mountain, Parkfield boxes (Figure 7). $P(C) = .1 * 3/365 = 8.2 \times 10^{-4}$, $\alpha = 0.5$, $\overline{\delta M_e} = .52$, $\sigma_e = .06$ and $\delta_1/T' = 3/365$.

Curve SMM: For events in the small Middle Mountain box with depths greater than 7.5 km. a and b are 1.41 and 0.54, (from Figure 6b), and A/A_C is set to 0.5 as a means of partitioning $P(C)$ equally between the small Middle Mountain box, and the balance of the Parkfield box.

Curve PKF1: Like SMM, except that events for the Parkfield box, minus those used in SMM a and b are 3.07 and 0.87 (from Figure 6a).

Curve PKF2: Like SMM, except that events for entire Parkfield box (combining events used in SMM and PKF1), a and b are 2.72 and 0.72, and A/A_C=1.0.

Figure Six

Cumulative Frequency-magnitude plots for three geographic regions in the Parkfield area. Annual rates are plotted, of the number of events greater than, or equal to, the indicated magnitude. The points from the NCSN catalog are for the period 1971-1995; this catalog is judged to be complete to about the M1.8 level, but contains relatively few events greater than M4.5. The rates from the Meagher et al. catalog cover the period 1934-1969; this catalog is judged to be complete to about the M4 level, and contains two Parkfield main events (1934 and 1966), their foreshocks and aftershocks.

a) Events are selected from the Parkfield box, excluding those events within the "small Middle Mountain box" (Figure 7), that located below a depth of 7.5 km.

b) Events are selected for the small Middle Mountain box, at depths of 7.5 km and greater.

c) Events are selected from the entire Parkfield box, with events added from the Cholame box of A&J.

Figure Seven

Map of the Parkfield region showing the epicenters of the earthquakes included in the magnitude frequency plots in Figure 6, and the outlines of the boxes used to select events for those figures (Adapted from Michael and Jones, 1995).

Figure Eight

A plot of $P(C|F \cup B)$ vs. M, for the Cholame segment. .

$P(C) = .3 * 3/(30 * 365) = 8.2 \times 10^{-5}$.

Curve A&J: For events in the Choame box of A&J, a and b are 1.41 and 0.54, (from their Table 1, using their "flat-foreshock distribution" model.

Curve Log-normal: Computed from equation (20'), for events in the combinded Parkfield and Cholame boxes (Figure 7), a=3.21, b=0.89.

Probability that an event within the Middle Mountain box is a foreshock vs. the event magnitude

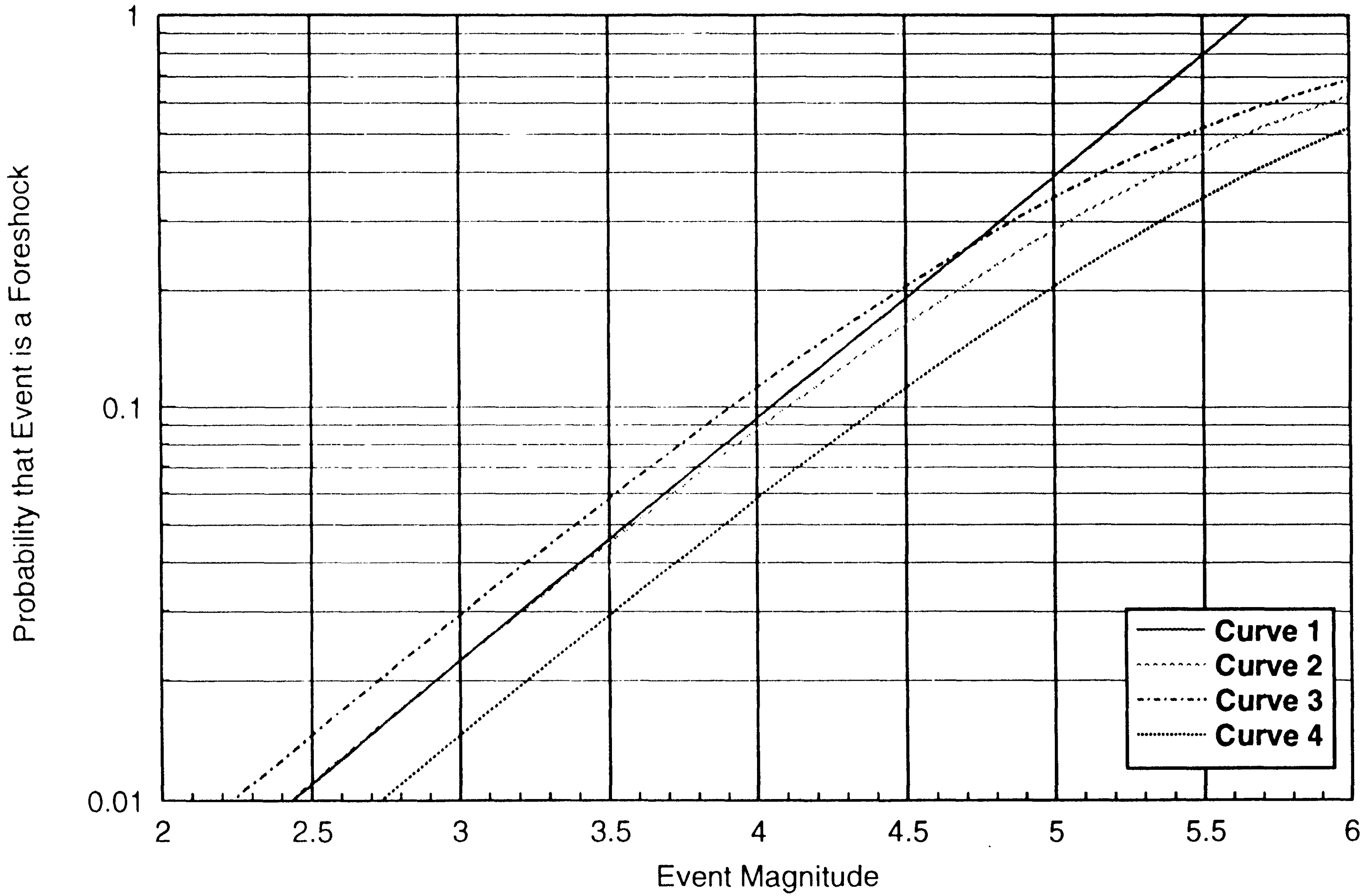


Figure 1-a

Mainshock probability within 72 hours, adjusted for probability gain against Poisson probability vs. candidate event magnitude

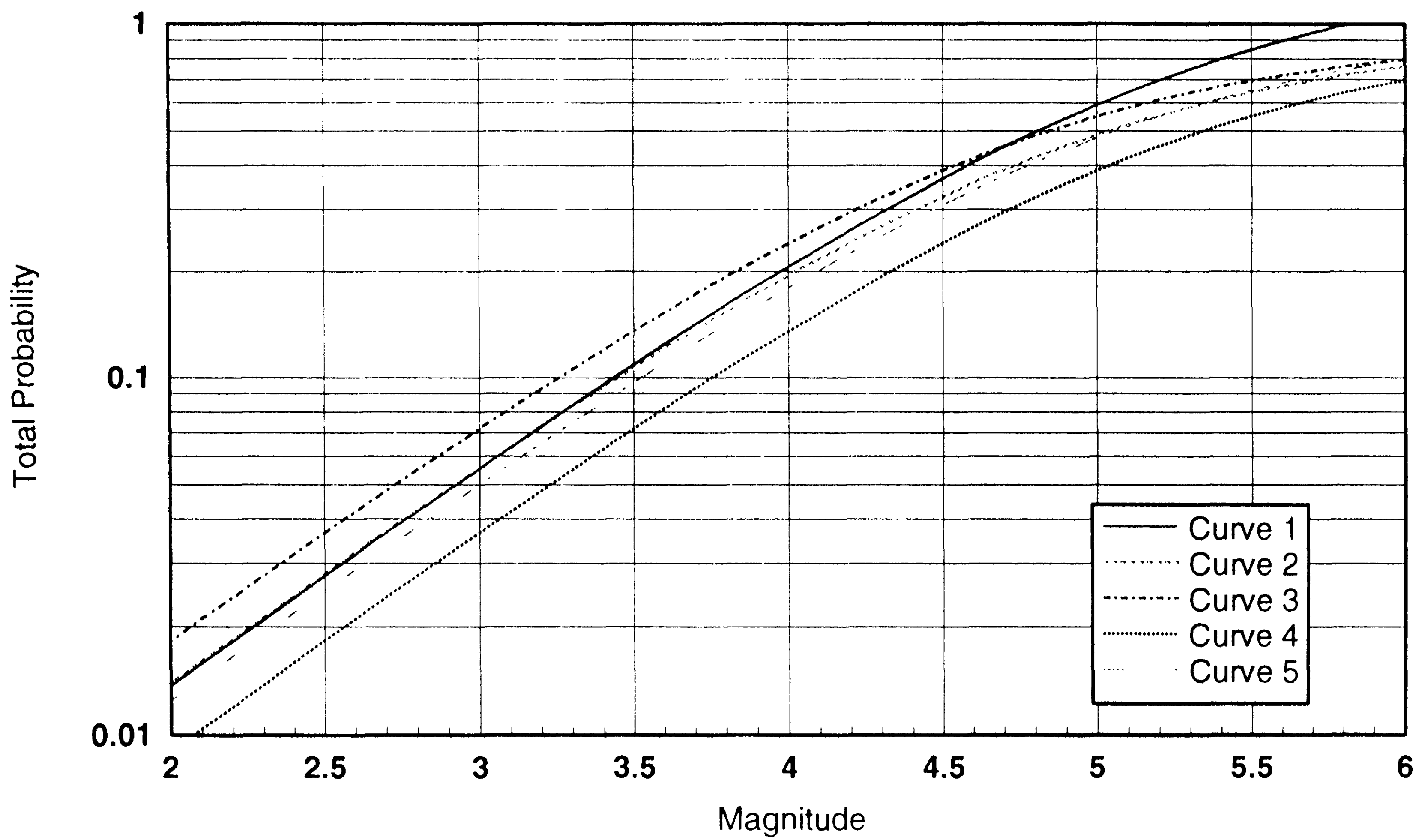
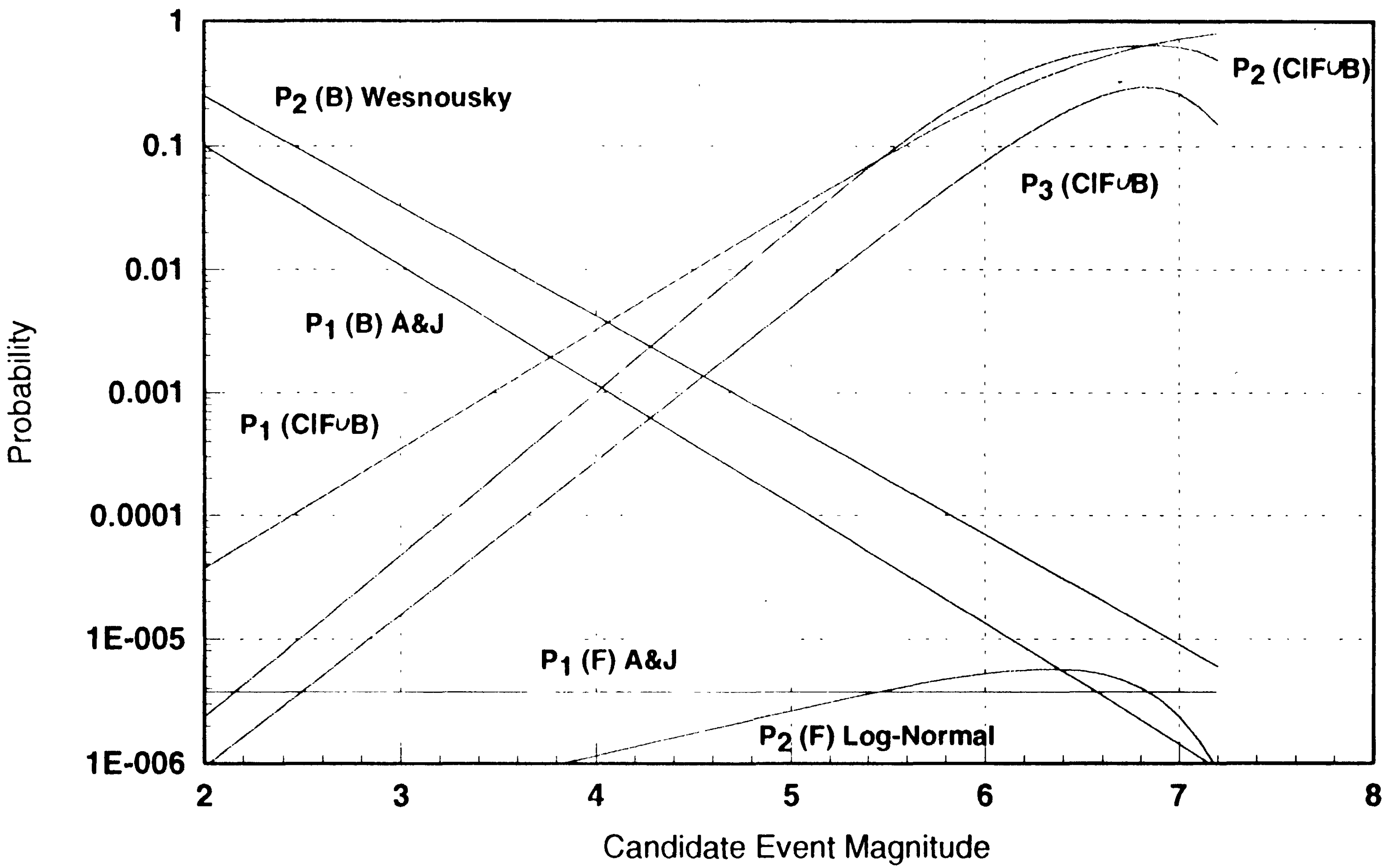


Figure 1-b

Foreshock, Background and Mainshock Probabilities vs. Candidate Event Magnitude



P(F) Foreshock probability
P(B) Background probability
P(CIF∪B) Mainshock probability given a foreshock candidate event has occurred

Figure 2

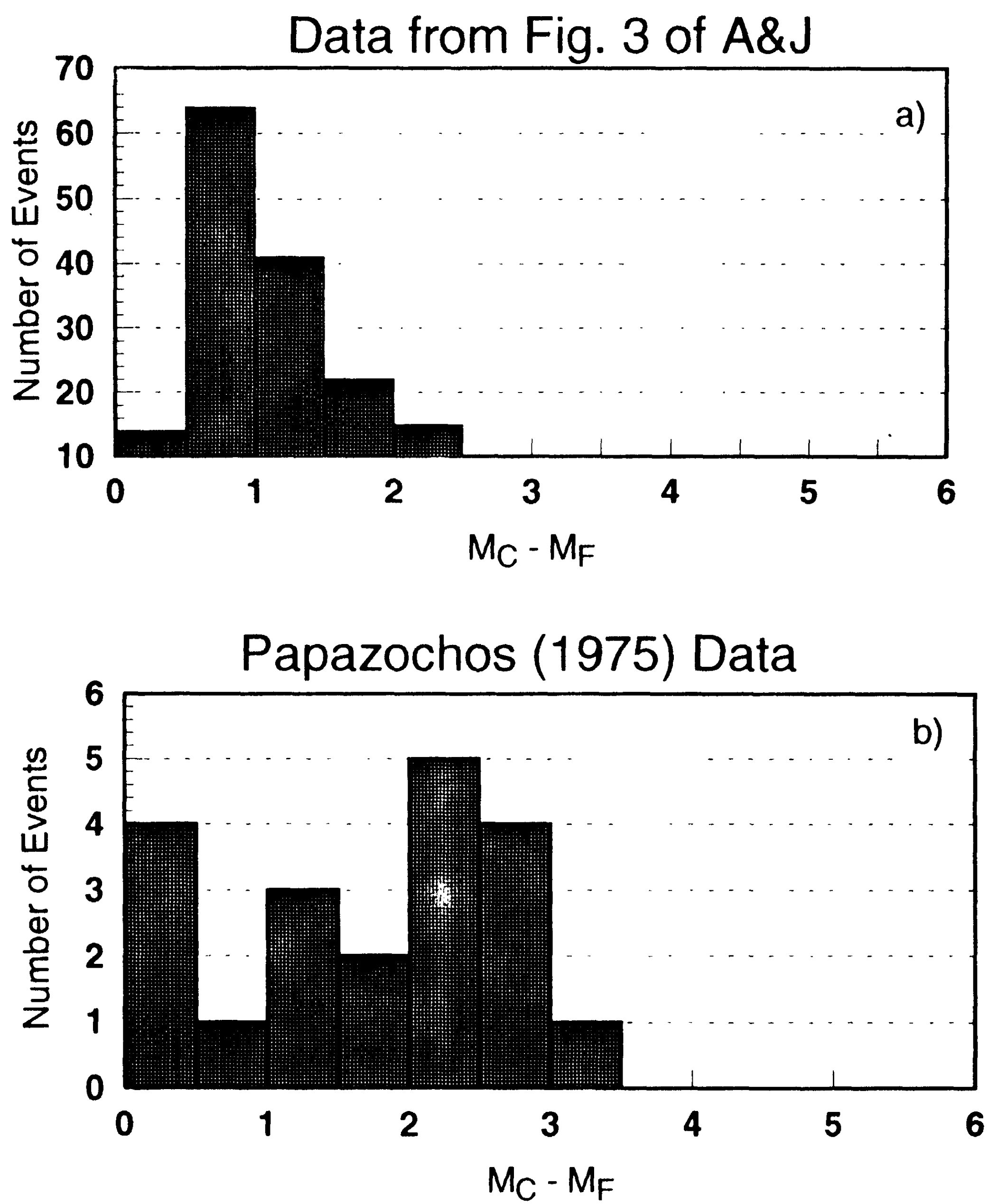

Data from Fig. 3 of A&J
a)
Number of Events
70
60
50
40
30
20
10
0
1
2
3
4
5
6
MC - MF
Papazochos (1975) Data
b)
Number of Events
6
5
4
3
2
1
0
0
1
2
3
4
5
6
MC - MF

Figure 3

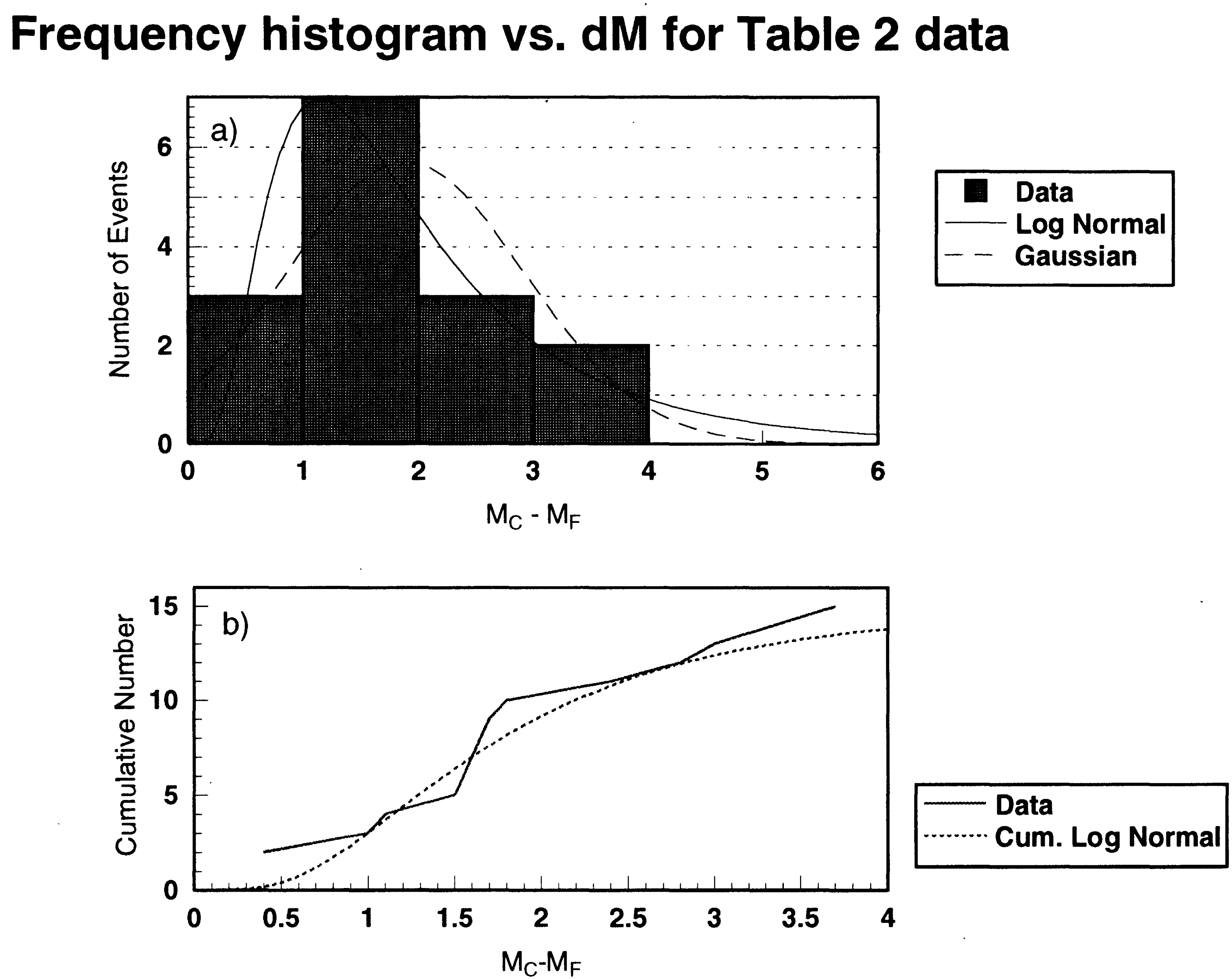
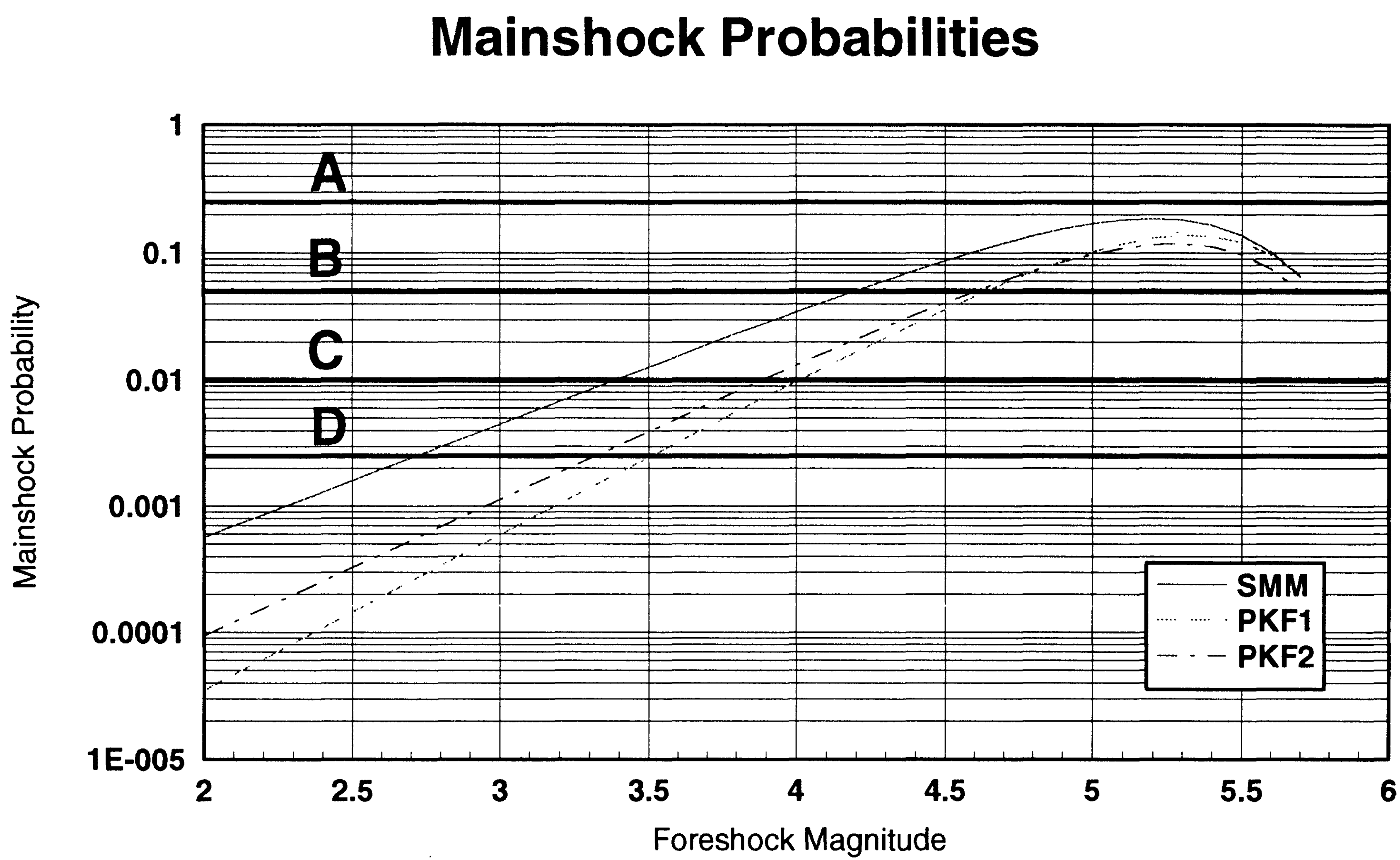

Frequency histogram vs. dM for Table 2 data
a)
Number of Events
0
2
4
6
0
1
2
3
4
5
6
M_C - M_F
Data
Log Normal
Gaussian
b)
Cumulative Number
0
5
10
15
0
0.5
1
1.5
2
2.5
3
3.5
4
M_C-M_F
Data
Cum. Log Normal

Figure 4



SMM Small Middle Mountain; a = 1.41, b = 0.54, $\delta M = 0.52$, $\sigma = 0.62$, $A/A_c = 0.5$

PKF1 Parkfield minus MM; a = 3.06, b = 0.87, $\delta M = .52$, $\sigma = .62$, $A/A_c = 0.5$

PKF2 Parkfield inc. MM; a = 2.72, b = 0.72, $\delta M = .52$, $\sigma = .62$, $A/A_c = 1.0$

Figure 5

Cumulative Frequency-Magnitude Distribution
Parkfield minus small Middle Mountain box
(Includes foreshocks and aftershocks)

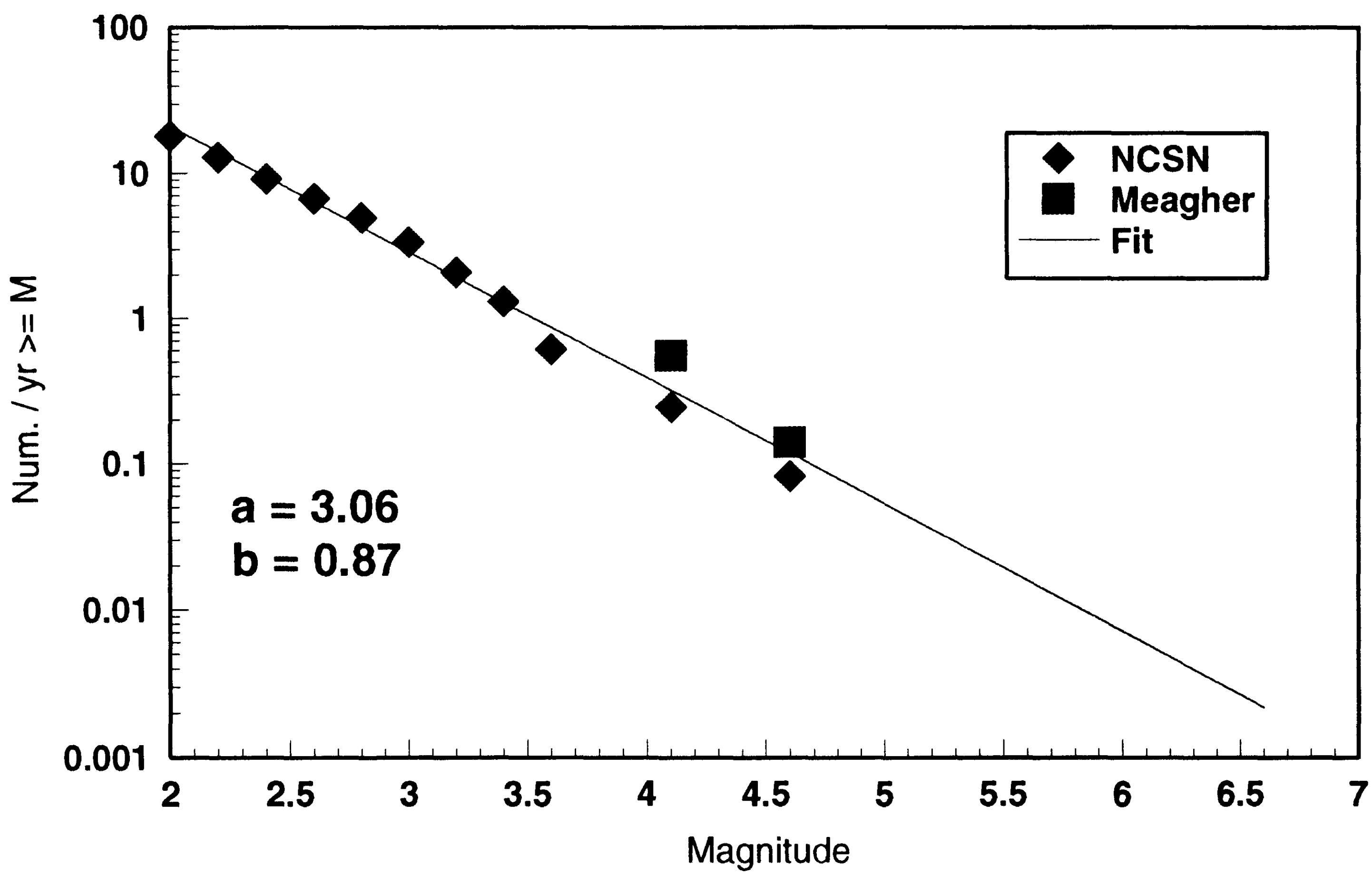


Figure 6a

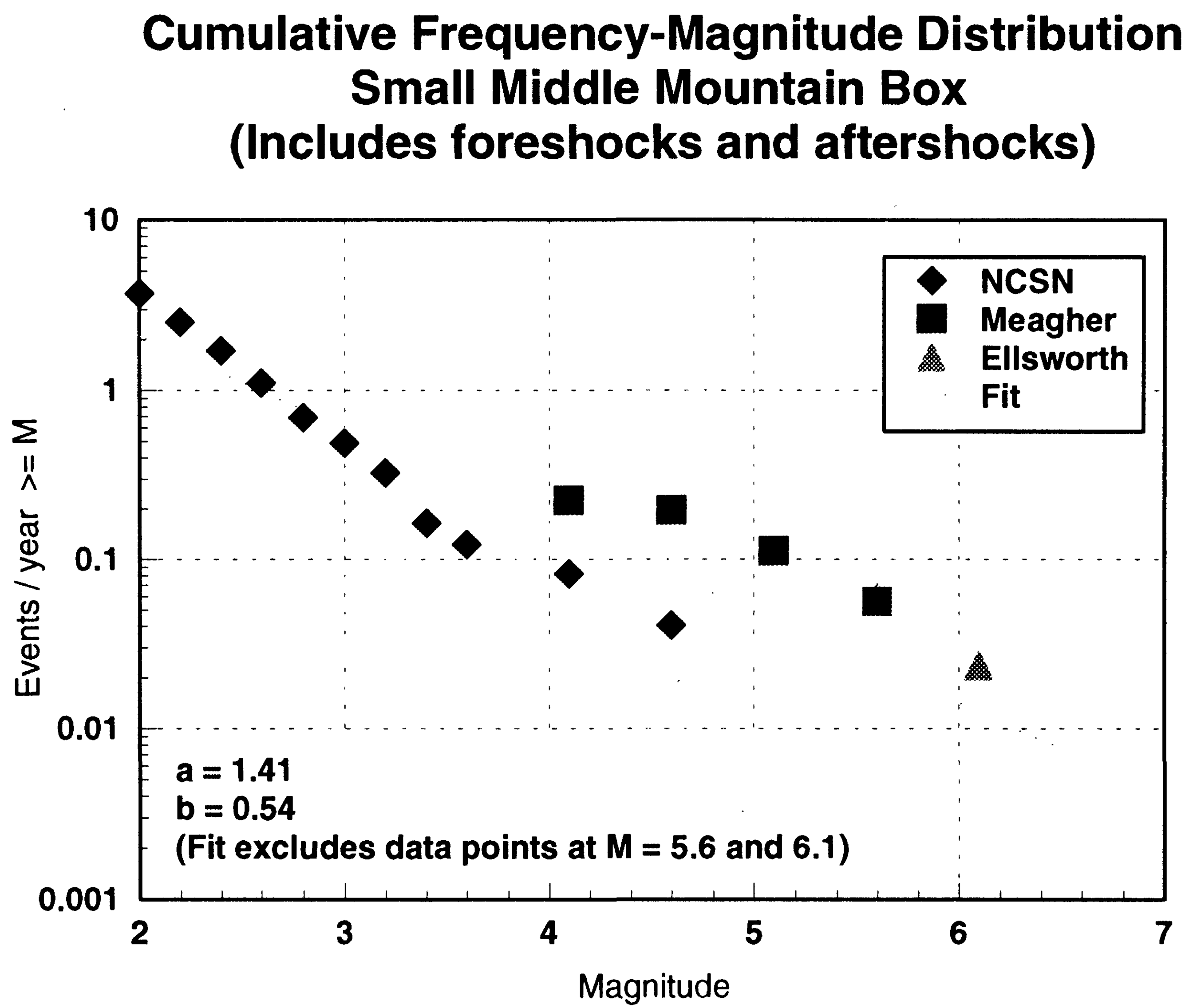

Cumulative Frequency-Magnitude Distribution
Small Middle Mountain Box
(Includes foreshocks and aftershocks)
NCSN
Meagher
Ellsworth
Fit
a = 1.41
b = 0.54
(Fit excludes data points at M = 5.6 and 6.1)
Events / year >= M
10
1
0.1
0.01
0.001
2
3
4
5
6
7
Magnitude

Figure 6b

Cumulative annual frequency-magnitude data for combined Parkfield and Cholame segments

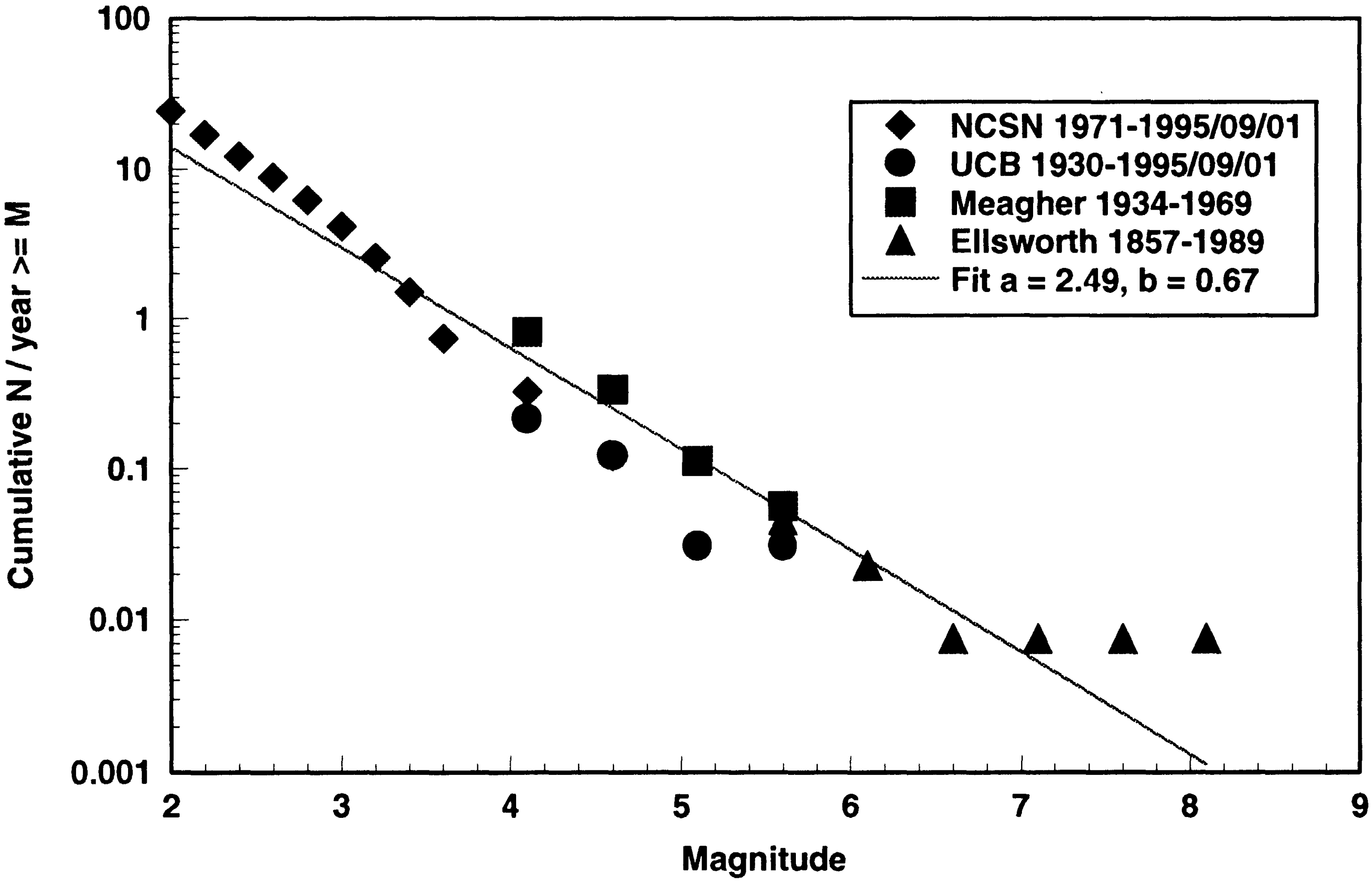


Figure 6c

Parkfield Seismicity 1982-1994 and Possible Alert Boxes

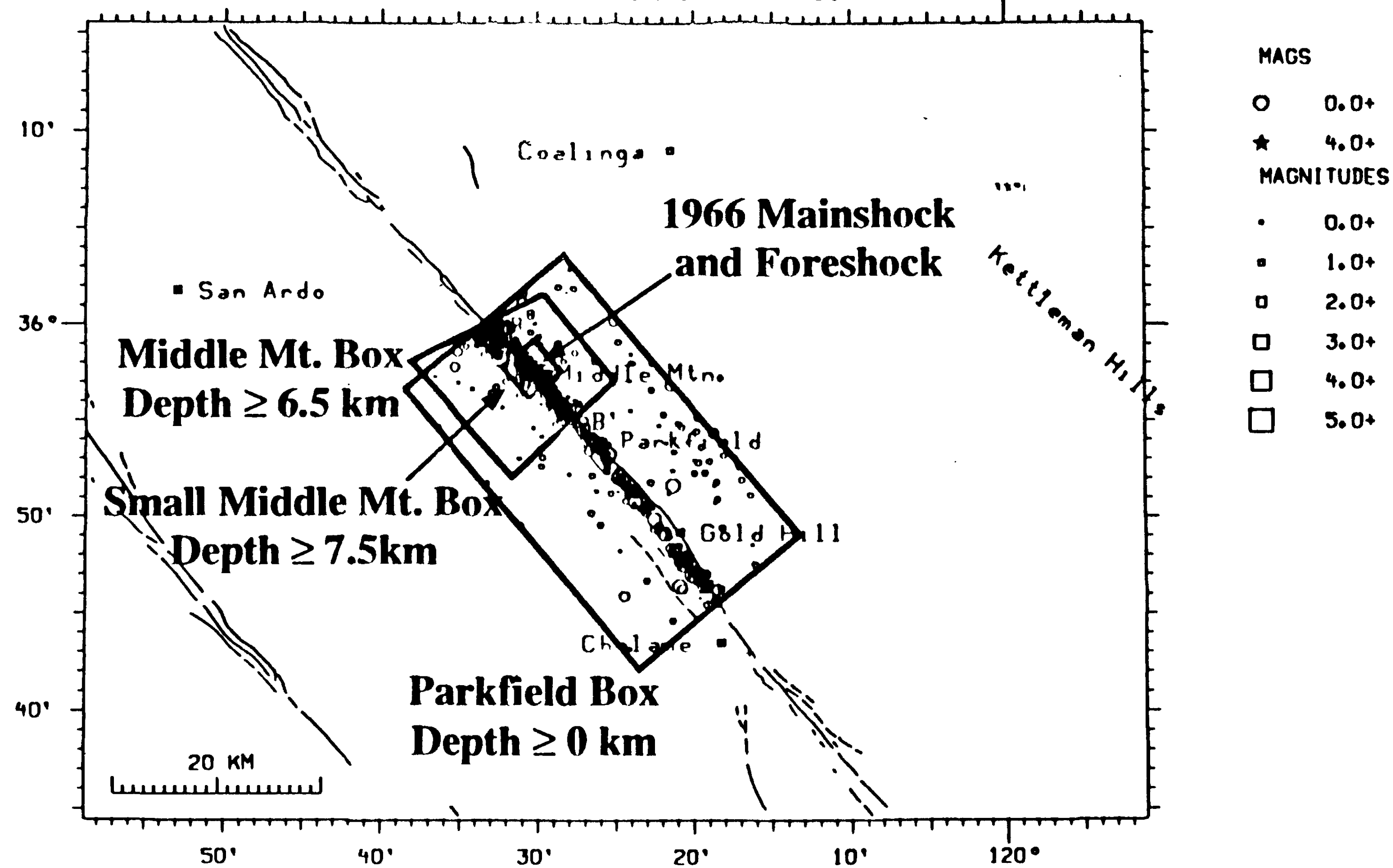


Figure 7.

Probability vs. candidate event magnitude for the Cholame segment

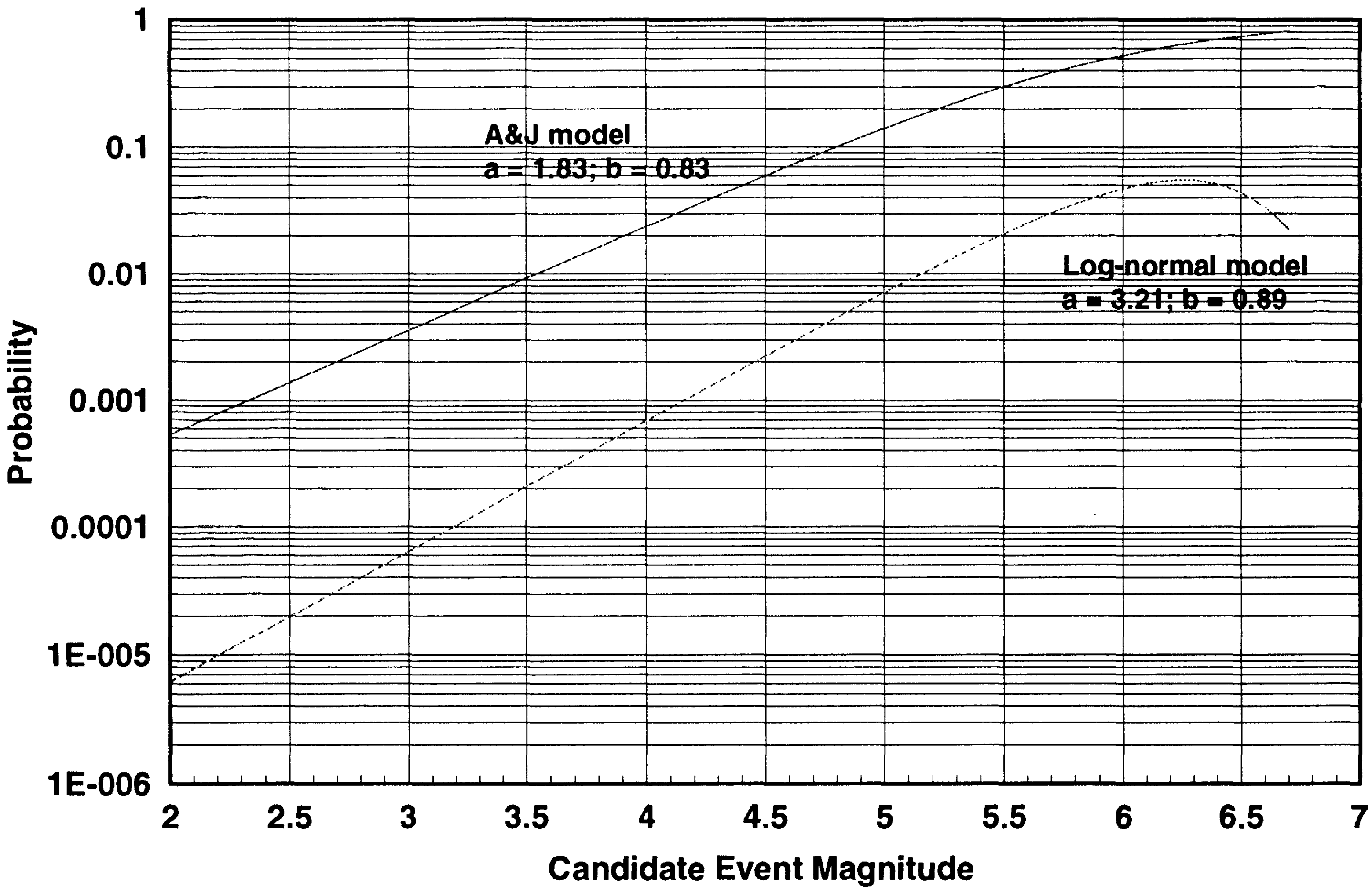


Figure 8