Appendix I: Calculating California Seismicity Rates

By Karen R. Felzer[1]

2008

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

California Department of Conservation
California Geological Survey

[1]U.S. Geological Survey, Pasadena, California

U.S. Department of the Interior
DIRK KEMPTHORNE, Secretary

U.S. Geological Survey
Mark D. Myers, Director

State of California
ARNOLD SCHWARZENEGGER, Governor

The Resources Agency
MIKE CHRISMAN, Secretary for Resources

Department of Conservation
Bridgett Luther, Director

California Geological Survey
John G. Parrish, Ph.D., State Geologist

U.S. Geological Survey, Reston, Virginia 2008

Contents

1 Introduction

Empirically the rate of earthquakes ≥ magnitude M is well fit by the Gutenberg-Richter relationship,

$$\log N = a - bM \qquad (1)$$

where N is the number of earthquakes ≥ M over a given time period, a is the number of $M \geq 0$ earthquakes over the same period, and b is a parameter that determines the ratio of larger to smaller earthquakes (Ishimoto and Iida 1939; Gutenberg and Richter 1944). Thus to characterize the seismicity rate, N, and risk in a given region we need to solve for the values of a and b. Here we are concerned with solving for the long term average values of these parameters for the state of California. My primary data source is a catalog of 1850-2006 $M \geq 4.0$ seismicity compiled with Tianqing Cao (Appendix H). Because earthquakes outside of the state can influence California I consider both earthquakes within the state and within 100 km of the state border (Figure 1).
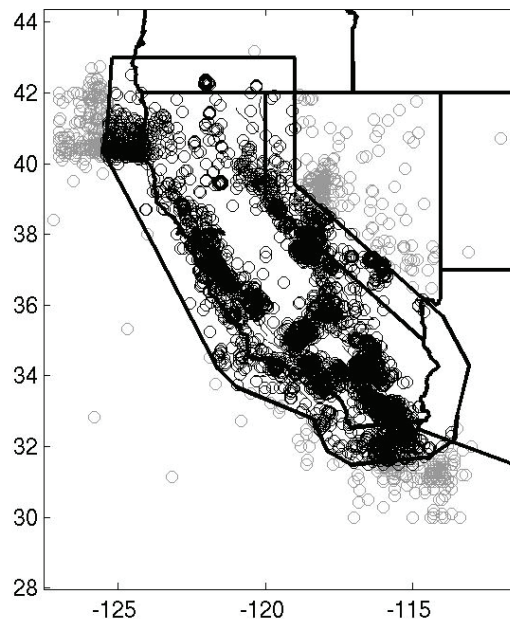
Figure 1: Map of the data used in this study, $M \geq 4$ earthquakes from 1850-2006. The catalog given in Appendix H is comprised of all the earthquakes plotted here (in black and gray), but only the earthquakes within the plotted polygon, which we refer to as the California region, are used for the rate calculation here.

The *a* and *b* values found here are calculated using methods employed by the 1996 and 2002 National Hazard Maps, with several revisions. These revisions include making corrections for magnitude error and rounding before calculating *a* values, using only modern instrumental data to calculate *b* value, and using a new comprehensive and spatially variable assessment of the magnitude completeness threshold as a function of time. We also calculate the seismicity rate in several different ways to account for the fact that the seismicity rate may change with time (for example, the higher seismicity rates in the San Francisco Bay Area before 1927 than after), and perform simulations to evaluate the accuracy with which the seismicity rate averaged over the last 156 years represents the true long term seismicity rate. Finally, the National Hazard Maps have traditionally only used the historical earthquake solutions of Toppozada, most recently compiled in Toppozada et al. (2002). We do our calculations both with the Toppozada solutions and with 84 of the Toppozada solutions substituted with historical earthquake solutions of Bakun (Bakun 1999; Bakun 2000; Bakun 2006). We find that this substitution creates an insignificant increase in the statewide seismicity rate of 0.6%, although it may produce larger differences on a regional level.

My final result, using an averaged Weichert method in which I allow the rate in the historic catalog (pre-1932) to be higher than the instrumental catalog rate, and in which I correct the rate upwards to account for the possibility of earthquakes as large as M 8.3 and associated higher seismicity rates in California over the long term, gives 7.5, -3.94, +3.0 $\geq$ 5.0 earthquakes/year for the full California catalog (98% confidence) and 4.17, -1.95, +1.67 $\geq$ 5.0 earthquakes/year for the declustered California catalog. The high errors result from high completeness magnitudes (and thus sparse useable data) in the historical part of the catalog. Rates solved for by using a straight Weichert method and by using direct catalog counts and without assuming M_{max} = 8.3 are also discussed and given in the text, tables, and figures.

2 Correcting for Magnitude Rounding

Most magnitudes in our catalog are rounded to the nearest 0.1. However a substantial proportion of the catalog in the early to mid 1900s is rounded to the nearest 0.5, and other parts are rounded to the nearest 0.01.

The maximum likelihood (MLE) solution for *a*, a robust method that has been used by the National Hazard Maps, is based on the total number of earthquakes $M \geq M_C$, where M_C is the magnitude above which the catalog is complete. If all earthquake magnitudes are rounded to the nearest 0.5 (magnitudes reported as 4.0, 4.5, 5.0, etc), then what is measured in the catalog as $M \geq M_C$ is actually $M \geq M_C$ - 0.25. If *b* =1.0, this causes the calculated rate of $M \geq M_C$ earthquakes to be $10^{0.25}$ = 1.8 times higher than the real rate.

When rounding is uniform throughout the catalog I can correct this overestimate by multiplying the number of $M \geq M_C$ earthquakes measured by $10^{-bRound/2}$ where *Round* = 0.5 for rounding to the nearest 0.5, *Round* = 0.1 for rounding to the nearest 0.1, and so forth, and *b* is the *b* parameter in the Gutenberg-Richter relationship. Since rounding is not uniform in the California catalog, however, we need an alternate solution. I make use of the distribution of real magnitudes corresponding to each rounded magnitude. For each earthquake reported as M 4.5 with *Round* = 0.5, for example, we know that it's true

magnitude lies between 4.25 and 4.75 with a truncated Gutenberg-Richter magnitude distribution between those values. Thus using the simple Monte Carlo routine given below (written in matlab) each rounded magnitude can be replaced with a magnitude from the real distribution.
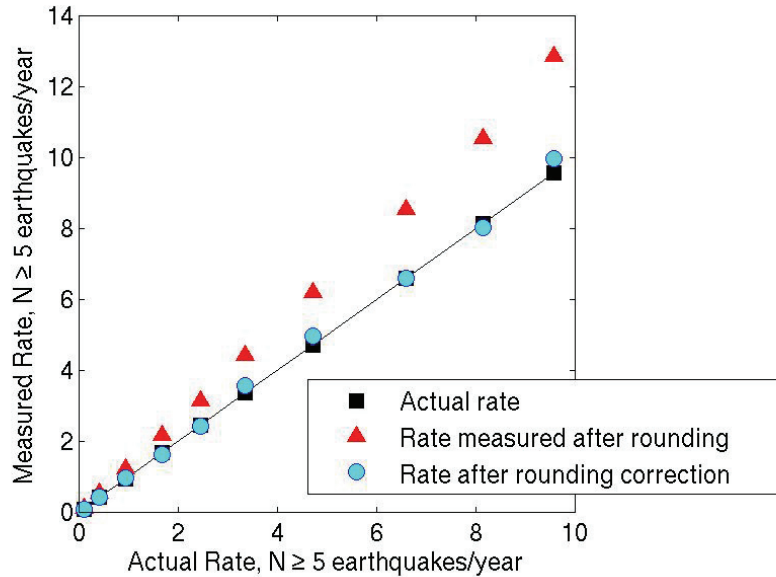
Figure 2: Correction for overestimation of seismicity rates due to rounding errors. We corrected for rounding by replacing the rounded magnitudes with resampled magnitudes from the appropriate distribution with a Monte Carlo routine (see text). To test our method we generated simulated catalogs, measured the real seismicity rates (black squares), rounded the magnitudes according to the amount that rounding occurs in the real 1850-2006 catalog, measured the seismicity rates in the rounded catalog (red triangles), and compared with the remeasured seismicity rates after applying our rounding correction (cyan circles). The coincidence of the black squares and cyan circles indicates that rates measured from the corrected catalogs agree well with the real rates.

Documentation for the Matlab Routine

1. round is a vector which contains the amount by which every individual magnitude in the catalog is rounded. For a magnitude that is rounded to the nearest 0.5 (reported as 4.0, 4.5, etc.) round = 0.5; for magnitudes reported to the closest 0.1 round = 0.1, etc.

2. mcat is a vector containing the magnitude of each earthquake in the catalog. The entries in mcat and round need to correspond, such that the first value in round is the amount by which the first value in mcat is rounded, etc.

3. rand is an internal matlab function which generates uniform random numbers between 0 and 1.

4. magsN is the list of new magnitudes.

Code Starts Here

```
rcat = round/2;

randR1 = rand(length(mcat),1);

randRb = randR1.×(10.^ -(b×(mcat+rcat)) - 10.^ -(b×(mcat-rcat))) + 10.^-(b×(mcat-
rcat));

magsN = -(1/b) log10(randRb);
```

I repeat this routine 500 times, calculating an *a* value for the whole catalog each time all of the magnitudes are replaced, and then average the 500 *a* values to get the the best corrected *a* value for the data set. Performing this solution on a simulated rounded catalog demonstrates that it accurately recovers the real seismicity rate (Figure 2).

3 Correcting for Magnitude Error

The *a* value can also be overestimated because of magnitude errors (Tinti and Mulargia 1985; Rhoades 1996). This is because Gaussian magnitude error is symmetric while the distribution of magnitudes is asymmetric. For example, say that we have an earthquake population of 2 M 5.2 earthquakes and 6 M 4.8 earthquakes and that our *a* value is based on the number of $M \geq 5$ earthquakes measured. Now we apply Gaussian magnitude error. Because Gaussian error is symmetric, the magnitudes have equal probabilities of being reported higher or lower than they actually are. Thus each individual M 5.2 earthquake has the same probability of being reported as M 4.8 as each M 4.8 earthquake has of being reported as M 5.2. There are 3 times as many M 4.8 as M 5.2 earthquakes, however. Thus, on average, for every 1 M 5.2 earthquake that is reported as an M 4.8, 3 M 4.8 earthquakes will be reported as M 5.2, resulting in a net increase of apparent $M \geq 5$ earthquakes.

As with rounding, if the amount of magnitude error is uniform throughout the catalog there is an easy fix, from Tinti and Mulargia (1985). If a_{GR} is the measured value of *a* then the true value of *a* is given by,

$$a = a_{GR} - \gamma^2 \log_{10}(e) \tag{2}$$

where

$$\gamma^2 = \frac{\beta^2 \sigma^2}{2} \tag{3}$$

where $\beta = b\ln(10)$ and $\sigma =$ the standard deviation of the magnitude error. Note from this equation that the effects of magnitude error may be quite large. An error with a standard deviation of 0.7, for example, which may apply to some historical earthquakes (Kagan et al. 2006) will cause an overesimate of the seismicity rate by more than a factor of 3.

Magnitude error is not uniform throughout the California catalog, however. In particular, magnitude errors tend to decrease in the more recent part of the catalog. To work with

variable magnitude error we note that the *a* value of a catalog may be reduced by reducing the *a* value for the entire catalog in a single step, as is done by Tinti and Mulargia (1985), or by subtracting some amount Δ*M* from each magnitude and then calculating *a* from the new magnitudes. For variable magnitude error the latter approach is advantageous because it allows an individual amount to be subtracted from each magnitude that is proportional to the earthquake's magnitude error. To find the amount, Δ*M*, that should be subtracted from the catalog magnitude the solution for *a* from Equation 2 can be substituted into the Gutenberg-Richter relationship and rearranged to get,

$$\Delta M = \frac{b^2 \sigma^2}{2 \log_{10}(e)}. \tag{4}$$

Correct *a* values can then be calculated from the catalog after Δ*M* is subtracted from each magnitude. The accuracy of recovering the correct *a* value with this method is demonstrated with simulated catalogs in Figure 3 and Figure 4.
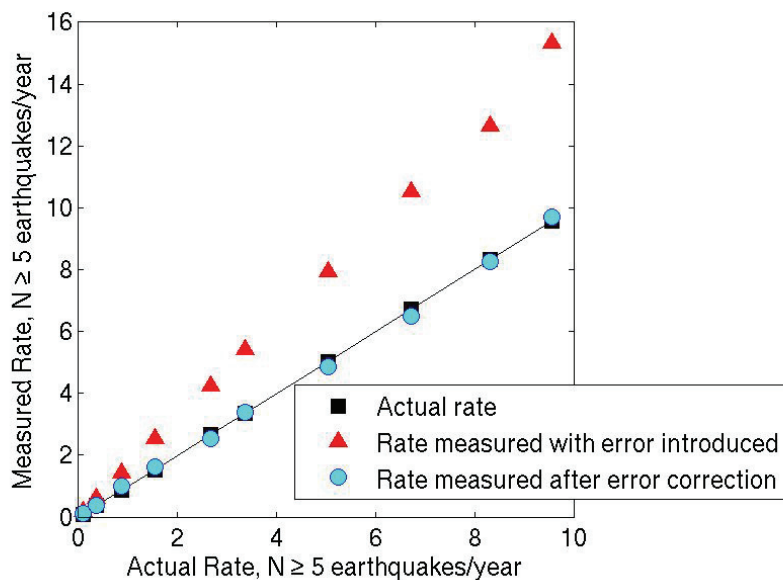
Figure 3: Correction for overestimation of seismicity rates due to magnitude errors. Magnitude errors are corrected by subtracting a correction from each magnitude that is proportional to the standard deviation of the error based on the work of Tinti and Mulargia (1985). To test our method we generated simulated catalogs, measured the real seismicity rates (black squares), introduced random Gaussian error to the magnitudes to the same extent that it occurs in the real 1850-2006 catalog, measured the seismicity rates in the catalog with errors introduced (red triangles), and compared with the remeasured seismicity rates after applying magnitude error correction (cyan circles). The coincidence of the black squares and cyan circles indicates that rates measured from the corrected catalogs agree well with the real rates.

5
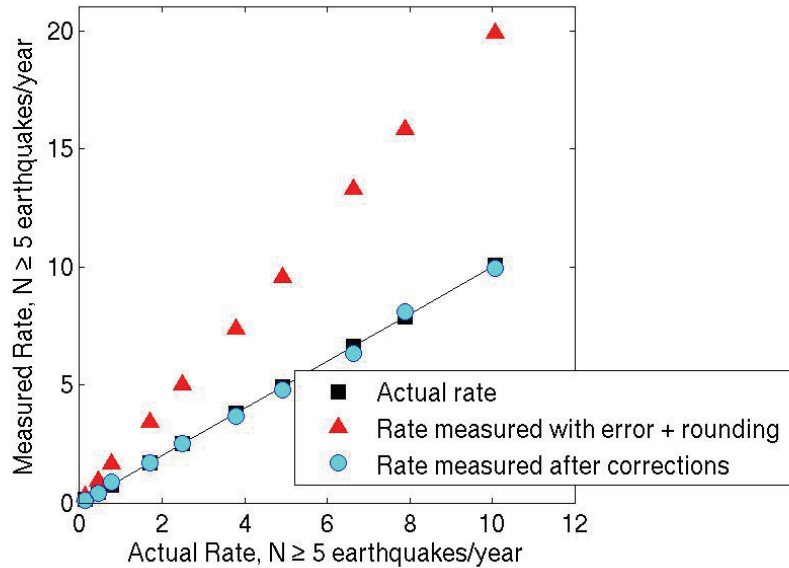
Figure 4: Test for our ability to correct for a combination of magnitude rounding and error. We generated simulated catalogs, measured the real seismicity rates (black squares), introduced random Gaussian error and then magnitude rounding to the same extent that it occurs in the real 1850-2006 catalog, measured seismicity rates in the catalog with the errors and rounding introduced (red triangles), and compared with the remeasured seismicity rates after applying our correction methods (cyan circles). The coincidence of the black squares and cyan circles indicates that rates measured from the corrected catalogs agree well with the real rates. The height of the red triangles above the black squares also indicates how significant the problem of combined magnitude error and rounding can be.

One problem with applying the above routine is that magnitude errors are not routinely provided for Southern California data. I solve for the errors by bootstrapping the station magnitudes used to calculate M_L (M_L is solved for by the data center by taking the median of the station magnitudes). The station magnitudes themselves, however, are not routinely saved in the data base after calculation. So I recalculate the station magnitudes for all Southern California 1932-2005 $M \geq 4$ earthquakes from the S wave amplitudes recorded at each station, which are available from the Southern California Earthquake Data Center.

For Northern California, the hypoinverse phase catalog format provides magnitude errors for some earthquakes occurring after 1970. For historical earthquakes which were solved for by both Toppozada and Bakun I used the errors given in Bakun (1999), Bakun (2000) or Bakun (2006), depending on which paper the earthquake was listed in, if the magnitude given by Toppozada was within the error range given by Bakun. If the earthquake was solved for by a different author, who provided a magnitude error range, this error was used. Otherwise a standard error of 0.333 was generally assigned to pre-1932 earthquakes (see Appendix H for further details). All Harvard CMT magnitudes were assigned a standard error of 0.09 following the recommendations of Kagan et al.

(2006). For earthquakes for which no other information was available, I assigned a standard error of 0.222 to earthquakes occurring from 1932-1972, and 0.111 to earthquakes occurring after 1972. These errors are comparable to other errors calculated for the same time period. The three significant digits for these errors make them easier to identify as assigned rather than calculated values.

4 Magnitude Completeness Thresholds

One of the greatest difficulties in calculating accurate seismicity rates from earthquake catalogs is that all catalogs are incomplete. Since catalog incompleteness is a magnitude dependent problem, it is usually dealt with by determining a magnitude completeness threshold or the magnitude, M_C, above which nearly all earthquakes are listed in the catalog. Only earthquakes larger than this completeness threshold are used for seismicity rate calculations.

There are many different methods that have been used to determine the magnitude completeness threshold, with varying success (see Woessner and Wiemer (2005)). Many methods depend on testing the agreement of the data with the Gutenberg-Richter magnitude frequency distribution. These methods often underestimate the true threshold, however, especially if the b value is not fixed, and the methods become particularly problematic if the catalog has significant magnitude and rounding errors. Thus I prefer the more objective and comprehensive method of Schorlemmer et al. (2006), who bases completeness thresholds in Southern California on observations of how frequently individual seismic stations detect earthquakes of different magnitudes and distances. From this information Schorlemmer et al. (2006) calculated a separate completeness magnitude for each point in space, based on its distance from surrounding seismic stations. The method of Schorlemmer et al. (2006) is entirely empirical and computationally intensive. Here I simplify the Schorlemmer et al. (2006) inversion somewhat, but I still consider the recording history of each instrument, tabulating for which earthquakes the station was used in the solution (e.g. for which earthquakes the station appears in the phase file) and for which earthquakes it was not. A grid of points is then made, spaced at 0.25 by 0.25 degrees over the state. The completeness magnitude at each point is set as the minimum magnitude that would produce an amplitude above the completeness amplitudes of at at least 4 stations. More details on this are given below. For the historic part of the catalog, before the seismic network was in operation, the seismic stations are replaced with cities and historic newspaper locations. More details on this are also given below.

Since the number of historic cities and number and location of seismic stations change with time, separate sets of completeness intervals are calculated at 5 year intervals. In addition, after completeness thresholds are calculated at each point the points are grouped into regions of similar completenesses and a summary completeness magnitude is assigned to each region. These regions are then used to calculate the seismicity rates. Further details on the method are given below.

Also, it is important to emphasize that although some of the completeness thresholds found are smaller than M 4.0, only $M \geq 4.0$ earthquakes are given in the catalog in Appendix H and actually used for the final rate calculations, such that completeness

thresholds smaller than M 4.0 are essentially 4.0 for our purposes. One reason for using only $M \geq 4.0$ earthquakes is that as earthquakes become smaller the M_L scale diverges from the M_W scale on which the large earthquakes are measured. Furthermore, many of the smaller earthquakes are not even measured in M_L but in M_d, M_c, or M_h, which may diverge even further from M_W.

4.1 Determining Completeness with Time at Points in Space

4.1.1 Determining Historical Magnitude Completeness Thresholds

As noted above I use the locations of cities and newspapers in place of seismic stations to determine catalog completeness for historical earthquakes. I use a total of 102 locations, which includes all cities in California that were incorporated before 1900, plus cities in Nevada and Arizona that are near the California border and a few unincorporated locations in California with long standing, continuous populations and a published newspaper. For each location I determine the first year in which there was significant population and the year in which continuous newspaper publishing commenced. Here continuous publication is defined as lasting without a break longer than one or two years past the year 1900. The start date of continuous newspaper coverage was determined from the newspaper publishing data of the California Newspaper Project, housed at the University of California at Riverside (http://cnp.ucr.edu/). Often the continuous coverage was provided not by a single paper but by multiple papers that either fully or nearly overlapped in time. Papers that were published less frequently than weekly or for which no copies currently exist were not considered. The list of cities, newspapers, and years are given in Table 1.

Table 1: Cities used to calculate historical earthquake completeness magnitudes in California.

City	Pop[a]	News[b]	Newspaper Source	Lat	Lon
Alameda	1853	1877	The Alameda Argus	37.765	-122.241
Alhambra	1874	1913	The Alhambra News	34.095	-118.126
Anaheim	1857	1875	Anaheim Gazette	33.835	-117.914
Antioch	1850	1870	The Antioch Ledger	38.005	-121.805
Arcata	1850	1886	Arcata Union	40.867	-124.082
Auburn	1848	1855	The Placer Herald	38.897	-121.076
Bakersfield	1869	1884	Daily Evening Gazette	35.373	-119.018
Barstow	1880	1910	Barstow Printer	36.815	-119.969
Berkeley	1873	1873	(UC Berkeley)	37.872	-122.272
Big Pine	1908	1908	Big Pine Herald	37.165	-118.289
Bishop	1862	1885	Inyo Register	37.364	-118.394
Bullhead City	1860	~1960	The Bullhead City Bee	35.148	-114.568
Calistoga	1862	1871	Calistoga Tribune	38.579	-122.579
Chico	1860	1900	Chico Weekly Enterprise	39.729	-121.836
Cloverdale	1872	1886	Cloverdale Weekly Revielle	38.806	-123.016
Coalinga	1889	1917	Coalinga Daily Record	36.140	-120.359

[a] The year in which the city was first populated.

[b] The year in which continuous newspaper coverage started.

City	Pop[a]	News[b]	Newspaper Source	Lat	Lon
Colusa	1862	1862	The Weekly Colusa Sun	39.214	-122.008
Colton	1887	1912	Colton Daily Courier	34.074	-117.751
Corona	1896	1896	The Corona Courier	33.875	-117.566
Coronado	1886	1912	The Coronado Strand	32.686	-117.182
Crescent City	1854	1906	Crescent City News	41.756	-124.201
Eureka	1850	1871	Daily Evening Signal	40.802	-124.163
Escondido	1888	1909	The Times-Advocate	33.119	-117.086
Etna	1878	1897	The Scotts Valley Advance	41.457	-122.894
Ferndale	1878	1878	Ferndale Enterprise	40.576	-124.263
Fresno	1872	1875	Expositor	36.748	-119.771
Ft. Bragg	1857	1889	The Ft. Bragg Advocate	39.446	-123.804
Gilroy	1850	1925	The Gilroy Evening Dispatch	37.000	-121.570
Healdsburg	1854	1878	Healdsburg Enterprise	38.611	-122.868
Hemet	1887	1893	The Hemet News	33.748	-116.971
Hollister	1858	1891	Freelance	36.853	-121.401
Independence	1862	1870	The Inyo Independent	36.803	-118.199
Julian	1869	1892	The Sentinel	33.079	-116.601
Lake Elsinore	1883	1890	The Elsinore Press	33.668	-117.326
Las Vegas	1909	1909	Clark County Review	36.175	-115.160
Lincoln	1859	1913	The News Messenger	38.892	-121.292
Livermore	1869	1891	The Livermore Echo	37.682	-121.767
Los Angeles	1781	1851	Los Angeles Star	34.052	-118.243
Los Gatos	1855	1881	The Los Gatos Weekly News	37.227	-121.974
Lompoc	1888	1919	The Lompoc Review	34.639	-120.457
Lone Pine	1870	1924	The Mt. Whitney Observer	36.606	-118.062
Long Beach	1888	1900	Long Beach Evening Tribune	33.767	-118.188
Madera	1877	1901	Madera Daily Tribune	36.961	-120.060
Marysville	1850	1860	The Marysville Daily Appeal	39.146	-121.590
Martinez	1849	1860	The Contra Costa Gazette	38.019	-122.133
Merced	1880	1880	The Merced Star	37.302	-120.482
Modesto	1870	1872	The Stanislaus County Weekly News	37.639	-120.996
Monrovia	1887	1937	The Monrovia Daily News Post	34.148	-117.998
Monterey	1770	1864	Monterey Weekly Gazette	36.600	-121.894
Morgan Hill	1899	1899	The Times	37.131	-121.653
Napa	1849	1853	Napa Register	38.297	-122.284
Oakland	1852	1867	Oakland Daily News	37.804	-122.270
Oceanside	1888	1892	Oceanside Blade	33.196	-117.379
Ontario	1882	1885	Ontario Record	34.063	-117.650
Oxnard	1898	1901	The Oxnard Courier	34.198	-119.176
Pacific Grove	1890	1890	Pacific Grove Review	36.618	-121.916
Palo Alto	1855	1891	(Founding of Stanford University)	37.442	-122.142
Pasadena	1874	1890	The Pasadena Daily Evening Star	34.148	-118.144
Paso Robles	1886	1895	The River News	38.156	-121.690
Petaluma	1849	1856	The Sonoma County Journal	38.233	-122.630
Phoenix	1868	1886	Daily Phoenix Herald	33.448	-112.073

[a] The year in which the city was first populated.

[b] The year in which continuous newspaper coverage started.

| City	Pop[a]	News[b]	Newspaper Source	Lat	Lon
Pomona	1888	1898	Daily Progress	34.055	-117.751
Quartzsite	1867	~1960	The Quartzsite Times	33.664	-114.229
Red Bluff	1865	1865	Red Bluff Independent	40.179	-122.235
Redding	1874	1891	The Daily Free Press	40.587	-122.391
Redwood City	1856	1859	San Mateo County Gazette	37.485	-122.235
Reno	1868	1868	The Reno Crescent	39.530	-119.813
Richmond	1905	1910	The Richmond Daily Independent	37.936	-122.347
Rio Vista	1862	1895	The River News	38.156	-121.690
Riverside	1873	1891	Riverside Daily Enterprise	33.953	-117.395
Rocklin	1855	1855	Rocklin Placer Herald	38.748	-121.235
Sacramento	1849	1857	The Daily Bee	38.582	-121.493
Salinas	1869	1869	The Salinas Standard	36.678	-121.654
San Bernardino	1851	1887	The Daily Courier	34.108	-117.289
San Diego	1850	1851	San Diego Herald	32.715	-117.156
San Francisco	1848	1850	The Alta California	37.775	-122.418
San Jacinto	1870	1889	San Jacinto Register	33.784	-116.958
San Jose	1850	1866	Evening News	37.299	-121.894
San Rafael	1861	1861	The Marin County Journal	39.974	-122.530
San Leandro	1855	1856	Alameda County Gazette	37.725	-122.155
San Luis Obispo	1856	1901	(Date of founding of Polytechnic University)	35.283	-120.659
Santa Ana	1869	1899	The Santa Ana Bulletin	33.746	-117.867
Santa Barbara	1782	1875	The Daily News	34.421	-119.697
Santa Cruz	1866	1884	The Santa Cruz Daily Sentinel	36.974	-122.030
Santa Monica	1875	1875	Santa Monica Outlook	34.019	-118.490
Santa Rosa	1850	1866	The Sonoma Democrat	38.441	-122.713
Sausalito	1885	1885	The Sausalito News	37.859	-122.484
Selma	1886	1886	Selma Enterprise	36.571	-119.611
Sonoma	1823	1899	Sonoma City Expositor	38.292	-122.457
St. Helena	1874	1874	St. Helena Star	38.505	-122.469
Stockton	1850	1885	The Mail	37.958	-121.290
Tehachapi	1876	1919	The Tehachapi News	35.132	-118.448
Ukiah	1856	1861	Mendocino Herald	39.150	-123.207
Vacaville	1852	1883	Vacaville Reporter	38.357	-121.987
Vallejo	1850	1868	Vallejo Evening Chronicle	38.104	-122.856
Ventura	1852	1898	The Ventura Independent	34.278	-119.292
Visalia	1852	1859	The Tulare Post	36.330	-119.291
Watsonville	1868	1868	Daily Recorder	36.910	-121.756
Winters	1874	1887	The Winters Express	38.525	-121.970
Woodland	1861	1868	The Yolo Mail	38.679	-121.772
Yreka	1851	1898	The Yreka Daily Reporter	41.736	-122.633
Yuma	1540	1872	The Arizona Centinel	32.725	-114.624

[a] The year in which the city was first populated.

[b] The year in which continuous newspaper coverage started.

We next need to determine how much shaking would be felt at the known locations of populations and newspapers as a result of earthquakes at each grid point in the state. To estimate this I use the empirical relationship between earthquake magnitude, distance, and Modified Mercalli Intensity (MMI) given by Bakun and Wentworth (1997):

$$MMI = 1.68 M_W - 3.29 - 0.0206D. \tag{5}$$

where D is the distance between the city and the hypothetical earthquake source. We next need to determine how high the MMI at different cities would need to be to ensure the recording of the earthquake. To solve for this I compare the intensities produced by the earthquakes that were recorded in the pre-1900 historic catalog with the full set of intensities that that we would expect the different cities to have actually experienced, based on how seismicity was distributed throughout the state in the 1945-2006 instrumental catalog and the Gutenberg-Richter magnitude frequency distribution. (Note that only pre-1900 earthquakes were used for the historical set because some instrumental solutions are present in the catalog after this date.) For intensities above the intensity at which an earthquake is highly likely to be recorded, the measured and anticipated intensity distributions should look the same (Figure 5). The similarity
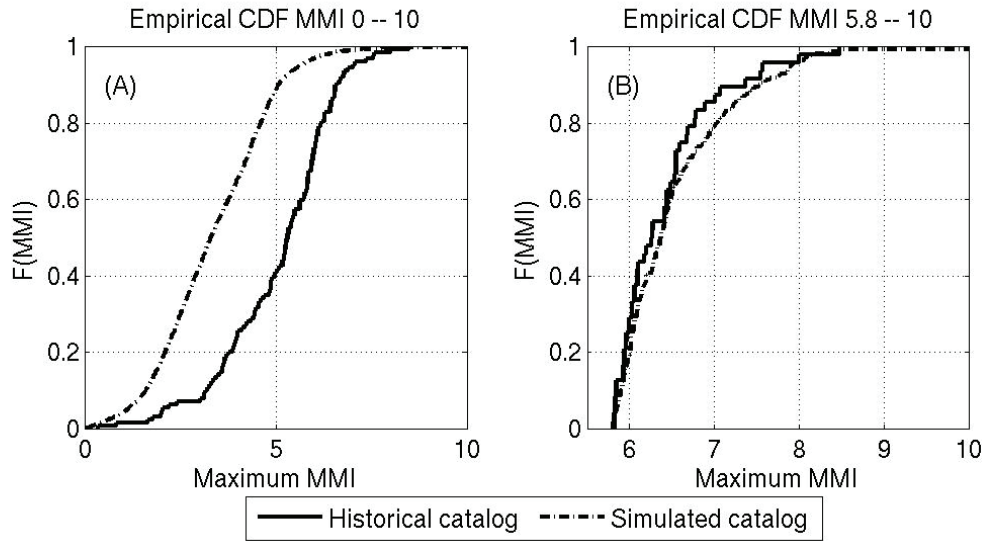
Figure 5: Distributions of maximum modified Mercalli Intensities (MMI) for earthquakes in the California 1850-1900 historic catalog (solid line) and a simulated catalog in which earthquake locations are based on earthquake densities in the 1945-2006 instrumental catalog and magnitudes follow a Gutenberg-Richter distribution M 5 – 8 (dashed-dotted line). The MMI intensities are calculated at all cities that had continuous newspaper coverage at the time that the earthquake occurred. (A) Empirical cdfs (cumulative density functions) are made from maximum MMI intensities between 0 and 10. The distributions for the simulated and historic catalogs are clearly different, with the simulated results containing many more smaller MMI, indicating that many earthquakes that produced small maximum MMIs were not included in the historic record, (B) MMIs between 5.8 and 10 only. Now the distributions for the simulated and real catalogs are statistically the same (Kolmogorov-Smirnoff test, 95% confidence), indicating that nearly all historical earthquakes that produced MMI 5.8 at least one location were recorded in the historic catalog.

of the distributions was statistically evaluated with the Kolmogorov-Smirnoff test. The distribution of intensities of MMI ≥ 5.8 produced by both simulated and real earthquakes at both newspaper or populated location was found to be the same. This indicates that all or nearly all earthquakes producing at least one MMI ≥ 5.8 measurement at a populated location were recorded in the historic catalog, which is reasonable since MMI 5.8 shaking is likely to cause some damage and substantial disarray. For lower levels of shaking the existence of a newspaper becomes more important; for example the distributions of the simulated and real intensities is the same if an earthquake produced at least 4 MMI ≥ 5.0 at cities with newspapers, or at least 6 MMI ≥ 5.0 intensities at populated locations that may or may not have had newspapers. Likewise at least 6 intensities of MMI ≥ 4.5 was sufficient at locations with newspapers for an earthquake to find its way into the historic record, but at least 10 MMI ≥ 4.5 observations were needed at general populated locations. The full list of intensity requirements found is given in Table 2.

Table 2: For each value of MMI (Modified Mercalli Intensity) this table provides the number of cities at which the MMI must be equaled or exceeded to ensure that the earthquake was be noted in the historic catalog.

MMI	# of Observations at Newspaper Cities[a]	# of Observations at Any Populated City[b]
5.8	1	1
5.6	2	2
5.0	4	6
4.5	6	10
4.0	10	13

[a] Number of locations with continuous newspaper coverage at which MMI must be equaled or exceeded.

[b] Number of simply populated locations, each of which may or may not have had their own newspapers at the time.

4.1.2 Determining Instrumental Magnitude Completeness Thresholds

Seismic instrumentation has existed at a few locations in California since the early 1900s, but a network of instruments suitable for recording local earthquakes and the systematic cataloging of earthquakes and assignment of local magnitudes did not begin in Southern California until 1932. An instrumental catalog was started in Northern California in 1942, but routine determination of magnitude did not begin until 1948 (Uhrhammer et al. 1996). Southern California instruments were capable of detecting some Northern California earthquakes, however, so I replace newspapers with instruments for completeness calculations statewide in 1932.

As noted above, an empirical approach is used to determine the completeness amplitude for each station for each decade of its operation. First I take all of the $M \geq 2.5$ earthquakes listed in the California catalog from 1932-2006, and estimate the Wood-Anderson seismograph (WAS) amplitude that each of the earthquakes would have

produced at each seismic station that was present at the time the earthquake occurred. This amplitude is estimated from the equation for local magnitude,

$$A = M_L - a_0 + n\log(D) + (k(\log(e \times D))) - S_c - 0.2; \qquad (6)$$

where A is Wood-Anderson amplitude in mm, $a_0 = 0.3173$, $k = -0.005$, $n = -1.14$, (constants currently being used in Southern California) and S_C = the station correction. For Southern California a list of seismic stations, the years for which each station was potentially in the field, are available on request from the Southern California Earthquake Data Center. Because the years an instrument is listed as being in the field in Southern California is not necessarily the time that the instrument was active and functional, however, I checked the station list against the stations listed as having recorded each $M \geq 4$ earthquake in Southern California. Only stations that actually appear on one or more of these earthquake lists for each five year period are counted as being active for that period. For Northern California I used the stations, station activity times, and station corrections given in Uhrhammer et al. (1996). The Northern California network also has a collection of new broadband triaxial stations, but these have not yet been assigned magnitude station corrections, and thus are not yet being used routinely to determine M_L. Since an earthquake cannot be assigned an accurate magnitude unless it is recorded at at least several M_L calculating stations, these broadband stations were not used for the present analysis. All of the stations included in the calculation are listed in Table 3.

Table 3: Seismic instruments used to calculate magnitude completeness thresholds after 1932.

Station	On Year	Off Year	Latitude	Longitude	Station Correction[a]
ARC	1952	2001	40.87772	-124.07738	0.2090
BAR	1955	1985	32.68005	-116.67215	-0.0600
BAR	2000	still in operation	32.68005	-116.67215	-0.0300
BC3	2000	still in operation	33.65515	-115.45366	0.1370
BKS	1962	still in operation	37.87622	-122.23558	-0.0350
BKR	2000	still in operation	35.26930	-116.07030	-0.3500
BRK	1934	still in operation	37.87352	-122.26099	0.1985
BRS	1990	1995	33.97145	-116.91265	-0.3200
BTC	2000	still in operation	33.01213	-115.21987	0.0840
BTP	2000	still in operation	34.68224	-118.57398	-0.3730
CIA	2000	still in operation	33.40186	-118.41372	-0.0330
CHF	2000	still in operation	34.33341	-118.02585	0.0680
CLC	2000	still in operation	35.81574	-117.59751	0.2860
CLI	1990	still in operation	33.14029	-115.52658	-0.1600
CLM	1990	1995	34.09613	-117.72297	-0.2000
CMB	1986	still in operation	38.03455	-120.38651	0.2400
CPP	2000	still in operation	34.06020	-117.80900	-0.4070
CWC	1965	1995	36.43988	-118.08016	-0.0100
CWC	2000	still in operation	36.43988	-118.08016	0.2860
DAN	2000	still in operation	34.63745	-115.38115	-0.1860
DEV	2000	still in operation	33.93597	-116.57794	-0.2380
DGR	2000	still in operation	33.65001	-117.00947	0.1000
DRC	2000	still in operation	32.80540	-115.44654	-0.2790

[a] Station correction for the calculation of magnitude.

Station	On Year	Off Year	Latitude	Longitude	Station Correction[a]
EDW	2000	still in operation	34.88303	-117.99106	0.1400
EWC	1990	2000	33.93724	-116.38216	0.1000
FPC	2000	still in operation	35.08200	-117.58267	0.0300
FUR	2000	still in operation	36.46703	-116.86322	-0.3300
GAV	1990	1995	34.02248	-117.50492	0.1500
GLA	1975	1980	33.05107	-114.82779	0.0500
GLA	2000	still in operation	33.05149	-114.82706	-0.0550
GR2	2000	still in operation	34.11830	-118.29940	-0.2660
GRH	1990	2000	34.30803	-118.55954	-0.5200
GSC	2000	still in operation	35.30177	-116.80574	-0.0800
GVR	1990	still in operation	34.04972	-118.11995	-0.3100
HAI	1932	1970	36.13664	-117.94753	-0.0300
HEC	2000	still in operation	34.82940	-116.33500	-0.1090
HOPS	1994	still in operation	38.99349	-123.07234	0.3240
ISA	1975	1980	35.66278	-118.47403	0.1600
ISA	1985	1990	35.66278	-118.47403	0.1400
ISA	2000	still in operation	35.66278	-118.47403	0.2200
JCS	2000	still in operation	33.08590	-116.59590	-0.0530
JRC	2000	still in operation	35.98230	-117.80760	-0.0600
JRSC	1994	still in operation	37.40373	-122.23868	0.1390
KCC	1995	still in operation	37.32363	-119.31870	0.3900
LJB	1990	2000	34.59092	-117.84890	0.1700
LJC	1932	1955	32.86340	-117.25414	-0.2900
LKL	2000	still in operation	34.61594	-117.82493	-0.4350
LUG	2000	still in operation	34.36560	-117.36683	-0.1380
LRL	2000	still in operation	35.47954	-117.68212	0.0530
MHC	1928	still in operation	37.34164	-121.64257	0.1280
MIN	1939	1999	40.34601	-121.60656	-0.1070
MLA	2000	still in operation	37.63019	-118.83605	-0.5670
MLS	2000	still in operation	34.00460	-117.56162	-0.2400
MPM	2000	still in operation	36.05799	-117.48901	0.1210
MTP	2000	still in operation	35.48434	-115.55320	0.1250
MWC	1932	1955	34.22362	-118.05832	-0.1100
MWC	1960	1965	34.22362	-118.05832	-0.1100
MWC	2000	still in operation	34.22362	-118.05832	0.0500
NEE	2000	still in operation	34.82490	-114.59941	-0.4400
NHL	1990	1995	34.39148	-118.59946	-0.2100
ORV	1992	still in operation	39.55451	-121.50036	0.4280
OSI	2000	still in operation	34.61450	-118.72350	-0.0130
PAS	1932	1995	34.14844	-118.17117	0.1700
PAS	2000	still in operation	34.14844	-118.17117	0.0500
PFO	2000	still in operation	33.61151	-116.45935	0.1400
PHL	2000	still in operation	35.40773	-120.54556	-0.0450
PLM	1970	1995	33.35361	-116.86265	-0.0500
PLM	2000	still in operation	33.35361	-116.86265	-0.0710
PLS	1990	still in operation	33.79530	-117.60906	-0.1300
RMM	1990	1995	34.64384	-116.62438	-0.4000
RPV	2000	still in operation	33.74346	-118.40412	-0.3500
RUS	2000	still in operation	34.05073	-118.08085	-0.3680
RVR	1932	1995	33.99351	-117.37545	0.0600
SAO	1988	still in operation	36.76403	-121.44722	0.3140

[a] Station correction for the calculation of magnitude.

Station	On Year	Off Year	Latitude	Longitude	Station Correction[a]
SBC	1932	1995	34.44076	-119.71492	-0.0900
SBC	2000	still in operation	34.44076	-119.71492	-0.2700
SBP	2000	still in operation	34.23240	-117.23484	-0.0690
SCI	2000	still in operation	32.97990	-118.54697	-0.0460
SCZ	2000	still in operation	33.99532	-119.63435	-0.5500
SDD	2000	still in operation	33.55259	-117.66171	-0.6380
SHO	2000	still in operation	35.89953	-116.27530	-0.4370
SLA	2000	still in operation	35.89095	-117.28332	-0.3470
SMF	1990	still in operation	34.02159	-118.44675	-0.2600
SMS	2000	still in operation	34.01438	-118.45617	-0.3300
SNC	2000	still in operation	33.24800	-119.52400	0.1620
SSC	1990	2000	33.99546	-119.63513	0.1000
SSW	2000	still in operation	33.17747	-115.61564	-0.2930
SOT	2000	still in operation	34.41600	-118.44900	-0.4480
STAN	1991	1994	37.40393	-122.17508	-0.2033
STO	1990	1995	34.69199	-117.11727	0.1700
SVD	2000	still in operation	34.10647	-117.09822	-0.2200
SWS	2000	still in operation	32.94080	-115.79580	-0.0610
SYL	1990	2000	34.35360	-118.45098	-0.3800
TA2	2000	still in operation	34.38203	-117.67822	-0.2450
TIN	1932	1995	37.05422	-118.23009	-0.2600
TIN	2000	still in operation	37.05422	-118.23009	-0.2410
THP	1990	1995	33.83172	-116.33896	-0.4000
THX	2000	still in operation	33.63495	-116.16402	-0.5300
TOV	2000	still in operation	34.15607	-118.82039	-0.0220
UPL	1990	1995	34.14817	-117.69940	-0.5000
USC	2000	still in operation	34.01919	-118.28631	-0.2700
VTV	2000	still in operation	34.56065	-117.32960	-0.3700
WDC	1992	still in operation	40.57988	-122.54113	0.4840
WDY	1955	1975	35.69998	-118.84421	0.1600
YBH	1993	still in operation	41.73193	-122.71038	0.4990

[a] Station correction for the calculation of magnitude.

After calculating the amplitudes that each earthquake would produce at each station, as noted above, I tabulated which stations were actually used in the processing of each earthquake, by finding which stations were listed in the earthquake's phase file. The initial goal was to find the seismic amplitude above which 95% of earthquakes attaining the amplitude were recorded at the station, but the percentages routinely reached a plateau before 95% was reached. Many stations present in 1932-1942, for example, recorded 89% of the earthquakes producing a WAS amplitude of 0.6 mm or higher at the station, and this percentage did not increase as the amplitude threshold was pushed higher and higher, up to 10 mm. Furthermore, with time, the percentage of recorded earthquakes at which a station reached a plateau was seen to steadily decrease, such that from 1982-1991 stations were rarely listed in the phase files of more than 30% of the earthquakes that produced amplitudes of $\geq$ 10 mm at them. This pattern suggests not that the stations actually didn't record the earthquakes, but that the more stations there are available, the choosier the analyst becomes. Thus with many stations to chose from, the analyst may routinely throw out up to 70% - 80% of the stations that did not produce the

clearest records, while in the 1930s, with many fewer stations available, only about 10% of recordings might routinely be thrown out or be unavailable for some reason.

Thus the goal became to look at a graph of the percentage of earthquakes that made it into a station's record, as a function of their amplitude, and find the amplitude at which the curve essentially leveled off. Since all of the records are observed to level off well before earthquake amplitude reaches 10 mm, I approximate the leveling off point as the amplitude at which the percentage of earthquakes being listed first reaches 95% of the percentage of earthquakes creating amplitudes of ≥ 10 mm that are listed at that station. The completeness amplitudes thus found tend to decrease from around 0.6 - 1.0 mm in the 1930s to 0.1 - 0.3 mm in the 1990s.

After the completeness amplitudes for each station were determined the state was divided into a 0.25 by 0.25 degree grid and the amplitudes that earthquakes of different magnitudes, occurring at each of the grid points, would produce at each of the stations was calculated. The completeness magnitude at each point was then set to the magnitude which surpassed the completeness amplitudes of at least 4 stations, since 4 stations are generally required for a robust epicentral and magnitude solution. A completeness magnitude was solved for for each point for each 5 year period from 1932-2006. Sometimes the completeness amplitudes at the stations show insignificant up and down variations from decade to decade; these are probably not real, but they lead to small and varied changes in the completeness magnitudes at the grid points, such that the completeness magnitude might be listed as 4.3, 4.2, 4.4, 4.3, etc., for neighboring time periods. In these cases, to simplify the results, when a group of consecutive completeness magnitudes show small and non-systematic variations they are uniformly replaced with the second-to-highest completeness magnitude in the group. The second to highest rather than highest value is used because of occasional short spikes in the computed completeness magnitudes.

4.2 Determining Completeness Magnitude with Time in Spatial Regions

We could use the completeness thresholds at points, solved for as described above, to calculate the seismicity rate. This could be done by determining seismicity rates in small boxes centered around each point and then summing to get the total rate. The problem is that each of these tiny boxes would have very few, if any, earthquakes, and having few observations leads to a large error on estimated rates. Thus in practice it appears best to combine groups of points with similar completenesses into larger regions. I make 8 regions in total, mapped in Figure 6: The North region (northern coast of California), the San Francisco Region, the Central Coast region, the Los Angeles Region (which also includes the San Diego area), the Mojave region, the Mid Region, and the Northeast region, which is characterized by very poor historical completeness but also apparently a low seismicity rate, and the rest of the state. The latitude/longitude vertices of the polygons that surround these regions are given in Table 4.

16
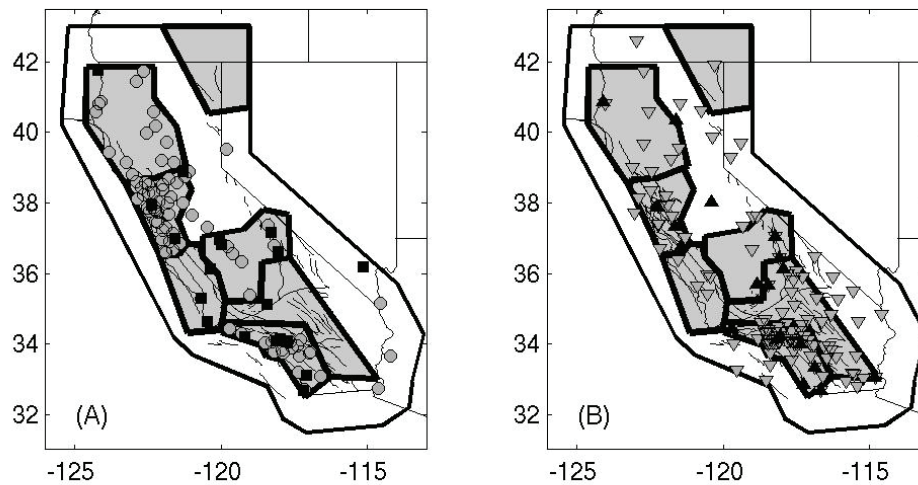
Figure 6: Different regions (shaded in gray) for which separate sets of completeness thresholds are calculated. Starting from the north and moving around counter-clockwise, the regions are the North region, the San Francisco region, the Central Coast Region, the Los Angeles region, the Mojave regions, and the Mid region, and the Northeast region. See Tables 5 through 12 for the magnitude completeness thresholds as a function of time in these regions and in the rest of the state. The regions are hand drawn around areas of similar completeness. In (A) cities with regular newspapers in print before 1900 are given in gray circles and cities used that had continuous newspaper coverage starting after 1900 are given by black squares. (B) Seismic stations operating before 1990 are given with black triangles and stations operating after 1990 are given with gray inverse triangles.

For the historical era I assign the completeness magnitude in a region to be the lowest magnitude to which 95% of the points are complete at the beginning of each five year completeness interval. 95% rather than 100% is used because earthquakes may also have been reported from mining camps, army forts, and other locations that are not included in the data base of cities and towns. For the instrumental era the completeness magnitude is assigned at the lowest magnitude to which 100% of the points in the region are complete at the beginning of the time interval. A tabulation of completeness results for each region is given in Tables 5, 6, 7, 8, 9, 10, 11, and 12.

17

Table 4: This table gives the vertices, in latitude and longitude, of the polygons that define the different regions used for the magnitude of completeness calculations. Vertices are given consecutively, in the clockwise direction. The different regions are mapped in Figure 6.

Region	Latitude/Longitude Limits
Full California Region	43.0, -125.2; 43.0 -119.0; 39.4, -119.0; 35.7, -114.0, 34.3, -113.1; 32.9, -113.5; 32.2, -113.6; 31.7, -114.5, 31.5, -117.1; 31.9, -117.9; 32.8, -118.4; 33.7, -121.0, 34.2, -121.6; 37.7, -123.8; 40.2, -125.4; 40.5, -125.4; 43.0, -125.2
North Region	38.507, -123.222; 40.232, -124.583; 41.862, -124.598; 41.851, -122.276; 40.952, -122.276; 40.188, -121.536; 39.060, -121.196; 38.703, -122.246; 38.668, -123.104; 38.507, -123.222
San Francisco Region	36.890, -121.148; 36.546, -121.422; 36.535, -122.099; 38.489, -123.207; 38.652, -123.090; 38.692, -122.243; 38.987, -121.344; 38.376, -121.005; 37.503, -121.461; 36.859, -121.135; 36.890, -121.148
Central Coast Region	34.406, -119.970; 34.320, -121.127; 36.519, -122.091; 36.547, -121.425; 36.828, -121.110; 36.814, -120.724; 36.463, -120.636; 35.982, -120.111; 35.212, -119.848; 34.378, -119.970; 34.406, -119.970
Los Angeles Region	33.043, -116.303; 32.498, -117.104; 33.577, -117.971; 33.649, -118.429; 33.973, -118.658; 34.328, -119.974; 34.619, -119.974; 34.557, -117.132; 33.043, -116.303
Mojave Region	33.022, -114.677; 33.073, -116.287; 34.551, -117.131; 34.627, -119.952; 35.193, -119.860; 35.230, -118.633; 36.313, -118.572; 36.436, -117.652; 33.022, -114.677
Mid Region	35.220, -119.839; 35.988, -120.085; 36.485, -120.619; 37.044, -120.619; 37.241, -119.223; 37.794, -118.565; 37.665, -117.703; 36.452, -117.662; 36.320, -118.606; 35.253, -118.647; 35.220, -119.839
Northeast Region	40.544, -120.410; 43.0, -121.984; 43.0 -119.011; 40.721, -119.011; 40.544, -120.410

Table 5: Completeness magnitudes for the North region

Starting Year	Ending Year	Magnitude of Completeness
1850	1855	7.3
1855	1860	7.1
1860	1865	6.7
1865	1875	6.4
1875	1880	6.3
1880	1890	6.2
1890	1932	6.1
1932	1942	5.6
1942	1952	5.2
1952	1957	5.1
1957	1997	4.7
1997	2007	3.4

Table 6: Completeness magnitudes for the San Francisco region.

Starting Year	Ending Year	Magnitude of Completeness
1850	1855	6.0
1855	1860	5.8
1860	1870	5.7
1870	1885	5.6
1885	1895	5.5
1895	1932	5.3
1932	1942	4.5
1942	1967	4.1
1967	1997	4.0
1997	2000	2.6
2000	2007	2.4

Table 7: Completeness magnitudes for the Central Coast region.

Starting Year	Ending Year	Magnitude of Completeness
1850	1855	7.4
1855	1860	7.3
1860	1870	6.6
1870	1890	6.5
1890	1905	6.4
1905	1932	6.3
1932	1987	4.1
1987	1992	3.8
1992	1997	3.5
1997	2000	2.9
2000	2007	2.7

Table 8: Completeness magnitudes for the Los Angeles region.

Starting Year	Ending Year	Magnitude of Completeness
1850	1855	6.9
1855	1870	6.4
1870	1875	6.2
1875	1890	6.0
1890	1905	5.8
1905	1932	5.7
1932	1993	3.9
1993	1997	2.8
1997	2000	2.6
2000	2007	2.1

Table 9: Completeness magnitudes for the Mojave region.

Starting Year	Ending Year	Magnitude of Completeness
1850	1855	8.0
1855	1865	7.4
1865	1870	7.3
1870	1875	7.1
1875	1880	7.0
1880	1890	6.9
1890	1895	6.8
1895	1910	6.7
1910	1932	6.6
1932	1993	4.1
1993	1997	3.0
1997	2000	2.9
2000	2007	2.2

Table 10: Completeness magnitudes for the Mid region.

Starting Year	Ending Year	Magnitude of Completeness
1850	1855	8.0
1855	1865	7.5
1865	1870	6.6
1870	1875	6.5
1875	1880	6.3
1880	1890	6.2
1890	1932	6.1
1932	1957	4.2
1957	1992	3.9
1992	1997	3.4
1997	2000	3.2
2000	2007	2.7

Table 11: Completeness magnitudes for the Northeast region.

Starting Year	Ending Year	Magnitude of Completeness
1850	1932	8.0
1932	1942	5.7
1942	1967	5.3
1967	1997	4.7
1997	2007	3.7

**Table 12:** Completeness magnitudes for the rest of the state.

Starting Year	Ending Year	Magnitude of Completeness
1850	1865	8.0
1865	1870	7.4
1870	1885	7.2
1885	1910	7.1
1910	1932	6.9
1932	1942	6.0
1942	1957	5.6
1957	1997	5.1
1997	2007	4.0

5 Calculation of the Gutenberg-Richter *b* Value

The accurate calculation of seismic risk is very sensitive to the Gutenberg-Richter *b* value. If *a* is based on the number of $M \geq 4$ earthquakes, for example, then *b* value error as small as 0.05 will cause the calculated rate of $M \geq 6.5$ earthquakes to be off by 25%, and an error of 0.1 will cause the $M \geq 6.5$ rates to be off by 50%.

Like *a* value calculation *b* value calculation is adversely affected by magnitude rounding and error, especially when errors vary systematically with magnitude. The calculation of *b* is also very sensitive to the completeness level of the catalog. For a *b* value to be accurate to the nearest 0.05, for example, 95% of the earthquakes that occurred above the minimum magnitude used, M_C, must be present in the catalog. If catalog completeness is estimated by eye from a cumulative magnitude frequency plot, as is often the case, M_C can easily be set too low to the degree that a *b* value underestimated by 0.1 to 0.2 will result (Figure 7). For historic and older catalogs it is usually difficult to find a value of M_C that both truly applies across the entire area and time covered and is low enough to provide a catalog large enough for good statistical analysis.
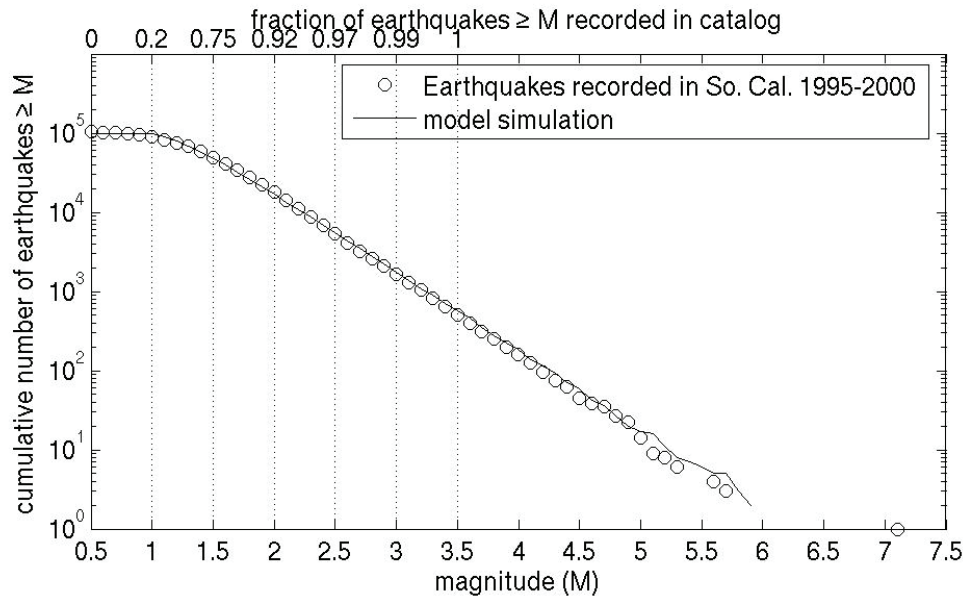
Figure 7: We model catalog incompleteness with $P = 1 - C10^{-M}$ for $C10^{-M} > 1$ and *P = 0.002* otherwise, where *P* is the probability that an earthquake of magnitude *M* will be recorded. For Southern California from 1995-2000 (a period when the network was relatively stable) we find *C = 8*. We use Monte Carlo to generate a simulated catalog with this incompleteness function and a GR distribution with *b* =1. The cumulative magnitude frequency distribution of the simulated catalog is plotted with a black line. Data from the Southern California catalog is given by circles. The bottom X axis gives earthquake magnitude; the top X axis gives the percentage of earthquakes occurring at that magnitude recorded in the simulated catalog. Note that the magnitude frequency curve visually appears quite complete even when only 75% of the earthquakes are recorded.

Fortunately, unlike the *a* value, calculating *b* does not require averaging over a long period of time as *b* has been observed to be time invariant. Thus we may avoid many of the above problems by only using the most modern and accurate catalog data to calculate *b*. Through a comprehensive instrument-based completeness magnitude analysis described below I find that the entire California region was complete to M 4.0 from 1997-2006. Then using only this part of the catalog, I use the maximum likelihood method (MLE) of Aki (1965) to obtain *b* = 1.02±0.11. Since this value is insignificantly different from the global average *b* value of 1.0 I use *b* =1.0 to solve for the mean expected seismicity rates, and the outer 98% confidence values of *b* (*b* = 0.91 and *b* = 1.13) to fill out the full error range of possible rates. I also solve for the *b* value for the declustered version of the 1997-2006 catalog, where declustering is accomplished with the method by (Gardner and Knopoff 1974), which has been traditionally used by the National Hazard Maps. Because this declustering method preferentially removes smaller earthquakes from clusters, it changes the magnitude frequency distribution. I find a *b* value of 0.85±0.13 for the declustered catalog. Because this value is insignificantly different from the *b* = 0.8 used for the declustered catalog by the 2002 National Hazard Maps I use *b* = 0.8, for the mean expected seismicity rates and the outer 98% confidence levels of *b* (*b* = 0.72 and *b* = 0.98) to find the full error range.

6 Calculating Seismicity Rates

With magnitudes corrected for rounding and errors, *b* value calculated, and completeness thresholds determined, average long term seismicity rates for California can now be calculated from the 1850-2006 catalog. The rate calculation is complicated, however, by the existence of the varying magnitude completeness thresholds in time and space, uncertainty about whether the seismicity rates have been constant in time over the entire 1850-2006 time period, and uncertainty about how well 156 years of catalog data accurately reflects true long term seismicity rates. To address these various uncertainties the rates are calculated with four different methods: direct observation, the Weichert method (Weichert 1980), the averaged Weichert method, and the averaged Weichert method with a long-term catalog correction. Summaries of each method, and the reason for using it, are given below.

6.1 Direct Observation

The direct observation method for measuring earthquake rates is the most straightforward. For each region of the state, the rate of earthquakes above a given magnitude threshold, *M*, is simply measured from the part of the catalog that is complete to *M*. The Los Angeles region, for example, is considered to be complete to M 5 from 1932, and so the number of $M \geq 5$ earthquakes from 1932-2006 are counted, and then divided by 2006 - 1932 = 74 to get the average annual $M \geq 5$ rate, and so on. The measured annual rates above each magnitude threshold for each region are then added together to get the estimated annual rates for the whole state. Direct observation is used to calculate seismicity rates in the full and declustered catalogs (Tables 13 and 14) and with corrections made for the systematic biases created by magnitude error and rounding (Tables 13 and 14) and without these corrections made (Tables 15 and 16). When the

magnitude error and rounding corrections are done they are done both with the preferred Gutenberg-Richter *b* values of 1.0 for the full catalog and 0.8 for the declustered catalog (Tables 13 and 14) and with the most extreme *b* values allowed by the 98% confidence intervals (Table 17). The corrections are implemented by creating 500 corrected earthquake catalogs (each corrected catalog is a bit different because of the randomization of the magnitude rounding correction – see above), making a direct observation rate calculation for each catalog, and then averaging the 500 rates.

Table 13: Cumulative calculated seismicity rates for the California region, *non-declustered* (full) catalog, corrected for rounding and magnitude errors.

Magnitude Range	Calculated with Weichert Method	Calculated with Averaged Weichert Method	Direct Observation Rate
$M \geq 5.0$	5.63 ± 0.6	6.8 ± 2.75	$4.73, -1.2 + 1.50$
$M \geq 5.5$	1.78 ± 0.19	2.15 ± 0.87	$2.15, -0.37 + 0.43$
$M \geq 6.0$	0.56 ± 0.06	0.67 ± 0.27	$0.71, -0.22 + 0.28$
$M \geq 6.5$	0.17 ± 0.019	0.21 ± 0.09	$0.24, -0.09 + 0.11$
$M \geq 7.0$	0.05 ± 0.006	0.06 ± 0.027	$0.074, -0.04 + 0.06$
$M \geq 7.5$	0.012 ± 0.0019	0.015 ± 0.008	$0.020, -0.016 + 0.035$

- Model projections are made with the Gutenberg-Richter relationship and b = 1.
- Errors are given at the 98% confidence level for the Weichert rates and 95% confidence for the direct observation rates.
- Observed rates are made from observations only from the parts of the catalog in time and space that are complete to the given magnitude.

Table 14: Cumulative calculated seismicity rates for the California region, *declustered* catalog, corrected for magnitude and rounding errors.

Magnitude Range	Calculated with Weichert Method	Calculated with Averaged Weichert Method	Direct Observation Rate
$M \geq 5$	3.23 ± 0.44	3.8 ± 1.2	2.78 ± 0.4
$M \geq 5.5$	1.28 ± 0.17	1.50 ± 0.48	1.34 ± 0.3
$M \geq 6$	0.49 ± 0.07	0.59 ± 0.19	$0.57, -0.2, +0.3$
$M \geq 6.5$	0.19 ± 0.02	0.22 ± 0.08	$0.22, -0.08, +0.11$
$M \geq 7$	0.067 ± 0.011	0.08 ± 0.03	$0.08, -0.04, + 0.06$
$M \geq 7.5$	0.02 ± 0.004	0.02 ± 0.012	$0.023, -0.017, + 0.035$

- A *b* value of 0.8 is used to correct for the magnitude errors and project the model rates with the Gutenberg-Richter relationship, which is truncated at a maximum magnitude of M 8.
- Errors are given at the 98% confidence level for the Weichert rates and 95% confidence for the direct observations rates.
- Observed rates are made from observations only from the parts of the catalog in time and space that are complete to the given magnitude.

Table 15: Cumulative calculated seismicity rates for the California region, *non-declustered* (full) catalog, <u>not</u> corrected for rounding and magnitude errors.

Magnitude Range	Calculated with Weichert Method	Calculated with Averaged Weichert Method	Direct Observation Rate
M ≥ 5	6.26 ± 0.64	8.3 ± 2.95	5.09, -1.8, +2.2
M ≥ 5.5	1.98 ± 0.20	2.63 ± 0.92	2.36 ±0.4
M ≥ 6.0	0.62 ± 0.064	0.83 ± 0.292	0.89, -0.3, +0.4
M ≥ 6.5	0.19 ± 0.02	0.256 ± 0.09	0.31, -0.1, +0.13
M ≥ 7.0	0.06 ± 0.006	0.075 ± 0.03	0.09, -0.05, + 0.07
M ≥ 7.5	0.01 ± 0.002	0.018 ± 0.009	0.026, -0.019, + 0.039

- Model rates are projected with the Gutenberg-Richter relationship and b = 1.
- Errors are given at the 98% confidence level for the Weichert rates and 95% confidence for the direct observation rates.
- Observed rates are made from observations only from the parts of the catalog in time and space that are complete to the given magnitude.

Table 16: Cumulative calculated seismicity rates for the California region, *declustered* catalog, <u>not</u> corrected for rounding and magnitude errors.

Magnitude Range	Calculated with Weichert Method	Calculated with Averaged Weichert Method	Direct Observation Rate
M ≥ 5.0	3.46 ± 0.29	4.07 ± 1.26	3.03, -0.8, +1.0
M ≥ 5.5	1.37 ± 0.12	1.61 ± 0.50	1.44, -0.29, +0.36
M ≥ 6.0	0.54 ± 0.05	0.63 ± 0.20	0.69, -0.16, +0.20
M ≥ 6.5	0.21 ± 0.019	0.24 ± 0.08	0.27, -0.09, +0.12
M ≥ 7.0	0.07 ± 0.007	0.09 ± 0.03	0.09, -0.04, + 0.07
M ≥ 7.5	0.02 ± 0.003	0.025 ± 0.013	0.026, -0.016, + 0.034

- A b value of 0.8 is used to correct for the magnitude errors and project the model rates with the Gutenberg-Richter relationship, which is truncated at a maximum magnitude of M 8.
- Errors are given at the 98% confidence level for the Weichert rates and 95% confidence for the direct observations rates.
- Observed rates are made from observations only from the parts of the catalog in time and space that are complete to the given magnitude.

Errors on each direct observation rate are calculated from the magnitude errors associated with each earthquake and from the Poissonian distribution, which, based on the length of the observation period, tells us how much random variability is expected in observed rates for a given underlying rate. The errors for the direct observation rates are calculated at the two-tailed 95% confidence rate such that 2.5% of the time the true rate will be less than the lower confidence limit and 2.5% of the time the true rate will be higher than the error bars.

As noted above, magnitude errors introduce two types of error to the seismicity rate calculation. The first is a systematic upwards bias of the seismicity rate; this bias and how it is corrected for is detailed above. The second component is random, Gaussian error. I estimate the standard deviation of this Gaussian error with Monte Carlo

simulations. For each earthquake population for which a direct observation rate is being measured, 500 simulated catalogs of the same size are generated. The magnitudes in each of these catalogs is then randomly perturbed according to the amount of magnitude error listed for each earthquake in the real catalog. The systematic error introduced by these magnitude errors is corrected for as described above. Finally the seismicity rates in all of the perturbed and corrected catalogs are compared to the rates in the original catalogs to solve for the standard deviation of the Gaussian component of the magnitude error induced seismicity rate error.

Table 17: The effect of using different *b* values to correct for magnitude rounding and error on direct count seismicity rate results. The 98% confidence range for *b* for the full catalog is 0.91 to 1.13 and for the declustered catalog is 0.72 to 0.98. This range of *b* values has a significant effect on the calculated model rates (Tables 18 and 19) but has very little effect on the direct count seismicity rates given here. The first column in the table gives magnitude; the other columns give the average annual rate of earthquakes larger than or equal to this magnitude, calculated from catalogs corrected with the *b* values given. An 'F' in the column title indicates that the full catalog is being measured, while a 'D' indicate that the declustered catalog is being used. Errors on the measured rates are given at 95% confidence.

Mag Range	b = 0.91, F.	b = 1.13 F.	b = 0.72, D.	b = 0.98, D.
M ≥ 5.0	4.75, -1.3 + 1.5	4.70, -1.2, +1.54	2.78 ± 0.4	2.77 ± 0.4
M ≥ 5.5	2.16 ± 0.4	2.14 ± 0.4	1.35 ± 0.3	1.34 ± 0.3
M ≥ 6.0	0.72, -0.2, +0.3	0.71, -0.22 + 0.28	0.57, -0.2 + 0.3	0.55, -0.2 + 0.3
M ≥ 6.5	0.24, -0.08 + 0.11	0.23, -0.08, +0.11	0.22, -0.08 + 0.11	0.21, -0.08, +0.11
M ≥ 7.0	0.07, -0.04, +0.06	0.07, -0.4, +0.6	0.08, -0.04 + 0.06	0.07, -0.04, +0.06
M ≥ 7.5	0.02, -0.016, +0.04	0.019, -0.017, +0.036	0.023, -0.017 + 0.035	0.023, -0.017 + 0.04

Once the effect of magnitude error on each data set is found, the complete error on each seismicity rate is calculated via an iterative approach. The goal is to find the lowest underlying seismicity rate that would produce an observed rate equal to or exceeding that observed 2.5% of the time and the highest underlying rate that would produce the number of observed earthquakes or less 2.5% of the time. These extreme underlying rates are found by first guessing a value for them, then using the Poissonian function to generate 10,0000 samples of the total number of earthquakes that the guessed rate might produce over the observation period, and then randomly perturbing each of these values to account for the effect of the Gaussian component of the magnitude error. Finally we calculate an annual rate from each simulated number of observed earthquakes. If the 10,000 simulated earthquake rates do not satisfy the 2.5% criteria then the guessed rate is adjusted and the simulations are run again.

The benefits of calculating seismicity rates via direct observation are that it is a straightforward method that does not require any assumptions about the underlying magnitude-frequency distribution. There are two drawbacks, however. The first if that if the Gutenberg-Richter magnitude-frequency distribution may be assumed for a

population of earthquakes (and this is a justifiable assumption, given the universal observation of the Gutenberg-Richter distribution in large and well-instrumented earthquake populations), then the rate of the largest earthquakes may be more accurately estimated by projecting upwards from the more robustly constrained rates of smaller earthquakes than by counting the number of larger earthquakes alone. The other issue is that catalog completeness times can be very short and non-representative for some of the smaller earthquakes. In some areas of California, for example, the catalog did not become reliably complete to M 5 until 1995, and the 1995-2006 seismicity rate happened to be lower in many areas of the state than over longer time periods. Even at M 5.5 and M 6.0 many regions of the state did not have complete coverage until 1932 or 1942 when the instrumental network started, so the rates measured for these magnitude ranges will not represent long-term rates to the same extent as those measured for larger magnitude cutoffs. The seismicity rate measurement methods that follow are designed to address one or both of these issues.

6.2 The Weichert Method

The National Seismic Hazard Maps of 1996 and 2002 used the method proposed by Weichert (1980) to calculate model seismicity rates. The Weichert method is a maximum likelihood algorithm designed to solve for the seismicity rate and the Gutenberg-Richter *b* value from a catalog whose completeness magnitude threshold changes with time. The Weichert method assumes that the Gutenberg-Richter magnitude frequency distribution holds, thus the uncertain rates estimated for the scarce large earthquakes and the smaller earthquakes with short recording times are made more accurate via projection from rates at other magnitude levels. I apply the Weichert method to calculate the average seismicity rate in each of the completeness regions (see section on completeness) and then sum to get an average seismicity rate for the entire state. Corrections for the systematic bias created by magnitude rounding and error is accomplished by creating 500 corrected catalogs (each corrected catalog is a bit different because of the randomization of the magnitude rounding correction – see above), running the Weichert calculation on each corrected catalog, and then averaging the resulting rates. One change that I make from the 1996 and 2002 NHM implementation of the Weichert method is that I fix the Gutenberg-Richter *b* value rather than allowing the algorithm to solve for it; *b* is fixed to 1.0 for the full catalog calculations and 0.8 for the declustered catalog (see calculations of the *b* value above). Trying to solve for *b* from the entire 1850-2006 catalog would be problematic because of magnitude rounding and errors which cannot be corrected for until *b* value is known.

Rates calculated with the Weichert method for the full and declustered catalogs, with corrections made for the systematic bias introduced by magnitude rounding and error, and with the preferred *b* values of 1.0 and 0.8, respectively, are given in Tables 13 and 14. The same calculations but without magnitude rounding and error corrections are given in Tables 15 and 16. Weichert rate calculations for the full and declustered catalogs with alternative Gutenberg-Richter *b* values (and with magnitude rounding and error corrections) are given in Tables 18 and 19. Errors on these rates are calculated with the method described by Weichert (1980) and are expressed at the 2σ, or about 98% confidence, level. 1σ values may be easily obtained by dividing the given errors in half.

Table 18: The effect of using different *b* values on the model results for the full (non-declustered) catalog. The earthquake catalog is corrected for rounding and magnitude error for these calculations. The 98% confidence range for the *b* value for the full data is 0.91 to 1.13. Calculations are done with the straight Weichert (columns with W on the top) and averaged Weichert (columns with av. W. on the top) methods, and the specified *b* value is used to correct for magnitude rounding and errors and to project $M \geq 5.0$ seismicity rates to other magnitudes. Note that changing the *b* value has opposite effectson the rates calculated with the Weichert and averaged Weichert methods. This is the because the Weichert method puts the most weight on the smallest earthquakes, near M 4, and on projecting the rate of $M \geq 5$ earthquakes up from these, so with a lower *b* value the projected rate of $M \geq 5$ earthquakes will be higher whereas with a higher *b* value it will be lower. The averaged Weichert method, on the other hand, puts more weight on the $M > 5.5$ earthquakes, to which parts of the historical record are complete, and on projecting the rate of $M \geq 5$ earthquakes down from these. Thus with a lower *b* value the rate of the down-projected M 5 earthquakes will be lower, while with a higher *b* value the rate will be higher.

Mag Range	*b* = 0.91, W.	*b* = 0.91, av. W.	*b* = 1.13, W.	*b* = 1.13, av. W.
M ≥ 5.0	6.020 ± 0.580	5.980 ± 1.900	5.190 ± 0.640	8.850 ± 4.900
M ≥ 5.5	2.100 ± 0.200	2.100 ± 0.640	1.400 ± 0.170	2.400 ± 1.300
M ≥ 6.0	0.730 ± 0.070	0.730 ± 0.220	0.380 ± 0.050	0.640 ± 0.350
M ≥ 6.5	0.250 ± 0.030	0.250 ± 0.080	0.100 ± 0.013	0.170 ± 0.096
M ≥ 7.0	0.080 ± 0.009	0.080 ± 0.030	0.030 ± 0.004	0.040 ± 0.030
M ≥ 7.5	0.020 ± 0.003	0.002 ± 0.010	0.006 ± 0.001	0.010 ± 0.007

Table 19: The effect of using different *b* values on the model results for the declustered catalog. The 98% confidence range for *b* for the declustered data is 0.72 to 0.98. These *b* values are used to calculate the seismicity rates with the Weichert method (columns headed by a "W") and with the averaged Weichert method (columns headed with "av. W").

Mag Range	*b* = 0.72, W.	*b* = 0.72, av. W.	*b* = 0.98, W.	*b* = 0.98, av. W.
M ≥ 5.0	3.20 ± 0.410	3.05 ± 0.84	2.850 ± 0.470	4.800 ± 2.500
M ≥ 5.5	1.39 ± 0.180	1.32 ± 0.37	0.920 ± 0.150	1.520 ± 0.790
M ≥ 6.0	0.59 ± 0.080	0.56 ± 0.16	0.300 ± 0.050	0.490 ± 0.250
M ≥ 6.5	0.25 ± 0.030	0.24 ± 0.07	0.090 ± 0.016	0.150 ± 0.080
M ≥ 7.0	0.10 ± 0.020	0.09 ± 0.03	0.030 ± 0.005	0.046 ± 0.030
M ≥ 7.5	0.03 ± 0.007	0.03 ± 0.01	0.007 ± 0.002	0.011 ± 0.009

The Weichert method is efficient and has relatively low calculation errors. The one significant drawback to the method is that it assumes that the seismicity rate is constant with time. Based on this assumption the method assigns each earthquake that can be counted equal weight in the rate calculations. Thus the seismicity rates during time

periods which have more countable earthquakes – e.g. time periods with lower completeness thresholds – end up influencing the total solution much more strongly than the seismicity rate at other times. For example, we consider a sample situation similar to that in the San Francisco Bay Area, in which, say, an area is complete to M 5.5 for one 50 year time period and complete to M 4.0 in a subsequent 50 year time period. The second time period will contain on the order of 30 times more useable earthquakes than the first. As a result the Weichert method will weight the second time period 30 times higher in the final rate calculation – or the final rate will be based ~97% on the second 50 year time period and only 3% on the first. If the average underlying seismicity rates during the two time periods are in fact the same then this weighting method is sound, as the rate measurement during the second time period will be far more accurate, based as it is on a much larger data base. If the underlying seismicity rates in the two time periods are different, however, the calculated rate will not reflect the averaged rate over the entire 100 year period, which is what we want, but rather, to a very large extent, only the seismicity rate over the second 50 year time period.

As a result of this, the Weichert rate calculation will produce an accurate long term (or at least an accurate 157-year averaged) seismicity for California only if the historical (1850-1932) and instrumental (1932-2006) parts of the earthquake catalog, which have substantially different magnitude completeness thresholds, have the same earthquake rate. Observationally, there is evidence that the rates are not in fact the same. In the San Francisco Bay Area, for example, the higher seismicity rate in the historic vs. instrumental era is well known (e.g. the ``stress shadow''). The same contrast can also be seen across the rest of the state, with the historical catalog more active than the instrumental. The only exceptions to this are the Northeast and Mojave regions, where completeness thresholds are high and the data is not sufficient to measure a historic rate (Table 20).

It is difficult to ascertain to what degree the differences between the historic and instrumental catalog rates are real. In fact, the historic completeness thresholds are so high that the error bars on the historic rate are often large enough to encompass the instrumental seismicity rate (Table 20). The magnitude errors provided for earthquakes in the historic catalog might also underestimate the true magnitude errors, which would cause overestimation of the seismicity rate. On the other hand, some of the largest earthquakes in the California record, including the M 7.8 1906 San Francisco earthquake, M 7.9 1857 Ft. Tejon earthquake, and the M ~7.6 (or quite possibly M 7.8 (Hough and Hutton 2006)) 1872 Owens Valley earthquake, occurred during the historic era, indicating that this may really have been a more seismically active time for California. If the higher seismicity rates from 1850-1932 are real, we need an alternative rate calculation method that will allow the historic part of the record to have influence despite its high completeness threshold. I call this method the ``averaged Weichert'' seismicity rate calculation.

Table 20: This table gives the average annual $M \geq 5$ earthquake rates in the historical (1850-1932) and instrumental (1932-2006) periods for each completeness region, or region in which a separate set of magnitude completeness thresholds are calculated. The regions are mapped in Figure 7. Rates are calculated using the Weichert method (Weichert 1980), a b value of 1.0, and corrections for magnitude and rounding error. Errors are given at the 98% confidence level. The Northeast region has very little seismicity, and no historical earthquakes above the magnitude completeness thresholds, so its historical seismicity rate cannot be estimated. Likewise the Mojave region had high completeness thresholds historically, and no earthquakes above its completeness threshold are present in the Toppozada version of the historic catalog, on which the numbers above are based. In the Bakun version of the historic catalog there is one earthquake, an M 7.2 in 1892, that is above the historical magnitude completeness threshold in the Mojave.

Region	1850-1932 Rate	1932-2006 Rate
North region	0.65, -0.46, +0.98,	0.42 ± 0.05
San Francisco region	1.35 ± 0.43	0.35 ± 0.05
Central Coast region	0.72, -0.55, +1.56	0.26 ± 0.04
Los Angeles region	0.97, -0.41, +0.71	0.42 ± 0.05
Mojave region	-	1.22 ± 0.09
Mid region	0.90, -0.51, +1.22	0.40 ± 0.05
Northeast region	-	0.02 ± 0.02
Rest of state	3.4, -2.92, +6.18	1.69 ± 0.22

6.3 The Averaged Weichert Method

The averaged Weichert method is designed to use the advantages of the Weichert rate calculation but to minimize the assumption that the parts of the catalog with significantly differing magnitude completeness thresholds all have the same seismicity rates – that is to drop the assumption that the long term average seismicity rate is correctly represented by only the most recent and complete part of the catalog. Dropping this assumption is accomplished by breaking the catalog at the points at which the magnitude completeness threshold drops significantly statewide, in 1932 and 1997 (due to the introduction and major expansion of instrumentation, respectively). Separate seismicity rates, using the Weichert routine, are calculated for the time periods 1850-1932, 1932-1997, and 1997-2006, and then the three rates are arithmetically averaged together, weighted by the number of years contained in each time period. Errors are also calculated separately for each time period and then propagated when the rates are combined.

Averaged Weichert rates for the full and declustered catalogs, using the preferred Gutenberg-Richter b values of 1.0 and 0.8, respectively, and with systematic rounding and magnitude error bias corrected for, are given in Tables 13 and 14. The same values without the rounding and magnitude error biases corrected for are given in Tables 15 and 16. Rates calculated with different b values, for the full and declustered catalogs, are given in Tables 18 and 19, respectively.

For the full catalog (corrected for rounding and magnitude error), the averaged Weichert rate is about 30% higher than the regular Weichert rate, but for the declustered catalog the difference is much more muted, at only a 7% increase. This contrast is probably in part because the Gardner and Knopoff (1974) declustering preferentially removes the smaller earthquakes, which dominate the Weichert seismicity rate calculation in the instrumental era when they are present.

6.4 Correcting Rates for Potentially Short Catalog Duration

As briefly noted in the introduction, one of the most serious problems that we face in calculating the long term average annual seismicity rate from the catalog is that the useable catalog covers only 156 years. Is this long enough to calculate the truly long term average rate?

Earthquake rates may be highly heterogeneous over time because of the strong tendency of earthquakes to cluster. The largest bursts of earthquake clustering, or aftershocks, tend to follow the largest earthquakes. Thus whether or not our catalog is long enough to smooth out the effects of clustering and recover an unbiased seismicity rate is dependent on whether the catalog is long enough to contain several earthquakes of around magnitude M_{max}, defined here as the magnitude of the largest earthquake possible in the state of California. The Gutenberg-Richter distribution over the whole state is expected, over long term observation, to continue smoothly up to M_{max}, where it is truncated. Thus the rate of the M_{max} earthquake is the Gutenberg-Richter relationship rate of earthquakes $M \geq M_{max}$, but earthquakes larger than M_{max} do not actually occur.

There are several earthquakes of M 7.8-7.9 in the 1850-2006 catalog. If $M_{max} \sim 7.9$, then the average seismicity rates and errors that we calculate are likely to be a fair representation. On the other hand, if we assume that M_{max} may equal 8.3 – a magnitude that might occur, for example, if the entire San Andreas Fault ruptured at once – then 1850-2006 does not represent a full ``seismic cycle'' and may systematically over or underestimate the true long term rate.

I use statistical ETAS earthquake simulations (Ogata 1988; Felzer et al. 2002), which model a constant background rate plus aftershock generation, to try to estimate how far off the 1850-2006 a value might be from the true long term a value if M_{max} is equal to 8.3. In these simulations all earthquakes $M \geq 2.5$ are modeled and may produce their own aftershocks. Because 10,000 years is too long for multiple timely simulations, however, simulations are run on much shorter time scales and then the results extrapolated to the needed times. Specifically, we want to find if the average recurrence period of M_{max} is P years, and we look at seismicity over a time period T, by what fraction might the a value observed over T differ from the long term a value? I investigate 1000 simulations of 30 days of earthquakes, 1000 simulations of 1 year of earthquakes, 500 simulations of 5 years, 500 simulations of 10 years, and 100 simulations of 50 years each. As the time period simulated increases, results for each trial become more stable, so the total number of trials may be decreased. The background seismicity rate for the trials was set at a rate based on 1932-2006 seismicity. From this the simulations produced a long term full catalog a value of $5.58 \geq 5.0$

earthquakes/year and a corresponding average $M_{max} = 8.3$ repeat time of about 1000 years.

From each simulation I calculate two values. One value, D, defined as follows, is a measure of how far away the a value for the sample (e.g. each single 30 day or 5 year time period simulated) is from the long term a value,

$$D = \frac{a_L - a_S}{a_S} \tag{7}$$

where a_L is the long term a value and a_S is the sample a value. The other value calculated is $R = T/P$, or the ratio of the sample time to the recurrence time of M_{max}. Not surprisingly, when R is low D is high. Values of a_S measured over short times tend to underestimate a_L the majority of the time, because they contain no sizeable earthquakes, and to strongly overestimate a_L when a particularly large earthquake and its aftershock sequence does occur. For the 1000 simulations of 30 days each I find an average D value of 1.3 ± 0.72 (1 σ), which, substituted into Equation 7, gives that on average a_S is 43% of a_L. As R increases the situation improves, although as R gets larger and larger the rate of improvement slows. Values of a_S measured over 50 year periods, for example, are on average 89% of a_L ($D = 0.12$, $\sigma = 0.3$).

We find that the following inverse power law functions can be fit to R and D and to R and σ to extrapolate to larger values of R,

$$D = 10^{-1.282} R^{-0.357}, \tag{8}$$

and

$$\sigma = 10^{-0.66} R^{-0.124}. \tag{9}$$

Using these equations for a 156 year time period I expect an average D of 0.10 and $\sigma = 0.28$, meaning that a_S is expected to be $91\% \pm 20\%$ of a_L, at one σ, or $91\% \pm 40\%$ of a_L at 98% confidence, or the long term rate may be calculated from the 155 year catalog rate, on average, by multiplying by 1/0.91, or 1.1. If the background seismicity rate is really faster than the one I have chosen for the simulations, meaning that the repeat time for an M 8.3 is less than 1000 years, then our 156 year sample will be closer to the long term rate – so the errors stated here represent the outer bounds of the real error.

One potential issue is that our simulations count earthquakes down to M 2.5, which provides higher earthquake counting accuracy than in the real catalog, which has high and varying magnitude completeness thresholds. To try to quantify the effect I recalculate errors on estimation of a_L from the simulation data when I only count $M \geq 5.5$ earthquakes. This creates significantly higher errors for some of the shorter time periods, but for the 50 year trials σ remains very similar ($\sigma = 0.33$ rather than 0.30), indicating that for time periods this long the error introduced by the clustering may be larger than the earthquake counting error. Of course in many areas in California there are periods of time that are not even complete to M 5.5. So, as an approximation, we simply use the larger of either the Poissonian or clustering-induced error.

It is important to emphasize that clustering produces bias and additional error *only* for the *a* value in the Gutenberg-Richter relationship, and not for the *b* value. Thus the whole magnitude frequency curve may be moved up or down in accordance with the changes in *a* value given above, but the shape of the curve does not change.

We consider our preferred calculated seismicity rate to be the averaged Weichert rate (calculated with the best fitting *b* values and corrected for magnitude and rounding error), adjusted by the factor of 1.1 to account for the possibility of M 8.3 earthquakes and higher seismicity rates over the long term (see further explanation for our choice of the preferred rate below). These preferred rates, along with error bars that are made wide enough to accommodate the rates calculated with all of the methods given above, are given in Table 21.

Table 21: This table provides the preferred cumulative California seismicity rates and conservative error bars. The preferred rates are averaged Weichert rates multiplied by 1.1 to adjust for the relatively short duration of the historic earthquake catalog and the possibility of earthquakes as large as M 8.3 and accompanying higher seismicity rates over the long term (see text). The error bars given encompass, at 95% confidence, the full range of mean rates and errors that are generated by calculating the seismicity rates with the direct observation, Weichert, averaged Weichert, and averaged-Weichert-with-long-term-corrections methods. Please see the Recommendations section of the text for further explanation.

Mag Range	Full Catalog Rate	Declustered Catalog Rate
M ≥ 5.0	7.5, -3.94, +3.0	4.17, -1.95, +1.67
M ≥ 5.5	2.4, -1.1, +0.95	1.65 ± 0.66
M ≥ 6.0	0.74, -0.34, +0.29	0.65, -0.26, +0.31
M ≥ 6.5	0.23, -0.11, +0.12	0.24 ± 0.11
M ≥ 7.0	0.07, -0.04, +0.06	0.09, -0.04, +0.06
M ≥ 7.5	0.017, -0.013, +0.022	0.02, -0.016, +0.024

7 Adjusting Seismic Moment Rate Estimates for Aftershocks and Declustering

As part of the seismic hazard analysis we will want to compare the average seismicity and seismic moment release rates solved for from the 1850-2006 catalog with the average seismic moment release rate inferred from geologically observed slip. It has long been tradition, however, to use a catalog-derived seismicity rate solved for from the declustered catalog (aftershocks and foreshocks removed). In the field, it is not possible to discriminate which parts of the geologic slip are from mainshocks, aftershocks, or foreshocks. Thus in order to compare the two rates we need to be able to estimate what percentage of the total seismic moment rate is released by aftershocks and foreshocks. This percentage can then be subtracted from the geologically inferred slip rates.

The seemingly most straightforward method to estimate the percentage of seismic moment which is released in aftershocks and foreshocks would be to measure the seismic moment of the entire 1850-2006 catalog, decluster it, and then measure the seismic moment of the earthquakes that remain. The task cannot actually be accomplished this simply, however, because of incompletenesses in the catalog that vary strongly with time and space. Thus in the parts of the catalog that are more incomplete more of the smaller foreshocks and aftershocks will be missing than elsewhere, making for an asymmetric calculation and an overall underestimation of the amount of seismic moment released in aftershocks and foreshocks. Thus instead of direct measurement I try several different statistical approaches.

The aftershock and foreshock definition traditionally used by the National Hazard Maps, and thus one that has been used throughout these calculations, is the one by Gardner and Knopoff (1974). In this definition clusters of earthquakes are captured by drawing boxes in space and time around each earthquake in the catalog, with the size and duration of the box varying with the magnitude of the potential mainshock, according to a chart empirically determined by Gardner and Knopoff (1974). The largest earthquake within each box is retained as the mainshock, while the other earthquakes are removed as either foreshocks or aftershocks. Because this algorithm prefers the larger earthquakes over the smaller ones, it produces different magnitude frequency distributions for the aftershocks/foreshocks and mainshocks. By declustering the relatively complete and accurate 1990-2005 California catalog with this method, I find empirically that the magnitude distribution of the Gardner and Knopoff (1974) mainshocks can be approximated as a Gutenberg-Richter distribution with $b = 0.92$ from M 2 to M 4 and $b = 0.8$ from M 4 to M 7. Above M 7 the full and declustered catalogs are the same. The same is true for $M \geq 7$ earthquakes in the 1850-2006 catalog. Over the long run, however, particularly with multiple mainshocks around M 8, we expect some $M \geq 7$ earthquakes to be classified as aftershocks and foreshocks. Since in the current catalog they are not, however, we do not know how their magnitudes will be distributed. This is a significant problem, for which three potential solutions are suggested below.

1) For the first solution, I assume that $M \geq 7$ earthquakes will never be classified as aftershocks. To implement this solution I first find that the declustered and rounding and magnitude error corrected catalog contains 48% as many $M \geq 5$ earthquakes as the full catalog does. Because of the magnitude-dependent nature of the declustering, this percentage will be higher at lower magnitude cutoffs. The declustered catalog without magnitude and rounding corrections contains 45.6% as many $M \geq 5$ earthquakes as the corresponding full catalog. Based on these values, and on the observed difference in b values, we can write the following expressions for the number of earthquakes in the declustered catalog as a function of magnitude. First I solve for two constants:

$$a_1 = \frac{N_C(5)}{10^{-5}} \tag{10}$$

where $N_C(5)$ is the number of earthquakes in the complete (non-declustered) catalog that are ≥ 5 and

$$a_2 = na_1, \tag{11}$$

where n is equal to $0.1 \times N_D(5)/N_C(5)$ and $N_D(5)$ is the number of $M \geq 5$ earthquakes in the declustered catalog. Thus $n = 0.048$ for the magnitude error corrected case and $n = 0.0456$ for the uncorrected case. Then we have that $N_D(M)$, the number of earthquakes in the declustered catalog that are $\geq M$ is given by:

$$N_D(M) = n10^{0.2M}(N_C(M) + a_1 10^{M_{max}}) - a_2 10^{-0.8M_{max}} \tag{11}$$

for $5 \leq M \leq 7$ and $N_D(M) = N_C(M)$ for $M > 7$. M_{max} is the maximum magnitude in the catalog, the point at which the cumulative Gutenberg-Richter distribution for the sample of magnitudes given in the catalog is truncated. For the 1850-2006 catalog, M_{max} may safely be set around M 8.0 (considering magnitude error). This gives us that from M 5 to M 8, 3% of the total seismic moment will be eliminated when the catalog is declustered.

2) Our second solution is that instead of truncating the aftershock population at M 7 we assume that over the long run the declustered catalog will form a continuous Gutenberg-Richter distribution from M 5 to M 8 with a uniform b value of 0.8. In this case a total of 9% of the moment will be removed by declustering.

3) Finally, we can eliminate assumptions about the maximum magnitude earthquake in California and the magnitude distribution of declustered $7 \leq M \leq 8$ earthquakes by doing a back of the envelope calculation using average aftershock statistical rules. First I will consider aftershocks, and then foreshocks.

For aftershocks, we have Båth's Law, which gives that, on average, the largest aftershock is 1.2 magnitude units smaller than its mainshock (Båth 1965). This rule has been shown to hold for California when aftershocks are counted within two fault lengths and 30 days of the mainshock. The law occurs statistically because the average number of aftershocks/mainshock is equal to $10^{b(M_{main}-1.2-M_{min})}$, while the aftershocks themselves have magnitudes randomly chosen from the Gutenberg-Richter distribution with a b value of 1.0. Here M_{main} is mainshock magnitude and M_{min} is the smallest aftershock counted (Felzer et al. 2002). If I set $M_{min} = M_{main}$ - 3, which allows me to go down to M 5 aftershocks with an M 8 mainshock, then I have on average 63 aftershocks/mainshock. If all of these aftershocks are smaller than the mainshock, then they will follow a Gutenberg-Richter distribution between M_{min} and M_{main}, with a Gutenberg-Ricther a value of $63/10^{-M_{min}}$. We then have that seismic moment is proportional to $10^{1.5M}$. Thus the seismic moment of the mainshock will be proportional to $10^{1.5M_{main}}$, while the seismic moment of the aftershocks will be proportional to $\int_{M_{min}}^{M_{main}} a10^{-M}10^{1.5M}\,dM$. Substituting in a and integrating, this gives that 5% of the total seismic moment of the mainshock-aftershock sequence will be contributed by the aftershocks. Nearly half of this aftershock seismic moment actually comes from the few aftershocks that are larger than the average (e.g. the sequences that have an aftershock larger than M_{main} - 1.2), so although 5% is the overall mean value, in the majority of sequences the aftershock contribution will be smaller.

We next turn our attention to foreshocks. In California, it has been found that about 45% of mainshocks have foreshocks within 3 magnitude units of themselves (Abercrombie an Mori 1996; Felzer et al. 2002), when aftershocks are searched for for 30 days and 5 km

around the mainshock epicenter. It has also been found that the magnitude of the largest foreshock in each sequence is not correlated to the magnitude of the mainshock (Jones and Molnar 1979; Reasenberg 1999), and that the largest foreshock magnitudes follow a uniform distribution. Thus if we consider only the moment of the largest foreshock in each sequence, set $M_{min} = M_{main}$ - 3 = the smallest magnitude foreshock counted, and have that the moment of the mainshock is proportional to $10^{1.5M_{main}}$, then the moment of the foreshocks is proportional to $\int_{M_{min}}^{M_{main}} 0.45 \times 10^{1.5M} \, dM$. Integrating this gives that 11% of the total seismic moment released by foreshocks and mainshocks is released by the foreshocks. Then combining into one equation (Moment of foreshocks + Moment of Aftershocks)/(Moment of foreshocks + Moment of aftershocks + Moment of the Mainshock) gives that a total of 15% of the seismic moment is released by the aftershocks and foreshocks combined.

The difference in the 15% and 9% values that I have calculated above may perhaps be because a Gutenberg-Richter relationship with b =0.8 is not really a perfect fit to the Knopoff-Gardner declustered data. In particular, if we had more $7 \leq M \leq 8$ earthquakes in our data set we might find a different relationship. The 9% figure was also based on functional fits to a limited data set, whereas the 15% was obtained from integrating statistical equations across all sample space – and the majority of the seismic moment will come from the rare large foreshocks and aftershocks. Furthermore, the Knopoff-Gardner declustering routine and the aftershock/foreshock definitions used in the back of the envelope calculation are not exactly the same.

In conclusion, I estimate that the seismic moment released by foreshocks and aftershocks should be between 3% and 15%.

8 Seismicity Rates in Southern and Northern California

In addition to calculating seismicity rates over the entire state and in the specified catalog completeness regions, for reasons of comparison direct observation seismicity rate calculations were performed for the Northern and Southern California polygons illustrated in Figure 8. The same completeness thresholds and regions were used for the calculation of these rates as for the whole state rate calculation. The rates and 95% errors for each region are given in Table 22 and the rates are plotted in Figure 9. Best fit Gutenberg-Richter relationships, with b set to 1.0, are also plotted with the data points. Note that the Gutenberg-Richter line does not fit the $M \geq 5$ data point for either data set; the rest of the points are fit within error bars, but the fit is more secure for Northern than Southern California. The lack of a single Gutenberg-Richter fit to all data points is not necessarily problematic, because, as discussed above, with the direct observation method the rates for different magnitudes are calculated over different time periods in accordance with when the catalog was complete to that magnitude. Thus if there is any variation of seismicity rate with time, the calculated rates at different magnitude levels may not agree with each other. This problem should be less significant for the larger earthquakes, which span more of the catalog (although to the extent that many of these earthquakes are historical they will be associated with higher magnitude and location errors!) and most severe for the smallest earthquakes. In some areas, for example, the

catalog is not considered complete to M 5.0 until 1995, and the 1995-2006 time period has a lower average seismicity rate than other periods.
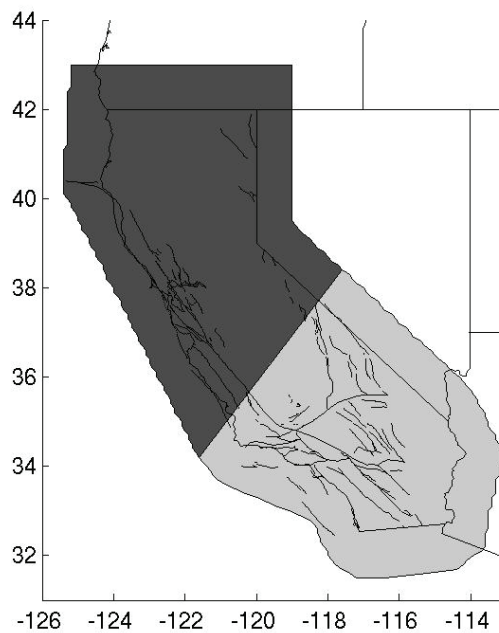
Figure 8: Regions used to calculate direct observation earthquake rates for Southern and Northern California. The rates are given in Table 22 and Figure 9.

Table 22: This table provides direct observation annual earthquake rates for polygons encompassing Southern and Northern California, respectively (the regions plotted are given in Figure 8). For this rate calculation the catalog was declustered and corrections were made for the systematic biases created by magnitude rounding and error. Calculation error is given at the 95% confidence level.

Mag Range	Southern California Rate	Northern California Rate
M ≥ 5.0	1.43, -0.27, +0.3	1.32, -0.26, +0.30
M ≥ 5.5	0.77, -0.21, +0.25	0.56, -0.18, +0.23
M ≥ 6.0	0.31, -0.13, +0.18	0.25, -0.13, +0.20
M ≥ 6.5	0.16, -0.06, +0.09	0.06, -0.03, +0.05
M ≥ 7.0	0.043, -0.03, +0.05	0.03, -0.02, +0.044
M ≥ 7.5	0.014, -0.012, +0.03	0.007, -0.0066, +0.026
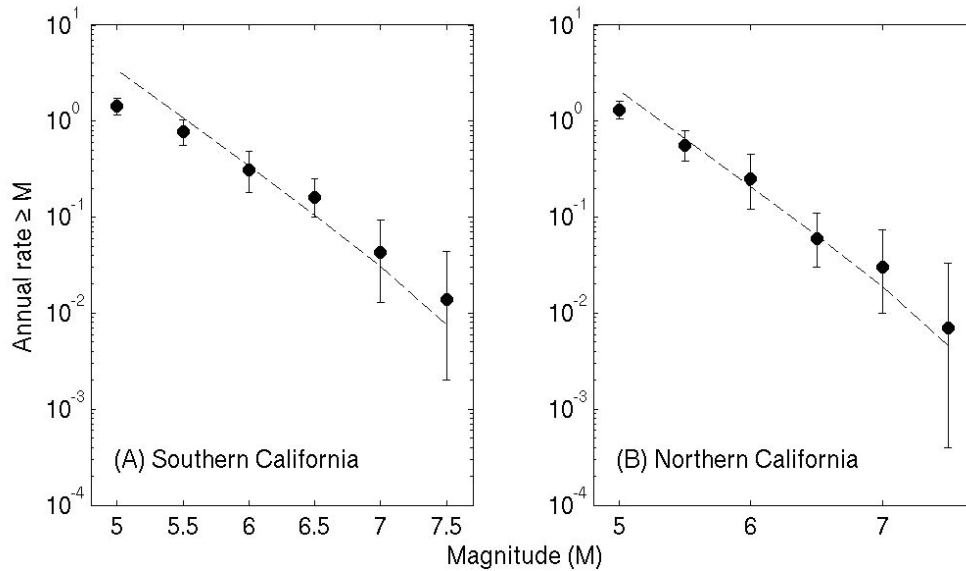
Figure 9: Graphs of the (A) Southern and (B) Northern California direct observation seismicity rates and 95% error bars, as given in Table 22.

9 Recommendations

In this appendix I have presented several different values for the average seismicity rates in California, each based on different assumptions that may be made in the calculations. Has the seismicity rate been constant from 1850-2006, meaning that the seismicity rate may be safely primarily based on only the most recent part of the instrumental catalog (e.g. the pure Weichert method), or was the historical era in fact more active than the instrumental era, thus justifying more weight on its seismicity rate? (e.g. the averaged Weichert method). Is M 7.9 as large as California earthquakes get, or might they get up to M 8.3, which would indicate that the total seismicity rate we observe in 1850-2006 might be biased low, and we need to add on a correction factor to account for this?

At present there are no definite answers to these questions. In the absence of such knowledge, the most general solution is to allow that things may happen. That is, given that we do not really know whether the historical rate was different than the instrumental one or not, we should use the solution, the averaged Weichert calculation, which allows the historical rate to be different. Likewise, since we don't know that California earthquakes can't be as large as M 8.3, we should chose the solution that allows that they may be this large.

At the same time, since any of our seismicity rate solutions may technically be the correct one, while our preferred solution is to go with the rates that allow for higher historical seismicity and the possibility of M 8.3 earthquakes, the preferred error bars should encompass all of the rates and their errors. These preferred rates and all-encompassing error bars are given in Table 21.

10 Seismic Moment Release Rate

In addition to calculating an average annual rate of $M \geq 5$ earthquakes we can also use the historic and instrumental catalogs to estimate the annual rate of seismic moment release by earthquakes in California. This value may be compared to the geologic estimate of the seismic moment rate, with the caveat that some portion of the geologic moment is released aseismically. Thus the seismic moment rate calculated from the catalog should be smaller than that inferred geologically.

We estimate the seismic moment release rate in two ways. The first is by simply converting the raw catalog magnitudes to seismic moments and then adding them in the eight California regions specified in Figure 6. For comparison I calculate two rates for each region, one for 1850-2006 inclusive and one with just the instrumental part of the catalog, from 1932-2006. The regional rates are also summed to estimate rates for the whole state. In doing these sums I do not correct for catalog incompleteness, magnitude rounding, or errors. The calculation of average seismicity rates is strongly influenced by small earthquakes, for which completeness issues are very important, but seismic moment is dominated by the few largest earthquakes, on which catalog incompleteness generally has a much smaller effect. Likewise I chose not to work with magnitude error corrections because the number of earthquakes dominating the calculation is so small. From our catalog we know the standard deviation of the magnitude error for each earthquake, but do not know whether each individual magnitude has been over or under-estimated, or by how much. With a large number of earthquakes this does not matter, as we can calculate that, on average, a magnitude with a given standard deviation has been overestimated by a given amount, and if we correct each earthquake by this amount and compile the results together, we will recover the correct seismicity rate. With a small number of earthquakes, however, doing the correction in this manner produces much more unpredictable results, despite the fact that it should produce improvement on average. So I calculate moment with raw magnitudes, with the caveat that the results are very much estimates, subject to high error. Because the seismic moment release rates are meant to be compared with geologic rate estimates, and mainshock and aftershock induced slip are indistinguishable geologically, I use the full rather than declustered catalog. Regional results are given in Table 23. Estimated statewide totals are 2.23×10^{19} N-m/year for 1850-2006 and 1.29×10^{19} N-m year for 1932-2006.

We also use the Gutenberg-Richter relationship to project a statewide seismic moment release rate estimate from the preferred statewide rate of 7.4 $M \geq 5.0$ earthquakes/year, assuming a maximum magnitude of M 8.3. That is, the Gutenberg-Richter relationship is extended uniformly up to magnitude 8.3 at which point it is truncated, such that a rate may be expressed for $M \geq 8.3$ earthquakes but any earthquake larger than M 8.3 is not allowed to exist. In this case catalog incompleteness and magnitude errors have been corrected for. This gives us an estimate of 3.64×10^{19} N-m/year, an estimate higher than the others because I assume a statewide maximum magnitude larger than any earthquake observed from 1850-2006. If I do the projection instead by using a rate of 6.74 $M \geq 5.0$ earthquakes/year, which is our rate measured with the averaged Weichert method for the full catalog corrected for rounding and magnitude error but not adjusted for the possibility of an M_{max} of 8.3, and combine this with an assumed M_{max} of 7.9 (the largest

38

magnitude actually observed between 1850 and 2006), I recover an average annual statewide seismic moment release rate of 2.07×10^{19} N-m/year.

Table 23: Estimated annual seismic moment release rates, by region, calculated by converting and summing the raw catalog magnitudes over the years indicated. All values are given in N-m/year.

Region	1850-2006	1932-2006
North region	7.87×10^{17}	1.32×10^{18}
San Francisco region	4.92×10^{18}	4.81×10^{17}
Central Coast region	6.31×10^{18}	2.30×10^{17}
Los Angeles region	7.30×10^{17}	4.73×10^{17}
Mojave region	3.08×10^{18}	6.06×10^{18}
Mid region	1.50×10^{18}	3.22×10^{17}
Northeast region	8.14×10^{14}	1.05×10^{15}
Rest of state	4.95×10^{18}	3.99×10^{18}
Total	2.23×10^{19}	1.29×10^{19}

11 Summary

We have taken the following steps to calculate the average seismicity rate from the 1850-2006 California area earthquake catalog:

1. Corrected magnitudes for rounding

2. Calculated and corrected for magnitude error

3. Re-calculated the *b* value from the 1997-2006 part of the California catalog that is complete, statewide, to $M \geq 4$.

4. Calculated completeness thresholds for different regions of the state for each successive 5 year period from 1850-2006 by using newspaper and city locations in the historic era and station locations in the instrumental era.

5. Calculated seismicity rates via direct observation, the Weichert method, and the averaged Weichert method. The first two methods assume that the seismicity rate is constant with time; the last method breaks the seismicity into three time periods and allows for some rate variability between the three eras.

6. Estimated the potential bias and error produced by earthquake clustering if our 156 year catalog length is only a fraction of the average seismic cycle length in California.

Our final result is that for the full corrected catalog, evaluated with the averaged Weichert method, and calculated under the assumption that 1850-2006 is long enough to fairly represent the long term seismicity rate, is 6.8 ± 2.75 $M \geq 5.0$ earthquakes/year, at 98% confidence for the full catalog and 3.8 ± 1.2 $M \geq 5$ earthquakes/year at 98% confidence for the declustered catalog. Alternatively, if we assume that the 1850-2006

catalog is much shorter than a full seismic cycle in California, and that M_{max}, the maximum magnitude that can occur in California, is about 8.3, then simulations indicate that estimates of the long term seismicity rate should be revised up to 7.5, -3.94 +3.0 for the full catalog and 4.17, -1.95 +1.67 for the declustered catalog.

12 Caveats

Shortly before the final deadline for this appendix it was brought to my attention that there were several different errors in the catalog, most notably that the ANSS (Advanced National Seismic System) catalog included some Nevada Test Site explosions as earthquakes in their catalog, which then inadvertently made it into the catalog given in Appendix H. These events almost all range from M 4 to 5.5 and stopped in 1992, so very few of them made it above the completeness threshold used for the ``Rest of State'' catalog, which is set at M 5.4 for 1955-1965 and M 5.3 from 1965-1995. The total effect on the calculated rates was judged to be < 1%, so the rates calculated herein were not updated, although the explosion events were eliminated from the official catalog given in Appendix H.

13 References

Abercrombie, R. E. and J. Mori (1996). Occurrence patterns of foreshocks to large earthquakes in the western united states. *Nature 381*, 303–307.

Aki, K. (1965). Maximum likelihood estimate of b in the formula log n = a - bm and its confidence limits. *Bull. Eq. Res. Inst. 43*, 237–239.

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

Bakun, W. H. (2000). Seismicity of California's North Coast. *Bull. Seis. Soc. Am. 90*, 797–812.

Bakun, W. H. (2006). Estimating locations and magnitudes of earthquakes in Southern California from Modified Mercalli intensities. *Bull. Seis. Soc. Am. 96*, 1278–1295.

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

Båth, M. (1965). Lateral inhomogeneities in the upper mantle. *Tectonophysics 2*, 483–514.

Felzer, K. R., T. W. Becker, R. E. Abercrombie, Göran Ekström, and J. R. Rice (2002). Triggering of the 1999 Mw 7.1 Hector Mine earthquake by aftershocks of the 1992 Mw 7.3 Landers earthquake. *J. Geophys. Res. 107*, 2190, doi:10.1029/2001JB000911.

Gardner, J. K. and L. Knopoff (1974). Is the sequence of earthquakes in southern California, with aftershocks removed, Poissonian? *Bull. Seis. Soc. Am. 64*, 1363–1367.

Gutenberg, B. and C. F. Richter (1944). Frequency of earthquakes in California. *Bull. Seis. Soc. Am. 4*, 185–188.

Hough, S. E. and K. Hutton (2006). Revisiting the 1872 Owens Valley, California, earthquake. in preparation.

Ishimoto, M. and K. Iida (1939). Observations of earthquakes registered with the microseismograph constructed recently. *Bull. Eq. Res. Inst., Univ. Tokyo 17*, 443–478.

Jones, L. M. and P. Molnar (1979). Some characteristics of foreshocks and their possible relationship to earthquake prediction and premonitory slip on faults. *J. Geophys. Res. 84*, 3596–3608.

Kagan, Y. Y., D. D. Jackson, and Y. Rong (2006). A new catalog of Southern California earthquakes, 1800-2005. *Seis. Res. Lett. 77*, 30–38.

Ogata, Y. (1988). Statistical models for earthquake occurrence and residual analysis for point processes. *J. Am. Stat. Assoc. 83*, 9–27.

Reasenberg, P. A. (1999). Foreshock occurrence before large earthquakes. *J. Geophys. Res. 104*, 4755– 4768.

Rhoades, D. A. (1996). Estimation of the Gutenberg-Richter relation allowing for individual earthquake magnitude uncertainties. *Tectonophysics 258*, 71–83.

Schorlemmer, D., J. Woessner, and C. Bachmann (2006). Probabilistic estimates of monitoring completeness of seismic networks. *Seis. Res. Lett. 77*, 233.

Tinti, S. and F. Mulargia (1985). Effects of magnitude uncertainties on estimating the parameters in the gutenberg-richter frequency-magnitude law. *Bull. Seis. Soc. Am. 75*, 1681–1697.

Toppozada, T. R., D. M. Branum, M. S. Reichle, and C. L. Hallstrom (2002). San Andreas fault zone, California: M ≥ 5.5 earthquake history. *Bull. Seis. Soc. Am. 92*, 2555–2601.

Uhrhammer, R. A., S. J. Loper, and B. Romanowicz (1996). Determination of local magnitude using BDSN broadband. *Bull. Seis. Soc. Am. 86*, 1314–1330.

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

Woessner, J. and S. Wiemer (2005). Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty. *Bull. Seis. Soc. Am. 95*, 684–698.