The magnitude distribution of earthquakes near Southern California faults

Morgan T. Page,[1] David Alderson,[2] and John Doyle[3]

Received 14 August 2010; revised 16 September 2011; accepted 28 September 2011; published 16 December 2011.

[1] We investigate seismicity near faults in the Southern California Earthquake Center Community Fault Model. We search for anomalously large events that might be signs of a characteristic earthquake distribution. We find that seismicity near major fault zones in Southern California is well modeled by a Gutenberg-Richter distribution, with no evidence of characteristic earthquakes within the resolution limits of the modern instrumental catalog. However, the *b* value of the locally observed magnitude distribution is found to depend on distance to the nearest mapped fault segment, which suggests that earthquakes nucleating near major faults are likely to have larger magnitudes relative to earthquakes nucleating far from major faults.

Citation: Page, M. T., D. Alderson, and J. Doyle (2011), The magnitude distribution of earthquakes near Southern California faults, *J. Geophys. Res.*, *116*, B12309, doi:10.1029/2010JB007933.

1. Introduction

[2] It is well known that earthquake magnitudes within large regions follow the Gutenberg-Richter (G-R) distribution. The Gutenberg-Richter magnitude distribution relates the cumulative number of earthquakes N above a given magnitude, M, by

$$\log(N) = a - bM, \tag{1}$$

where a and b are constants [*Ishimoto and Iida*, 1939; *Gutenberg and Richter*, 1944]. The b value is generally approximately 1 [*Frohlich and Davis*, 1993], which means, in combination with constant stress drop scaling [e.g., *Aki*, 1972], that the number of earthquakes in a given magnitude range is proportional to the reciprocal of the fault rupture area. For California, $b = 1$ matches the modern catalog well [*Felzer*, 2008; *Hutton et al.*, 2010].

[3] While the Gutenberg-Richter distribution is used to model seismicity in large regions, there is some question as to whether it applies to earthquakes in individual fault zones. The characteristic magnitude distribution [*Wesnousky et al.*, 1983; *Schwartz and Coppersmith*, 1984; *Wesnousky*, 1994] alternatively holds that large earthquakes in major fault zones occur at a higher rate relative to smaller earthquakes than the Gutenberg-Richter distribution would predict. The characteristic magnitude distribution has been suggested in part because of an apparent mismatch between paleoinferred rates of large earthquakes on major faults and rates of smaller earthquakes from the instrumental catalog for a narrow region surrounding the fault. In this work we consider only the modern instrumental catalog, for which hypocenters are known and magnitudes are well characterized. This choice will limit the size of the catalog and therefore the highest magnitudes available; however, including data from many faults throughout California, rather than studying a single fault zone, improves the power of our tests considerably.

[4] The characteristic magnitude distribution is often used in seismic hazard analysis [e.g., *Working Group on California Earthquake Probabilities*, 1990a, 1990b, 1995, 1999; *Field et al.*, 2008]. However, the use of the characteristic earthquake model can lead to some difficulty in matching regional catalog rates. On a state-wide basis, magnitudes are G-R distributed, and it can be difficult to produce an overall catalog that matches the G-R distribution when seismicity on individual faults is modeled with a characteristic distribution. Previous statewide hazard models for California have contained discrepancies between historic earthquake rates and rates given by the model between magnitudes 6 and 7 [*Field et al.*, 1999; *Petersen et al.*, 2000]. Significant tinkering with model parameters has been required to alleviate what has colloquially become known as the "battle of the bulge" [*Field et al.*, 2008].

[5] Inherent in the characteristic earthquake hypothesis is a scaling break between the large and small events on a given fault. Southern California is a good place to look for such a scaling break, if it exists, since there are earthquake catalogs of well-located earthquakes with well-characterized magnitudes and digital models of 3-D fault surfaces in the region. We investigate the magnitude distribution of earthquakes near major fault zones in Southern California to determine if the largest events in fault zones are larger than would be predicted by a Gutenberg-Richter distribution. We begin by examining seismicity near the Parkfield section of the San Andreas Fault before extending our analysis to all major mapped fault zones in Southern California. We also look for changes in the magnitude distribution with distance

[1]U.S. Geological Survey, Pasadena, California, USA.
[2]Operations Research Department, Naval Postgraduate School, Monterey, California, USA.
[3]Control and Dynamical Systems, California Institute of Technology, Pasadena, California, USA.

0148-0227/11/2010JB007933

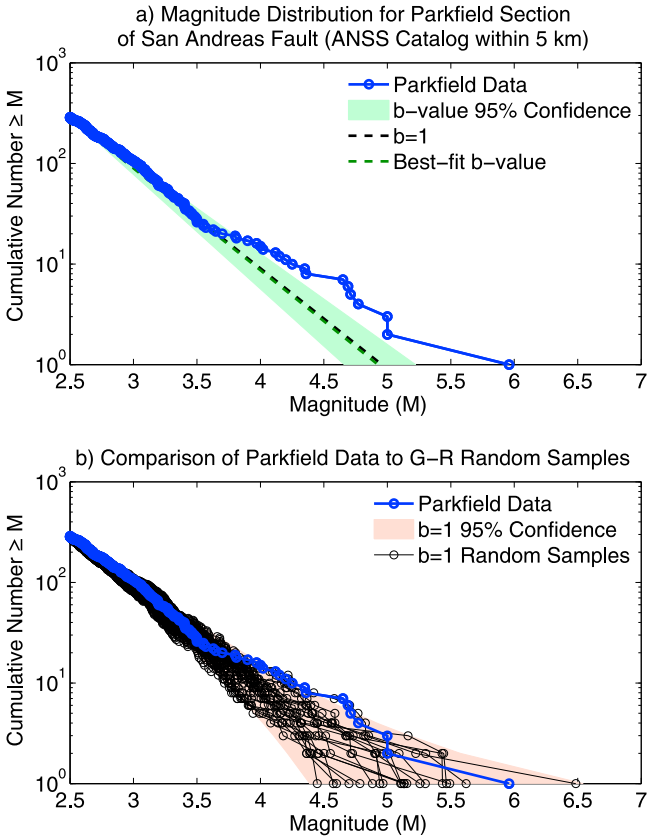

Figure 1. The cumulative magnitude distribution for earthquakes within 5 km of the Parkfield section of the San Andreas Fault is shown in blue. (a) An analysis of the b value error alone could lead to the erroneous conclusion that the largest events in this zone violate Gutenberg-Richter (G-R) behavior. (b) However, random samples drawn from a G-R distribution (black lines) demonstrate considerable scatter. The largest event is within the scatter predicted from random G-R samples and thus does not violate the null hypothesis that Parkfield earthquake magnitudes are drawn from a G-R distribution with a b value of 1.

from major fault zones, to see if the catalog contains differences between major fault seismicity and regional seismicity that are often assumed in seismic hazard models.

2. Seismicity Near Parkfield

[6] The Parkfield section of the San Andreas Fault has been hypothesized to rupture in quasiperiodic "characteristic" events of approximately magnitude 6 [*Bakun and Lindh*, 1985; *Jackson and Kagan*, 2006]. We do not consider time dependence in this study, but focus instead on the magnitude distribution for this fault section. Is there an increase in $M6$ earthquakes near Parkfield, beyond what would be consistent with G-R statistics?

[7] Figure 1a shows the cumulative magnitude distribution for Parkfield section earthquakes. Events are included from the ANSS catalog, 1984–2007, within 5 km of the fault trace, as defined by the Working Group on California Earthquake Probabilities [*Field et al.*, 2008]. Comparing directly to a best fit G-R curve, the G-R distribution appears to severely underpredict the rate of $M6$ earthquakes. Even accounting for b value error (95% confidence bounds for the b value are determined by maximum likelihood [*Aki*, 1965] does not account for this apparent overprediction. It is easy to see from Figure 1a one reason for the development of the characteristic earthquake hypothesis.

[8] However, this simple analysis fails to account for the inherent variability in the tail of the distribution where sampling error becomes important. We can see this by generating 20 random sets of magnitudes drawn from the Gutenberg-Richter distribution, each with the same number of events as the Parkfield data set (Figure 1b). These samples show that variability in the tail of the power law distribution is the rule rather than the exception. In fact, considering this variability, the rate of $M6$ events in Parkfield is consistent with a $b = 1$ G-R distribution at 95% confidence (the exact 95% confidence bounds for each point in the curve are shown with the shading in Figure 1b).

3. Magnitude Distributions in Individual Fault Zones

[9] We systematically extend our analysis to all major mapped faults in Southern California. We assign earthquakes, in 3D, to the nearest fault in the Southern California Earthquake Center (SCEC) Community Fault Model (CFM), version 3.0. This is similar to the rCFM earthquake database [*Woessner and Hauksson*, 2006; *Hauksson*, 2010] which assigns earthquakes to the nearest fault as defined by the rectilinear CFM version 2.5. The SCEC CFM version 3.0 that we use in this analysis contains triangulated, nonplanar fault surfaces. Like the rCFM catalog, we use events from the Southern California Seismic Network from 1981 to 2004 (inclusive), relocated using a double-difference method [*Hauksson and Shearer*, 2005]. We adjust the relocated catalog in two ways: (1) we replace the magnitudes with more recent magnitudes from the Southern California Seismic Network (SCSN), and (2) we add missing events that are in the SCSN catalog but absent from the older rCFM database. This gives a total of 26,479 earthquakes above magnitude 2.5 and within 20 km of the CFM fault segments. Importantly, the revised data set includes the 1992 $M7.3$ Landers earthquake, which is absent from the original relocated catalog. This earthquake is the largest earthquake in the revised data set. The addition of missing events and the use of newer SCSN magnitudes do not significantly change the results we present here.

[10] We separate the earthquakes into bins on the basis of the fault zone to which each is assigned, which is the closest fault in the CFM (we calculate the closest distance in 3D, taking into account the depths of the events and the nonplanar fault sources of the CFM 3.0). The faults themselves are chosen ("segmented") just as they are defined in the CFM. There is certainly some subjectivity in how segments are defined; this cannot be avoided, but we do not personally modify the faults as defined by the community consensus CFM representation. The largest earthquakes in the catalog may indeed rupture multiple segments, or even have hypocenters located by the catalog some distance from the primary rupture. We consider hypocenters only and do not use extended sources or assign large earthquakes to multiple segments; this also prevents data selection on our part as

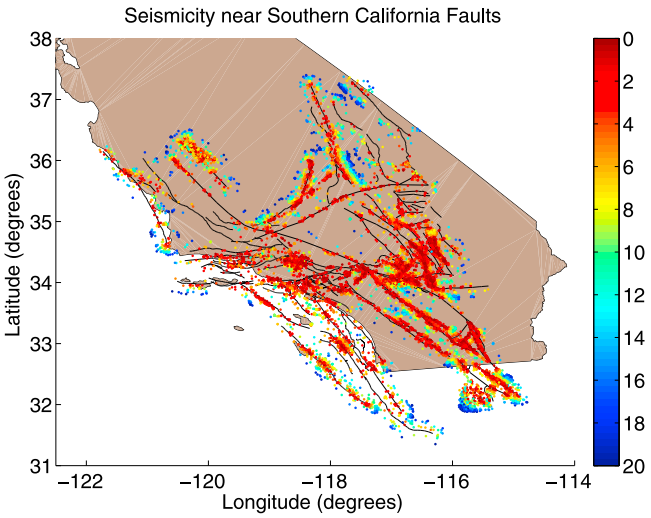

Figure 2. Epicenters for earthquakes (M ≥ 2.5, 1981–2004) within 20 km of the Community Fault Model (CFM) 3.0 faults (surface traces shown with black lines) are colored according to their distance, in kilometers, from the nearest CFM fault plane.

faults involved in particular events are subject to debate (whereas hypocenter distance to the CFM is well defined).

[11] Fault traces and epicenters for the earthquakes in our data set are shown in Figure 2. For the following analysis we include events within 5 km of each fault plane segment and with a magnitude above 2.5 to ensure completeness. Of the 163 faults in the CFM database, 155 faults have associated earthquakes within 5 km and above the minimum magnitude.

[12] On the basis of the magnitude of the largest earthquake in each fault zone bin and the number of earthquakes in that bin, we can obtain a p value. This gives the probability of observing a largest earthquake at least as extreme as in that fault zone, provided that the null hypothesis is correct. Our null hypothesis is that earthquake magnitudes within each fault zone follow a G-R distribution. The p value for a set of N earthquakes with a largest observed magnitude $M_{\text{max\_observed}}$ is

$$p = 1 - \left(1 - 10^{M_{\text{min}} - M_{\text{max\_observed}}}\right)^N, \qquad (2)$$

if the set is complete down to magnitude M_{min} and $b = 1$. Note that equation (2) neglects the upper magnitude cutoff of the G-R relationship. This is valid if there are not enough events in the catalog to "see" this cutoff. Thus our null hypothesis is that the magnitudes are selected from a G-R distribution with a b value of 1, and that the maximum magnitude cutoff is significantly larger than the largest event in our data set (which has a magnitude of 7.3). A b value of 1 is found to fit Southern California seismicity as a whole [*Hutton et al.*, 2010].

[13] We also test a p value statistic that can incorporate spatially variable b values [see, e.g., *Woessner and Hauksson*, 2006], given by $p = 1 - (1 - 10^{b(M_{\text{min}} - M_{\text{max\_observed}})})^N$. The p value statistic given in equation (2) may have greater power in situations where an anomalously large event could be fit by relaxing the b value; however, if the true b value is much greater than 1.0, a truly anomalous event may be missed. We apply the second p value statistic to the binned data, assuming the maximum likelihood b value for each bin. However, we find that this statistic does not significantly change our results. We will thus focus on the results for the simpler statistic in equation (2).

[14] We calculate p values for the seismicity in each fault bin, with $M_{\text{min}} = 2.5$ and assuming $b = 1$. Only 15 of the 155 faults (9.7%) have maximum observed earthquake magnitudes beyond the 90% confidence level (one sided); 12 of the faults (7.7% of the faults) have maximum events beyond the 95% confidence level. Furthermore, only two of the faults have events larger than the 99% confidence interval ($p < 0.01$).

[15] The number of events for each segment versus the magnitude of the largest event is shown in Figure 3a, along with confidence intervals for 90%, 99%, and 99.9%. In Figure 3b we generate similar results for synthetic faults. We synthetically model faults by drawing events randomly from a G-R distribution with a b value of 1 and no upper magnitude cutoff. The synthetic faults are constrained to have the same total-number-of-events distribution as the real faults. The largest-event distribution of the real faults and synthetic faults are not significantly different (see Figure 3). We therefore have a null result: the largest events in CFM fault zones are not larger than would be expected were they pulled randomly from a G-R distribution.

[16] Given that faults form a complex network, and that catalog events have location errors, it is not always clear which earthquakes should be assigned to a given fault. However, varying the distance we include around a fault surface has little effect on the results. Including events within 1 km of the fault plane and 20 km of the fault plane, for example, gives three and two faults, respectively, with $p < 1\%$. This number of fault bins with $p < 1\%$ is not statistically significant (five such fault bins would be required for this to be statistically significant at one-sided 95% confidence).

[17] Even though the magnitude distributions for some faults show events that appear large by eye, this variation is to be expected for power law distributions. The largest magnitudes in this data set are well modeled by a Gutenberg-Richter distribution. The faults with the three largest maximum events are shown in Figure 4. The largest earthquake in Figure 4a, which shows seismicity assigned to the Lavic Lake fault, is the 1999 M7.1 Hector Mine earthquake. This fault segment has 1039 earthquakes, and applying equation (2) gives a p value of 0.026. The 1992 M7.3 Landers earthquake is assigned to the Johnson Valley Fault (Figure 4b), which with 763 events has a p value of 0.012. Figure 4c shows the seismicity nearest to the Santa Monica fault. For this fault zone there are 38 earthquakes; the maximum event in this bin is the 1994 Northridge M6.7 earthquake, which because of its depth and the northward dip of the Santa Monica fault is actually closer to the Santa Monica fault, as defined by the CFM, than to the Northridge thrust fault. Because of the small number of earthquakes in this bin and the large maximum event, the p value for this fault segment is 0.0024. Had the Northridge earthquake been assigned to the Northridge thrust fault, which contains 706 other earthquakes (Figure 4d), many of them aftershocks of the 1994 Northridge main shock, the p value would be 0.044.

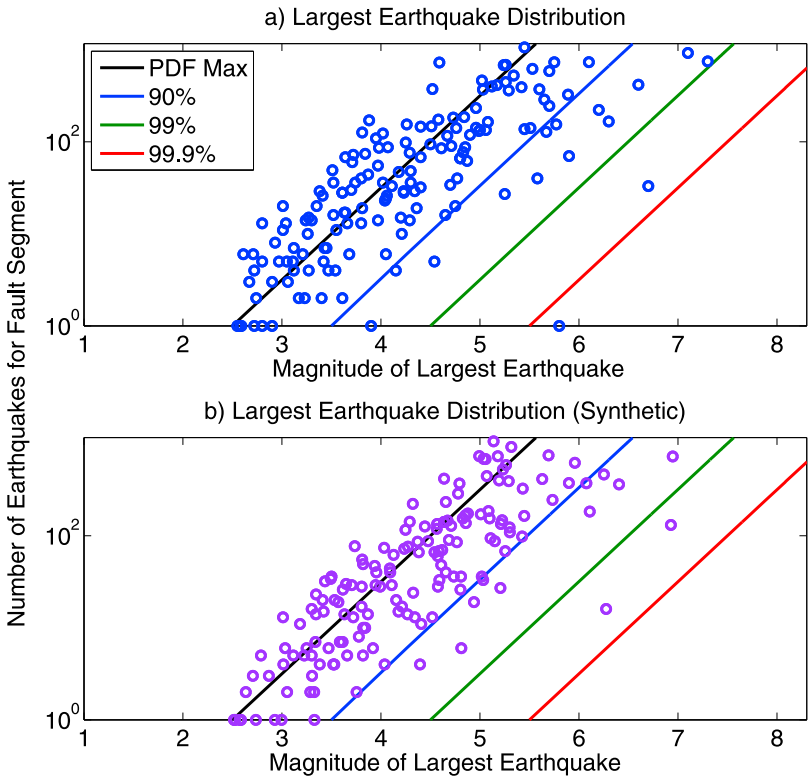

Figure 3. The largest-event distribution for (a) faults in Southern California and (b) synthetic faults with magnitudes following a G-R distribution and the same number-of-events distribution as the real faults. Each circle represents one fault section. The black line shows the G-R extrapolation from small events, which is typically considered the G-R expectation. In fact, as we discuss in the text, this extrapolation, which corresponds to the peak of the maximum event distribution (the black curve shown in Figure 5a), is below the mean expectation for the maximum observed event. The blue, green, and red lines show the 90%, 99%, and 99.9% confidence bounds, respectively, on the magnitude of the largest event, given the number of events assigned to the fault. Only two of the 155 faults are beyond the 99% confidence bounds, which shows that on the whole the faults are not more anomalous in terms of their largest events than synthetic faults with earthquake magnitudes sampled from a G-R distribution.

[18] Besides the Santa Monica fault, there is only one other fault bin that has a p value below 0.01. This is the Clamshell-Sawpit Canyon fault, which contains only one event above magnitude 2.5, the 1991 $M5.8$ Sierra Madre earthquake. While the Sierra Madre main shock is closest to the Clamshell fault as defined by the CFM, the aftershocks of this event are closest to the Sierra Madre fault. Thus the Clamshell fault has only one event but a large maximum event, which gives this fault the largest p value in the data set of 5.0×10^{-4}.

[19] Even though the largest event for both of the faults shown in Figure 4 is much larger than the G-R expectation, the fault zones, considered as a whole, do not contain anomalously large events. We expect approximately this proportion of faults to have maximum observed magnitudes higher than the G-R extrapolation from small magnitudes (and we would expect the lowest p values to occur on the faults that contain the largest earthquakes). This is because the largest-event distribution for a power law is skewed, as shown in Figure 5. While the cumulative rate of small events matches the G-R curve very well, at high magnitudes the individual samples show much more scatter. The last point of each sample, when the magnitudes are plotted cumulatively, is more likely than not to be above a straight-line extrapolation from lower magnitudes. Put another way, the distribution of the last point, shown in black in Figure 5a, is skewed to the right, and both the median and mean of this largest-event distribution are higher than the mode, which corresponds to where the G-R probability distribution function (plotted in red on the same plot) intersects the x axis. (Analytically the black probability distribution function for the largest observed magnitude is equivalent to the derivative of $p(M_{\text{max\_observed}})$ in equation (2).)

[20] Thus it is not unexpected that a magnitude-frequency curve, when plotted cumulatively, will have a perceived characteristic "bump" in the tail that is actually a byproduct of power law statistics. In fact, power law statistics implies that most random draws from a pure G-R distribution will have a maximum observed magnitude that is higher than a log linear extrapolation of the low-magnitude rate (for example, the red line in Figure 5a) appears to predict.

[21] The extreme variability in the tail of a power law is also evident in subsets of the data, as pointed out by *Howell* [1985] and shown in Figure 6. Sorting the data by distance

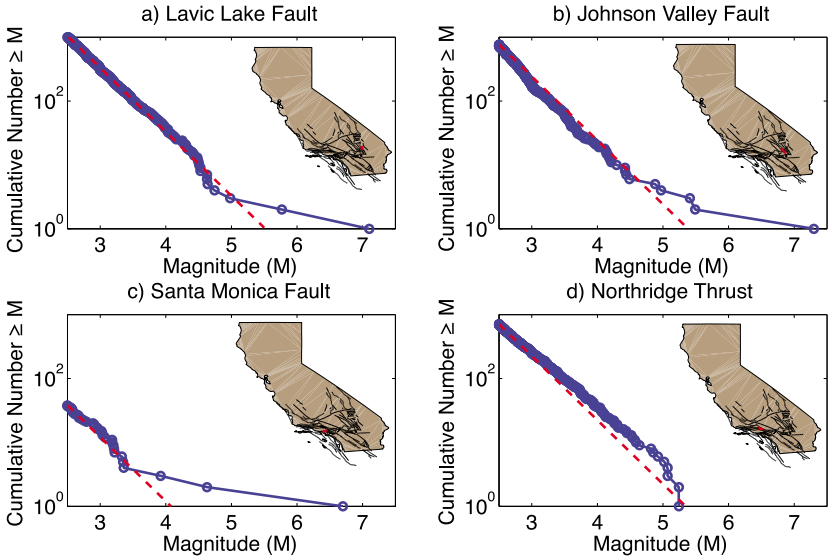

Figure 4. (a–c) Cumulative magnitude-frequency distributions for the three faults with the largest earthquakes in the data set. The G-R extrapolation with $b = 1$ is shown in red. While by eye these faults appear to contain anomalously large events, on the whole the largest-event distribution among the CFM fault sections is consistent with G-R behavior. (d) The 1994 *M*6.7 Northridge earthquake is located closer to the Santa Monica fault than to the Northridge Thrust fault, as defined by the CFM 3.0; this results in a magnitude-frequency distribution for the Santa Monica fault that appears more anomalous because the smaller aftershocks of the Northridge earthquake primarily locate on the Northridge Thrust. It should be noted that picking out the faults with the largest earthquakes will naturally result in distributions that appear to violate G-R behavior; however, this is a result of data selection, and when we analyze all the faults, we find that the largest earthquakes do not violate G-R behavior.

from the CFM faults results in similar variability as the random subsets shown in Figure 5a. In fact, this variability is necessary; that is, one of the 10 subsets, on average, must have a largest event a magnitude unit higher than a G-R extrapolation from lower magnitudes if the total set of earthquakes is to follow a G-R distribution as well (since a data set 10 times the size will have, on average, a largest observed event 1 magnitude unit higher when $b = 1$). This variability in the tail, interestingly, does not decrease with more data until the data set is large enough to be affected by the maximum possible magnitude for that region. This stability of the largest-event distribution with respect to sample size is shown in Figure 5b.

4. Magnitudes on the Fault Versus in the Bulk

[22] To what extent are earthquakes that nucleate on large, mapped faults different than earthquakes that nucleate on smaller faults in the "bulk"? Certainly many of the faults are readily apparent in seismicity locations; however, are the large, mapped faults apparent from other features of the seismicity, namely the magnitude distribution? Although the major faults in California may accommodate much of the strain release, it is another question whether large earthquakes nucleate near the major faults, given the propensity for faults to rupture together in single ruptures [*Wesnousky*, 2008].

[23] The extent to which magnitudes are sensitive to nucleation location has important implications for hazard analysis. If, for example, larger earthquakes are more likely to nucleate closer to mapped faults, this would suggest increased hazard from potential foreshocks located close to major faults relative to other regions [see, e.g., *Agnew and Jones*, 1991]. Furthermore, if the magnitude distribution is sensitive to the size of nearby faults, it also suggests that the G-R magnitude distribution we observe could be an effect of the fault network geometry.

[24] The magnitude distribution of our catalog does, in fact, change with distance from the CFM faults, as shown in Figure 6. The maximum likelihood b value for the 10% of earthquakes that are closest to the fault is 1.08 ± 0.04 at 95% confidence. By contrast the furthest bin from the CFM faults has a b value ranging from 1.15 to 1.24, at 95% confidence. The correlation (we use Pearson's linear correlation in this paper) between the b values for the 10 bins shown in Figure 6a and the distance of the bins from the CFM faults is statistically significant. In addition, we can obtain an extremely statistically significant result which does not rely on binning at all by calculating the correlation between the magnitude of each earthquake in the data set and its distance to the closest CFM fault. This correlation is -0.026 (it is negative because magnitudes tend to increase as distance from the fault decreases). While this correlation may seem small in the absolute sense (as is to be expected given that the b values are not dramatically different) it is, in fact, significant at $p = 3 \times 10^{-5}$, as determined from randomly shuffled versions of the catalog. By increasing the minimum magnitude (see Figure 6b), we can remove enough earthquakes so as to lose statistical significance; however, this does not happen for any $M_{\min} \leq 3.1$.

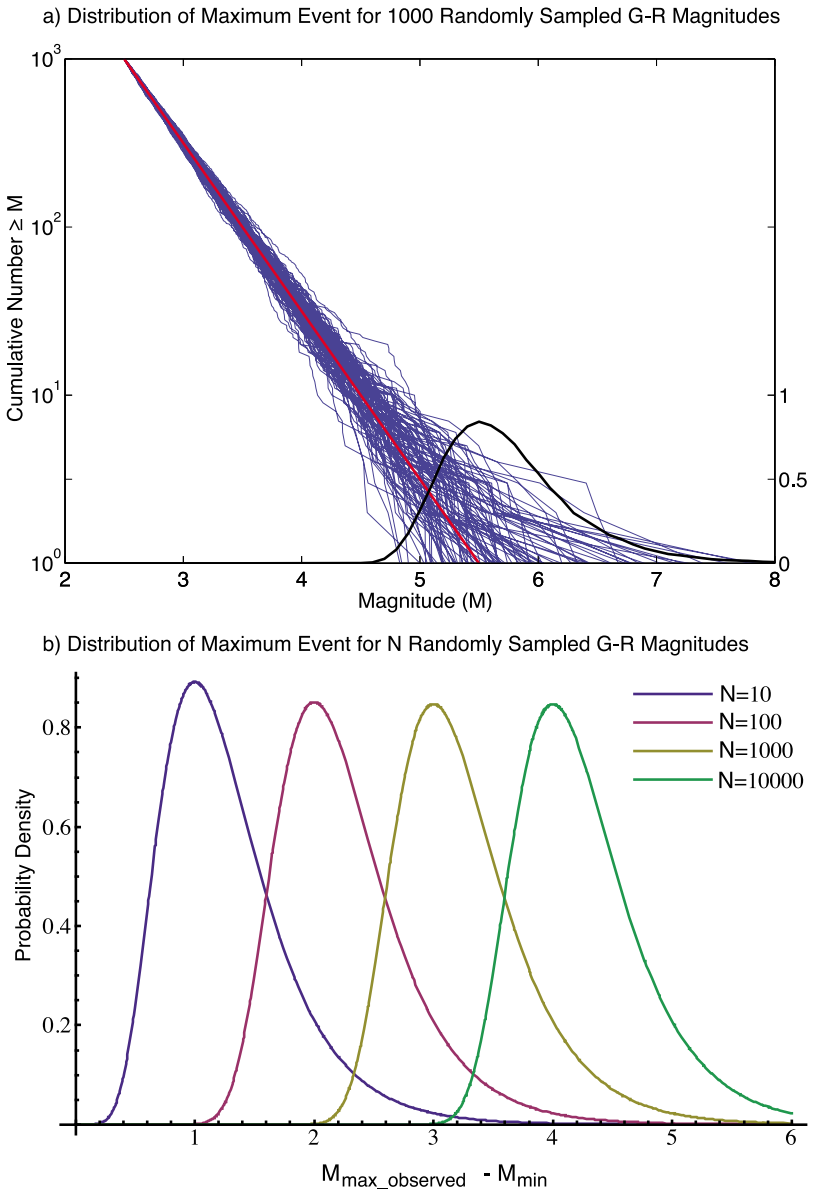

Figure 5. (a) One hundred samples of 1000 random G-R events are shown in blue, and input G-R distribution is shown in red (note that the G-R distribution function extends below the *x* axis). As expected, the largest deviations from the red line are at the high magnitudes. Furthermore, the distribution of the maximum observed event in each sample (black) is a skewed distribution. Even though the maximum of the distribution agrees with the red G-R curve, both the mean and median of the distribution are higher. This shows an important fact of power law statistics: it is more likely than not that the largest event in a G-R sample will be larger than an extrapolation from rates of smaller earthquakes. (b) The variation from sample to sample in the largest event is stable for samples with $N \gtrsim 10$ events. Thus, provided that there are not enough events to sample the maximum possible magnitude, obtaining more data does not reduce the scatter expected in the tail of the distribution.

[25] One effect that could be causing the *b* value change is short-term catalog incompleteness following large earthquakes. Since we are testing for nonstationarity in the magnitude distribution function, it is extremely important that any time intervals included in the catalog be complete down the cutoff magnitude, which we choose to be $M_{\min} = 2.5$. It is well known that because of aftershocks and coda waves, catalogs are incomplete immediately following earthquakes. This phenomenon is known as short-term aftershock incompleteness [*Kagan*, 2004]. We account for this known effect by removing time periods of the catalog following each event; the amount of time removed depends on the magnitude of the event.

[26] Events are removed within a time interval of

$$t = \max \begin{cases} 10^{(M-M_{\min}-4.5)/0.76} & \text{days} \\ 30 & \text{sec} \end{cases}, \qquad (3)$$

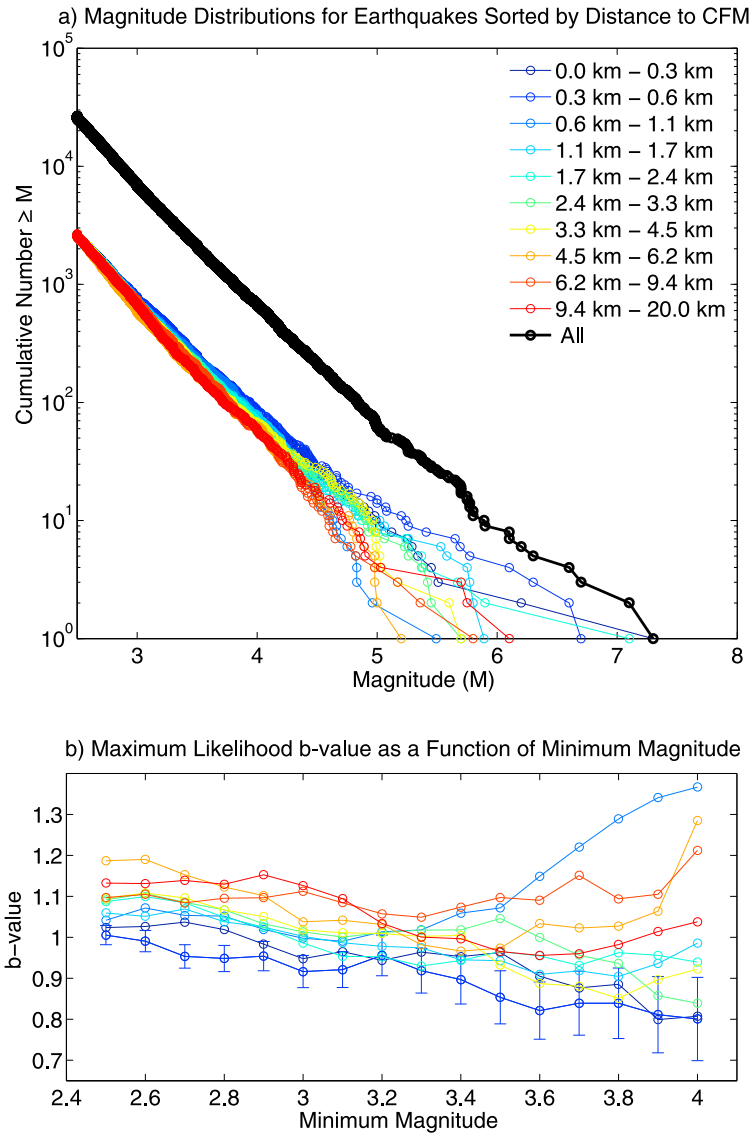

Figure 6. (a) All earthquakes within 20 km of the CFM faults (black) are sorted by distance from the nearest CFM fault and then divided into 10 sets of equal size. While the variation in the tails of the subset magnitude distributions is not unusual, there is variation in the magnitude distribution with distance from the faults that can be seen. The nonuniform *b* values of these subsets as well as the correlation between the magnitude and distance to the CFM are both statistically significant. (b) The *b* values for the 10 subsets as a function of minimum magnitude are shown. The 95% confidence error bars are shown for the closest bin to the faults; these error bars are approximately the same width for the other bins since each bin contains an equal number of earthquakes. The correlation between magnitude and distance from the CFM faults is statistically significant for a minimum magnitude between 2.5 and 3.1.

following each earthquake with magnitude M. The top expression is taken from [*Helmstetter et al.*, 2006]. In addition, for conservatism, we have added a minimum time window removal of 30 sec for all events. This leaves only 2438 earthquakes in the catalog; however, magnitude is still observed to be negatively correlation with distance to the CFM ($p = 3.8 \times 10^{-4}$). The relation of *Helmstetter et al.* [2006] was developed using many aftershock sequences and thus may be an underestimate for particular sequences, as shown by *Woessner et al.* [2011]. However, even a more strict criterion that removes a time interval of $10t$ for each event, as defined by equation (3), leaves a statistically significant negative correlation. Therefore we conclude that the observed *b* value change is not caused by short-term aftershock incompleteness.

[27] It should be noted that the CFM faults have been established using seismicity (in addition to surface traces, seismic reflection profiles, and wellbore data) [*Plesch et al.*, 2007]. If many of the faults in the CFM are drawn such that they are close to the largest earthquakes in the catalog, then it is not surprising that the magnitude distribution would change with distance from the faults. However, this result is just as significant when we restrict our catalog to $M < 4$. Therefore we can conclude that the largest earthquakes in

the catalog ($M \geq 4$) are not driving this result. The best test to ensure that there is no circularity between the development of the CFM and changes in the magnitude distribution with respect to it would be to perform these tests on the portion of the catalog developed post-CFM. Unfortunately, this version of the CFM was last updated in January 2004, which is roughly coincident with the end of our catalog. Newer relocated catalogs are not yet available; in the future an updated catalog will allow one to test whether this effect is observed in seismicity not used to develop the CFM.

5. Discussion

[28] We observe highly statistically significant changes in *b* value with distance from the major faults in Southern California, as defined by the CFM 3.0. If these observed *b* value changes are persistent in time (and, importantly, are not caused by the process by which the CFM is defined) and extend up to high magnitudes then it suggests that earthquakes nucleating near major faults have greater potential to become large earthquakes. Importantly, this *b* value change is a different phenomenon than spatial maximum-magnitude variation; it suggests that even if the maximum possible magnitude is sufficiently large, it is still less likely that an earthquake nucleating 10–20 km from a major fault in Southern California to be large in magnitude compared to an earthquake nucleating within, for example, 1 km from a major mapped fault. This result also suggests that the lengths of local faults influence the magnitude distribution (to the extent that the faults represented in the CFM are the longest and most well defined), which is surprising given that faults are known to "link up" as a fault network, as evidenced by earthquakes that rupture multiple faults.

[29] The more complex, nonplanar fault surfaces as defined in the CFM 3.0 are necessary to clearly see the observed *b* value variation. The changes in *b* values observed between different bins when earthquakes are sorted with distance from the CFM, as shown in Figure 6, are not statistically significant for the same earthquakes sorted by distance to the CFM version 2.5, although the correlation between magnitude and distance from the fault for the unbinned data is borderline statistically significant (this test has greater power).

[30] Interestingly, while we do see evidence of fault geometry influencing the magnitude distribution, we do not see evidence of non-G-R behavior for the largest earthquakes. Could faults have characteristic behavior beyond the magnitudes available in the instrumental catalog? We limited our analysis to the modern catalog, which does limit our ability to constrain the magnitude distribution at magnitudes greater than 7. However, if the regional G-R relationship is a result of the power law distribution of fault lengths, as suggested by *Wesnousky* [1999], this suggests that smaller faults have characteristic behavior at smaller magnitudes. Our data set contains segments from long faults such as the San Andreas fault, as well as far smaller faults; still, we see no evidence of anomalously large events for any of the fault zones.

[31] It seems that fault geometry does influence the magnitude distribution, but through the *b* value (which changes the rates at all magnitudes) rather than at characteristic magnitudes. Magnitude distributions may appear "characteristic" by eye, but this is, in fact, due to the large intrinsic variability of samples from a power law distribution. This work shows that the available data is consistent with the null hypothesis of G-R scaling near major faults.

6. Conclusion

[32] Many seismic hazard products rely on the assumption that earthquakes that nucleate on a major fault are different (e.g., likely to be larger in magnitude) than those that nucleate "in the bulk" (i.e., on smaller, unmapped faults). We do, in fact, see changes in the magnitude distribution with distance from the major faults, however, they are not of the "characteristic" variety typically included in such models. We see evidence for changes in *b* value but do not see evidence for non-G-R behavior for the largest events. Still, these changes in *b* value, although small, can have a large effect on rates at high magnitudes and are therefore important for seismic hazard analysis.

[33] **Acknowledgments.** We would like to thank J. Woessner and E. Hauksson for providing their rCFM catalog and K. Felzer and J. M. Carlson for many helpful discussions.

References

Agnew, D. C., and L. M. Jones (1991), Prediction probabilities from foreshocks, *J. Geophys. Res.*, *96*, 11,959–11,971, doi:10.1029/91JB00191.

Aki, K. (1965), Maximum likelihood estimate of *b* in the formula log $N = a - bM$ and its confidence limits, *Bull. Earthquake Res. Inst. Tokyo Univ.*, *43*, 237–239.

Aki, K. (1972), Earthquake mechanism, *Tectonophysics*, *13*(1–4), 423–446, doi:10.1016/0040-1951(72)90032-7.

Bakun, W. H., and A. G. Lindh (1985), The Parkfield, California, earthquake prediction experiment, *Science*, *229*(4714), 619–624, doi:10.1126/science.229.4714.619.

Felzer, K. R. (2008), Appendix I: Calculating California seismicity rates, *Tech. Rep. OFR 2007-1437I*, U.S. Geol. Surv., Reston, Va.

Field, E. H., D. D. Jackson, and J. F. Dolan (1999), A mutually consistent seismic-hazard source model for southern California, *Bull. Seismol. Soc. Am.*, *89*(3), 559–578.

Field, E. H., et al. (2008), The Uniform California Earthquake Rupture Forecast, version 2 (UCERF 2), *Tech. Rep. OFR 2007–1437*, U.S. Geol. Surv., Reston, Va.

Frohlich, C., and S. D. Davis (1993), Teleseismic *b* values; or much ado about 1.0, *J. Geophys. Res.*, *98*(B1), 631–644, doi:10.1029/92JB01891.

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

Hauksson, E. (2010), Spatial separation of large earthquakes, aftershocks, and background seismicity: Analysis of interseismic and coseismic seismicity patterns in southern California, *Pure Appl. Geophys.*, *167*, 979–997, doi:10.1007/s00024-010-0083-3.

Hauksson, E., and P. Shearer (2005), Southern California hypocenter relocation with waveform cross-correlation, Part 1: Results using the double-difference method, *Bull. Seismol. Soc. Am.*, *95*, 896–903, doi:10.1785/0120040167.

Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2006), Comparison of short-term and time-independent earthquake forecast models for Southern California, *Bull. Seismol. Soc. Am.*, *96*, 90–106, doi:10.1785/0120050067.

Howell, B. F. (1985), On the effect of too small a data base on earthquake frequency diagrams, *Bull. Seismol. Soc. Am.*, *75*, 1205–1207.

Hutton, K. L., J. Woessner, and E. Hauksson (2010), Earthquake monitoring in Southern California for seventy-seven years, *Bull. Seismol. Soc. Am.*, *100*, 423–446.

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

Jackson, D. D., and Y. Y. Kagan (2006), The 2004 Parkfield earthquake, the 1985 prediction, and characteristic earthquakes; lessons for the future, *Bull. Seismol. Soc. Am.*, *96*, S397–S409, doi:10.1785/0120050821.

Kagan, Y. Y. (2004), Short-term properties of earthquake catalogs and models of earthquake source, *Bull. Seismol. Soc. Am.*, *94*, 1207–1228, doi:10.1785/012003098.

Petersen, M. D., C. H. Cramer, M. S. Reichle, A. D. Frankel, and T. C. Hanks (2000), Discrepancy between earthquake rates implied by historic earthquakes and a consensus geologic source model for California, *Bull. Seismol. Soc. Am.*, *90*, 1117–1132, doi:10.1785/0119990008.

Plesch, A., et al. (2007), Community Fault Model (CFM) for Southern California, *Bull. Seismol. Soc. Am.*, *97*, 1793–1802, doi:10.1785/0120050211.

Schwartz, D. P., and K. J. Coppersmith (1984), Fault behavior and characteristic earthquakes: Examples from the Wasatch and San Andreas fault zones, *J. Geophys. Res.*, *89*, 5681–5698.

Wesnousky, S. G. (1994), The Gutenberg-Richter or characteristic earthquake distribution, which is it?, *Bull. Seismol. Soc. Am.*, *84*, 1940–1959.

Wesnousky, S. G. (1999), Crustal deformation processes and the stability of the Gutenberg-Richter relationship, *Bull. Seismol. Soc. Am.*, *89*, 1131–1137.

Wesnousky, S. G. (2008), Displacement and geometrical characteristics of earthquake surface ruptures: Issues and implications for seismic-hazard analysis and the process of earthquake rupture, *Bull. Seismol. Soc. Am.*, *98*, 1609–1632.

Wesnousky, S. G., C. H. Scholz, K. Shimazaki, and T. Matsuda (1983), Earthquake frequency distribution and the mechanics of faulting, *J. Geophys. Res.*, *88*, 9331–9340.

Working Group on California Earthquake Probabilities (1990a), Probabilities of large earthquakes occurring in California on the San Andreas fault, *Tech. Rep. OFR 88–398*, U.S. Geol. Surv., Reston, Va.

Working Group on California Earthquake Probabilities (1990b), Probabilities of large earthquakes in the San Francisco Bay Region California, *U.S. Geol. Surv. Circ.*, *1053*, 51 pp.

Working Group on California Earthquake Probabilities (1995), Seismic hazards in southern California: Probable earthquakes, 1994–2024, *Bull. Seismol. Soc. Am.*, *85*, 379–439.

Working Group on California Earthquake Probabilities (1999), Earthquake probabilities in the San Francisco Bay Region: 2000–2030—A summary of findings, *Tech. Rep. OFR 99–517*, U.S. Geol. Surv., Reston, Va.

Woessner, J., and E. Hauksson (2006), Associating Southern California seismicity with late Quaternary faults: Implications for seismicity parameters, paper presented at Southern California Earthquake Center Annual Meeting, Palm Springs, Calif.

Woessner, J., S. Hainzl, W. Marzocchi, M. J. Werner, A. M. Lombardi, F. Catalli, B. Enescu, M. Cocco, M. C. Gerstenberger, and S. Wiemer (2011), A retrospective comparative forecast test on the 1992 Landers sequence, *J. Geophys. Res.*, *116*, B05305, doi:10.1029/2010JB007846.

D. Alderson, Operations Research Department, Naval Postgraduate School, Monterey, CA 93943, USA. (dlalders@nps.edu)

J. Doyle, Control and Dynamical Systems, California Institute of Technology, 103 Steele, M/C 107–81, Pasadena, CA 91125, USA. (doyle@cds.caltech.edu)

M. T. Page, U.S. Geological Survey, 525 South Wilson Ave., Pasadena, CA 91106-3212, USA. (pagem@caltech.edu)