DEPARTMENT OF THE INTERIOR

U.S. GEOLOGICAL SURVEY

A Reevaluation of the Seismicity Alert Probabilities at Parkfield, California

Andrew J. Michael [1]

and

Lucile M. Jones [2]

August 11, 1995

OPEN-FILE REPORT 95-630

[1] U.S. Geological Survey
345 Middlefield Road - MS 977
Menlo Park, California 94025

[2] U.S. Geological Survey
525 So. Wilson Dr.
Pasadena , California 91106

Abstract

For eight years, the US Geological Survey has used the Parkfield Earthquake Prediction Experiment scenario document to estimate the probability that earthquakes observed at Parkfield will turn out to be foreshocks and thus followed by the magnitude 6 mainshock. Much has been learned in that time about the seismic regime at Parkfield, the long-term probability of the Parkfield mainshock and about how to estimate these types of probabilities. The probabilities for potential foreshocks at Parkfield are reexamined and revised in light of these advances. Compared to the earlier assessment, the new estimates of the long-term probability of the Parkfield mainshock are lower, our estimate of the rate of background seismicity is higher and we find that the assumption that foreshocks at Parkfield occur in a unique way is not statistically significant at the 95% confidence level. While the exact numbers can vary depending on the assumptions that are made, the new alert probabilities are lower than previously estimated.

Due to the public policy issues associated with the Parkfield alert probabilities, we stress that the content of this paper represents our opinions and does not currently represent the official position of the USGS.

Introduction

The Parkfield Earthquake Prediction Experiment scenario document (Bakun et al., 1987; hereafter referred to as OFR 87-192) includes estimates of the probability that earthquakes observed in the Parkfield area, and therefore possible foreshocks, will be followed by the magnitude 6 Parkfield mainshock. It also takes these probabilities and relates them to alert levels A through D which can be used to summarize the experiments status. This system has been exercised since the writing of OFR 87-192 including A-level alerts in October, 1992 and November, 1993.

The probability estimates in OFR 87-192 should be reevaluated for three reasons. First, these probabilities depend on the long-term probability of the mainshock occurring. In OFR 87-192 this was based on the model of Bakun and Lindh (1985) which included the prediction that the mainshock would occur before 1993 with 95% confidence. This prediction was not fulfilled and therefore the probabilities should be reestimated.

Second, simply because time has passed, we can use several more years of data to determine the rate of background seismicity. Moreover, OFR 87-192 limits the area in which the higher level alerts can occur based on the observations of foreshocks in 1934 and 1966. Based on the experience of the past few years there are arguments both to further restrict this area and to greatly relax these restrictions that should be considered.

Third, since the time of writing of OFR 87-192, several studies have advanced the art of computing these types of probabilities. These have revealed some subtle inconsistencies in the mathematical approach used in OFR 87-192 to describe the rate of foreshocks before mainshocks and the magnitude distribution of those events. For instance, although an attempt was made to correlate the Parkfield foreshocks with the average rate in California, the rate of large foreshocks assumed in OFR 87-192, based on the events in 1901, 1922, 1934, and 1966, is three times as high at large magnitudes as that observed on strike slip faults throughout the San Andreas physiographic province (as defined by Zoback and Zoback, 1980).

Thus, to complete our reevaluation of the Parkfield alert probabilities, we must reevaluate 1) the rate of foreshocks before mainshocks as a function of magnitude and the appropriate mathematical form to express that, 2) the appropriate alert area (possible location for the foreshock), and 3) the long-term probability of the mainshock. Evaluating each of these factors requires making assumptions, and each assumption has some effect on the results. We will examine the basis for each assumption and their effect on the results. Our final result will be a preferred set of assumptions and the foreshock probabilities that follow from them. This paper supersedes the abstract Jones and Michael (1994).

These probabilities can be used to revise the alert structure used for the Parkfield earthquake prediction experiment. However, that revision is not within the scope of this paper nor can it be done be us. The Parkfield alert structure represents a method for declaring Geologic Hazards Warnings by the USGS. These warnings can only be issued by the Director of the USGS or under a scenario document that has been approved by the Director. We stress that the content of this paper represents our opinions and does not currently represent the official position of the Director or the

USGS. However, it is also our opinion that the current official USGS policy as set out in OFR 87-192 needs to be revised.

Original Methodology

The original Parkfield report established four alert levels associated with the probability that the mainshock would occur in the next 3 days. This probability was calculated by first determining the probability that an earthquake of some magnitude under Middle Mountain will be a foreshock to the Parkfield mainshock and then assigning that probability to a corresponding level (1.5 for Level D, 2.5 for C, 3.5 for B, and 4.5 for A) . OFR 87-192 is too terse to determine exactly how these probabilities were derived, so the following discussion also relies on an unpublished manuscript by Jones and Lindh from 1986.

OFR 87-192 assumes that the probability that a given earthquake was a foreshock was the ratio between the rate at which foreshocks occur and the rate at which non-foreshocks or background earthquakes occur. The rate at which background earthquakes occur was estimated from the cumulative magnitude distribution for declustered earthquakes in the Middle Mountain area (Figure 1) from 1971 to 1984. The catalog was declustered by removing events that were not the largest within 3 days of their occurrence. The result was

$$N/yr = 73.1 \cdot 10^{(-0.62 \pm 0.15)M}$$

The method for determining the rate of the largest foreshocks within a 3 day window is not described in OFR 87-192. Jones and Lindh assumed that, based on average behavior of strike-slip events in California (Jones, 1984) half of Parkfield mainshocks would have a foreshock sequence. Further they assumed that the largest foreshocks in the Parkfield sequences would be an M5 event. This was based on the observations of foreshocks, the largest of which was M5, in 1934 and 1966. We note, that this latter assumption was included in the math used in OFR 87-192 but was not stated in the text which instead said that the magnitude of the foreshocks was unspecified. Given a 21.7 year repeat time for Parkfield mainshocks, the cumulative rate of the largest foreshock, for magnitudes less than 5, was presented as:

$$N_f/yr = 0.5/21.7 = 0.023$$

To get the probability that a candidate event is a foreshock they took:

$$P_f(M) = \frac{N_f/yr}{N/yr} = 3.2 \cdot 10^{-4} \cdot 10^{0.62M}$$

One problem with this method becomes apparent because the probability can exceed one when the rate of background earthquakes is less than the rate of foreshocks. This shows that $P_f(M)$ is not a correctly computed probability. OFR87-192 avoided this problem because they did considered the largest possible foreshock to be magnitude 5, based on the 1934 and 1966 sequences. At this magnitude the probability remains under 1.

The P_f is the probability that the candidate event is a foreshock. To get the probability that the mainshock would occur in the next 3 days two more steps were taken as described in OFR 87-192. First, P_f was taken to be an integral from the origin time of the candidate event to infinity. To convert this to a three day probability a decaying exponential with time was used and this reduces the probability by a factor 0.79. Then the foreshock probability, and the long-term conditional probability as determined from Bakun and Lindh (1985) and Lindh (1983) was combined with the Poisson rate of the mainshock occurring by the formulation of Utsu (1979). These latter two steps take $P_f = 0.2$ and get a probability that the mainshock will occur in the next 72 hours of 0.37.

In retrospect, we see two subtle inconsistencies in the OFR 87-192 methodology. First is the use of cumulative distributions for N/yr and N_f/yr. With this method, one is actually computing the probability that an earthquake of magnitude M or greater is a foreshock. Using cumulative distributions results in a larger probability because the probability is averaged over the higher magnitudes instead of evaluated at a point.

The second problem is in the assumed magnitude distribution of the foreshocks. Jones and Lindh write that they assumed that half of the Parkfield earthquakes would have foreshocks like the average of California strike-slip earthquakes, citing Jones (1984), and that the largest foreshock at Parkfield would have a magnitude of 5. However, as Agnew and Jones (1991) point out, any evaluation of the rate of foreshocks must include the magnitude range for potential foreshocks. Thus, the Jones (1984) results show that 44% of the strike-slip earthquakes have foreshocks within three units of magnitude. By assuming that Parkfield foreshock sequences will include an M5 event, Jones and Lindh used a much higher rate of large foreshocks than the average for California.

Moreover, assuming that the foreshock sequence will always include an M5 presents problems when considering the risk associated with smaller earthquakes. The catalog is declustered by taking only the largest events in ±3 day windows. This makes the form of the foreshock distribution versus magnitude 0 below M5, 0.5 at M5, and 0 above M5. This gives the same cumulative distribution for N_f as used by Jones and Lindh but makes it difficult to see how to determine the risk associated with M ≠ 5 events. Using the cumulative distributions allowed computing probabilities for M<5 events but it is now unclear to us what these probabilities mean. It is also unclear what probabilities are associated with M>5 potential foreshocks.

Preferred Methodology

Since the enactment of OFR 87-192, Agnew and Jones (1991) considered the general problem of deriving the probability that the characteristic earthquake on a fault will occur after a smaller earthquake near that fault, either a background earthquake or a foreshock. This derivation uses rate densities instead of cumulative rates so that the resulting probabilities are for a magnitude rather than for that magnitude or greater. Also, this derivation leaves the assumed magnitude distribution of the foreshocks as a variable so we can choose to use a Parkfield specific or average Californian rate. We prefer the Agnew and Jones (1991) methodology for reevaluating the Parkfield alerts.

Agnew and Jones (1991) assumed that the earthquakes of interest could be divided into three classes: background events, foreshocks, and mainshocks. By this classification foreshocks are always followed by mainshocks and background events are never followed by mainshocks. The problem is that we can not determine whether an event is a background event or a foreshock at the time the candidate event occurs. Once the mainshock has occurred we can tell, but then the question is of little practical use. The useful information that can be determined is that, given that either a background event or a foreshock has occurred what is the probability that it is a foreshock and therefore the mainshock will follow. We call this the alert probability.

They showed that if:

$P(C)$ = Probability of the characteristic mainshock

$P(B)$ = Probability of a background earthquake

$P(F)$ = Probability of a foreshock

$P(C|F \cup B)$ = Probability of a characteristic mainshock given either a foreshock or a background event

$P(F|C)$ = Probability of a foreshock given a mainshock, i.e., the rate of foreshocks before mainshocks

Then:

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

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

This means that the probability that an event is a foreshock is the ratio of foreshocks to total events: foreshocks and background events. The second step assumes that the rate of foreshocks is the rate of mainshocks times the proportion of mainshocks preceded by foreshocks. It satisfies common sense because if either the probability of the mainshock or the percentage of mainshocks preceded by foreshocks ($P(C)$ or $P(F|C)$) is zero, so is the probability of the mainshock. Also, if all events are foreshocks ($P(B)$=0), the probability is 1. Thus, the probability that an earthquake will be a foreshock can be calculated when we know the background rate for that earthquake $P(B)$, the probability of having the Parkfield mainshock independent of any potential foreshocks, $P(C)$, and the rate at which foreshocks precede the mainshock $P(F|C)$.

Input Assumptions

Three values must be determined in order to apply equation 1 to Parkfield. They are the probability that the mainshock will occur, the probability that the mainshock will be preceded by foreshocks, and the rate at which background events occur. Each of these needs to be reexamined.

Mainshock Probability

OFR 87-192 used the model of Bakun and Lindh (1985) to compute the long term probability that the Parkfield mainshock will occur. This model predicted that the Parkfield mainshock would occur before 1993 with 95% confidence, but that prediction was not fulfilled. A review of the Parkfield Earthquake Prediction Experiment

(NEPEC Working Group, 1993) under the auspices of the National Earthquake Prediction Evaluation Panel concluded that a variety of current models suggest that the annual probability that the Parkfield mainshock will occur is around 10%. The models discussed in the NEPEC report are purely statistical analyses of the sequence of six events that occurred in 1857, 1881, 1901, 1922, 1934, and 1966. These models use a variety of distributions that all assume some sort of semi-periodic behavior. An alternative would be to use a Poisson model which would lower the annual probability of the mainshock occurring to 4%. We currently prefer the semi-periodic models because they have 95% confidence regions for the time of the mainshock that do not end until 2001 to 2003 (Savage, 1993). Thus, we do not reject the semi-periodic nature of the Parkfield mainshocks at this time. Choosing higher or lower values would increase or decrease the output alert probabilities respectively. For small long term probabilities the relationship is approximately linear with a slope of 1. Hence, doubling or halving the long term probability would approximately double or halve the alert probabilities respectively. This approximation holds as long as it is much more likely that a candidate event is a background earthquake and not a foreshock.

The OFR 87-192 alert probabilities used a long term probability for the mainshock of 15% per year. This was based on the Bakun and Lindh (1985) model at the beginning of 1986. Applying their model to the current time would result in a larger result (OFR 87-192), thus we need to use a different model. The change we propose lowers the alert probabilities by about one third.

To make the alert probabilities testable we need to define the expected mainshock. In OFR 87-192 the Parkfield mainshock is defined to be a magnitude 6 earthquake along the San Andreas fault near Parkfield, CA. However, the magnitude of past Parkfield earthquakes have not been exactly 6 and in OFR 87-192 a sample warning message modifies this to be "about 6". This is a vague definition because some of the terms are loosely defined such as "about 6", "near Parkfield", and in light of the 1989 Loma Prieta earthquake even the term "along the San Andreas fault" can be open to interpretation.

The long term probability is an input to the alert probabilities so the definition of the expected mainshock should correspond to the catalog of events used to determine the long term probability. As one goes back in time less and less is known about the individual mainshocks in the Parkfield catalog. The teleseismic records show that the last three are all about moment-magnitude 6. And all six events are known to have produced surface rupture along the San Andreas fault system (including the Southwest Fracture) near Parkfield (Bakun, 1988). Thus we define the Parkfield mainshock to be an event with $M_w \geq 5.7$ that produces surface rupture along the San Andreas fault, and/or the Southwest Fracture zone, between 36° N and 35° 45'N and within 5 km of the mapped trace. Coordinates of this box are shown in Table 1 and it is the same length, but half the width, of the Parkfield box shown in Figure 1. Events that produce additional surface rupture outside of this box or on other faults are also Parkfield mainshocks. We do not have a strong preference for the source of the M_w determination. The Harvard CMT catalog, the University of California at Berkeley catalog, and the Caltech-USGS catalog are all reasonable candidates.

We choose to use the surface rupture to describe Parkfield mainshocks because it is a known characteristic of the past six Parkfield mainshocks and because of problems with two other possible measures: the mainshock hypocenter and the centroid of the moment release determined from instrumental means. While the hypocenter of previous Parkfield earthquakes appears to repeat, it seems possible that a hypocenter under Middle Mountain could be accompanied by rupture to the northwest along the San Andreas fault. Or a hypocenter at the southeast end of any Parkfield region could rupture further southeast. In either case an event could have an hypocenter within a defined Parkfield box but primarily release moment elsewhere. This would lead to us defining an event as a Parkfield event while not believing it is really a repeat of the previous events. To improve on the hypocenter we could use a moment tensor centroid, but it is possible that location errors in these determinations would provide a wrong answer until long after the mainshock occurs. If accurately located (errors of less than 5km) moment tensor centroids could be determined quickly they may be a better choice than using the surface rupture.

Foreshock Rate

The next value that must be decided upon is the rate at which mainshocks are preceded by foreshocks. For strike-slip earthquakes in California, Agnew and Jones (1991) used a distribution in which half of all $M \geq 5$ mainshocks are preceded by one or more $M \geq 2$ foreshocks within 3 days, based on the results of Jones (1984). Including the 1966 Parkfield mainshock and foreshock sequence, Jones (1984) studied 16 strike-slip mainshocks with $M \geq 5$, from 1966 to 1980, within the San Andreas physiographic province as defined by Zoback and Zoback (1980). Seven of the 16 sequences had foreshocks within the three days preceding the mainshock and within the three units of magnitude below the mainshock magnitude. To update this, we identified 17 additional strike-slip sequences with $M \geq 5$ mainshocks through the end of 1994. Of these 10 have $M \geq 2$ foreshocks within three days of origin time (Table 1). This gives a total of 17 of 33 sequences that have foreshocks, which is extremely close to the original result of Jones (1984) and therefore there is no reason to update the foreshock rate used in Agnew and Jones (1991).

For the distribution of foreshocks versus magnitude Agnew and Jones (1991) used a flat distribution; any equal sized interval in magnitude, less than the mainshock magnitude, has the same probability of containing a foreshock. This means that one sixth of the events have a foreshock within 1 unit of magnitude, a quarter of the events have a foreshock within 1.5 units of magnitude and all events have a foreshock within 6.5 units of magnitude, based on an analysis of several hundred foreshock-mainshock pairs recorded in southern California. Obviously this flat distribution has problems when using large magnitude differences but this was judged not to be a practical problem. They also noted that the form of the magnitude distribution of foreshocks was very uncertain.

The several hundred foreshock-mainshock pairs used to define the form of the distribution by Agnew and Jones (1991) included mainshocks as small as magnitude 3. Applying this distribution to larger events like Parkfield requires a belief in the self-similarity of the foreshock process. This is because the data set compiled in Jones (1984) and updated in this paper is a small data set and therefore has little data to describe the behavior of large foreshocks. To provide more data we searched the Caltech Southern California catalog back to 1933 and the U.C. Berkeley Northern California catalog back to 1950 for magnitude 5 or greater mainshocks in the San

Andreas physiographic province. We then removed events that were known to have thrust or normal faulting mechanisms and determined the magnitude of the largest foreshock, if any existed. The resulting catalog of 97 events is fraught with problems. First, it is likely to contain an unknown number of thrust fault events which will lower the observed foreshock rate (Jones, 1984). Second, these catalogs are certainly incomplete below magnitude 3 and may be incomplete below magnitude 4. This is difficult to determine because we expect that the time around the larger events got more attention than other time periods so that the completeness of the possible foreshocks may be greater than the overall catalog completeness.

Our contention is that this larger catalog can not be used to determine the overall rate of foreshock occurrence due to uncertainties about the focal mechanisms involved (Jones, 1984), but can be used to shed light on the form of the magnitude distribution of foreshocks. Figure 2 shows a cumulative number plot of the number of events versus the difference in the mainshock and foreshock magnitudes. The distribution used by Agnew and Jones should be a straight line on this plot and down to a magnitude difference of two this appears to be acceptable based on the admittedly sparse data set. We do note, however, that based on this sparse data set other distributions could be chosen. However, these other distributions may violate the Agnew and Jones (1991) data set that includes smaller magnitudes.

The question is should we use this generic distribution of foreshock behavior for strike-slip earthquakes in the San Andreas physiographic province for the Parkfield case. At Parkfield half of the past 4 mainshocks had foreshocks with $M \geq 5$, or within 1 unit of the mainshock's magnitude. Thus the observations at Parkfield suggest that this fault segment has more large foreshocks than other areas, but is this difference significant at the 95% confidence level?

Given that 17 out the 33 sequences have foreshocks the distribution in Agnew and Jones (1991) predicts that 15% of the sequences should have foreshocks within one unit of mainshocks's magnitude. This is because in any one unit of magnitude the fraction of sequences that have foreshocks should be N_m from Agnew and Jones (1991). If we remove the 1966 Parkfield mainshock from the data set of 33 sequences this number is reduced to 14.6%, an insignificant change. The probability that f or more of the n sequences had foreshocks even though the rate of foreshocks should only be p=0.15 can be determined through the binomial distribution:

$$\sum_{i=f}^{n} P_n^i = \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i}$$

The probability of getting 2 or more foreshocks in the 4 sequences is 0.11. Therefore the confidence that the Parkfield foreshock behavior is different than the average behavior is only 89%. Based on this analysis, we choose to apply the average rate of foreshock occurrence as determined by Agnew and Jones (1991) to the Parkfield case. If one chooses to apply some higher rate based on the Parkfield history the output alert probabilities will be higher, however they will not be statistically justifiable at the 95% level. Moreover, making a distribution of foreshock rate versus foreshock magnitude will be difficult based on the small data set. The rate of foreshocks affects the result in a similar manner to the long-term mainshock probability. Hence, as discussed above, doubling or halving the rate of foreshocks would approximately double or halve the alert probabilities respectively.

Background Rate

The final input needed to calculate the alert probabilities is the rate at which background events occur. Doing this requires choosing a spatial region in which earthquakes are considered part of the Parkfield background and foreshock process. The distribution of foreshocks from Agnew and Jones (1991) is based on the having the foreshock epicenter within 10km of the mainshock epicenter. For most earthquakes this requires a region that contains the area within 10km of the expected rupture area. However, for Parkfield the epicenters of the past two mainshock and foreshock hypocenters are well constrained to a small volume under Middle Mt (Cole and Ellsworth, 1995) If this behavior is expected to continue then a smaller volume could be used. For instance, if the mainshock hypocenter is expected to remain in the same spot then the region used could be a 10km radius circle centered on the mainshock epicenter. If the foreshock is also expected to remain in the same spot then an even smaller area could be used. When we do this we are assuming that based on the past observations the behavior of Parkfield foreshocks is different from other areas: that the rate of foreshocks may be the same as in other areas but that these foreshocks are located in a smaller region. This approach was used in OFR 87-192. While lower level alerts could occur due to earthquakes anywhere along the Parkfield segment of the San Andreas fault, the same magnitude earthquakes caused higher level alerts if they occurred near Middle Mt. Further, A and B level alerts could only occur due to events near Middle Mt.

The Middle Mt. Box used in OFR 97-192 was based on expecting the foreshocks to occur near the previous foreshocks but was enlarged to allow for possible location errors in both the previous foreshocks and the preliminary locations for current activity. A depth limit was placed to lower the background rate of earthquakes because the creeping segment of the San Andreas fault overlaps the Parkfield hypocenter at shallow depths. After the first A level alert in 1992 there was some discussion that the box extended too far southeast along the fault and that events in this location were not that worrisome (Figure 3). In addition our ability to rapidly and accurately locate events has exceeded expectations. Rarely do locations move more than 1 km from the first solution reviewed by a person. Thus, if one chooses to believe that future foreshocks will occur where they did in 1934 and 1966 (and these locations have been better constrained by Cole and Ellsworth, 1995) then a smaller alert box may be justified. This would increase the alert probabilities because it would decrease the background seismicity rate.

The assumption that the next Parkfield foreshock-mainshock sequence will resemble the last two sequences is worrisome. The Middle Mt. area is an interesting spot. Not only did it contain the previous two epicenters and foreshocks, but it has displayed unusually sensitivity to stresses applied by remote sources such as the 1983 Coalinga earthquake (Poley et al., 1987). This may be explained by the high pore pressures inferred by a combination of three-dimensional velocity models, gravity data, and electrical resistivity observations (Eberhart-Phillips and Michael, 1993) and a high Vp/Vs ratio (Michelini et al., 1991). These factors weigh in favor of considering the Middle Mt. area to be a special spot and limiting the foreshock region to that area.

While Middle Mt. has unusual characteristics and has been the initiation zone in past earthquakes, it may not be the initiation zone in the future. In the 1934 (Segall and Du, 1993) and 1966 (Segall and Harris, 1987) mainshocks most of the moment was released from an area 8 to 25 km southeast of Middle Mt (Figure 3). This is the same area that Segall and Harris (1987) demonstrated was storing strain that could be released in a future event. It seems reasonable that rupture in this area could be triggered by a propagating rupture that did not begin under Middle Mt., but somewhere else along the edge of the patch of stored strain. The mainshock catalog includes six events. Of these we know that the last two of them nucleated under Middle Mt. and that the 1922 event nucleated either there or to the northwest of Middle Mt. For the previous three events we do not know where the hypocenter was.

We simply do not have enough information about the complete earthquake history, the rupture process, the material properties along the fault, or the state of stress on the fault to come to a firm conclusion about where the next sequence will initiate. Given this uncertain state of affairs we prefer to equally consider the consider the possibility of foreshocks in the the larger area used by Agnew and Jones (1991). OFR87-192 did allow for C and D level alerts based on events outside the Middle Mt. box, however this appears to have been done in an approximate manner and not by using additional calculations for the larger areas considered. Using the larger area will slightly reduce the probabilities associated with possible foreshocks, but it will be a small effect because most of the background seismicity occurs near Middle Mt. This is especially true of the larger magnitude events. In the rest of the paper we will compute the background rates and probabilities for the Parkfield box and the two Middle Mt. boxes to illustrate the effect of making this choice and to make comparisons between the new method and that of OFR 87-192.

To compute the background seismicity rate the Northern California Seismographic Network catalog for the years 1982 through February, 1995 was declustered by keeping only the largest events in ±3 day windows. By declustering the catalog we prevent the aftershock process from increasing the seismicity rate over the actual rate of independent background events. Agnew and Jones (1991) used the declustering method of Reasenberg (1985) but the simpler method chosen here is easier to apply in real time. If another larger event occurs before the three day window is over, the current alert will either be extended at the current level or moved to a higher alert level. This does mean that when seismicity increases in magnitude over time alerts will be declared based on events that would not be in the declustered catalog. However, these alerts will generally be low level ones.

We could have declustered the catalog by only removing events that had a larger one in the three days before them, but then the background catalog would not correspond well to the chosen foreshock distribution. The foreshock distribution with respect to magnitude is based on only the largest event in the three days before the observed mainshocks. Events that were followed by a larger one are not included in the foreshock distribution. We feel it is more important for the background distribution to correspond to the foreshock distribution than to avoid slightly underpredicting the frequency, and slightly overpredicting the probability, of the lower level alerts. We tested the affect of making this choice on the alert probabilities and it is on the order of 10%. This is small compared to the other possible errors in the data and assumptions.

Distributions of the declustered seismicity rate versus magnitude are shown in Figure 4 for three possible areas: our preferred Parkfield box, the original Middle Mt. box, and a possible smaller Middle Mt. box. The definitions of these boxes are shown in Figure 1 and Table 1. The Small Middle Mt. box was defined by having an area that was ±2 km from the epicenter of the 1966 foreshock. For the two larger boxes the linear fits to the distributions were determined by using a maximum likelihood method (Aki, 1965).

For the Small Middle Mt. box the observed distribution departs too far from a straight line to obtain a satisfactory fit by this objective method. To fit this distribution we first constrained the activity at the largest magnitude by using the longer term catalog compiled for $M \geq 3.7$ events by Cole and Ellsworth (1995). In the Small Middle Mt. box, since 1930, they observed the 1934 and 1966 foreshocks, mainshocks, and 3 previous repetitions of the Nov. 14, 1993 M=4.8 event on June 5, 1934; Dec. 28, 1939, and Nov. 16, 1959. The last three occurrences of this event are in the background classification and were the only ones used to measure the background rate over this time period. This gives a rate of occurrence of 0.046 M=4.8 events per year in the Small Middle Mt. box. We constrained the linear fit to intersect this point and then tested slopes from b=0.3 to b=0.8. A slope of b=0.5 fits the data decently, especially at the higher magnitude levels where the most important alerts will occur.

The difficulty fitting a straight line to the data in the Small Middle Mt. box illustrates another problem with using such a small volume for the alert system. The methodology of Agnew and Jones (1991) is based on applying the linear fit to the cumulative magnitude distribution of the background seismicity. With such a small box this is difficult and leads to higher uncertainties in the alert probabilities.

Single Event Probabilities and Alert Levels

Given the assumptions discussed above we can compute the probability that an observed event is a foreshock and that therefore the mainshock will follow within the next three days. These values are computed for the three alert zones we considered (Figure 5). Also shown in Figure 5 are the values computed by the method of OFR 87-192, that assume that the long term mainshock probability is 15% per year. These values are useful to compare the results of the new method to the scenario document currently in use.

For the Middle Mt. box which was studied using both methods the new methodology gives probabilities that are lower than OFR 87-192 by a factor of 2.5 to 5 with the biggest difference for events between M4 and M5. Thus the difference is large in the most important magnitude range for public policy considerations. If the Parkfield box is used for the new methodology then the differences are as large as a factor of 14, with differences between M4 and M5 about a factor of 8 to 10. This comparison is between our preferred application of the new methodology and the existing scenario document, but is not a fair comparison between the old and new methods.

To directly compare the old and new methods we also determined the results using the method of OFR 87-192 for the Middle Mt. box but with the long term mainshock probability reduced to 10% per year. The modified OFR 87-192 results are a factor of 1.7 to 3.7 greater than the new results for the Middle Mt. box. As before this difference is largest in the range from M4 to M5.

The values determined in this report are lower than those in OFR 87-192 for a few reasons: the lower long-term probability of the mainshock occurring used here, the lower rate of foreshocks assumed for each mainshock, and the new methodology used. Another factor is that since October, 1992 the Parkfield area has increased in activity, especially at the higher magnitude levels. Including this data into our estimate of the background seismicity rates has increased these rates and therefore decreased the probability that any event is a foreshock.

Discussion

The probability of the mainshock occurring in a 3 day window associated with an A level alert in OFR 87-192 is at least 37%. Under the new formulation this level can not be reached in either the Parkfield or Middle Mt. boxes. For the small Middle Mt. box it would require a M5.7 event, large enough that it could fit within our definition of the mainshock if it produces ground rupture. The lowest magnitude that could trigger an A level alert under OFR 87-192 is 4.5. At this magnitude the new formulation would give a probability of only 4% for the Parkfield box, 7% for the Middle Mt. box, and 13% for the Small Middle Mt. box. If we continue to associated M4.5 events with A level alerts then the meaning of the alert levels has greatly changed. Hence, we do not suggest keeping the same alert rules with respect to magnitude and only changing the associated probabilities. To do so would destroy the communication value of the words "A level alert." This should be reserved for cases where the associated probabilities are high enough to warrant action on the part of groups concerned with earthquake response.

One possibility is to use the structure set up for Southern California (Jones et al., 1991). There D level alerts occur when the probability of a mainshock occurring in the next 3 days reaches 0.1%. Higher level alerts include C level at 1% and B level at 5%. The B level is stated to extend from 5% to 25% and they did not implement the A level that would occur above 25%. For Parkfield we will use A level when the probability exceeds 25%. For the three boxes this would result in the following alert criteria for single foreshocks:

Alert level	Mainshock Probability	Magnitude Required for Parkfield	Middle Mt.	Small Middle Mt.
A	25%	5.9	5.8	5.0
B	5%	4.7	4.2	3.3
C	1%	3.6	2.8	2.0
D	0.1%	1.9	0.8	0.0

Two points should be noted about this alert structure. First, it is only possible to reach an A level alert if the Small Middle Mt. box is used. However, we emphasize that it is our opinion that using such a small box may miss possible foreshocks if events do not evolve exactly as they did in 1934 and 1966 (which is why OFR87-192 included C and D level alerts for a larger Parkfield box). And even for the Small Middle Mt. box a repeat of the 1934 or 1966 foreshock would only be an A level alert if the preliminary magnitude is not underestimated. We also note that the threshold to get to a D level alert is lower than the USGS Northern California Seismic Networks completion level for the Middle Mt. box and below its detection threshold for the Small Middle Mt. box. This presents an operational problem.

A key question is how much are we willing to bet on the next sequence containing an almost exact repeat of the 1934 and 1966 M5 foreshocks which occurred 17 minutes before the mainshock. W. Ellsworth (pers. comm., 1994) has suggested an alert scheme that would require a repeat of this foreshock, as determined through waveform cross-correlation, to go to an A level alert. This event has only been observed twice and both times it was a foreshock, hence it is impossible to assess the background rate for this event and therefore it is not meaningful to assess the alert probability associated with it. If this event is never a background earthquake and always a foreshock then the alert probability should be 100%, although our confidence in this value should be low because it is based on only two observations. Perhaps to implement such an alert scheme requires developing a much deeper physical understanding of the foreshock process than now exists. Until that happens we must rely on statistical analyses. Given that there are only a three decently recorded repeats of the Parkfield earthquake sequence, and the mainshock catalog includes three events that we know very little about, we prefer to avoid making specific assumptions based on the 1934 and 1966 sequences. Instead we suggest using the larger Parkfield box as the new alert boundaries for seismicity.

The magnitudes in the table shown above are the minimum magnitudes required to exceed a given probability. For specific alerts we suggest giving the letter designation and the probability associated with the specific event as shown in Figure 5. This has the advantage that if the magnitude of an event changes by a small amount such that it goes from one alert level to another the letter may change abruptly but the probability announced will change only by a small amount.

Instead of using a single alert box, one could use multiple boxes in the alert scheme; for example the Small Middle Mt. box and the Parkfield box. To properly use these two boxes you have to assign a probability, called q, that the foreshock, if one occurs, will occur in the Small Middle Mt. box. Then the probability of a foreshock in the Small Middle Mt. box is the product of q and the generic foreshock distribution. The alert probabilities would have to be recomputed for the Small Middle Mt box based on this new foreshock distribution and will be lower than shown in Figure 5 and above. Then the remaining chance that a foreshock will occur, $1-q$, is assigned to the larger Parkfield box which will also now have a foreshock distribution lower than the generic one and therefore the alert probabilities will also decrease. However, the background rate for the larger box must be decreased by the rate of earthquakes in the Small Middle Mt. box. This latter effect will increase the alert probabilities by some amount that depends on the relative rates of activity between the two boxes.

While we are suggesting that multiple alert boxes could be used and a way of properly doing these computations we have chosen not to show an example because we do not know how to set the probability q or which boxes should be used. One could even extend this method to use all three. Ideally these choices should be made independent of the final results. We are therefore reluctant to show an example.

Conclusions

We have reevaluated the probability that observed earthquakes at Parkfield will be foreshocks to the Parkfield mainshock. This reevaluation includes reducing the long-term probability of the mainshock occurring, decreasing the rate at which foreshocks are assumed to precede Parkfield mainshocks, and using the methodology of Angew and Jones (1991) to correct problems associated with cumulative magnitude distributions. Further, we suggest using the Parkfield box as the alert area in which foreshocks are assumed to occur because of uncertainties in how the first three Parkfield earthquakes sequences occurred and how the next Parkfield earthquake sequence will unfold.

The alert probabilities computed with this new formulation are lower than those in OFR 87-192 and currently in use. This requires a new alert structure for seismicity. For uniformity, we suggest using one similar to that adopted for southern California. Here the probability that any earthquake will be followed by the mainshock is calculated and an alert is then declared if the probability exceeds a given value (25% for A, 5% for B, 1% for C, and 0.1% for D).

Acknowledgements

This paper has benefited from reviews and comments from Duncan Agnew, Bill Bakun, Bill Ellsworth, Dave Jackson, Allan Lindh, and Paul Reasenberg.

-- --

REFERENCES

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

Aki, K., Maximum-likelihood estimate of b in the formula $N = a\,10^{-bM}$ and its confidence limits, *Bull. Earth. Res. Inst.,, 45,* 237-239, 1965.

Bakun, W., History of significant earthquakes in the Parkfield area, *Earthquakes and Volcanoes, 20,* 45-51, 1988.

Bakun, W.H., and A.G. Lindh, The Parkfield, California, earthquake prediction experiment, *Science, 229,* 619-624, 1985.

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. Schultz, P. Segall, and W. Thatcher, Parkfield, California, Earthquake prediction scenarios and response plans, *U.S. Geol. Surv., Open-File Report, 87-192,* 45 pp, 1987.

Cole, A.T., and W.L. Ellsworth, Earthquake clustering and the long-term evolution of seismicity near Parkfield, California, 1931-1994 (abstract), *Seism. Res. Lett.,* 28, 1995.

Eberhart-Phillips, D., and A.J. Michael, Three-dimensional velocity structure, seismicity, and fault structure in the Parkfield region, central California, *J. Geophys. Res., 98,* 15,737-15-758, 1993.

Jones, L.M. et al., Short term earthquake hazards assessment for the southern San Andreas fault, *U.S. Geol. Surv., Open-File Report, 91-32,* 1991.

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

Jones, L.M., and A. J. Michael, A re-evaluation of the Parkfield alert level probabilities (abstract), *Eos Trans. AGU, 75,* 469, 1994.

Lindh, A.G., Preliminary assessment of long-term probabilities for large earthquakes along selected fault segments of the San Andreas System in California, *U.S. Geol. Surv., Open-File Report, 83-63,* 15 pp, 1983.

Michelini, A., and T.V. McEvilly, Seismological studies at Parkfield, I, Simultaneous inversion for velocity structure and hypocenters using cubic b-splines parameterization, *Bull. Seismol. Soc. Am., 81,* 524-552, 1991.

NEPEC Working Group to Evaluate the Parkfield Prediction Experiment, Earthquake Research at Parkfield - 1993 and Beyond, *U.S. Geol. Surv., Open-File Report, 93-622,* 19 pp, 1993.

Poley, C.M., A.G. Lindh, W.H. Bakun, and S.S. Schulz, Temporal changes in microseismicity and creep near Parkfield, California, *Nature, 327,* 134-137, 1987.

Reasenberg, P., Second-order moment of Central California seismicity, 1969-1982, *J. Geophys. Res., 90,* 5479-5496, 1985.

Savage, J.C., The Parkfield prediction fallacy, *Bull. Seismol. Soc. Am., 83,* 1-6, 1993.

Segall, P., and Du T.V. McEvilly, Jr., How similar were the 1934 and 1966 Parkfield earthquakes?, *J. Geophys. Res., 98,* 4527-4538, 1993.

Segall, P., and R. Harris, Earthquake deformation cycle on the San Andreas fault near Parkfield, California, *J. Geophys. Res., 92,* 10,511-10,525, 1987.

Zoback, M.L., and M.D. Zoback, State of stress in the conterminous United States, *J. Geophys. Res., 85,* 6113-6156, 1980.

Table 1

Vertices of Regions

Region	Minimum Depth (km)	Latitude (N)	Longitude (W)
San Andreas Fault		35 58.26	120 35.86
(mainshock definition)		36 01.74	120 30.74
		35 47.25	120 15.72
		35 43.77	120 20.82
Parkfield	0	35 56.53	120 38.41
		36 03.47	120 28.19
		35 48.98	120 13.17
		35 42.04	120 23.37
Middle Mountain	6.5	36 01.5	120 29.5
		35 57.0	120 25.0
		35 52.0	120 31.5
		35 58.0	120 38.0
Small Middle Mountain	7.5	35 56.07	120 30.42
		35 57.45	120 28.38
		35 59.11	120 30.10
		35 57.73	120 32.14

Table 2

$M \geq 5$ Strike-Slip Mainshocks, 1981-1994, in the San Andreas Physiographic Province and their Foreshock Behavior

NUM	DATE	ORIGIN	LAT	LONG	MAG	FORE
1	810426	1209 47.02	33 05.91	115 37.90	5.7	4.1
2	810904	1550 50.13	33 39.09	119 05.58	5.5	
3	840123	0540 00.00	36 21.19	121 54.51	5.2	
4	840424	2115 00.00	37 18.81	121 39.39	6.2	2.3
5	860126	1920 00.00	36 48.56	121 17.29	5.7	
6	860331	1155 00.00	37 28.02	121 41.52	5.5	2.6
7	861121	2333 00.00	40 21.30	124 25.63	5.1	2.3
8	870207	0345 14.85	32 23.28	115 18.27	5.4	3.0
9	870731	2356 00.00	40 24.52	124 24.43	5.5	3.2
10	871124	1315 56.71	33 00.87	115 51.10	6.6	6.2
11	880220	0839 00.00	36 47.68	121 18.65	5.1	
12	881203	1138 26.44	34 09.06	118 07.81	5.0	
13	900116	2008 00.00	40 14.63	124 23.04	5.4	
14	900228	2343 36.75	34 08.62	117 41.84	5.4	3.7
15	920423	0450 23.22	33 57.67	116 19.05	6.1	4.6
16	920628	1157 34.13	34 12.01	116 26.20	7.3	3.6
17	920711	1814 16.15	35 12.60	118 03.94	5.7	

NUM is the number of earthquake as referred to in text. DATE is in the form yymmdd; where yy=year, mm=month, and dd=date. ORIGIN is the origin time in the form hhmm sec; where hh=hour, mm=minute, sec=seconds in GMT. LAT is the epicentral latitude in degrees and minutes north. LONG is the epicentral longitude in degrees and minutes west. MAG is the mainshock magnitude. FORE is the magnitude of a foreshocks if there was a $M \geq 2$ foreshock in the 3 days before the listed mainshock and within 10 km of the mainshock hypocenter.

Figure Captions

Figure 1. Map showing the background seismicity at Parkfield from 1982 through 1994, three possible boxes on which to base the alert structure, and a box defined to contain the predicted mainshock hypocenter.

Figure 2. The cumulative number of foreshocks with respect to increasing magnitude difference between the mainshock and the foreshock. The data set used are 38 sequences with $M \geq 5$ mainshocks and that had foreshocks. This data set includes events in the San Andreas physiographic province from the Caltech-USGS catalog from 1933 to 1994, the U.C. Berkeley catalog from 1950-1994, the U.S.G.S. Northern California catalog from 1969-1994, Jones (1984), and the compilation shown in Table 2. Sequences with known dip-slip mainshock mechanisms were removed.

Figure 3. Along fault cross-section showing seismicity from 1982 (dots), on highlighting the 1966 mainshock and foreshock (stars), the $M \geq 3$ events since 1992 (boxes), and the regions of primary moment release in the 1934 and 1966 mainshocks (hatched areas).

Figure 4. Cumulative seismicity plots versus magnitude for the three alert boxes considered for the time period from 1982 through February, 1995.

Figure 5. Probabilities of the mainshock occurring within the 3 days after a single possible foreshock occurs. Curves are shown for the three alert boxes considered, the results of OFR 87-192, and the method of OFR 87-192 modified as explained in the text.