

**REPRODUCED FROM BEST AVAILABLE COPY**

**U.S. Department of the Interior  
U.S. Geological Survey**

**Short-Term Earthquake Hazard Assessment  
for the San Andreas Fault in Southern California**

**Working Group:**

**Chairmen: Lucile M. Jones<sup>1</sup> and Kerry E. Sieh<sup>2</sup>**

**Members:**

**Duncan Agnew<sup>3</sup>, Clarence Allen<sup>2</sup>, Roger Bilham<sup>4</sup>,  
Mark Ghilarducci<sup>5</sup>, Bradford Hager<sup>2,6</sup>, Egill Hauksson<sup>7,8</sup>,  
Kenneth Hudnut<sup>9,8</sup>, David Jackson<sup>10</sup>, and Arthur Sylvester<sup>11</sup>**

**Corresponding Members:**

**Keiti Aki<sup>7</sup> and Frank Wyatt<sup>3</sup>**

**Open-file Report 91-32**

**This report is preliminary and has not been reviewed for conformity with U. S. Geological Survey editorial standards or with the North American Stratigraphic Code. Any use of trade, product or firm names is for descriptive purposes only and does not imply endorsement by the U. S. Government.**

---

<sup>1</sup> U. S. Geological Survey, 525 S. Wilson Avenue, Pasadena, CA 91106

<sup>2</sup> Division Geological and Planetary Sciences, Calif. Inst. of Tech., Pasadena, CA 91125

<sup>3</sup> Inst. for Geophysics and Space Physics, Univ. of California, La Jolla, CA 92092

<sup>4</sup> Dept. of Geological Sciences, Univ. of Colorado, Boulder, CO 80309

<sup>5</sup> Calif. Governor's Office Emergency Services, 2151 E. D St., Ontario, CA 91764-4452

<sup>6</sup> now at Dept. Earth, Atmos. and Plan. Sci., Mass. Inst. of Tech., Cambridge, MA 02139

<sup>7</sup> Dept. of Geological Sci., Univ. of Southern California, Los Angeles, CA 90089-0740

<sup>8</sup> now at Division Geol. and Planet. Sci., Calif. Inst. of Tech., Pasadena, CA 91125

<sup>9</sup> Lamont-Doherty Geological Observatory of Columbia Univ., Palisades, NY 10964

<sup>10</sup> Dept. of Earth and Space Sciences, Univ. of California, Los Angeles, CA 90024

<sup>11</sup> Dept. of Geological Sciences, Univ. of California, Santa Barbara, CA 93106

## Table of Contents

Preface	
Letter from the chairman of NEPEC .....	i
Letter from the chairman of CEPEC .....	ii
Executive Summary .....	1
I. Introduction .....	3
II. Short-term Earthquake Hazard Assessment .....	4
III. Possible Earthquake Precursors .....	5
III.1 Summary of Current Instrumentation.....	5
III.2 Foreshocks.....	7
III.2.1 Theory .....	7
III.2.2 Hazard Levels for Foreshocks .....	10
TABLE 1. Magnitudes of Potential Foreshocks.....	10
TABLE 2. Alternate Solution Using Davis et al. (1989) .....	11
III.3 Aseismic Fault Slip .....	11
III.3.1 Steady State Creep.....	12
III.3.2 Triggered Creep.....	12
III.3.3 Evidence for Premonitory Creep on California Faults .....	12
III.3.4 Hazard Levels for Surficial Creep .....	13
III.4. Strain .....	14
III.4.1 Available Data.....	14
III.4.2 Criteria for Strain .....	14
III.4.3 Hazard Levels for Strain.....	15
III.5 Combined Hazard Levels .....	15
IV. Response Plan for the USGS .....	16
V. Need for Improved Instrumentation.....	17
V.1 Centralized Recording and Analysis .....	18
V.2 Seismological Data.....	18
V.3 Strain and Creep Data.....	19
V.4 Fundamental Understanding of the Southern San Andreas Fault .....	20
References.....	22
Appendix A .....	24

November 30, 1990

Dr. Dallas Peck, Director  
U. S. Geological Survey  
National Center  
12201 Sunrise Valley Drive  
Reston, VA 22092

Dear Dr. Peck,

This proposed plan for the southern San Andreas fault, with ongoing earthquake hazard assessment and communication of any inferred increases in hazard, has been recommended by the National Earthquake Prediction Evaluation Council (NEPEC) for implementation by the U.S.G.S.. Modeled in its general structure of alert levels and response scenarios after the system in place for the Parkfield Prediction Experiment (Bakun *et al.*, U.S.G.S. Open-file Report 87-192, 1987), this plan relies for decision making on alert levels largely on the past record of foreshock occurrences throughout California. The two highest levels (C and B, in a D, C, B level range) are attainable only by the occurrence of foreshocks. Other observations of deformation can produce only the lowest D-level alert (probability of 0.1-1% for a M7.5 mainshock in 72 hours).

Such a formal assessment and communication procedure is important to have in place for southern California in advance of the more significant (M5+) potential foreshocks or other anomalous phenomena, in order to preclude inconsistent announcements to the public. The system has been effective in this way at Parkfield, and this proposed plan is appropriate and timely for implementation on the southern San Andreas fault.

Thomas V. McEvilly  
Chairman, NEPEC




GOVERNOR'S OFFICE OF EMERGENCY SERVICES  
OFFICE OF EARTHQUAKE PROGRAMS  
2151 E. D. ST., SUITE 203A  
ONTARIO, CA 91764  
714-391-4485 FAX 714-391-3984



November 21, 1990

TO: BILL BAKUN  
USGS, MENLO PARK

FROM: RICHARD ANDREWS   
CHIEF DEPUTY DIRECTOR

SUBJECT: WORKING GROUP REPORT ON SOUTHERN SAN ANDREAS

The attached letter from Jim Davis summarizes the conclusions of CEPEC regarding the referenced document.

I concur with Davis' conclusions about the report and its release.

cc: J. Davis  
L. Jones, USGS, Pasadena 

# Memorandum

To : Richard Andrews, Deputy Director  
Governor's Office of Emergency Services

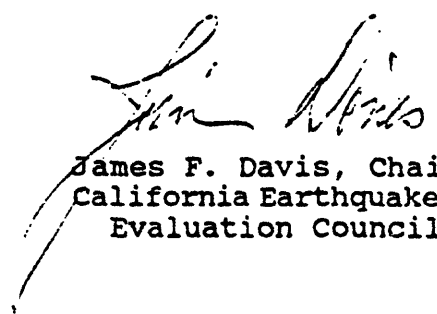
Date: Nov. 14, 1990

From : James F. Davis

Department of Conservation  
Division of Mines and Geology  
1416-9th Street, Sacramento 95814

Subject: CEPEC evaluation of the USGS Working Group Open-File Report  
entitled, Short-Term Earthquake Hazard Assessment for the Southern  
San Andreas Fault

At its meeting on August 29, 1990, CEPEC considered the Working Group report on the southern San Andreas fault. The Council heard a presentation on the report from Working Group member Duncan Agnew. CEPEC requested that several changes be made in the report which would avoid any confusion regarding the distinction between actions taken at certain alert levels by the USGS of a scientific nature and those of the State of a public safety nature. The Working Group agreed to this. A revised version has been reviewed by CEPEC and we are recommending that this report now be accepted by OES as the basis for OES communicating with the USGS and the public when events above the threshold magnitude levels occur along the southern San Andreas fault. We suggest that if you concur, you advise the USGS so that they can release the report. Please let me know if you have any questions.

  
James F. Davis, Chair  
California Earthquake Prediction  
Evaluation Council (CEPEC)

## Executive Summary

The historically dormant southernmost 200 km of the San Andreas fault (from Cajon Pass, northwest of San Bernardino, southeast to Bombay Beach on the Salton Sea) is the segment most likely to produce an earthquake of magnitude 7.5 or greater within the near future. Such an earthquake would cause widespread damage in San Bernardino, Imperial, Riverside, Orange and Los Angeles counties, which together have over 12 million inhabitants. If anomalous earthquake or other geophysical activity were to occur near the southern San Andreas fault, scientists would be expected to advise government officials on the likelihood that a major earthquake is forthcoming. The primary purpose of this report is to present a system for quantifying and communicating information about short term increases in the earthquake hazard from the southern San Andreas fault.

The system we have adopted follows that used for the Parkfield earthquake prediction experiment in central California. It includes several levels, each corresponding to a different range of estimated short-term hazard. The responses of the U. S. Geological Survey (USGS) will be similar to those defined for the Parkfield experiment. The probabilities that the predicted earthquake will occur within the 72 hours of the estimate being made are comparable to the probabilities defined for the alerts at Parkfield, but the criteria for reaching each level necessarily differ from those at Parkfield. The Working Group felt that there was presently no way to produce meaningful estimates of probabilities above 25%, but reserved an additional level (A) to allow for this becoming possible in the future. For now, this level is unattainable. The defined levels are:

Level	Probability of M>7.5 earthquake in next 72 hours	Expected time between occurrences	USGS action
D	0.1% to 1%	6 months	Notify scientists involved in data collection and OES Ontario office
C	1% to 5%	5 years	As for level D, and also notify Comm. Officer of OES in Sacramento, and OEVE chief.
B	5% to 25%	28 years	As for levels C and D, and also notify USGS Director, and CDMG State Geologist, and start intensive monitoring.

The different levels can be reached because of earthquakes, creep events (rapid aseismic surficial slip on faults) and strain events (anomalous deformation of the crust).

Our hazard estimates are based primarily upon the observation that half of magnitude 5.0 or greater strike-slip earthquakes in California have been preceded by immediate foreshocks (defined as earthquakes within 3 days and 10 km of the mainshock). Therefore, the next major earthquake produced by the southern San Andreas could well be preceded by one or more foreshocks. This report describes a method for estimating the probability of the next major earthquake, given the occurrence of a possible foreshock. To be considered a possible foreshock, the rupture zone of the earthquake must come within 10 km of the southern San Andreas fault. The table below gives the magnitude of possible foreshock needed to reach a specified probability level for four microseismic regions of the southern San Andreas fault.

Level	B	C	D
Probability of M7.5 in 72 hr	5-25%	1-5%	0.1-1%
San Bernardino	5.8	5.0	3.9
San Geronio	6.1	5.3	4.2
Palm Springs	5.2	4.5	3.4
Mecca Hills	4.9	4.2	3.1

Anomalous creep and strain episodes are also possible precursors to the next major earthquake along the southern 200 km of the San Andreas fault. Exact probabilities cannot be calculated for these possible strain precursors, because the data are inadequate to quantify the relationship between precursory slip or strain and large earthquakes. Moreover, unlike Parkfield (where several types of strain and creep meters are densely arrayed along the fault), only one strainmeter and four creepmeters are deployed near the southern San Andreas fault. Therefore, only one hazard level is defined for strain and aseismic slip; this is arbitrarily set equal to the lowest seismic level (D). This level is reached if an amount of aseismic slip or strain occurs that is unprecedented in the history of recording along the southern San Andreas fault.

The reliability of any estimate of short-term hazard is limited by inadequacies in the data now being recorded along the southern San Andreas fault. For example, continuous measurements of ground deformation are limited to one strainmeter and four creepmeters. Because seismic stations are sparsely distributed and the automatic processing rudimentary, the depth and rupture size of most earthquake sources cannot be resolved, earthquakes above about magnitude 3.5 are not recorded on scale, and their spectral characteristics cannot be determined properly. Furthermore, the available data are not all recorded in one place. Therefore, this report recommends improvements in data management, instrumentation, and research that would increase the ability of scientists to advise on the short-term likelihood of a great southern California earthquake. We should:

*Implement centralized recording and analysis.* A chief scientist for the southern San Andreas fault should be appointed and supported by the chief of the Office of Earthquakes, Volcanoes and Engineering (USGS) with the authority to undertake the actions described here. Deformation data now available from southern California should be given in real time to the Pasadena office of the USGS and Caltech to be evaluated together with the seismic data. Such evaluation will be an assigned task of the Pasadena office of the USGS and the Seismological Laboratory.

*Improve seismic data.* Expand the real-time earthquake analysis system to cover all the existing seismic network, add procedures for quickly estimating the magnitudes of large earthquakes, and improve the quality and quantity of seismic stations along the southern San Andreas fault.

*Improve creep and strain data.* An increased number of telemetered creepmeters along the southern San Andreas fault and auxiliary faults would enhance the evaluation of possible precursors. Additional deformation measurements would also be desirable, but will require careful planning. We suggest that a group of university and USGS scientists develop such a plan.

*Improve our fundamental understanding of the fault.* Better data would improve our ability to estimate short-term probabilities, as would a better understanding of the behavior of the fault. We therefore recommend that additional geodetic, paleoseismic, and seismologic research be undertaken to better understand the nature of the fault zone.

## I. Introduction

The southernmost 200 km of the San Andreas fault in California, from Cajon Pass southeast to Bombay Beach on the Salton Sea (Figure 1), has not produced a major earthquake within the historic record. Both geodetic evidence of continuing strain accumulation (Savage *et al.*, 1986) and the occurrence of recent prehistoric large earthquakes (Sieh, 1986; Sieh and Williams, 1990), however, lead us to conclude that this fault segment will eventually produce great earthquakes that pose one of the greatest hazards to southern California. An estimated 1.0-1.5 million people now live adjacent to the San Andreas fault within the projected zone of severe shaking for such an earthquake. A magnitude 7.5 to 8.0 earthquake on this segment would also cause widespread damage to San Bernardino, Imperial, Riverside, Orange, and Los Angeles counties, which together have over 12 million inhabitants. For these reasons, the Southern San Andreas Fault Working Group was formed in 1989 to recommend how the scientific community might best respond to anomalous geophysical activity along the fault, increase our understanding of regional seismotectonics, and offer timely scientific advice to state and local governments.

The southernmost 100 km of the San Andreas fault, the Coachella Valley segment from the Salton Sea to San Geronimo Mountain, was identified by the Working Group on California Earthquake Probabilities (WGCEP, 1988) as the segment of the San Andreas fault zone most likely to produce a major earthquake of magnitude 7.5 or greater within the near future. That group estimated the conditional probability of such an event to be 40% within the next 30 years. The latest large earthquake on the Coachella Valley segment of the San Andreas fault occurred about 300 years ago (Sieh, 1986; Sieh and Williams, 1990), and it is both realistic and prudent to assume that the next large event there will occur within our lifetimes.

The Coachella Valley segment abuts the San Bernardino segment which extends from the southern San Bernardino Mountains to Cajon Pass (Figure 1). The geologic record of earthquakes for the San Bernardino segment is more poorly understood than that of the Coachella Valley and the time of the last earthquake on that segment is not known. For this reason, the WGCEP (1988) considered the San Bernardino segment separately from the Coachella Valley segment and assigned it a 30-year probability of 20%. However, it is not known, at present, how much of the southern San Andreas fault will be involved in the next great earthquake. The present Working Group thought it possible or even likely that faulting in the next earthquake in the Coachella Valley will extend at least through the San Bernardino segment (over 200 km) producing a magnitude 7.5-8 earthquake and could continue to rupture through to the northwest into the Mojave segment (over 350 km) with a magnitude 8 or greater earthquake. Because of uncertainty about the final length of the next great earthquake, the section of the fault to be considered in this study was at the discretion of the Working Group. We chose to include only those sections of the fault that have not slipped in the historic record and thus excluded the Mojave segment. The region considered includes the Coachella Valley and San Bernardino segments as defined by WGCEP (1988) and extends from the Salton Sea to Cajon Pass, a distance of 210 km.

Moderate earthquakes and creep events have been recorded over the last fifty years on the southern San Andreas fault and will be again. When that happens, seismologists will be expected to advise state and local officials about the potential for further activity on the fault. In particular, they will be asked if the activity could be a precursor to the "Big One." It seems prudent to consider the most likely scenarios for such "earthquake crises" in advance, so that we can, with time available for careful evaluation, agree on appropriate answers to such questions. While experiences in public safety situations elsewhere have shown that scenario and response plan exercises often do not anticipate the details of subsequent events, they lead to more rapid and rational responses; conversely, lack of planning can be a recipe for fiasco. Thus, the primary goal of the Southern San Andreas Fault Working Group is to develop a system for quantifying and communicating information about short term increases in the earthquake hazard from the southern San Andreas fault.



# Southern San Andreas Fault

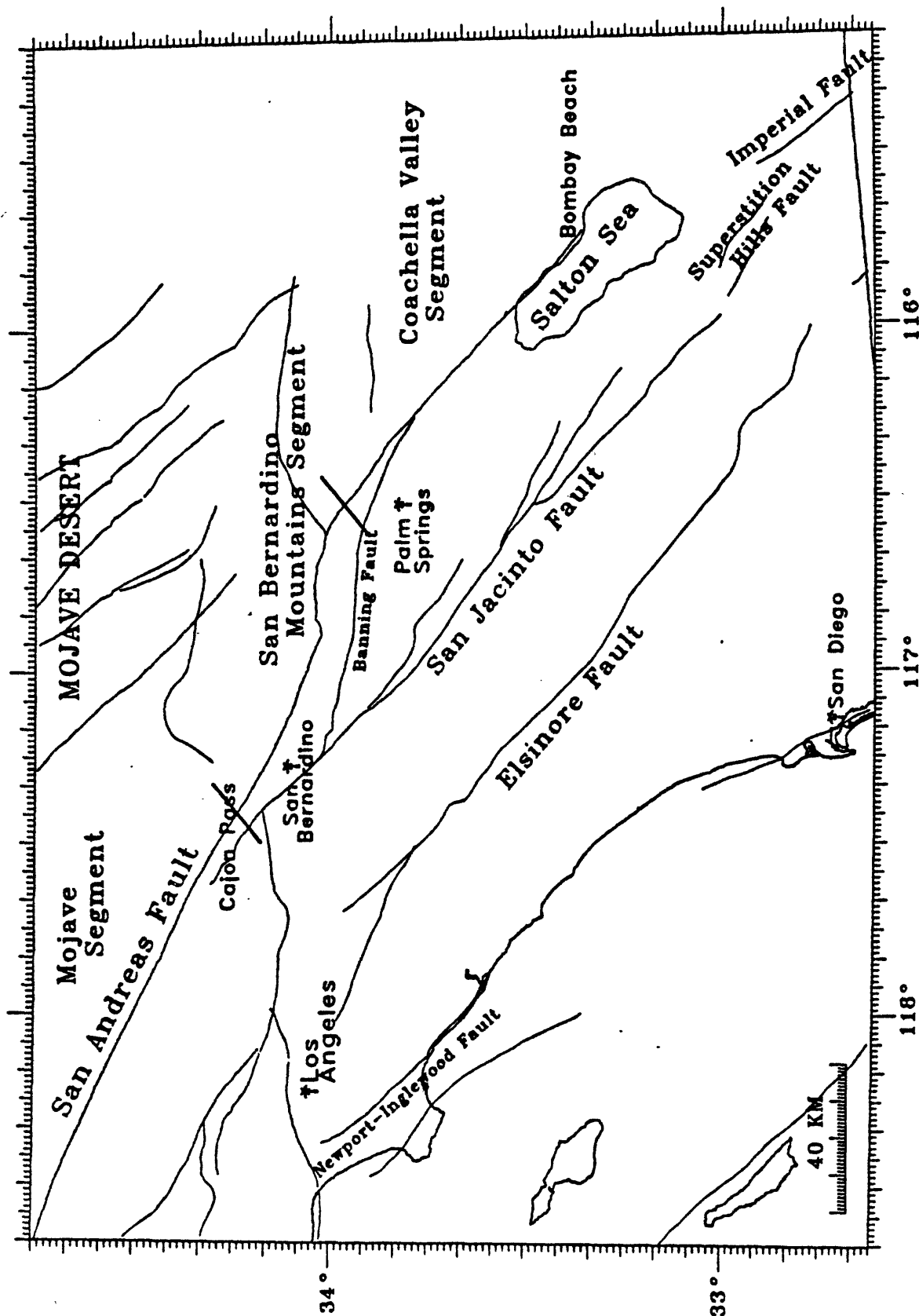


Figure 1. A map of the southernmost segment of the San Andreas fault in California, from Cajon Pass southeast to near Bombay Beach on the Salton Sea, showing locations referred to in the text. Towns are shown by palm trees. The Coachella Valley and San Bernardino Mountains segments of the San Andreas fault as defined by WGCEP (1988) are indicated. The Mecca Hills and Palm Springs microseismic regions lie within the Coachella Valley segment and the San Geronimo and San Bernardino regions lie within the San Bernardino Mountains segment of the WGCEP.

A system for short-term warnings was developed for the Parkfield segment of the central San Andreas fault (Bakun *et al.*, 1987). At Parkfield, magnitude 6 earthquakes have occurred every 22 years on average, with the last one in 1966, making that section most likely to produce a moderate earthquake within the next decade (Bakun and Lindh, 1985). Few people are at risk from that earthquake, but the greater chance of having an earthquake within a limited period of time makes Parkfield an ideal site for experiments in prediction. The U. S. Geological Survey has installed many instruments at Parkfield in an attempt to issue a short-term warning for the next Parkfield earthquake. An alert system has also been established for quantifying and communicating hazard information to the state of California (Bakun *et al.*, 1987). The Parkfield system provides a prototype for developing an alert system for the southern San Andreas fault.

In devising this system, it became clear to the Working Group that, along the southern San Andreas fault, the quality of the data now recorded is very poor, both for the immediate purpose of making short-term hazard assessments and for the longer-term goal of improving our ability to do so. Members of the Working Group unanimously agreed that improved instruments and data management would increase the chance that a useful warning could be issued before the next great earthquake. The Working Group therefore decided to recommend specific improvements to the instrumentation, data management, and research effort in southern California. These proposed improvements are aimed at significantly increasing our ability to recognize and understand changes in the physical properties of the fault that might precede a great earthquake. The improved instrumentation would also increase the scientific knowledge to be gained when the great earthquake itself occurs.

This document describes a system for estimating the short-term hazard of a great earthquake on the southern San Andreas fault. Section II outlines the procedure followed in defining different levels and how they will be declared to have started and ended. Section III is the core of the document, and describes the different precursors that might be recognized and how they would determine a hazard assessment. Section IV describes the actions the U. S. Geological Survey will take in response to each level. Section V presents recommendations for improving geophysical recording on the southern San Andreas fault.

## II. Short-term Earthquake Hazard Assessments

The Parkfield earthquake prediction experiment provides a prototype for scientific response and communication systems for short term earthquake anomalies. A system of earthquake alerts that last for 72 hours has been established to respond to short term changes in geophysical properties of the San Andreas fault near Parkfield (Bakun *et al.*, 1987). Four levels of short term alerts, labeled D, C, B and A, have been defined for increasing probabilities of the Parkfield earthquake occurring within the time of the alert. Actions by certain designated scientists in the USGS are mandated for each alert level.

We adopt a similar system for the southern San Andreas fault. We define "short-term" to be, as at Parkfield, 72 hours, and establish a system of hazard levels such that actions at each level on the part of the USGS are similar to those defined for Parkfield. The phenomena that determine the levels are different for the southern San Andreas fault than for Parkfield but the probabilities that the forecast earthquake will occur within the 72-hour period are comparable. Because the social consequences of a M8 earthquake in a region with 12 million inhabitants are quite different from those of a M6 earthquake at a town with 34 inhabitants, the social response to a given level on the southern San Andreas fault is expected to differ greatly from that at Parkfield.

Although the levels are defined by the probability over 72 hours, the probability of the mainshock occurring is not constant over this time period. The hazard is highest immediately

following the possible precursor, and decreases with time. However, one alarm that lasts for a fixed time is preferred by public officials who will be responding. The 72 hour period is chosen because it is long enough to include the great majority of possible mainshocks but short enough to have a probability of an earthquake occurring that is significantly greater than the background probability. A hazard level will lapse 72 hours after it began if no further activity commensurate with that level occurs within that time. If further activity does occur, the level will continue for 72 hours from the time of the later activity.

A major difference between the system described here and that developed for Parkfield is the absence of a level A. At Parkfield, a geologic hazard warning will be issued immediately and automatically by the USGS at level-A. This statement warns of approximately a 1 in 2 chance of a M6 Parkfield earthquake occurring within 72 hours and is in essence a formal earthquake prediction. We do not feel that the level of understanding of the behavior of the southern San Andreas fault allows probabilities as high as 50% to be determined. As described below, we feel the highest probabilities that can be estimated for the southern San Andreas fault are on the order of 10-20%. Therefore, at the present time, the equivalent of a level-A alert cannot be reached for the southern San Andreas fault. We allow the definition to remain so as not to preclude the possibility of more certainty in the future as our knowledge increases.

### **III. Possible Earthquake Precursors**

The Working Group considered three types of phenomena as possible earthquake precursors - anomalous earthquake activity, surface creep on faults, and changes in strain as recorded on strainmeters. Of these, only earthquakes, as potential foreshocks to great earthquakes, are well enough recorded and understood to provide a formal estimate of conditional probabilities; creep and strain must be evaluated more subjectively. While other phenomena besides these three, such as ground water geochemistry or geoelectricity, might show precursory activity, they are not well enough recorded along the southern San Andreas fault nor is their relationship with large earthquakes well enough understood to be used at this time for short-term earthquake hazard assessment.

We first summarize the equipment currently deployed to record these phenomena. Then, for each possible precursor, we discuss (1) the evidence for that phenomenon as a short-term precursor to large earthquakes, (2) its recorded history along the southern San Andreas fault and (3) appropriate levels of concern for different possibly precursory activities.

#### **III.1 Summary of Current Instrumentation**

Earthquakes in southern California are recorded by the Southern California Seismic Network, a joint project of the California Institute of Technology (Caltech) and the southern California office of the United States Geological Survey (USGS), in Pasadena. The average station spacing near the southern San Andreas fault is about 20 km, so that all earthquakes above magnitude 1.8 are recorded in the southern California catalog (Figure 2). Most of the stations consist of a single short-period vertical seismometer, so that S-wave arrival times cannot usually be determined. Two three-component, force-balance accelerometers and three high-gain three-component seismometers are located within 50 km of the southern San Andreas fault (Figure 2). Because earthquakes in the Coachella Valley tend to be shallow (above 10 km), the lack of S-wave readings and the 20 km station spacing mean that the depths of these earthquakes cannot usually be resolved within 5 km. Ten stations within 50 km of the southern San Andreas fault have an extra vertical component with a low gain setting; all other stations saturate at about magnitude 2.5-3.0.

# Southern California Seismic Network

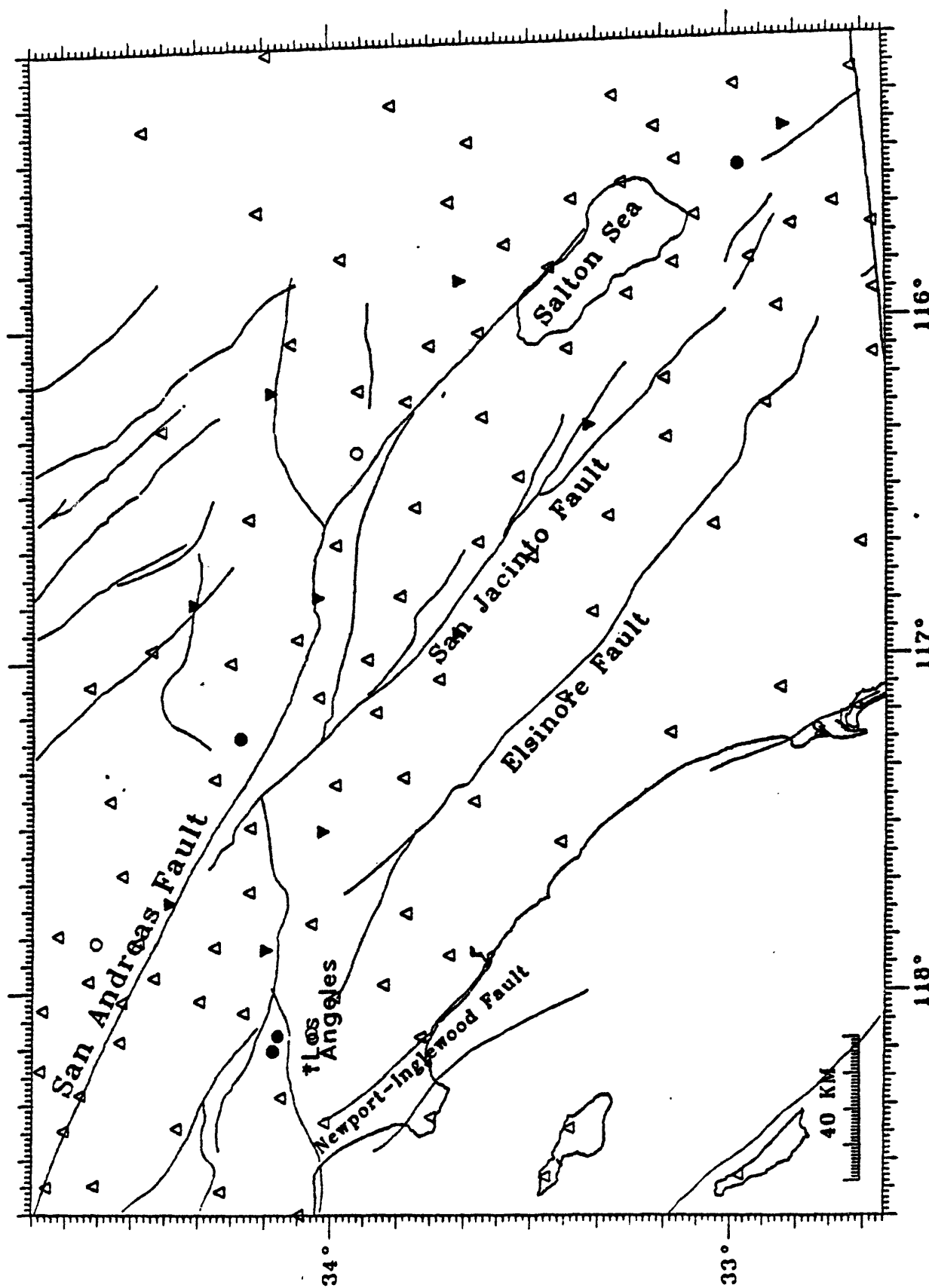


Figure 2. A map of the stations of the Southern California Seismic Network. Sites with one short-period vertical component are shown by triangles. Two-component stations (one high-gain and one low-gain vertical sensors) are shown by filled inverted triangles. Sites with two horizontal components in addition to vertical sensors are shown by circles: filled

The analog data from the seismic stations are first telemetered to Pasadena by microwave and leased telephone lines, and then digitized and recorded by a central recording computer. All of the data are processed and analyzed within one to three days. One quarter of the stations (64 of the 280 for all of southern California) are analyzed by a real-time picker (RTP) (Allen, 1982). This system provides the location of any earthquake of magnitude greater than 2.2, within 5 minutes of its initiation. For earthquakes of magnitude less than 4.1, the magnitude is also determined. A new software system is being developed to provide real-time locations and magnitudes for all earthquakes with magnitudes between 1.8 and 6.5. This system is expected to be operational by 1990 or 1991.

There are relatively few measurements of ground deformation in southern California. Existing instrumentation includes alignment arrays, geodetic nets, creepmeters, several strainmeters and tiltmeters at the Pinon Flat Observatory, and a water-level tilt network in the Salton Sea (Figure 3). Alignment arrays are sets of monuments installed over a small area (typically less than 1 km<sup>2</sup>) that are repeatedly surveyed. Alignment arrays and geodetic nets around the southern San Andreas fault are supplemented with Global Positioning Satellite (GPS) measurements. However, these arrays and networks are unlikely to provide information on short term precursors to large earthquakes, because the measurements are made too infrequently, often at yearly intervals. A permanent GPS network is being planned that could be used continuously.

Creepmeters are instruments installed to measure surface slip across the trace of a fault. Caltech operates four creepmeters, two on the San Andreas fault and two on the Imperial fault. One Imperial fault creepmeter is recorded on site; data from the others are telemetered to Pasadena. Several digital creepmeters (up to 10) will be placed along the San Andreas and San Jacinto faults over the next few years in a cooperative project between the University of Colorado and Caltech. As planned, the resulting data will be recorded on site only. Without telemetry, these instruments cannot be used for short term earthquake hazard assessment.

The only continuous, high-precision strain measurements are made at Pinon Flat Observatory (PFO), within 40 km of much of the southern San Andreas fault, but 75 km away from the southern end at Bombay Beach and the northern end at Cajon Pass (Figure 3). The instrumentation at PFO includes long-base strainmeters and tiltmeters, a borehole dilatometer, a borehole tensor strainmeter, and several borehole tiltmeters. These provide very high sensitivity recordings; however, different instruments have different time periods over which they give the best results, and different degrees of processing required to attain these results. The most easily interpreted instrument is the borehole dilatometer, because it is subject to the least environmental disturbance. The long-base instruments produce better data, but processing and interpreting these data require someone familiar with the idiosyncrasies of the instruments. Expert involvement is also desirable to interpret data from the borehole tensor strainmeter.

A closer but less sensitive record of crustal deformation is provided by the water-level recorders operated around the Salton Sea by the Lamont-Doherty Geological Observatory. The difference in water-level between stations gives a measure of tilt between them. These data also require an expert for processing and interpretation, especially because a wide range of environmental effects may cause apparent tilts. Moreover, meaningful signals cannot be resolved for periods of less than 2 days because of seiches and thermal noise, so that data from this system cannot be used for short-term analysis.

Data relevant to short-term earthquake prediction on the southern San Andreas fault are thus recorded by several different organizations. Seismic data are recorded by the cooperative Caltech/USGS southern California seismic network in Pasadena. Creepmeters on the southern San Andreas fault are recorded on site and retrieved by Caltech (2 instruments) and University of Colorado (2 instruments). Strain data from PFO are recorded on-site, along with a computer connection to the University of California at San Diego. The Salton Sea data are stored on site by a

A map of the Salton River and Delta region, showing the river's course and its distribution into the Salton Sea. The map includes several labeled locations: San Bernardino, Palm Springs, Pinon Flat, and the Salton Sea. It also identifies major faults: the San Jacinto Fault and the Elsinore Fault. Seismic stations are marked with triangles (Digital) and squares (Telemetered). A scale bar indicates 20 KM. The map is framed by latitude and longitude coordinates.

**Figure 3.** A map of sites at which strain, creep or tilt are measured in southern California. The Pinon Flat Strain Observatory is shown by a large triangle. Creepmeters are shown by squares and noted if they are telemetered to Pasadena or digitally recorded. Alignment arrays are shown by small triangles. The Lamont water level gauge network is shown by circles.

computer and accessed by modem by scientists at Lamont in New York. In addition, two dilatometers in the Mojave Desert (50-200 km from the southern San Andreas fault) have satellite telemetry to Menlo Park. A central recording and analysis facility for southern California has not been established.

### III.2 Foreshocks

Half of the strike-slip earthquakes in California have been preceded by immediate foreshocks within 3 units of magnitude (Jones, 1984), including the 1857 magnitude 8 Fort Tejon earthquake on the southern San Andreas fault. Two of the four moderate earthquakes on the southern San Andreas fault in the last six decades have also had foreshocks.

Thus, the next southern San Andreas mainshock could well be preceded by one or more immediate foreshocks. An immediate foreshock is defined as an earthquake, smaller than the mainshock, that occurs less than 3 days before it and within 10 km of the mainshock's epicenter (Jones, 1985). Although immediate foreshocks are well-documented, they can only be identified after the later, larger earthquake occurs. So far, no characteristic has been found that distinguishes foreshocks from background earthquake activity. Therefore, when a small to moderate earthquake occurs on the southern San Andreas fault, we cannot tell if it is a foreshock, but the possibility that it increases the probability that a major earthquake could soon occur.

This increase in the seismic hazard following moderate earthquakes has been recognized and used for a few short-term earthquake advisories (e.g., Goltz, 1985). These warnings have been based on a regional level of foreshock occurrence (Jones, 1985), applicable anywhere in southern California. Applying such a formula to the southern San Andreas ignores both the existence of an estimate of the long-term probability for the large event and the substantial spatial variations in background activity along this fault segment. Thus, the Working Group felt that we needed a formal method for estimating the probability of a large earthquake, given the occurrence of a possible foreshock near a major fault. A method has been developed and is described in Appendix A. In Section III.2.1 we give a relatively nontechnical discussion of the procedure used, emphasizing the reasoning behind the estimate rather than the formal mathematics (given fully in Appendix A). Section III.2.2 describes our conclusions regarding the foreshock magnitudes needed to reach particular levels.

#### III.2.1 Theory

In determining short-term probabilities, we assume that foreshocks and mainshocks are theoretically (but not necessarily in practice) separable from background seismicity. We then suppose that some earthquake has occurred, either a background event or a foreshock, though we do not know which. If this "candidate event" is a foreshock, the mainshock will by definition soon follow. To see the reasoning used, a simple example may help. Leaving out the complications of magnitude, location, and so on, suppose that mainshocks occur on the average every 500 years, and that half of them have foreshocks (in this example, defined as being within a day of the mainshock); then we expect a foreshock every 1000 years. Suppose further that a background event occurs on average every year. Then, given a potential foreshock, there is very nearly one chance in 1000 that it is a foreshock. This makes the probability of a mainshock in the next day 0.1%. While this is low, it is far above the background probability, which is (assuming a Poisson process) 1 in 500 times 365, or 0.00055%.

What we have done here is to compute the probability that a mainshock will soon occur, given a foreshock *or* background earthquake; that is, a conditional probability. Appendix A gives the complete formula for this conditional probability, dependent on the same quantities we have just used: the probabilities of a background earthquake, of a mainshock, and of a foreshock if a mainshock has actually happened (which in our simple case is the fraction of mainshocks having

foreshocks). In the example, all of these probabilities are assumed to have been estimated from a very long record of seismicity. In reality, we get these quantities from very different sources:

**Background Seismicity.** The probability of a background earthquake is derived from the magnitude-frequency relation and spatial distribution of earthquakes above magnitude 1.8 recorded over the last 11 years by the Southern California Seismic Network. The rate of background seismicity varies considerably along the southern San Andreas fault, from the highest rate for the whole San Andreas system at San Geronio Pass, to one of the lowest in the Mecca Hills. We have divided the southern segment into four microseismic zones to account for these variations (Figure 4). The Mecca Hills and Palm Springs microseismic regions make up the Coachella Valley segment and the San Geronio and San Bernardino microseismic regions make up the San Bernardino Mountains segment of WGCEP (1988)

A critical assumption in using this catalog data is that the last 11 years of earthquake activity represents the long-term rate. The magnitude-frequency distribution determined from the earthquakes above magnitude 3.0 since 1932 is comparable to that determined from the past 11 years, suggesting the 11 year interval is typical. If the rate of seismic activity along the southern segment were to change, the probabilities determined here should be modified.

**Long-term Probability of Mainshocks.** The long-term probability of a mainshock occurring on the Coachella Valley segment of the southern San Andreas fault is a complicated, controversial quantity that has already been the topic of another Working Group, the Working Group on California Earthquake Probabilities (WGCEP, 1988). We use here the results of WGCEP (1988), a probability of 40% over the next 30 years for the Coachella segment and 20% over 30 years for the San Bernardino Mountains segment. The committee has adopted these results because they have already been reviewed and accepted by the National and California Earthquake Prediction Evaluation Councils. Davis *et al.* (1989) have recently made a case for a much lower probability for the Coachella Valley segment: 9% over the next 30 years (they did not consider the San Bernardino segment). Probabilities have been calculated using both values to show the effect of the different assumed values for long term probability in the Coachella Valley.

We also assume that all sections of the southern San Andreas fault are equally likely to contain the epicenter of the mainshock. It has been suggested that mainshocks are more likely to occur at points of complication on the fault. However, at the gross scale at which we are analyzing the southern San Andreas fault, each region has numerous points of complication, and further refinement is not supported by our present state of knowledge. Another possibility we rejected was to assume the mainshock more likely to occur in regions with a high rate of background seismicity. One clear lesson from 50 years of seismic recording in southern California is that large earthquakes do not preferentially occur at the sites of small earthquakes.

**Conditional Probability of Foreshocks.** The third quantity needed is the conditional probability of a foreshock given that a mainshock has occurred. In Appendix A, we call this a "reverse transition probability" because, unlike most conditional probabilities, it goes backwards in time. We use the chance that an earlier event precedes a later one, rather than the more customary approach of discussing the chance that one type of event will be followed by another. This does not violate causality; we are simply assuming that the two types of events (foreshocks and mainshocks) are interrelated.

If we had a record of the foreshocks for many Coachella Valley mainshocks, or even many San Andreas mainshocks, we could estimate the conditional probability directly. Since we do not, we assume that the average properties and probabilities of foreshocks to moderate and large earthquakes on many southern California faults adequately approximate the temporal average over many mainshocks on the southern San Andreas fault. The simple model discussed at the start of this section presented only one type of foreshock and mainshock, so that the reverse transition probability was the fraction of mainshocks preceded by foreshocks. In actuality, both foreshocks and mainshocks come with additional "labels" such as location and magnitude. We must extend



# Southern San Andreas Fault 1977-1987 $M > 1.8$ Declustered

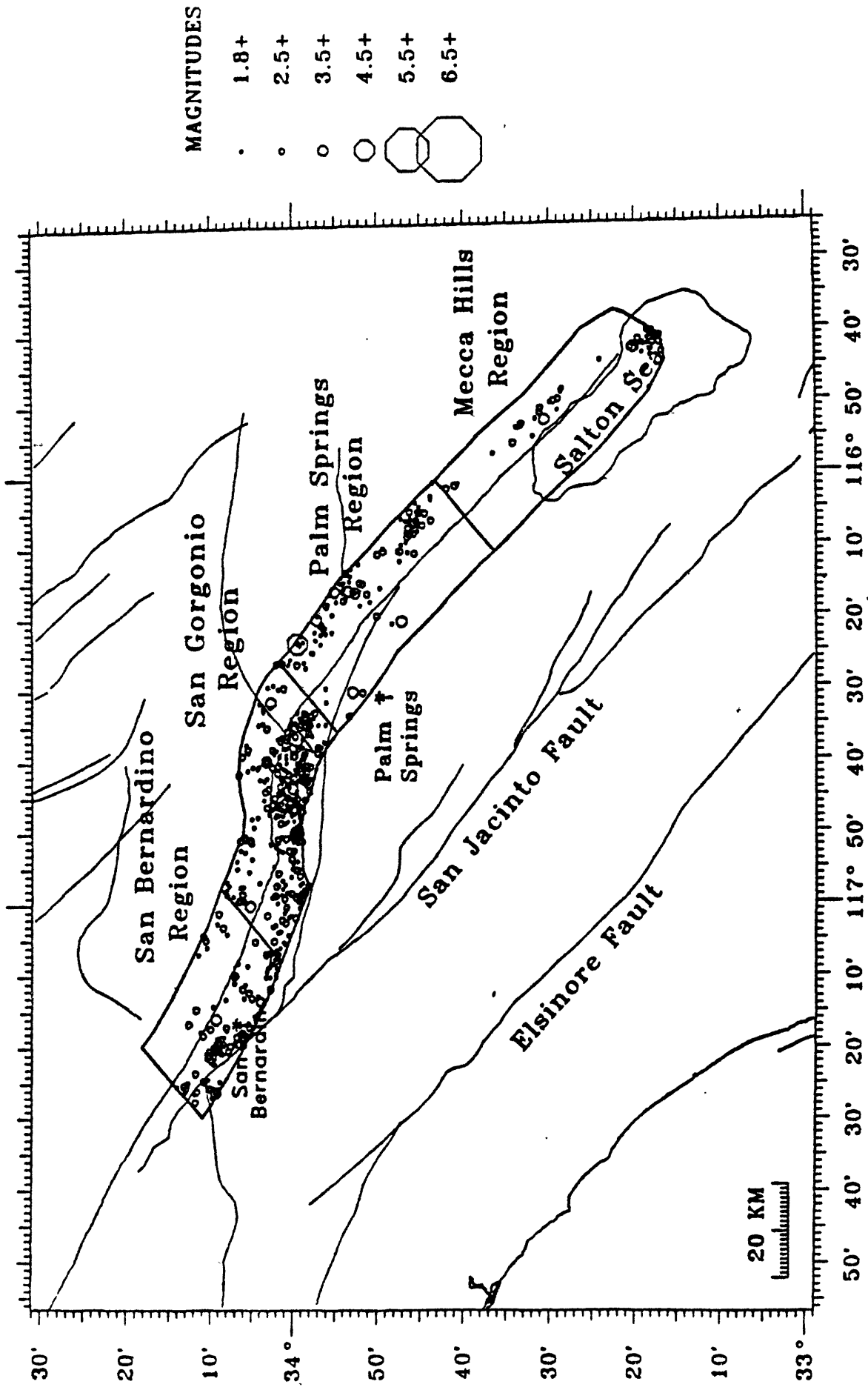


Figure 4. A map of magnitude 1.8 and greater earthquakes located within 10 km of the Coachella Valley segment of the southern San Andreas fault recorded in the Caltech catalog between 1977 and 1987.

the conditional probability to allow for these. Again, Appendix A gives the full details, which we summarize here. Foreshocks are definable once the mainshock occurs and the average characteristics of California foreshocks are briefly described and used to define the reverse transition probability for potential San Andreas foreshocks.

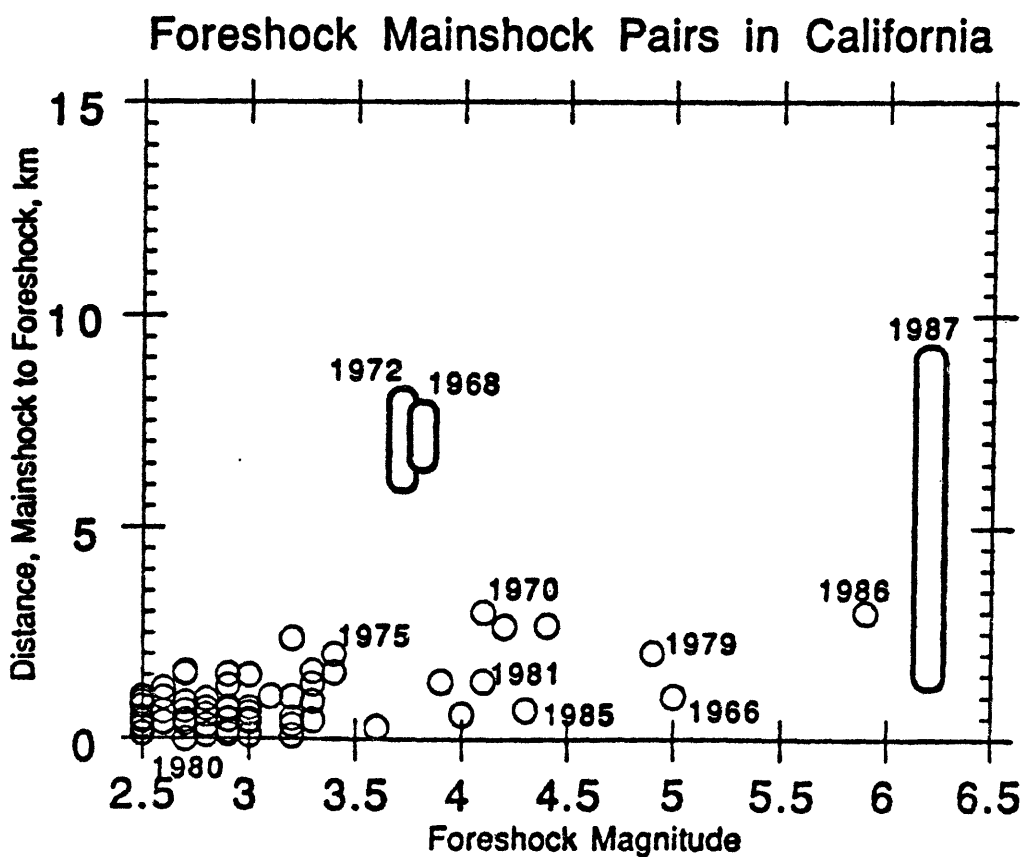
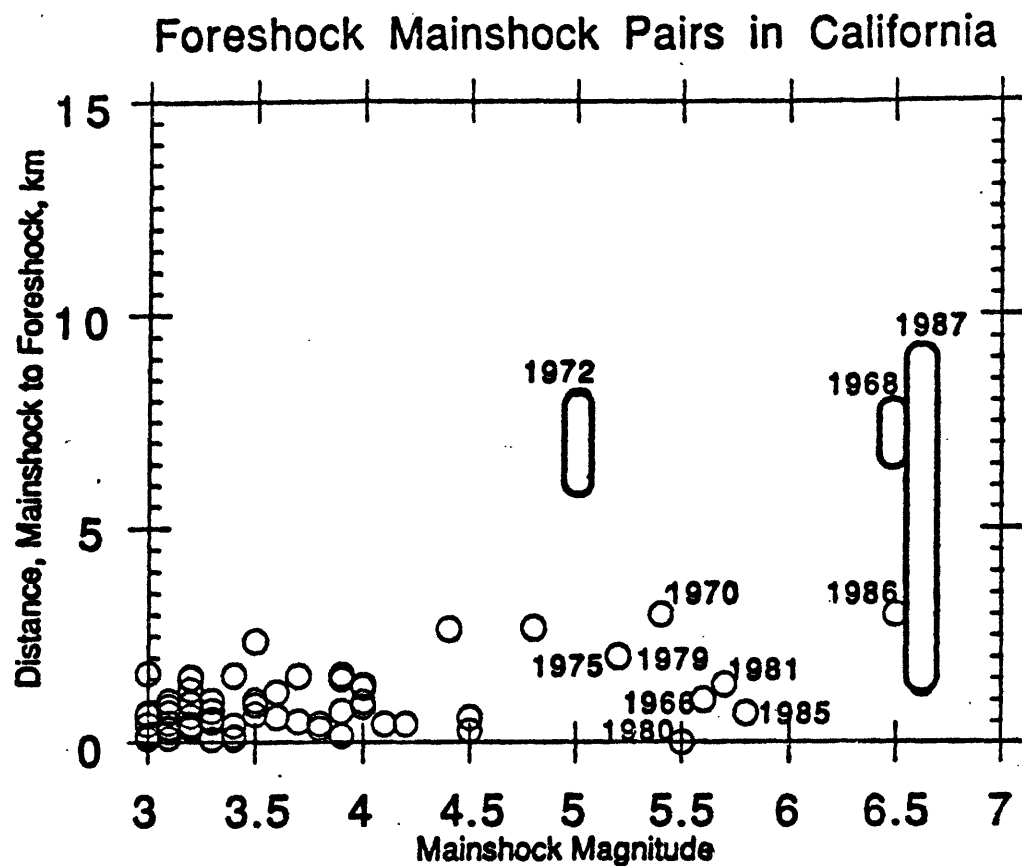
*Temporal Dependence.* If a foreshock occurs, it is more likely to happen just before the mainshock than some greater time before it (Jones, 1985; Jones and Molnar, 1979). The distribution of foreshock-mainshock intervals,  $t$ , varies roughly as  $1/t$ . As a consequence, the maximum conditional probability of a mainshock occurs just after the potential foreshock, and diminishes rapidly with time. (As time elapses with no mainshock, it becomes more probable that the potential foreshock was just a background earthquake). We have not included this temporal change directly in our levels, but simply leave the probability unchanged for our chosen 72-hour span. This interval is approximately the time within which 95% of mainshocks will have occurred.

*Location.* Foreshocks occur close in space as well as close in time to the mainshock. All well-recorded foreshocks in southern California have had epicenters within 10 km of their mainshocks' epicenters (Figure 5; Appendix A). No dependence of this distance on magnitude of mainshock or foreshock has been seen (Figure 5). However, a significant minority of these foreshocks have occurred on a different fault from their mainshock so an earthquake need not be on the southern San Andreas fault to be considered a potential foreshock. The Working Group has chosen a somewhat more generous definition of foreshock and required only that some part of the rupture zone of the foreshock lie within 10 km of the southern San Andreas fault. Defining the distance from the fault in terms of the rupture zone of the potential foreshock allows the monitoring seismologists some flexibility in evaluating a particular earthquake sequence.

As noted above, we have assumed that the mainshock epicenter is equally likely anywhere along the southern San Andreas fault. We have also assumed that foreshocks are equally likely to occur anywhere along the fault. In particular, we discussed and rejected the hypothesis that foreshocks are preferentially located at sites of high background activity. Although data on this subject are limited, what modern data we have do not support this hypothesis (Jones, 1984). One example is the lack of foreshocks on the Calaveras fault despite a rate of background seismicity that is one of the highest in California.

*Magnitude Dependence.* The least certain part of the transition probability is how it depends on mainshock and foreshock magnitude. Our data on this are inevitably incomplete because a much lower magnitude threshold must be used for foreshocks than for mainshocks to consider the magnitude distribution of all possible foreshocks to a given mainshock. The southern California data suggest that for any narrow range of mainshock magnitudes all foreshock magnitudes are equally likely (except of course that foreshocks are always smaller). We have therefore assumed a flat distribution with magnitude of the foreshocks and used Jones' (1984) finding that half of the strike-slip earthquakes in California were preceded by foreshocks within 3 units of magnitude.

We have treated each of the above factors (time, location, and magnitude) separately, because the data available do not suggest any correlation among them. Likewise, we have not included any other parameters upon which the reverse transition probability might depend. For instance, while we might suspect that foreshocks would have focal mechanisms similar to that of the mainshock, we lack the data to evaluate this properly. Once more data have been accumulated, differences in probability depending on focal mechanism, number of aftershocks to the potential foreshock, tectonic regime, or other criteria can be accommodated by the method described in Appendix A. But at this point, none are sufficiently well documented for inclusion.



**Figure 5.** Distance between foreshock and mainshock epicenters versus the (a) magnitude of the mainshock and (b) magnitude of the foreshock for foreshock-mainshock sequences (foreshocks of magnitude 2.5 and greater and mainshocks of magnitude 3.0 and greater) recorded in the Caltech catalog between 1977 and 1987. Sequences that have been relocated in special studies are also plotted and include 1966 Parkfield, 1968 Borrego Mountain, 1970 Lytle Creek, 1972 Bear Valley, 1975 Galway Lakes, 1979 Homestead, 1980 Livermore, 1981 Westmoreland, 1985 Kettleman Hills, 1986 Chalfant Valley, and 1987 Superstition Hills.

### III.2.2 Hazard Levels from Foreshocks

Because we can now formally determine the probability of a large earthquake occurring after a potential foreshock, we can define minimum probabilities for each of the levels we have chosen. We define minimum probabilities that a mainshock will occur within the 72 hour interval after an earthquake along the two southern segments of the San Andreas fault of 5% for level-B, 1% for level-C, and 0.1% for level-D. We assume that if the rupture zone of the potential foreshock is within 10 km of the southern San Andreas fault, then the probability increases as outlined below. By defining the distance between the potential foreshock and the San Andreas fault in terms of the rupture zone, we require subjective judgement by the seismologists monitoring the fault in determining the extent of the rupture zone. In particular, the documented tendency of earthquakes within the Brawley Seismic Zone (just south of the southern end of the San Andreas fault) to have rupture areas much larger than normally associated with earthquakes of the same magnitude (Johnson and Hill, 1982) and the presence of northeast trending faults in the same area (Hudnut et al., 1989) need to be taken into account.

Appendix A derives the conditional probability of a mainshock occurring given a potential foreshock (Equation 28). This conditional probability is a function of (a) the time window over which the probability is evaluated, (b) the long-term probability of the mainshock in that time window, (c) the length of the fault, (d) the rate density of background earthquakes (as a function of magnitude) over that length of the fault, and (e) the percentage of mainshocks preceded by foreshocks within the time window defined in (a). As described in the Appendix, we have used Jones' (1984) finding that half of the strike-slip earthquakes in California were preceded by foreshocks within 3 units of magnitude and assumed a flat distribution with magnitude of the foreshocks for (e).

We have defined levels for two of the segments of the WGCEP (1988). They estimated the 30-year probability of a mainshock of  $M = 7.5 - 8.0$  to be 40% for the Coachella Valley segment and 20% for the San Bernardino segment (WGCEP, 1988). The corresponding long term probabilities for any 72-hour interval are 0.011% and 0.0055%. The length of the two segments are 110 and 100 km, respectively. Table 1 gives the magnitudes of potential foreshocks needed to reach the chosen probabilities for characteristic mainshocks in the four microseismic zones of the southern San Andreas, given the rates of background activity detailed in Appendix A.

TABLE 1. Magnitudes of Potential Foreshocks  
for the Southern San Andreas Fault

Level Probability of M7.5 in 72 hr	B 5-25%	C 1-5%	D 0.1-1%	Background Probability
San Bernardino	5.8	5.0	3.9	0.0055%
San Gorgonio	6.1	5.3	4.2	0.0055%
Palm Springs	5.2	4.5	3.4	0.011%
Mecca Hills	4.9	4.2	3.1	0.011%

The information in Table 1 is displayed graphically in Figure 6. The increase in probability with greater magnitudes can be seen as well as how the magnitudes needed to reach a given level vary between the different microseismic zones. The background probabilities of the characteristic mainshocks on the San Bernardino Mountains and Coachella Valley segments are also shown. A level D represents a factor of 10 increase on the Coachella Valley segment but a factor of 20 increase on the San Bernardino Mountains segment compared to the background probability of the characteristic mainshock.

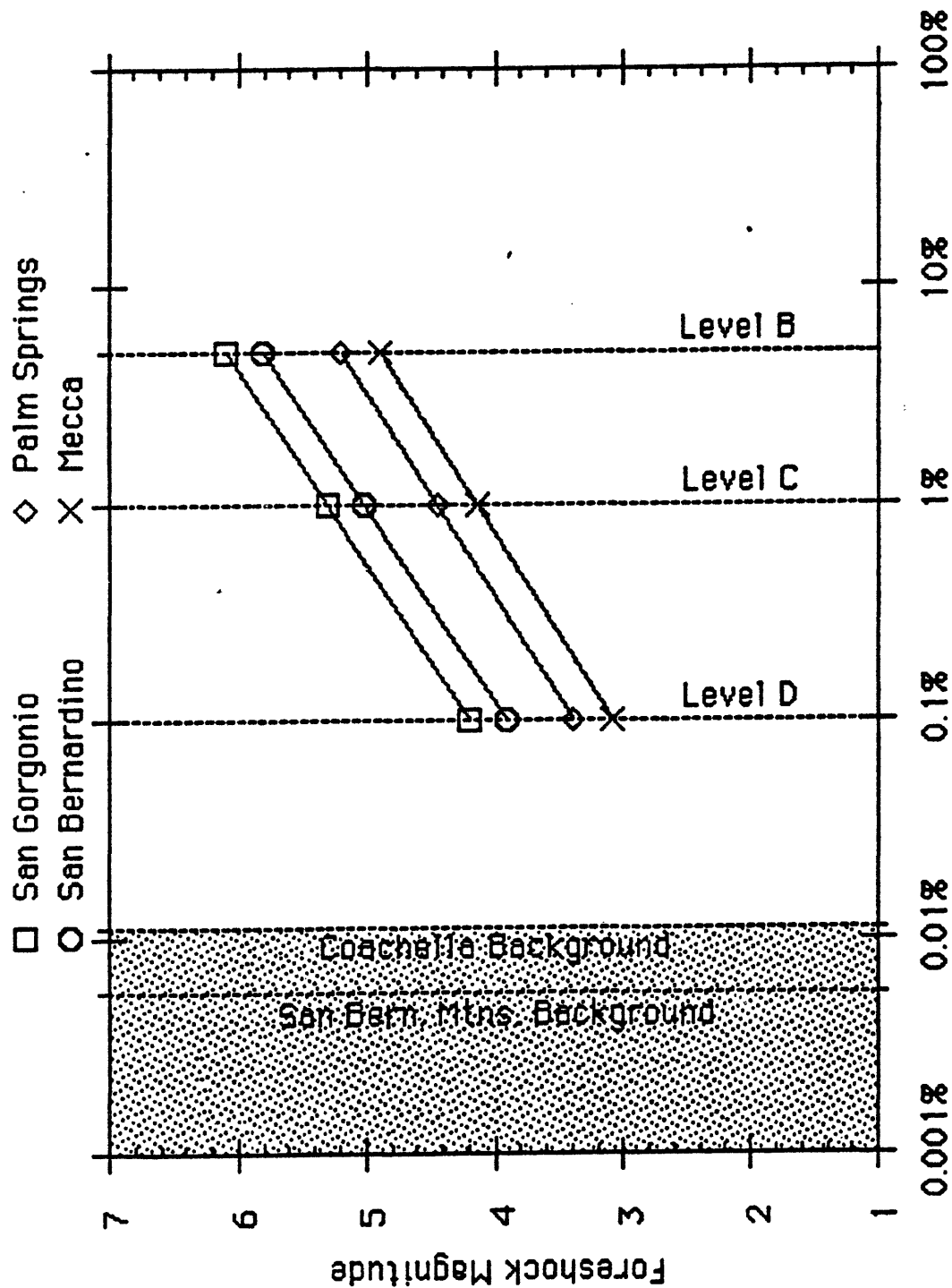


Figure 6. The probability of a major earthquake occurring on the southern San Andreas fault within 72 hours after a potential foreshock versus the magnitude of the potential foreshock for the four microseismic regions shown in Figure 4, San Bernardino (circles), San Gorgonio (squares), Palm Springs (diamonds), and Mecca Hills (crosses). The background probabilities for 72 hours based on the long-term probabilities of WGCEP (1988) are also shown for the Coachella Valley and San Bernardino Mountains segments.

Expected false alarm rates for these levels are calculated in Appendix A. On the Coachella Valley segment, the present rate of background seismicity is expected to produce a level-B false alarm once every 28 years, a level-C false alarm every 5 years, and a level-D false alarm once every 6 months. On the San Bernardino Mountains segment, the present rate of background seismicity is expected to produce a level-B false alarm once every 57 years, a level-C false alarm every 10 years, and a level-D false alarm once a year. These false alarm rates are compatible with the stated probability levels. For a probability of 0.05, nineteen level-B false alarms should be issued for every successful prediction. The mean recurrence time of large earthquakes is about 250 years (WGCEP, 1988), and we assumed that half of these would be preceded by foreshocks. We should thus successfully predict once every 500 years during which time 18 false alarms would be issued (at 1 per 28 years). In the last 60 years of recorded earthquakes, one earthquake (the 1948 Desert Hot Springs local magnitude 6.5 earthquake) was large enough to produce a level-B hazard estimate.

The magnitudes in the above table are determined using the results of WGCEP (1988) which give a 30-year probability for the Coachella Valley segment of a  $M=7.5-8.0$  earthquake to be 40%. Davis *et al.* (1989) have recently made a case for a much lower 30-year probability of 9%. This Working Group felt that which, if either, of these values was correct is yet to be conclusively decided; however, to provide a consistent approach to both the San Bernardino Mountains and the Coachella Valley segments, we have adopted the results of WGCEP (1988). We feel that this is the least certain part of the analysis and that further work on this topic is important to reduce the uncertainties. The effect of the long-term probabilities on the short-term results can be seen by recalculating the magnitudes of potential foreshocks for each level, using the 30-year probabilities of Davis *et al.* (1989), shown in Table 2. The magnitude needed to reach each level increases by 0.7 units for the Davis *et al.* (1989) probability as compared to the WGCEP (1988) values.

TABLE 2. Alternate Solution Using Davis *et al.* (1989)  
Magnitudes of Potential Foreshocks for the Coachella Valley Segment

Level Probability of $M7.5$ in 72 hr	B 5-30%	C 1-5%	D 0.1-1%	Background Probability
Palm Springs	5.9	5.2	4.1	0.0025%
Mecca Hills	5.6	4.8	3.8	0.0025%

The false alarm rates for these alternate values are one level-B false alarm every 126 years, one level-C false alarm every 23 years and one level-D false alarm every 2.2 years.

### III.3 Aseismic Fault Slip

Many theoretical analyses of fault rupture predict that the sudden, unstable slip of an earthquake should be preceded by some amount of stable slip on the fault (e. g., Stuart, 1986; Rudnicki, 1988; Lorenzetti and Tullis, 1989). The amount of slip depends upon the model but most models predict a measurable amount at the surface for the largest earthquakes. Fortuitous recordings from some earthquakes (described in Section III.3.3) also suggest that faults can start to move before the earthquake. Current earthquake prediction experiments like those at Parkfield and the Tokai Gap in Japan therefore include detailed recordings of ground deformation. However, for surface fault creep, we lack the detailed, historic data needed to make a formal calculation of conditional probabilities, as we did for foreshocks. We have instead considered both the general evidence for creep as a precursor to large earthquakes and the history of creep on the southern San

Andreas fault, and from these factors developed subjective criteria for evaluating creep episodes along the southern San Andreas fault.

These criteria are restricted by the limited number of creepmeters installed along the southern San Andreas fault. At the present time, only one creepmeter is telemetered to Pasadena. If more data were available with reasonably dense spacing along the fault, then we would have required any recognized creep episode to be recorded on at least two creepmeters within 10 km. With present data, we do not have the luxury of redundancy.

### III.3.1 Steady State Creep

Measurements of fault-crossing features in the Coachella Valley indicate slow aseismic surface creep. Observations of offset geological features since 1907, offset man-made features since 1950, and geodetic measurements of creep since 1970 all indicate that creep of 2-3 mm/yr has gone on for the last 80 years (Sieh and Williams, 1990). Where this aseismic creep has been monitored continuously (Figure 3), it mostly occurs in episodes lasting less than a day and having amplitudes less than 1 cm (Louie *et al.*, 1984). These episodes seem to occur randomly, but the long term rate of 2-3 mm/yr (determined on baselines of less than 20 m) appears to be steady, at least in the current century and possibly for a longer period. Geodetic data across the Coachella Valley (from baselines longer than 30 km) indicate a dextral shear rate greater than 20 mm/yr (King and Savage, 1983). A simple elastic model of the Coachella Valley suggests that the observed creep and shear strain data are consistent with an effectively frictionless fault zone in the uppermost 3-4 km of the fault, and a locked fault below that depth (Bilham and King, 1989).

### III.3.2 Triggered Creep

Creep also occurs on the southern San Andreas fault at the time of, or shortly after, large local earthquakes. In 1968, 1979, and again in 1986, surface displacements of 2-20 mm occurred along segments of the fault after earthquakes with magnitude 6 or more. What causes such creep is not clear. Observed triggered creep of 22 mm at one location in the Mecca Hills in 1968 (Figure 3) may indicate that the maximum creep event amplitude may be larger than that so far observed by the few available creepmeters. The timing of the 1968 creep event, however, is not well known, and the observed displacement of 22 mm may represent several smaller creep events. The triggered creep is not necessarily coseismic; creep in 1986 occurred on Durmid Hill, 60 km from the North Palm Springs mainshock (Figure 3) and 17 hours after the mainshock (Williams *et al.*, 1989).

### III.3.3 Evidence for Premonitory Creep on California Faults

There are two known cases in which creep may have occurred at the surface prior to a mainshock at depth:

*Parkfield 1966:* En echelon cracks were observed along the fault trace in the days preceding the 1966 Parkfield earthquake, and a steel irrigation pipe across the fault broke nine hours before the mainshock (Wallace and Roth, 1967).

*Superstition Hills 1987:* Six observations of fault creep as it developed in the hours to months following the 1987 Superstition Hills earthquake could be fit to a smooth model if 4-14 cm of creep had occurred on the northernmost 4 km of the fault before the mainshock (Sharp *et al.*, 1989).

Neither of these examples is completely satisfactory. The failed pipe at Parkfield could be a coincidence, and the surface cracks might be related to similar seasonal cracking subsequently observed in this area. The Superstition Hills evidence is better documented, but complicated by the

foreshock. A large, magnitude 6.2, foreshock on the Elmore Ranch fault preceded this magnitude 6.6 earthquake by 11.4 hours. The inferred precursory creep occurred close to the intersection of the fault with the Elmore Ranch fault. When this creep occurred on the Superstition Hills fault is uncertain, and it could have been coseismic with and mechanically related to the first earthquake.

In the 1979 Imperial Valley earthquake (magnitude 6.5) on the Imperial fault, a creepmeter was in place across the fault well before the earthquake. The data from this instrument showed no fault motion until after the earthquake had begun (Cohn *et al.*, 1982). Thus precursory surface slip might be recordable at present levels prior to some, but certainly not all strike-slip earthquakes.

### III.3.4 Hazard Levels from Surficial Creep

We thus cannot ignore the possibility of a fault slipping aseismically before a strike-slip mainshock. Even scientists who believe that creep will not precede the next major earthquake still think that *if* a large amount of creep were seen, it should raise our expectations of a major earthquake. However, as was noted above, we lack the kind of data for creep needed to formally define the increase in mainshock probability. The Working Group therefore decided to use only one level for creep arbitrarily set equal to a seismic level D. This level would be achieved whenever we observe creep greater than that so far recorded on the southern San Andreas fault, a more stringent requirement than for the seismic data (for which level D will be reached annually). However, the unclear connection between creep and large earthquakes makes it appropriate to require a larger signal for a comparable level.

The amount of creep that will be considered anomalous is defined differently for aseismic creep and creep episodes accompanying earthquakes. We distinguish three classes of creep:

(1) *Single aseismic creep events:* The largest previous creep event recorded on the southern San Andreas fault was less than 1 cm (Louie *et al.*, 1984). Therefore, a single creep event exceeding 1 cm within 1 day will be considered anomalous and produce a level D.

(2) *Multiple aseismic creep events:* Triggered and aseismic creep combine to provide 2-3 mm/yr of creep on the southern San Andreas fault, a rate that appears constant over at least the last century. A significant increase in rate would be unusual. Therefore, if several creep events of less than 1 cm were to occur within 1 year such that the yearly rate exceeds 2 cm, the last creep event would produce a level D.

(3) *Triggered creep:* The documented occurrence of triggered slip following local, moderate earthquakes requires a higher slip threshold for triggered than aseismic slip. The largest previous creep event was 22 mm in the Mecca Hills following the 1968 Borrego Mountain earthquake. Triggered slip on the southern San Andreas fault will produce a level D if it exceeds 25 mm of creep at any one site or 20 mm over at least 20 km.

These levels should be regarded as the best educated guess until more extensive case histories permit stricter quantification.



### III.4 Strain

#### III.4.1 Available Data

Strainmeters are not widely distributed in southern California. As described in Section III.1, only two installations measure strain within 100 km of the Coachella Valley: the Pinon Flat Observatory (PFO), 20 km south of Palm Springs, and water level monitors around the Salton Sea that can be used as a less sensitive tiltmeter (Figure 3). Short term strains on the order of one part in  $10^9$  can be resolved with the instruments at PFO while the Salton Sea installation can only resolve vertical deformation of one microradian per 2 days.

#### III.4.2 Criteria for Strain

Theory and some observations suggest that fault slip, like creep, can begin before an earthquake occurs. Thus, clear evidence of deep-seated slip on the southern San Andreas fault would be extremely anomalous and the basis for an earthquake alert. The problem is obtaining "clear evidence." Creepmeters measure surface offsets that may not be related to slip at depth where the earthquakes start. Strainmeters will respond to slip at depth but measurements of strain at one place cannot determine which fault the slip might be on. Indeed, a single record of strain change cannot show whether the strain reflects displacement along a distant fault, some kind of broad-scale deformation, or a small local displacement.

With only one set of sensitive strainmeters near the southern San Andreas fault, a strain anomaly cannot by itself indicate slip on that fault. However, strain measurements can be used to supplement data recorded by the seismic network or creepmeters. Strain measurements can limit models proposed on the basis of creep or seismicity data because over short time periods crustal response to fault slip is that of an elastic halfspace, as demonstrated by observations of coseismic strain. For example, if a large creep event were observed along a given fault, then far-field strain data may show whether it was caused by shallow or deep movement.

Declaring a strain anomaly is slightly complicated at PFO, because of the particular mix of instruments now in use there. Moreover, because data are available from only one site, a trade-off will always exist between the amount of deformation and the distance to the deformation event when evaluating the possible source of a recorded anomaly. Rather than attempting to set precise levels of anomalous behavior, we propose here to define an anomaly as a signal unprecedented in the history of the instrument, as judged by someone familiar with it. Routine monitoring would probably use the borehole instruments at PFO, because of greater simplicity in processing the data, but any anomaly seen on these should be regarded as tentative until confirmed by the PFO long-base instruments. An anomaly on the latter must be taken seriously, because these instruments have a long history of stability and are largely immune to local disturbances that might affect the borehole instruments. They are also much more accessible for testing if a problem with the instrument is suspected.

A strain anomaly would itself reach only level D because of the ambiguity in interpreting a strain signal from only one site. However, the location of such an anomaly could be estimated from creep or seismicity if either should occur. In the latter case, the known location and strain anomaly size would give an estimate of the source moment. To give some idea of the possible numbers, the detectable level of change in strain over 10 hours is 1-5 nanostrain depending on the instrument (if the earth tides were automatically removed). For slip along the southern part of the Coachella fault segment, this strain level at PFO corresponds to what would be seen for a magnitude 5 "slow earthquake." A smaller event farther north along the fault would give the same signal, and of course a more rapid event would be more easily detected.

### III.4.3 Hazard Levels from Strain

Borehole dilatometers used for routine monitoring of PFO strain will only be considered anomalous if confirmed by the long baseline instruments. Strain anomalies are treated differently depending on whether or not they occur together with signals from the seismic or creep networks. As for creep, large uncertainties in strain measurement and in the relation between strain and large earthquakes have led the Working Group to use only one level for strain, arbitrarily set equal to a seismic level D.

(1) *Aseismic Strain Signals*: An aseismic strain change observed at PFO will reach level D if the signal is unprecedented in the history of the instrument as interpreted by someone familiar with it. This unsatisfying definition appears to be the best now available.

(2) *Strain Accompanied by Fault Slip*: Strain data from PFO can be used to delimit the type and amount of deformation when the location of the strain source can be determined, such as the deformation associated with a magnitude 5 or greater earthquake or an aseismic creep event along the southern San Andreas fault. Level D is reached if strain signals are detected that indicate anomalous fault slip at PFO by both borehole and surface instruments. "Anomalous" could mean unusually deep (greater than 8 km) or unusually large.

Because of the low sensitivity of the water level recorders at the Salton Sea, any tectonic tilt recorded at the Salton Sea should also be recorded by the more sensitive instruments at PFO. Therefore, signals from the tiltmeter network will not be used for short term hazard assessment.

### III.5 Combined Hazard Levels

If more than one anomaly were recorded at one time, then the situation would be considered more threatening. For instance, as discussed in the strain section, strain anomalies accompanying a magnitude 5 earthquake that suggest abnormally large slip at greater depths (where the great earthquake is expected to begin) would be much more ominous than the magnitude 5 earthquake by itself. Indeed, many of the strain anomalies are defined as occurring with some seismic activity. Some way of combining the levels must be adopted.

Because the strain and creep anomalies reach only level D, the combination rules can be rather simple. We have adopted a simplified version of the Parkfield combination rules. Thus a level D occurring during the 72 hours of a preexisting level C or D will raise the assessment by one level: the level C would become level B and the level D would become level C. For instance, a magnitude 4 earthquake along the Coachella Valley segment would by itself reach level D. If creep greater than 25 mm were to accompany or occur within 3 days of that earthquake (Creep level D #3), then the combined level would be C. However, we feel that the relationship between possibly precursory strain and earthquakes is not well enough understood to justify raising a seismic level B any higher because of a strain or creep anomaly.

#### IV. Response Plan for the USGS

The purpose of our system is to quantify and communicate information about temporary increases in the earthquake hazard. When a level is reached, the scientists in data acquisition, both inside and outside the USGS, of course must assure the integrity of the data recording systems. But the USGS must also communicate this assessment of the earthquake hazard to interested parties, both scientific and governmental. The response plan for the USGS detailed here is essentially the same as agreed upon for Parkfield, considering the different organizational structures of its southern and northern Californian operations. This plan involves only the scientific response to a given level and notification of the Governor's Office of Emergency Services of the State of California (OES).

The Chief of the USGS Office of Earthquakes, Volcanoes and Engineering (OEVE) must appoint and support a chief scientist for the southern San Andreas fault. All short-term hazard assessments for the southern San Andreas fault will be made by this chief scientist. Data from three different projects, the seismic network, the creepmeters and the Pinon Flat strain observatory, will be used for hazard assessment, but only one of these projects, the seismic network, is even partially operated by the USGS. If a central data recording center is established as recommended in the next section, operations of that center will be coordinated so that the chief scientist for the southern San Andreas will be notified of anomalies in any recorded phenomena. Until such time, the seismic data are monitored by USGS scientists, but university scientists must report anomalies in the other phenomena by telephone to the chief scientist. When an alert is declared in any of the three categories, the chief scientist will ask the researchers in all three projects to check their data to (1) look for other possible anomalies and (2) assure the integrity of the data recording and analysis systems. At a minimum, this system should insure that data on the great earthquake not be lost because of easily fixable, but unnoticed equipment problems.

The specific scientific response by the USGS to the three levels are given below.

##### Level D

Level D means that the probability of a great earthquake occurring within 72 hours is on the order of 0.1-1% (the exact probability for a strain or creep generated level D is not known). The appropriate response to this level is awareness. As described above, the chief scientist will notify all groups actively monitoring the southern San Andreas and request a check on other possible anomalies and the integrity of the data recording systems. The chief scientist will notify the scientist-in-charge of the southern California office of the USGS in Pasadena, and the chiefs of the Branches of Seismology and Tectonophysics in Menlo Park. Scientists outside the USGS doing research on the southern San Andreas fault could make arrangements to receive notification by fax or electronic-mail. The chief scientist will also notify the southern California office of the OES. At these low probabilities, no further action is warranted.

##### Level C

Level C means that the probability of a great earthquake occurring within 72 hours is on the order of 1-5%. The appropriate response to this level is precaution. In addition to the activities undertaken for level D, the chief scientist will also notify the chief of the USGS Office of Earthquakes Volcanoes, and Engineering (OEVE) and the office of the Director of OES in Sacramento. The USGS will also request that available field geologists go to the southern San Andreas fault to check for surface offsets and set baselines for measuring any future offsets.

## Level B

Level B means that the probability of a great earthquake occurring within 72 hours is on the order of 5-25%. The appropriate response to this level is preparation. In addition to the activities undertaken for levels D and C, the office chief of OEVE will also notify the Director of the USGS and the State Geologist of California. An intensive scientific monitoring effort will be undertaken, coordinated by the scientist-in-charge of the southern California office of the USGS.

The extent of the intensive monitoring effort will depend on the resources available at the time. The present plan calls for notifying the chief of the Branch of Engineering Seismology and requesting deployment around the southern San Andreas fault of several portable high dynamic range, digital seismic stations. In addition, all portable high gain and strong motion instruments available in southern California (at present 3 strong motion and 1 high gain portables) should be deployed. A geodetic resurvey of all geodetic nets on the southern San Andreas and deployment of portable GPS receivers will be requested.

## V. Need for Improved Instrumentation

In preparing this report, the Working Group was struck by the inadequacy of the information available from the southern San Andreas fault. Strain is recorded at only one site and creep at only 4 sites. Seismic station spacing is so sparse that the depths of most earthquakes cannot be resolved, and the dynamic range of the telemetered stations is so limited that earthquakes above about magnitude 3.5 are not recorded on scale. Analog telemetry so limits the dynamic range and bandwidth that questions about the spectral characteristics cannot be addressed. The data are recorded at many different sites with limited coordination between the different organizations. These inadequacies reduce the chance that a useful warning about the next great southern San Andreas earthquake will be issued and indeed raised doubts within the working group about the feasibility of even a simple alert system. However, the charge of the Working Group was "to recommend ways in which the scientific community might best keep abreast of the changing situation along the fault, increase its understanding of the regional seismotectonics, and offer appropriate scientific advice to local governmental agencies." Some system is necessary because even with the present inadequacies, scientific advice will be needed by local government. But to complete the full charge of the Working Group, we strongly recommend that the recording and analysis of geophysical data from the southern San Andreas fault be improved.

Earthquake precursors, especially foreshocks, can occur within a very short time, minutes to hours, before the mainshock. Thus, for information to be useful for short-term warnings, it must be immediately available to scientists; however, very few data in southern California are accessible in real time. Many of the recommendations below should improve the real-time flow of data to a central recording site.

Improving the quality of the data and not just its accessibility would also enhance our ability to make short-term hazard assessments in southern California. Almost all instrumentation near the southern San Andreas fault was installed in the 1970's. Since that time, both the instrument quality and the scientific understanding of data from those instruments have greatly improved. As seismology has developed, we have found that information beyond the fact of earthquake occurrence could be used to assess the likelihood that an earthquake is a foreshock to a great event. Immediate questions that arise include:

1. What are the time, location, depth, magnitude and focal mechanism of the potential foreshock event(s)?
2. On which fault did the potential foreshock occur?

3. Did the potential foreshock rupture toward or away from the San Andreas fault?
4. Did surface rupture take place?
5. Is creep or slip occurring above or below the seismogenic zone?
6. What were the dynamic and static stress drops of the potential foreshock?
7. Do continuous strain data suggest significant aseismic fault slip?
8. Where and when did triggered slip occur on nearby faults?

These questions must be answered within a few minutes or at least a few tens of minutes after the potential foreshock. Unfortunately, present instrumentation near the southern San Andreas fault and current scientific understanding of the tectonics and seismicity of the fault are inadequate to answer these questions accurately. Thus, the following sections discuss short-term and long-term improvements to the existing system to provide a more detailed analysis capability for this critical section of the San Andreas fault. Within each type of operation, the recommendations easiest to implement are listed first.

## V.1 Centralized Recording and Analysis

*Coordination and Response.* Because so many different organizations are involved in recording data in southern California, coordination and communication between the different groups has been limited. The present organization of the USGS in southern California provides no mechanism for undertaking the actions described in this report.

**Recommendation 1:** As an organizational first step, appoint a chief scientist for the southern San Andreas fault to coordinate response. This person would monitor ongoing seismic activity and coordinate scientific investigations as has been done for Parkfield and Mammoth Lakes. This task would include developing the scientific expertise needed for short term earthquake hazard assessment using both seismic and deformation data.

*Recording Center.* Some instruments presently in the area record data only on site. Just the Salt Creek and North Shore creepmeters in the Coachella Valley are telemetered (intermittently via satellite) to Pasadena and the data are not routinely available for real-time analysis. Similarly, numerous strain and tilt instruments at Pinon Flat and USGS dilatometers in the Mojave Desert are recorded locally. In some cases, data are transmitted to Menlo Park via satellite, and a simple E-mail command code would permit timely transmission of these data from Menlo Park to Pasadena.

A central recording site is urgently needed where the relevant creep and strain data may be analyzed in near real-time with the seismic data. Because the seismic data are recorded in Pasadena, this is a logical site for a southern California center. In many cases, personnel in Pasadena may not have the necessary expertise to evaluate the strain data, but they should be available for display to develop such expertise, and the necessary experts can be consulted over the telephone.

**Recommendation 2:** Install the necessary software and telemetry so that creep and strain data can be received and displayed in real-time in Pasadena. Begin with borehole strainmeter and air pressure data from Pinon Flat Observatory.

## V.2 Seismological Data

*Real-time Analysis.* At present, only a small subset of data is easily available in real-time in Pasadena from the southern California seismic network. A 64-channel real-time processor (RTP) is now used to determine real-time earthquake locations and magnitudes. Because signals from only 64 stations of the 280 stations now operating in southern California can be processed in real time, and the area being monitored is all of southern California, not all available stations along the southern San Andreas fault are utilized to calculate the location and magnitude of each earthquake. With this limitation, only about 25% of the network is being used to determine the locations, so that depths cannot be determined accurately; focal mechanisms are unreliable or indeterminate; and the location errors of the epicenters are large.

If a magnitude 5-6.5 earthquake were to occur near the southern San Andreas fault, the present system would provide an epicentral location accurate only to about 5 km. The depth and focal mechanism of the earthquake would not be known for at least one hour, perhaps much longer. It would also be difficult to monitor the spatial development of its aftershock sequence, because epicentral locations of low quality tend to smear over a large area. If data from all currently operating seismograph stations in southern California were analyzed by a RTP, then the uncertainty in the hypocenters could be reduced from approximately 5 km to 1-2 km, and focal mechanisms could be determined with reasonable accuracy.

**Recommendation 3:** Upgrade the real-time earthquake processing capability for southern California from 64 to all 256 seismic stations.

*Magnitudes.* The present RTP can determine duration magnitudes only up to about magnitude 4. This hardware limitation results from signal clipping associated with the exclusive use of high-gain seismographs.

**Recommendation 4:** Implement available methods to determine magnitudes of large earthquakes in real time, using force-balance accelerometers and low gain seismometers already in place.

*Station Density.* The spacing of high-gain, short period seismic stations in southern California is about 20 km. This spacing is inadequate for obtaining high quality hypocenters and for correlating hypocenters with the mapped trace of the San Andreas fault or nearby orthogonal faults. Currently no stations are located immediately west of the fault in the Coachella Valley sediments, where borehole installations would be required to avoid near-surface noise and attenuation. Data from new borehole stations would provide high quality hypocenters and source parameters, which, in turn, would allow monitoring of the stress around stuck patches of the fault (e. g., Malin *et al.*, 1989). Meaningful monitoring of rupture direction and migration of hypocenters would also become possible. With digital telemetry, these stations would have sufficient band-width for many waveform studies.

**Recommendation 5:** Upgrade the existing high-gain short period network by adding about 40 new three-component stations, some installed in boreholes for improved dynamic range. Data should be digitally transmitted for high fidelity signal recording.

### V.3 Strain and Creep Data

*Creep and Slip Data.* Following a magnitude 5 earthquake, geologists will drive to the Coachella Valley to look for surface rupture and triggered slip. They will require 2-4 hours (presuming no major traffic delays) to reach various stretches of the Coachella Valley segment by automobile from Pasadena. Creepmeters and slip meters could provide immediate information about surface fault displacement if they were installed with 5-10 km spacing across the San Andreas fault and nearby secondary faults and telemetered to the central facility.

**Recommendation 6:** Deploy an array of at least 20 (1 every 10 km) creep and slip meters along the southern San Andreas fault and candidate complementary faults. Data from these instruments should be telemetered using channels on the planned microwave link that will also transmit the data for the seismic network to Pasadena.

*Strain and Tilt Data.* No borehole strainmeter is currently deployed close to the southern San Andreas fault. The utility of strain measurements in any alarm system is greatly increased if the strainmeters are deployed at more than one site. Data from at least one additional borehole strainmeter near the San Andreas fault, in conjunction with PFO strain data and Salton Sea tilt data would greatly help us in determining alert thresholds. Obviously a number of borehole or long baseline strainmeters along the San Andreas fault, although perhaps outside the actual fault zone itself, would better define possible slip models than a single borehole strainmeter. These data may be acquired in a variety of ways. However, any installation of deformation-measuring instruments will require large capital costs and a long-term commitment to operations, so the task must be well organized and coordinated.

**Recommendation 7:** A group of university and USGS scientists should begin the planning for the establishment of deformation measuring instrumentation to monitor strain and tilt along the southern San Andreas fault. This plan should be coordinated with new seismic equipment for a balanced expenditure of funds and an integrated field program.

### V.4 Fundamental Understanding of the Southern San Andreas Fault

The above recommendations will improve the data available for estimating the short-term probability of a major earthquake, based on existing knowledge of the San Andreas fault and the behavior of past earthquakes. In addition, the improved understanding of the earthquakes, geologic history, and seismotectonics of the San Andreas fault expected to evolve from the improved data will improve our ability to use the data. The Working Group has found that many aspects of the southern San Andreas fault are not well understood and this impairs our ability to respond. We therefore recommend that more fundamental studies of the fault be carried out. These studies should include:

*Geodetic Measurements.* Because any earthquake is the result of a cycle of accumulated strain, measurements of the regional strain field and changes in that field are essential to a physical understanding of it. Measuring how the strain field close to the fault interacts with the more distant strain field (on both long and short time scales) is particularly important. At present, one large aperture and seven small aperture networks of geodetic monuments cross the southern San Andreas fault. The new satellite based measurements (GPS - Global Positioning System) are the most reliable and efficient system for regional geodetic measurements while traditional geodetic techniques are useful for smaller scale measurements.

**Recommendation 8:** Establish fixed networks of GPS receivers and augment the dense arrays of geodetic monuments to study strain buildup and release around the southern San Andreas fault.

*Improved Probability Estimates.* As Tables 1 and 2 show, the value for the long-term probability of a major earthquake is important in determining short-term probabilities after a potential foreshock. For the southern San Andreas fault, this long-term probability is extremely uncertain for two reasons. First, the geologic data applicable to this question are now limited to only one paleoseismic site. Also, there is currently disagreement (described above) on how long-term probability should be estimated from these data. These are not, however, the only factors that could be improved. We could also use information on how the frequency of foreshocks depends on both the variables we have used and on others (such as the focal mechanism) that we have not included.

**Recommendation 9:** Expand paleoseismic and geologic studies of the southern San Andreas fault to improve our estimates of the times and surface slip distributions of previous major earthquakes.

**Recommendation 10:** Continue research on the best methods for determining long-term probabilities of major earthquakes from limited data on recurrence times of previous earthquakes. Develop more complete data sets for foreshocks, and improved ways to examine their statistics.

*General Seismological Studies.* In addition to a dense short-period network, broad-band, high-dynamic range seismometers provide detailed information, especially about the spectrum of an earthquake, to study its physics. Studies of dynamic and static stress drops around asperities on faults, combined with high quality hypocenters from the high-gain downhole network recommended above, are promising research areas in fault zone physics.

**Recommendation 11:** Install several wide dynamic range, broad-band seismometers in southern California and use their data to study source and path effects.

We feel that relatively inexpensive options should be implemented quickly (Recommendations 1, 2, 3, 4 and 10). If additional funding were to become available for operations along the southern San Andreas fault, a reasoned, careful approach should be undertaken to make the most cost-effective use of those funds.



## References

- Allen, R. V., 1982, Automatic phase pickers: Their present use and future prospects, *Bull. Seismol. Soc. Amer.*, **72**, S225-S242.
- Bakun, W. H., and A. G. Lindh, 1985, The Parkfield, California, earthquake prediction experiment, *Science*, **229**, 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.* 86-365, 59 pp.
- Bilham, R., and G. King, 1989, The morphology of strike-slip faults: Examples from the San Andreas fault, California, *J. Geophys. Res.*, **89**, 10,204-10,216.
- Cohn, S. N., C. R. Allen, R. Gilman, and N. R. Goulty, 1982, Pre-earthquake and post-earthquake creep on the Imperial fault and Brawley Seismic Zone, in *The Imperial Valley, California Earthquake of October 15, 1979*, U.S. Geol. Surv. Prof. Paper 1254, U.S. Government Printing Office, 183-191.
- Davis, P. M., D. D. Jackson, and Yan Y. Kagan, 1989, The longer it has been since the last earthquake, the longer it will be to the next?, *Bull. Seismol. Soc. Amer.*, **79**, 1439-1456.
- Goltz, J., 1985, The Parkfield and San Diego earthquake predictions: a chronology., Special Report by the Southern California Earthquake Preparedness Project, Los Angeles, CA, 23 pp..
- Hudnut, K. W., L. Seeber, and J. Pacheco, 1989, Cross fault triggering in the November 1987 Superstition Hills Earthquake Sequence, southern California, *Geophysical Research Letters*, **16**, 199-202.
- Johnson, C. E., and D. P. Hill, 1982, Seismicity of the Imperial Valley, in *The Imperial Valley, California Earthquake of October 15, 1979*, U.S. Geol. Surv. Prof. Paper 1254, U.S. Government Printing Office, 15-24.
- Jones, L. M., 1984, Foreshocks (1966-1980) in the San Andreas system, California, *Bull. Seismol. Soc. Amer.*, **74**, 1361-1380.
- Jones, L. M., 1985, Foreshocks and time-dependent earthquake hazard assessment in southern California, *Bull. Seismol. Soc. Amer.*, **75**, 1669-1680.
- Jones, L. M., 1988, Focal mechanisms and the state of stress on the San Andreas fault in southern California, *J. Geophys. Res.*, **93**, 8869-8891.
- Jones, L. M., and P. Molnar, 1979, Some characteristics of foreshocks and their possible relationship to earthquake prediction and premonitory slip on faults, *J. Geophys. Res.*, **84**, 3596-3608.
- King, N. E., and J. C. Savage, 1984, Regional deformation near Palmdale, California, 1973 - 1983, *J. Geophys. Res.*, **89**, 2471-2477.
- Lorenzetti, E., and T. E. Tullis, 1989, Geodetic predictions of a strike-slip model: Implications for intermediate- and short-term earthquake prediction, *J. Geophys. Res.*, **94**, 12,343 -12,361.
- Louie, J. N., C. R. Allen, D. C. Johnson, P. C. Haase, and S. N. Cohn, 1985, Fault slip in southern California, *Bull. Seismol. Soc. Amer.*, **75**, 811-833.
- Malin, P. E., S. N. Blakeslee, M. G. Alvarez, and A. G. Martin, 1989, Microearthquake imaging of the Parkfield asperity, *Science*, **244**, 557-559.

- Rudnicki, J. W., 1988, Physical models of earthquake instability and precursory processes, *Pure and Appl. Geophys.*, **126**, 531-554.
- Savage, J. C., W. H. Prescott, and G. H. Gu, 1986, Strain accumulation in southern California, 1973-1984, *J. Geophys. Res.*, **91**, 7455-7474.
- Sharp, R. V., K. E. Budding, J. Boatwright, M. J. Ader, M. G. Bonilla, M. M. Clark, T. E. Fumal, K. K. Harms, J. J. Lienkamper, D. M. Morton, B. J. O'Neill, C. L. Ostergren, D. J. Ponti, M. J. Rymer, J. L. Saxton, and J. D. Sims, 1989, Surface faulting along the Superstition Hills fault zone and nearby faults associated with the earthquakes of 24 November 1987, *Bull. Seismol. Soc. Amer.*, **79**, 252-281.
- Sieh, K. E., 1986, Slip rate across the southern San Andreas and prehistoric earthquakes at Indio, California, *Trans. Amer. Geophys. U.*, **67**, 1200.
- Sieh, K. E., and P. L. Williams, 1990, Behavior of the southernmost San Andreas fault in the past 300 years, *J. Geophys. Res.*, **95**, 6629-6645.
- Stuart, W. D., 1986, Forecast model for large and great earthquakes in southern California, *J. Geophys. Res.*, **91**, 13771-13786.
- Wallace, R. E., and E. F. Roth, 1967, The Parkfield-Cholame California earthquakes of June-August 1966: Rates and patterns of progressive deformation, *U. S. Geol. Surv. Prof. Pap.* **579**.
- Williams, P. L., McGill, S. F., Sieh, K. E., Allen, C. R., and J. N. Louie, 1988, Triggered slip along the San Andreas fault after the 8 July 1986 North Palm Springs earthquake, *Bull. Seismol. Soc. Amer.*, **78**, 1112-1122.
- Working Group on California Earthquake Probabilities, 1988, (D. C. Agnew, C. R. Allen, L. S. Cluff, J. H. Dieterich, W. L. Ellsworth, R. L. Keeney, A. G. Lindh, S. P. Nishenko, D. P. Schwartz, K. E. Sieh, W. Thatcher, and R. L. Wesson), Probabilities of large earthquakes occurring in California on the San Andreas fault, *U.S. Geol. Surv. Open-file Rep.* **88-398**, 62 pp.

United States Geological Survey  
525 S. Wilson Avenue  
Pasadena, CA 91106

January 14, 1991

**Memorandum**

To: Southern San Andreas Working Group  
From: Lucy Jones *Lucy*  
Subject: The Working Group Report

After almost two years since the first meeting of our working group, our report has been open-filed. I am sending you the final version of the report with this memo. I have not made too many changes since the version I sent you in April. Some minor changes were suggested by the USGS reviewer (Paul Reasenber), the chairmen of NEPEC and CEPEC have written letters of endorsement and Dick Andrews requested that we remove terminology about alerts because (in 25 words or less) the USGS determines the hazard and OES alerts the public. We now have a system of hazard levels to be declared.

Thus our task as a working group, to make recommendations to the USGS, has been completed. The next step is for Dr. Peck to accept the recommendations, after which we will officially begin operating the hazard level system in Pasadena.

Two aspects of the hazard level system, once it becomes operational, may be of direct interest to you. If you remember, we recommended that an electronic mail system be devised to notify all researchers working on the southern San Andreas fault when a hazard level has been declared. If you wish to be one of those notified, please send me your electronic mail address. You will receive any notification of an official hazard level and may also, if you wish, receive automatic messages about earthquakes near the San Andreas fault at lower magnitude levels. The second aspect is for those of you recording creep or strain data near the southern San Andreas fault. Please contact me to make arrangements for communicating information about possible anomalies.

Thank you for all of the work you put into the working group. I think our work will make a difference.

**Appendix A**  
**Prediction Probabilities from Foreshocks**

Duncan Carr Agnew  
Institute for Geophysics and Space Physics  
University of California  
La Jolla, CA 92092

and

Lucile M. Jones  
U. S. Geological Survey  
525 S. Wilson Avenue  
Pasadena, CA 91106

submitted to  
Journal of Geophysical Research  
November 1990

# Prediction Probabilities From Foreshocks

DUNCAN CARR AGNEW

*Institute of Geophysics and Planetary Physics,  
University of California, La Jolla*

LUCILE M. JONES

*U.S. Geological Survey, Pasadena, California*

When any earthquake occurs, the possibility that it might be a foreshock increases the probability that a larger earthquake will occur nearby within the next few days. Clearly, the probability of a very large earthquake ought to be higher if the candidate foreshock were on or near a fault capable of producing that very large mainshock, especially if the fault is towards the end of its seismic cycle. We derive an expression for the probability of a major earthquake characteristic to a particular fault segment, given the occurrence of a potential foreshock near the fault. To evaluate this expression, we need: (1) the rate of background seismic activity in the area, (2) the long-term probability of a large earthquake on the fault, and (3) the rate at which foreshocks precede large earthquakes, as a function of time, magnitude, and spatial location. For this last function we assume the average properties of foreshocks to moderate earthquakes in California: (1) the rate of mainshock occurrence after foreshocks decays roughly as  $t^{-1}$ , so that most foreshocks are within three days of their mainshock, (2) foreshocks and mainshocks occur within 10 km of each other, and (3) the fraction of mainshocks with foreshocks increases linearly as the magnitude threshold for foreshocks decreases, with 50% of the mainshocks having foreshocks with magnitudes within three units of the mainshock magnitude (within three days). We apply our results to the San Andreas, Hayward, San Jacinto, and Imperial faults, using the probabilities of large earthquakes from the report of the Working Group on California Earthquake Probabilities (1988). The magnitude of candidate event required to produce a 1% probability of a large earthquake on the San Andreas fault within three days ranges from a high of 5.3 for the segment in San Geronio Pass to a low of 3.6 for the Carrizo Plain.

Probably the most evil feature of an earthquake is its suddenness. It is true that in the vast majority of cases a severe shock is heralded by a series of preliminary shocks of slight intensity .... [but] only after the havoc has been wrought does the memory recall the sinister warnings of hypogene action.

C. G. Knott (1908, p. 10)

## 1. INTRODUCTION

Many damaging earthquakes have been preceded by smaller earthquakes that occur within a few days and a few kilometers of the mainshock [e.g., Jones and Molnar, 1979]; these are referred to as immediate foreshocks. If such foreshocks could be recognized before the mainshock, they would be very effective for short-term earthquake prediction; but so far no way has been found to distinguish them from other earthquakes. Even without this, the mere existence of foreshocks provides some useful predictive capacity. When any earthquake occurs, the possibility that it might be a foreshock increases the probability that a larger earthquake will soon happen nearby. For southern California, Jones [1985] showed that after any earthquake there is a 6% probability that a second one equal to or larger than the first will follow within five days and 10 km of the first. The probability is much lower for a second earthquake much larger than the first; for example, the probability of an earthquake two units of magnitude larger is only 0.2%. Using

these results, the U.S. Geological Survey has issued four short-term earthquake advisories after moderate earthquakes [e.g., Goltz, 1985]. A more recent study by Kagan and Knopoff [1987] developed a model for the clustering of earthquakes which could indicate areas of space and time in which larger events might follow smaller ones. The size of these areas depended on the probability gain, the ratio of probability of an earthquake given the occurrence of a possible precursor (such as a foreshock) to the probability in the absence of such a precursor [Kagan and Knopoff, 1977; Vere-Jones, 1978; Aki, 1981]. For low levels of probability gain, Kagan and Knopoff [1987] found that one-third of all earthquakes with magnitudes 4 and above fell within their predicted regions.

These results are from studies of earthquake catalogs; Jones [1985] used a catalog for southern California, and Kagan and Knopoff [1987] used one for central California. As a consequence, both papers give generic results about pairs of earthquakes, without much regard for other factors. But it ought to be possible to do better: the probability of a very large earthquake should be higher if the candidate foreshock were to occur near a fault capable of producing that mainshock than if it were located in an area where we believe such a mainshock to be very unlikely. Moreover, the chance of a candidate earthquake actually being a foreshock should be higher if the rate of background (nonforeshock) activity were low.

In this study we derive an expression for the probability of a major earthquake following a possible foreshock near a major fault from the basic tenets of probability theory. 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. We then apply this expression to

the San Andreas fault system to develop short-term probabilities for possible earthquake warnings based on possible foreshocks.

## 2. MODELS FOR PROBABILITIES FROM FORESHOCKS

Because of the nature of seismicity along major fault systems such as the San Andreas fault, we have been led to address certain fundamental issues about the relationship between foreshocks and large earthquakes. These major faults illustrate in an extreme form the "maximum magnitude" model introduced by *Wesnowsky et al.* [1983], in which the frequency of the largest earthquakes on a fault zone is much higher than would be predicted by the extrapolation of the frequency-magnitude distribution for background earthquakes. For many parts of the San Andreas fault this is a straightforward consequence of the low level of present-day seismicity. For instance, along the Coachella Valley segment of the San Andreas fault (Figure 4) an extrapolation of present seismicity to higher magnitudes predicts a magnitude 7.5 earthquake every 2900 years, whereas the recurrence rate estimated from slip-rate data is 200–300 years.

This behavior implies that the large characteristic earthquakes on a fault zone are not simply the largest members of the total population of earthquakes there, but are somehow derived from a different population. Foreshocks to such events can thus reasonably be regarded as also being a separate class of events from the background earthquakes. A physical model that might underlie this is that some special failure process takes place before characteristic earthquakes, with an enhanced rate of small earthquakes and eventual failure on a large scale both being a result of it. It is of course also possible that no such process occurs; a moderate shock might, depending on the details of stress nearby, trigger only smaller events (in which case it is a mainshock) or larger ones (making it a foreshock), as suggested by *Brune* [1979]. There would then be no innate difference between background events and foreshocks; but we believe that it remains fruitful (as will be shown) to make at least a conceptual division.

That we make this division does not mean that there are any characteristics that can distinguish between foreshocks and other earthquakes; indeed, if there were, we would not have had to consider the second model above. We can only identify foreshocks, like aftershocks, by virtue of their association with a larger event; and, as our opening quotation suggests, for foreshocks such identification can only be retrospective. Such classification by association means that any particular shock might have been classified "incorrectly", and actually have been a background shock that just happened to fall close to a larger event. In our present state of knowledge this is unavoidable, and it may always remain so.

### 2.1. Zero-Dimensional Model

Starting from the assumption that foreshocks are a separate class of earthquake from background earthquakes, we can set out a formal probabilistic scheme for finding the probability of a large shock, given the occurrence of a possible foreshock. For clarity we begin with a "zero-dimensional" model, ignoring spatial variations, magnitude dependence, and other complications, which will be added in later sections. With these simplifications, a numerical example will illustrate the reasoning. Suppose that mainshocks occur every 500 years (on average), and that half of them have foreshocks (defined as being

within a day of the mainshock); then we expect a foreshock every 1000 years. If a comparable background earthquake occurs, on average, annually, we get 1000 background earthquakes per foreshock. If an earthquake occurs that could be either one, we then would assume the probability to be  $10^{-3}$  that it is a foreshock, and so will be followed by a mainshock within a day. This is low, but still far above the background one-day probability of  $5.5 \times 10^{-6}$ .

For a formal treatment we begin by defining events (in the probability-theory meaning of the term):

$B$ : A background earthquake has occurred.

$F$ : A foreshock has occurred.

$C$ : A large (characteristic) earthquake will occur.

As noted above, if a small background shock were to happen by coincidence just before the characteristic earthquake, we would certainly class it as a foreshock. Thus,  $B$  and  $C$  cannot occur together: they are disjoint. The same holds true for  $B$  and  $F$ : we can have a foreshock or a background earthquake, but not both.

The probability that we seek is the conditional one of  $C$ , given either  $F$  or  $B$ , because we do not know which has occurred. This is, by the definition of conditional probability,

$$P(C|F \cup B) = \frac{P(C \cap (F \cup B))}{P(F \cup B)} \quad (1)$$

Because  $F$  and  $B$  are disjoint, the probability of their union is the sum of the individual probabilities, allowing us to write the numerator of (1) as

$$P((C \cap F) \cup (C \cap B)) = P(C \cap F) + P(C \cap B) = P(C \cap F)$$

where the disjointness of  $C$  and  $B$  eliminates the  $P(C \cap B)$  term. From the definition of conditional probability,

$$P(C \cap F) = P(F|C)P(C)$$

where  $P(F|C)$  is the probability that a mainshock is preceded by a foreshock. Again using the disjointness of  $F$  and  $B$ , we can write the denominator as

$$P(F \cup B) = P(F) + P(B) \quad (2)$$

Because a foreshock cannot, by definition, occur without a mainshock, the intersection of  $C$  and  $F$  is  $F$ , and therefore

$$P(F) = P(F \cap C) = P(F|C)P(C) \quad (3)$$

We can use (2) and (3) to write (1) as

$$P(C|F \cup B) = \frac{P(F)}{P(F) + P(B)} = \frac{P(C)P(F|C)}{P(F|C)P(C) + P(B)} \quad (4)$$

For  $P(B) \gg P(F|C)P(C)$  this expression is small (the candidate event is probably a background earthquake), while for  $P(B) = 0$ , the expression becomes equal to one: any candidate earthquake must be a foreshock.

The second form of expression in (4) is a function of three quantities, which in practice we obtain from very different sources.  $P(B)$ , the probability of a background earthquake, would be found from seismicity catalogs for the fault zone.  $P(C)$ , the probability of a characteristic earthquake, would be

found from calculations of the type presented by the *Working Group on California Earthquake Probabilities* [1988]. If we had a record of the seismicity before many such characteristic earthquakes, we could evaluate  $P(F|C)$  (which we shall hereafter call  $\Phi_{FC}$ ) from it directly. (For this simple model,  $\Phi_{FC}$  is just the fraction of large earthquakes preceded by foreshocks.) Of course, we do not have such a record, and so are forced to make a kind of reverse ergodic assumption, namely that the time average of  $\Phi_{FC}$  over many earthquakes on one fault is equal to the spatial average over many faults. This may not be true, but it is for now the best we can do.

## 2.2. One-Dimensional Model

As a simple extension to the previous discussion, suppose that we have  $N$  "regions" and that  $C_i$ ,  $B_i$ , and  $F_i$  denote the occurrence of an event in the  $i$ th region, with  $C$  (for example) now being the occurrence of a large earthquake in any possible region. These regions can be sections of the fault or (as we will see below) volumes in a multidimensional space of all relevant variables. The quantity of interest is now  $P(C|F_i \cup B_i)$ : we have a candidate foreshock in one region, and want the probability of a large earthquake starting anywhere. Assuming that the occurrences  $C_i$  are disjoint (the epicenter can only be in one place), we then have that the probability of a foreshock in the  $i$ th region can be written as

$$P(F_i) = \sum_{j=1}^N \Phi_{FC}(i, j) P(C_j) \quad (5)$$

where  $\Phi_{FC}(i, j) = P(F_i|C_j)$ . We may regard  $\Phi_{FC}$  as the probability of a foreshock in region  $i$  given a large earthquake in region  $j$ . We call this the precurent probability because it refers to the probability of an event preceding a second one (not, it should be noted, with an implication of violated causality). As a simple example, we could take  $\Phi_{FC}(i, j) = \alpha \delta_{ij}$ , which would imply that large earthquakes are preceded by foreshocks only in the same region, and even then only a fraction  $\alpha$  of them have foreshocks at all.

We can then easily revise (4) above to get the probability we seek; simply adding subscripts to the candidate event yields

$$P(C|F_i \cup B_i) = \frac{P(F_i) + P(C)P(B_i)}{P(B_i) + P(F_i)} \quad (6)$$

Equations (5) and (6) are the basic ones we shall use in the more general case. Equation (5) shows us how to compute the probability of a foreshock happening in the location of our candidate earthquake, by summing over all possible mainshocks. The use of the precurent probability  $\Phi_{FC}$  is the key to this approach; we can (and in the next section shall) design it to embody our knowledge and assumptions about the relation between foreshocks and the earthquakes they precede. Having found the foreshock probability, we then use (6) to find the conditional probability of a large earthquake.

An important consequence of (5) is that we may sum over all possible foreshocks (again assuming disjointness) to get

$$P(F) = \sum_{i=1}^N \sum_{j=1}^N \Phi_{FC}(i, j) P(C_j) \quad (7)$$

giving us the overall probability of a foreshock somewhere in the total region. This must satisfy  $P(F) = \alpha P(C)$ , where  $\alpha$  is

the fraction of mainshocks with foreshocks; this and equation (7) together constrain the normalization of  $\Phi_{FC}$ .

Next to the probability level itself, the socially most interesting quantities would seem to be the chance of an alert being a false alarm, and the rate at which false alarms occur for a given probability level. The probability that an alert is a false alarm is  $P(\bar{C}|F_i \cup B_i)$ , which is just  $1 - P(C|F_i \cup B_i)$ : if we have a 10% chance of having a mainshock, we have a 90% chance of not having one. The rate of false alarms is equivalent to the probability of a false alarm happening in some given time, and this is just the probability that an alert is a false alarm times the probability of the event that triggers it, namely

$$\sum_{i=1}^N [1 - P(C|F_i \cup B_i)] [P(B_i) + P(F_i)]$$

As will be shown in section 4, we would in practice usually choose the probability of a mainshock given a small event,  $P(C|F_i \cup B_i)$  to have a fixed value (e.g., 1%), which we denote by  $S$ , for all regions. This value of  $S$  then sets the value of  $P(B_i)$  for the  $i$ th region; from (6),  $P(B_i) = P(F_i)[(1 - S)/S]$ , which makes the probability of a false alarm

$$\frac{(1 - S)}{S} \sum_{i=1}^N P(F_i) = \frac{(1 - S)}{S} \sum_{i=1}^N \sum_{j=1}^N \Phi_{FC}(i, j) P(C_j)$$

where we have used (5). For fixed  $S$  and  $\Phi_{FC}$  this expression is proportional to  $P(C_j)$  only: the rate of false alarms for a given probability depends only on the rate of mainshocks and not on the rate of background activity. In terms of the simple example at the beginning of section 2.1, fixing a probability level of 0.1% means that we would set the magnitude level of candidate events such that there would be 1000 background events for each actual foreshock; but the absolute rate of such background earthquakes (and thus of false alarms) is then determined only by the rate of foreshocks, and thus of mainshocks.

## 3. A MULTIDIMENSIONAL MODEL FOR FORESHOCKS

We now develop an expanded version of (5), which contains more variables. The first step is to define our events more thoroughly:

- $B$ : A background earthquake has occurred at coordinates  $(x_0 \pm e_0, y_0 \pm e_0)$ , during the time period  $[t, t + \delta_0]$ , with magnitude  $M \pm \mu$ . (All of the quantities  $e_0$ ,  $\delta_0$ , and  $\mu$  are small and are included because we will be dealing with probability density functions; as will be seen below, they cancel from the final expression).
- $F$ : A foreshock has occurred, with the same parameters as in event  $B$ .
- $C$ : A major earthquake will occur somewhere in the region of concern, which we denote by  $A_C$  (also using this variable for the area of this region). This earthquake will happen during the time period  $[t + \Delta, t + \Delta + \delta_1]$ , with magnitude between  $M_C$  and  $M_C + \mu_C$ .

We assume that we are computing the probability at some time in the interval  $(t + \delta_0, t + \Delta)$ ; the possible foreshock has happened, but the predicted mainshock is yet to come.

### 3.1. Rate Densities of Earthquake Occurrence

We begin by defining a rate of occurrence for the background seismicity (in the literature on point processes this would be

called an intensity, a term we avoid because of existing seismological usage). This rate (or, strictly speaking, rate density) we call  $\Lambda(x, y, M)$ ; it is such that the probability of  $B$  is

$$P(B) = \delta_0 \int_{x_0-\epsilon_0}^{x_0+\epsilon_0} dx \int_{y_0-\epsilon_0}^{y_0+\epsilon_0} dy \int_{M-\mu}^{M+\mu} dM \Lambda(x, y, M) \quad (8)$$

By not making  $\Lambda$  dependent on the time  $t$  we make the occurrence of background earthquakes into a Poisson process. If we assume that at any location the Gutenberg-Richter frequency-magnitude relation holds, we may write

$$\Lambda(x, y, M) \equiv \Lambda_s(x, y) e^{-\beta(x, y)M} \quad (9)$$

where  $\beta$  is 2.3 times the usual  $b$  value. (While common rather than natural logarithms are conventional in this area, they lead to messier expressions, and we have therefore not used them). If  $\beta$  is constant over a region of area  $A$ , and during a time interval  $T$  the cumulative number of earthquakes of magnitude  $M$  or greater is given by the usual formula

$$N(M) = 10^{a-bM} \quad (10)$$

then, since the expected value of  $N(M)$  is

$$E[N(M)] = T \int_M^\infty dM \iint_A dx dy \Lambda_s(x, y) e^{-\beta M} \quad (11)$$

we have that  $\Lambda_s = (10^a \beta)/(AT)$  for  $\Lambda_s$  constant within the region.

Similarly, we can define a rate density for the occurrence of large earthquakes,

$$\Omega(x, y, M, t) \equiv \Omega_s(x, y, t) e^{-\beta'(x, y)M} \quad (12)$$

where  $\beta'$  is 2.3 times the  $b$  value for these events. In this case, we introduce a dependence on time  $t$  because the occurrence of large earthquakes is often formulated as a renewal process [e.g., Nishenko and Buland, 1987], with time being measured relative to the last earthquake. The probability of  $C$  is then

$$P(C) = \delta_1 \iint_{A_C} dx dy \int_{M_C}^{M_C+\mu_C} dM \Omega_s(x, y, M, t+\Delta) \quad (13)$$

where  $A_C$  is the area of concern, i.e., the particular segment of a fault.

For lack of better information we would usually take  $\Omega_s$  to be a constant, but we could choose to make it spatially varying. Such variation could include increases near fault jogs and terminations if we think that rupture nucleation is more likely there, or a proportionality to  $\Lambda_s$  if we suspect that background earthquakes are (on the average) the likely triggers of large ones (both issues are discussed in section 3.2.2). For  $\Omega_s$  constant, we have that

$$\Omega_s = \frac{P(C)\beta'}{A_C \delta_1 e^{-\beta' M_C} (1 - e^{-\beta' \mu_C})} \quad (14)$$

Note that while we have regarded both  $A$  and  $A_C$  as two-dimensional regions (and hence also as the areas of such

regions), we may in fact make them three-dimensional or one-dimensional if we so choose, making sure that we adjust the numbers of the integrals in (8) and (13) accordingly. The one-dimensional model is easiest to develop analytical expressions for, and may be an adequate approximation for the case of a long fault zone. In this case, of course, we need to project the background seismicity (out to some distance away) onto the fault zone.

### 3.2. Computation of the Foreshock Probability

We are now in a position to write the formal expression for the foreshock probability  $P(F)$  in the same way as was done in (5) for the discrete one-dimensional case. In this case,  $\Phi_{FC}$  becomes a density function over all the variables involved, its value indicating with what relative frequency foreshocks with different parameters occur before mainshocks with particular ones. Instead of a single sum, as in (5), we have a multiple integral:

$$P(F) = \int_t^{t+\delta_0} dt \int_{x_0-\epsilon_0}^{x_0+\epsilon_0} dx \int_{y_0-\epsilon_0}^{y_0+\epsilon_0} dy \int_{M-\mu}^{M+\mu} dM \int_{t+\Delta}^{t+\Delta+\delta_1} dt' \iint_{A_C} dx' dy' \int_{M_C}^{M_C+\mu_C} dM' \Phi_{FC}(t, t', x, y, x', y', M, M') \Omega_s(x', y', t') e^{-\beta'(x', y')M'} \quad (15)$$

Of these eight integrals, the last four are the integration of the precurent probability density times the density of mainshock occurrence over the space of possible mainshocks and are the equivalent of the sum in (5). But this gives only the rate density for foreshocks, which must in turn be integrated over the space of the candidate event (the first four integrals) to produce the actual probability  $P(F)$ .

Equation (15) is clearly quite intractable as it stands. To render it less so, we assume that we can separate the behaviors of  $P(F)$  in time, magnitude, and location. This implies the following assumptions:

- 1:  $\beta'$  does not depend on  $x'$  or  $y'$ .
- 2: Over the range of integration,  $\Omega_s$  does not depend on  $t'$ .
- 3: The functional forms of the precurent probability density for time, space, and magnitude are independent, so that we can write the function as the product of the marginal distributions:

$$\Phi_{FC} = \Phi_s(x, y, x', y') \Phi_t(t, t') \Phi_m(M, M')$$

Of these assumptions, the third seems the least likely to be valid, since the dependence on both distance and time might be correlated with the magnitude of either the mainshock or the candidate foreshock. The most likely correlation, with mainshock magnitude, does not matter very much, since our range of integration of this variable is small.

These assumptions made, we can divide the integral in (15) into a product of three integrals (in space, time, and magnitude):

$$P(F) = \int_t^{t+\delta_0} dt \int_{t+\Delta}^{t+\Delta+\delta_1} dt' \Phi_t(t, t') \int_{M-\mu}^{M+\mu} dM \int_{M_C}^{M_C+\mu_C} dM' \Phi_m(M, M') e^{-\beta' M'} \int_{x_0-\epsilon_0}^{x_0+\epsilon_0} dx \int_{y_0-\epsilon_0}^{y_0+\epsilon_0} dy \iint_{A_C} dx' dy' \Phi_s(x, y, x', y') \Omega_s(x', y') \quad (16)$$



### 3.3. Functional Forms for the Foreshock Density

To evaluate the integrals in (16), we need to know the three precurrent probability densities  $\Phi_t$ ,  $\Phi_s$ , and  $\Phi_m$ . Our expressions for these incorporate our knowledge and assumptions about foreshocks. In the following sections, we describe in some detail what is known about the temporal, spatial, and magnitude dependences of foreshocks. From these data, we find functions for the relevant  $\Phi$ ; these functions must include both the actual dependence on the variables and a normalization. The nature of the normalization can be seen if we imagine extending the range of the first four integrals in (15) to cover all possible foreshocks (however we chose to define them); the resulting  $P(F)$  must then be equal to  $\alpha P(C)$ , where  $\alpha$  is, as for the one-dimensional model, the fraction of mainshocks preceded by foreshocks. In deriving our expressions we have aimed for simplicity rather than attempting to find a function that can be shown to be statistically optimal.

**3.3.1 Time.** Most foreshocks occur just before the mainshock. An increase in earthquake occurrence above the background rate has only been seen for a few days [Jones, 1984; 1985; Reasenberg, 1985] to a week [Jones and Molnar, 1979] before mainshocks. For 26% of Californian mainshocks, the foreshocks are most likely to occur within 1 hour of the mainshock; the rate of foreshock occurrence before mainshocks (Figure 1) varies with the  $t^{-1}$  type behavior also seen in Omori's law for aftershocks [Jones, 1985; Jones and Molnar, 1979]. This variation can be well fit by the function that Reasenberg and Jones [1989] found for California aftershock sequences:

$$\Phi_t(t, t') = \frac{N_t}{t' - t + c} \quad (17)$$

where  $t$  is the foreshock time and  $t'$  the mainshock time;  $c$  is a constant, found by Reasenberg and Jones [1989] to be 200 s for aftershocks. The relevant integral from (16) is then

$$\int_t^{t+\delta_0} dt \int_{t+\Delta}^{t+\Delta+\delta_1} dt' \Phi_t(t, t') = \delta_0 N_t \ln[1 + \delta_1/(\Delta + c)] \quad (18)$$

where we have assumed that  $\delta_0$  (the uncertainty of the time of

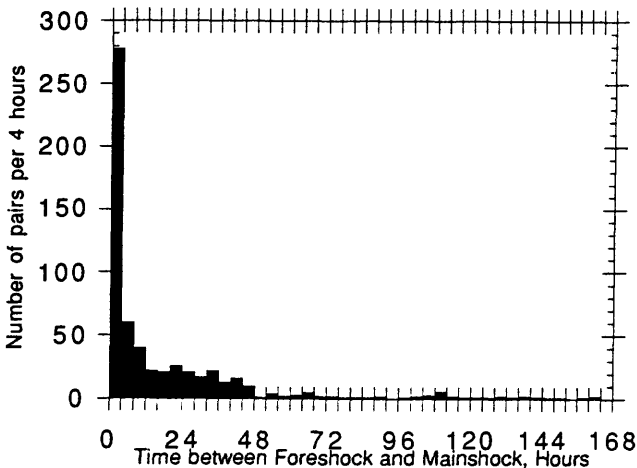


Fig. 1. The number of foreshock-mainshock pairs recorded in southern California versus the time between foreshock and mainshock in hours for foreshocks  $M \geq 2.0$  and mainshocks  $M \geq 3.0$  recorded between 1932 and 1987.

the candidate earthquake) is small. The normalization is determined by the requirement that

$$\int_{t+\delta_0}^{t+\delta_0+t_w} dt' \Phi_t(t, t') = 1 \quad (19)$$

where  $t_w$  is the total time window within which we admit preceding earthquakes to be foreshocks. This then gives

$$\begin{aligned} \int_t^{t+\delta_0} dt \int_{t+\Delta}^{t+\Delta+\delta_1} dt' \Phi_t(t, t') &= \delta_0 \frac{\ln[1 + \delta_1/(\Delta + c)]}{\ln[1 + t_w/(\delta_0 + c)]} \\ &\equiv \delta_0 I_t(\Delta, \delta_1) \end{aligned} \quad (20)$$

where, with an eye to future simplifications, we have separated out the  $\delta_0$  term. Note that (17) predicts a finite rate for all times, whereas the assumption of a limited time window automatically forces the rate to fall to zero beyond some time; we can easily modify  $\Phi_t$  to allow for this.

**3.3.2 Location.** Foreshocks not only occur close in time to the mainshock, but are also nearby in space. Jones and Molnar [1979] found that epicenters of mainshocks ( $M \geq 7$ ) and their foreshocks in the National Earthquake Information Center (NEIC) catalog were almost all within 30 km of each other, approximately the location error for the NEIC catalog. Jones [1985], with the more accurate locations of the California Institute of Technology (Caltech) catalog, found that epicenters of mainshocks ( $M \geq 3$ ) and their foreshocks were almost all within 10 km of each other; this result also held for foreshocks of  $M \geq 5$  mainshocks within the San Andreas system [Jones, 1984] if relative relocations were used. Even the largest foreshocks ( $M \geq 6$  at Mammoth Lakes and Superstition Hills) have had epicenters within 10 km of the epicenters of their mainshocks.

We have assembled a data set of sequences with high-quality locations to examine the dependence of the distance between foreshocks and mainshocks on the magnitudes of the earthquakes. This data set includes all foreshock-mainshock pairs with  $M_{\text{fore}} \geq 2.5$  and  $M_{\text{main}} \geq 3.0$  recorded in southern California since 1977 (the start of digital seismic recording), and several sequences relocated in special studies, with relative location accuracy of at least 1 km. Figure 2 shows the distance between foreshock and mainshock versus magnitude of the mainshock (2a) and magnitude of the foreshock (2b). The epicentral separation between foreshock and mainshock does not correlate strongly with either magnitude. Rather, the data seem to group into two classes: foreshocks that are essentially at the same site as their mainshock ( $< 3$  km) and foreshocks that are clearly separated from their mainshocks. Only foreshocks to larger mainshocks ( $M_{\text{main}} \geq 5.0$ ) occur at greater epicentral distances (5–10 km). Of these spatially separate foreshocks some (but not all) ruptured towards the epicenter of the mainshock (the rupture zones are shown by the ovals in Figure 2). The greatest reported distance between foreshock and mainshock epicenters is 8.5 km; the greatest reported distance between foreshock rupture zone and mainshock epicenter is 6.5 km. It would therefore seem that, whatever other behavior  $\Phi_s$  may have, it can be taken to be zero for distances greater than 10 km.

It is possible (and allowed for in our choice of variables for  $\Phi_s$ ) for foreshocks to be preferentially located in some sections

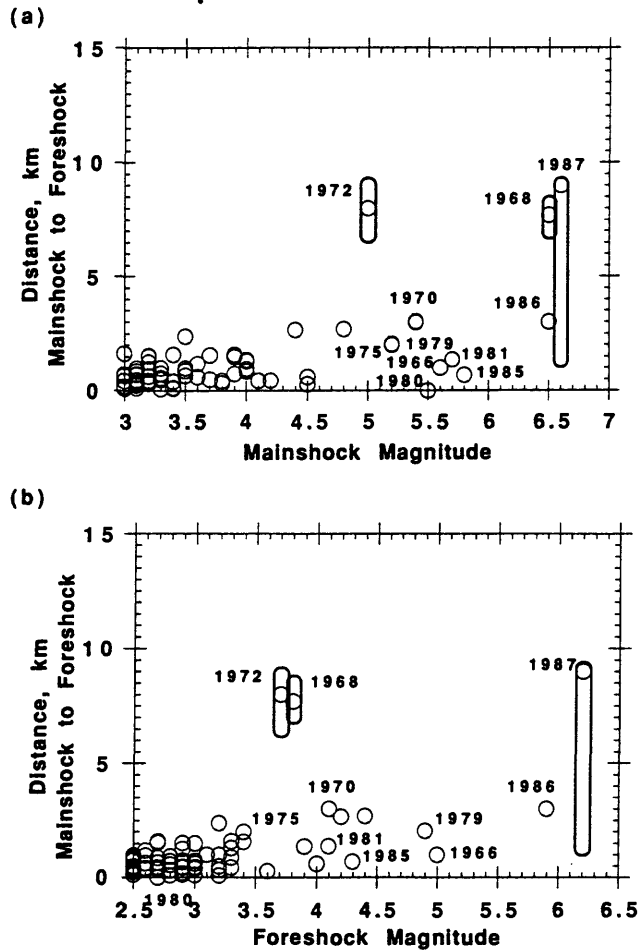


Fig. 2. Distance between foreshock and mainshock epicenters versus the (a) magnitude of the mainshock and (b) magnitude of the foreshock for foreshock-mainshock sequences (foreshocks  $M \geq 2.5$  and mainshocks  $M \geq 3.0$ ) recorded in the Caltech catalog between 1977 and 1987. Sequences that have been relocated in special studies are also plotted and include 1966 Parkfield, 1968 Borrego Mountain, 1970 Lytle Creek, 1972 Bear Valley, 1975 Haicheng ( $M = 7.3$ ), 1975 Galway Lakes ( $M = 5.2$ ), 1979 Homestead, 1980 Livermore, 1981 Westmoreland, 1985 Kettleman Hills, 1986 Chalfant Valley, and 1987 Superstition Hills. For the three foreshock sequences with known rupture zones the distance range of foreshock rupture zone to the mainshock epicenter is shown by the elongated ovals; the circles inside these show the distance for the foreshock epicenter.

of major faults. Jones [1984] suggested that foreshocks are more common at areas of complication along faults; this would require that either  $P(C)$  or  $\Phi_s$  (or both) be larger at such places. An increase of  $P(C)$  would be in accordance with the notion that epicenters of mainshocks are mostly at such points [King and Nabelek, 1985; Bakun et al., 1986]. While this seems like a valid refinement, in practice differentiating between the many possible complex sites and the "smooth" parts of the fault would require gridding at the kilometer scale, a level of detail that does not seem justified by our present level of knowledge. One further choice would be to make  $\Phi_s$  proportional to the local rate of background activity  $\Lambda_s$ , thus asserting that most mainshocks with foreshocks occur in areas with high background seismicity. The data on foreshocks to moderate earthquakes in California [Jones, 1984] does not support this: while the fraction of earthquakes with foreshocks does vary by region, it does not appear to be related to background seismicity.

city. For example, the Calaveras fault in central California has a relatively high rate of background activity and no foreshocks.

Foreshocks and mainshocks thus clearly occur close together in space, within 10 km of each other in all resolvable cases—but show no other clear dependence on location. We therefore have made  $\Phi_s$  depend only on  $\rho$ , the distance between candidate foreshock and possible mainshock epicenters ( $\rho = [(x - x')^2 + (y - y')^2]^{1/2}$ ). The condition for  $\Phi_s$  to be properly normalized is

$$\iint_{A_C} dx dy \iint_{A_C} dx' dy' \Phi_s(x, y, x', y') \Omega_s(x', y') = \iint_{A_C} dx' dy' \Omega_s(x', y') \quad (21)$$

which in general can be done only numerically, even for  $\Omega_s$  constant and  $\Phi_s$  having a simple dependence on  $\rho$ . If, however, we make the simplification, mentioned in section 3.1, of making our spatial integrals one-dimensional (with  $A_C$  then being the length of the fault), assume  $\Omega_s$  constant, and make  $\Phi_s$  constant for  $\rho \leq \rho_w$  and zero for larger  $\rho$ , we find that  $\Phi_s$  is

$$\frac{1}{\rho_w(1 - \rho_w^2/4A_C)} \quad \text{if } \rho \leq \rho_w \\ 0 \quad \text{if } \rho > \rho_w \quad (22)$$

We use  $\rho_w = 10$  km to agree with the data presented above. Then, provided that the location  $x_0$  of the candidate earthquake is more than a distance  $\rho_w$  from an end of the fault zone and that  $\Omega_s(x')$  is constant over a distance  $2\rho_w$ , the integral needed in (16) is

$$\int_{x_0 - \rho_w}^{x_0 + \rho_w} dx \int_{A_C} dx' \Phi_s(x, x') \Omega_s(x) = 2e_0 \frac{\Omega_s(x_0)}{1 - \rho_w^2/4A_C} \equiv 2e_0 I_s(x_0) \quad (23)$$

where we have defined  $I_s$  in a parallel way to  $I_t$ ; the dependence on  $x_0$  comes through the dependence on the value of  $\Omega_s$  near the candidate earthquake.

**3.3.3. Magnitude.** The functional form for  $\Phi_m(M, M')$  is probably the least certain part of  $\Phi_{FC}$ . Plots of the difference in foreshock and mainshock magnitudes with a uniform magnitude threshold for foreshocks and mainshocks [e.g., Jones, 1985] show the magnitude difference to be a negative exponential distribution. However, to consider all possible foreshocks to a given mainshock, the completeness threshold for the foreshocks should be much lower than for the mainshocks. A bivariate plot of foreshock and mainshock magnitudes for all recorded foreshocks in southern California (Figure 3) suggests that for any given 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.

Because of the simplicity of this last assumption, we have used it here by making  $\Phi_m$  constant; we set  $\Phi_m(M, M') \equiv N_m$ , a normalizing factor. The normalization of  $\Phi_m$  is in general set by

$$\int_{M_B}^{M'} \int_{M_D}^{M'} \Phi_m(M, M') dM dM' = \alpha \int_{M_B}^{\infty} e^{-\beta M'} dM' \quad (24)$$

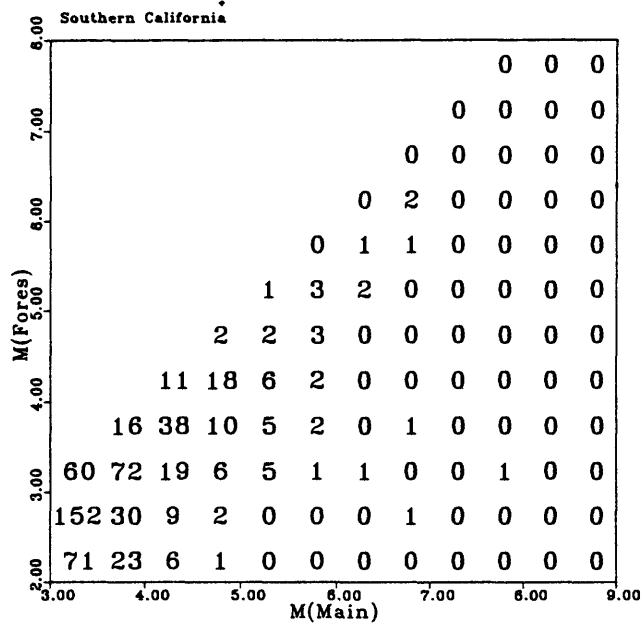


Fig. 3. The number of foreshock-mainshock pairs in half unit of magnitude bins for the magnitudes of foreshock and mainshock. Data included all  $M \geq 2.0$  foreshocks and  $M \geq 3.0$  mainshocks recorded between 1932 and 1987 in southern California.

Equation (24) says that if we look before all mainshocks with magnitudes greater than  $M_B$  for foreshocks above a cutoff magnitude of  $M_D$ , we find that a fraction  $\alpha$  of the mainshocks have foreshocks. Note that we have chosen to normalize  $\Phi_f$  and  $\Phi_s$  to integrate to 1, so  $\Phi_m$  contains the information about the total fraction of mainshocks with foreshocks.

Making  $\Phi_m$  constant implies that the fraction of mainshocks preceded by foreshocks will increase as the magnitude threshold for foreshocks decreases. This is consistent with reported foreshock activity, since the data suggest that foreshocks are relatively common before major strike-slip earthquakes. Jones and Molnar [1979] found that 30% of the  $M \geq 7.0$  earthquakes occurring outside of subduction zones were preceded by foreshocks in the NEIC catalogue ( $M \geq 4.5$ – $5.0$ ) and almost 50% had foreshocks  $M \geq 2$  reported in the literature. Jones [1984] showed that half of the  $M \geq 5.0$  strike-slip earthquakes in California were preceded by  $M \geq 2.0$  foreshocks. (Foreshocks were less common on thrust faults.)

For  $\Phi_m$  constant and equal to  $N_m$ , (24) implies that

$$N_m = \frac{\alpha\beta'}{1 + \beta'(M_B - M_D)} \quad (25)$$

The data presented by Jones [1984], with  $M_B = 5.0$  and  $M_D = 2.0$ , gave  $\alpha$  equal to 0.5 for strike-slip earthquakes. Adopting this value, with a  $\beta'$  of 2.3, gives  $N_m = 0.15$ . A consequence of taking  $\Phi_m$  constant is then that all earthquakes should have foreshocks within 6.5 units of magnitude of the mainshock. Holding  $\Phi_m$  constant for all  $M$  would of course lead to the absurd result that more than 100% of mainshocks have foreshocks within, say, 8 magnitude units. For the smaller range of magnitudes considered here a constant  $\Phi_m$  does not present any difficulties.

The integral needed for (16) is then

$$\int_{M-\mu}^{M+\mu} dM \int_{M_C}^{M_C+\mu_C} dM' \Phi_m(M, M') e^{-\beta'M'} = 2\mu N_m e^{-\beta'M_C} \frac{1 - e^{-\beta'\mu_C}}{\beta'} \equiv 2\mu I_m(M_C, \mu_C) \quad (26)$$

where we have assumed  $\mu$  small, and again separated it out from the rest of the expression.

### 3.4. Mainshock Probability

We now can combine the integrals in (18), (23), and (26) into (16) to get the foreshock probability:

$$P(F) = 4\delta_0\mu e_0 I_f I_s I_m$$

Solving the integral in (8) for the background event gives

$$P(B) = 4\delta_0\mu e_0 \Lambda_s(x_0) e^{-\beta M}$$

We substitute these values of the background and foreshock probabilities into (6) to obtain:

$$P(C|F \cup B) = \frac{I_s I_f I_m}{I_s I_f I_m + \Lambda_s(x_0) e^{-\beta M}} \quad (27)$$

The candidate earthquake errors  $\delta_0$ ,  $e_0$ , and  $\mu$  have canceled out.

For making calculations, it is also useful to set  $I_f$  equal to 1 (solve for the probability in a fixed time interval) and (for the case of a linear fault) take  $I_s$  in (23) to be equal to  $\Omega_s(x_0)$ . If we take  $\Omega_s$  to be constant and combine (14) and (26), we find that the dependence on  $M_C$  and  $\mu_C$  cancels out, and we are left with

$$P(C|F \cup B) = \frac{(N_m P(C)/A_C \delta_1)}{(N_m P(C)/A_C \delta_1) + \Lambda_s(x_0) e^{-\beta M}} \quad (28)$$

## 4. APPLICATION TO THE SAN ANDREAS FAULT SYSTEM, CALIFORNIA

We now have an expression for the conditional probability of a characteristic earthquake on a fault segment given the occurrence of an earthquake that is either a background event or a foreshock. To evaluate this, we need the long-term probability of the characteristic mainshock (the terms involving the actual magnitude of the characteristic earthquake have canceled out), the length of the fault segment, and the rate density of background seismicity for that segment. To show how this works, we now apply this to the San Andreas fault system in California, because the long-term probabilities for characteristic earthquakes that we need have been estimated for the major faults of this system, the San Andreas, Hayward, San Jacinto and Imperial faults. This was first done by Lindh [1983] and Sykes and Nishenko [1984], and more recently by the Working Group on California Earthquake Probabilities [1988], hereafter referred to as WGCEP-88.

Our division of the fault into segments and our values of  $P(C)$  for each segment come largely from WGCEP-88. One exception is that the lengths of the Southern Santa Cruz Mountains and the San Francisco Peninsula segments have been altered to match the rupture zone of the 1989 Loma Prieta earth-

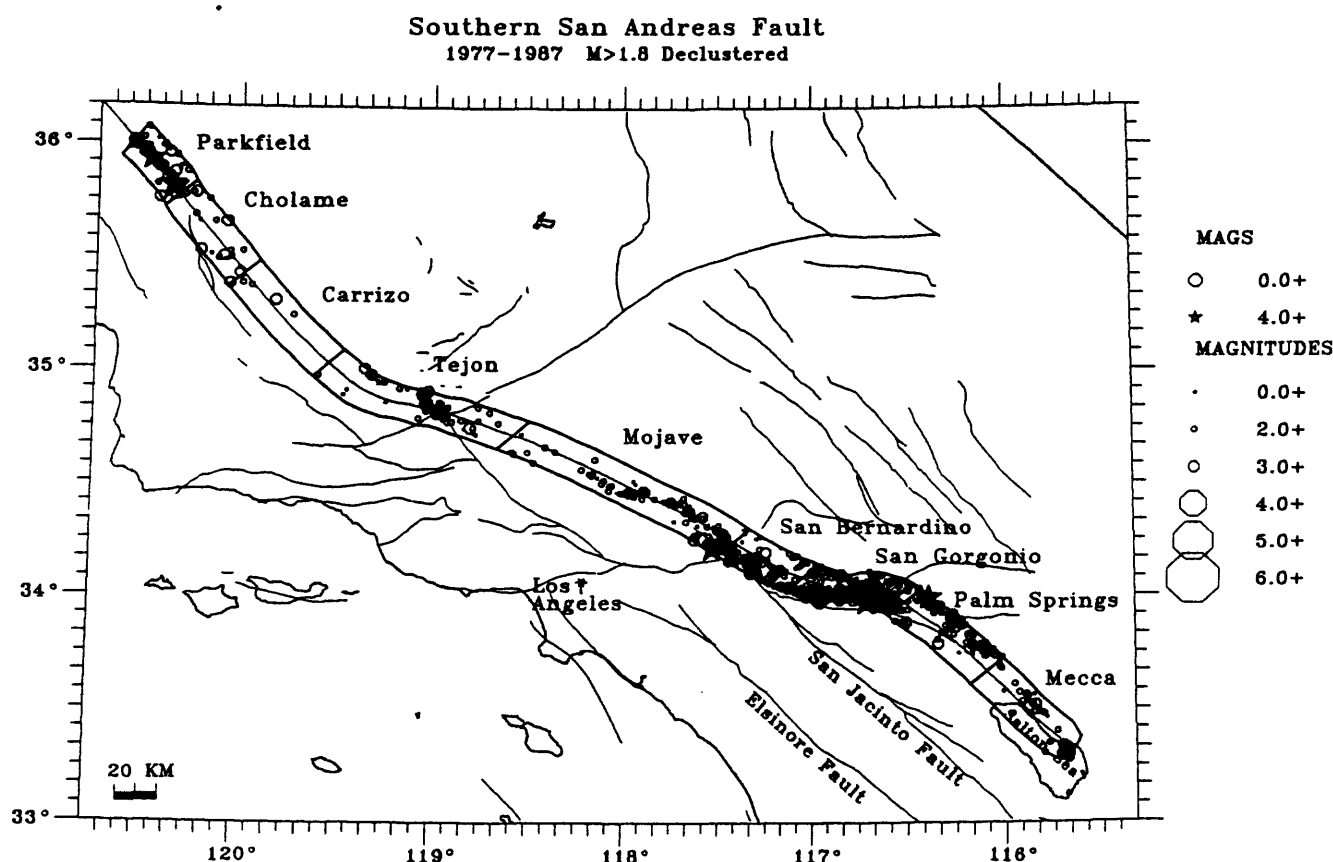


Fig. 4. A map of  $M \geq 1.8$  declustered earthquakes located within 10 km of the southern San Andreas fault recorded in the Caltech catalog between 1977 and 1987 and (for Parkfield) the CALNET catalog between 1975 and 1989.

quake (A. Lindh, personal communication, 1990). We took  $P(C)$  to be constant along each segment; as noted in section 3.2.1, we have not tried to include the possibility that nucleation points (and higher values of  $P(C)$ ) are more likely at points of complication. We have also not altered the distribution of  $P(C)$  to account for any possible relationship between nucleation point and level of background activity.

The rate density for the background seismicity is determined from the microearthquakes recorded between 1977 and 1987 by the Caltech/U.S. Geological Survey Southern California Seismic Network [Given *et al.*, 1988] for southern California and between 1975 and 1989 by CALNET, the U.S. Geological Survey Central California Seismic Network (P. Reasenberg, personal communication, 1990), for northern California. Background seismicity can be defined in many ways; it is important in this application that it be defined in the same way as the foreshocks will be. Because foreshocks can be up to 10 km from their mainshock (Figure 2), background seismicity up to 10 km from the surface trace of the San Andreas fault is included in the background rate.

Another issue is how to handle temporal clustering in the catalog. We assume that if an earthquake of  $M = 6$  (for instance) were to occur on the southern San Andreas fault with an aftershock sequence, we will only evaluate the probability that the  $M = 6$  earthquake is a foreshock, and not individually determine the probabilities that the  $M = 6$  and each of its aftershocks is a foreshock and then sum them. For consistency we therefore want to determine the background seismicity using a

catalog from which aftershock sequences and swarms have been removed. In such a declustered catalogue, sequences are recognized by some algorithm and replaced in the catalogue with one event at the time of the largest earthquake in the sequence, which is given a magnitude equivalent to the summed moment of all the earthquakes in the sequence. To produce our declustered catalogs, we used the algorithm of Reasenberg [1985].

The resulting background seismicity within 10 km of the faults is shown in Figures 4-6. It is clear from these that the rate of background seismicity can vary significantly within the fault segments defined by WGCEP-88. For example, the Coachella Valley segment of the San Andreas includes the active region around Desert Hot Springs (including a  $M = 6.5$  event in 1948) and a very quiet region (near the Salton Sea) where the largest earthquake in 55 years has been  $M = 3.5$ . To account for this variation, we have divided some of the WGCEP-88 segments into smaller regions, which are shown in Figures 4-6 and listed in Table 1.

Table 1 provides the data needed for each segment. To use (28) we also need the time period  $\delta_1$ , which we set to 3 days ( $1.09 \times 10^5$  s), to match the recent usage of the U.S. Geological Survey and the California's Governor's Office of Emergency Services in issuing earthquake advisories. Alert levels for such advisories are defined to correspond to certain probabilities; the magnitudes of earthquakes needed to trigger those alert levels can then be computed from (28), and are also given in Table 1. Figure 7 shows the probability as a function of the magnitude of

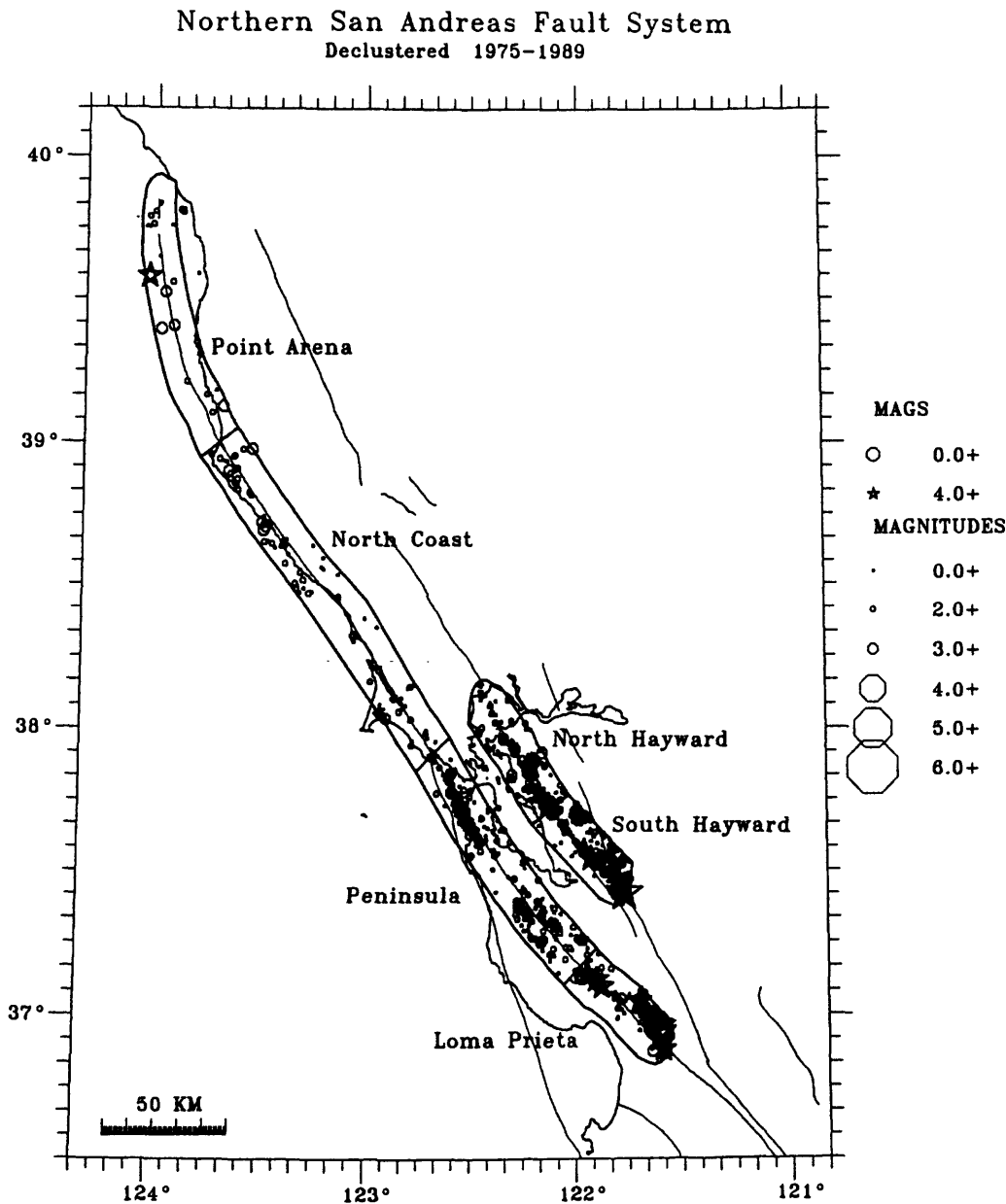


Fig. 5. A map of  $M \geq 1.8$  declustered earthquakes located within 10 km of the San Jacinto fault recorded in the Caltech catalog between 1977 and 1987.

the candidate earthquake for each segment.

We have treated the Parkfield segment in two different ways. In the Table 1 listing for Parkfield, we treat it in the same way as the other segments, regarding the foreshock as equally likely anywhere along the segment, and taking  $P(C)$  from WGCEP-88. These assumptions give short-term probabilities much lower than those estimated by *Bakun et al.* [1987] for the Parkfield earthquake prediction experiment. *Bakun et al.* [1987] used a somewhat different methodology and also used different assumptions in two areas: their value of  $P(C)$  is 1.5 times that of WGCEP-88, and they assume that the foreshock will be located in a small region under Middle Mountain, making a smaller area for defining background seismicity. (Their assumption that 50% of Parkfield mainshocks will be preceded by foreshocks agrees with our choice in section 3.2.3). For a better comparison we have used the *Bakun et al.* assumptions to deter-

mine short-term probabilities with our methodology and given these in Table 1 as Middle Mountain probabilities. These remain lower than the *Bakun et al.* results; for example, a magnitude 1.5 shock gives a probability of 0.1% from our methodology and 0.68% (Level D alert) according to *Bakun et al.*

As with the long-term probabilities of major earthquakes, these short-term foreshock-based probabilities are better seen as a means of ranking the relative hazard from different sections of the faults than as highly accurate absolute estimates. The probabilities are as uncertain as the data used to calculate them, which in some cases are uncertain indeed. For example, the values of  $P(C)$  found by WGCEP-88 are up to a factor of 4 larger than those found by *Davis et al.* [1989]; this would lead to similarly large differences in the short-term probabilities.

The relative short-term probabilities for different segments shown in Table 1 and Figure 7 are strongly affected by both the

**San Jacinto and Imperial Faults**  
1977-1988  $M > 1.8$  Declustered

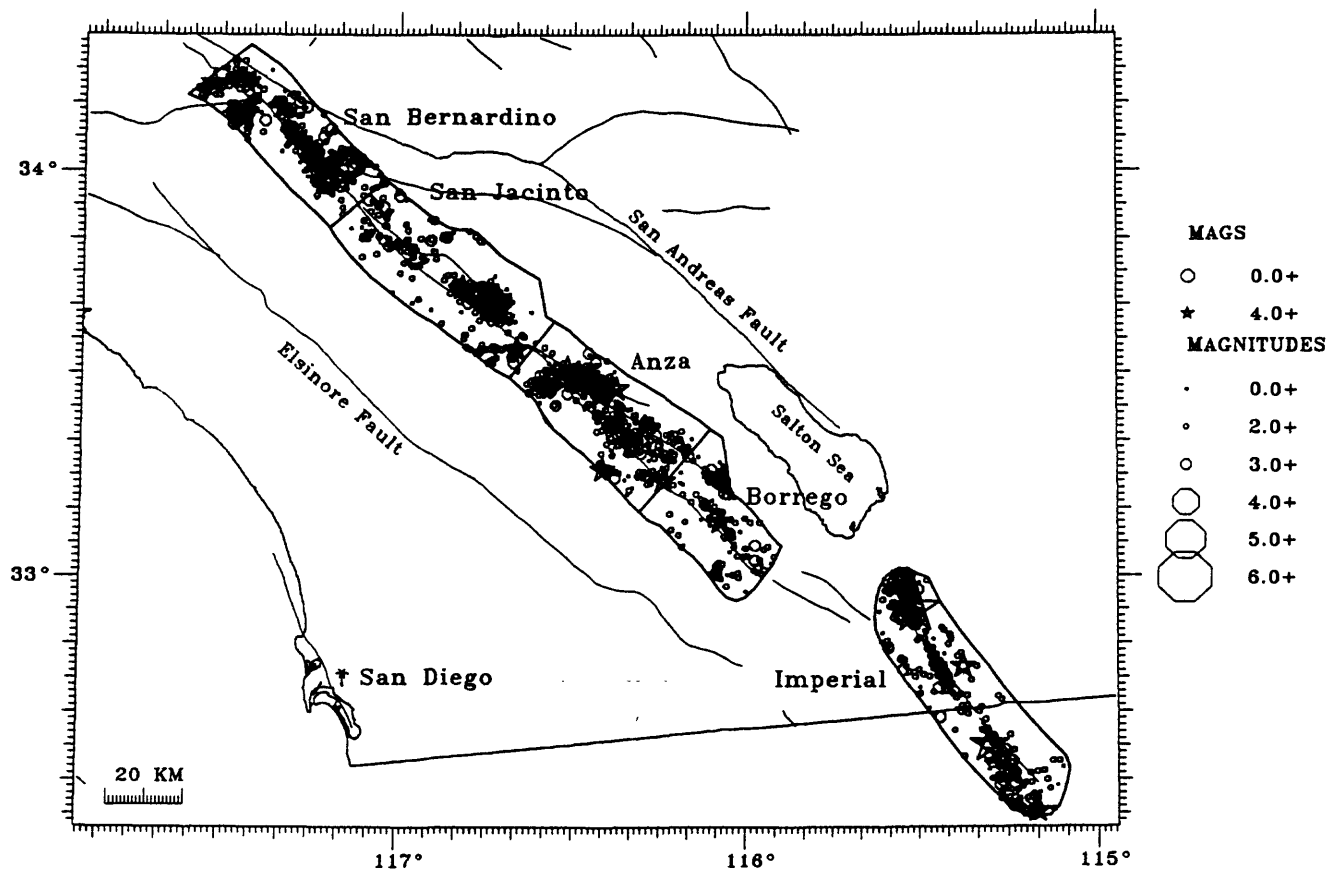


Fig. 6. A map of  $M \geq 1.5$  declustered earthquakes located within 10 km of the northern San Andreas and Hayward faults recorded in the CALNET catalog between 1975 and 1989.

TABLE 1. Parameters and Magnitude Levels

Segment	Length, km	$P_C$ /3 days	$a$	$b$	$\Lambda_s$ , events/km s	$\beta$	Magnitude for		
							0.1%	1%	10%
<i>San Andreas Fault</i>									
Mecca	60	$1.1 \times 10^{-4}$	3.67	0.95	$4.91 \times 10^{-7}$	2.18	3.1	4.2	5.3
Palm Springs	50	$1.1 \times 10^{-4}$	4.00	0.97	$1.29 \times 10^{-6}$	2.23	3.5	4.5	5.6
San Gorgonio	60	$5.5 \times 10^{-5}$	4.46	0.94	$2.99 \times 10^{-6}$	2.16	4.2	5.3	6.4
San Bernardino	40	$5.5 \times 10^{-5}$	3.95	0.92	$1.36 \times 10^{-6}$	2.12	4.0	5.1	6.2
Mojave	100	$8.2 \times 10^{-5}$	3.85	0.90	$4.22 \times 10^{-7}$	2.07	3.3	4.4	5.6
Tejon	100	$2.7 \times 10^{-5}$	3.49	0.88	$1.80 \times 10^{-7}$	2.02	3.7	4.9	6.1
Carrizo	60	$2.7 \times 10^{-5}$	2.58	1.03	$4.32 \times 10^{-8}$	2.37	2.6	3.6	4.6
Cholame	50	$8.2 \times 10^{-5}$	2.87	0.83	$8.15 \times 10^{-8}$	1.91	2.3	3.6	4.8
Parkfield	35	$8.2 \times 10^{-4}$	4.17	0.87	$1.79 \times 10^{-6}$	2.00	2.5	3.6	4.8
Middle Mountain	20	$1.2 \times 10^{-3}$	3.40	0.74	$4.52 \times 10^{-7}$	1.70	1.5	2.9	4.3
Loma Prieta	50	$8.2 \times 10^{-5}$	4.41	1.01	$2.52 \times 10^{-6}$	2.32	3.4	4.4	5.4
Peninsula	100	$5.5 \times 10^{-5}$	4.57	1.15	$2.08 \times 10^{-6}$	2.64	3.3	4.2	5.1
North Coast	150	$1.4 \times 10^{-5}$	3.26	0.88	$6.09 \times 10^{-8}$	2.02	3.7	4.9	6.1
Point Arena	100	$1.4 \times 10^{-5}$	2.95	0.69	$1.40 \times 10^{-8}$	1.59	3.8	5.3	6.8
<i>San Jacinto Fault</i>									
San Bernardino	50	$5.5 \times 10^{-5}$	4.58	0.98	$4.94 \times 10^{-6}$	2.25	4.0	5.0	6.1
San Jacinto	65	$2.7 \times 10^{-5}$	4.49	1.01	$3.18 \times 10^{-6}$	2.32	4.1	5.1	6.1
Anza	50	$8.2 \times 10^{-5}$	4.57	0.95	$4.68 \times 10^{-6}$	2.18	3.9	5.0	6.1
Borrego	40	$1.4 \times 10^{-5}$	4.05	0.99	$1.84 \times 10^{-6}$	2.28	4.0	5.0	6.1
<i>Hayward Fault</i>									
North Hayward	60	$5.5 \times 10^{-5}$	4.24	0.99	$1.39 \times 10^{-6}$	2.28	3.5	4.5	5.5
South Hayward	50	$5.5 \times 10^{-5}$	4.41	1.01	$2.52 \times 10^{-6}$	2.32	3.6	4.6	5.6
<i>Imperial Fault</i>									
Imperial	50	$1.4 \times 10^{-4}$	4.59	0.96	$4.95 \times 10^{-6}$	2.21	3.6	4.7	5.8

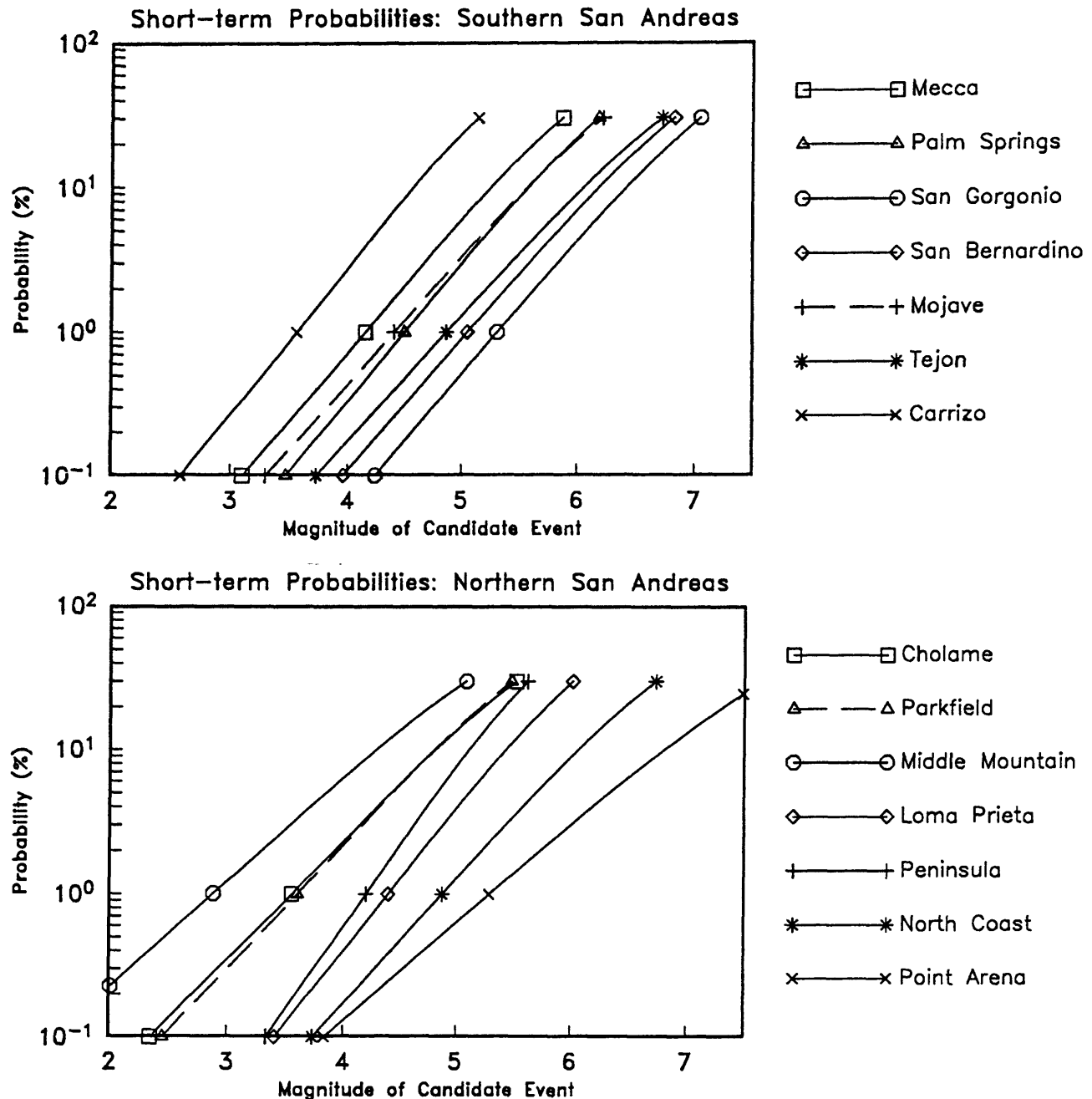
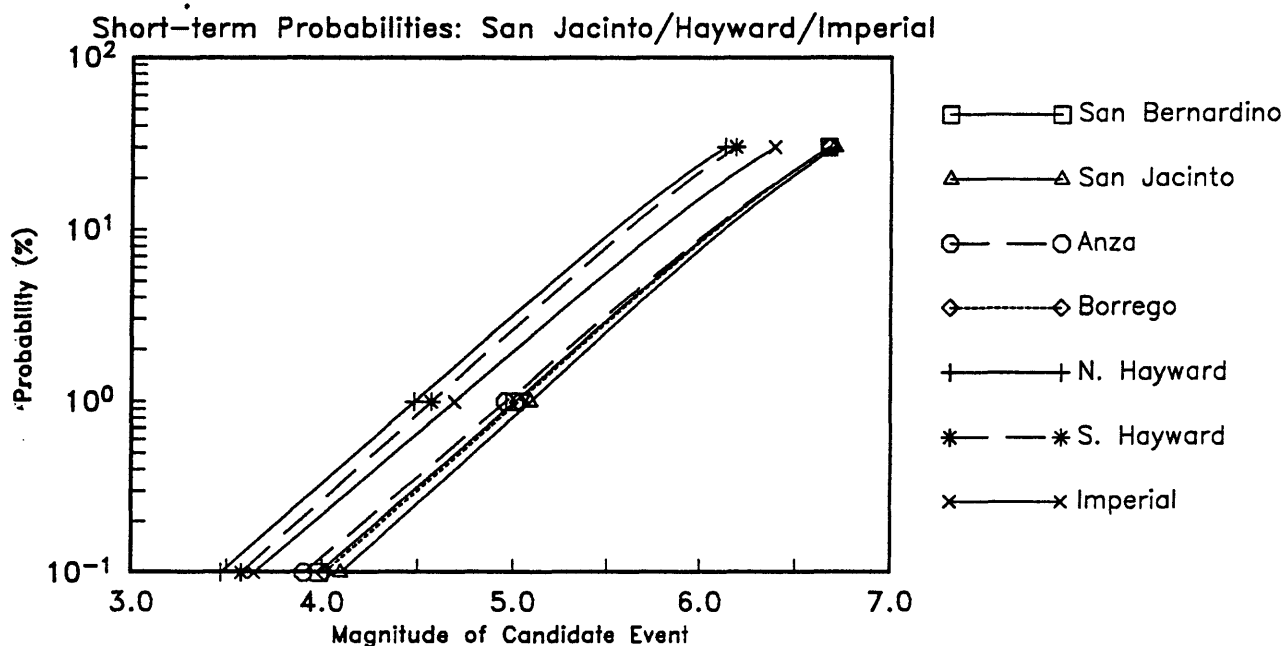


Fig. 7. The probability that an earthquake of given magnitude is a foreshock to a characteristic mainshock, plotted against magnitude for each fault segment listed in Table 1. Shown are results for the (a) southern San Andreas, (b) northern San Andreas, and (c) other faults of the San Andreas system.

long-term probability  $P(C)$  and the rate of background seismicity. Outside of the Parkfield and "Middle Mountain" segments, which have very high  $P(C)$ , the highest short-term probabilities are from the Carrizo and Cholame segments; even though the 30-year probability is only 10% on the Carrizo segment, the background seismicity there is almost nonexistent. The San Francisco Peninsula and the San Bernardino Mountain segments both have a 30-year probability of 20%, but the probabilities in the San Gorgonio subregion are much lower than those near San Francisco because of higher background seismicity. At high magnitudes, the lowest probabilities are for the Point Arena segment because of its very low  $\beta$ , which may be a result of catalog incompleteness at low magnitudes.

The possibility of the next Parkfield earthquake triggering a larger earthquake on the Cholame segment has been much discussed. Our procedure gives a magnitude 6 in Cholame a 52% chance of being a foreshock to a characteristic mainshock on that segment; but this result comes from the low background rate for the Cholame segment itself. Since this rate predicts a magnitude 6 shock every 1400 years, not every 22 years as at Parkfield, this high probability does not apply to a possible Parkfield trigger. We can, however, use (3) of our zero-dimensional model to roughly estimate the probability that a Parkfield earthquake will be a foreshock to a larger earthquake at Cholame. The WGCEP-88 probability of a Cholame earthquake is 30% in 30 years, while the background rate for



Parkfield mainshocks (and hence candidate Cholame foreshocks) is one every 21.7 years. To determine the short-term probability with (3) we need to assume a value for  $P(F|C)$ , the rate at which Cholame mainshocks have Parkfield mainshocks as foreshocks. If we assume that, as for an average magnitude 7 earthquake, 15% of Cholame earthquakes are preceded by magnitude 6 foreshocks, then the short-term probability of a Cholame earthquake after a Parkfield earthquake is 3%. At the other extreme, if we assume that 50% of Cholame earthquakes are preceded by Parkfield earthquakes (the only type of foreshock it has is Parkfield earthquakes), the probability becomes 10%.

As discussed in section 2.2 above, the rate of false alarms depends on the background probability for the characteristic earthquake. A cumulative false-alarm rate for the whole San Andreas fault is thus dominated by the contribution from Parkfield, for which a 10% probability level occurs every 8.4 years. By comparison, for the Coachella Valley segment of the San Andreas fault the false alarm rate for a 10% probability is once every 63 years. For 0.1% it is once every 5.5 months, but this probability level is only 9 times the background one.

## 5. DISCUSSION

The procedure developed here can be made more general than has been appropriate for the above application to the San Andreas fault. As discussed in sections 3 and 4, we could include a different dependence of  $\Phi_i$  on time or make  $P(C)$  include information about the most likely epicenters for the mainshock (such as fault jogs or terminations). Another extension would be to set  $P(C)$  from an extrapolation of the frequency-magnitude relation; while a violation of the maximum-magnitude model, this would allow application of this technique to many more regions. The greatest flexibility comes from the precurent probability density  $\Phi_{FC}$ , since we can, as the data warrant, alter this function to include additional data types. For example, there is evidence to suggest that most foreshocks have focal mechanisms similar to that of their

mainshock [Jones and Lindh, 1987]. If that relationship were parameterized,  $\Phi_{FC}$  and the integration in (15) could include variables describing the difference in focal mechanisms; thus normal- or thrust-faulting earthquakes would be given a lower probability of being a foreshock to a San Andreas mainshock. If any other characteristics are recognized as being more common in foreshocks than background earthquakes (such as number of aftershocks), we can rigorously include this information in our computation of the conditional probabilities. Another direction to go is in improving our estimates for the precurent probability beyond the rather simple forms described above. Considerable work has been done in the last few years on how to estimate multivariate density functions, which is precisely the problem at hand [Silverman 1986]. An obvious question is whether the estimated densities differ significantly between regions; if so, this could reflect significant differences in the nucleation and triggering of large earthquakes.

Of course, nothing in the derivations of section 2 is specific to foreshocks; this procedure can be used for any potential earthquake precursor. Equation (6) shows that what is needed is a long-term mainshock probability  $P(C)$ , a rate for background events  $P(B)$ , and a precurent probability  $\Phi_{FC}$ , which would in many cases just be the fraction of mainshocks with precursors. At present, these data are not available for any precursor but foreshocks. For instance, the background rate of creep events can be determined for some sections of the San Andreas fault system, but we have almost no data on the fraction of mainshocks preceded by such events.

There have been a number of earlier papers on estimating the probabilities of earthquakes in the presence of precursors [Kagan and Knopoff 1977; Vere-Jones 1978; Guagenti and Scirocco 1980; Aki 1981; Anderson 1982; Grandori et al. 1984]. Most of these take a slightly different definition of events from the ones we have used. Rather than distinguishing between background events (independent of large earthquakes) and precursors (always followed by a large earthquake), these papers assume that all possible precursors fall into one class of events, with some probability of a possible precursor not being followed



by an earthquake. (For example, Anderson [1982] computes the probabilities of a precursor being useful or useless). For seismicity, a division into background and precursory events appears to be a better approximation to the likely physics. Most of these papers also deal with the case (not discussed here) of how possible multiple precursors could increase the conditional probability above that for a single precursor. The discussion above suggests that this will usually be a moot point, since only rarely do we have the information needed to estimate the conditional probabilities. With the exception of the work of Kagan and Knopoff [1987] and (in part) Anderson [1982], there does not seem to have been much consideration of any multidimensional cases of the kind described in section 3. The Kagan and Knopoff study is closest to the approach presented here, though the functional form employed by them is derived from a fracture mechanics model, whereas ours is more purely empirical. The models also differ considerably in their specification of long-term probability. In the Kagan and Knopoff model, this is given by a Poisson rate derived from the frequency-magnitude relation (10), whereas here it can be independent of that. As noted in section 2, such independence appears to be a more satisfactory representation of the seismicity of an active fault zone.

## 6. CONCLUSIONS

We have shown that the probability that an earthquake that occurs near a major fault will be a foreshock to the characteristic mainshock depends on the rate of background earthquake activity on that segment, the long-term probability of the mainshock, and the rate at which the mainshocks are preceded by foreshocks, which we call the precurrent probability. Assuming certain reasonable forms for the density function of this probability (as a function of time, location, and magnitude) we have found an expression for the short-term probability that an earthquake is a foreshock, and applied it to the faults of the San Andreas system. Because the rate of foreshocks before mainshocks is assumed to be the same for all segments, the differences in short-term probabilities between segments arise from differences in background rate of seismicity and in long-term probabilities. The background rates are more variable between regions and lead to larger variations in short-term probabilities. For the San Andreas fault the two extremes are the nearly aseismic Carrizo Plain, where a 1% probability for a characteristic earthquake would be found for a magnitude 3.6 candidate event, and the highly seismic San Geronio region, where it would take a magnitude 5.3 to reach this level.

**Acknowledgements.** We thank the members of the Working Group on Short-term Earthquake Alerts for the Southern San Andreas Fault, especially Brad Hager and Dave Jackson, for raising some of the issues that led to this paper. We have benefited greatly from reviews by Andy Michael, Dave Jackson (again), Al Lindh, and especially Mark Mathews. We also thank Paul Reasenberg for comments and for providing the declustered CALNET data. Preparation of this paper was in part supported by U.S. Geological Survey grant 14-08-0001-G1763.

## REFERENCES

- Aki, K., A probabilistic synthesis of precursory phenomena, in *Earthquake Prediction: An International Review*, edited by P. Richards, pp. 566-574, American Geophysical Union, Washington D. C., 1981.
- Anderson, J. G., Revised estimates for the probabilities of earthquakes following the observation of unreliable precursors, *Bull. Seismol. Soc. Amer.*, 72, 879-888, 1982.
- Bakun, W. H., K. S. Breckenridge, J. Bredehoeft, R. O. Burford, W. L. Ellsworth, M. J. S. Johnston, L. Jones, A. G. Lindh, C. Mortenson, R. J. Mueller, C. M. Poley, E. Roeleffs, S. Schulz, P. Segall, and W. Thatcher, Parkfield, California, earthquake prediction scenarios and response plans, *U. S. Geol. Surv. Open-file Rep.* 86-365, 1987.
- Bakun, W. H., G. C. P. King, and R. S. Cockerham, Seismic slip, aseismic slip, and the mechanics of repeating earthquakes on the Calaveras fault, California, in *Earthquake Source Mechanics (Geophysical Monograph 37)*, edited by C. Scholz, pp. 195-207, American Geophysical Union, Washington D. C., 1986.
- Brune, J. N., Implications of earthquake triggering and rupture propagation for earthquake prediction based on premonitory phenomena, *J. Geophys. Res.*, 84, 2195-2197, 1979.
- Davis, P. M., D. D. Jackson, and Y. Kagan, The longer it has been since the last earthquake, the longer the expected time till the next?, *Bull. Seismol. Soc. Amer.*, 79, 1439-1456, 1989.
- Given, D. D., L. K. Hutton, L. Stach, and L. M. Jones, The Southern California Network Bulletin, January-June, 1987, *U.S. Geol. Surv. Open-file Rep.* 88-408, 1988.
- Goltz, J., The Parkfield and San Diego earthquake predictions: a chronology, *Special Report by the Southern California Earthquake Preparedness Project*, Los Angeles, Calif, 1985.
- Grandori, G., E. Guagenti, and F. Perotti, Some observations on the probabilistic interpretation of short-term earthquake precursors, *Earthq. Eng. Struct. Dynam.*, 12, 749-760, 1984.
- Guagenti, E. G. and F. Scirocco, A discussion of seismic risk including precursors, *Bull. Seismol. Soc. Amer.*, 70, 2245-2251, 1980.
- Jones, L. M., Foreshocks (1966-1980) in the San Andreas System, California, *Bull. Seismol. Soc. Amer.*, 74, 1361-1380, 1984.
- Jones, L. M., Foreshocks and time-dependent earthquake hazard assessment in southern California, *Bull. Seismol. Soc. Amer.*, 75, 1669-1680, 1985.
- Jones, L. M. and A. G. Lindh, Foreshocks and time-dependent conditional probabilities of damaging earthquakes on major faults in California, *Seismol. Res. Letters*, 58, 21, 1987.
- Jones, L. M. and P. Molnar, Some characteristics of foreshocks and their possible relationship to earthquake prediction and premonitory slip on faults, *J. Geophys. Res.*, 84, 3596-3608, 1979.
- Kagan, Y. and L. Knopoff, Earthquake risk prediction as a stochastic process, *Phys. Earth Planet. Int.*, 14, 97-108, 1977.
- Kagan, Y. and L. Knopoff, Statistical short-term earthquake prediction, *Science*, 236, 1563-1567, 1987.
- King, G. C. P. and J. Nabelek, Role of fault bends in the initiation and termination of rupture, *Science*, 228, 984-987, 1985.
- Knott, C. G., *The Physics of Earthquake Phenomena*, 278 pp., Clarendon Press, Oxford, 1908.
- Lindh, A. G., Preliminary assessment of long-term probabilities for large earthquakes along selected segments of the San Andreas fault system in California, *U.S. Geol. Surv. Open-File Rep.* 83-63, pp. 1-15, 1983.
- Nishenko, S. P. and R. Buland, A generic recurrence interval distribution for earthquake forecasting, *Bull. Seismol. Soc. Amer.*, 77, 1382-1399, 1987.
- Reasenberg, P., Second-order moment of Central California seismicity, 1969-1982, *J. Geophys. Res.*, 90, 5479-5496, 1985.
- Reasenberg, P. A. and L. M. Jones, Earthquake hazard after a mainshock in California, *Science*, 243, 1173-1176, 1989.
- Silverman, B., *Density Estimation*, 175 pp., Chapman and Hall, London, 1986.
- Sykes, L. and S. P. Nishenko, Probabilities of occurrence of large plate-rupturing earthquakes for the San Andreas, San Jacinto, and Imperial faults, California 1983-2003, *J. Geophys. Res.*, 89, 5905-5927, 1984.
- Vere-Jones, D., Earthquake prediction—a statistician's view, *J. Phys. Earth*, 26, 129-146, 1978.
- Wesnousky, S. G., C. H. Scholz, K. Shimazaki, and T. Matsuda, Earthquake frequency distribution and faulting mechanics, *J. Geophys. Res.*, 88, 9331-9340, 1983.
- Working Group on California Earthquake Probabilities, Probabilities of large earthquakes occurring in California on the San Andreas fault, *U.S. Geol. Surv. Open-File Rep.* 88-398, 1988.
- D. C. Agnew, Institute of Geophysics and Planetary Physics, University of California, San Diego, La Jolla, CA 92093-0225.
- L. M. Jones, U.S. Geological Survey, 525 S. Wilson Avenue, Pasadena, CA 91106

(Received August 17, 1990;  
revised January 14, 1991;  
accepted January 18, 1991.)

# Prediction Probabilities From Foreshocks

DUNCAN CARR AGNEW

*Institute of Geophysics and Planetary Physics,  
University of California, La Jolla*

LUCILE M. JONES

*U.S. Geological Survey, Pasadena, California*

When any earthquake occurs, the possibility that it might be a foreshock increases the probability that a larger earthquake will occur nearby within the next few days. Clearly, the probability of a very large earthquake ought to be higher if the candidate foreshock were on or near a fault capable of producing that very large mainshock, especially if the fault is towards the end of its seismic cycle. We derive an expression for the probability of a major earthquake characteristic to a particular fault segment, given the occurrence of a potential foreshock near the fault. To evaluate this expression, we need: (1) the rate of background seismic activity in the area, (2) the long-term probability of a large earthquake on the fault, and (3) the rate at which foreshocks precede large earthquakes, as a function of time, magnitude, and spatial location. For this last function we assume the average properties of foreshocks to moderate earthquakes in California: (1) the rate of mainshock occurrence after foreshocks decays roughly as  $t^{-1}$ , so that most foreshocks are within three days of their mainshock, (2) foreshocks and mainshocks occur within 10 km of each other, and (3) the fraction of mainshocks with foreshocks increases linearly as the magnitude threshold for foreshocks decreases, with 50% of the mainshocks having foreshocks with magnitudes within three units of the mainshock magnitude (within three days). We apply our results to the San Andreas, Hayward, San Jacinto, and Imperial faults, using the probabilities of large earthquakes from the report of the Working Group on California Earthquake Probabilities (1988). The magnitude of candidate event required to produce a 1% probability of a large earthquake on the San Andreas fault within three days ranges from a high of 5.3 for the segment in San Geronio Pass to a low of 3.6 for the Carrizo Plain.

Probably the most evil feature of an earthquake is its suddenness. It is true that in the vast majority of cases a severe shock is heralded by a series of preliminary shocks of slight intensity .... [but] only after the havoc has been wrought does the memory recall the sinister warnings of hypogene action.

C. G. Knott (1908, p. 10)

## 1. INTRODUCTION

Many damaging earthquakes have been preceded by smaller earthquakes that occur within a few days and a few kilometers of the mainshock [e.g., Jones and Molnar, 1979]; these are referred to as immediate foreshocks. If such foreshocks could be recognized before the mainshock, they would be very effective for short-term earthquake prediction; but so far no way has been found to distinguish them from other earthquakes. Even without this, the mere existence of foreshocks provides some useful predictive capacity. When any earthquake occurs, the possibility that it might be a foreshock increases the probability that a larger earthquake will soon happen nearby. For southern California, Jones [1985] showed that after any earthquake there is a 6% probability that a second one equal to or larger than the first will follow within five days and 10 km of the first. The probability is much lower for a second earthquake much larger than the first; for example, the probability of an earthquake two units of magnitude larger is only 0.2%. Using

these results, the U.S. Geological Survey has issued four short-term earthquake advisories after moderate earthquakes [e.g., Goltz, 1985]. A more recent study by Kagan and Knopoff [1987] developed a model for the clustering of earthquakes which could indicate areas of space and time in which larger events might follow smaller ones. The size of these areas depended on the probability gain, the ratio of probability of an earthquake given the occurrence of a possible precursor (such as a foreshock) to the probability in the absence of such a precursor [Kagan and Knopoff, 1977; Vere-Jones, 1978; Aki, 1981]. For low levels of probability gain, Kagan and Knopoff [1987] found that one-third of all earthquakes with magnitudes 4 and above fell within their predicted regions.

These results are from studies of earthquake catalogs; Jones [1985] used a catalog for southern California, and Kagan and Knopoff [1987] used one for central California. As a consequence, both papers give generic results about pairs of earthquakes, without much regard for other factors. But it ought to be possible to do better: the probability of a very large earthquake should be higher if the candidate foreshock were to occur near a fault capable of producing that mainshock than if it were located in an area where we believe such a mainshock to be very unlikely. Moreover, the chance of a candidate earthquake actually being a foreshock should be higher if the rate of background (nonforeshock) activity were low.

In this study we derive an expression for the probability of a major earthquake following a possible foreshock near a major fault from the basic tenets of probability theory. 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. We then apply this expression to

found from calculations of the type presented by the *Working Group on California Earthquake Probabilities* [1988]. If we had a record of the seismicity before many such characteristic earthquakes, we could evaluate  $P(F|C)$  (which we shall hereafter call  $\Phi_{FC}$ ) from it directly. (For this simple model,  $\Phi_{FC}$  is just the fraction of large earthquakes preceded by foreshocks.) Of course, we do not have such a record, and so are forced to make a kind of reverse ergodic assumption, namely that the time average of  $\Phi_{FC}$  over many earthquakes on one fault is equal to the spatial average over many faults. This may not be true, but it is for now the best we can do.

## 2.2. One-Dimensional Model

As a simple extension to the previous discussion, suppose that we have  $N$  "regions" and that  $C_i$ ,  $B_i$ , and  $F_i$  denote the occurrence of an event in the  $i$ th region, with  $C$  (for example) now being the occurrence of a large earthquake in any possible region. These regions can be sections of the fault or (as we will see below) volumes in a multidimensional space of all relevant variables. The quantity of interest is now  $P(C|F_i \cup B_i)$ : we have a candidate foreshock in one region, and want the probability of a large earthquake starting anywhere. Assuming that the occurrences  $C_i$  are disjoint (the epicenter can only be in one place), we then have that the probability of a foreshock in the  $i$ th region can be written as

$$P(F_i) = \sum_{j=1}^N \Phi_{FC}(i, j) P(C_j) \quad (5)$$

where  $\Phi_{FC}(i, j) = P(F_i|C_j)$ . We may regard  $\Phi_{FC}$  as the probability of a foreshock in region  $i$  given a large earthquake in region  $j$ . We call this the precurent probability because it refers to the probability of an event preceding a second one (not, it should be noted, with an implication of violated causality). As a simple example, we could take  $\Phi_{FC}(i, j) = \alpha \delta_{ij}$ , which would imply that large earthquakes are preceded by foreshocks only in the same region, and even then only a fraction  $\alpha$  of them have foreshocks at all.

We can then easily revise (4) above to get the probability we seek; simply adding subscripts to the candidate event yields

$$P(C|F_i \cup B_i) = \frac{P(F_i) + P(C)P(B_i)}{P(B_i) + P(F_i)} \quad (6)$$

Equations (5) and (6) are the basic ones we shall use in the more general case. Equation (5) shows us how to compute the probability of a foreshock happening in the location of our candidate earthquake, by summing over all possible mainshocks. The use of the precurent probability  $\Phi_{FC}$  is the key to this approach; we can (and in the next section shall) design it to embody our knowledge and assumptions about the relation between foreshocks and the earthquakes they precede. Having found the foreshock probability, we then use (6) to find the conditional probability of a large earthquake.

An important consequence of (5) is that we may sum over all possible foreshocks (again assuming disjointness) to get

$$P(F) = \sum_{i=1}^N \sum_{j=1}^N \Phi_{FC}(i, j) P(C_j) \quad (7)$$

giving us the overall probability of a foreshock somewhere in the total region. This must satisfy  $P(F) = \alpha P(C)$ , where  $\alpha$  is

the fraction of mainshocks with foreshocks; this and equation (7) together constrain the normalization of  $\Phi_{FC}$ .

Next to the probability level itself, the socially most interesting quantities would seem to be the chance of an alert being a false alarm, and the rate at which false alarms occur for a given probability level. The probability that an alert is a false alarm is  $P(\bar{C}|F_i \cup B_i)$ , which is just  $1 - P(C|F_i \cup B_i)$ : if we have a 10% chance of having a mainshock, we have a 90% chance of not having one. The rate of false alarms is equivalent to the probability of a false alarm happening in some given time, and this is just the probability that an alert is a false alarm times the probability of the event that triggers it, namely

$$\sum_{i=1}^N [1 - P(C|F_i \cup B_i)] [P(B_i) + P(F_i)]$$

As will be shown in section 4, we would in practice usually choose the probability of a mainshock given a small event,  $P(C|F_i \cup B_i)$  to have a fixed value (e.g., 1%), which we denote by  $S$ , for all regions. This value of  $S$  then sets the value of  $P(B_i)$  for the  $i$ th region; from (6),  $P(B_i) = P(F_i)[(1 - S)/S]$ , which makes the probability of a false alarm

$$\frac{(1 - S)}{S} \sum_{i=1}^N P(F_i) = \frac{(1 - S)}{S} \sum_{i=1}^N \sum_{j=1}^N \Phi_{FC}(i, j) P(C_j)$$

where we have used (5). For fixed  $S$  and  $\Phi_{FC}$  this expression is proportional to  $P(C_j)$  only: the rate of false alarms for a given probability depends only on the rate of mainshocks and not on the rate of background activity. In terms of the simple example at the beginning of section 2.1, fixing a probability level of 0.1% means that we would set the magnitude level of candidate events such that there would be 1000 background events for each actual foreshock; but the absolute rate of such background earthquakes (and thus of false alarms) is then determined only by the rate of foreshocks, and thus of mainshocks.

## 3. A MULTIDIMENSIONAL MODEL FOR FORESHOCKS

We now develop an expanded version of (5), which contains more variables. The first step is to define our events more thoroughly:

- $B$ : A background earthquake has occurred at coordinates  $(x_0 \pm e_0, y_0 \pm e_0)$ , during the time period  $[t, t + \delta_0]$ , with magnitude  $M \pm \mu$ . (All of the quantities  $e_0$ ,  $\delta_0$ , and  $\mu$  are small and are included because we will be dealing with probability density functions; as will be seen below, they cancel from the final expression).
- $F$ : A foreshock has occurred, with the same parameters as in event  $B$ .
- $C$ : A major earthquake will occur somewhere in the region of concern, which we denote by  $A_C$  (also using this variable for the area of this region). This earthquake will happen during the time period  $[t + \Delta, t + \Delta + \delta_1]$ , with magnitude between  $M_C$  and  $M_C + \mu_C$ .

We assume that we are computing the probability at some time in the interval  $(t + \delta_0, t + \Delta)$ ; the possible foreshock has happened, but the predicted mainshock is yet to come.

### 3.1. Rate Densities of Earthquake Occurrence

We begin by defining a rate of occurrence for the background seismicity (in the literature on point processes this would be

### 3.3. Functional Forms for the Foreshock Density

To evaluate the integrals in (16), we need to know the three precurrent probability densities  $\Phi_t$ ,  $\Phi_s$ , and  $\Phi_m$ . Our expressions for these incorporate our knowledge and assumptions about foreshocks. In the following sections, we describe in some detail what is known about the temporal, spatial, and magnitude dependences of foreshocks. From these data, we find functions for the relevant  $\Phi$ ; these functions must include both the actual dependence on the variables and a normalization. The nature of the normalization can be seen if we imagine extending the range of the first four integrals in (15) to cover all possible foreshocks (however we chose to define them); the resulting  $P(F)$  must then be equal to  $\alpha P(C)$ , where  $\alpha$  is, as for the one-dimensional model, the fraction of mainshocks preceded by foreshocks. In deriving our expressions we have aimed for simplicity rather than attempting to find a function that can be shown to be statistically optimal.

**3.3.1. Time.** Most foreshocks occur just before the mainshock. An increase in earthquake occurrence above the background rate has only been seen for a few days [Jones, 1984; 1985; Reasenberg, 1985] to a week [Jones and Molnar, 1979] before mainshocks. For 26% of Californian mainshocks, the foreshocks are most likely to occur within 1 hour of the mainshock; the rate of foreshock occurrence before mainshocks (Figure 1) varies with the  $t^{-1}$  type behavior also seen in Omori's law for aftershocks [Jones, 1985; Jones and Molnar, 1979]. This variation can be well fit by the function that Reasenberg and Jones [1989] found for California aftershock sequences:

$$\Phi_t(t, t') = \frac{N_t}{t' - t + c} \quad (17)$$

where  $t$  is the foreshock time and  $t'$  the mainshock time;  $c$  is a constant, found by Reasenberg and Jones [1989] to be 200 s for aftershocks. The relevant integral from (16) is then

$$\int_{t-\delta_0}^{t+\delta_0} dt \int_{t-\Delta}^{t+\Delta+\delta_1} dt' \Phi_t(t, t') = \delta_0 N_t \ln[1 + \delta_1/(\Delta + c)] \quad (18)$$

where we have assumed that  $\delta_0$  (the uncertainty of the time of

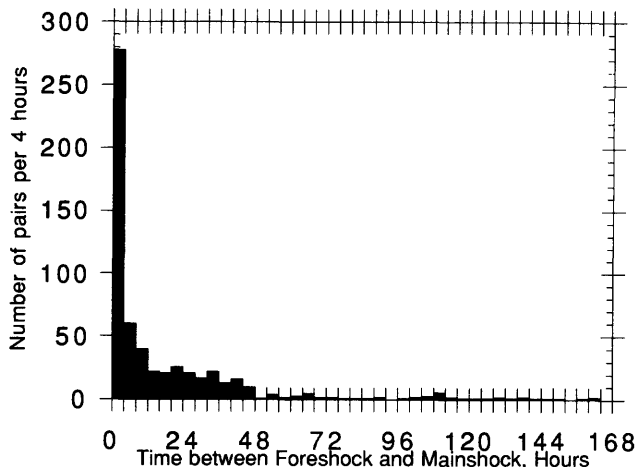


Fig. 1. The number of foreshock-mainshock pairs recorded in southern California versus the time between foreshock and mainshock in hours for foreshocks  $M \geq 2.0$  and mainshocks  $M \geq 3.0$  recorded between 1932 and 1987.

the candidate earthquake) is small. The normalization is determined by the requirement that

$$\int_{t+\delta_0}^{t+\delta_0+t_W} dt' \Phi_t(t, t') = 1 \quad (19)$$

where  $t_W$  is the total time window within which we admit preceding earthquakes to be foreshocks. This then gives

$$\begin{aligned} \int_{t-\delta_0}^{t+\delta_0} dt \int_{t-\Delta}^{t+\Delta+\delta_1} dt' \Phi_t(t, t') &= \delta_0 \frac{\ln[1 + \delta_1/(\Delta + c)]}{\ln[1 + t_W/(\delta_0 + c)]} \\ &\equiv \delta_0 I_t(\Delta, \delta_1) \end{aligned} \quad (20)$$

where, with an eye to future simplifications, we have separated out the  $\delta_0$  term. Note that (17) predicts a finite rate for all times, whereas the assumption of a limited time window automatically forces the rate to fall to zero beyond some time; we can easily modify  $\Phi_t$  to allow for this.

**3.3.2. Location.** Foreshocks not only occur close in time to the mainshock, but are also nearby in space. Jones and Molnar [1979] found that epicenters of mainshocks ( $M \geq 7$ ) and their foreshocks in the National Earthquake Information Center (NEIC) catalog were almost all within 30 km of each other, approximately the location error for the NEIC catalog. Jones [1985], with the more accurate locations of the California Institute of Technology (Caltech) catalog, found that epicenters of mainshocks ( $M \geq 3$ ) and their foreshocks were almost all within 10 km of each other; this result also held for foreshocks of  $M \geq 5$  mainshocks within the San Andreas system [Jones, 1984] if relative relocations were used. Even the largest foreshocks ( $M \geq 6$  at Mammoth Lakes and Superstition Hills) have had epicenters within 10 km of the epicenters of their mainshocks.

We have assembled a data set of sequences with high-quality locations to examine the dependence of the distance between foreshocks and mainshocks on the magnitudes of the earthquakes. This data set includes all foreshock-mainshock pairs with  $M_{\text{fore}} \geq 2.5$  and  $M_{\text{main}} \geq 3.0$  recorded in southern California since 1977 (the start of digital seismic recording), and several sequences relocated in special studies, with relative location accuracy of at least 1 km. Figure 2 shows the distance between foreshock and mainshock versus magnitude of the mainshock (2a) and magnitude of the foreshock (2b). The epicentral separation between foreshock and mainshock does not correlate strongly with either magnitude. Rather, the data seem to group into two classes: foreshocks that are essentially at the same site as their mainshock ( $<3$  km) and foreshocks that are clearly separated from their mainshocks. Only foreshocks to larger mainshocks ( $M_{\text{main}} \geq 5.0$ ) occur at greater epicentral distances (5–10 km). Of these spatially separate foreshocks some (but not all) ruptured towards the epicenter of the mainshock (the rupture zones are shown by the ovals in Figure 2). The greatest reported distance between foreshock and mainshock epicenters is 8.5 km; the greatest reported distance between foreshock rupture zone and mainshock epicenter is 6.5 km. It would therefore seem that, whatever other behavior  $\Phi_s$  may have, it can be taken to be zero for distances greater than 10 km.

It is possible (and allowed for in our choice of variables for  $\Phi_s$ ) for foreshocks to be preferentially located in some sections

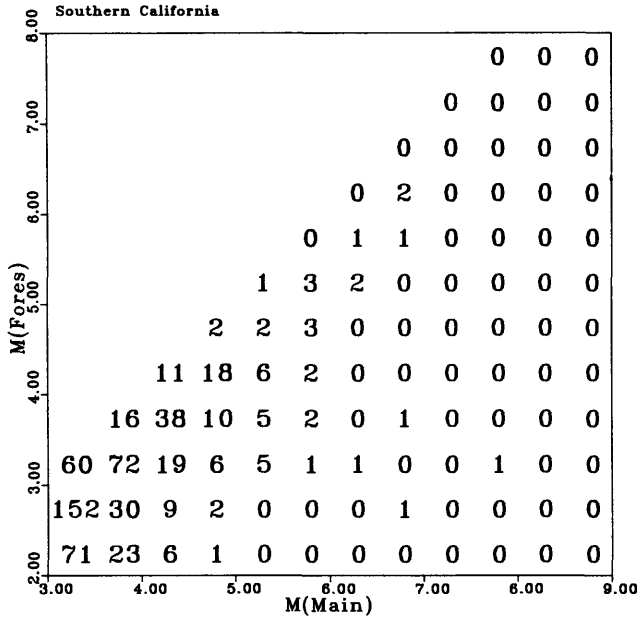


Fig. 3. The number of foreshock-mainshock pairs in half unit of magnitude bins for the magnitudes of foreshock and mainshock. Data included all  $M \geq 2.0$  foreshocks and  $M \geq 3.0$  mainshocks recorded between 1932 and 1987 in southern California.

Equation (24) says that if we look before all mainshocks with magnitudes greater than  $M_B$  for foreshocks above a cutoff magnitude of  $M_D$ , we find that a fraction  $\alpha$  of the mainshocks have foreshocks. Note that we have chosen to normalize  $\Phi_i$  and  $\Phi_s$  to integrate to 1, so  $\Phi_m$  contains the information about the total fraction of mainshocks with foreshocks.

Making  $\Phi_m$  constant implies that the fraction of mainshocks preceded by foreshocks will increase as the magnitude threshold for foreshocks decreases. This is consistent with reported foreshock activity, since the data suggest that foreshocks are relatively common before major strike-slip earthquakes. Jones and Molnar [1979] found that 30% of the  $M \geq 7.0$  earthquakes occurring outside of subduction zones were preceded by foreshocks in the NEIC catalogue ( $M \geq 4.5$ – $5.0$ ) and almost 50% had foreshocks  $M \geq 2$  reported in the literature. Jones [1984] showed that half of the  $M \geq 5.0$  strike-slip earthquakes in California were preceded by  $M \geq 2.0$  foreshocks. (Foreshocks were less common on thrust faults.)

For  $\Phi_m$  constant and equal to  $N_m$ , (24) implies that

$$N_m = \frac{\alpha\beta'}{1 + \beta'(M_B - M_D)} \quad (25)$$

The data presented by Jones [1984], with  $M_B = 5.0$  and  $M_D = 2.0$ , gave  $\alpha$  equal to 0.5 for strike-slip earthquakes. Adopting this value, with a  $\beta'$  of 2.3, gives  $N_m = 0.15$ . A consequence of taking  $\Phi_m$  constant is then that all earthquakes should have foreshocks within 6.5 units of magnitude of the mainshock. Holding  $\Phi_m$  constant for all  $M$  would of course lead to the absurd result that more than 100% of mainshocks have foreshocks within, say, 8 magnitude units. For the smaller range of magnitudes considered here a constant  $\Phi_m$  does not present any difficulties.

The integral needed for (16) is then

$$\int_{M-\mu}^{M+\mu} dM \int_{M_C}^{M_C+\mu_C} dM' \Phi_m(M, M') e^{-\beta'M'} = 2\mu N_m e^{-\beta'M_C} \frac{1 - e^{-\beta'\mu_C}}{\beta'} \equiv 2\mu I_m(M_C, \mu_C) \quad (26)$$

where we have assumed  $\mu$  small, and again separated it out from the rest of the expression.

### 3.4. Mainshock Probability

We now can combine the integrals in (18), (23), and (26) into (16) to get the foreshock probability:

$$P(F) = 4\delta_0\mu e_0 I_t I_s I_m$$

Solving the integral in (8) for the background event gives

$$P(B) = 4\delta_0\mu e_0 \Lambda_s(x_0) e^{-\beta M}$$

We substitute these values of the background and foreshock probabilities into (6) to obtain:

$$P(C|F \cup B) = \frac{I_s I_t I_m}{I_s I_t I_m + \Lambda_s(x_0) e^{-\beta M}} \quad (27)$$

The candidate earthquake errors  $\delta_0$ ,  $e_0$ , and  $\mu$  have canceled out.

For making calculations, it is also useful to set  $I_t$  equal to 1 (solve for the probability in a fixed time interval) and (for the case of a linear fault) take  $I_s$  in (23) to be equal to  $\Omega_s(x_0)$ . If we take  $\Omega_s$  to be constant and combine (14) and (26), we find that the dependence on  $M_C$  and  $\mu_C$  cancels out, and we are left with

$$P(C|F \cup B) = \frac{(N_m P(C)/A_C \delta_1)}{(N_m P(C)/A_C \delta_1) + \Lambda_s(x_0) e^{-\beta M}} \quad (28)$$

## 4. APPLICATION TO THE SAN ANDREAS FAULT SYSTEM, CALIFORNIA

We now have an expression for the conditional probability of a characteristic earthquake on a fault segment given the occurrence of an earthquake that is either a background event or a foreshock. To evaluate this, we need the long-term probability of the characteristic mainshock (the terms involving the actual magnitude of the characteristic earthquake have canceled out), the length of the fault segment, and the rate density of background seismicity for that segment. To show how this works, we now apply this to the San Andreas fault system in California, because the long-term probabilities for characteristic earthquakes that we need have been estimated for the major faults of this system, the San Andreas, Hayward, San Jacinto and Imperial faults. This was first done by Lindh [1983] and Sykes and Nishenko [1984], and more recently by the Working Group on California Earthquake Probabilities [1988], hereafter referred to as WGCEP-88.

Our division of the fault into segments and our values of  $P(C)$  for each segment come largely from WGCEP-88. One exception is that the lengths of the Southern Santa Cruz Mountains and the San Francisco Peninsula segments have been altered to match the rupture zone of the 1989 Loma Prieta earth-

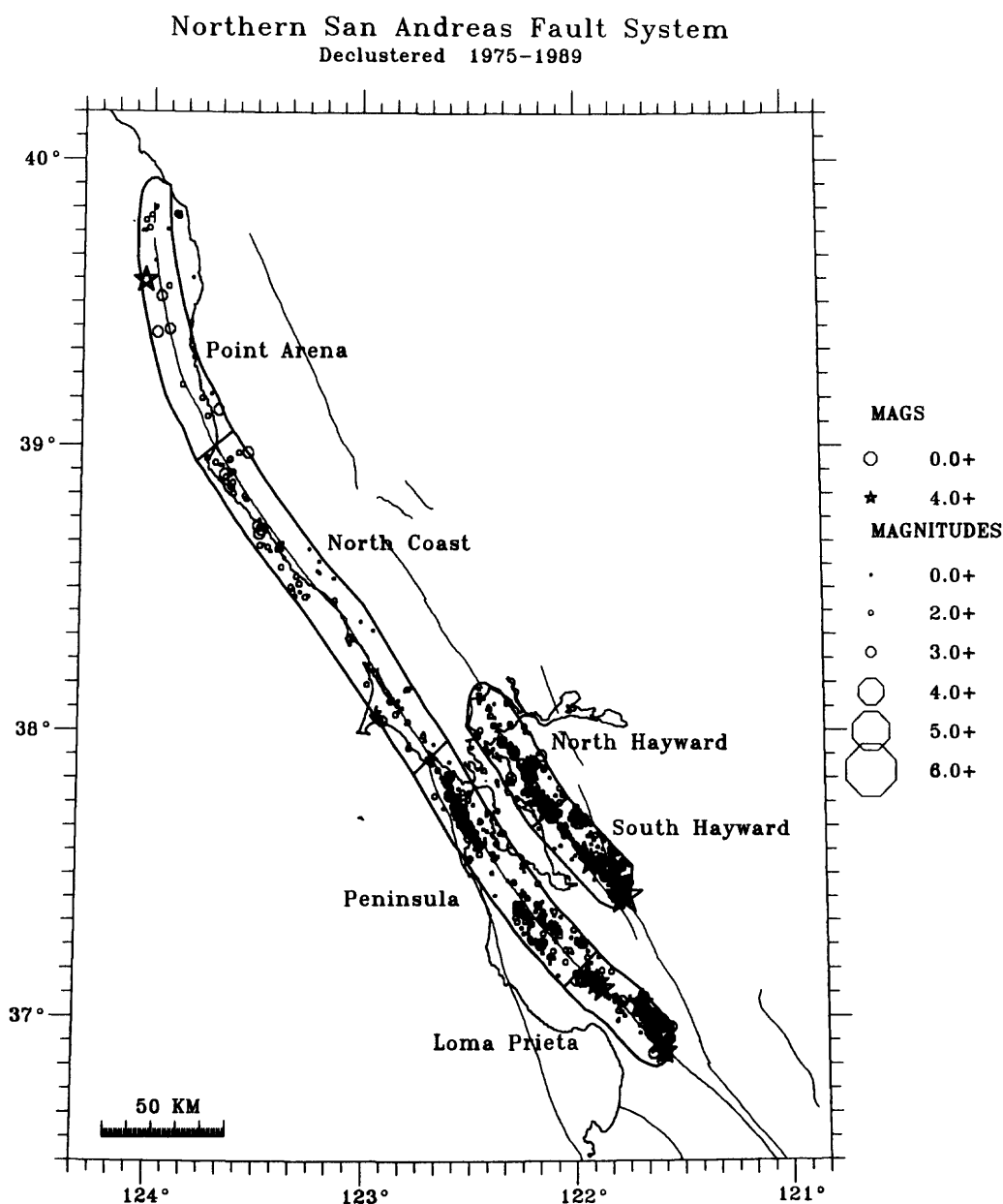


Fig. 5. A map of  $M \geq 1.8$  declustered earthquakes located within 10 km of the San Jacinto fault recorded in the Caltech catalog between 1977 and 1987.

the candidate earthquake for each segment.

We have treated the Parkfield segment in two different ways. In the Table 1 listing for Parkfield, we treat it in the same way as the other segments, regarding the foreshock as equally likely anywhere along the segment, and taking  $P(C)$  from WGCEP-88. These assumptions give short-term probabilities much lower than those estimated by Bakun *et al.* [1987] for the Parkfield earthquake prediction experiment. Bakun *et al.* [1987] used a somewhat different methodology and also used different assumptions in two areas: their value of  $P(C)$  is 1.5 times that of WGCEP-88, and they assume that the foreshock will be located in a small region under Middle Mountain, making a smaller area for defining background seismicity. (Their assumption that 50% of Parkfield mainshocks will be preceded by foreshocks agrees with our choice in section 3.2.3). For a better comparison we have used the Bakun *et al.* assumptions to deter-

mine short-term probabilities with our methodology and given these in Table 1 as Middle Mountain probabilities. These remain lower than the Bakun *et al.* results; for example, a magnitude 1.5 shock gives a probability of 0.1% from our methodology and 0.68% (Level D alert) according to Bakun *et al.*

As with the long-term probabilities of major earthquakes, these short-term foreshock-based probabilities are better seen as a means of ranking the relative hazard from different sections of the faults than as highly accurate absolute estimates. The probabilities are as uncertain as the data used to calculate them, which in some cases are uncertain indeed. For example, the values of  $P(C)$  found by WGCEP-88 are up to a factor of 4 larger than those found by Davis *et al.* [1989]; this would lead to similarly large differences in the short-term probabilities.

The relative short-term probabilities for different segments shown in Table 1 and Figure 7 are strongly affected by both the

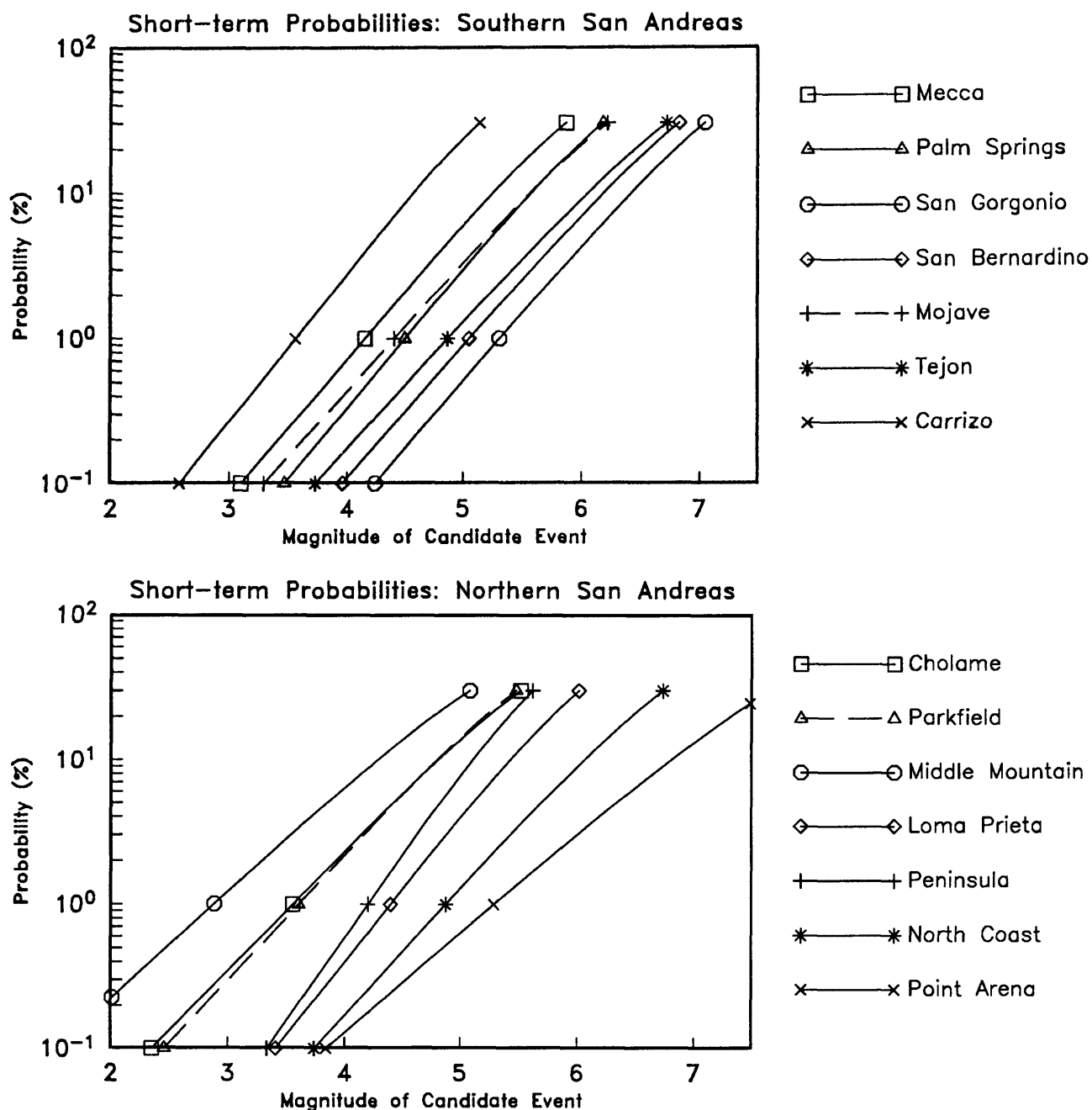


Fig. 7. The probability that an earthquake of given magnitude is a foreshock to a characteristic mainshock, plotted against magnitude for each fault segment listed in Table 1. Shown are results for the (a) southern San Andreas, (b) northern San Andreas, and (c) other faults of the San Andreas system.

long-term probability  $P(C)$  and the rate of background seismicity. Outside of the Parkfield and "Middle Mountain" segments, which have very high  $P(C)$ , the highest short-term probabilities are from the Carrizo and Cholame segments; even though the 30-year probability is only 10% on the Carrizo segment, the background seismicity there is almost nonexistent. The San Francisco Peninsula and the San Bernardino Mountain segments both have a 30-year probability of 20%, but the probabilities in the San Gorgonio subregion are much lower than those near San Francisco because of higher background seismicity. At high magnitudes, the lowest probabilities are for the Point Arena segment because of its very low  $\beta$ , which may be a result of catalog incompleteness at low magnitudes.

The possibility of the next Parkfield earthquake triggering a larger earthquake on the Cholame segment has been much discussed. Our procedure gives a magnitude 6 in Cholame a 52% chance of being a foreshock to a characteristic mainshock on that segment; but this result comes from the low background rate for the Cholame segment itself. Since this rate predicts a magnitude 6 shock every 1400 years, not every 22 years as at Parkfield, this high probability does not apply to a possible Parkfield trigger. We can, however, use (3) of our zero-dimensional model to roughly estimate the probability that a Parkfield earthquake will be a foreshock to a larger earthquake at Cholame. The WGCEP-88 probability of a Cholame earthquake is 30% in 30 years, while the background rate for

by an earthquake. (For example, *Anderson* [1982] computes the probabilities of a precursor being useful or useless). For seismicity, a division into background and precursory events appears to be a better approximation to the likely physics. Most of these papers also deal with the case (not discussed here) of how possible multiple precursors could increase the conditional probability above that for a single precursor. The discussion above suggests that this will usually be a moot point, since only rarely do we have the information needed to estimate the conditional probabilities. With the exception of the work of *Kagan and Knopoff* [1987] and (in part) *Anderson* [1982], there does not seem to have been much consideration of any multidimensional cases of the kind described in section 3. The Kagan and Knopoff study is closest to the approach presented here, though the functional form employed by them is derived from a fracture mechanics model, whereas ours is more purely empirical. The models also differ considerably in their specification of long-term probability. In the Kagan and Knopoff model, this is given by a Poisson rate derived from the frequency-magnitude relation (10), whereas here it can be independent of that. As noted in section 2, such independence appears to be a more satisfactory representation of the seismicity of an active fault zone.

## 6. CONCLUSIONS

We have shown that the probability that an earthquake that occurs near a major fault will be a foreshock to the characteristic mainshock depends on the rate of background earthquake activity on that segment, the long-term probability of the mainshock, and the rate at which the mainshocks are preceded by foreshocks, which we call the precurrent probability. Assuming certain reasonable forms for the density function of this probability (as a function of time, location, and magnitude) we have found an expression for the short-term probability that an earthquake is a foreshock, and applied it to the faults of the San Andreas system. Because the rate of foreshocks before mainshocks is assumed to be the same for all segments, the differences in short-term probabilities between segments arise from differences in background rate of seismicity and in long-term probabilities. The background rates are more variable between regions and lead to larger variations in short-term probabilities. For the San Andreas fault the two extremes are the nearly aseismic Carrizo Plain, where a 1% probability for a characteristic earthquake would be found for a magnitude 3.6 candidate event, and the highly seismic San Geronio region, where it would take a magnitude 5.3 to reach this level.

**Acknowledgements.** We thank the members of the Working Group on Short-term Earthquake Alerts for the Southern San Andreas Fault, especially Brad Hager and Dave Jackson, for raising some of the issues that led to this paper. We have benefited greatly from reviews by Andy Michael, Dave Jackson (again), Al Lindh, and especially Mark Mathews. We also thank Paul Reasenberg for comments and for providing the declustered CALNET data. Preparation of this paper was in part supported by U.S. Geological Survey grant 14-08-0001-G1763.

## REFERENCES

- Aki, K., A probabilistic synthesis of precursory phenomena, in *Earthquake Prediction: An International Review*, edited by P. Richards, pp. 566–574, American Geophysical Union, Washington D. C., 1981.
- Anderson, J. G., Revised estimates for the probabilities of earthquakes following the observation of unreliable precursors, *Bull. Seismol. Soc. Amer.*, 72, 879–888, 1982.
- Bakun, W. H., K. S. Breckenridge, J. Bredehoeft, R. O. Burford, W. L. Ellsworth, M. J. S. Johnston, L. Jones, A. G. Lindh, C. Mortenson, R. J. Mueller, C. M. Poley, E. Roeleffs, S. Schulz, P. Segall, and W. Thatcher, Parkfield, California, earthquake prediction scenarios and response plans, *U. S. Geol. Surv. Open-file Rep.* 86-365, 1987.
- Bakun, W. H., G. C. P. King, and R. S. Cockerham, Seismic slip, aseismic slip, and the mechanics of repeating earthquakes on the Calaveras fault, California, in *Earthquake Source Mechanics (Geophysical Monograph 37)*, edited by C. Scholz, pp. 195–207, American Geophysical Union, Washington D. C., 1986.
- Brune, J. N., Implications of earthquake triggering and rupture propagation for earthquake prediction based on premonitory phenomena, *J. Geophys. Res.*, 84, 2195–2197, 1979.
- Davis, P. M., D. D. Jackson, and Y. Kagan, The longer it has been since the last earthquake, the longer the expected time till the next?, *Bull. Seismol. Soc. Amer.*, 79, 1439–1456, 1989.
- Given, D. D., L. K. Hutton, L. Stach, and L. M. Jones, The Southern California Network Bulletin, January–June, 1987, *U.S. Geol. Surv. Open-file Rep.* 88-408, 1988.
- Goltz, J., The Parkfield and San Diego earthquake predictions: a chronology, *Special Report by the Southern California Earthquake Preparedness Project*, Los Angeles, Calif, 1985.
- Grandori, G., E. Guagenti, and F. Perotti, Some observations on the probabilistic interpretation of short-term earthquake precursors, *Earthq. Eng. Struct. Dynam.*, 12, 749–760, 1984.
- Guagenti, E. G. and F. Scirocco, A discussion of seismic risk including precursors, *Bull. Seismol. Soc. Amer.*, 70, 2245–2251, 1980.
- Jones, L. M., Foreshocks (1966–1980) in the San Andreas System, California, *Bull. Seismol. Soc. Amer.*, 74, 1361–1380, 1984.
- Jones, L. M., Foreshocks and time-dependent earthquake hazard assessment in southern California, *Bull. Seismol. Soc. Amer.*, 75, 1669–1680, 1985.
- Jones, L. M. and A. G. Lindh, Foreshocks and time-dependent conditional probabilities of damaging earthquakes on major faults in California, *Seismol. Res. Letters*, 58, 21, 1987.
- Jones, L. M. and P. Molnar, Some characteristics of foreshocks and their possible relationship to earthquake prediction and premonitory slip on faults, *J. Geophys. Res.*, 84, 3596–3608, 1979.
- Kagan, Y. and L. Knopoff, Earthquake risk prediction as a stochastic process, *Phys. Earth Planet. Int.*, 14, 97–108, 1977.
- Kagan, Y. and L. Knopoff, Statistical short-term earthquake prediction, *Science*, 236, 1563–1567, 1987.
- King, G. C. P. and J. Nábelek, Role of fault bends in the initiation and termination of rupture, *Science*, 228, 984–987, 1985.
- Knott, C. G., *The Physics of Earthquake Phenomena*, 278 pp., Clarendon Press, Oxford, 1908.
- Lindh, A. G., Preliminary assessment of long-term probabilities for large earthquakes along selected segments of the San Andreas fault system in California, *U.S. Geol. Surv. Open-File Rep.* 83-63, pp. 1–15, 1983.
- Nishenko, S. P. and R. Buland, A generic recurrence interval distribution for earthquake forecasting, *Bull. Seismol. Soc. Amer.*, 77, 1382–1399, 1987.
- Reasenberg, P., Second-order moment of Central California seismicity, 1969–1982, *J. Geophys. Res.*, 90, 5479–5496, 1985.
- Reasenberg, P. A. and L. M. Jones, Earthquake hazard after a mainshock in California, *Science*, 243, 1173–1176, 1989.
- Silverman, B., *Density Estimation*, 175 pp., Chapman and Hall, London, 1986.
- Sykes, L. and S. P. Nishenko, Probabilities of occurrence of large plate-rupturing earthquakes for the San Andreas, San Jacinto, and Imperial faults, California 1983–2003, *J. Geophys. Res.*, 89, 5905–5927, 1984.
- Vere-Jones, D., Earthquake prediction—a statistician's view, *J. Phys. Earth*, 26, 129–146, 1978.
- Wesnowsky, S. G., C. H. Scholz, K. Shimazaki, and T. Matsuda, Earthquake frequency distribution and faulting mechanics, *J. Geophys. Res.*, 88, 9331–9340, 1983.
- Working Group on California Earthquake Probabilities, Probabilities of large earthquakes occurring in California on the San Andreas fault, *U.S. Geol. Surv. Open-File Rep.* 88–398, 1988.
- D. C. Agnew, Institute of Geophysics and Planetary Physics, University of California, San Diego, La Jolla, CA 92093-0225.
- L. M. Jones, U.S. Geological Survey, 525 S. Wilson Avenue, Pasadena, CA 91106

(Received August 17, 1990;  
revised January 14, 1991;  
accepted January 18, 1991.)