Association of Earthquakes and Faults in the San Francisco Bay Area Using Bayesian Inference

by Robert L. Wesson, William H. Bakun, and David M. Perkins

Abstract   Bayesian inference provides a method to use seismic intensity data or instrumental locations, together with geologic and seismologic data, to make quantitative estimates of the probabilities that specific past earthquakes are associated with specific faults. Probability density functions are constructed for the location of each earthquake, and these are combined with prior probabilities through Bayes' theorem to estimate the probability that an earthquake is associated with a specific fault. Results using this method are presented here for large, preinstrumental, historical earthquakes and for recent earthquakes with instrumental locations in the San Francisco Bay region. The probabilities for individual earthquakes can be summed to construct a probabilistic frequency–magnitude relationship for a fault segment. Other applications of the technique include the estimation of the probability of background earthquakes, that is, earthquakes not associated with known or considered faults, and the estimation of the fraction of the total seismic moment associated with earthquakes less than the characteristic magnitude. Results for the San Francisco Bay region suggest that potentially damaging earthquakes with magnitudes less than the characteristic magnitudes should be expected. Comparisons of earthquake locations and the surface traces of active faults as determined from geologic data show significant disparities, indicating that a complete understanding of the relationship between earthquakes and faults remains elusive.

Introduction

Estimation of future earthquake hazard is commonly approached by characterizing the frequency–magnitude behavior of faults and, for time-dependent estimates, the time and other parameters of the most recent large earthquake. Often both geologic and seismologic information are available to characterize the frequency–magnitude behavior of a fault, but there is wide variation in the character of the information and the certainty of its interpretation.

Geologic information on surface fault rupture is perhaps the strongest evidence that a particular fault was responsible for a particular earthquake. However, it is well known that large earthquakes can occur without being accompanied by surface fault rupture, and cases of "sympathetic" or "triggered" slip have been reported along fault traces distant from the causative fault (cf. Sieh, 1982; McClellan and Hay, 1990). Commonly, then, the association of a particular earthquake with a particular fault is an exercise in inference. One powerful and disciplined approach, which also has the advantage of producing a numerical result, is Bayesian inference. In terms familiar to geophysicists, Bayesian inference is a method for using the statistics of the forward problem to solve the inverse problem. That is, if we can estimate the probability of obtaining a given set of observations for each

of a set of hypotheses, then using Bayes' theorem, we can estimate the probabilities, given the observed data, that each of the hypotheses is the correct one. Examples of the application of Bayesian inference in other contexts may be found in von Winterfeldt and Edwards (1986) and Watson and Buede (1987).

The components of information we have to consider for past earthquakes include

- Location, geometry, and slip rates of recognized faults,
- Observations of earthquake intensity,
- Calculated locations of earthquakes and corresponding uncertainty,
- Hypotheses based on worldwide observations and theoretical and laboratory studies that may have bearing, including ideas about characteristic earthquakes, segmentation, and frequency–magnitude relations,
- Information from the first motions at a set of seismograph stations (or from the focal mechanism solution).

The association of earthquakes with geologic faults depends on what we mean by a fault and on the purpose of the analysis. Understanding the detailed tectonics of a fault zone

1306

may require determination of earthquake focal mechanisms and the identification of a variety of surfaces within a fault zone. In contrast, for purposes of a seismic hazard analysis, it may be sufficient to associate all nearby earthquakes with a broad fault zone. In this work, we base our associations primarily upon the spatial relationship of the earthquake and fault. We do not use focal mechanism data in the following.

The principal motivation for this work is to determine the probabilities that earthquakes can be assigned to specific faults for purposes of earthquake hazards assessment.

Active Faults in the San Francisco Bay Region

The region considered in this study (Fig. 1) is a rectangular box oriented subparallel to the San Andreas fault system including the San Francisco Bay region considered by the Working Group on California Earthquake Probabilities (1999) (WG99). We focus primarily on the description of active fault segments used by WG99 (Table 1). From the viewpoint of the characteristic earthquake model, a fault segment is the basic section of fault that will rupture in a characteristic earthquake. For each fault segment, we use an estimate of the characteristic magnitude and recurrence time (or mean annual frequency of occurrence) from WG99. A few examples are based on the set of faults considered by the Working Group on Northern California Earthquake Potential (1996) (WG96).

Application of Bayesian Inference

Bayesian inference can be conveniently applied through a discretization of the earthquake association problem. Divide the region of interest into grid cells, k, and denote the event that the epicenter of the earthquake is located in the kth cell as H_k. This strategy enables us to use discrete probabilities and matrix operations for computational simplicity. Suppose that we have some observations, O, that enable us to estimate the probability that the epicenter (or hypocenter) of the earthquake is located in the kth cell, that is $P(H_k \mid O)$. Assuming that the association may be made upon location alone, then by Bayes' rule (cf. Kennedy and Neville, 1986), the probability that the earthquake is associated with the ith fault or fault segment, given that its epicenter is located in the kth grid cell, $P(F_i \mid H_k)$, can be written as

$$P(F_i|H_k) = \frac{P(H_k|F_i)P(F_i)}{\sum_i P(H_k|F_i)P(F_i)}, \tag{1}$$

where $P(H_k \mid F_i)$ is the probability that the earthquake would locate in the kth grid cell, given that it occurs on the ith fault or fault segment; $P(F_i)$ is the prior probability that the earthquake occurred on fault or fault segment i. Assuming that the epicenter (or hypocenter) of the earthquake does indeed lie within the grid, then summing over all the grid cells, we
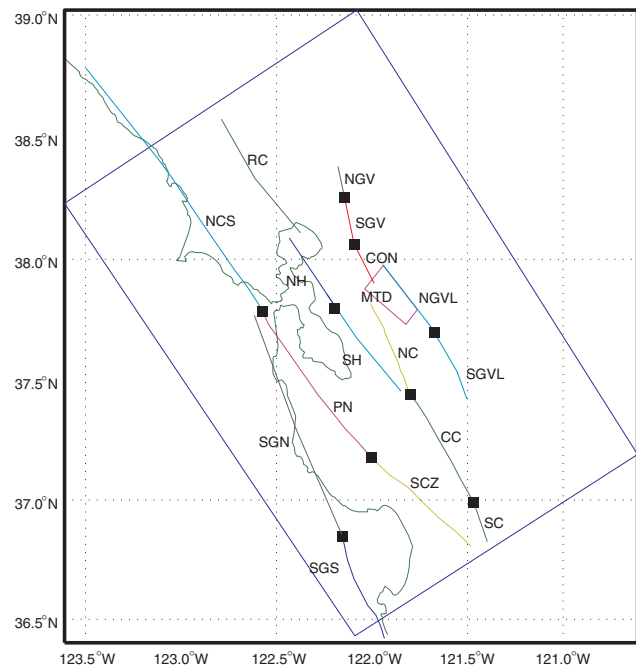

Figure 1. Fault segmentation scheme used in the association analysis based on WG99. Faults assumed to be vertical are shown as line segments. Dipping faults are represented by polygons showing the vertical projection of the fault onto the surface. The abbreviated names for fault segments are expanded in Table 1.

can obtain the probability that the earthquake is associated with the ith fault given the observations, $P(F_i \mid O)$, by the total probability theorem (cf. Kennedy and Neville, 1986) as

$$P(F_i|O) = \sum_k P(F_i|H_k)P(H_k|O). \tag{2}$$

An example of the overall inference scheme as applied to intensity observations is shown schematically in Figure 2. In the following we consider several approaches to the determination of the components of equations (1) and (2).

Throughout, we maintain the possibility that the earthquake is not associated with any of the identified fault segments, that is, in the "background."

$P(H_k \mid O)$, Probability that the Epicenter (or Hypocenter) is Located in Grid Cell k Given the Observations

Case I: Instrumental Location Available. Suppose that the location x,y (or x,y,z) of the earthquake epicenter (or hypocenter) is given together with an error ellipsoid centered at the mean location, with the orientations of the axes corresponding to the principal axes of the solution and the lengths of the semiaxes to the standard errors (cf. Klein, 1978; Lahr, 1984). From these parameters we can determine the probability density function for the location. The probability that

Table 1
Segment Parameters and Equal and Characteristic (or Activity-Based) Prior Probabilities for WG99 Fault Segments

WG99 Fault or Segment	Abbeviation	Mean Frequency (no./yr)	Prior Probabilities	
			Equal	Characteristic
San Andreas, Santa Cruz Mtns.	SCZ	4.577E-03	0.0471	0.0428
San Andreas, Peninsula	PN	4.490E-03	0.0471	0.0420
San Andreas, North Coast South	NCS	4.644E-03	0.0471	0.0435
San Andreas, North Coast North	NCN*	4.619E-03	0.0000	0.0000
Hayward/RC, Southern Hayward	SH	5.256E-03	0.0471	0.0492
Hayward/RC, Northern Hayward	NH	5.629E-03	0.0471	0.0527
Hayward/RC, Rogers Creek	RC	4.236E-03	0.0471	0.0396
Calaveras, Southern	SC	1.399E-02	0.0471	0.1309
Calaveras, Central	CC	1.778E-02	0.0471	0.1664
Calaveras, Northern	NC	4.855E-03	0.0471	0.0454
Concord/GV, Concord	CON	3.810E-03	0.0471	0.0357
Concord/GV, Southern Green Valley	SGV	3.866E-03	0.0471	0.0362
Concord/GV, Northern Green Valley	NGV	4.165E-03	0.0471	0.0390
San Gregorio, Southern	SGS	1.603E-03	0.0471	0.0150
San Gregorio, Northern	SGN	2.245E-03	0.0471	0.0210
Greenville, Southern	SGVY	1.100E-03	0.0471	0.0103
Greenville, Northern	NGVY	1.108E-03	0.0471	0.0104
Mount Diablo	MTD	2.135E-03	0.0471	0.0200
Background	BKGD		0.2000	0.2000
Total		0.08549	1.0000	1.0000

*The North Coast North segment is entirely outside the area considered.

the epicenter is located in a particular grid cell is simply the integral of the probability density function over that grid cell.

Let the location of the earthquake in the principal axes of the solution, x_i, be given as μ_i, $i = 1,2$ or $i = 1,2,3$ with standard errors, σ_i, and assume that the probability density function for the location is a multivariate normal distribution:

$$p(x_i) = \prod_{i=1}^{2\,\text{or}\,3} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right). \qquad (3)$$

The probability that the focus of the earthquake is in any two- or three-dimensional grid cell is the integral of equation (3), which may be easily computed with error functions. Relations between the specification of location errors in two and three dimensions are discussed in Appendix A.

Suppose that the principal axes of the location solution coincide with the "geographical" axes and we want to calculate the probability that the focus of the earthquake occurs in an equidimensional grid cell centered at x^k with dimension $2\Delta x$. This quantity is the integral of equation (3) over the area or volume of the cell,

$$P(H_k|O) = \prod_{i=1}^{2\,\text{or}\,3} \frac{1}{2}\left(\text{erf}\left(\frac{x_i^k + \Delta x}{\sigma_i\sqrt{2}}\right) - \text{erf}\left(\frac{x_i^k - \Delta x}{\sigma_i\sqrt{2}}\right)\right). \quad (4)$$

In practice the principal axes of the location solution will not coincide with the geographic axes, and we approximate the integral in equation (4) by translating and rotating the

grid to find the center point of the grid cell in the system of the principal axes of the solution, then using equation (4) directly, without translating and rotating all the limits of integration. In essence, the integration is performed over an equal area (or volume) and centered at the same location, but over a region with a different orientation. Because the probability density is smoothly varying, the integral can be satisfactorily approximated by the integral over an equivalent rectangle (or cube), centered at the same point, but aligned with the system of coordinates of the grid.

Case II: Only Intensity Information is Available. Instrumental determinations of earthquake epicenters exist only for the last several decades, but data on the intensities of important earthquakes may be available going back centuries. We present here a location scheme for earthquakes based on Bayesian inference that requires only intensity data and is compatible with the overall inference scheme for fault association.

We want to estimate $P(H_k \mid O)$, the probability, given the observations of intensity, that the epicenter of the earthquake occurred in the kth grid cell. We can use Bayes' rule to estimate this probability as

$$P(H_k|O) = \frac{P(O|H_k)P(H_k)}{\sum_k P(O|H_k)P(H_k)}, \qquad (5)$$

where $P(O \mid H_k)$ is the probability of the observations, given the set of hypotheses, H_k, and $P(H_k)$ is the prior probability
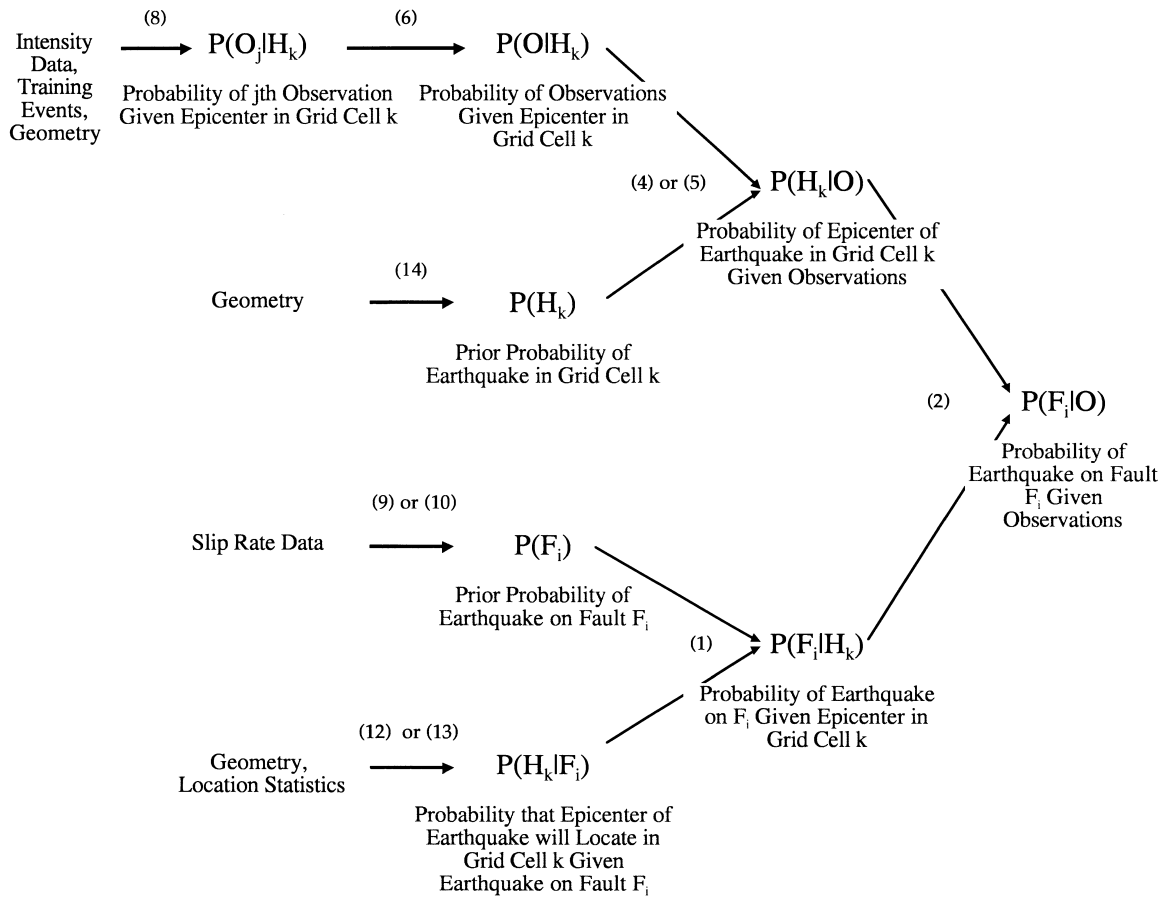
Figure 2. Schematic diagram of inference scheme as applied to observations of intensity. The numbers in parentheses indicate the equations in the text enabling the particular step. The scheme as applied to instrumental location data is similar, except that determination of the probability of the epicenter in grid cell *k* is replaced by integration of a probability density function (estimated from the errors in the location) over the grid cell, as described in the text.

that the epicenter is located in the *k*th grid cell. The observations, *O*, are actually made up of a set of observations, O_j. Assuming that we can take these observations to be independent, then

$$P(O|H_k) = \prod_j P(O_j|H_k). \qquad (6)$$

To estimate $P(O_j \mid H_k)$, [that is] the probability of the *j*th observation given the grid, we use the data set of Bakun (1998). Using the 11 earthquakes from Bakun's data set with instrumentally determined hypocenters and moment magnitudes greater than 5.5, observations of Modified Mercalli intensity (MMI) were fit to obtain a regression model, obtaining

$$\text{MMI}_j = -2.263 + 1.405 M_{\text{w}} - 0.01552\Delta_j, \qquad (7)$$

where MMI_j is the observed MMI reported by Bakun and Wentworth (1997), including their site corrections, M_{w} is the magnitude, and Δ_j is the epicentral distance to the *j*th obser-

vation. The standard error of the MMI_j is estimated to be 0.74. Our model then is that the *j*th observation is normally distributed with a mean given by equation (7) and a standard deviation of 0.74. If the actual observation of MMI corrected for site conditions is MMI_{obs}, then the probability of obtaining an observation within the one integer unit interval including this value is

$$P(O_j \mid H_k) = N(\text{MMI}_{\text{obs}}^+, 0.74) - N(\text{MMI}_{\text{obs}}^-, 0.74), \qquad (8)$$

where *N* is the cumulative normal distribution for the intensity given the M_{w} of the event as estimated by Bakun (1999) and the epicentral distance, $\text{MMI}_{\text{obs}}^- = \text{int}(\text{MMI}_{\text{obs}})$, and $\text{MMI}_{\text{obs}}^+ = \text{int}(\text{MMI}_{\text{obs}}) + 1$. Schematically this is shown in Figure 3.

As a test, this location procedure was applied to subsets of intensity data observed from the 18 October 1989 Loma Prieta earthquake. Figure 4a shows the location (or intensity center, as termed by Bakun [1999]) obtained using the 200 closest observations. Contours show the cumulative proba-
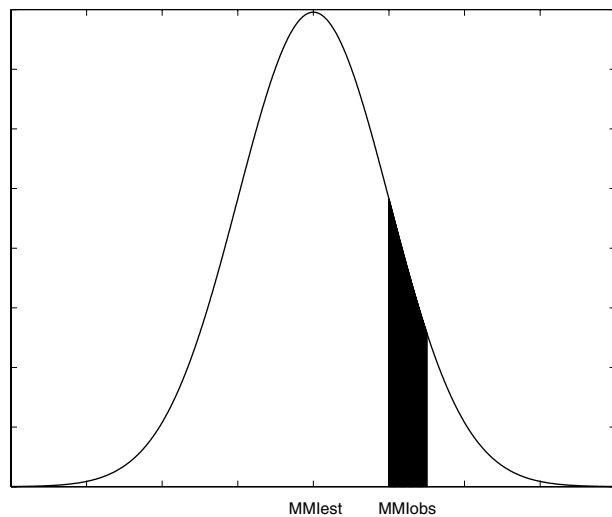
Figure 3.   The probability of the ith observation of intensity given that the earthquake occurred in the kth grid cell, $P(O_j | H_k)$, can be estimated as the area under the normal curve corresponding to the one interval unit intensity including the observation as in equation (8). The normal curve is centered on the predicted intensity given by equation (7) (i.e., the mean) using the magnitude given by Bakun (1999) and the distance to the kth grid cell. The standard deviation was determined from the regression to Bakun's data, as described in the text.

bility that the earthquake is located within the contour as determined by the previously described procedure. The cumulative probability is estimated by sorting the probabilities for each grid cell in descending order, then performing a cumulative sum and contouring the result. The center of these contours is taken as the "intensity center" corresponding to the usage of Bakun (1999). The intensity center for the Loma Prieta earthquake in Figure 4a is to the north of the instrumentally determined epicenter, but it agrees well with the reach of the fault believed to have ruptured during the earthquake as interpreted from aftershocks and waveform inversions (Dietz and Ellsworth, 1990; Wald *et al.*, 1991). Figure 4b–d shows locations as determined from samples of the data chosen to simulate the observations for specific preinstrumental earthquakes. Although a location for an earthquake using only four observations, such as for the earthquake in 1836, is very imprecise, one can develop a sense of the resolution of the method with such a small number of data. The data set chosen to simulate the observation set of the 1868 is clearly biased in that the great majority of the observations are well to the east of the Loma Prieta earthquake, and the resulting location is skewed to the west. As expected the locations improve significantly as the number of observations increase and as the data are evenly distributed with azimuth.

Locations of eight preinstrumental earthquakes are shown in Figure 5. The data used to determine these locations are the same as those used by Bakun (1999). The lo-

cations are obtained assuming the earthquake magnitudes determined by Bakun (1999) and are generally in good agreement with those he obtained using his least-squares technique. The Bayesian technique does not minimize a particular quantity as does the least-squares. In contrast, it estimates the probability, given the data and assumptions, that the epicenter of the earthquake is located in each grid cell.

$P(F_i)$, Prior Probability that the Earthquake Occurred on Fault i

The prior probability of an earthquake being associated with a particular fault is an estimate of that probability absent any information about the location of the earthquake, except that it is in the region of interest. Our prior probabilities include, either implicitly or explicitly, information about the magnitude of the earthquake. One way of thinking about the priors is to suppose that you are in a city a great distance from the San Francisco Bay Area. You hear on the radio that an earthquake of about magnitude 7 has occurred in the Bay Area, but you know nothing else about its location. What is the probability that the earthquake occurred on each of the various faults or fault segments in the region?

Suppose we recognize N faults that we believe are capable of generating an earthquake of this magnitude. We also recognize that there is some probability, $p_{background}$, that the earthquake is not associated with any of the recognized faults in the region. The most simple-minded estimate of the probability that the earthquake occurred on the ith fault, F_i, is

$$P(F_i) = (1 - p_{background})/N. \qquad (9)$$

Fault priors estimated in this way will be referred to as "equal" priors.

The degree of fault activity may be taken into account through several possible approximate approaches. The simplest approach is to use the long-term geologic slip rate and to estimate the rate of characteristic earthquakes. That is, if fault F_i is assumed to have a recurrence rate of T_i years, then the mean annual frequency of characteristic earthquakes on that fault is $\mu_i = 1/T_i$ and the mean annual frequency of characteristic earthquakes in the region is $\sum_i \mu_i$. Thus a "characteristic" estimate of the prior probability that a given earthquake occurred on fault F_i is given by

$$P(F_i) = \frac{(1 - p_{background})\mu_i}{\sum_j \mu_j}.$$

A second simplified approach is to assume that each fault segment is characterized by an incremental Gutenberg–Richter frequency–magnitude distribution; the probability that an earthquake of a given magnitude is on the ith fault is
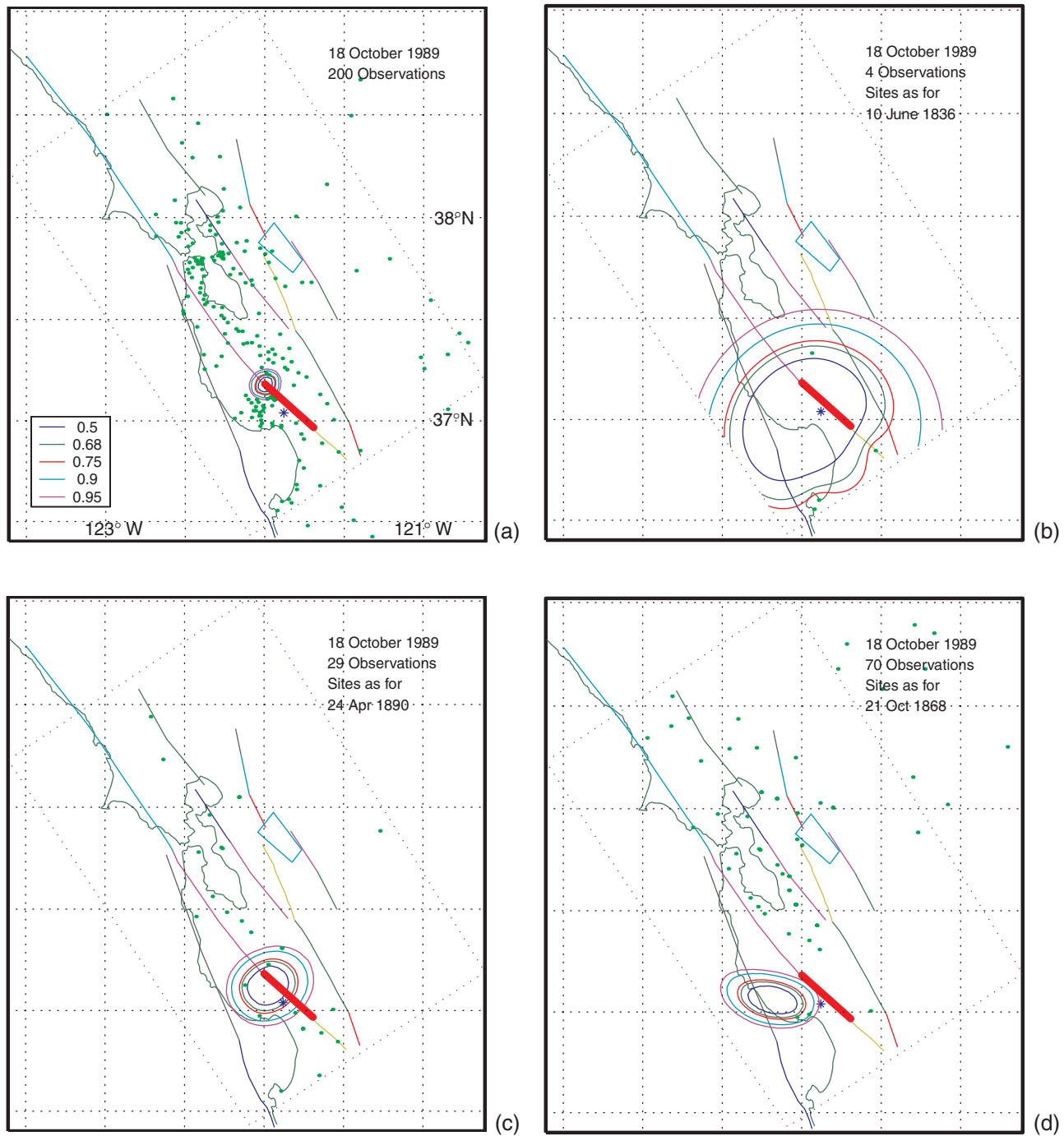
Figure 4.   Examples of locations of the 1989 Loma Prieta earthquake using intensity data with different subsets of the data: (a) the 200 closest observations, (b) 4 observations selected to simulate the locations of the observations of the 1836 earthquake, (c) 29 observations selected to simulate the locations of the observations of the 1890 earthquake, and (d) 70 observations selected to simulate the locations of the observations of the 1868 earthquake. Locations of intensity observations are shown as green dots. Contours are the values of the cumulative probability that the epicenter is contained within the given contour. Fault segments from WG99 are also shown to give an intuitive feel of how the probability can be concentrated on one or a few segments for well-located earthquakes or distributed over many segments for poorly located events.
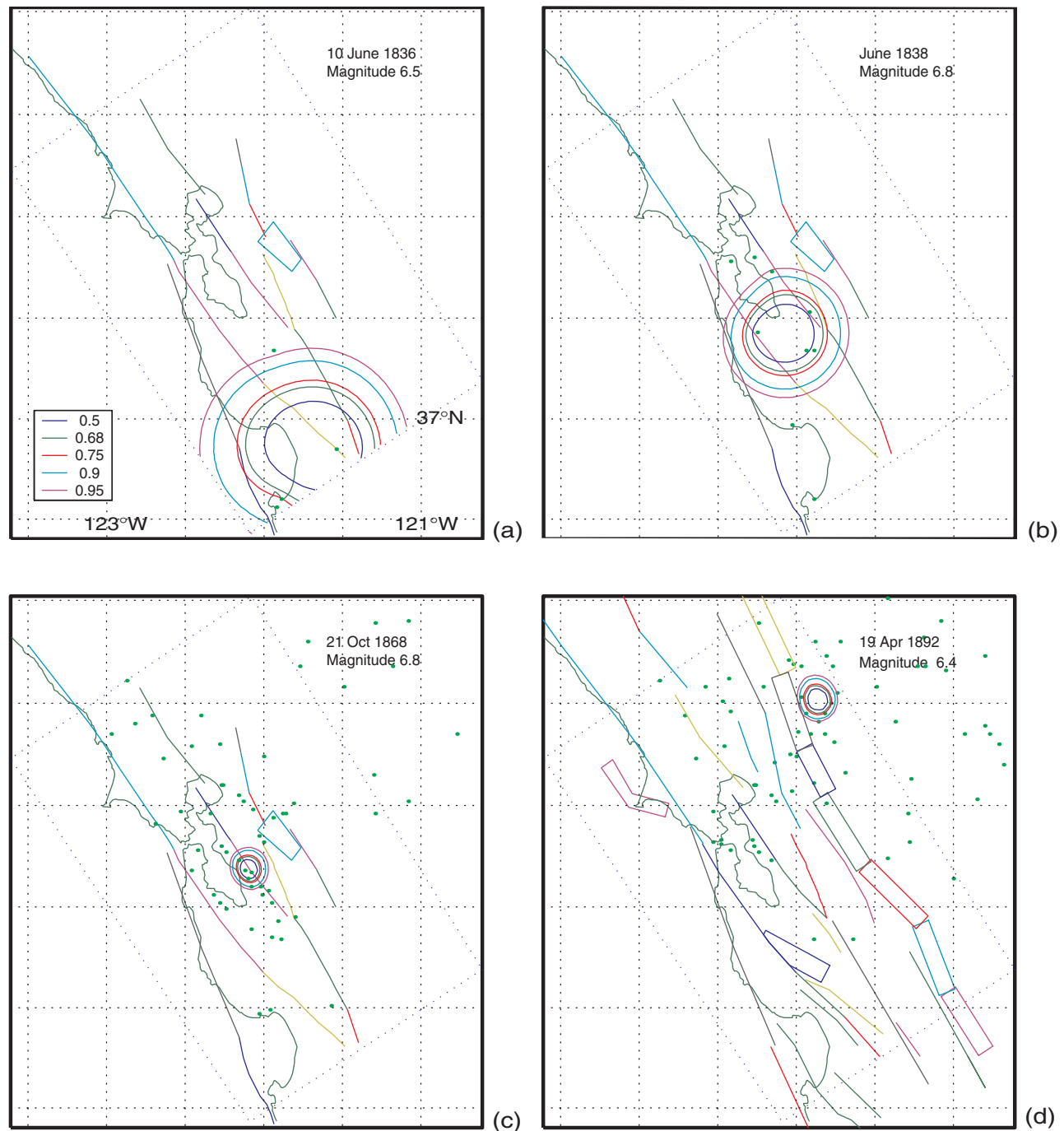
Figure 5.   Examples of locations of eight preinstrumental earthquakes determined from intensity data. Locations of intensity observations are shown as green dots. Contours are the values of the cumulative probability that the epicenter is contained within the given contour. Fault segments are also shown to give an intuitive feel of how the probability can be concentrated on one or a few segments for well-located earthquakes or distributed over many segments for poorly located events. Fault segments in (d) and (e) are from WG96, which includes the faults along the western margin of the Great Valley. Otherwise fault segments are from WG99.
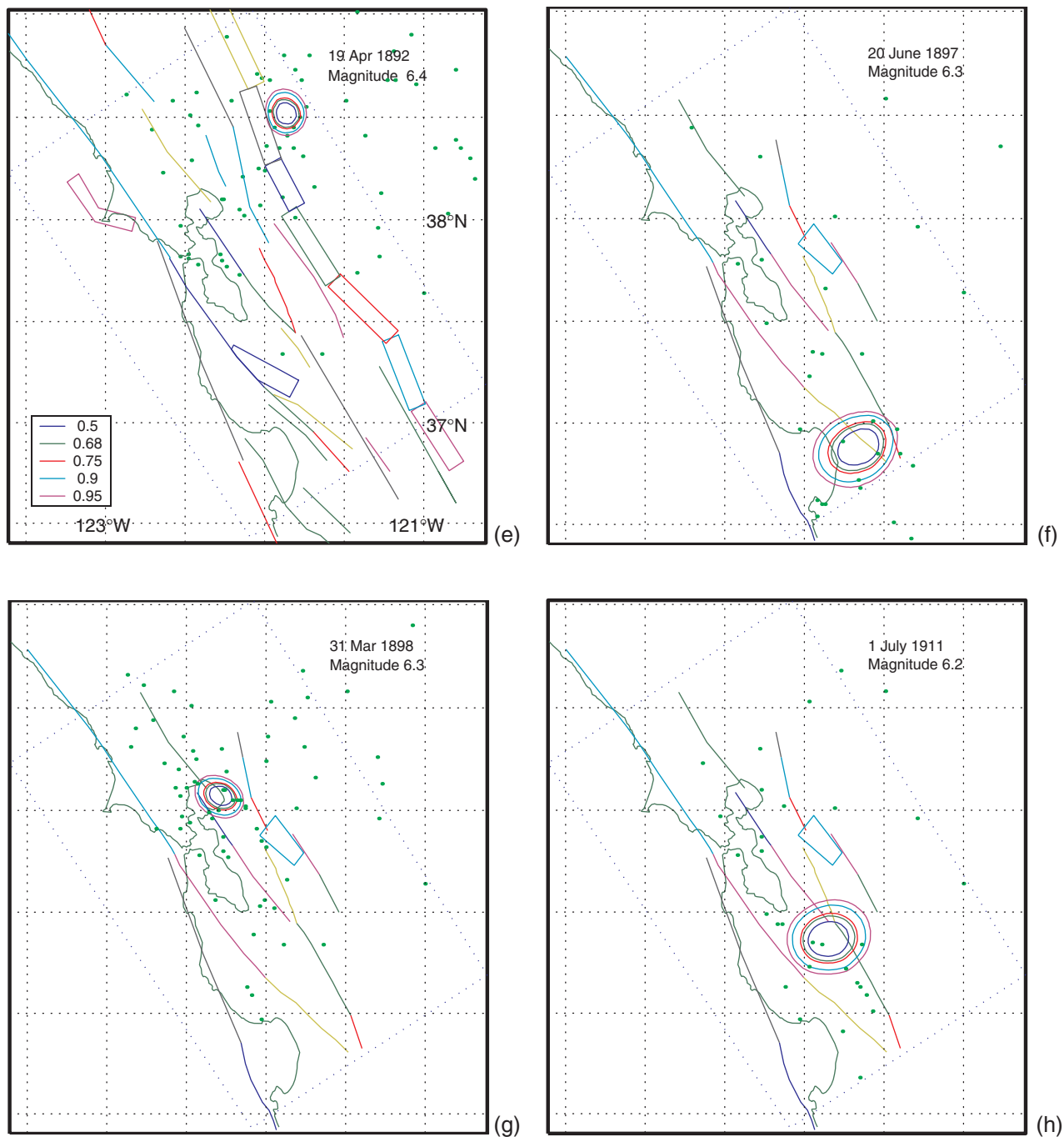
Figure 5. (*Continued*)

$$P(F_i) = \frac{(1 - p_{\text{background}})10^{a_i - b_i M}}{\sum_j 10^{a_j - b_j M}}.$$

If, for simplicity, we assume that all faults in the region share a common b value, then this expression can be simplified to give a simplified "Gutenberg–Richter" prior:

$$P(F_i) = \frac{(1 - p_{\text{background}})10^{a_j}}{\sum_j 10^{a_j}}.$$

Note that under the assumption of common b values, this prior is also independent of magnitude. Of course the a_j's are what we hope to determine as the result of the analysis.

The effect of selecting the equal prior, characteristic prior, or priors based on moment rate, slip rate, and fault area is investigated in Appendix B. For well-located earthquakes, the influence of this prior is typically overwhelmed by the spatial information. For poorly located earthquakes, its effect becomes more important.

It might be argued that shorter fault segments should be

eliminated from consideration as the possible sources of larger earthquakes. However, for earthquakes in the magnitude 6–7 range, using the Wells and Coppersmith (1994) relation for subsurface rupture length to magnitude considering two standard deviations about the mean, or 95% confidence limits, the rupture lengths are 30–117 km for a magnitude 7 and 14–58 km for a magnitude 6.5 earthquake. Thus, comparing the lower limits of the confidence levels with the segment lengths in Table 1, and also considering the possibility of multisegment rupture, a segment length limitation on the magnitude of associated earthquakes was not imposed.

$P(H_k \mid F_i)$, Probability that the Earthquake Would Locate in the kth Grid Cell Given that it Occurs on the ith Fault

The probability that the epicenter of the earthquake will appear to be located in the kth cell given that it occurs on the ith fault is related to the geometric extent of both the earthquake and the fault and the uncertainty of the location process. Thus this probability will depend on the nature and extent of the seismic network, timing errors, processing techniques, and so on. We can estimate this probability as a function of distance from an assumed fault by considering a set of hypocenter location data (Fig. 6). We fit the distance from the epicenters to the fault with a normal distribution centered at the median of the data. The resulting normal distribution with a standard deviation of 0.46 km, together with bracketing normal distributions of 0.4 and 0.5 km, is shown in Figure 6. Thus we can conclude that for well-located earthquakes and relatively simple fault structures, the standard deviation for the distance between a located earthquake and the fault surface is about 0.5 km. In parts of the network where the station density is less, or where the fault zones may be more complex, this distance would be greater. However, so as to be conservative in assigning earthquakes to faults, we will adopt 0.5 km as the distance characterizing the standard deviation of the probability density function below. Obviously, the larger this distance, the more earthquakes will be associated with the faults and the fewer will be associated with the background. This tradeoff will be examined more fully below and in Appendix B.

We can consider the fit to the histogram in the form

$$n(r) = \exp\left(-\frac{r^2}{2\sigma_f^2}\right), \qquad (11)$$

where $n(r)$ is the number of earthquakes at distance r and σ_f is a measure of the dispersion, the standard deviation. Assume that the probability density for the distance of an earthquake location from a fault, given that it is associated with the ith fault, is proportional to equation (11). The probability that the earthquake will locate in a particular grid cell will be the integral of the density function over that grid cell. If the fault extends beyond the grid, we consider only the por-
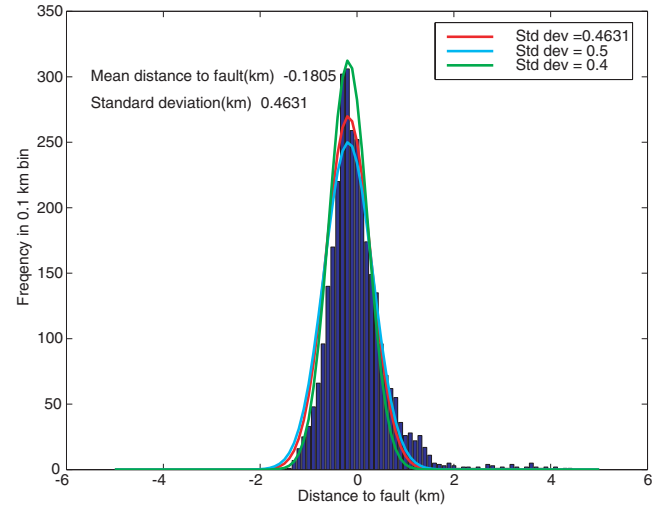

Figure 6.   Histogram of distances of 3135 NCSN epicenters from a best-fit polynomial curve representing the San Andreas fault for a set of earthquakes near San Juan Bautista. Although the data show a dominant central peak, they are not particularly well fit by the normal distribution, largely because of the data on the right-hand tail, which indeed may have a real geologic cause. Since we do not want the data on the right-hand tail to overly influence the standard deviation, we choose a normal distribution with a peak centered at the median of the data.

tion of the fault within the grid. Thus, an estimate of the probability that the earthquake will locate in grid cell k given that it is associated with the ith fault is

$$P(H_k|F_i) = \alpha_i \iint_{\substack{2D \text{ or } 3D \text{ grid cell}}} \exp\left(-\frac{x_{ki}^2}{2\sigma_f^2}\right), \qquad (12)$$

where

$$\alpha_i = \left(\sum_k \iint_{3D \text{ grid cell}} \exp\left(-\frac{x_{ki}^2}{2\sigma_f^2}\right)\right)^{-1}.$$

The normalizing factors, α_i, ensure that the probabilities will sum to 1 for each fault.

In contrast to the situation for instrumental locations, for the intensity data it is necessary to estimate σ_f in a more *ad hoc* manner. A value of 10 km has been adopted in the following. A sensitivity analysis (Appendix B) shows that this assumption is reasonable.

Many of the faults in the San Francisco Bay Area can be assumed to be vertical, and the horizontal distance from the fault to the grid cell is the only parameter needed. For other faults, if the depths of the earthquakes are available, then x_{ki} is the slant distance from the center of the grid cell to the fault. For the intensity data, or other situations where the depths of the earthquakes are not available, x_{ki} is the

shortest distance along the ground surface from the center of the grid cell to the line or polygon on the surface defined by projecting the fault surface vertically upward. The distance would be zero, for example, at a site located directly above the surface of a dipping fault.

Prior Probabilities for Grid Cells and Background

We now estimate prior probabilities for the occurrence of an earthquake in each grid cell, the prior probability that the earthquake is in the background (that is, it is not associated with any of the identified fault segments), and the probability that the earthquake will locate in a particular grid cell given that it is in the background.

A simple-minded analysis of the catalog of large earthquakes and the locations of faults in the San Francisco Bay region suggests, as a first approximation, that the probability of a large earthquake on an unknown fault is the order of 10%–20%, depending on the number of faults or fault segments included in the analysis. As will be shown below, the results are not sensitive to this assumption.

In constructing estimates of the prior probabilities for each grid cell and for each grid cell given that the earthquake is in the background, we are constrained by the basic postulates of probability and the total probability theorem. Obviously

$$\sum_k P(H_k|B) = 1, \quad \sum_k P(H_k) = 1,$$
$$P(H_k|B) \geq 0, \text{ and } P(H_k) \geq 0.$$

By the total probability theorem,

$$P(H_k|B) = \frac{P(H_k) - \sum_i P(H_k|F_i)P(F_i)}{P(B)}. \quad (13)$$

This imposes the additional constraint that

$$P(H_k) - \sum_i P(H_k|F_i)P(F_i) \geq 0.$$

All these constraints can be met by the following construction. Let

$$\delta = \left[1 - \sum_k \sum_i P(H_k|F_i)P(F_i)\right]/n_{\text{cells}} \text{ and}$$
$$P(H_k) = \delta + \sum_i P(H_k|F_i)P(F_i); \quad (14)$$

then determine $P(H_k \mid B)$ from equation (13). Note that in addition to satisfying the constraints, this construction has the desirable characteristic that at great distance from all faults, $P(H_k)$ tends to the nonzero constant, δ.

Examples

Example calculations of the prior and conditional probabilities per grid cell as described previously are shown in Figure 7 for $P(H_k \mid F_i)$, $P(H_k)$, $P(H_k \mid B)$, and $P(F_i \mid H_k)$. The calculations are done on a 2×2 km grid with a fault standard deviation of 10 km. $P(H_k \mid F_i)$ and $P(F_i \mid H_k)$ are shown for the Santa Cruz Mountains segment of the San Andreas fault ($i = 1$). $P(H_k \mid F_i)$ (Fig. 7a) is controlled only by distance from the fault trace. $P(H_k)$ (Fig. 7b) is larger for cells close to faults with larger fault priors, especially the southern Calaveras fault, and where several faults are in close proximity. $P(H_k \mid B)$ (Fig. 7c) is constant at about 1.1×10^{-4} per grid cell. $P(F_i \mid H_k)$ is shown in Figure 7d. $P(F_i \mid H_k)$ for the Santa Cruz mountains segment is obviously dominated by distance from the fault segment, but note that where the Santa Cruz segment approaches the southern Calaveras fault, the probabilities are decreased. This reflects the increased probability that an earthquake in those cells is associated with the southern Calaveras fault.

Association of Earthquakes and Faults

Earthquake Catalogs

The methods described above have been applied to three earthquake catalogs from the San Francisco Bay region. The first is a catalog of large events and associated intensity data for the period 1836–2000 assembled by Bakun (1999). Only earthquakes with magnitudes greater than or equal to 5.5 are included. (No additional earthquakes in this magnitude range occurred in the region from the completion of Bakun's work in 1998 through the end of 2000.) The second catalog is the University of California (UC) Berkeley Historic Earthquake Relocation Project (HERP) covering the period from 1951 to 1997 (Uhrhammer *et al.*, 1999). This catalog includes 1311 earthquakes in the study area and is complete above about magnitude 3. The third catalog is the U.S. Geological Survey–UC Berkeley Northern California Seismic Network (NCSN) catalog covering the period from 1984 to 1997. This catalog contains 36,956 earthquakes in the study area and is complete, depending on the location within the region, above magnitudes of between 1 and 2.5. Only earthquakes with high-quality locations were included, that is, location solutions with the number of stations greater than or equal to 8, the azimuth gap less than or equal to 200°, and the root mean square error less than or equal to 0.3 sec. The HERP and NCSN catalogs are also considered after the application of Reasenberg's declustering algorithm (Reasenberg, 1985) set to exclude earthquakes with magnitudes less than 1.0 and with the parameter RFACT set to 5.0. The declustered HERP and NCSN catalogs contain 1015 and 14,735 earthquakes in the study area.

In doing the associations, a relatively coarse, two-dimensional grid was used for the earthquakes for which only intensity observations are available, while a finer, three-dimensional grid was used for the modern earthquake cata-
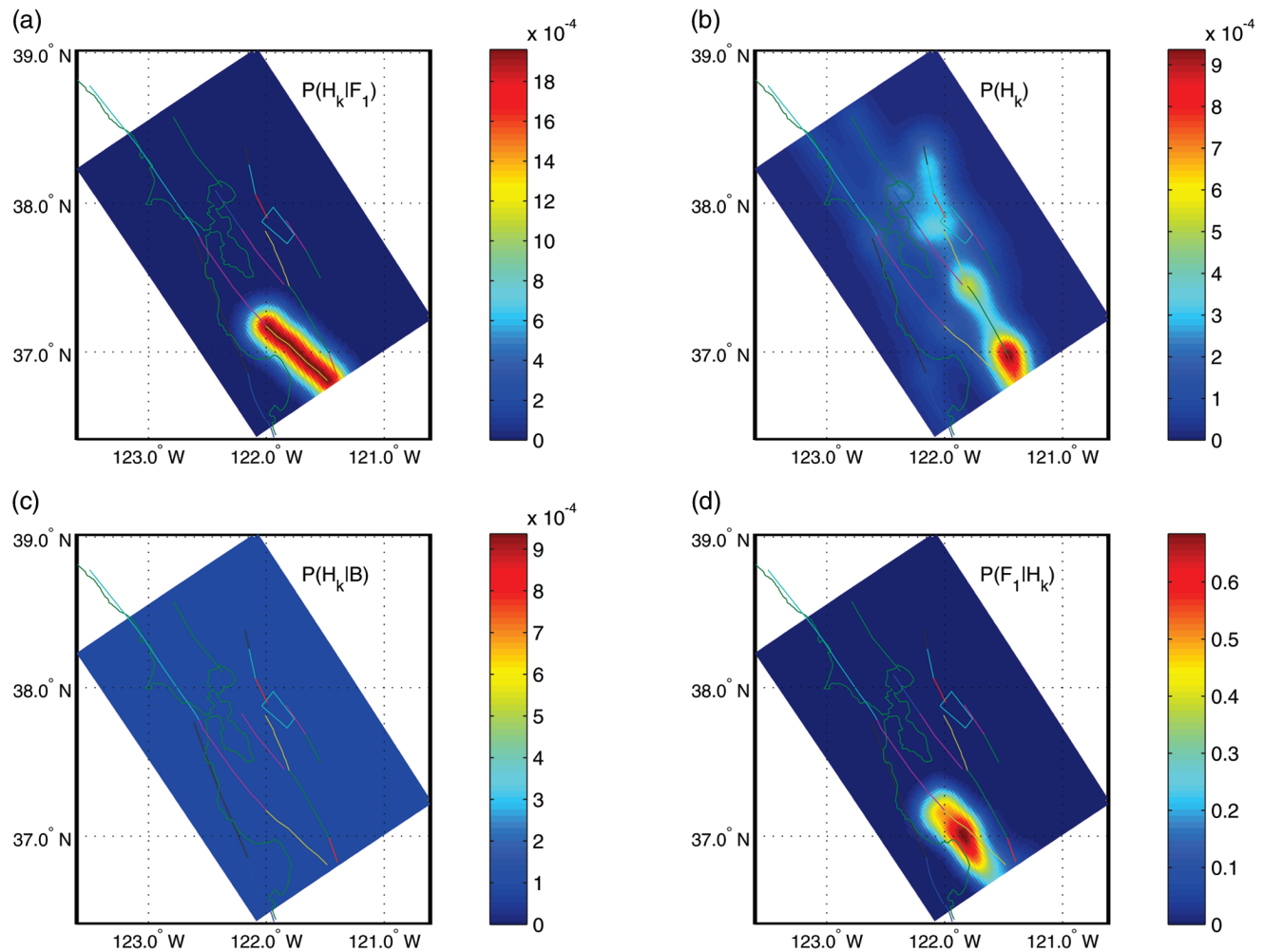
Figure 7.   (a) $P(H_k \mid F_i)$, (b) $P(H_k)$, (c) $P(H_k \mid B)$, and (d) $P(F_i \mid H_k)$ calculated for the WG99 fault model using a 2×2 km grid, a fault standard deviation of 10 km, and characteristic priors. $P(H_k \mid F_i)$ and $P(F_i \mid H_k)$ are shown for the Santa Cruz Mountains segment of the San Andreas fault ($i = 1$).

logs. A grid of 2 km $\times$ 2 km cells was assumed for the intensity-based catalog and of 1 km $\times$ 1 km $\times$ 1 km extending to a depth of 25 km for the modern catalogs. The probability that the epicenter or hypocenter of each earthquake was located in each of the grid cells was calculated using equation (4). Then the posterior, or "association," probabilities were calculated for each earthquake using characteristic fault priors.

Association for Preinstrumental Earthquakes

Results for the associations of the earthquakes in the Bakun catalog (excepting 1906) (Table 2) raise a number of interesting issues. If we believe, for example, that the southern Calaveras fault does indeed have a high recurrence rate, then it must be clearly considered as a candidate for several of the older earthquakes. A particularly illuminating example is given by the association probabilities for the 1868 earthquake, for which we have strong geologic evidence that

the earthquake occurred on the southern Hayward fault. The procedure estimates a 0.46 probability of association with the southern Hayward fault, 0.13 with the northern Hayward fault, 0.21 with the northern Calaveras fault, and 0.10 with the background. In view of the geologic evidence, however, in the subsequent analysis the earthquake is assigned to the southern Hayward fault with a probability of 1.0.

Association for the HERP and NCSN Catalogs

Association probabilities were calculated for the HERP and NCSN catalogs. The results for the NCSN catalog within the region of interest are shown in Figure 8. In this figure each event is plotted at its epicenter, and the color and symbol indicate the dominant association probability, that is, the fault segment (or background) for which the association probability is greater than 0.5. If there is no dominant probability, a black circle is plotted at the epicenter. Again, it is worth emphasizing that these associations are based primar-

Table 2

Association Probabilities for Earthquakes, $M \geq 5.5$, for WG99 Segments

Date	Mag	SAF - Santa Cruz Mountains	SAF - Peninsula	SAF - North Coast South	SAF - North Coast North	Southern Hayward	Northern Hayward	Rogers Creek	Southern Calaveras	Central Calaveras	Northern Calaveras	Concord	Southern Green Valley	Northern Green Valley	San Gregorio South	San Gregorio North	South Greenville	North Greenville	Mount Diablo	Background
10 June 1836	6.5	0.20	0.03	0.00	0.00	0.00	0.00	0.00	0.23	0.16	0.00	0.00	0.00	0.00	0.09	0.04	0.00	0.00	0.00	0.25
June 1838	6.8	0.06	0.27	0.00	0.00	0.18	0.01	0.00	0.00	0.18	0.09	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.17
27 Aug 1855	5.5	0.00	0.01	0.11	0.00	0.01	0.08	0.21	0.00	0.00	0.01	0.03	0.06	0.13	0.00	0.01	0.00	0.01	0.01	0.33
2 Jan 1856	5.5	0.07	0.10	0.02	0.00	0.03	0.02	0.00	0.04	0.06	0.02	0.01	0.00	0.00	0.09	0.13	0.01	0.01	0.01	0.41
15 Feb 1856	5.8	0.01	0.26	0.05	0.00	0.14	0.10	0.00	0.00	0.02	0.04	0.02	0.01	0.00	0.00	0.10	0.03	0.01	0.01	0.23
26 Nov 1858	6.1	0.04	0.13	0.00	0.00	0.17	0.01	0.00	0.00	0.30	0.13	0.00	0.00	0.00	0.00	0.02	0.03	0.00	0.02	0.14
17 Apr 1860	5.5	0.11	0.00	0.00	0.00	0.00	0.05	0.00	0.22	0.07	0.00	0.05	0.00	0.00	0.21	0.02	0.03	0.04	0.00	0.37
4 July 1861	5.8	0.00	0.05	0.00	0.00	0.21	0.00	0.00	0.00	0.12	0.23	0.00	0.01	0.00	0.00	0.01	0.03	0.01	0.08	0.12
26 Feb 1864	5.9	0.17	0.05	0.00	0.00	0.01	0.00	0.00	0.13	0.14	0.01	0.00	0.00	0.00	0.13	0.07	0.02	0.01	0.00	0.30
5 Mar 1864	5.9	0.13	0.09	0.00	0.00	0.03	0.00	0.00	0.09	0.21	0.02	0.00	0.00	0.00	0.04	0.07	0.03	0.00	0.00	0.28
21 May 1864	5.6	0.08	0.10	0.01	0.00	0.05	0.01	0.00	0.04	0.19	0.04	0.00	0.00	0.00	0.01	0.06	0.01	0.00	0.01	0.35
24 May 1865	5.9	0.16	0.02	0.00	0.00	0.00	0.00	0.00	0.23	0.20	0.04	0.00	0.00	0.00	0.07	0.03	0.00	0.00	0.00	0.27
8 Oct 1865	6.5	0.26	0.31	0.00	0.00	0.06	0.00	0.00	0.00	0.18	0.04	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.15
26 Mar 1866	5.6	0.19	0.01	0.00	0.00	0.00	0.00	0.00	0.30	0.16	0.00	0.00	0.00	0.00	0.00	0.03	0.01	0.00	0.00	0.23
15 July 1866	5.9	0.02	0.00	0.00	0.00	0.46	0.13	0.00	0.00	0.09	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.77
21 Oct 1868*	6.8	0.27	0.01	0.00	0.00	0.02	0.00	0.00	0.01	0.01	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.10
17 Feb 1870	5.8	0.00	0.19	0.00	0.00	0.19	0.23	0.00	0.00	0.17	0.02	0.00	0.00	0.00	0.03	0.08	0.00	0.00	0.00	0.20
2 Apr 1870	5.6	0.00	0.09	0.05	0.00	0.01	0.00	0.02	0.09	0.01	0.09	0.08	0.04	0.01	0.00	0.02	0.00	0.01	0.04	0.12
10 Apr 1881	6.1	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.10	0.02	0.00	0.00	0.00	0.00	0.00	0.15	0.04	0.01	0.66
28 June 1882	5.9	0.30	0.13	0.00	0.00	0.01	0.00	0.00	0.05	0.19	0.01	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.20
30 Mar 1883	5.9	0.26	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.12	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.24
26 Mar 1884	6.1	0.15	0.03	0.00	0.00	0.00	0.01	0.31	0.07	0.07	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.35
19 May 1889	6.0	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.06	0.31	0.00	0.01	0.22	0.10	0.00	0.09	0.21	0.19
31 July 1889	5.6	0.00	0.10	0.03	0.00	0.28	0.29	0.00	0.20	0.18	0.07	0.04	0.12	0.00	0.00	0.02	0.00	0.00	0.03	0.12
24 Apr 1890	6.3	0.45	0.01	0.00	0.00	0.00	0.00	0.00	0.34	0.15	0.06	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.15
2 Jan 1891	5.7	0.19	0.17	0.00	0.00	0.00	0.00	0.00	0.01	0.35	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.00	0.13
12 Oct 1891	5.8	0.00	0.00	0.00	0.00	0.00	0.20	0.31	0.00	0.00	0.00	0.02	0.13	0.19	0.00	0.00	0.00	0.00	0.00	0.15
19 Apr 1892	6.4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.09	0.00	0.00	0.00	0.00	0.00	0.90
21 Apr 1892	6.4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.97
30 Apr 1892	5.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.13	0.00	0.00	0.00	0.00	0.00	0.85
13 Nov 1892	5.8	0.27	0.02	0.00	0.00	0.00	0.00	0.00	0.18	0.13	0.00	0.00	0.00	0.00	0.09	0.04	0.00	0.00	0.00	0.27
9 Aug 1893	5.5	0.00	0.01	0.00	0.00	0.00	0.06	0.39	0.00	0.00	0.00	0.00	0.01	0.03	0.00	0.00	0.01	0.00	0.00	0.29
20 June 1897	6.3	0.31	0.00	0.20	0.00	0.00	0.00	0.19	0.13	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18
31 Mar 1898	6.3	0.00	0.00	0.00	0.00	0.02	0.38	0.00	0.00	0.00	0.00	0.13	0.15	0.04	0.00	0.00	0.00	0.00	0.00	0.09
30 Apr 1899	5.9	0.28	0.01	0.00	0.00	0.00	0.00	0.03	0.00	0.06	0.02	0.00	0.00	0.00	0.13	0.05	0.00	0.00	0.00	0.34
2 June 1899	5.5	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.51	0.00	0.00	0.03	0.02	0.01	0.00	0.05	0.00	0.00	0.00	0.15
6 July 1899	5.9	0.08	0.16	0.12	0.00	0.00	0.27	0.00	0.00	0.16	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.24
19 May 1902	5.9	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.01	0.00	0.03	0.00	0.07	0.16	0.00	0.00	0.00	0.00	0.00	0.76
11 June 1903	6.1	0.32	0.20	0.00	0.00	0.03	0.00	0.00	0.00	0.25	0.11	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.00	0.14
3 Aug 1903	6.1	0.06	0.19	0.00	0.00	0.17	0.00	0.00	0.00	0.32	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.13
11 Mar 1910	5.8	0.22	0.05	0.00	0.00	0.14	0.00	0.00	0.12	0.11	0.13	0.00	0.00	0.00	0.13	0.08	0.00	0.00	0.00	0.28
01 Jul 1911	6.2	0.04	0.05	0.00	0.00	0.00	0.00	0.00	0.05	0.55	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08
09 Nov 1914	5.5	0.33	0.11	0.00	0.00	0.00	0.00	0.00	0.02	0.13	0.00	0.00	0.00	0.00	0.07	0.07	0.00	0.00	0.00	0.22
22 Oct 1926	6.1	0.50	0.13	0.00	0.00	0.00	0.00	0.00	0.02	0.15	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.17
15 Feb 1927	5.9	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.29	0.09	0.00	0.00	0.00	0.00	0.14	0.01	0.00	0.00	0.00	0.31
05 Sep 1955	5.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.49	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.51
02 Oct 1969	5.6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
02 Oct 1969	5.7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.98
06 Aug 1979	5.7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.81
24 Jan 1980	5.8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.60	0.07	0.33
24 Apr 1984	6.2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
31 Mar 1986	5.6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
18 Oct 1989	6.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Probabilities for earthquakes prior to 1955 are determined using locations derived from intensity data. Probabilities for later earthquakes are determined using locations in the HERP catalog. Probabilities greater than or equal to 0.1 and less than 0.2 are surrounded by light borders. Probabilities greater than or equal to 0.2 are surrounded by bold borders.

*In subsequent analysis the 21 October 1868 earthquake is assumed to be associated with the southern Hayward fault with a probability of 1 based on geologic observations.
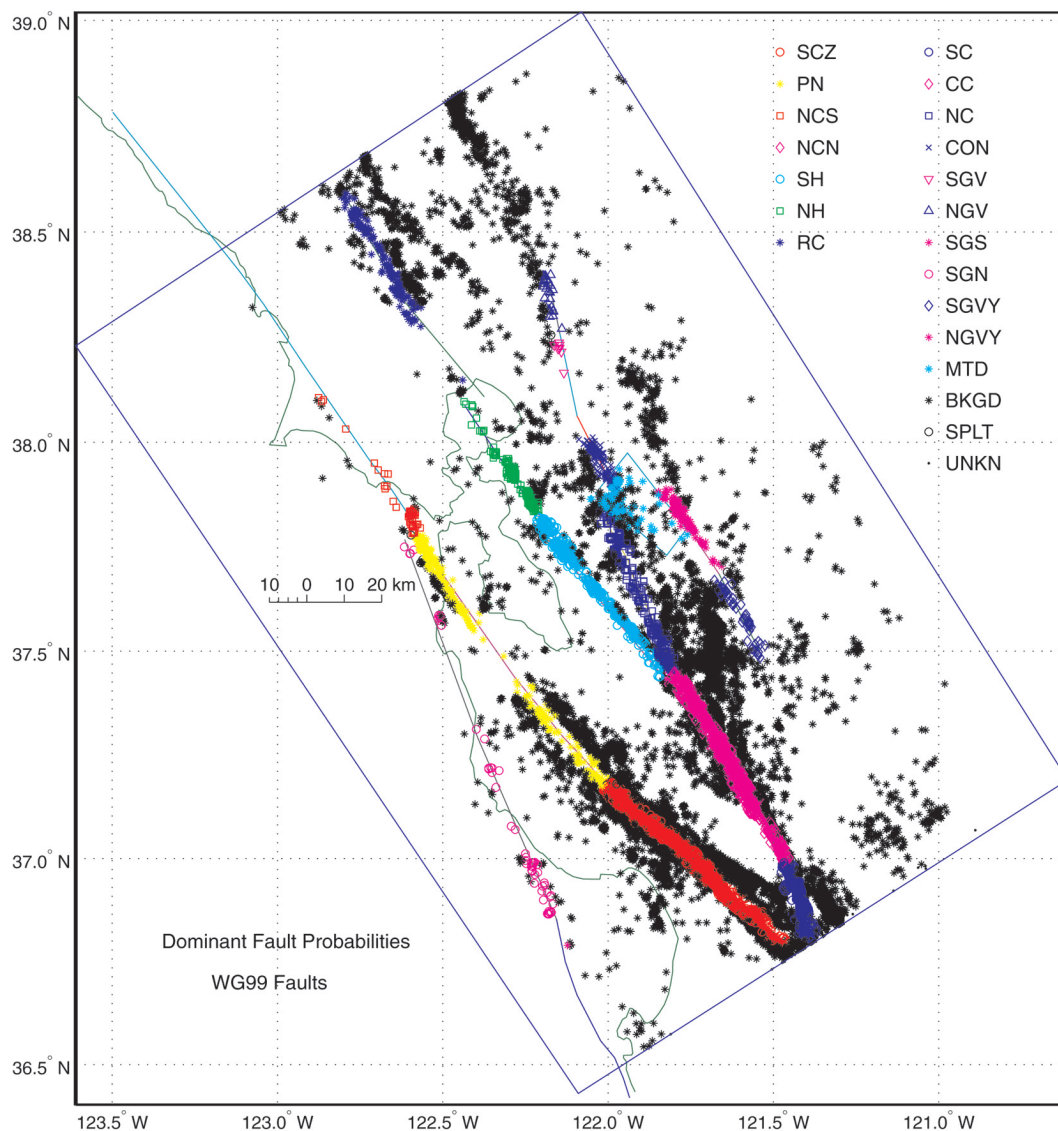
1317

Figure 8.   Earthquakes in the NCSN catalog coded to show the dominant probability of association using the fault segments from WG99. A symbol corresponding to a fault segment or the background is assigned to an earthquake when the corresponding association probability is greater than or equal to 0.5. When there is no dominant association probability, that is the largest association probability for the earthquake is less than 0.5, an open circle (SPLT) is assigned. If for some reason the association process failed (usually because the earthquake is too close to the edge of the grid), the earthquake is shown as a small solid dot (UNKN).

ily on distance from faults. The results of a detailed tectonic analysis of events may well yield a different result, as discussed subsequently. The three-dimensional character of the association process can be seen more clearly in Figure 9.

Figure 8 provides a number of interesting observations about seismicity and its relation to faults in the San Francisco Bay region. In general the bands of earthquakes north of the bay are less concentrated than those along the San Andreas, Hayward, and Calaveras fault systems. It remains to be determined whether this difference in the apparent character of

the seismicity is a reflection of a difference in fault behavior or a result of earthquake location issues. The more diffuse nature of the seismicity along the San Andreas fault along the Santa Cruz Mountain and Peninsula segments is well established. It may be argued that the distribution of earthquakes around fault segments that are creeping (southern Santa Cruz Mountains, southern and central Calaveras, northern and southern Hayward) is more tightly compacted about the fault, as contrasted with fault segments that are not creeping. It seems clear, however, that our assumption
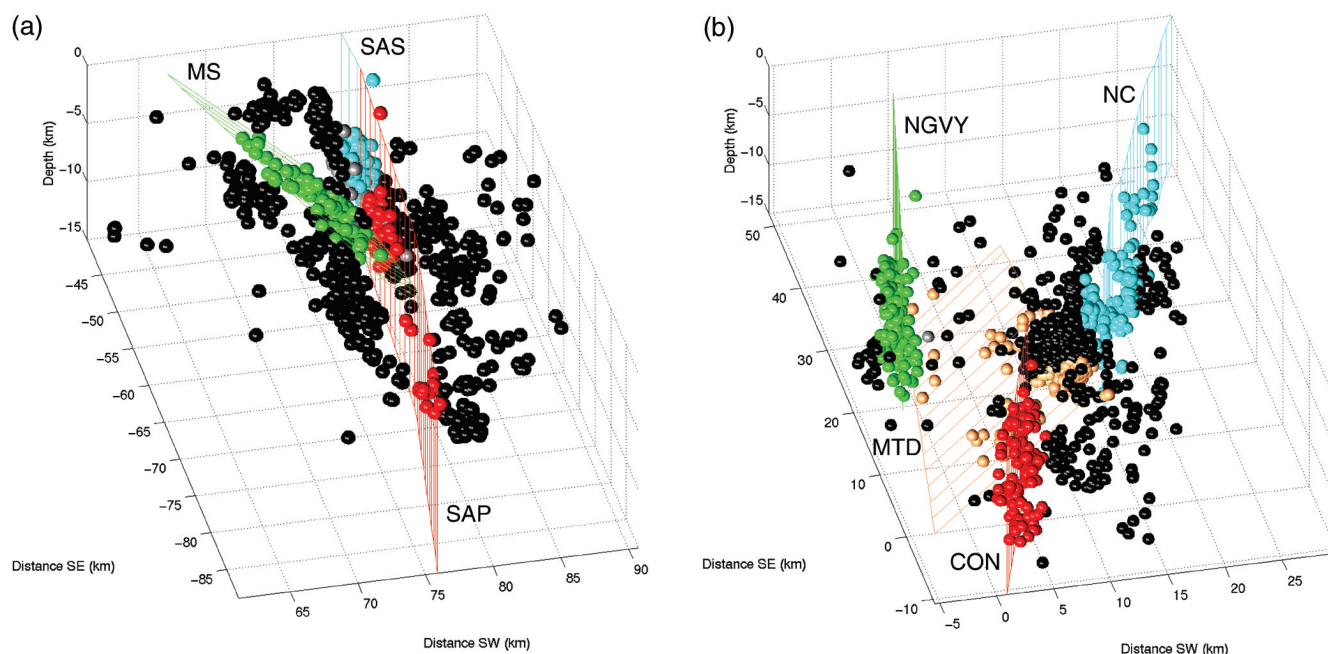
Figure 9. (a) A detail of the three-dimensional association of earthquakes in the NCSN catalog using the more detailed set of fault segments given by WG96 along the peninsula segment of the San Andreas fault. View looking southeast. Earthquakes associated with the peninsula segment (SAP) are shown in red, those with the Santa Cruz Mountains segment (SAS) in blue, those with the Monte Vista–Shannon fault (MS) in green, and those with the background in black. Earthquakes with no dominant association probability (that is, the maximum probability loss than 0.5) are shown in gray. (b) A detail of the three-dimensional association of earthquakes in the NCSN catalog with the WG99 fault set in the vicinity of Mount Diablo. Earthquakes associated with the Concord fault (CON) are shown in red, with the Mount Diablo fault (MTD) in copper, with the northern Calaveras fault (NC) in blue, the northern Green Valley fault (NGVY) in green, and the background in black. Earthquakes with no dominant association probability (that is, the maximum probability loss than 0.5) are shown in gray.

of one value of the fault standard deviation for the entire region is simplistic. There are several concentrations of earthquakes that are not associated with a fault in the data set. Presumably many of these concentrations of earthquakes will eventually be associated with faults still to be discovered and will thus be removed from the background.

How are the observed earthquakes apportioned among the candidate faults and the background? Figure 10 answers this question for the NCSN and HERP catalogs. The largest fraction of earthquakes in these catalogs (more than 50%) is associated with the background. The second largest fraction, about 20%, is associated with the central Calaveras segment. The third largest fraction, between 6% and 14%, is associated with the Santa Cruz Mountains segment. The fraction of earthquakes associated with each of the other segments is 4% or less. Note that this result is not affected much by whether the full or declustered catalog is used.

Also shown in Figure 10 are the fault prior probabilities. The low level of correlation between the fault priors and the results supports the assertion that the results are relatively insensitive to the priors.

Of course, as mentioned above, the fraction of earthquakes associated with the background versus the candidate faults depends on the value of the parameter selected for the fault standard deviation. The larger the fault standard deviation, the larger will be the fraction of earthquakes associated with the faults and the smaller the fraction associated with the background. The sensitivity of the percentage of earthquakes associated with the background as a function of the fault standard deviation for the NCSN catalog and the WG99 set of faults is examined in Figure 11. Notice first that the curve is rather steep in the vicinity of the preferred value of 0.5 km. Nonetheless, notice also that as the fault standard deviation is varied over a range of a factor of 20, the percentage of earthquakes associated with the background varies by only about a factor of 2. Indeed, an unreasonably large value of the fault standard deviation would have to be assumed to associate less than about one-third of the earthquakes with the background. Considering that the prior probability for background was assumed to be 0.2, these results indicate both the lack of sensitivity to that assumption and that, at least for a fault standard deviation of
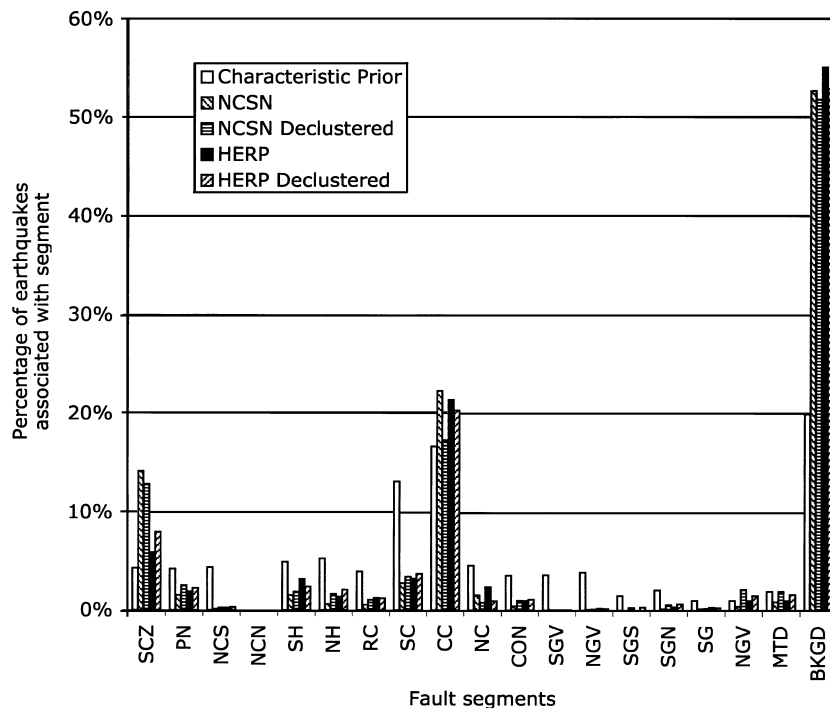
Figure 10. Percentage of earthquakes in the NCSN and HERP catalogs associated with the fault segments in the WG99 model and the background. Also shown are the characteristic fault priors used in the associations. Note that there is little correlation between the priors and the resulting percentages, indicating that the results are not particularly sensitive to the priors.
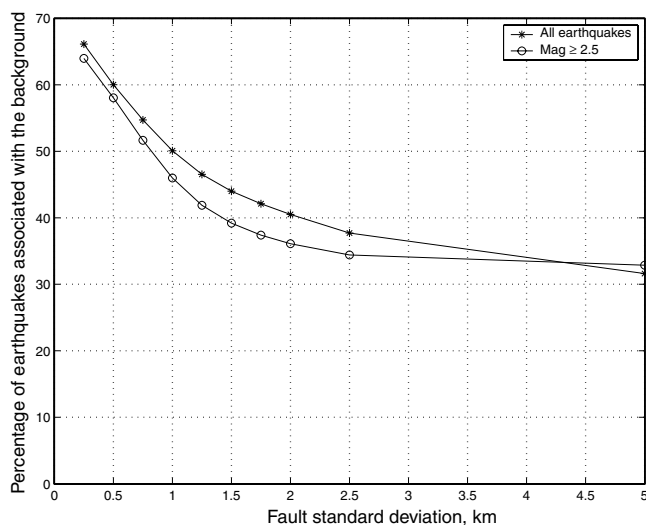

Figure 11. Percentage of earthquakes in the NCSN catalog associated with the background (relative to the WG99 set of fault segments) as a function of the assumed value of the fault standard deviation parameter.

0.5 km, a value closer to 0.5 would be more appropriate. The sensitivity of the results for the locations of the older earthquakes based on intensity is discussed in Appendix B.

Implications for Association and Discussion

The inference scheme as implemented here does not overcome certain limitations of the data and of our understanding regarding the character of faults and their relationship to earthquakes. The principal issue is a mixing of the

lack of knowledge of the geometry of the fault zones at depth and the amount and direction of systematic biases in the calculation of earthquake hypocenters. In addition to the 1906 earthquake, five historical earthquakes have reports of ground failures that can be reasonably ascribed to surface offset on a mapped fault trace (Bakun, 1999). These are June 1838 and 24 April 1890 on the San Andreas fault, 4 July 1861 and 20 June 1897 on the Calaveras fault, and 21 October 1868 on the Hayward fault. For each of these events, the maximum association in Table 2 is with the fault segment where cracks on the fault trace were reported, although the difference between the 0.23 (northern Calaveras segment) and 0.21 (southern Hayward segment) associations for the 1861 event are not significant. If we admit the uncertainty of the WG99 fault segmentation model and sum the associations over a fault, then the five earthquakes are clearly preferentially associated with the fault where ground failures were reported. However, given the concentration of active faults in the region, we must acknowledge considerable uncertainty about associating historical earthquakes and faults using intensity data only.

It is interesting to compare the associations based on the modern, instrumental data with the associations based on intensity data for those earthquakes for which we have both. Table 3 shows the association probabilities for selected modern earthquakes determined from intensity data alone. In contrast to the probabilities for these earthquakes determined from the instrumental locations (Table 2), which favor the background for most of these events, the probabilities determined from the intensity-derived locations much more strongly favor the faults. This results from the much larger fault standard deviation adopted for the intensity locations.

Table 3
Association Probabilities for Selected Modern Earthquakes Determined from Intensity Data Alone

Date	Mag	SAF - Santa Cruz Mountains	SAF - Peninsula	SAF - North Coast South	SAF - North Coast North	Southern Hayward	Northern Hayward	Rogers Creek	Southern Calaveras	Central Calaveras	Northern Calaveras	Concord	Southern Green Valley	Northern Green Valley	San Gregorio South	San Gregorio North	South Greenville	North Greenville	Mount Diablo	Background
06 Aug 1979	5.7	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.49	0.46	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04
24 Jan 1980	5.8	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.11	0.12	0.00	0.00	0.00	0.00	0.00	0.26	0.21	0.09	0.17
24 Apr 1984	6.2	0.02	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.86	0.02	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.06
31 Mar 1986	5.6	0.01	0.01	0.00	0.00	0.12	0.00	0.00	0.00	0.66	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05
18 Oct 1989	6.9	0.48	0.34	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.13

Probabilities greater than or equal to 0.1 and less than 0.2 are surrounded by light borders. Probabilities greater than or equal to 0.2 are surrounded by bold borders.

In general the intensity-based association probabilities are more satisfying in that they reflect associations of faults and earthquakes more in line with detailed geologic and seismologic studies of individual earthquakes, such as for the 1979 and 1984 earthquakes with the Calaveras fault and the 1980 earthquake with the Greenville faults. The exceptions are illuminating. The 1986 earthquake occurred near Mount Lewis and is most definitely not associated with the Calaveras fault; however, the intensity-based association draws the earthquake to the Calaveras. Similarly, the intensity-based location associates the Loma Prieta earthquake with the San Andreas fault. Of course the association process used here does not take into account the arguments, discussed later, for not assigning this earthquake to the San Andreas.

The associations of the 1979 Coyote Lake and 1989 Loma Prieta earthquakes illustrate the difficulties. Using the standard instrumental locations and descriptions of the fault segments, neither of these earthquakes were associated with the nearby faults. The location of the Coyote Lake earthquake in the HERP catalog is 3–4 km northeast of the trace of the Calaveras fault. This location has endured intensive analysis (cf. Reasenberg and Ellsworth, 1982). The general assumption throughout the current work is that the strike-slip faults in the region can be considered to be vertical. Detailed studies of the aftershocks by Reasenberg and Ellsworth, together with the hypocenter and focal mechanism of the main shock, suggest that the Calaveras fault at depth in this region is composed of three subparallel surfaces within a zone of about 5 km width, each dipping at near vertical to 76° northeast. This complexity is obviously not reflected in the simple, vertical planar fault plane with fault standard deviation of 0.5 km assumed in the association.

The hypocenter of the 1989 Loma Prieta earthquake is located at 18 km depth about 5 km to the southwest of the San Andreas fault. Aftershocks of the earthquake lie in a southwest-dipping zone, which approaches the trace of the San Andreas fault at shallow depth. However, here the San Andreas fault proper is interpreted to be near vertical based on the rupture in 1906, exposures in a tunnel, and seismicity and velocity studies (Eberhart-Phillips and Michael, 1998). While it has been argued that the Loma Prieta earthquake did not occur on the same fault surface as that which ruptured in 1906 (Prentice and Schwartz, 1991), the 1906 and 1989 ruptures are so close together that from the point of view of seismic hazard, they might be considered to have originated on the same fault.

It now seems clear that the ability to estimate relative earthquake locations far exceeds both our understanding of the systematic biases and our understanding of the fault geometry at depth. One approach to this problem would be to apply a regionally variable fault standard deviation. In any case caution must be exercised in the interpretation of the current results from this point of view.

As can be seen in Figure 9a,b, the association process works in complicated geometries. What must be remembered is that the current method associates earthquakes with

the fault segments without regard to factors other than the distance, the uncertainties in the location, and the priors. In the case of Figure 9a, some of the earthquakes associated with the peninsula segment of the San Andreas fault are those for which focal mechanism solutions have been used by Zoback *et al.* (1999) to argue that the earthquakes are not San Andreas earthquakes. In contrast, in Figure 9b earthquakes seem to be well associated with the Mount Diablo and northern Calaveras faults. Unfortunately, detailed studies, including focal mechanisms, by Walter *et al.* (1998) indicate that most, if not all, of these earthquakes are more properly associated with east–west–striking faults not included in the fault set. In short, the procedure associates earthquakes with the fault set it is given and cannot detect faults on its own.

Applications

Frequency–Magnitude Relations

Probabilistic annual frequency–magnitude curves were calculated for each fault segment and the background by summing the association probabilities for the segment in 0.25 magnitude bins centered on the quarter magnitudes (i.e., 6.0, 6.25, 6.5, . . .). Results from the NCSN catalog were used for the magnitude range $1.25 < M < 3.5$ for the time period 1984–1997, from the HERP catalog for the magnitude range $3.5 \leq M < 5.5$ for the time period 1951–1998, and from the Bakun (1999) and HERP catalogs for $M \geq 5.5$ for the time period 1836–2000. Bakun claims completeness for his catalog at M 5.5 beginning in 1850. Although earthquakes with $M \geq 5.5$ in the time period 1836–1850 may be missing, we extend the period back to 1836 because of the importance of the 1836 and 1838 earthquakes. The 1906 earthquake, although not considered in the associations, is included in the frequency–magnitude graphs for the entire region.

Bakun and many other authors have noted the significant difference in the rate of occurrence of large earthquakes in the San Francisco Bay region prior to and following the 1906 earthquake. Commonly, the observation of relatively frequent large earthquakes in the 70 years prior to 1906 and the relative absence of large earthquakes in the 70 years following 1906 are interpreted in terms of an "earthquake cycle," in which the activity prior to 1906 reflects a high level of stress throughout the region and the absence of large events after 1906 reflects the low level of stress following the release of stress in 1906. Consequently, it may be argued that the piecing together of frequency–magnitude curves from different time periods is inappropriate. The aim here is to make the best estimate of the long-term average behavior; therefore it is appropriate to use all the data available.

Magnitudes for the earthquakes with $M \geq 5.5$ are those given by Bakun (1999). Where a large number of earthquakes is associated with a segment, the frequency–magnitude relationship is reasonably well fit with the classic

Gutenberg–Richter frequency–magnitude relationship,

$$\log_{10} n = a - bM,$$

where n is the number of earthquakes in each magnitude bin and a is the incremental a value. [The incremental a value is related to the cumulative a value by $a_{inc} = a_{cum} + \log_{10}(10^b \Delta^M - 10^{-b \Delta^M})$, where ΔM is the width of the magnitude bins (Herrmann, 1977).] Estimates for a and b were calculated for each segment for both the complete and declustered catalogs using the maximum likelihood method (Weichert, 1980), weighted least-squares on the incremental frequency, and least-squares on the cumulative frequency. In the weighted least-squares analysis, the observations for each bin were weighted by the number of observations within the bin. Results are shown for a sampling of fault segments and the background in Figure 12. Where data are plentiful, all these techniques yield similar results.

The frequency–magnitude curves also suggest that most of the fault segments are appropriately modeled by the Gutenberg–Richter relationship from low magnitudes to very close to the characteristic magnitude. Thus, while the rate of characteristic earthquakes may be as much as an order of magnitude greater than the rate of earthquakes in the magnitude bin just below the characteristic magnitude, the rates of potentially damaging earthquakes with magnitudes below that of the characteristic earthquake are significant.

Probability of a "Background Earthquake"

An important issue in many earthquake hazard studies is the determination of the probability of a background earthquake of sufficiently large magnitude to be of concern. We can estimate the rate of earthquakes with magnitude of, say, 6.5 and greater directly. The maximum likelihood predictor of the rate of earthquakes in a given magnitude range is simply the observed number of earthquakes in that magnitude range divided by the time period of observation.

We could also extrapolate the Gutenberg–Richter relations obtained from the Bayesian analysis (Fig. 12e,f) to 6.5. We are cautious about applying this extrapolation, especially from the Gutenberg–Richter relations from much lower magnitudes, because of the uncertainty in the b values and because the magnitude–frequency relation may not be linear in this range, as we are near the maximum expected magnitude for the background.

As discussed above, the rate of seismicity in the San Francisco Bay region appears to have varied significantly through time. For our purpose, we will ignore these temporal variations and consider estimates of the long-term average based on the entire period of observation, in our case 165 years (1836–2000). (For extensive discussion of inferences about the variability of the seismicity rate, see WG99.)

Obviously the numbers of observed events of magnitude greater than or equal to 6.5 are small. Our challenge is to use the available data in the most sensible and effective manner in constructing our estimates. Within the Bakun catalog
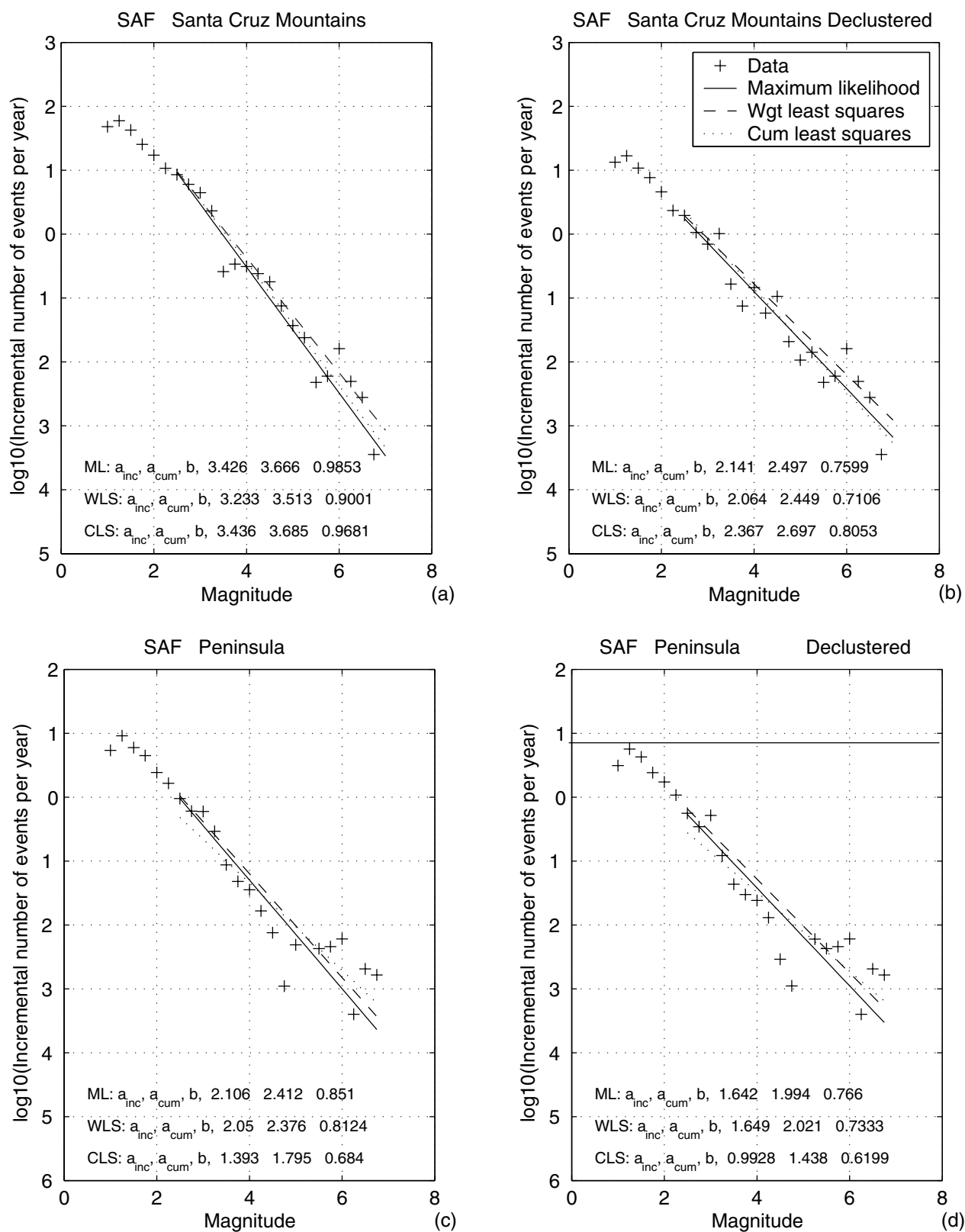
Figure 12.   Frequency–magnitude plots for the WG99 segment model from observed and declustered data sets showing annual incremental summed probabilities in each magnitude bin and maximum likelihood and least-squares fits to Gutenberg–Richter parameters. Calculations were done assuming a prior background probability of 0.2,   (*continued*)
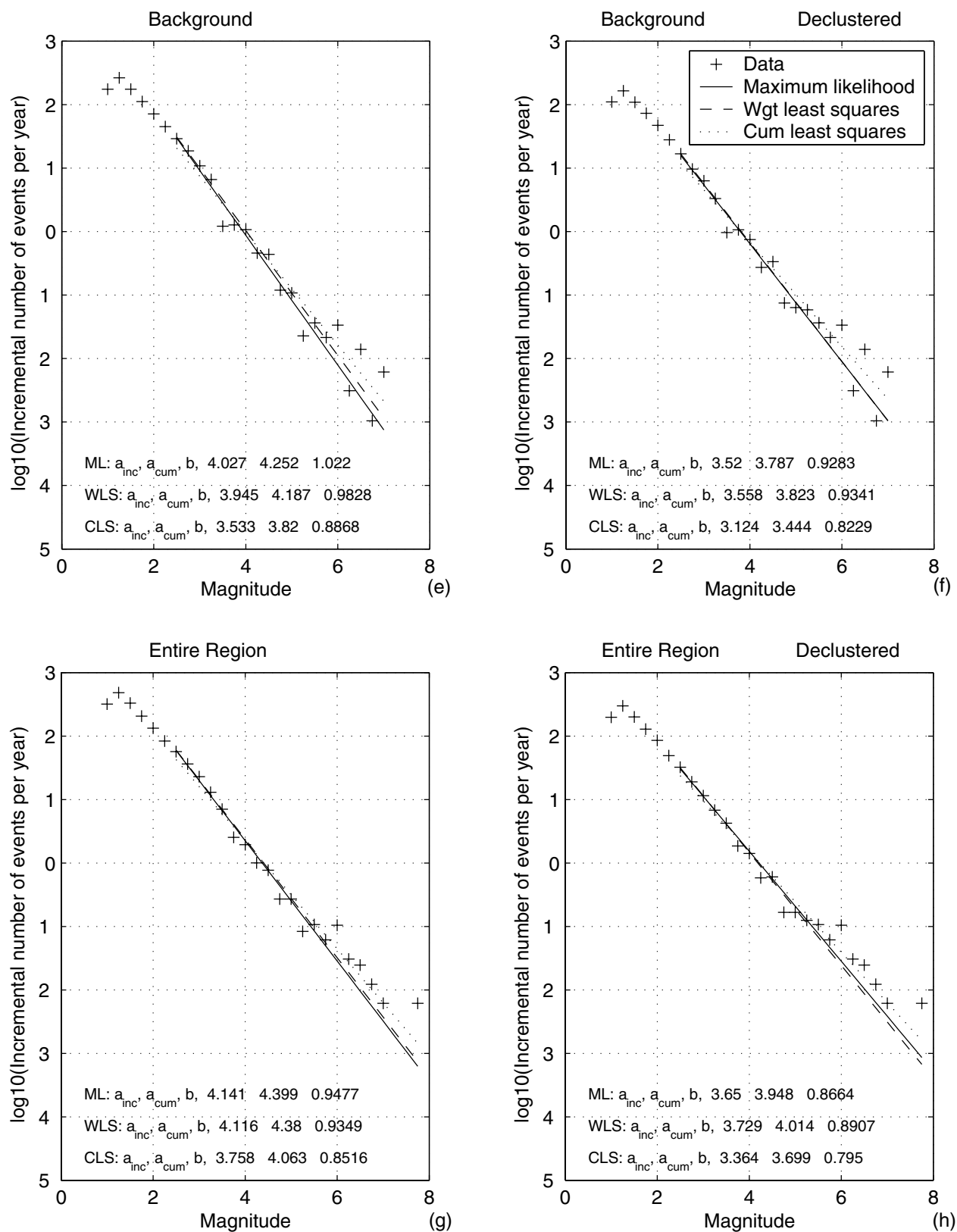
Figure 12. (*Continued*)   a fault standard deviation of 0.5 km for the HERP and NCSN data, and of 10 km for the earthquakes located using intensity data. Estimates of *a* and *b* were calculated using the techniques of maximum likelihood (ML), weighted least sqares on the incremental frequency (WLS), and least squares on the cumulative frequency (CLS).

for the area under consideration there are a total of six earthquakes in 165 years with magnitudes greater than 6.5. We estimate in probabilistic terms from the Bayesian analysis that 1.57 of those earthquakes occurred in the background (Table 4). Is this a reasonable number? This number includes the probabilities that the 1838, 1836, and 1865 earthquakes occurred in the background (0.17, 0.25, and 0.15, respectively).

However, the Bayesian analysis does not adequately capture the uncertainties related to the association of the Loma Prieta earthquake and the magnitudes of the 1892 earthquakes. The Loma Prieta earthquake, based on the NCSN location, the assumption of a vertically dipping San Andreas fault, and the assumption of a fault "half-width" of 0.5 km, is associated with the background with a probability of 1.0. All the other earthquakes with magnitudes greater than or equal to 6.5 were located and associated on the basis of intensity observations with an assumed fault half-width of 10 km. If the Loma Prieta earthquake were considered to be associated with the Santa Cruz Mountains segment of the San Andreas fault, then only 0.57 earthquakes would be associated with the background.

In contrast, the two earthquakes in 1892 are most likely associated with the background (background probabilities of 0.90 and 0.97), but they are both assigned magnitudes of 6.4 and therefore neither are included in the count!

As shown in Table 4, the estimated 30-year probabilities from this direct approach for a magnitude 6.5 or larger event thus range from 10%, assuming Loma Prieta is *not* in the background and that the two 1892 earthquakes are appropriately excluded, to 46%, assuming Loma Prieta *is* in the background and that the two 1892 earthquakes should also be included.

The Exponential Tail

Another important issue in many earthquake hazard studies is the question of the "exponential tail," that is, the frequency of potentially damaging earthquakes with magnitudes less than the characteristic magnitude. Put another way, the exponential tail is that part of the magnitude–frequency distribution of independent earthquakes (i.e., not aftershocks) on characterized faults that is not accounted for in the characteristic earthquakes. What part of the available moment rate is released in these smaller independent events and not available to produce the larger characteristic events? The historical seismicity (Bakun, 1999) and the association of earthquakes with faults can be used to answer this question. Suppose that each fault, for purposes of a hazard study, has a threshold magnitude above which the earthquakes will be characteristic in the sense that they rupture either the entire fault segment or a large fraction of it. Table 5 shows estimates of this threshold for the WG99 faults. Then using the association probabilities in Table 2, for each fault we can sum the portion of moment for each earthquake attributed to that fault for all earthquakes below the threshold magnitude. That is, we sum the products of the association probabilities times the earthquake moment, fault by fault, for events below the threshold magnitude.

Then the seismic moment rate for earthquakes in the exponential tail for each fault is this sum divided by 150 years, the period of completeness of Bakun's catalog (Table 5). Whereas 150 years is not long enough to sample many characteristic earthquakes, it may be long enough to sample the character of smaller events on each fault. If the threshold magnitude is much larger than 5.5, then the moment for missing magnitude less than 5.5 events can be ignored. The moment for these events on the Calaveras fault cannot be ignored, so that the exponential tail estimates for the Calaveras fault are not appropriate.

Also shown in Table 5 are WG99 estimates of the annual seismic moment rate for each fault, taking into account WG99 estimates of the lengths, widths, geologic slip rate, and corrections for aseismic slip. The ratio of the estimated annual seismic moment in the exponential tail over the last 150 years, divided by the WG99 estimate of the long-term seismic moment rate, provides an estimate of the fraction of seismic moment in the exponential tail (Table 5). For example, 9% of the model moment rate is in the exponential tail for that part of the San Andreas fault contained within the WG99 area. The mean of these ratios (excluding the Calaveras fault) is 6%, which agrees well with the estimate of Youngs and Coppersmith (1985) for the San Francisco Bay Area.

This analysis suggests a variation to the strategy used by Frankel *et al.* (1996) and Peterson *et al.* (1996) in preparing probabilistic seismic hazard maps. Their approach combined a fault-specific approach at higher magnitudes together with smoothed seismicity at the lower magnitudes. This analysis suggests an approach for associating smaller-magnitude earthquakes with faults. This offers the possibility that some of the hazard distributed regionally by the smoothed seismicity approach could be more appropriately assigned to faults, thereby increasing the hazard adjacent to faults and decreasing the hazard between and distant from faults.

Table 4
Probabilistic Estimates of Numbers of Earthquakes from
the Bakun Catalog in the Background

	Number	Fraction in Background	Annual Rate	30-Year Probability
Background as calculated	1.57	26%	0.010	25%
Background less Loma Prieta	0.57	10%	0.003	10%
Background plus two 1892 earthquakes	3.44	43%	0.021	46%
Background less Loma Prieta plus two 1892 earthquakes	2.44	31%	0.015	36%
Entire region (without 1892 earthquakes)	6		0.036	66%

Earthquakes with $M \geq 6.5$ that occurred in the period 1836–2000.

Table 5
Probabilistic Estimates of the Seismic Moment in the "Exponential Tail" for Fault Systems in the San Francisco Bay Area

Fault System	San Andreas	Hayward/ Rogers Creek	Calaveras	Concord/ Green Valley	San Gregoriao	Greenville	Mount Diablo
Threshold magnitude	6.65	6.19	5.56	5.95	6.65	5.95	6.4
Cumulative moment below threshold magnitude $(N - m)$	1.32E + 19	1.86E + 18	2.23E + 17	5.70E + 17	2.87E + 18	4.52E + 17	4.29E + 17
Annual moment rate below threshold magnitude $(N - m)$	8.77E + 16	1.24E + 16	1.48E + 15	3.80E + 15	1.91E + 16	3.01E + 15	2.86E + 15
WG99 estimate of total moment rate $(N - m/\text{yr})$	9.82E + 17	3.73E + 17	2.49E + 17	5.70E + 16	3.69E + 17	6.57E + 16	3.20E + 16
Percent of moment in exponential tail	9%	3%	1%	7%	5%	5%	9%

Real-Time Systems

It would be relatively straightforward to implement this approach in a real-time seismic system to give an immediate estimate of the fault segment with which an earthquake is associated, without, in the case of a large earthquake, waiting for aftershocks.

Conclusion

Bayesian inference provides a practical and feasible method for associating earthquakes with faults, taking into account location uncertainty and geologic data such as fault slip rates. As applied to the faults of the San Francisco Bay region, the resulting earthquake–fault associations are generally reasonable, both for historical earthquakes (from intensity data) and for recent earthquakes with instrumental locations. Exceptions tend to illuminate deviations from the simplistic view of the relationship between faults and earthquakes. The technique provides the ability to construct probability-based frequency–magnitude curves for specific faults and perform other applications potentially relevant to probabilistic seismic hazard analysis.

The techniques developed within this article are based on the idealized correspondence of earthquake locations with the surface traces of active faults as determined from geologic data, a correspondence that is often taken for granted. In fact there are quite significant disparities between these two sources of data, even in the San Francisco Bay region. While on a gross scale the earthquakes clearly associate with the major faults in the region, the details of this association are complex. Understanding the circumstances in which earthquakes locate directly on the fault surface, as contrasted with those in which the earthquakes locate near the main fault, but in fact on subsidiary faults typically with different orientations, will give important clues about the geology and physics of earthquakes.

References

Bakun, W. H. (1998). Modified Mercalli intensities for some recent California earthquakes and historic San Francisco Bay region earthquakes, *U.S. Geol. Surv. Open-File Rept. 98-584,* 175 pp.

Bakun, W. H. (1999). Seismic activity of the San Francisco Bay Region, *Bull. Seism. Soc. Am.* **89,** 764–784.

Bakun, W. H., and C. M. Wentworth (1997). Estimating earthquake location and magnitude from seismic intensity data, *Bull. Seism. Soc. Am.* **87,** 1502–1521.

Dietz, L. D., and W. L. Ellsworth (1990). The October 17, 1989, Loma Prieta, California, earthquake and its aftershocks: geometry of the sequence from high-resolution locations, *Geophys. Res. Lett.* **17,** 1417–1420.

Eberhart-Phillips, D., and A. J. Michael (1998). Seismotectonics of the Loma Prieta, California, region determined from three-dimensional V_p, V_p/V_s, and seismicity, *J. Geophys. Res.* **102,** 21,099–21,210.

Frankel, A., C. Mueller, T. Barnhard, D. Perkins, E. Leyendecker, N. Dickman, S. Hanson, and M. Hopper (1996). National seismic-hazard maps: documentation June 1996, *U.S. Geol. Surv. Open-File Rept. 96-532,* 110 pp.

Herrmann, R. A. (1977). Recurrence relations, *Earthquake Notes* **48,** 47–49.

Kennedy, J. B., and A. M. Neville (1986). *Basic Statistical Methods for Engineers and Scientists*, Harper and Row, New York, 613 pp.

Klein, F. W. (1978). Hypocenter location program HYPOINVERSE. I. Users guide to versions 1, 2, 3, and 4, *U.S. Geol. Surv. Open-File Rept. 78-694,* 114 pp.

Lahr, J. C. (1984). Description of the weighted regression and quality estimation used in the earthquake location program HYPOELLIPSE, *U.S. Geol. Surv. Open-File Rept. 84-766,* 38 pp.

McClellan, P. H., and E. A. Hay (1990). Triggered slip along the Calaveras fault during the magnitude 7.1 Loma Prieta, California, earthquake, *Geophys. Res. Lett.* **17,** 1227–1230.

Peterson, M. D., W. A. Bryant, C. H. Cramer, T. Cao, M. S. Reichle, A. D. Frankel, J. J. Lienkaemper, P. A. McCrory, and D. P. Schwartz (1996). Probabilistic seismic hazard map for the state of California, California Division of Mines and Geology Open-File Report 96-08, published jointly as *U.S. Geol. Surv. Open-File Rept. 96-706*, 59 pp.

Prentice, C. S., and D. P. Schwartz (1991). Reevaluation of 1906 surface faulting, geomorphic expression, and seismic hazard along the San Andreas fault in the southern Santa Cruz Mountains, *Bull. Seism. Soc. Am.* **81,** 1424–1479.

Reasenberg, P. A. (1985). Second-order moment of central California seismicity, 1969–1982, *J. Geophys. Res.* **90,** 5479–5495.

Reasenberg, P. A., and W. L. Ellsworth (1982). Aftershocks of the Coyote Lake, California, earthquake of August 9, 1979, *J. Geophys. Res.* **87,** 10,637–10,655.

Sieh, K. E. (1982). Slip along the San Andreas Fault associated with the earthquake, in *U.S. Geol. Surv. Prof. Pap. 1254*, 155–159.

Uhrhammer, R. A., J. Fink, and S. Ford (1999). San Francisco Bay region historical earthquake relocation project, *Seim. Res. Lett.* **70,** no. 2, 271–272.

von Winterfeldt, D., and W. Edwards (1986). *Decision Analysis and Behavioral Research*, Cambridge University Press, New York, 604 pp.

Wald, D. J., D. V. Helmberger, and T. H. Heaton (1991). Rupture model of the 1989 Loma Prieta earthquake from the inversion of strong-motion and broadband teleseismic data, *Bull. Seism. Soc. Am.* **81,** 1540–1572.

Walter, S. R., D. H. Oppenheimer, and R. I. Mandel (1998). Seismicity maps of the San Francisco and San Jose 1 × 2 degree quadrangles, California, for the period 1967–1993, U.S. Geol. Surv. Miscellaneous Investigations Map, I-2580, scale 1:250,000.

Watson, S. R., and D. M. Buede (1987). *Decision Synthesis: The Principles and Practice of Decision Analysis*, Cambridge U Press, New York, 299 pp.

Weichert, D. H. (1980). Estimation of the earthquake recurrence parameters for unequal observation periods for different magnitudes, *Bull. Seism. Soc. Am.* **70,** 1337–1346.

Wells, D. L., and K. J. Coppersmith (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement, *Bull. Seism. Soc. Am.* **84,** 974–1002.

Working Group on California Earthquake Probabilities (1999). Earthquake probabilities in the San Francisco Bay Region, 2000–2030: a summary of findings, *U.S. Geol. Surv. Open-File Rept. 99-517*, 60 pp.

Working Group on Northern California Earthquake Potential (1996). Database of potential sources for earthquakes larger than magnitude 6 in northern California, *U.S. Geol. Surv. Open-File Rept. 96-705*, 52 pp.

Youngs, R. R., and K. J. Coppersmith (1985). Implications of fault slip rates and earthquake recurrence models to probabilistic seismic hazard estimates, *Bull. Seism. Soc. Am.* **75,** 939–964.

Zoback, M. L., R. C. Jachens, and J. A. Olson (1999). Abrupt along-strike change in tectonic style: San Andreas fault zone, San Francisco Peninsula, *J. Geophys. Res.* **104,** 10,719–10,742.

Appendix A

Relation between Two- and Three-Dimensional Specifications of Standard Errors in Earthquake Locations

The relationship between two- and three-dimensional specifications of errors in earthquake locations has been discussed by Klein (1978) and Lahr (1984) for the earthquake location programs HYPOINVERSE and HYPOELLIPSE,

respectively. In view of the subtle differences in the approach of those two authors and to ensure clarity in the derivation of the probability density functions, the approach taken in this work is summarized here. The outputs from HYPOINVERSE include the lengths, azimuths, and dips of the three standard errors of the solution in the principal coordinates of the solution. These standard errors are determined as the square roots of the eigenvalues of the 3×3 covariance matrix of solution for the earthquake location (that is, with the origin time removed) (Klein, 1978). The earthquake focus has a 32% probability of lying within the ellipsoid so defined. The output of HYPOELLIPSE includes the same parameters, except that the lengths are scaled such that the probability of the earthquake focus lying within the ellipsoid so defined is 67% (Lahr, 1984). The three standard errors from HYPOINVERSE can be used directly to determined the probability density function (in the principal coordinate system) from equation (3).

If we are given the three standard errors in the principal coordinate system of the solution and want to determine the standard errors of the two horizontal coordinates, then we must consider the "shadow" of the error ellipsoid on the horizontal plane (Lahr, 1984). Following Lahr, but maintaining the lengths of the semimajor axes as the standard errors, σ_i, let

$$s_p = \begin{bmatrix} 1/\sigma_1^2 & 0 & 0 \\ 0 & 1/\sigma_2^2 & 0 \\ 0 & 0 & 1/\sigma_3^2 \end{bmatrix}$$

be a tensor in the coordinate system of the principal axes. The tensor can be rotated into a new coordinate system specified by a unit vector of direction cosines, $\hat{l}$, that is, $s = \hat{l}^{\mathrm{T}} s_p \hat{l}$. The radius of the error ellipsoid is $r = 1/\sqrt{s}$. Let s_{ij} be the components of s in the "geographic" coordinate system with z vertical. Following Lahr, the condition for the shadow ellipse in the horizontal plane is $\nabla(1/\sqrt{s}) \cdot \hat{k} = 0$, where is the vertical unit vector. Then the equation for the shadow ellipse can be written as $ax^2 + 2bxy + cy^2 = 1$, where $a = s_{11} - (s_{13}^2/s_{33})$, $b = s_{12} - (s_{13}s_{23}/s_{33})$, and $c = s_{22} - (s_{23}^2/s_{33})$. The semimajor axes of the error ellipse, that is the standard errors in the horizontal plane, are

$$s_1 = 1/\sqrt{(a \cos^2 \theta_0 + b \cos\theta_0 \sin\theta_0 + c \sin^2\theta_0)},$$

$$\begin{aligned} s_2 = 1/\surd(& a \cos^2(\theta_0 + \pi/2) + b \cos(\theta_0 + \pi/2) \\ & \sin(\theta_0 + \pi/2) + c \sin^2(\theta_0 + \pi/2)), \end{aligned}$$

and these principal axes are rotated an angle, $\theta_0 = 1/2 \tan^{-1}(b/a - c)$, relative to the geographic axes. These two standard errors can then be used in the two-dimensional version of equation (3).

Appendix B

Sensitivity Analysis

The sensitivity of the results was tested relative to input parameters of the prior probabilities for the faults, the prior probability for the background, and the standard deviation of location distance from a fault using the older earthquakes located from intensity data. For the purposes of these tests the 1868 earthquake was not constrained to be associated with the southern Hayward fault. Results are shown for five selected fault segments: Santa Cruz Mountains–San Andreas (SCZ), peninsula–San Andreas (PN), southern Hayward (SH), southern Calaveras (SC), central Calaveras (CC), and for the background (BKGD).

Prior Probability for the Faults

In addition to the segment rate fault priors shown in Table 1, four other sets of fault priors were constructed. Three sets of priors were scaled according to moment rate, the seismogenic area of the fault segment, and slip rate. There are no particularly strong physical arguments to favor the moment rate, area, or slip rate priors over the segment priors, as discussed in the text. Indeed these priors are intentionally perverse and constructed to try to determine the sensitivity of the results to a range of inputs. The final set of priors were equal priors. These sets of priors are listed in Table B1. As shown in Figure B1, the results for the background are almost unaffected by the selection among this set

of priors. The largest variations are seen for the smaller fault segments, especially the southern Calaveras, which tend to have fewer earthquakes associated with them with the moment rate and area priors. Correspondingly the larger fault segments tend to have more earthquakes associated with them. In general the relative similarity of the results from the most physically based segment rate priors (i.e., those adopted in the text) and the unbiased equal priors give confidence that the prior fault probability is not a dominant factor in the results.

Prior Probability for the Background

The prior probability for the background can be estimated from the historical large earthquakes by making best guesses about what fault is responsible for each, if possible, then calculating the fraction of unassociated earthquakes. Based on this line of reasoning, the initial guess of the prior probability for the background was 20% for the 18-fault-segment WG99 model. To test the sensitivity of the results to this parameter, associations for the older, larger earthquakes were performed using prior probabilities for the background of 0.01, 0.1, 0.2, 0.3, 0.4, and 0.5. The calculated segment priors are shown in Table B2; results are shown in Figure B2. The frequencies for the fault segments are modestly affected by the selection of the background probability. Note that, as expected, the lower the background prior, the higher the segment frequencies. Although within the range of reasonable values, say 0.1–0.2, the variation

Table B1
Alternative Prior Fault Segment Probabilities

WG99 Fault or Segment	Abbreviation	Mean M_w	Mean Frequency (no./yr)	Characteristic	Moment Rate	Area	Slip Rate	Equal
San Andreas, Santa Cruz Mtns.	SCZ	7.03	4.5770E-03	0.0428	0.0844	0.0575	0.0752	0.0471
San Andreas, Peninsula	PN	7.15	4.4900E-03	0.0420	0.1217	0.0683	0.0913	0.0471
San Andreas, North Coast South	NCS	7.45	4.6440E-03	0.0435	0.3250	0.1292	0.1289	0.0471
San Andreas, North Coast North	NCN*	7.29	4.6190E-03	0.0000	0.0000	0.0000	0.0000	0.0000
Hayward/RC, Southern Hayward	SH	6.67	5.2560E-03	0.0492	0.0291	0.0386	0.0483	0.0471
Hayward/RC, Northern Hayward	NH	6.49	5.6290E-03	0.0527	0.0147	0.0260	0.0483	0.0471
Hayward/RC, Rogers Creek	RC	6.98	4.2360E-03	0.0396	0.0441	0.0467	0.0483	0.0471
Calaveras, Southern	SC	5.80	1.3990E-02	0.1309	0.0081	0.0129	0.0805	0.0471
Calaveras, Central	CC	6.23	1.7783E-02	0.1664	0.0252	0.0401	0.0805	0.0471
Calaveras, Northern	NC	6.78	4.8550E-03	0.0454	0.0205	0.0362	0.0322	0.0471
Concord/GV, Concord	CON	6.25	3.8100E-03	0.0357	0.0041	0.0198	0.0215	0.0471
Concord/GV, Southern Green Valley	SGV	6.24	3.8660E-03	0.0362	0.0050	0.0190	0.0268	0.0471
Concord/GV, Northern Green Valley	NGV	6.02	4.1650E-03	0.0390	0.0032	0.0121	0.0268	0.0471
San Gregorio, Southern	SGS	6.96	1.6030E-03	0.0150	0.0154	0.0489	0.0161	0.0471
San Gregorio, Northern	SGN	7.23	2.2450E-03	0.0210	0.0643	0.0876	0.0376	0.0471
Greenville, Southern	SGVY	6.60	1.1000E-03	0.0103	0.0142	0.0677	0.0107	0.0471
Greenville, Northern	NGVY	6.66	1.1080E-03	0.0104	0.0142	0.0677	0.0107	0.0471
Mount Diablo	MTD	6.65	2.1350E-03	0.0200	0.0069	0.0219	0.0161	0.0471
Background	BKGD			0.2000	0.2000	0.2000	0.2000	0.2000
Total			0.08549	1.0000	1.0000	1.0000	1.0000	1.0000

*The North Coast segment is outside the area considered.

Table B2
Alternative Prior Fault Segment Probabilities Assuming Different Prior Background Probabilities

WG99 Fault or Segment	Abbreviation	Mean M_w	Mean Frequency (no./yr)	Prior Probabilities Segment Rate Varying Background					
San Andreas, Santa Cruz Mtns.	SCZ	7.03	4.5770E-03	0.0530	0.0482	0.0428	0.0375	0.0321	0.0268
San Andreas, Peninsula	PN	7.15	4.4900E-03	0.0520	0.0473	0.0420	0.0368	0.0315	0.0263
San Andreas, North Coast South	NCS	7.45	4.6440E-03	0.0538	0.0489	0.0435	0.0380	0.0326	0.0272
San Andreas, North Coast North	NCN*	7.29	4.6190E-03	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Hayward/RC, Southern Hayward	SH	6.67	5.2560E-03	0.0609	0.0553	0.0492	0.0430	0.0369	0.0307
Hayward/RC, Northern Hayward	NH	6.49	5.6290E-03	0.0652	0.0593	0.0527	0.0461	0.0395	0.0329
Hayward/RC, Rogers Creek	RC	6.98	4.2360E-03	0.0491	0.0446	0.0396	0.0347	0.0297	0.0248
Calaveras, Southern	SC	5.80	1.3990E-02	0.1620	0.1473	0.1309	0.1145	0.0982	0.0818
Calaveras, Central	CC	6.23	1.7783E-02	0.2059	0.1872	0.1664	0.1456	0.1248	0.1040
Calaveras, Northern	NC	6.78	4.8550E-03	0.0562	0.0511	0.0454	0.0398	0.0341	0.0284
Concord/GV, Concord	CON	6.25	3.8100E-03	0.0441	0.0401	0.0357	0.0312	0.0267	0.0223
Concord/GV, Southern Green Valley	SGV	6.24	3.8660E-03	0.0448	0.0407	0.0362	0.0317	0.0271	0.0226
Concord/GV, Northern Green Valley	NGV	6.02	4.1650E-03	0.0482	0.0438	0.0390	0.0341	0.0292	0.0244
San Gregorio, Southern	SGS	6.96	1.6030E-03	0.0186	0.0169	0.0150	0.0131	0.0113	0.0094
San Gregorio, Northern	SGN	7.23	2.2450E-03	0.0260	0.0236	0.0210	0.0184	0.0158	0.0131
Greenville, Southern	SGVY	6.60	1.1000E-03	0.0127	0.0116	0.0103	0.0090	0.0077	0.0064
Greenville, Northern	NGVY	6.66	1.1080E-03	0.0128	0.0117	0.0104	0.0091	0.0078	0.0065
Mount Diablo	MTD	6.65	2.1350E-03	0.0247	0.0225	0.0200	0.0175	0.0150	0.0125
Background	BKGD			0.0	0.1	0.2	0.3	0.4	0.5
Total			0.08549	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

*The North Coast segment is outside the area considered.

is not particularly significant. The frequency–magnitude curves for the background are, not unexpectedly, more sensitive the prior background probability. When the obviously extreme value of 0.01 is excluded, the variation is considerably less. Admittedly the selection of the preferred value of 0.2 is somewhat arbitrary, and the uncertainty in the frequency–magnitude curves for the background resulting from the uncertainty in this prior must be accepted.

Standard Deviation of Location Distance from the Fault

As discussed in the text, the standard deviation of location distance from the fault (the parameter σ_f in equation 12) can be estimated from data for recent earthquakes. In contrast, the estimate for the preinstrumental data must be made relatively *ad hoc*. Frequency–magnitude curves calculated for fault standard deviations of 1, 5, 10, 15, and 20 km are shown in Figure B3. Again when the obviously ex-

treme value of 1 km is rejected, there is considerably less variation in the results. A value of 10 km is adopted, and the uncertainties resulting from the uncertainty in this choice must be accepted.

U.S. Geological Survey
Central Region Geologic Hazards Team
P.O. Box 25046
Denver, Colorado 80225
rwesson@usgs.gov
 (R.L.W., D.M.P.)

U.S. Geological Survey
Earthquake Hazards Team
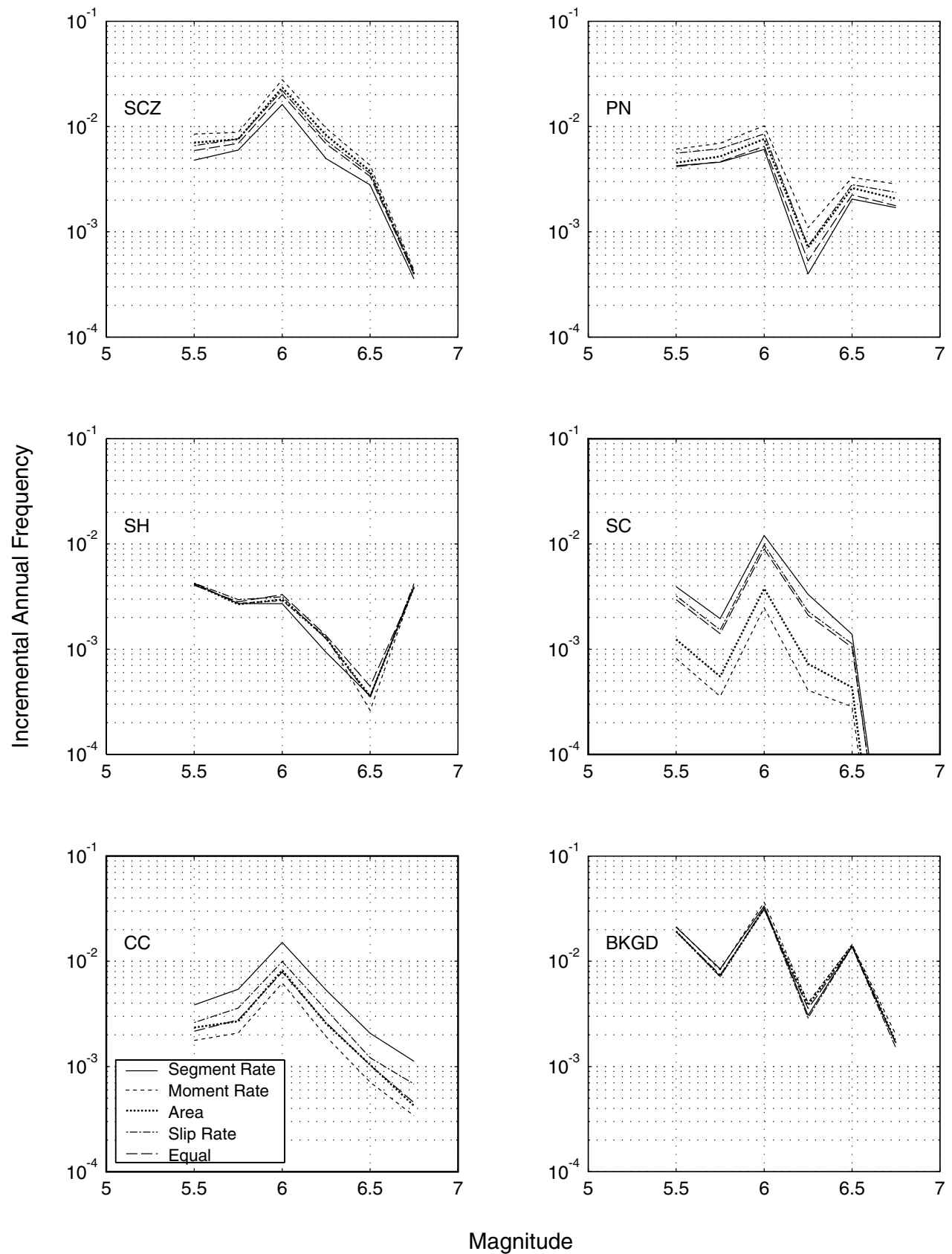345 Middlefield Road
Menlo Park, California 94025
 (W.H.B.)

Figure B1. Effect of assumed fault prior probability on the frequency–magnitude relation for selected fault segments and the background determined from the earthquakes located based on intensity observations.
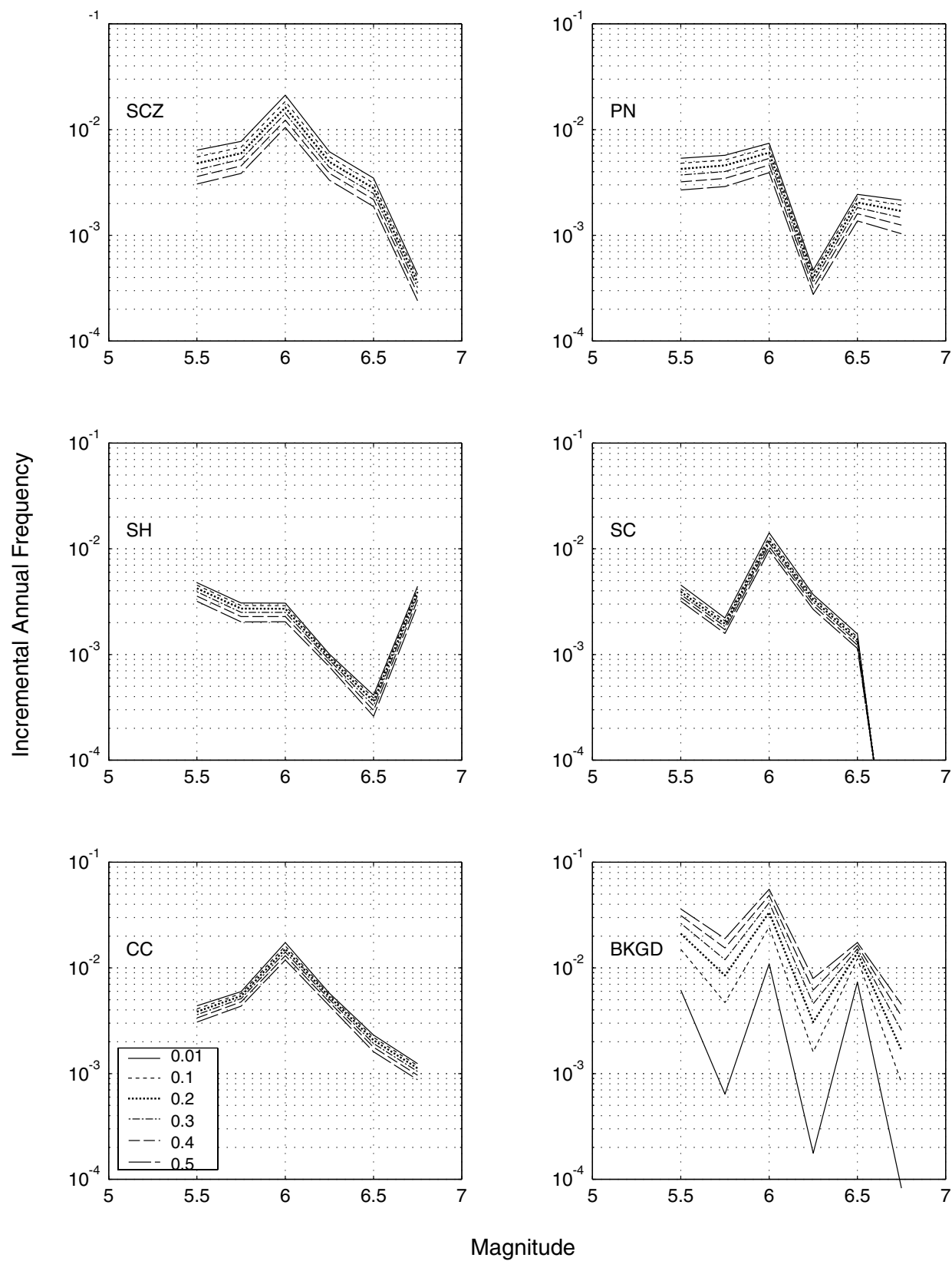
Figure B2.   Effect of assumed background prior probability on the frequency–magnitude relation for selected fault segments and the background determined from the earthquakes located based on intensity observations.
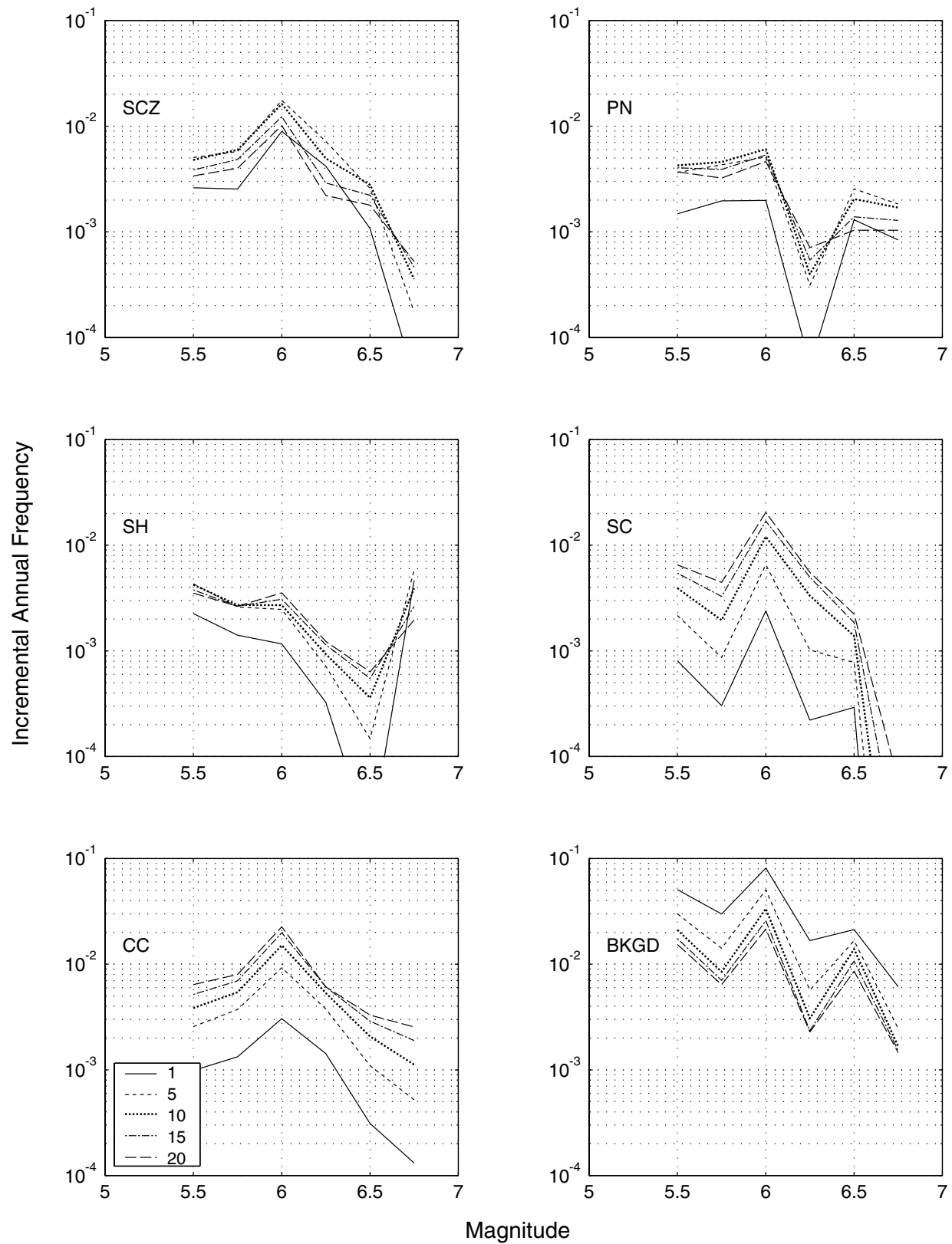
Figure B3.   Effect of assumed "fault standard deviation" on the frequency–magnitude relation for selected fault segments and the background determined from the earthquakes located based on intensity observations.