U.S. DEPARTMENT OF THE INTERIOR

U.S. GEOLOGICAL SURVEY

Computer Programs ASPAR, GSAS and ENAS and APROB For the Statistical Modeling of Aftershock Sequences and Estimation of Aftershock Hazard

by

Paul A. Reasenberg

Open-File Report 94-221

1994

Menlo Park, CA 94025

TABLE OF CONTENTS

PAGE

INTRODUCTION

The programs ASPAR, ENAS, APROB and GSAS use a statistical model to represent an aftershock sequence. The model, which is described in the report "Earthquake Hazard After a Main Shock in California" (Reasenberg and Jones, 1989; also see Reasenberg and Jones, 1994), is based on the Gutenberg-Richter relation for the magnitude distribution of aftershocks and on Omori's Law for the dependence of the rate of aftershocks on time. These reports, which are included here in Appendix I, present the model and describe its application to California aftershock sequences, from which a "generic California model" was developed. The nomenclature used in these reports is adopted here.

The computer programs ASPAR, ENAS, APROB and GSAS were developed as a group during the late 1980's and early 1990's. Versions of ENAS and APROB dated earlier than February, 1994, are known to include a computational error, so be sure you are using current versions of the programs. This report is not intended to fully document every aspect of these programs, but rather to briefly describe their function so that a seismologist may be able to decide whether the programs suit a particular need. Because these programs are research tools, their performance may depend on the data to which they are applied. For example, I have come across a small number aftershock sequences that hang-up the subroutine that estimates the Omori's Law parameters. This routine uses a method of successive approximation that, for some particular set of aftershock times apparently fails to converge; interruption of the program is needed in these few cases. Other unexpected and possibly undesirable behavior is possible. Therefore, it is vital that users of these programs scrutinize every result.

Graphical output is provided for this purpose, and each model determined should be checked closely for reasonableness by comparison with the data and other computational means. In other words, **these programs are not intended for production use**.

As with most research developments, these programs build on the works of others. The model is based on statistical relations describing earthquake occurrence introduced by Omori and Utsu (1970, 1971, 1972, 1974), and Gutenberg and Richter (1954), Richter (1958). The statistical estimation of the model parameters is based on maximum likelihood techniques introduced by Aki (1965) and Ogata (1983). Carl Kisslinger (personal communication) contributed significantly to the computer implementation of Ogata's method. Bob Page (personal communication) wrote an earlier version of the routine for estimating the Omori's law parameters that I included in early versions of ASPAR, including the version used in preparing Reasenberg and Jones (1989). The estimation of b-values using Aki's method is based on code written by Fred Klein (personal communication).

The names of the programs are acronyms:

ASPAR: Aftershock Sequence model PARameter estimation

GSAS: Generation of Synthetic Aftershock Sequences

ENAS: Expected Number of AfterShocks

APROB: Aftershock PROBabilities

Programs ASPAR, ENAS, APROB and GSAS have been used at the U.S. Geological Survey in Menlo Park and Pasadena, and other seismological centers, for four distinct purposes, which I briefly describe below.

1. **Short-term hazard assessment.** The programs provide rapid estimates of the probability of aftershocks (including potentially damaging ones) for hazard assessment purposes. These estimates of probabilities have been released as hazard advisories to the California Office of Emergency Services and as forecasts to the public and news media after significant earthquakes in California. The advisories and forecasts have been useful in communicating to the public the hazard associated with aftershocks. For example, the aftershock hazard forecasts were used to guide businesses and individuals in deciding when and whether to reoccupy certain damaged buildings after the 1989 Loma Prieta (M7.1) and 1994 Northridge (M6.7) California earthquakes.

The effective translation of aftershock probabilities into understandable hazard forecasts is critical. Some people apparently do not understand the meaning of statements involving probabilites, and considerable public confusion has arisen when the news media have reported our aftershock hazard forecasts. For example, one individual thought the statement "There is a 50 percent chance of a damaging (M>5) aftershock in the next three weeks" meant that it would be safe for the next three weeks, after which time a damaging aftershock might occur. A discussion of the pitfalls encountered and lessons learned in communicating hazard forecasts after the 1989 Loma Prieta earthquake is given in Reasenberg (1990b). Since then, our experience with the 1994 Northridge earthquake (Reasenberg and Jones, 1994) has taught us new lessons in hazard communication, and we are still very much in the early stages of this learning process.

2. **Logistical planning of portable seismograph deployments.** Program ENAS provides estimates of the expected number of aftershocks within a specified size range during a specified time interval. This information has been used by seismologists in planning the deployment of portable seismograph equipment in the epicentral area after a large earthquake. Such estimates may guide both the choice of instrument type and the length of deployment. For example, on March 8, 1994, the following table was generated by program ENAS to guide a second-wave deployment of portable seismographs in the epicentral region of the Northridge, California (17 January, 1994; M=6.7) earthquake:

Table 1

1994 Northridge (M 6.7) Earthquake
Expected Number of Aftershocks
(Based on a Reasenberg-Jones Model Using M>3 Events)

Model parameters: $a = -1.31$, $b = 0.91$, $p = 1.20$, $c = 0.18$

Observation Period (1994)	M>3	M>3.5	M>4	M>4.5
3/9 - 3/31	18	6	2	0.8
3/9 - 4/30	34	12	4	1
3/9 - 5/31	45	16	6	2

3. **Intermediate-term hazard assessment.** The programs provide an estimate of the probability of an equal- or larger-magnitude earthquake in the intermediate term (months to years) following a large earthquake. For example, after the 1992 Landers (M 7.4) earthquake in southern California, program ASPAR was used to estimate the probability of additional $M \geq 7$ and $M \geq 7.5$ earthquakes in southern California for use in an assessment of seismic hazards in southern California (SCEC Phase I report, 1992).

4. **Research.** Programs ASPAR, ENAS, GSAS and APROB have been used to systematically study the characteristics of and variations among aftershock sequences as a function of time, location and tectonic environment. For example, Reasenberg (1990, 1990a) and Kisslinger and Jones (1991) found a positive correlation between the Omori's law parameter p and the regional heatflow. White and Reasenberg (1991) found variations in p that they tentatively related to stress heterogeneity in the Garm region, USSR. Reasenberg et al. (1990) found that the generic aftershock sequence in the shallow crust in central Japan is virtually the same as that in California. Arabasz and Hill (1994) found the aftershock sequences in the Intermountain Seismic Belt in Utah to be less productive, by a factor of approximately 4, compared to the generic California model.

Program ASPAR

Program ASPAR estimates the parameters b, p, a, K and c (see Appendix I for definitions of model parameters) using maximum likelihood estimation techniques and displays the data and the estimated model in graphical form. This program is an interactive, graphical program using X-windows graphics. It currently runs on a Sun Sparcstation computer under the Openwindows graphical user interface. The main components of the program are written in Fortran, and in principle could be converted to another platform. The main challange in such a conversion will be that of interfacing with another graphical user interface.

Program ASPAR is a UNIX shell script program that runs all the software components, controls the user interaction, and handles the file management. From within this script the following programs are called, according to the interactive responses: Program ASPAR3X, a Fortran program that performs the basic modeling; Utility program AI2PS, a shell script that converts the graphics file produced by ASPAR3X to the PostScript language; and program APROB4B, a Fortran program that computes earthquake probabilites. It is not necessary to use the shell script ASPAR; instead, the component programs may be run separately. Usually, however, the shell script ASPAR is used.

Program ASPAR3X is the basic modeling program used to model an aftershock sequence and estimate probabilites for future aftershocks. The input to this program is a single file containing, in chronological order, the times and magnitudes of all the events in one aftershock sequence. The first record must be the main shock. Following the main shock are the data for the aftershocks. Several formats of data are accepted by the program, including hypo71, Caltech, U.C.Berkeley, hypoinverse and others, selectable interactively. When running ASPAR3X, the user must (1) name the hypocenter file (also called the "summary file"); (2) specify the data format type; (3) specify the minimum magnitude cutoff (aftershocks listed in the input file with magnitudes smaller than the cutoff will not be used); (4) specify the amount of time that the graphics will be left on the screen after calculation is complete. After this, the program will open a new X-window for plotting and display the earthquake data in the form of a "stick plot" (magnitudes vs. time) and a graph of earthquake rate versus time (both using log-time axes). At this time, the user must interactively select the time interval defining the data that will be used in the model. If it is believed that the data set is incomplete, for example, during the first day, or if a strong, secondary aftershock sequence occured at a later time, these features can be excluded from the model in this step. If all the data are acceptable, you may select all the data by spanning a time interval that includes all the data in the plot. To do this, place the cursor at the time of the earliest data to use, and click the "select" mouse button (left mouse button). Then, move the cursor to the time of the latest data to use, and again click the "select" mouse button. Next, the program calculates the Omori's law parameters p and c and displays their values on the seismicity rate graph. Next the program plots the magnitude distribution. Now you must select the minimum magnitude earthquake to use in the estimation of the Gutenberg-Richter parameter b. Select a magnitude above which the data appear to be complete by placing the cursor at the desired magnitude and clicking the "select" mouse button. Now, the program draws the time decay model and magnitude distribution model curves and makes a cumulative number plot of the earthquake sequence (with log-time axis). This completes the modeling procedure.

If you are running the shell script ASPAR, the graphics display will remain on the screen for the time period selected in step (4) above. After the X-window closes, a prompt appears in the text window asking if you want to repeat the modeling step. Answer "y" if you want to repeat the modeling procedure, "n" if you are satisfied with the current model. If you answer "y", a new X-window will appear, and the above steps will be repeated.

Next, a prompt appears in the text window asking "Do you want a hardcopy of the model graphs? (y/n)". Answer "y" to produce a hard copy and automatically send it to "lpr" (UNIX name for the default printer device. In the current setup, this device must be a PostScript-compatible printer.).

Next, a prompt appears in the text window asking "Do you want to print the model summary table? (y/n)". Answer "y" to produce a one-page summary of the calculated model parameters and automatically send it (a text file) to "lpr".

Next, a prompt appears in the text window asking "Do you want to calculate earthquake probabilities? (y/n)". If you answer "y" here, a new program will be run (program APROB4B) that produces a table of earthquake probabilites. Program APROB4B calculates interval probabilites as defined in Equation 4 in the *Science* report. APROB4B reads the model parameters that were just produced by program ASPAR3X from a temporary file called junkpar.tmp, which ASPAR3X creates. When APROB4B (or any other version of APROB) runs, you must first specify a title for the table; then you will have to specify when (in days relative to the mainshock) the table should begin. For example, if you are running shell script ASPAR two weeks after a mainshock, you might specify that the table begin at 14 days, so the results will correspond to contemporary intervals. The table will then include probabilites for earthquakes during intervals starting 14 to 21 days after the mainshock. There are two modes of use for program APROB4B. In the first mode, the parameters calculated in program ASPAR3X are used directly in the probability calculation. However, as described in the *Science* report, it is also possible to use the Bayesian estimates of the model parameters. The Bayesian estimates of the model parameters make use of the "Generic Model" parameters determined for California, which are coded into program ASPAR3X, and reduce the uncertainty in modeling during the early stage of an earthquake sequence (see Reasenberg and Jones, 1989). For this reason I recommend that the Bayesian parameters be used in all short-term hazard assessment applications.

As mentioned above, the generic model parameters for California are coded into program ASPAR3X. The generic model parameters were obtained by modeling 62 aftershock sequences between 1933 and 1987, as described in the Reasenberg and Jones (1989). The generic model parameters for California are $a = -1.67$, $b = 0.91$, $p = 1.08$, and $c = 0.05$.

Program ENAS

Program ENAS estimates the number of aftershocks expected in an aftershock sequence in a specified time interval and magnitude range. The default values for the model parameters provided in the prompts reflect the generic model for California.

To run Program ENAS, answer each prompt. The first 4 prompts are used for setting the model paramters a, p, b and c. The default values, corresponding to the generic model for California, are obtained by hitting CR (carriage return). The next two prompts specify the magnitude range of earthquakes to be estimated, in magnitude units relative to the main shock. For example, the default values of -3.0 and 0.0 specify earthquakes with magnitudes between the mainshock magnitude and 3 units smaller than the mainshock. The values 0.0 and 9.0 would, effectively, specify all earthquakes equal to or larger than the mainshock. The next two prompts specify the time interval, in days relative to the main shock. The default values specify the first week after the main shock.

Program GSAS

Program GSAS generates synthetic aftershocks sequences. This program is not necessary for hazard estimation. However, it is a useful research tool and can be used to provide a calibration for the parameter estimation routines in program ASPAR. Program GSAS employs random number generators in the IMSL mathematical subroutines library to generate artificial aftershock sequences. The magnitudes of events in the sequences are random with exponential magnitude distribution, while the event times are generated by a time-dependent Poisson process with intensity following the modified Omori's law. The sequences generated by GSAS can be read by program ASPAR with the data format option "Synthetic".

Program APROB

The first part of program APROB is similar to program ENAS in that it starts by asking for the four model parameters a, p, b and c, and will supply the generic model parameters by default. Then it goes on to calculate the probabilities for the occurrence of one or more earthquakes in specified magnitude ranges.

Graphical Output: Four Illustrative Examples

On the following pages are shown illustrative examples of graphical output from ASPAR. I have marked explanations on the pages by hand.

Figure 1

This is an earthquake sequence following a M 5.7 earthquake in Japan. The times selected for analysis completely span all the data. Note the small tick marks at the bottom of the "stick" plot that indicate the selected minimum and maximum times, and the corresponding vertical lines in the rate plot. Note that the time scales on the "stick" plot and cumulative number plot are the same as the one shown under the "rate plot". Above the "stick" plot are shown the main shock summary record, as listed in the data input file, and the number, magnitude cutoff and time period of data read from that file. In the upper right corner of the rate plot, the model parameters p, c, K and a and their standard deviations are given. At the bottom of the rate plot two estimates of the goodness of fit between the model (curve) and the data (triangles) are given. These measures have are sensitive to different departures of the model from the data. Chi-squared is sensitive to mean-squared distance between the points and model (as might, for example, be produced by a secondary aftershock sequence), while Kolmogorov-Smirnov is relatively tolerant of such discrepancies but more sensitive to the overall similarity of the shapes of the model curve and data. On the first line, the value of Chi-squared (CHI2), number of degrees of freedom (DF) and confidence (p) for accepting the model are given. (This is a different use of the symbol "p" from the

Omori parameter). On the next line the value of the Kolmogorov-Smirnov statistic (K/S), number of points used in the calculation (N), and confidence (p) for accepting the model are given. Both the chi-squared and Kolmogorov-Smirnof statistic indicate an acceptable fit when $p \geq 0.05$. Scrutinizing both measures of goodness of fit is necessary, and if either measure indicates poor fit, an attempt to remodel the data with different time and magnitude cutoffs should be made.

Above the magnitude distribution plot the estimate of the parameter b and its standard deviation are given. The line below this indicates the magnitude cutoff level chosen for this estimate (M 2.0), and the number of earthquakes satisfying this criterion (218). The line below this indicates how many earthquakes appear to be "missing" (assuming the modeled magnitude distribution), between the data cutoff specified when the data were initially read into the program ($M \geq 1.0$ in this case) and the b-value estimation magnitude cutoff ($M \geq 2.0$ in this case).

Figure 2

This is the first of three examples based on the Loma Prieta, California, M 7.1 earthquake sequence. In this example, all of the available data were selected (see the little tick marks below the "stick" plot). Notice that the fit of the model to the data is not acceptable: Chi-squared test rejects the model at $p < 0.001$ (a rounded off "p=0.000" is printed), and a Kolmogorov-Smirnov test rejects the model at $p < 0.001$. The reason for the poor fit is mainly the strong secondary aftershock sequence following the M5.7 earthquake on April 18, 1990 (157 days after the mainshock).

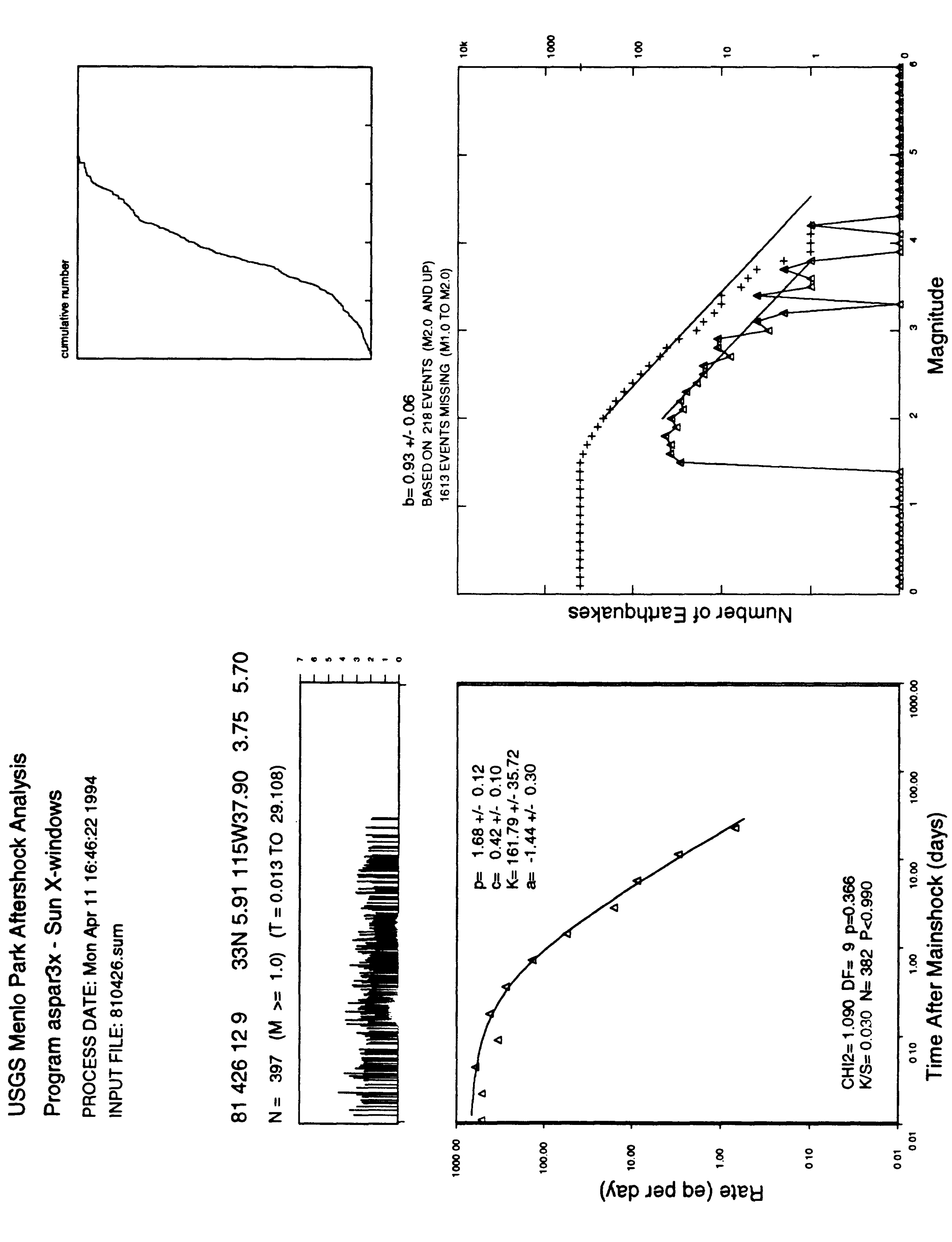
Figure 3

This is the same example shown in Figure 2, except data after April 18, 1990 were excluded (see tick marks under "stick" plot and vertical lines in rate plot, which define the selected interval). With the elimination of the secondary aftershock sequence, the fit of the model to the data is improved (Kolmogorov-Smirnov test accepts the model; but Chi-squared test still rejects). It seems Chi-squared rejects the model owing to discrepancies in the first half-day after the main shock. Often, data are incomplete in network processing in the first day after a large earthquake, so this is not surprising. Note that excluding the secondary aftershock sequence raised the estimated value of p. This is not surprising, since it is well known that inclusion of secondary aftershock sequences artificially lowers the apparant value of the Omori parameter p.

Figure 4

This is the same example as shown in Figure 3, except that data before approximately 0.2 days are excluded. Now both Kolmogorov-Smirnov and Chi-squared tests accept the model.

FIGURE 1



USGS Menlo Park Aftershock Analysis
Program aspar3x - Sun X-windows
PROCESS DATE: Mon Apr 11 16:46:22 1994
INPUT FILE: 810426.sum
81 426 12 9 33N 5.91 115W37.90 3.75 5.70
N = 397 (M >= 1.0) (T = 0.013 TO 29.108)
cumulative number
p= 1.68 +/- 0.12
c= 0.42 +/- 0.10
K= 161.79 +/- 35.72
a= -1.44 +/- 0.30
CHI2= 1.090 DF= 9 p=0.366
K/S= 0.030 N= 382 P<0.990
Rate (eq per day)
Time After Mainshock (days)
b= 0.93 +/- 0.06
BASED ON 218 EVENTS (M2.0 AND UP)
1613 EVENTS MISSING (M1.0 TO M2.0)
Number of Earthquakes
Magnitude

FIGURE 2

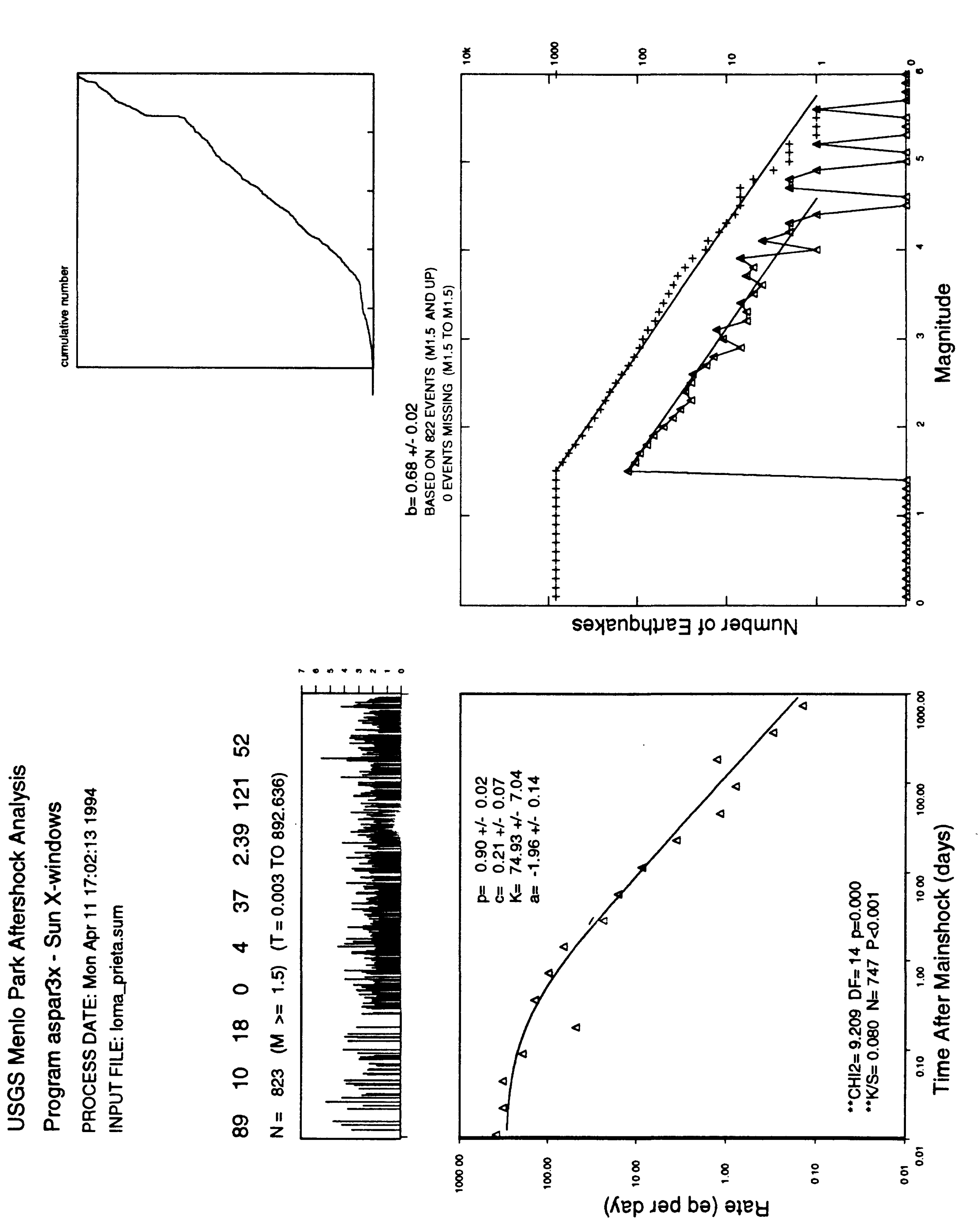

USGS Menlo Park Aftershock Analysis
Program aspar3x - Sun X-windows
PROCESS DATE: Mon Apr 11 17:02:13 1994
INPUT FILE: loma_prieta.sum
89 10 18 0 4 37 2.39 121 52
N = 823 (M >= 1.5) (T = 0.003 TO 892.636)
p= 0.90 +/- 0.02
c= 0.21 +/- 0.07
K= 74.93 +/- 7.04
a= -1.96 +/- 0.14
**CHI2= 9.209 DF= 14 p=0.000
**K/S= 0.080 N= 747 P<0.001
Rate (eq per day)
Time After Mainshock (days)
cumulative number
b= 0.68 +/- 0.02
BASED ON 822 EVENTS (M1.5 AND UP)
0 EVENTS MISSING (M1.5 TO M1.5)
Number of Earthquakes
Magnitude

FIGURE 3

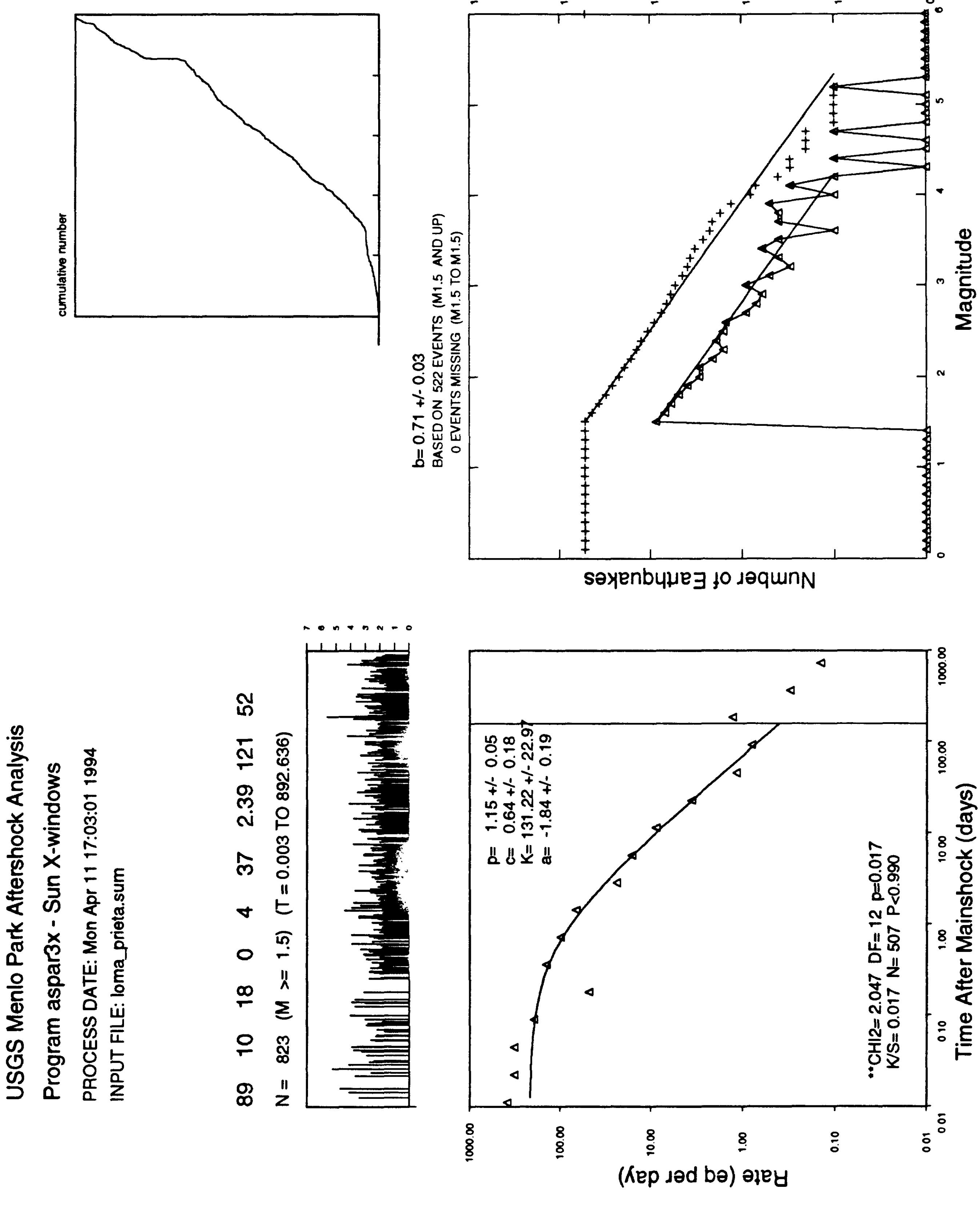
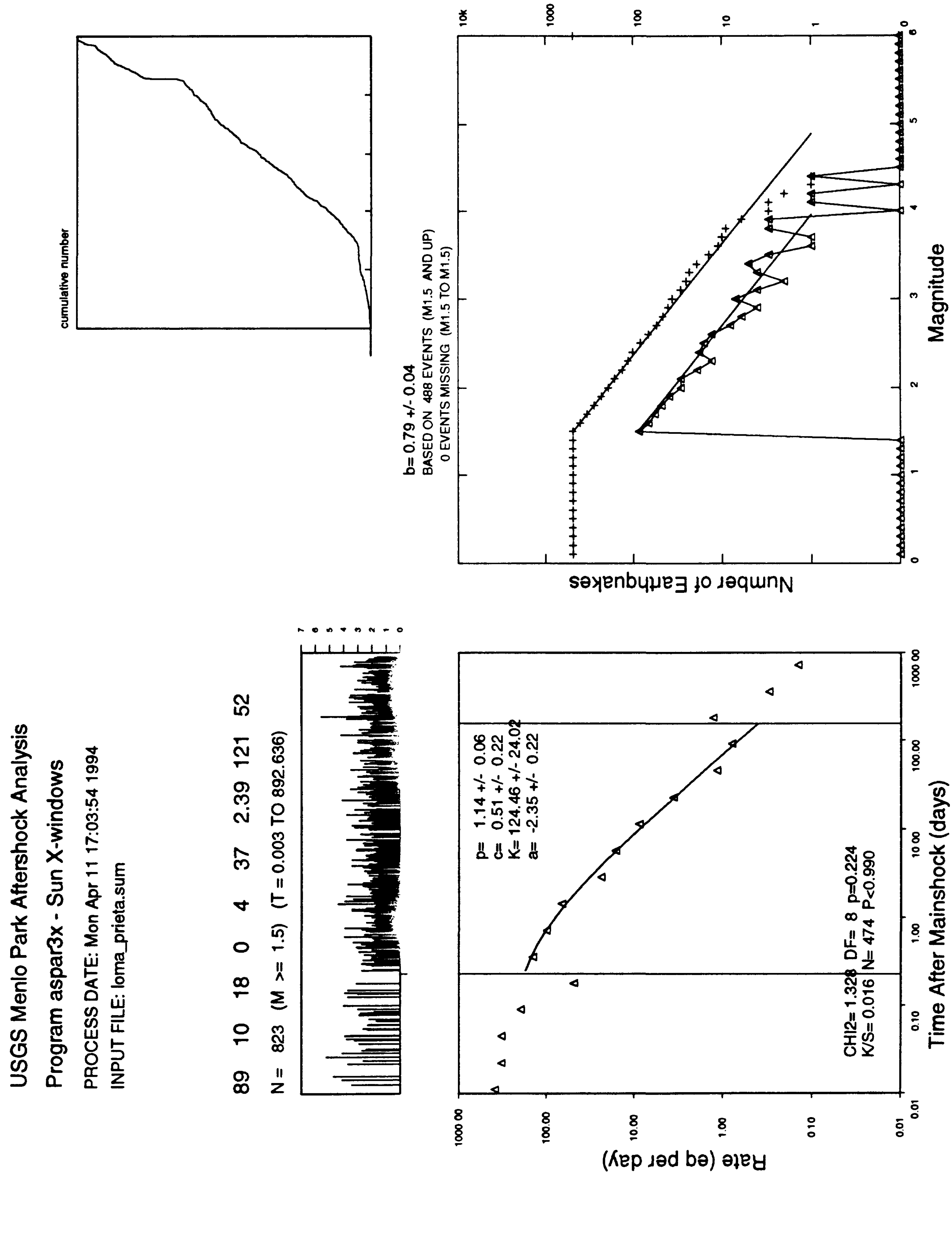

USGS Menlo Park Aftershock Analysis
Program aspar3x - Sun X-windows
PROCESS DATE: Mon Apr 11 17:03:01 1994
INPUT FILE: loma_prieta.sum
89 10 18 0 4 37 2.39 121 52
N = 823 (M >= 1.5) (T = 0.003 TO 892.636)
cumulative number
b= 0.71 +/- 0.03
BASED ON 522 EVENTS (M1.5 AND UP)
0 EVENTS MISSING (M1.5 TO M1.5)
Number of Earthquakes
Magnitude
p= 1.15 +/- 0.05
c= 0.64 +/- 0.18
K= 131.22 +/- 22.97
a= -1.84 +/- 0.19
**CHI2= 2.047 DF= 12 p=0.017
K/S= 0.017 N= 507 P<0.990
Rate (eq per day)
Time After Mainshock (days)

FIGURE 4


USGS Menlo Park Aftershock Analysis
Program aspar3x - Sun X-windows
PROCESS DATE: Mon Apr 11 17:03:54 1994
INPUT FILE: loma_prieta.sum
89 10 18 0 4 37 2.39 121 52
N = 823 (M >= 1.5) (T = 0.003 TO 892.636)
cumulative number
p= 1.14 +/- 0.06
c= 0.51 +/- 0.22
K= 124.46 +/- 24.02
a= -2.35 +/- 0.22
CHI2= 1.328 DF= 8 p=0.224
K/S= 0.016 N= 474 P<0.990
Rate (eq per day)
Time After Mainshock (days)
b= 0.79 +/- 0.04
BASED ON 488 EVENTS (M1.5 AND UP)
0 EVENTS MISSING (M1.5 TO M1.5)
Number of Earthquakes
Magnitude

Tabular Output: Summary and Earthquake Probability Tables

Table 2 summarizes the modeling of the Loma Prieta earthquake shown in Figure 4. The model parameters determined from the aftershock data in Figure 4 are listed in the "post" column (a posteriori values), while the generic California parameter values are listed in the "prior" column. The result of combining these sets of parameters using the Bayesian formula given in Reasenberg and Jones (1989) is given in the column "bayes". These are the parameters that will be used in the calculation of probabilities in Table 3.

Table 3 gives probabilities for aftershocks of the Loma Prieta earthquake. The model parameters used in generating this table are printed on the top of the page. Bayesian parameters were interactively requested at the time the program APROB4B was run (see Table 2). The table consists of 8 blocks. Each block gives probabilities for the occurrence of one or more aftershocks equal to or greater than a given magnitude ($M \geq 4.0$, $M \geq 4.5$, $M \geq 5.0$, $M \geq 5.5$, $M \geq 6.0$, $M \geq 6.5$, $M \geq 7.0$ and $M \geq 7.5$). The time interval for each probability is defined by the row and column, as follows. The time period is (S,T), in days. S is the beginning of the time period; T is the end. T-S is the duration (in days) of the time period, and is given by the row. For example, looking at the block for $M \geq 5.0$ aftershocks, we see that for the 7-day period beginning at the time of the mainshock (T - S = 7 and S = 0), the probability value 0.705 is circled. This means that, during the 1-week period immediately after the Loma Prieta earthquake, there was a 70% chance of one or more $M \geq 5$ aftershocks.

Also, in the 60-day period beginning 5 days after the main shock, there was a 35.5% chance of one or more $M \geq 5$ aftershocks.

Also, in the 30-day period beginning at the time of the main shock, there was a 3.7% chance of one or more $M \geq 7$ aftershocks.

The times corresponding to "S" (i.e., the columns) are selected interactively when program APROB4B is running, with the question "Starting how many days after main shock?".

Table 2

```
***********************************************************

                 *** program aspar3 ***

input file: loma_prieta.sum
process date: Mon Apr 11 17:03:54 1994
mainshock: 89    10    18    0    4    37    2.39  121   52.64 16.74 7.1

as of 153.10 days after the mainshock...

            --------parameter estimates--------      --weights---
            prior       post       bayes    s.d.       prior  post
            ----------------------------------------------------
 p:          1.07       1.14        1.13    0.06      (0.066 0.934)
 b:          0.90       0.79        0.80    0.04      (0.042 0.958)
 a:         -1.76      -2.35       -2.27    0.20      (0.123 0.877)
 c:          0.05       0.51        0.07    0.05      (0.950 0.050)

 **********************************************************
```

Table 3

*** PROGRAM APROB4B ***

Loma Prieta Earthquake
MAINSHOCK MAGNITUDE = 7.10
A = -2.27 P = 1.13 B = 0.80 C = 0.073

Bayesian model parameters were selected.

M > 4.00

(T-S) \ S:	0.00	1.00	2.00	3.	4.	5.	6.	7.	8.
1.	0.994	0.627	0.423	0.313	0.247	0.202	0.171	0.148	0.129
3.	0.999	0.852	0.702	0.588	0.502	0.436	0.385	0.343	0.309
7.	1.000	0.937	0.854	0.776	0.707	0.647	0.596	0.551	0.512
30.	1.000	0.986	0.963	0.937	0.912	0.886	0.861	0.836	0.813
60.	1.000	0.992	0.980	0.966	0.951	0.936	0.921	0.906	0.892
90.	1.000	0.995	0.986	0.976	0.965	0.954	0.943	0.932	0.921
365.	1.000	0.998	0.995	0.992	0.988	0.984	0.980	0.977	0.973
1000.	1.000	0.999	0.998	0.996	0.994	0.992	0.990	0.988	0.986

M > 4.50

(T-S) \ S:	0.00	1.00	2.00	3.	4.	5.	6.	7.	8.
1.	0.867	0.325	0.197	0.139	0.107	0.086	0.072	0.062	0.054
3.	0.928	0.534	0.383	0.298	0.243	0.205	0.176	0.155	0.137
7.	0.953	0.669	0.536	0.449	0.387	0.340	0.303	0.274	0.249
30.	0.975	0.816	0.731	0.669	0.620	0.579	0.545	0.515	0.488
60.	0.981	0.858	0.791	0.741	0.701	0.667	0.637	0.611	0.588
90.	0.984	0.877	0.818	0.774	0.739	0.709	0.682	0.659	0.638
365.	0.989	0.921	0.883	0.854	0.830	0.810	0.792	0.776	0.762
1000.	0.992	0.940	0.910	0.888	0.870	0.855	0.841	0.829	0.818

M > 5.00

(T-S) \ S:	0.00	1.00	2.00	3.	4.	5.	6.	7.	8.
1.	0.553	0.145	0.084	0.058	0.044	0.035	0.029	0.025	0.022
3.	0.650	0.263	0.175	0.132	0.105	0.087	0.074	0.065	0.057
7.	0.705	0.357	0.264	0.212	0.178	0.153	0.134	0.120	0.108
30.	0.772	0.492	0.408	0.357	0.320	0.292	0.270	0.251	0.234
60.	0.795	0.541	0.465	0.417	0.382	0.355	0.333	0.314	0.298
90.	0.806	0.567	0.494	0.448	0.415	0.389	0.367	0.349	0.333
365.	0.837	0.636	0.575	0.536	0.507	0.485	0.466	0.450	0.436
1000.	0.854	0.674	0.618	0.583	0.558	0.537	0.520	0.506	0.493

Table 3, continued

M > 5.50

(T-S) \ S:	0.00	1.00	2.00	3.	4.	5.	6.	7.	8.
1.	0.275	0.061	0.034	0.024	0.018	0.014	0.012	0.010	0.009
3.	0.342	0.115	0.074	0.055	0.043	0.036	0.030	0.026	0.023
7.	0.386	0.161	0.115	0.091	0.075	0.064	0.056	0.050	0.045
30.	0.445	0.237	0.189	0.162	0.143	0.129	0.118	0.109	0.101
60.	0.468	0.267	0.221	0.194	0.175	0.161	0.149	0.140	0.132
90.	0.480	0.284	0.238	0.211	0.193	0.178	0.167	0.157	0.149
365.	0.516	0.332	0.289	0.264	0.246	0.233	0.222	0.212	0.204
1000.	0.536	0.360	0.319	0.295	0.278	0.265	0.254	0.245	0.238

M > 6.00

(T-S) \ S:	0.00	1.00	2.00	3.	4.	5.	6.	7.	8.
1.	0.120	0.025	0.014	0.010	0.007	0.006	0.005	0.004	0.004
3.	0.154	0.047	0.030	0.022	0.018	0.014	0.012	0.011	0.009
7.	0.177	0.068	0.048	0.037	0.031	0.026	0.023	0.020	0.018
30.	0.210	0.102	0.080	0.068	0.060	0.054	0.049	0.045	0.042
60.	0.223	0.117	0.095	0.082	0.074	0.068	0.062	0.058	0.055
90.	0.230	0.125	0.103	0.090	0.082	0.075	0.070	0.066	0.063
365.	0.251	0.149	0.127	0.115	0.107	0.100	0.095	0.091	0.087
1000.	0.264	0.163	0.142	0.130	0.122	0.116	0.110	0.106	0.103

M > 6.50

(T-S) \ S:	0.00	1.00	2.00	3.	4.	5.	6.	7.	8.
1.	0.050	0.010	0.006	0.004	0.003	0.002	0.002	0.002	0.001
3.	0.065	0.019	0.012	0.009	0.007	0.006	0.005	0.004	0.004
7.	0.075	0.028	0.019	0.015	0.012	0.011	0.009	0.008	0.007
30.	0.090	0.042	0.033	0.028	0.024	0.022	0.020	0.018	0.017
60.	0.096	0.048	0.039	0.034	0.030	0.028	0.025	0.024	0.022
90.	0.099	0.052	0.042	0.037	0.034	0.031	0.029	0.027	0.025
365.	0.109	0.062	0.053	0.048	0.044	0.041	0.039	0.037	0.036
1000.	0.115	0.069	0.059	0.054	0.051	0.048	0.046	0.044	0.042

Table 3, continued

M > 7.00

(T-S) \ S:	0.00	1.00	2.00	3.	4.	5.	6.	7.	8.
1.	0.020	0.004	0.002	0.002	0.001	0.001	0.001	0.001	0.001
3.	0.026	0.008	0.005	0.004	0.003	0.002	0.002	0.002	0.001
7.	0.031	0.011	0.008	0.006	0.005	0.004	0.004	0.003	0.003
30.	0.037	0.017	0.013	0.011	0.010	0.009	0.008	0.007	0.007
60.	0.039	0.020	0.016	0.014	0.012	0.011	0.010	0.010	0.009
90.	0.041	0.021	0.017	0.015	0.014	0.012	0.012	0.011	0.010
365.	0.045	0.025	0.021	0.019	0.018	0.017	0.016	0.015	0.014
1000.	0.048	0.028	0.024	0.022	0.020	0.019	0.018	0.018	0.017

M > 7.50

(T-S) \ S:	0.00	1.00	2.00	3.	4.	5.	6.	7.	8.
1.	0.008	0.002	0.001	0.001	0.000	0.000	0.000	0.000	0.000
3.	0.011	0.003	0.002	0.001	0.001	0.001	0.001	0.001	0.001
7.	0.012	0.004	0.003	0.002	0.002	0.002	0.001	0.001	0.001
30.	0.015	0.007	0.005	0.004	0.004	0.003	0.003	0.003	0.003
60.	0.016	0.008	0.006	0.005	0.005	0.004	0.004	0.004	0.004
90.	0.016	0.008	0.007	0.006	0.005	0.005	0.005	0.004	0.004
365.	0.018	0.010	0.009	0.008	0.007	0.007	0.006	0.006	0.006
1000.	0.019	0.011	0.010	0.009	0.008	0.008	0.007	0.007	0.007

REFERENCES CITED

Aki, K., 1965, Maximum likelihood estimate of b in the formula log N = a - bM and its confidence limits, Bulletin of the Earthquake Research Institute, **43**, 237-239.

Arabasz, W. J., Hill, S. J., 1994, Aftershock temporal behavior and earthquake clustering in the Utah region, Seismol. Research Lett. **65**, 32.

Gutenberg, B. and Richter, C. F., 1954, Seismicity of the Earth, Princeton University Press, 2nd ed., 16-25.

Kisslinger, C., Jones, L. M., 1991, Properties of aftershock sequences in southern California, J. Geophys. Res. **96**, 11,947-11,958.

Ogata, Y., 1983, Estimation of the parameters in the modified Omori formula for aftershock frequencies by the maximum likelihood procedure, J. Phys. Earth, **31**, 115-124.

Reasenberg, P. A., 1990, Variations in the Omori decay parameter for composite earthquake clusters in central California, Seismol. Research Lett. **61**, 44.

Reasenberg, P. A., 1990a, Microclusters and compound clusters, EOS **71**, 1457.

Reasenberg, P. A., 1990b, Applications of the USGS aftershock sequence model and guidelines for drafting aftershock forecasts, U.S.G.S. Open-File Report 90-341.

Reasenberg, P. A. and Jones, L. M., 1989, Earthquake hazard after a mainshock in California, Science, 243, 1174-1176.

Reasenberg, P. A. and Jones, L.M., 1994, Earthquake Aftershocks: Update, Science, 265, 1251-1252.

Richter, C. F., 1958, Elementary Seismology, W. H. Freeman, San Francisco.

Southern California Earthquake Center, 1992, Future seismic hazards in southern California, Phase I: Implications of the 1992 Landers earthquake sequence, California Division of Mines and Geology.

Utsu, T., 1969, Journal of the Faculty of Sciences, Hokkaido University Series 7, Vol. 3, No. 3.

Utsu, T., 1970, Journal of the Faculty of Sciences, Hokkaido University Series 7, Vol. 3, No. 4.

Utsu, T., 1971, Journal of the Faculty of Sciences, Hokkaido University Series 7, Vol. 3, No. 5.

Utsu, T., 1972, Journal of the Faculty of Sciences, Hokkaido University Series 7, Vol. 4, No. 1.

White, R. A., Reasenberg, P. A., 1991, Variations in the Omori decay rate in the Garm, USSR: possible dependence on changing stress, EOS **72**, 332.

APPENDIX I

In 1994, an error was discovered in earlier versions of the programs ENAS and APROB. Unfortunately, the erroneous version of program APROB was used in preparing Tables 1 and 2 in Reasenberg and Jones (1989). Thus, these tables in the 1989 Science Report are in error and should not be used. The corrected version of Table 1 in the 1989 paper is provided in Reasenberg and Jones (1994), which is reprinted below together with the original (1989) Science Report.

Reprint Series
3 March 1989, Volume 243, pp. 1173–1176

SCIENCE

Earthquake Hazard After a Mainshock in California

PAUL A. REASENBERG AND LUCILE M. JONES

Earthquake Hazard After a Mainshock in California

Paul A. Reasenberg and Lucile M. Jones

After a strong earthquake, the possibility of the occurrence of either significant aftershocks or an even stronger mainshock is a continuing hazard that threatens the resumption of critical services and reoccupation of essential but partially damaged structures. A stochastic parametric model allows determination of probabilities for aftershocks and larger mainshocks during intervals following the mainshock. The probabilities depend strongly on the model parameters, which are estimated with Bayesian statistics from both the ongoing aftershock sequence and from a suite of historic California aftershock sequences. Probabilities for damaging aftershocks and greater mainshocks are typically well-constrained after the first day of the sequence, with accuracy increasing with time.

In the immediate aftermath of a large earthquake in a populated region, numerous decisions will have to be made concerning the suspension and resumption of critical services, including the operation of utilities, industrial processes, transportation facilities, and schools. The need to resume these activities and to reoccupy structures that may have been weakened or partially damaged in the mainshock must be tempered by the expectation that one or more additional damaging earthquakes, including either a second, larger mainshock or one or more strong aftershocks, may occur (*1, 2*). Although most of the structural damage associated with an earthquake sequence occurs during the mainshock shaking, significant additional damage and loss of life has been sustained during strong aftershocks, particularly in structures weakened by the mainshock. Reliably assessing the extent of structural damage sustained in the mainshock for a particular structure may take several weeks or more. However, the need to reoccupy that structure may be urgent. To approach rationally the questions of when to resume certain activities and which structures to reoccupy, we must be able to assess the probabilities for the occurrence of both a larger mainshock and strong aftershocks.

The probability that a larger earthquake will follow an earthquake of a given magnitude has been estimated empirically for the southern California region from the occurrence rate of foreshocks (*3*). State and federal hazard evaluation and emergency response officials have included this assessment of the enhanced probability of a larger earthquake in responding to recent moderate events in California (*4*). We have developed a parametric model in which we describe stochastically an earthquake sequence and derive a probability for the occurrence of either a larger mainshock or a strong aftershock. Our model is based on data from California earthquakes, but can be applied elsewhere.

The distributions of aftershocks in space, time, and magnitude follow well-known stochastic laws (*2, 5–9*). Indeed, aftershocks can be identified only in a statistical fashion; they bear no known characteristics differentiating themselves from other earthquakes. In general, the rate of occurrence of earthquakes increases abruptly after a mainshock, and then decreases with time after the mainshock according to a power-law decay, while the earthquake magnitudes have an exponential distribution that is stationary in time (Fig. 1). We use these relations to model earthquake sequences and to estimate probabilities for the occurrence of strong aftershocks or larger mainshocks in any given time interval. We consider the combined probability that one or more additional earthquakes (strong aftershock or larger mainshock) will occur in a given magnitude range and time interval. We do not distinguish between the case of one such event occurring and that of more than one occurring; we assume that virtually all questions of public policy would have the same outcome in either case.

We model the aftershock process as a nonhomogeneous Poisson process in time with intensity, $N(t)$, obeying the modified Omori law (*7*)

$$N(t) = \frac{K}{(t+c)^p} \quad (1)$$

where t is time after the mainshock, and K, c, and p are constants. We model the magnitude distribution following the Gutenberg-Richter relation

$$N(M) = A \cdot 10^{-bM} \quad (2)$$

where M is the aftershock magnitude, and A and b are constants. Then the rate, λ, of aftershocks with magnitude M or larger, at the time t following a mainshock of magnitude M_m, may be expressed as

$$\lambda(t,M) = 10^{a + b(M_m - M)}(t+c)^{-p} \quad (3)$$

where a, b, p, and c are constants. The probability, P, of one or more earthquakes occurring in the magnitude range ($M_1 \le M < M_2$) and time range ($S \le t < T$) is (*10*)

$$P = 1 - \exp\left[-\int_{M_1}^{M_2}\int_S^T \lambda(t,M)\,dt\,dM\right] \quad (4)$$

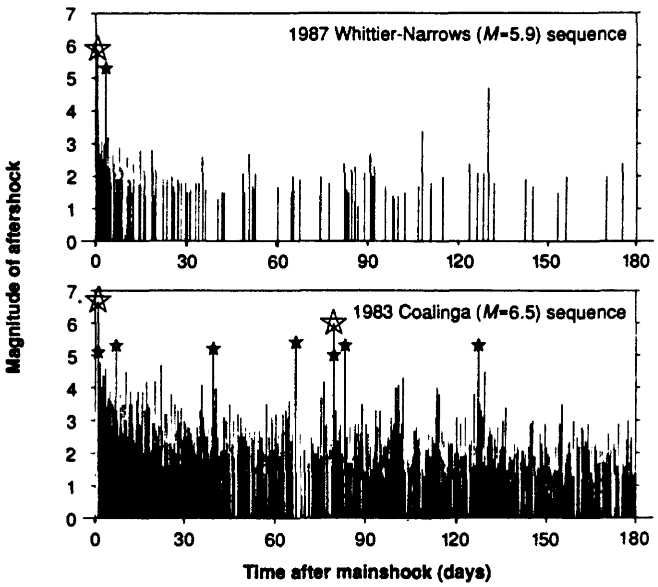


Fig. 1. Aftershock activity following two recent California earthquakes. (**A**) 1 October 1987 ($M = 5.9$) Whittier-Narrows earthquake. (**B**) 2 May 1983 ($M = 6.5$) Coalinga earthquake. Small stars indicate $M \ge 5.0$ events; large stars, $M \ge 5.5$ events.

P. A. Reasenberg, U.S. Geological Survey, 345 Middlefield Road, Mail stop 977, Menlo Park, CA 94025.
L. M. Jones, U.S. Geological Survey, 525 South Wilson Avenue, Pasadena, CA 91106.

We estimate the interval probabilities $P(M_1, M_2, S, T)$ by evaluating Eq. 4 over selected time and magnitude intervals, using point estimates of the constant model parameters. Probabilities for aftershocks are obtained when $M_2 = M_m$. Probabilities for a larger mainshock are obtained when $M_1 = M_m$ and $M_2 = \infty$ (Tables 1 and 2).

We have estimated the parameters in Eq. 3 using earthquake data from California (*11–14*). We identified 62 aftershock sequences ($M_m \geq 5$) occurring from 1933 to 1987 using a cluster recognition algorithm (*10, 15*). Model parameters were estimated separately for each sequence with the method of maximum likelihood. We used all aftershocks with $M \geq M_m - 3$ to determine the fit to Omori's Law (parameters a and p); we used all aftershocks with $M \geq 2$ to determine parameter b (*16*). Mean parameter values determined for these 62 sequences are $\bar{b} = 0.90 \pm 0.02$, $\bar{p} = 1.07 \pm 0.03$, and $\bar{a} = -1.76 \pm 0.07$ (*17*) (Fig. 2). These values are similar to those obtained from comparable aftershock sequences worldwide. Ranges and median value of b are 0.51 to 1.33, median 0.83 for 13 sequences in Japan; 0.46 to 1.00, median 0.82 for 10 sequences in Southern California; and 0.56 to 1.36, median 0.82 for 10 sequences in Greece (7). The range of most commonly reported values of p worldwide is ~1.0 to ~1.4. Earthquake sequences in eastern California had significantly higher values of a than their counterparts in both the compressional regime of southern California and the strike-slip regime of central California, which implies that there is a higher probability for aftershocks in eastern California sequences (*18*). We refer to the distributions of parameter values determined for the 62 historic California sequences as the a priori distributions. The set of model parameters consisting of the medians of the a priori distributions ($a = -1.67$, $b = 0.91$, $p = 1.08$, $c = 0.05$) is termed the "generic California" model (Fig. 2; Table 1).

Estimated interval probabilities for the generic sequence indicate that most large aftershocks (those with magnitude one unit below the mainshock or greater) occur within a few weeks of the mainshock, and are approximately seven times as likely as a greater mainshock in any given interval (Table 1). For example, the estimated probability that at least one $M \geq 5.5$ earthquake will follow a $M = 6.5$ mainshock in a generic sequence during the 1-week interval beginning 0.01 day after the mainshock is 0.34. After 15 days, the 1-week probability drops to 0.03. The estimated probability for the occurrence of a larger mainshock in the 30-day interval beginning 0.25 days after the mainshock is 0.04 (*19*).

Primary support for the validity of the generic model for earthquakes with magnitude larger than the mainshock is obtained independently from the empirical frequency of foreshocks. During the first 7-day interval following $M \geq 5.0$ earthquakes in southern California, the probability (determined from the foreshock occurrence rate) that another earthquake of equal or greater magnitude will occur is 0.056 (*20*). The corresponding probability estimated with the generic California model is 0.049 (Table 1). The agreement between these estimates for the immediate probability of a larger mainshock provides some confidence that our model is approximately valid in this extended magnitude range. Thus, the generic model provides a useful starting point for estimating post-mainshock hazard in the absence of any information about a particular sequence other than the mainshock magnitude. However, departures from this generic behavior are expected in any given aftershock sequence.

Two recent earthquake sequences serve to illustrate such departures: the 1983 ($M = 6.5$) Coalinga earthquake and the 1987 ($M = 5.9$) Whittier-Narrows earthquake (*21–23*). The magnitude distributions for these sequences differed slightly ($b = 0.73$ for Whittier-Narrows, $b = 0.89$ for Coalinga). The Coalinga sequence was more productive in aftershocks ($a = -1.47$) than the Whittier-Narrows sequence ($a = -1.60$), and the decay in its rate of aftershocks was slower ($p = 1.06$ for Coalinga; $p = 1.50$ for Whittier-Narrows). These contrasts in model parameters account for substantial differences in the resulting probability estimates, both between these sequences and relative to the generic sequence, and illustrate the variation of hazard among California earthquake sequences (Table 2) (*24*). For example, the calculated probability for the occurrence of one or more $M \geq 4.9$ events at Whittier-Narrows during the 1-week beginning 1 day after the mainshock was 0.10 (Table 2); one aftershock in this magnitude range occurred 2.8 days after the Whittier-Narrows mainshock (Fig. 1A). At Coalinga, the estimated probability for one or more $M \geq 5.5$ events during the 90-days beginning 1 day after the mainshock was 0.39; one strong aftershock ($M = 5.8$) occurred at Coalinga 80 days after the mainshock (Fig. 1B).

A much more practical use of the model is the calculation of interval probabilities for aftershocks or larger mainshocks in real time during an ongoing aftershock sequence. The model parameters for an ongoing earthquake sequence can be estimated with Bayes rule (*25, 26*). We assume that the a priori estimates of each parameter, θ, are normally distributed with some mean value θ_0 and variance σ_0^2, and that the a posteriori estimate of the parameter, determined from a sample of size n, is normally distributed with some mean $\hat{\theta}$ and variance σ^2. Then the Bayesian estimate of θ, for a mean squared error loss function, is given by

$$\hat{\theta}_B = \left(\frac{\sigma_0^2}{\sigma_0^2 + \sigma^2/n}\right)\hat{\theta} + \left(\frac{\sigma^2/n}{\sigma_0^2 + \sigma^2/n}\right)\theta_0 \quad (5)$$

Thus, Bayesian estimates, $\hat{\theta}_B$, of the model parameters can be obtained throughout the sequence, with accuracy increasing with time after the mainshock. Immediately after the mainshock, the calculation of $\hat{\theta}_B$ heavily weights the a priori mean parameter value; during the course of the aftershock sequence, the a posteriori parameter estimates are increasingly weighted as the current data become more numerous and σ^2/n becomes small compared to σ_0^2. Monte Carlo simula-

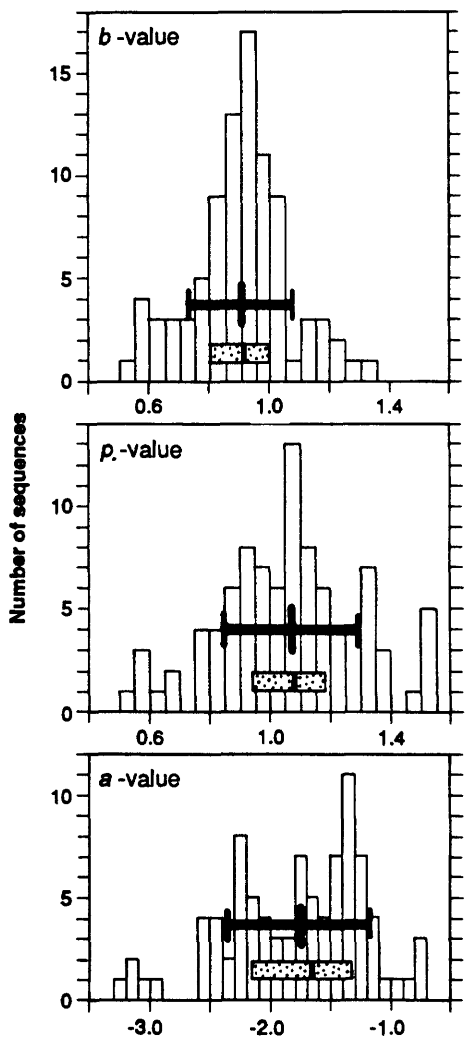


Fig. 2. Distributions of parameters (b, p, and a) determined for aftershock sequences following 62 ($M \geq 5.0$) mainshocks in California from 1933 to 1987. Solid bar indicates mean ± 1 sd. Shaded bar indicates median (central line) and upper and lower quartiles (end points) of distribution.

tions indicate that, for the generic California sequence, the a posteriori parameter estimates receive more than half the total weight within approximately 24 hours. Thus, immediately useful and increasingly accurate estimates of probabilities for aftershocks or larger mainshocks can be obtained during an ongoing earthquake sequence.

Our statistical model is completely general, and can be easily extended to other geographic or tectonic regions; only the a priori parameter values are particular to California. The ability to estimate parameters for an ongoing sequence, however, obviously depends on the availability of network processing with the capability to locate epicenters and to estimate magnitudes accurately in real time.

In the present model, the estimated values of the parameters are essentially determined from the smaller magnitude earthquakes. Justification for extending the model to larger magnitudes is provided by the close agreement between the estimated probability for larger mainshocks that we determined and the observed foreshock frequency in southern California. Furthermore, the model should be applicable at larger magnitudes for a self-similar process, and California seismicity is apparently self-similar over a wide range of magnitudes (*27*). Although there is some evidence that the Gutenberg-Richter magnitude relation may systematically underestimate the number of larger magnitude earthquakes worldwide (*7*), it adequately accounts for the California data.

We have adopted a simple inverse power-law time decay to describe aftershock rate. More sophisticated models with more parameters—such as trigger and epidemic models, models allowing for secondary or multiple aftershock sequences, and those based on a combination of power-law and exponential time decays—may be appropriate for modeling some complete sequences that include numerous observations (*28, 29*). However, we preferred to develop a simple model to ensure that the estimation of parameters is stable during the early hours of an ongoing aftershock sequence when precious few data are available from which to infer a larger number of parameters.

The simplification of the spatial distribution of aftershocks described above precludes any inference of the detailed spatial distribution of aftershocks or larger mainshock (*30*). However, from the standpoint of early hazard evaluation, detailed spatial resolution of the expected earthquake activity may be effectively limited by a lack of knowledge about the mainshock faulting process. As such data become available in the days following the mainshock, appropriate corrections to the isotropic results could be applied.

Table 1. Interval probabilities, $P(M_1, M_2, S, T)$ for the generic California aftershock sequence for strong aftershocks or larger mainshocks ($M_1 = M_m - 1$, $M_2 = \infty$), and for larger mainshocks only ($M_1 = M_m$, $M_2 = \infty$). Time intervals are described by S (interval start time, in days after the mainshock) and $(T - S)$ (duration, in days). Model parameters for the generic sequence are ($b = 0.91$, $p = 1.08$, $a = -1.67$, $c = 0.05$).

$(T-S)$	S: 0.01	0.25	0.50	1	3	7	15	30	60
	Earthquakes with $M \geq M_m - 1$								
1	0.234	0.119	0.083	0.052	0.021	0.009	0.004	0.002	0.001
3	0.296	0.181	0.140	0.100	0.049	0.024	0.012	0.006	0.003
7	0.338	0.227	0.186	0.144	0.083	0.046	0.025	0.013	0.007
30	0.399	0.297	0.258	0.217	0.152	0.104	0.068	0.042	0.024
60	0.424	0.326	0.289	0.249	0.185	0.136	0.096	0.064	0.039
90	0.437	0.342	0.305	0.267	0.203	0.154	0.113	0.079	0.051
365	0.479	0.390	0.357	0.320	0.261	0.214	0.173	0.137	0.103
1000	0.504	0.420	0.388	0.353	0.297	0.252	0.212	0.177	0.142
	Earthquakes with $M \geq M_m$								
1	0.032	0.015	0.011	0.007	0.003	0.001	0.001	0.000	0.000
3	0.042	0.024	0.018	0.013	0.006	0.003	0.001	0.001	0.000
7	0.049	0.031	0.025	0.019	0.011	0.006	0.003	0.002	0.001
30	0.061	0.042	0.036	0.030	0.020	0.013	0.009	0.005	0.003
60	0.066	0.047	0.041	0.035	0.025	0.018	0.012	0.008	0.005
90	0.068	0.050	0.044	0.037	0.028	0.020	0.015	0.010	0.006
365	0.077	0.059	0.053	0.046	0.036	0.029	0.023	0.018	0.013
1000	0.083	0.065	0.059	0.052	0.042	0.035	0.029	0.024	0.019

Table 2. Interval probabilities, $P(M_1,M_2,S,T)$, for strong aftershocks or a larger mainshock ($M_1 = M_m - 1$, $M_2 = \infty$), following the 1987 ($M = 5.9$) Whittier-Narrows, CA, earthquake and the 1983 ($M = 6.5$) Coalinga, CA, earthquake. Time intervals are described by S (interval start time, in days after the mainshock) and $(T - S)$ (duration, in days). Model parameters for the Whittier-Narrows earthquake data were $a = -1.60$, $b = 0.73$, $p = 1.50$, and $c = 0.05$ and for the Coalinga earthquake data were $a = -1.47$, $b = 0.89$, $p = 1.06$, and $c = 0.05$.

$(T-S)$	S: 0.01	0.25	0.50	1	3	7	15	30	60
	Whittier-Marrows ($M = 5.9$) *Sequence; Earthquakes with* $M \geq 4.9$								
1	0.393	0.141	0.084	0.044	0.012	0.004	0.001	0.000	0.000
3	0.431	0.185	0.123	0.074	0.026	0.010	0.004	0.001	0.000
7	0.448	0.208	0.146	0.095	0.040	0.017	0.007	0.003	0.001
30	0.465	0.232	0.171	0.120	0.062	0.034	0.017	0.009	0.004
60	0.470	0.238	0.178	0.127	0.069	0.040	0.023	0.012	0.006
90	0.472	0.241	0.181	0.130	0.073	0.043	0.025	0.015	0.008
365	0.476	0.248	0.188	0.138	0.080	0.051	0.033	0.021	0.013
1000	0.478	0.250	0.191	0.141	0.083	0.054	0.036	0.024	0.016
	Coalinga ($M = 6.5$) *Sequence; Earthquakes with* $M \geq 5.5$								
1	0.330	0.176	0.125	0.081	0.033	0.015	0.007	0.003	0.002
3	0.413	0.265	0.209	0.153	0.077	0.039	0.020	0.010	0.005
7	0.467	0.330	0.276	0.218	0.129	0.074	0.040	0.022	0.011
30	0.545	0.427	0.378	0.324	0.234	0.165	0.109	0.069	0.039
60	0.577	0.466	0.420	0.370	0.283	0.214	0.154	0.105	0.066
90	0.593	0.487	0.443	0.394	0.310	0.242	0.181	0.130	0.086
365	0.643	0.550	0.511	0.468	0.393	0.332	0.274	0.221	0.169
1000	0.673	0.588	0.552	0.513	0.444	0.387	0.334	0.283	0.233

REFERENCES AND NOTES

1. Utsu (*2*) defines "aftershock" as follows: "It is often observed that a number of earthquakes occur in a group within a limited interval of time and space. The largest earthquake in such a series is called the mainshock, and smaller ones occurring before and after the mainshock are called foreshocks and aftershocks respectively." Such a retrospective definition requires observation of the entire sequence (so that the largest earthquake in the series can be determined). In this study, we assume that a large earthquake has recently occurred, and refer to it as the "mainshock." We refer to smaller earthquakes that may follow it as "aftershocks," and any larger earthquake that may follow as a "larger mainshock."
2. T. Utsu, *J. Fac. Sci. Hokkaido Univ. Ser.* 7 **3**, No. 3 (1969).
3. L. M. Jones, *Bull. Seismol. Soc. Am.* **75**, 1669 (1985). Foreshocks in this study were limited to earthquakes within 10 km of the mainshock epicenter.
4. J. Goltz, *The Parkfield and San Diego Earthquake Predictions: A Chronology*, *Report of the Southern California Earthquake Preparedness Project* (Governor's Office of Emergency Services, Sacramento, CA, 1985).
5. T. Utsu, *Geophysics* **30**, 521 (1961).
6. ———, *J. Fac. Sci. Hokkaido Univ. Ser.* 7 **3**, No. 4 (1970).
7. ———, *ibid.* **3**, No. 5 (1971).
8. ———, *ibid.* **4**, No. 1 (1972).
9. ———, *J. Phys. Earth* **22**, 71 (1974).

10. P. A. Reasenberg, *J. Geophys. Res.* **90**, 5479 (1985).
11. Data were obtained from catalogs prepared separately for southern and eastern California (*12*), northern California before 1971 (*13*), and northern California after 1971 (*14*). A variety of magnitude scales have been used in California. For mainshocks ($M \geq 5.0$), M_W (moment magnitude) was used when available; otherwise M_L (local magnitude) was used. For southern and eastern California aftershocks, we used M_L for ($M \geq 3.0$), and M_{CA} (coda amplitude magnitude) for ($M < 3.0$). For northern California aftershocks, we used M_L before 1970 and M_D (coda duration amplitude) after 1970.
12. D. D. Given, L. K. Hutton, L. M. Jones, *U.S. Geol. Surv. Open-File Rep. 87-488* (1987).
13. R. B. Darragh *et al.*, *Bull. Seismographic Stations Univ. Calif., Berkeley* **55**, 1–2 (1985).
14. S. L. Kirkman-Reynolds and F. W. Lester, *U.S. Geol. Surv. Open-File Rep. 86-157* (1987).
15. This algorithm identifies clusters of earthquakes in time and space. By defining the identified clusters as the set of aftershocks, we effectively eliminate the spatial part of the problem. Thus, from this point on, the analysis considers aftershocks as two-dimensional (time-magnitude) vectors.
16. We used $c = 0.05$ days, the value that minimizes χ^2 for the post-1970 data.
17. Errors are ±1 standard deviation of the mean.
18. Subsets of sequences occurring in each tectonic regime in California were compared with the two-sample *t* test for difference in the mean of each parameter; $\bar{a}_{east} > \bar{a}_{north}$, $p < 0.02$ and $\bar{a}_{east} > \bar{a}_{south}$, $p < 0.005$.
19. Sensitivity of the calculated probabilities to variations in the model parameters was investigated. A 10% increase in *a*, *b*, *c*, or *p*, relative to the generic value, leads to probabilities for strong aftershocks at $S = 1$, $(T - S) = 365$ of 0.44, 0.35, 0.32, and 0.25, respectively, compared to the generic probability of 0.32 in Table 1. Corresponding probabilities for larger mainshocks are 0.068, 0.042, 0.046 and 0.035, compared to the generic probability 0.046 in Table 1.
20. L. M. Jones and P. A. Reasenberg, *Eos* **69**, 1305 (1988).
21. J. Bennett and R. Sherburne, Eds., *Calif. Div. Mines Geol. Spec. Pub 66* (1983).
22. E. Hauksson *et al.*, *Science* **239**, 1409 (1988).
23. W. Ellsworth and M. Rymer, Eds., *U.S. Geol. Surv. Prof. Pap. 1487* (1988).
24. We obtained the probability estimates given in Table 2 using parameters estimated for the entire earthquake sequences; thus, they are generally better-determined than those that would have been obtained in real-time during the earthquake sequences. We present these best model estimates of aftershock probability rather than the real-time or "limited knowledge" estimates in order to best demonstrate the time- and magnitude-dependence of the calculated probabilities and to provide a uniform comparison of results for these contrasting earthquake sequences.
25. R. T. Hogg and A. T. Craig, *Introduction to Mathematical Statistics* (Macmillan, New York, 1978).
26. A. H. Bowker and G. J. Lieberman, *Engineering Statistics* (Prentice Hall, Englewood Cliffs, New Jersey, ed. 2, 1972).
27. Y. Y. Kagan and L. Knopoff, *J. Geophys. Res.* **86**, 2853 (1981).
28. Y. Ogata, *J. Phys. Earth* **31**, 115 (1983).
29. ———, *J. Am. Stat. Assoc.* **83**, 9 (1988).
30. The empirical (isotropic) spatial distribution of strong aftershocks in California was determined from the California data. The median distance between $M \geq 5.0$ aftershocks and their mainshock epicenter is 5 km. The 80 percentile distance is 9 km.

31. We thank M. V. Matthews for providing technical assistance throughout the study, B. Ellsworth and Y. Ogata for helpful discussions and suggestions, and R. D. Brown, for initially stimulating our interest in this problem.

20 September 1988; accepted 23 December 1988

broid systematics because it provides a phylogenetic signal over an issue where morphology is equivocal due to homoplasy. In a maximum parsimony analysis of all informative nucleotide sites (*1*), billfishes composed one clade, and all other scombroids composed a separate clade. *Gasterochisma* was nested within the nonbillfish clade. In placing billfishes so distant from the scombrids, and thus the butterfly mackerel, our study provides strong evidence for two conclusions: cranial endothermy evolved two times, and it evolved independently in very distantly related lineages.

In 300 replications of the bootstrap procedure with the use of a heuristic search on all informative nucleotide sites (5), a grouping of all cranial endotherms (billfishes + *Gasterochisma*) did not occur. This finding (3) represents direct evidence against the monophyly of cranial endotherms (6). In a parsimony analysis of all informative amino acid sites (*7*), the strict consensus of 96 equally most parsimonious trees indicates separation of the billfish clade from *Gasterochisma*. Furthermore, a 10% increase in tree length is required to produce a topology that indicates monophyly of cranial endotherms (tree length increased from 111 to 122 amino acid substitutions). This difference in length represents highly significant statistical evidence against the monophyly of cranial endotherms according to the topology-dependent cladistic permutation test for nonmonophyly (8). Our phylogeny and that of Collette *et al.* (3) support the same conclusion about how many times these evolved but differ significantly from the morphological phylogeny of Johnson (2).

Beyond counting how many times endothermy has evolved we seek to understand the selective pressures that have favored the evolution of endothermy and the preadaptations that may have permitted its evolution in the Scombroidei. Thus, we must identify the ectothermic sister groups of the endothermic lineages.

The morphological hypotheses (*1*, *2*) consider billfishes to be derived scombroids that share a most recent common ancestry with members of the family Scombridae. *Gasterochisma* resides within the Scombridae in one of these studies (2). The molecular data (3) indicate that billfishes lie outside of a clade composed of all other scombroids, suggesting that cranial endothermy evolved independently in two very distant lineages. The morphological data of Collette *et al.* (2) suggest that cranial endothermy evolved twice within a group of closely related fishes: the clade composed of billfishes plus Scombridae.

We have recently completed a second molecular analysis on scombroid relationships based on the nuclear gene *lactate dehydrogenase b* (9). The *LDH b* nucleotide trees are similar to the *cytochrome b* trees and refute the monophyly of cranial endotherms with robust statistical support.

Johnson and Baldwin state that the addition of taxa could weaken the conclusions of our molecular phylogenetic analysis. This criticism could theoretically be leveled at any phylogenetic hypothesis. However, the addition (*10*) of taxa to the molecular phylogenetic analysis, including the wahoo, *Acanthocybium solandri*, a species which Johnson proposes is the sister group to billfishes, reinforces our conclusion (3) that the billfishes are distantly related to other scombroid fishes (Fig. 1). Furthermore, our analysis of this enlarged *cytochrome b* data set rejects the hypothesis by Johnson (*1*) that *Acanthocybium* is the sister-group of billfishes and is consistent with the placement of the wahoo made by Collette *et al.* (2). This conclusion is also strongly supported by the *LDH b* analysis.

Molecular data provide an important source of phylogenetic information for the Scombroidei, primarily because it complements existing morphological data and is informative in instances where morphological hypotheses conflict. We believe that historical patterns are best elucidated when a combination of different types of data, morphological and molecular, is used to corroborate and test phylogenetic hypotheses. We hope our study encourages such a synthesis.

John R. Finnerty
Barbara A. Block*
Department of Organismal Biology and Anatomy, University of Chicago, Chicago, IL 60637, USA

*Present address: Stanford University, Biological Sciences, Hopkins Marine Station, Pacific Grove, CA 93950, USA.

REFERENCES AND NOTES

1. G. D. Johnson, *Bull. Mar. Sci.* **39**, 1 (1986).
2. B. B. Collette, T. Potthoff, W. J. Richards, S. Ueyanagi, J. L. Russo, Y. Nishikawa, in *Ontogeny and Systematics of Fishes*, H. G. Moser *et al.*, Eds. (Spec. Publ. no. 1, Supplement to *Copeia*, American Society of Ichthyologists and Herpetologists, Lawrence, KS, 1984), pp. 591–620.
3. B. A. Block, J. R. Finnerty, A. F. R. Stewart, J. Kidd, *Science* **260**, 210 (1993).
4. T. Potthoff, S. Kelley, S. Ueyanagi, *Bull. Mar. Sci.* **30**, 329 (1980); I. Nakamura, *Publ. Seto Mar. Biol. Lab.* **28**, 255 (1983); T. Potthoff and S. Kelley, *Fish. Bull. U.S.* **84**, 647 (1986).
5. J. Felsenstein, *Evolution* **39**, 783 (1985).
6. In the same way that a high bootstrap value for a particular node indicates support for the monophyly of a group of taxonomic units, a low value represents evidence that a group of taxonomic units is not monophyletic (D. P. Faith, personal communication).
7. The *cytochrome b* nucleotide sequences used to construct the phylogeny in our report [figure 1 of (*1*)] were translated into amino acid sequences. A heuristic search for the most parsimonious tree was performed on these data with the use of *PAUP*, version 3.0s [Phylogenetic Analysis Using Parsimony, Swofford, D. L., 1990; tree-bissection and reconnection (TBR) branch swapping were performed on ten starting trees generated through random addition of taxa]. Ninety-six equally most parsimonious trees identified in the search indicated that *Gasterochisma* is nested within a clade containing members of the scombroid families Scombridae and Gempylidae and that billfishes fall outside of this clade. The "enforce topological constraints" option of PAUP was used to identify the most parsimonious topology that supported the monophyly of cranial endotherms, *Gasterochisma*, and billfishes. This topological constraint resulted in an 11-step increase in tree length over the most parsimonious tree (122 over 111 steps).
8. D. P. Faith, *Syst. Zool.* **27**, 401 (1991). We conducted topology-dependent cladistic permutation (T-PTP) tests for non-monophyly of cranial endotherms with the use of the inferred amino acid sequences from the data of our report (*1*). Each data set was permuted 99 times such that the observed character states of each character were shuffled randomly among taxa. The permuted data sets were searched for the most parsimonious tree under two constraints: (i) include and (ii) do not include billfishes plus butterfly mackerel as a monophyletic group. The length difference between trees never equalled or exceeded the length difference from the unpermuted *cytochrome b* data (5) (T-PTP = 0.01).
9. Briefly, degenerate primers were designed to the *lactate dehydrogenase b* locus with the use of amino acid alignments of all isozymes of LDH sequenced in vertebrates, a subset of the taxa examined in the *cytochrome b* study (*1*) were used in the second molecular study. The consensus aligned sequence for the scombroid fishes that were examined, after the insertion of alignment gaps, is 628 base pairs long (Finnerty and Block, unpublished results).
10. J. R. Finnerty and B. A. Block, *Fish. Bull. U.S.*, in press.
11. A. Knight and D. P. Mindell, *Syst. Biol.* **42**, 18 (1993).

18 January 1994; accepted 29 March 1994

Earthquake Aftershocks: Update

Since 1989, the U.S. Geological Survey has provided public forecasts of expected aftershock activity following major earthquakes in California, based on a stochastic model (*1*). The model represents the rate of aftershocks of magnitude M or larger as

$$\lambda(t, M) = 10^{a + b(M_m - M)}(t + c)^{-p}$$

where t is time after the mainshock, M_m is the mainshock magnitude, and a, b, p and c are constant parameters. Forecasts based on this model typically have been posed in probabilistic terms, such as, "There is a 50 percent chance of one or more magnitude 5 or larger earthquakes in the next 7 days." While such probabilistic statements may be clearly understood by scientists and emergency response officials, they often have created confusion and miscommunication among the press and general public. In an effort to more effectively communicate the aftershock hazard after the 17 January 1994

Northridge earthquake, we also provided the expected daily numbers of aftershocks of magnitude 3 and larger. We believe that this approach was better understood; it certainly reduced our need to explain some apparent paradoxes stemming from the public's unfamiliarity with statistical modeling. For example, it was frequently asked why our forecasts of probabilities of large aftershocks did not decrease after the occurrence of a large aftershock. The answer, which involves a discussion of the assumptions made in modeling aftershocks as a stochastic renewal process, is difficult to communicate in a press conference. The new forecasts of expected numbers of aftershocks more naturally conveyed a sense of how the Northridge aftershock sequence was decaying, and was expected to decay, with time. This approach, together with an explanation of the expected constant ratio in the numbers of large and small events, helped to communicate a sense of the temporal decay in earthquake hazard associated with large aftershocks.

The Northridge earthquake sequence was slightly more productive than the generic California sequence, given its mainshock magnitude of 6.7 (*2*). This characteristic was reflected in all our models. Estimates of the parameter *a* ranged from -1.1 ± 0.2 to -1.3 ± 0.2 during the first 10 days of the sequence, settling at -1.3, approximately 1 SD above the generic value -1.67. The decay rate and magnitude distribution parameters for the Northridge sequence ($p = 1.2$; $b = 0.9$) are both close to generic values of 1.08 and 0.91, respectively. To track the models' predictive success, we compared the model-predicted daily earthquake counts to the actual daily counts. Models obtained with the use of data from the first 1, 2, 5, and 10 days after the main shock were used to calculate the expected number of $M \geq 3$ aftershocks on each of the first 12 days of the sequence (Fig. 1). The actual counts of aftershocks in this period were generally well-predicted by those models based on data from two or more days. The model for the first 24 hours of the sequence overestimated *a*, underestimated *p*, and thus significantly overestimated the number of aftershocks in the days to follow. Such a lack of model constraint in the first 24 hours was expected on the basis of our earlier Monte Carlo experiments (*1*).

In the process of compiling modeling data (Fig. 1), we corrected an error in our formulation for calculating earthquake probabilities and expected numbers of aftershocks (*4*). The error arose from our incorrectly treating $\lambda(t, M)$ as a density function, when in fact it is a density with respect to *t* and a rate with respect to M. Thus, λ should not be integrated with respect to M to obtain the interval probabilities, as was indicated in equation 4 of our original report (*1*). This error, which entered into the calculation of tables 1 and 2 in our original report and in all estimates of aftershock probabilities to date (*3*), resulted in our underestimation of probabilities by up to a factor of approximately 2. However, the error did not affect the estimation of model parameters or the generic model in Reasenberg and Jones (*1*). Corrected probabilities for the generic California model, as defined in (*1*), have been calculated (Table 1).

Table 1. Corrected version of table 1 in Reasenberg and Jones (*1*). Interval probabilities, $P(M_1, M_2, S,$ and $T)$, defined as the probability of one or more earthquakes occurring in the magnitude range ($M_1 \leq M < M_2$) and time range ($S \leq t < T$), for the generic California aftershock sequence. Top part gives probabilities for strong aftershocks or larger mainshocks ($M_1 = M_m - 1$, $M_2 = \infty$); bottom part gives probabilities for larger mainshocks only ($M_1 = M_m$, $M_2 = \infty$). Time intervals are described by *S* (interval start time, in days after the mainshock) and ($T-S$) (duration, in days). Model parameters for the generic sequence are ($b = 0.91$, $p = 1.08$, $a = -1.67$, and $c = 0.05$).

($T-S$)	S								
	0.01	0.25	0.50	1	3	7	15	30	60
				Earthquakes with $M \geq M_m - 1$					
1	0.428	0.233	0.166	0.107	0.044	0.019	0.009	0.004	0.002
3	0.520	0.341	0.271	0.199	0.101	0.051	0.025	0.012	0.006
7	0.578	0.417	0.350	0.278	0.165	0.095	0.051	0.027	0.014
30	0.656	0.522	0.465	0.402	0.292	0.206	0.137	0.085	0.049
60	0.685	0.563	0.510	0.451	0.348	0.264	0.190	0.130	0.081
90	0.700	0.584	0.534	0.478	0.378	0.296	0.223	0.150	0.105
365	0.745	0.645	0.603	0.555	0.469	0.397	0.328	0.265	0.203
1000	0.770	0.681	0.643	0.599	0.522	0.456	0.394	0.335	0.275
				Earthquakes with $M \geq M_m$					
1	0.066	0.032	0.022	0.014	0.005	0.002	0.001	0.001	0.000
3	0.086	0.050	0.038	0.027	0.013	0.006	0.003	0.002	0.001
7	0.101	0.064	0.052	0.039	0.022	0.012	0.006	0.003	0.002
30	0.123	0.087	0.074	0.061	0.042	0.028	0.018	0.011	0.006
60	0.132	0.097	0.084	0.071	0.051	0.037	0.026	0.017	0.010
90	0.138	0.102	0.090	0.077	0.057	0.042	0.030	0.021	0.014
365	0.155	0.120	0.1-7	0.095	0.075	0.060	0.048	0.037	0.028
1000	0.165	0.131	0.119	0.106	0.087	0.072	0.060	0.049	0.039

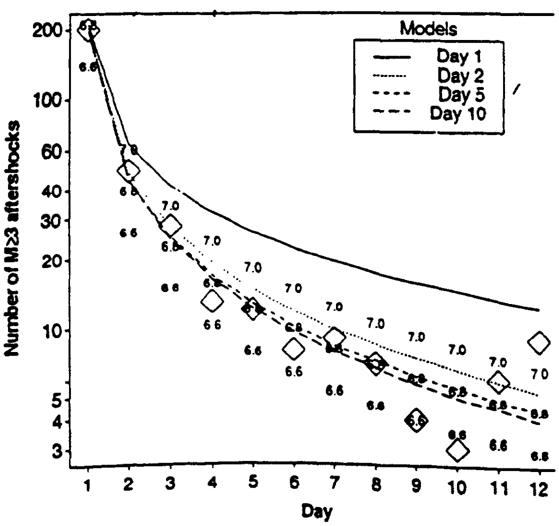


Fig. 1. Northridge aftershock modeling. Observed numbers of ($M \geq 3.0$) aftershocks during the first 12 days of the Northridge earthquake sequence (diamonds) compared with expected daily counts of aftershocks predicted with statistical models. Models based on aftershock data for the first 1, 2, 5, and 10 days after the mainshock are shown with solid and broken lines. Corresponding numbers of aftershocks expected for generic sequences following magnitude 6.6, 6.8, and 7.0 mainshocks are shown for comparison.

P. A. Reasenberg
U.S. Geological Survey, 345 Middlefield Road, Menlo Park, CA 94025, USA
L. M. Jones
U.S. Geological Survey, 525 South Wilson Boulevard, Pasadena, CA 91106, USA

REFERENCES AND NOTES

1. P. A. Reasenberg and L. M. Jones, *Science* **243**, 1173 (1989).
2. The moment magnitude, M_w, is 6.7. The earthquake produced an aftershock sequence roughly comparable, in numbers and magnitudes of aftershocks, to a generic sequence after a $M_w = 6.8$ mainshock.
3. The modeling results presented here were obtained with the corrected versions of our programs.
4. We are grateful to J. C. Pechmann for pointing out the error. We apologize to our colleagues who have used the erroneous programs.

25 April 1994; accepted 28 April 1994