Anderson, J. G., Cotton, F., Bindi, D. (2021): A ground motion based procedure to identify the earthquakes that are the most relevant for probabilistic seismic hazard analysis. - Earthquake Spectra, 37, 2, 762-788.

https://doi.org/10.1177/8755293020981987

A ground motion based procedure to identify the earthquakes that are the most relevant for probabilistic seismic hazard analysis

John G Anderson, M.EERI[1] , Fabrice Cotton[2,3], and Dino Bindi[2]

Abstract

A method is proposed to identify within seismic catalogs those earthquakes that are most relevant to the seismic hazard. The approach contrasts with the classical approach to decluster the seismic catalog with the expectation that the remaining main shocks will be the relevant events for the seismic hazard analysis. We apply a time window like in the window declustering approach of Gardner and Knopoff, but the time window is motivated by relevance to engineering. A ground motion criterion replaces the spatial window. An event in the time window is included in the ''Maximum Shaking Earthquake Catalog (MSEQ catalog)'' if the median ground motion at its epicenter exceeds the predicted median ground motion there from the main shock, using a locally appropriate ground motion prediction equation. Ground motion can be measured by any parameter that is estimated by a ground motion prediction equation. We consider peak acceleration and spectral amplitude (SA) at periods of 0.2, 1.0, and 3.0 s. The longer period parameters systematically remove more small events. The purpose is not to produce a declustered catalog, in which each group of physically related earthquakes is represented by its largest event. Statistical properties of the MSEQ catalog somewhat resemble the corresponding declustered catalog in three tested regions, but the MSEQ catalogs all retain more large-magnitude earthquakes. The MSEQ catalog may better represent the potential hazard in a region, and thus might be considered as an alternative to a declustered catalog in developing the seismicity model for probabilistic seismic hazard analysis.

Keywords

Probabilistic seismic hazard, declustering, seismicity, Maximum Shaking Earthquake Catalog, most significant earthquake, hazard curve, Gardner and Knopoff, Poissonian seismicity catalog, Kolmogorov–Smirnov critical values

Introduction

Probabilistic seismic hazard analysis (PSHA) has become a very common tool to characterize the seismic hazard. The outcome of PSHA is a hazard curve for a site, giving either the rate or the probability of exceedance of an amplitude of ground motion, considering all of the most relevant earthquakes. A central assumption is that a true hazard curve exists for any location (e.g. Anderson and Biasi, 2016; Marzocchi and Jordan, 2014, 2017). The hazard curve is estimated by the use of fundamental methods of probability theory, combining a seismicity model with a model for the ground motion.

The term ''seismicity model'' is used here to describe the entire set of relevant earthquakes, including their locations, magnitudes, and annual occurrence rates. The criteria defining the relevance of an earthquake for PSHA studies are usually not defined or clarified in modern PSHA studies. Most PSHA studies involve a declustering step which removes foreshocks and aftershocks to build a declustered catalog which will be used to develop the seismicity model. The classical justification of this declustering step is that PSHA considers that seismicity rates should be Poissonian (e.g. CEUS-SSC, 2012, Chapter 2; Frankel et al., 1996; Petersen et al., 2008). To achieve this, dependent earthquakes (i.e. foreshocks and aftershocks) should be removed so that the retained earthquakes follow a Poisson distribution. This declustering is intended to achieve a stable estimate of background rates as the resulting catalog contains events which can be considered as independent. Spatial biases in areas that have experienced clusters in the past with respect to regions that did not are also reduced by declustering, but generally remain, and are dealt with by the use of area sources or spatial smoothing (e.g. McGuire, 2004).

It is generally thought that this declustering process will naturally remove earthquakes that have a negligible contribution in terms of ground shaking and earthquake hazard, and therefore that declustering has a negligible effect for engineering practice (e.g. McGuire, 2004). In other words, declustering processes based on the goal to select independent earthquakes are thought to be an efficient way to exclude those earthquakes from the catalog that are not relevant for hazard evaluations.

[1] Nevada Seismological Laboratory, University of Nevada, Reno, NV, USA
[2] GFZ German Research Center for Geosciences, Potsdam, Germany
[3] Institute of Geosciences, University of Potsdam, Potsdam, Germany

Corresponding author:
John G Anderson, Nevada Seismological Laboratory, University of Nevada, Reno, NV 89557, USA. Email: jga.seismo@gmail.com

Indeed, foreshocks and aftershocks of large earthquakes are not generally considered relevant to the seismic hazard estimate. This is because by definition they have smaller magnitudes, generally cause weaker ground motions, and for the most part do not contribute to the damage.

A widely used method of declustering is that of Gardner and Knopoff (1974). This method (here referred to as the ''GK method'') gives a rule for the radius of an aftershock zone as a function of magnitude and a second rule for the duration of the aftershock sequence also as a function of magnitude (Table 1). Earthquakes inside of this time-distance window are removed from the catalog. This conceptual model is widely used, with local adjustments to the time and spatial windows.

Several other methods for recognition, modeling, and removal of aftershocks have been proposed in the seismological literature. These include Savage (1972), Prozorov and Dziewonski (1982), Ogata (1988), Reasenberg (1985), Frohlich and Davis (1990), Molchan and Dmitrieva (1992), Ogata (1999), Zaliapin et al. (2008), Marsan and Lengline (2008), Console et al. (2010), Zaliapin and Ben-Zion (2011), and Anderson and Nanjo (2013). In these papers, the concept of clusters includes all of the events that are part of the geologically brief release of accumulated elastic strain in which the main shock is the climax. The diversity of approaches arises because, as described for example by Epistemic-Type Aftershock Sequence (ETAS) models (e.g. Ogata, 1988. 1999), every earthquake has a probability that it is independent. Thus, it is impossible to unambiguously classify every earthquake as either a dependent or independent event. It follows, as noted by Zaliapin and Ben-Zion (2020), that the declustering method can usefully be tailored for the application. If the classification of each earthquake is recognized as subject to epistemic uncertainty, then this uncertainty should be minimized by aligning the declustering procedure with the requirements of hazard studies.

To achieve this alignment, we propose that an essential characteristic of the earthquake catalog used to build the seismicity model for PSHA is consistency with the ''true'' hazard curve. Recent earthquakes sequences (discussed subsequently), and several studies (e.g. Marzocchi and Taroni, 2014) show that the key assumptions of declustering methods may not be consistent with this proposition. We suggest to reverse the classical assumption (''selection of independent earthquakes is an efficient way to identify relevant earthquakes'') and develop a method with the goal to select the most relevant earthquakes, regardless of whether or not they are independent. The resulting catalog can be compared with the catalogs that result from declustering, and tested for Poissonian behavior, and we carry out these comparisons and tests in this article, but our goal is not to obtain a declustered or Poissonian catalog. Regardless of the result of that comparison, one can evaluate whether the catalog of most relevant earthquakes might be used, instead of a declustered catalog, for probabilistic seismic hazard assessment purposes.

Indeed, the criteria defining the relevance of an earthquake for PSHA studies are usually not defined or clarified in modern PSHA studies. From an engineering point of view and seismic building code perspective, relevant earthquakes are earthquakes that cause ground motions sufficient to cause significant damage of intact buildings. These engineering considerations are then essential. They explain why we keep only earthquakes of magnitude larger than $M > 4.5$ for the hazard computation and why most of the aftershocks and foreshocks (which are producing smaller shaking than the main shocks) should be removed. Unimportant earthquakes should, however, not be recognized according to their magnitude, time, and location only; they should rather be defined as events which are less damaging than main shocks (for intact buildings).

There are numerous examples of aftershocks that could cause considerable direct (and not only cumulative) damage. We discuss four such earthquake sequences here. First, on June 28, 1992, the Mw7.27 Landers, California, earthquake at (11:57 coordinated universal time [UTC]) was followed by the Mw6.48 Big Bear earthquake (15:05 UTC, 3.1 h later) on a completely different fault about 36 km west of the Landers rupture. The Mw6.14 Joshua Tree earthquake (April 23, 1992) preceded the Landers earthquake by 66 days. The Joshua Tree earthquake, whose 10 to 12 km long aftershock zone is south of, but continuous with the aftershocks of, the Landers rupture, apparently occurred on a different fault that is separated from the Landers rupture by the east–west Pinto Mountain fault (Sieh et al., 1993). Using the GK parameters (Table 1), these three earthquakes are represented by only the Landers event in the National Seismic Hazard Model declustered catalog (Petersen et al., 2014). Second, the November 12, 1999 Duzce, Turkey earthquake (M7.2) extended the rupture of the August 17, 1999 Izmit, Turkey earthquake for 30 km eastward along the North Anatolian Fault system, and caused significant additional damage and loss of life (e.g. Anderson et al., 2000). The epicentral separation of about 110 km is large enough that the Duzce earthquake would be considered an independent event by the GK criterion (Table 1), but if the Izmit epicenter had been about 30 km farther east along the August rupture, the Duzce event would be considered an aftershock, highlighting an arbitrary property of the procedure. Our third example comes from New Zealand, where the Mw7.1 Darfield earthquake (September 4, 2010) was followed 171 days later (February 22, 2011) by the destructive Mw6.1 Christchurch earthquake, with epicenter 42 km to the east. By the GK declustering parameters (Table 1), the Christchurch earthquake it is part of the Darfield cluster. Finally, the central Italy seismic sequence of 2016 consisted of three main events: August 24, Mw6.0 (Amatrice); October 26, Mw5.9 (Ussita); and October 30, Mw6.5 (Norcia); and over 55,000 smaller events in the first 5 months (Pizzi et al., 2017). Each of these three largest earthquakes would have been a disaster in its own right. The second epicenter is about 26 km away and 63 days following the first, compared to the GK thresholds of 54 km and 510 days (Table 1). The October 30 epicenter was located between the first two, and its larger magnitude makes the Norcia event the main shock and sole representative of the sequence, using the GK approach to declustering.

Table 1. Distance and time windows refrenced in this paper.

	M_W											
	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0
R_{GK}, km[1]	19.5	22.5	26	30	35	40	47	54	61	70	81	94
T_{GK}, d[2]	6	11.5	22	42	83	155	290	510	790	915	960	985
T_{EU}, d[3]	14.6	27.2	48.4	82.6	137	220	346	533	807	924	950	977

1. Distance criterion used by Gardner and Knopoff (1974) and the U. S. Geological Survey.
2. Time criterion used by Gardner and Knopoff (1974) and the U. S. Geological Survey.
3. Time criterion used to decluster the European catalog (see text).

From the seismological perspective, all four of these earthquake sequences represent a set of related earthquakes, and thus in the context of the physics of earthquakes it is legitimate that each be considered a cluster identified by their main shock. From the engineering perspective, however, each of these sequences include foreshocks or aftershocks that cause ground motions that controlled the local damage. Thus, classical declustering methods can result in the removal of some events that, while physically related to others, would also have important engineering implications. We explore an alternative procedure that leads to a catalog that includes all of these significant shocks. We call the resulting catalog a ''Maximum Shaking Earthquake Catalog'' (MSEQ catalog) to distinguish it from declustering that is focused on identifying clusters based on a common geophysical event.

Developing the MSEQ catalog

Developing an MSEQ catalog is straightforward. Every earthquake in the catalog is initially labeled as a ''main shock.'' Subsequently, starting with the largest earthquake and working downward in magnitude, every earthquake afterward in the time window specified for that magnitude is identified for review. The review is based on the amplitude of ground motion at the epicenter of the reviewed event. If the estimated peak acceleration, for instance, of the shock under review is smaller than the estimated peak acceleration from the main shock for that location, then it is flagged as a ''subshock.'' Otherwise, it continues to be considered a main shock. We note that by this method, the MSEQ process can recognize subshocks in old and in incomplete catalogs.

Obviously, the peak accelerations have not been measured, so they are estimated by a ground motion prediction equation. In the implementation described in this article, each earthquake is tested by four ground motion parameters: peak acceleration (PGA), and pseudo-acceleration response spectral amplitude (SA) for oscillators with periods of 0.2, 1.0, and 3.0 s. This results in four different MSEQ catalogs, which are identified as MSEQPGA, MSEQ-SA (0.2 s), MSEQ-SA (1.0 s), and MSEQ-SA (3.0 s), respectively. Of course, other ground motion parameters can be chosen.

Time window

Consistent with our intent to develop a catalog in a manner consistent with user applications, we sought application examples from multiple sources. Here are three examples. Insurance Company A states, in its earthquake insurance rider, that any earthquake within 3 days of the main shock will be considered an aftershock. A second insurance company, which we will call Insurance Company B, has a time window that is 15 days long, independent of magnitude. A third opinion, from a civil engineer, is that an aftershock is any event that occurs before a structure is repaired.

These three definitions of aftershocks are obviously very different. The MSEQ catalog can be developed for any of these time windows. This study uses the GK time window, and a modified version that has been preferred for European applications (Table 1). The GK window increases with magnitude, from ~10 days for M = 3 events, to ~150 days for M = 5, and ~3 years for M = 7. We suggest that this window may be considered a proxy for the time required for repairs after an earthquake. It also provides continuity with previous probabilistic seismic hazard studies where it has seen extensive use, and isolates the effect of the ground motion prediction equation as the alternative spatial window. The short time windows used by the insurance companies can be implemented and explored in a later study. It is expected that these shorter windows will remove fewer earthquakes than the GK window, and consequently we expect that the hazard estimates using those catalogs would be higher than hazard estimates using a GK declustering time window.

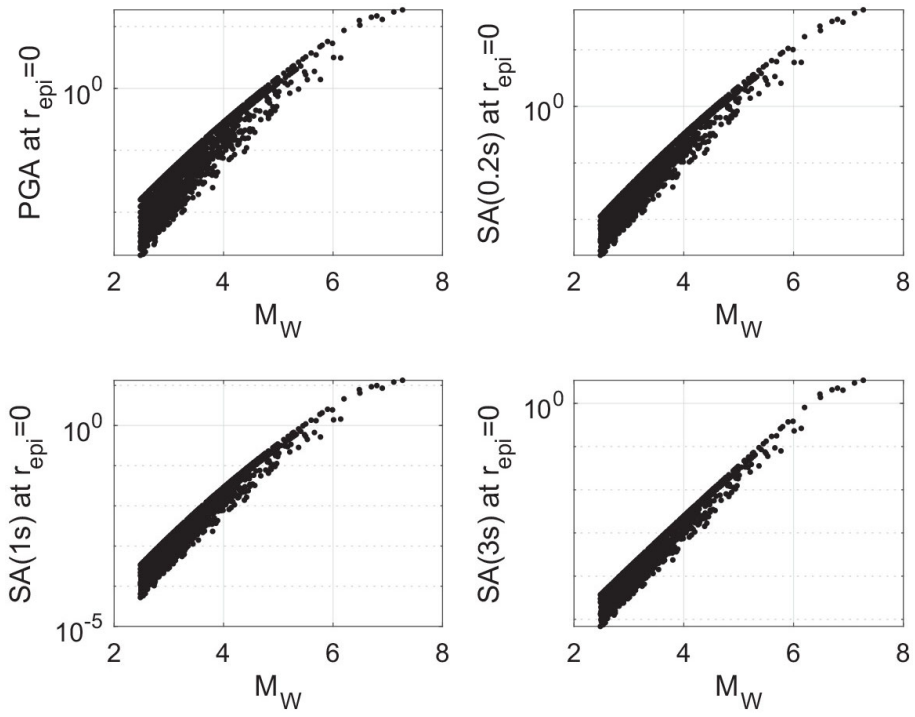

Figure 1. *Median estimates of PGA, SA (0.2 s), SA (1.0 s), and SA (3.0 s), as labeled in respective frames, at the epicenter of each earthquake in the USGS C2 catalog in the Landers region. Earthquake locations are shown in Figure 2d. Each estimate uses V_{S30} = 800 m/s and is based on Bindi et al. (2017). These parameters depend on the hypocentral distance, so the variable depths result in multiple values for each magnitude.*

Ground motion prediction equation

The model of Bindi et al. (2017) was selected for this application. This model is based on the NGA West 2 database (Ancheta et al., 2014) but uses a simple functional form. It is based on the hypocentral distance, and thus is consistent with the GK distance measure, making it convenient for this application. Using this model, Figure 1 shows peak accelerations at the epicenters of all earthquakes in the USGS catalog in the Landers test area, as will be discussed subsequently. The Bindi et al. (2017) model is based on earthquakes with Mw > 3.5, so it is extrapolated to smaller magnitudes to generate this figure. In this extrapolation, the logarithm of the peak acceleration is nearly proportional to magnitude at the small magnitudes. This is reasonable considering the fundamental definition of the magnitude scale (Richter, 1958): $M_L = \log A - \log A_0$.

Estimating the ground motion based on the epicentral distance of a large earthquake, for which the rupture can be tens or even hundreds of kilometers in length, is not entirely consistent with the goal of MSEQ catalog development to identify subshocks based on whether or not the small earthquake has a larger motion than the main shock. At sites close to the fault but far from the epicenter, some subshocks may not be recognized as such. To deal with this, the catalog is supplemented with additional ‘‘virtual hypocenters’’ for the largest earthquakes. The ‘‘virtual hypocenters’’ are treated as independent earthquakes in the MSEQ algorithm. Thus, for instance, earthquakes that occurred in the selected time window after the 1992 Landers, California, earthquake (Mw7.27) are compared with ground motions from an Mw7.27 earthquake in five different locations within the aftershock zone. If the reviewed earthquake has motions smaller than expected motions from any of the virtual hypocenters, then it is not included in the MSEQ catalog. The alternative, of course, would be to use the distance to the identified fault rupture, together with a ground motion prediction equation (GMPE) based on rupture distance. We considered this alternative, but the use of the virtual hypocenters, identified by an interactive review of the catalog, together with the Bindi et al. (2017) model based on hypocentral distance is more easily implemented. Our judgment is that the greater precision of using the actual rupture would not make much practical difference, but of course it would be more rigorous if this method is actually applied in a PSHA using a GMPE that is based on rupture distance.

To avoid confusion in calculating magnitude–frequency relations from the MSEQ catalog, enhanced with virtual hypocenters, we introduce a parameter N*. Although inspired in part by the parameter N* introduced in CEUS-SSC (2012, Chapter 3), the application here is distinctly different (but can be made compatible). For most earthquakes, N* = 1. However, when a large earthquake is represented by n multiple locations along the fault, each of these locations are included in the catalog with N* = 1/n. With this numerical procedure, computing Gutenberg–Richter relations for the MSEQ catalog remains straightforward.

This approach does not consider variability in the GMPE, which of course has been quantified by Bindi et al. (2017). Considering this variability, the criterion for building the MSEQ catalog could be formulated as follows: an event is not included in the MSEQ catalog if it has a probability of less than 50% of having the ground motion parameter at its epicenter larger than the median ground motion estimated from the main shock.

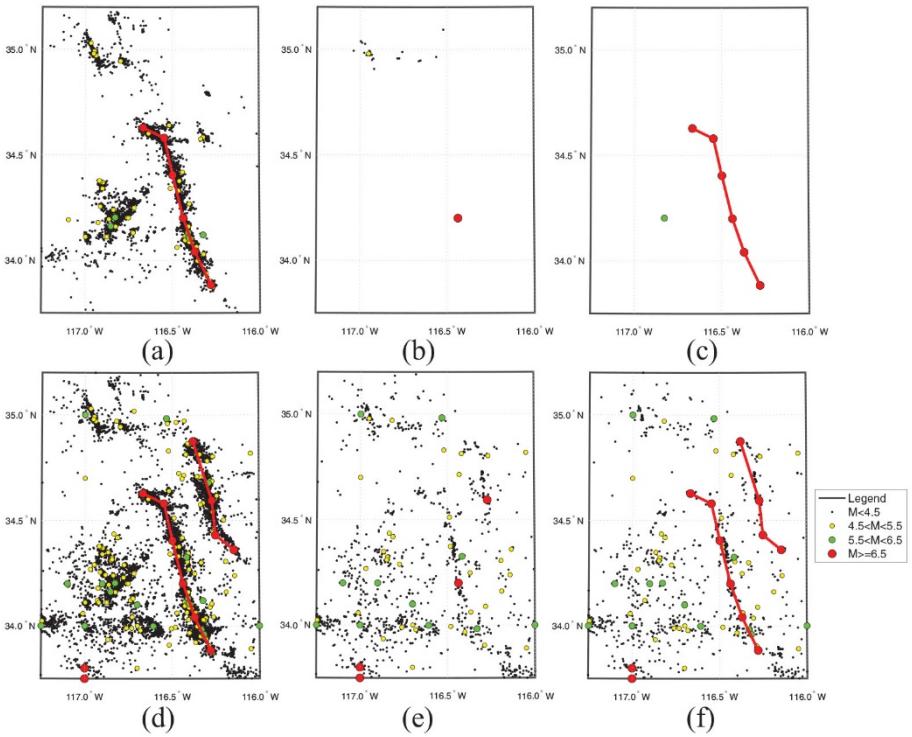

Figure 2. *Earthquakes in the Landers, California region. Earthquakes on maps (a), (b), and (c) occur in the aftershock time window, with duration of 939 days starting one day before the Landers main shock of June 28, 1992. Maps (d), (e), and (f) are drawn from the complete catalog, which includes events from 1769 to 2014. (a) Events in the USGS catalog C2, which has been reviewed to remove explosions and other non-tectonic events, and includes events with Mw ø 2.0 (Petersen et al., 2014). The Landers main shock is represented by six connected points in this figure, as described in the text. (b) Earthquakes in the USGS C3 catalog, which has been declustered using the Gardner and Knopoff (1974) algorithm. Here, the Landers main shock is represented only by a single point at its epicenter. (c) Earthquakes in the MSEQ-SA (3.0 s) catalog for the aftershock time window. The remaining earthquakes are the Landers main shock, represented by the six connected points, and the Mw6.48 Big Bear earthquake. (d) All earthquakes in the USGS Catalog C2. (e) All earthquakes in the USGS Catalog C3. (f) All MSEQ-SA (3.0 s) earthquakes. In maps (d) and (f), the October 16, 1999, Hector Mine earthquake (Mw7.11) east of the Landers rupture is represented by four points.*

Foreshocks

In the algorithm tested in this article, an earthquake will be retained in the MSEQ catalog if, for some interval of the time, it is the strongest source of ground motion at its epicenter. Thus, foreshocks are not removed. In the real-time perspective of the resident in the area, the foreshock is the earthquake; it is only recognized as a foreshock in retrospect. This approach is further justified in the ‘‘Discussion’’ section of this article.

Examples

Landers area

The MSEQ approach is illustrated, and compared with the GK method, in Figure 2. This illustration uses the earthquake catalogs C2 and C3 that were used by the US Geological Survey (USGS) to develop the 2014 National Seismic Hazard Model (Petersen et al., 2014; Shumway, 2019). Catalog C2 is a complete catalog, and catalog C3 includes only the events retained after declustering using the GK method. Figure 2a shows the post-mainshock catalog of earthquakes with Mw > 3 for the region around the Landers earthquake. The Landers main shock has Mw = 7.27, so based on the GK parameters, all earthquakes in a time window of 939 days, and within 76 km of the epicenter of the Landers epicenter, would be removed. After declustering using this algorithm, all earthquakes have been removed near the epicenter, but some remain in the northern portion of this region, as seen in Figure 2b. The Landers main shock has been identified by six virtual hypocenters, which can be seen in Figure 2a. The virtual hypocenters used to represent the main shock are also shown in Figure 2c, in which the MSEQ procedure has been applied. In the MSEQ catalog, the northern group of small events has been removed. However, the M6.48 Big Bear earthquake, represented by the green-colored point in Figure 2c, remains.

Figure 2d to f shows the corresponding catalogs for their entire time period. Comparison of Figure 2e and f, with the GK declustering and MSEQ catalogs, respectively, suggests at first glance that a similar number of earthquakes have been removed. Inspection finds, however, that Figure 2f has more events with Mw in the range 5.5–6.5, while Figure 2e appears to have more events in the smallest magnitude range.

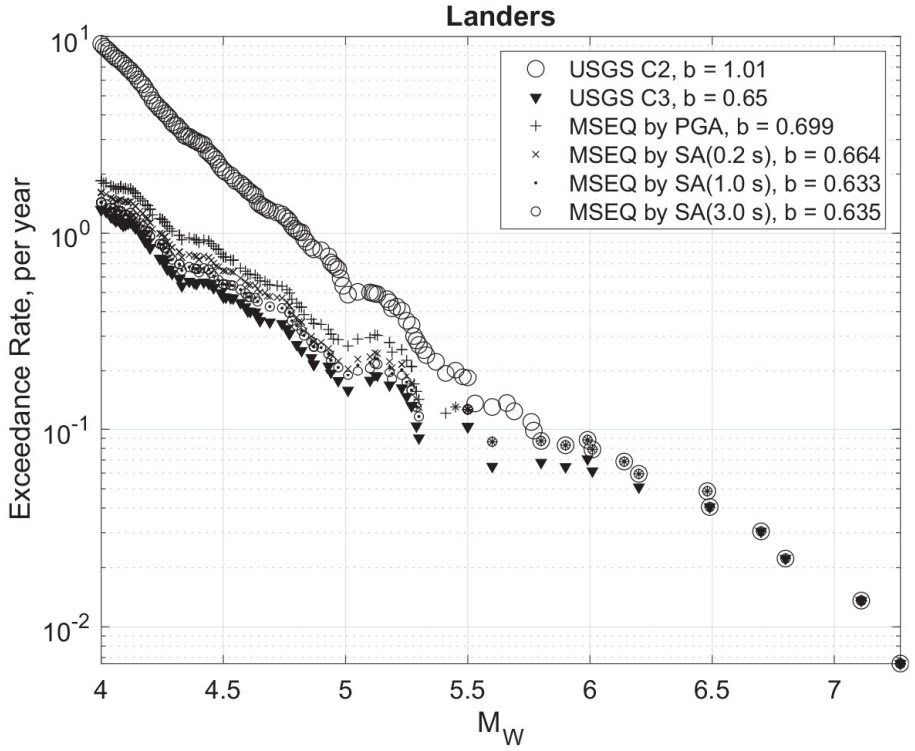

Figure 3. *Magnitude–frequency relationships for the Landers area (Figure 2). The estimated rate at each magnitude is determined as described in Online Supplement 1. The time windows used to make rate estimates are increased as the magnitude increases, as determined from the complete catalog. The variable time window and the relatively small size of this catalog explains why the points are not a monotonic function of magnitude, and why rates of large magnitudes may be overestimated. The b values are determined using linear least squares applied to the event rates in the magnitude range 4.0 ≤ Mw ≤ 7.0; extrapolation outside of this range is not defensible.*

This impression is confirmed in Figure 3, which shows the exceedance rates for six considered models: the full, un-declustered catalog, the catalog declustered by the USGS (C3), and the MSEQ catalogs using the GMPE for peak acceleration, SA (0.2 s), SA (1.0 s), and SA (3.0 s). The progressively longer period GMPE criteria remove progressively greater numbers of small events (this trend is more clearly seen in Figure 10 of Appendix 2). This follows from scaling properties of the seismic source spectrum, in which the amplitude at long periods decreases more rapidly as magnitude decreases than it does at high frequencies (e.g. Anderson, 2015). The MSEQ catalogs are most similar to the C3 catalog at M~3–3.5 (not shown in Figure 3). For smaller magnitudes, the MSEQ catalogs have fewer events than the GK method. Above M~4.5, all four of the MSEQ catalogs include larger numbers of events than the GK approach. For M~5–6, catalog C3 has an annual rate that is about 30% smaller than the MSEQ catalogs. These catalogs merge with the full catalog at about M5.8. The b values in Figure 3 are determined as in the Online Supplement. This method uses increasing catalog durations for larger magnitudes. In the Online Supplement and Appendix 2, these durations give a well-behaved magnitude–frequency distribution (MFD) for the Western US catalog. The complete catalog for the Western United States (Online Supplement 1) currently appears to be bilinear with an increase in the slope (b value) at about M;7. In the small Landers region, the rate estimates are scattered. Still, the b values fitted to these MFDs illustrate how the MSEQ catalogs are increasingly depleted in medium magnitude events, but cannot be used to extrapolate to either larger or smaller events.

Although our goal is not to create a declustered catalog of independent events, we tested the MSEQ catalogs for whether they are Poissonian in time. Figure 4 illustrates the procedure, where the Kolmogorov–Smirnov (K-S) test (e.g. Fisz, 1963; Luen and Stark, 2012) itself is summarized in the figure caption. Critical values given in elementary statistics books are not valid in our case because the mean rate is determined from the data. The correct approach is given by Lilliefors (1969), and the critical values used in this study are given in Appendix 1. For compactness, one can plot the difference between the distribution of time intervals in the catalog and the comparable Poissonian distribution normalized by the critical value from Appendix 1, and this is done in Figure 5.

For the Landers earthquake region in Figure 5, one can see that at magnitudes 3, 4, 5, and 6, the original catalog (C2) is not Poissonian. For M ≥ 3, the C3 catalog and the MSEQ-SA (3.0 s) both fail the test for Poissonian behavior by similar margins. At M ≥ 4, in contrast, the MSEQ-SA (3.0 s) catalog passes the K-S test, as does the C3 catalog. Although not shown, it is typical that the MSEQ-SA (3.0 s) catalog is closer to Poissonian than MSEQ catalogs for the shorter period GMPEs. For M ≥ 5, both the MSEQ and the declustered USGS catalog are consistent with Poissonian behavior. For M ≥ 6 + earthquakes, the power of the test is smallest, but Poissonian behavior is inconsistent with the timing of these largest earthquakes for all catalogs. In all cases where the maximum difference between the observed cumulative distribution function and the best-fitting exponential distribution is smaller than the critical value (i.e. the peak on Figure 5 is less than 1.0), the Kolmogorov–Smirnov test does not prove

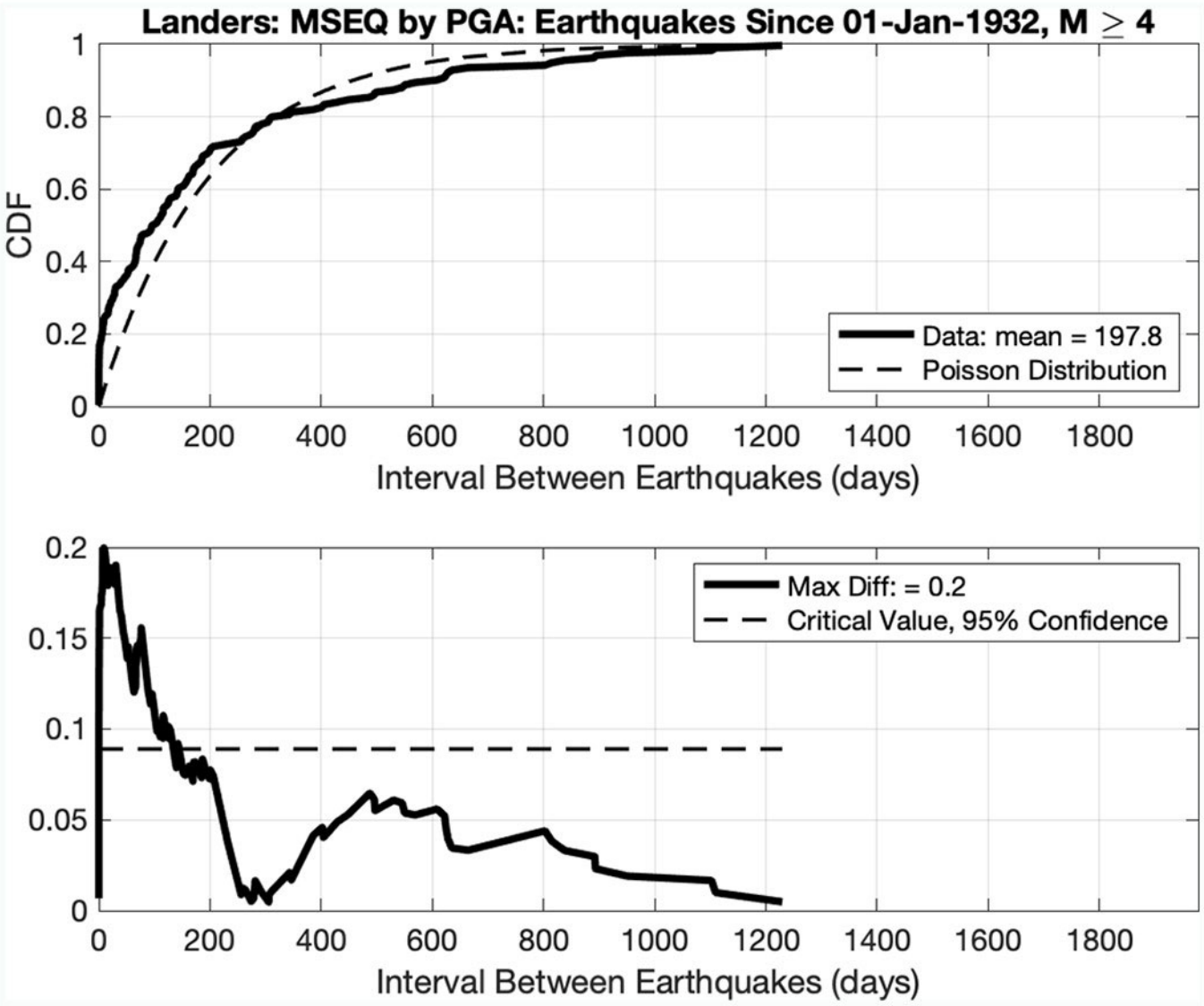

Figure 4. *Poissonian test example. This is for the MSEQ-PGA catalog for the Landers area, for earthquakes with magnitude M ≥ 4. In both frames, the horizontal axis is the interval between subsequent earthquakes in the catalog. The time intervals are determined, then sorted into increasing order, and the cumulative distribution of intervals smaller than each time interval is plotted in the upper frame. The mean interval is found, and a Poissonian distribution with that mean is also shown in the top frame. In the lower frame, the absolute value of the difference between these two curves is plotted, that is, DCDF = $|CDF_{Data} - CDF_{Poisson}|$ The Kolmogorov–Smirnov test, as modified by Lilliefors (1969), provides a critical value for 95% confidence, $CV_{K\text{-}S}$, and if max(D_{CDF}) > $CV_{K\text{-}S}$, there is 95% confidence that this distribution of time intervals between earthquakes is not Poissonian. The critical value, given in Appendix 1, depends on the number of earthquakes.*

that a set is Poissonian in time, but only that we cannot reject the Poisson hypothesis. Also, of course, this test does not evaluate whether the spatial distribution of the events is random.

Italy and Germany

Next, we consider a catalog of European earthquakes. This single catalog contains a flag identifying each earthquake as a main shock, a foreshock, or an aftershock based on the GK procedure with modified time and distance windows. MSEQ catalogs for PGA and SA (0.2 s), SA (1.0 s), and SA (3.0 s) were developed for the complete catalog. In Figure 6, the MSEQ-SA (3.0 s) catalog is compared with the complete and the declustered catalogs for Germany and for Italy. Germany is a region with relatively low overall seismic activity. Italy can be considered a region of high seismicity, with typical large event sizes in the M6–7 range. The catalog used for this study is primarily the European-Mediterranean Earthquake Catalog (EMEC; Grünthal and Wahlström, 2012). We employ here the temporal extension up to 2014 described in Grünthal et al. (2018; see ‘‘Data and resources’’ section).

A comparison of the MFDs from Figure 6 is shown in Figure 7. Finally, Figure 8 tests these catalogs for Poissonian behavior. The reduction in the numbers of earthquakes in these regions is not as great as in the Landers case, so the differences among the maps in Figure 6 are less obvious, and the reduction in the b value in Figure 7 is also small. However, the tendencies seen for Landers are reproduced. The MSEQ catalog contains progressively fewer earthquakes in the sequence from PGA to SA (3.0

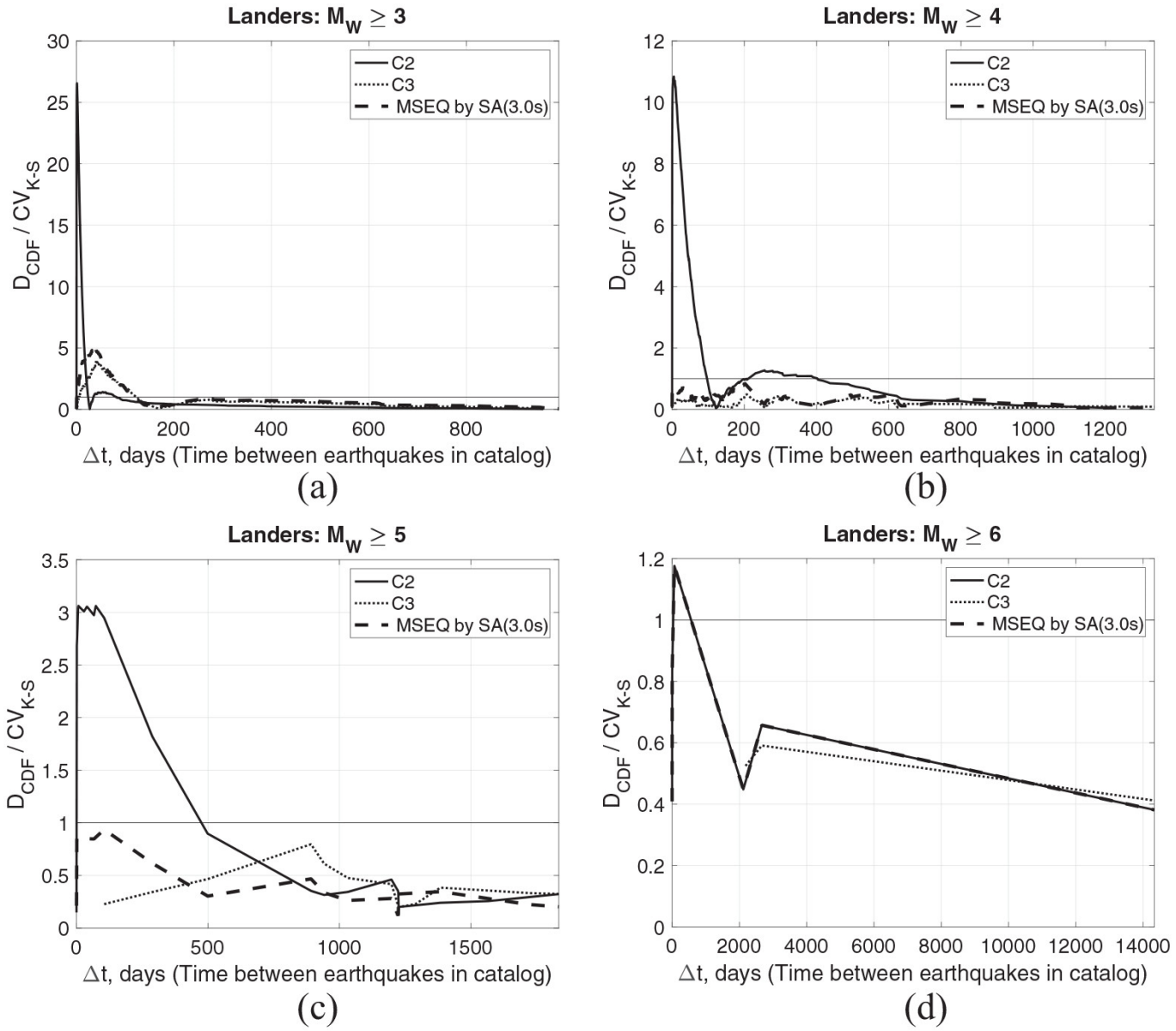

Figure 5. *Normalized Kolmogorov–Smirnov tests for three of the discussed catalogs for the Landers area, for magnitudes greater than a) 3.0, b) 4.0, c) 5,0, and d) 6.0. Catalogs for which the normalized difference is greater than 1.0 are, with 95% confidence, not Poissonian in time. Maps of these earthquakes are shown in Figure 2d to f, and magnitude–frequency distributions corresponding to these data are shown in Figure 3.*

s). In the German catalog, the GK procedure removes more events than the MSEQ procedure, even at the smallest magnitude displayed in Figure 7. In the Italian case, as in Landers, the MSEQ catalog using SA (3.0 s) includes fewer events than the GK catalog. In terms of achieving a Poissonian catalog, the GK catalog is successful at the M3 level in Germany, while MSEQ is not. Neither GK declustering nor the MSEQ process at the M3 level generates a Poissonian catalog in Italy, recalling the behavior in the Landers catalog. For M > 4 and M > 5, Poissonian behavior of the GK declustered catalog cannot be rejected in either the German or the Italian catalogs, but in the MSEQ catalog it is rejected with 95% confidence in Germany for M > 4 and in Italy for both M > 4 and M > 5, although it may be ‘‘close.’’ In conclusion, these tests do not show that MSEQ catalogs achieve a Poissonian distribution in time, but they indicate that the MSEQ catalog is moved toward being Poissonian in a manner similar to the effects of declustering.

Discussion

We propose that the use of an MSEQ catalog in PSHA is an alternative to the use of a declustered catalog. Declustering has often seemed like a rather arbitrary procedure; as noted, the mathematical problem is non-unique. Perhaps the reason the GK method has been used for 46 years and counting is that there has not been a convincing rationale to switch to any of the alternatives. We suggest that the MSEQ catalog is based on a different but reasonable rationale that is related to the use to which the hazard curve is applied. We have found that the MSEQ catalog resembles a declustered catalog, from which one would, for instance, determine a similar b value and rate of earthquakes. We anticipate that the seismic hazard analysis based

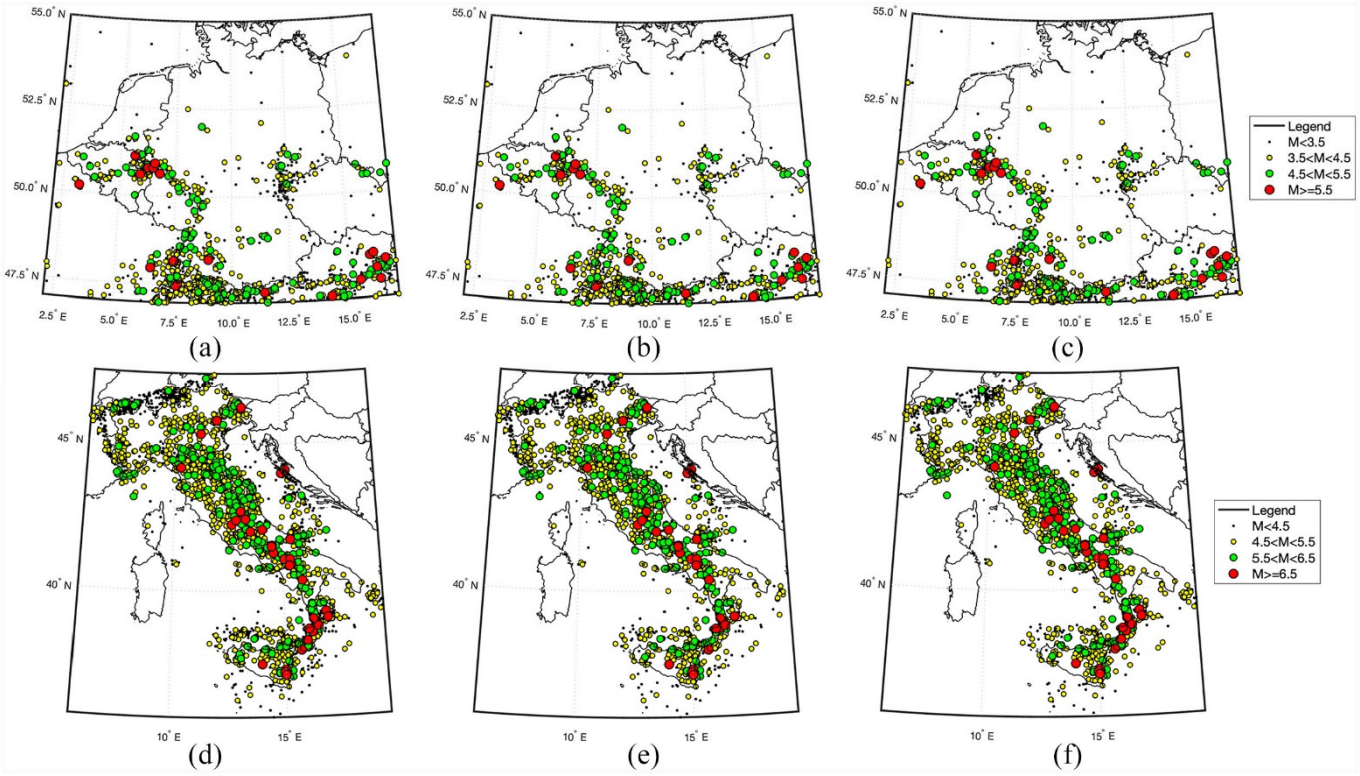

Figure 6. *Italy and Germany seismicity maps. Maps (a), (b), and (c) show respectively the complete catalog around Germany, the catalog declustered using a modified GK procedure with a different time window as shown in Table 1, and the events in the MSEQ catalog based on SA (3.0 s). Maps (d), (e), and (f) are equivalent for an oblique region that includes most of Italy.*

on an MSEQ catalog would incorporate internal consistency through use of the same GMPE for the MSEQ catalog development and the subsequent hazard calculation. Decisions need to be made on what ground motion parameter to use, or even if it would make sense to use different ground motion parameters depending on the application, but these are secondary if this rationale is adopted. Work remains to be done, also, to evaluate the consequences of this alternative to estimates of the hazard, and evaluate whether this approach can obtain acceptance from the users of PSHA.

To this end, Appendix 2 shows a simplified hazard calculation for a site in central Nevada. The hazard at this site is primarily controlled by the declustered catalog, but regional faults contribute at long periods. The hazard for peak acceleration is determined using only the MSEQ-PGA catalog, and the hazard for SA (1.0 s) is determined using only the MSEQ-SA (1.0 s) catalog. Compared with the hazard estimates for peak acceleration using the USGS model (Petersen et al., 2014), the hazard estimated using the MSEQ catalog is higher by about 40% at the exceedance rate of 4 3 10_{24} per year. For SA (1.0 s), the result is ambiguous because the calculation leaves out the regional faults, but the contribution to the hazard from the background is between 20% and 30% higher.

Retaining large aftershocks in PSHA background seismicity catalogs

We assume the existence of a true (Anderson and Biasi, 2016) or data-generating (Marzocchi and Jordan, 2014, 2017) hazard curve at every location. This curve is only theoretically knowable by a conceptual experiment that has recorded a long duration of seismic records from the site (e.g. 10^6 years) and uses the relevant records (e.g.M ≥ 5) to calculate the rates at which various levels of ground motions are exceeded. If we imagine a constructed structure in Christchurch, the M_W6.1 Christchurch earthquake is highly relevant, even though it can be considered an aftershock, and so should be included in such a ‘‘measured’’ hazard curve. The procedure that is presented in this article is designed to retain such events. We propose that there is a second reason why events such as the Christchurch should be retained. That reason is that it could have happened separately from the preceding M_W7.1 Darfield earthquake. The fact that it did not is perhaps a consequence of the fact that main shocks are not entirely Poissonian in time, but rather some underlying physics (e.g. triggering) can cause them to be closely spaced in time. But if one allows that either of these two events could have happened at separate, random times, then they should both be counted in the conceptual experiment to measure a hazard curve. By the same reasoning, this article does not remove the strongest foreshocks. We suggest that, for instance, the Joshua Tree event that preceded the Landers earthquake could have occurred separately, and thus should be counted in the conceptual experiment to measure the hazard curve.

More generally, one might consider the large earthquakes that are retained by the MSEQ approach in the context of the distinction between triggered earthquakes and aftershocks. The distinction is often ambiguous. Remote triggering of small to

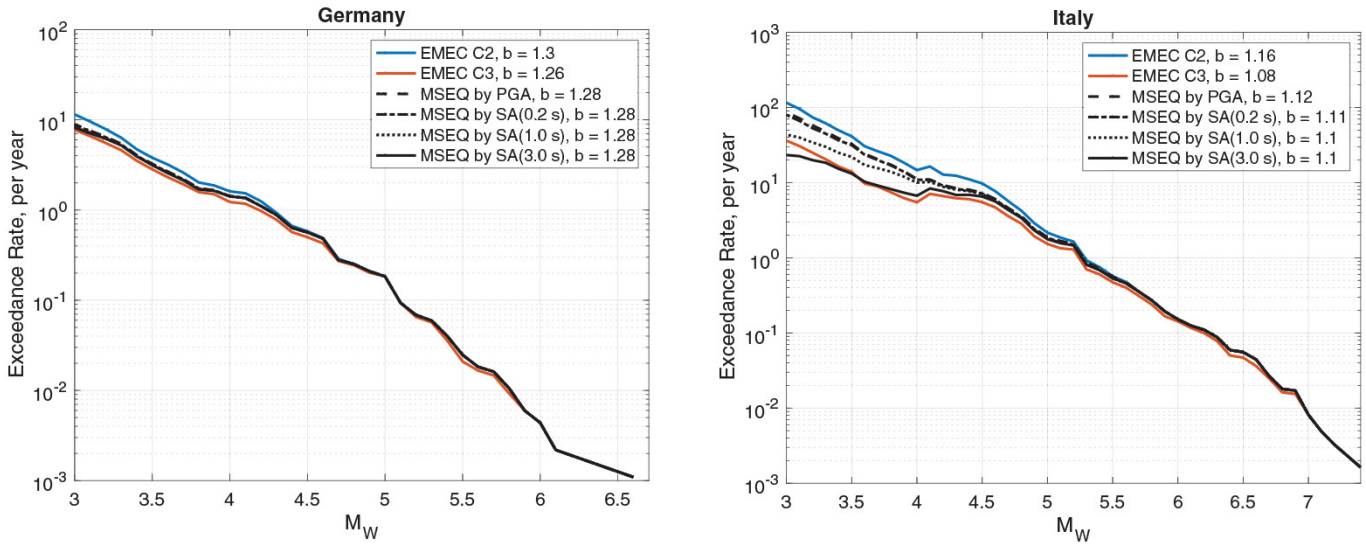

Figure 7. *Magnitude–frequency distributions for the original and modified catalogs from Germany (left) and Italy (right). For the Italian catalogs, b values are estimated for magnitudes greater than about 4.5. The legend retains the abbreviated terminology from Figure 3, in which C2 represents the complete, clean catalog, and C3 represents the corresponding catalog declustered by the GK method.*

moderate earthquakes to distances of over 700 km was obvious after the Landers earthquake (e.g. Anderson et al., 1994). Remote triggering of distant independent earthquakes has been recognized after several additional earthquakes (e.g. Hough et al., 2003; Jiang et al., 2010; Meltzner and Wald, 2003; Papadopoulos, 2002; Prejean et al., 2004). Of course, if triggered earthquakes can occur at large distances, then there is no surprise if they also occur at short distances. Felzer et al. (2004) propose that there is no statistical distinction between a triggered main shock and an aftershock, except that the triggered main shock happens to be larger with a probability controlled by a Gutenberg–Richter relationship. From these perspectives, the large events that are retained in the MSEQ catalog can legitimately be considered triggered main shock earthquakes rather than dependent events in a single cluster represented by a single main shock.

How important is Poissonian behavior in time?

As seen in Figures 5 and 8, the MSEQ catalogs developed in this study are generally not as close to being Poissonian as the catalogs developed by GK procedures. In recognizing this, we consider how important it is to have a Poissonian catalog for PSHA. The assumption that the earthquakes occur at random times is used for the transition from a hazard curve giving the annual rate of earthquakes to the annual probability (e.g. Cornell, 1968; McGuire, 2004). Let Y be an amplitude of ground motion, and let g(Y) be a hazard curve giving the annual rate of exceedance of Y. Then with the Poissonian assumption, the probability of an earthquake causing ground motions that exceed Y in a time interval of duration T is $p(Y, T) = 1 – \exp(–g(Y)T)$. However, neither the early history of PSHA, nor recent studies, supports the idea that a Poissonian behavior must be strictly enforced. For instance, Cornell (1968), Algermissen and Perkins (1976), and Algermissen et al. (1982) recognize that earthquake catalogs are not Poissonian when small magnitudes are included, but apparently consider it sufficient that the large events ‘‘tend to be’’ Poissonian. For instance, Algermissen et al. (1982) writes,

> The concept of hazard mapping used here is to assume that earthquakes are exponentially distributed with regard to magnitude and randomly distributed with regard to time. The exponential magnitude distribution is an assumption based on empirical observation. The distribution of earthquakes in time is assumed to be Poissonian. The important observation is that the occurrence of large shocks tends to be Poissonian while small shocks often are not. However, the ground motions associated with small shocks are of only marginal interest in engineering applications (Cornell, 1968).

Figures 5 and 8 confirm that in the MSEQ catalogs considered here, the large events have distributions in time that are close to Poissonian even when they fail the Kolmogorov–Smirnov/Lilliefors test, while, as found by Luen and Stark (2012) as well as here, declustered catalogs that are complete to smaller magnitudes are not always purely Poissonian.

The consequences of the failure to be Poissonian are not necessarily a problem. Cornell and Winterstein (1988) found that for regular repetition of large earthquakes, the error introduced by the Poissonian assumption is acceptably small except when the events are highly regular and the elapsed time since the last event exceeds the recurrence interval. More generally, Marzocchi and Taroni (2014) indicated that more realistic clustered distributions of earthquakes lead to a distribution of exceedances that is still well approximated by a Poisson distribution at the low probabilities of primary concern for engineering design. In most of PSHA studies, where exceedance probabilities less than 10% in 50 years are required and ground motion parameters are not very small, the Poisson distribution could be assumed regardless of time independence of earthquakes.

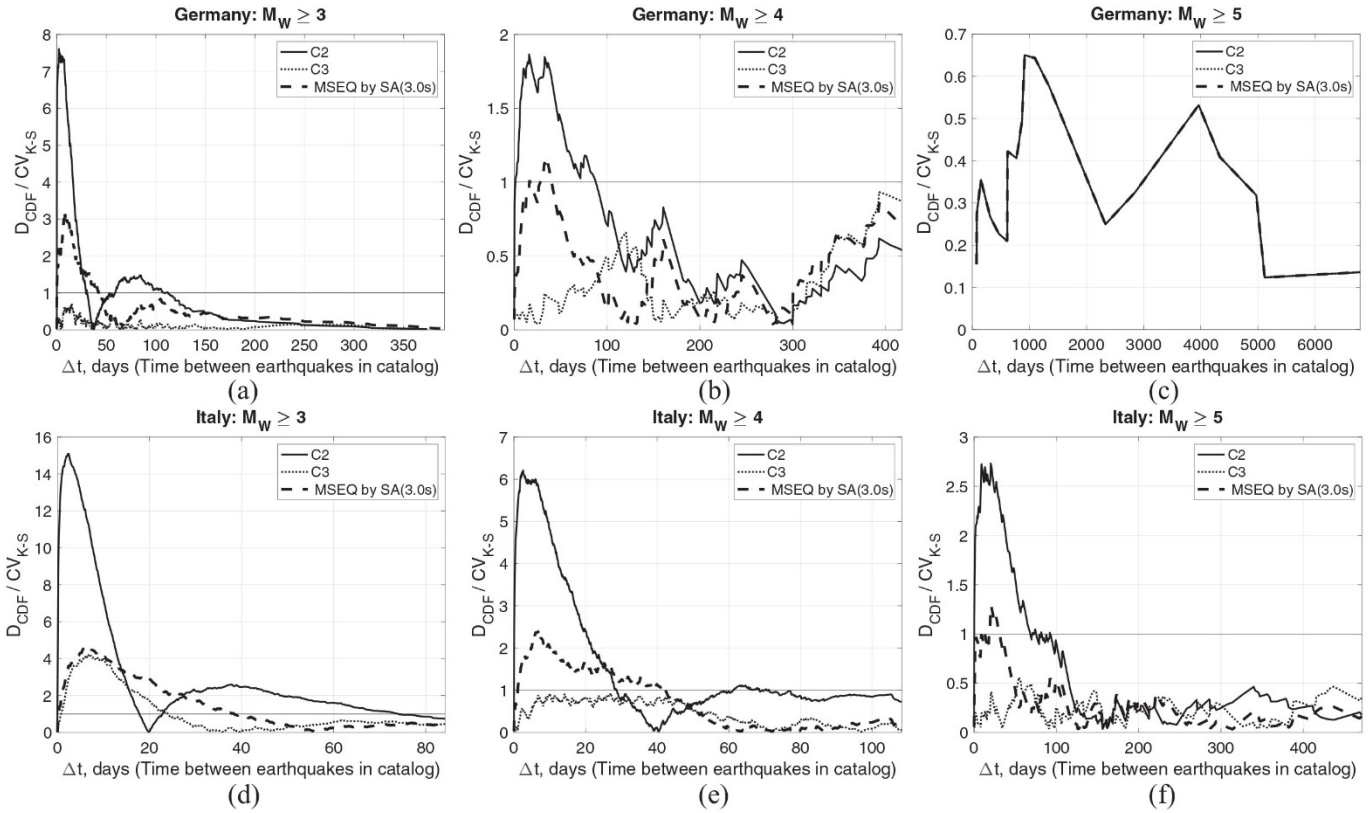

Figure 8. *Results of the Kolmogorov–Smirnov test for Poissonian behavior for the discussed catalogs for Germany and Italy. Frames (a–c) are for the Germany region of the European catalog, as shown in Figure 6a, for the cataloged events with magnitudes greater than 3.0, 4.0, and 5.0, respectively. Part (c) tests the MSEQ-SA (3.0 s) catalog. Parts (d–f) are the equivalent for the Italian part of the European catalog shown in Figure 6d. The legend retains the terminology from Figure 7, in which C2 represents the complete, clean catalog, and C3 represents the corresponding catalog declustered by the GK method.*

Alternatives for handling large dependent earthquakes

The standard PSHA approach, basing the background seismicity on a declustered catalog, is the reference approach, against which all suggestions of modified procedures will inevitably be compared. An oversimplified expectation is that adding dependent events inflates the overall earthquake rate, even though they have no significant impact on the true hazard, since the main shock does the main damage.

However, some studies (e.g. Boyd, 2012; Iervolino, 2019; Iervolino et al., 2014) suggest that declustering underestimates the hazard. Given that dependent events are sometimes damaging, several papers have considered alternative approaches to include the effects of aftershocks. The first alternative is to include all events in the working catalog, or to use some alternative approach, such as the one suggested by Marzocchi and Taroni (2014), to end up with a seismicity model with total earthquake rates that match the original catalog. Boyd (2012) makes the case that without declustering, the result would give the highest hazard estimates.

The second alternative is to adjust the hazard model for the effect of the aftershocks in the declustered catalog. Yeo and Cornell (2009) developed a model for the temporally changing hazard in an aftershock zone. Models that combine this aftershock PSHA (APSHA) with a random sequence of main shocks are presented by Boyd (2012), Iervolino et al. (2014), and Iervolino (2019). Effectively, these methods aim for the intensity measure for each independent cluster, represented by its main shock, to be adjusted upward from the value for the main shock alone so that aftershocks are also statistically included. Boyd (2012) finds that this approach results in a 10% to 20% increase in the hazard for a site in California using stochastic realizations of the aftershock distribution. Iervolino et al. (2014) come to a similar conclusion with an analytical formulation that they call Sequence-Based PSHA (SPSHA).

The strength of these ideas is that the clusters might then retain the same level of approximation to a Poissonian distribution of events that present methods achieve. However, these methods still imply the need to decluster the earthquakes catalog, identify independent main shocks, and describe quantitatively the spatial and time distributions of foreshocks and aftershocks. All these steps may be associated to large epistemic uncertainties, both from the declustering approach and the potential overestimate of the uncertainties in current GMPEs, contributed by the ergodic assumption (Anderson and Brune, 1999). Despite the possible drawbacks, this set of methods are statistical approaches to incorporate foreshocks and aftershocks, with the result being some increase in the hazard compared to the standard declustering approach.

This article suggests another approach. Since intermediate-sized events are retained in the MSEQ catalog but removed by declustering, we might expect that if this catalog is used in PSHA in the place of a declustered catalog, on average the seismic hazard will be increased somewhat. The results in Appendix 2 are consistent with that expectation. However, those results are quite limited: we used the same b value as used by Petersen et al. (2014) rather than fitting b to the MFD of the MSEQ catalogs in the magnitude range that matters, and have not rigorously evaluated whether a Gutenberg–Richter relationship is appropriate. Inclusion of intermediate-sized events may steepen the magnitude–frequency curve (i.e. produce a higher b value), which, counterintuitively, may lower the rate of higher magnitudes, which in turn lowers the hazard as seen in a hazard curve. The impact of the use of the MSEQ catalog thus remains to be systematically tested.

Spatial considerations

Of course, any individual catalog is a realization of the seismicity for one time interval, and other time intervals will likely have the seismicity in other locations. Thus, as shown by Marzocchi and Taroni (2014), methods using a non-declustered seismic catalog, a declustered seismic catalog, or the MSEQ catalog, directly to estimate the background seismicity in a PSHA computation, may overestimate the seismicity rates for the regions that recently experienced a strong seismic cluster with respect to areas that did not experience clusters in the time interval covered by the catalog. Such potential spatial biases due to earthquake clustering coupled with the limited duration of earthquake catalogs are the motivation for either smoothing the background seismicity (e.g. Frankel et al., 1996, 2002; Petersen et al., 2008, 2014) or using large areas source zones (e.g. Cinti et al., 2004). With a million years of seismicity and strong motion records, perhaps all the active sources in a region would be sampled, and smoothing would not be so necessary.

Conclusion

The MSEQ approach approximates the thought experiment to build an empirical hazard curve. It aims to retain all earthquakes that are most relevant for earthquake engineering applications. Many large events, in spite of being unambiguously foreshocks or aftershocks in an earthquake cluster, remain in the MSEQ catalog and can be considered independent, triggered earthquakes that could occur separately. For PSHA, it seems reasonable that these events are included in the catalog that is used to set the rate and define the shape of the MFD of earthquakes and, along with geological observations (e.g. Field et al., 2013; Petersen et al., 2014; Wesnousky, 1986), contribute to the complete seismicity model. For all of these reasons, we propose that the MSEQ approach is appropriate to be considered as a legitimate methodology in PSHA. Also note that if declustering is preferred, the MSEQ catalog identifies earthquakes that perhaps should be constrained to be included in the declustered catalog.

Data and resources

The US Geological Survey earthquake catalogs are available from https://www.sciencebase.gov/catalog/item/5db9be62e4b06957974eb5ca, last accessed March 17, 2020. The European-Mediterranean Earthquake Catalog (EMEC) used for this project is available under request (contact fcotton@gfz-potsdam.de)

Acknowledgments

We appreciate a careful review and helpful comments by Christian Bosse. John Anderson is grateful for the opportunity to visit the German Research Center for Geosciences (GFZ) on two occasions, in 2017 and again in 2019, enabling development of this article. We greatly appreciate the thoughtful comments and recommendations of three anonymous reviewers and the associate editor. This input led to significant improvements in this article.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

John G Anderson https://orcid.org/0000-0002-8882-6039
Dino Bindi https://orcid.org/0000-0002-8619-2220

Supplemental material

Supplemental material for this article is available online.

References

Abrahamson NA, Silva WJ and Kamai R (2014) Summary of the ASK14 ground motion relation for active crustal regions. Earthquake Spectra 30(3): 1025–1055.

Algermissen S and Perkins D (1976) A probabilistic estimate of maximum ground acceleration in the contiguous United States. Open File Report 76-416. US Geological Survey. Available at: https://doi.org/10.3133/ofr76416 (accessed January 2021).

Algermissen ST, Perkins DM, Thenhaus PC, Hanson SL and Bender BL (1982) Probabilistic estimates of maximum acceleration and velocity in rock in the contiguous United States. Open File Report 82-1033. US Geological Survey. Available at: https://doi.org/10.3133/ofr821033 (accessed January 2021).

Ancheta TD, Darragh RB, Stewart JP, Seyhan E, Silva WJ, Chiou BS-J, Wooddell KE, Graves RW, Kottke AR, Boore DM, Kishida T and Donahue JL (2014) NGA-West2 database. Earthquake Spectra 30(3): 989–1005.

Anderson JG and Biasi GP (2016) What is the basic assumption for probabilistic seismic hazard assessment? Seismological Research Letters 87(2A): 323–326.

Anderson JG and Brune JN (1999) Probabilistic seismic hazard analysis without the ergodic assumption. Seismological Research Letters 70(1): 19–28.

Anderson JG. Physical processes that control strong ground motion. In: Schubert G (ed) Treatise on geophysics. 2nd ed. Oxford: Elsevier, 2015, pp. 505–557.

Anderson JG and Nanjo K (2013) Distribution of earthquake cluster sizes in the Western United States and in Japan. Bulletin of the Seismological Society of America 103(1): 412–423.

Anderson JG, Brune JN, Louie JN, Zeng Y, Savage M, Yu G, Chen Q and dePolo D (1994) Seismicity in the western Great Basin apparently triggered by the Landers, California, earthquake, 28 June 1992. Bulletin of the Seismological Society of America 84(3): 863–891.

Anderson JG, Sucuoglu H, Erberik A, Yilmaz T, Inan E, Durukal E, Erdik M, Anooshehpoor R, Brune JN and Ni S-D (2000) Implications for seismic hazard analysis. Earthquake Spectra 16(S1): 113–137.

Bindi D, Cotton F, Kotha SR, Bosse C, Stromeyer D and Grünthal G (2017) Application-driven ground motion prediction equation for seismic hazard assessments in non-cratonic moderate seismicity areas. Journal of Seismology 21: 1201–1218.

Boyd OS (2012) Including foreshocks and aftershocks in time-independent probabilistic seismic hazard analyses. Bulletin of the Seismological Society of America 102(3): 909–917.

CEUS-SSC (2012) Central and Eastern United States seismic source characterization for nuclear facilities. Technical report EPRI, US DOE, and US NRC, Palo Alto, CA, July.

Cinti FR, Faenza L, Marzocchi W and Montone P (2004) Probability map of the next M ≥ 5.5 earthquakes in Italy. Geochemistry, Geophysics, Geosystems 5(11): Q11003.

Console R, Jackson DD and Kagan YY (2010) Using the ETAS model for catalog declustering and seismic background assessment. Pure and Applied Geophysics 167: 819–830.

Cornell CA (1968) Engineering seismic risk analysis. Bulletin of the Seismological Society of America 58(5): 1583–1606.

Cornell CA and Winterstein SR (1988) Temporal and magnitude dependence in earthquake recurrence models. Bulletin of the Seismological Society of America 78(4): 1522–1537.

Felzer KR, Abercrombie RE and Ekström G (2004) A common origin for aftershocks, foreshocks, and multiplets. Bulletin of the Seismological Society of America 94(1): 88–98.

Field E, Biasi G, Bird P, Dawson T, Felzer K, Jackson D, Johnson K, Jordan T, Madden C, Michael A, Milner K, Page M, Parsons T, Powers P, Shaw B, Thatcher W, Weldon R II and Zeng Y (2013) Uniform California earthquake rupture forecast, version 3 (UCERF3)—The time-independent model. Open-File Report 2013–1165, 5 November. Menlo Park, CA: US Geological Survey. Available at: http://pubs.usgs.gov/of/2013/1165/ (accessed January 2021).

Fisz M (1963) Probability Theory and Mathematical Statistics. 3rd ed. New York: John Wiley & Sons, p. 677.

Frankel A, Mueller C, Barnhard T, Perkins D, Leyendecker EV, Dickman N, Hanson S and Hopper M (1996) National seismic-hazard maps: documentation June 1996. Open File Report 96-532, 6 April. Reston, VA: US Geological Survey.

Frankel AD, Petersen MD, Mueller CS, Haller KM, Wheeler RL, Leyendecker EV, Wesson RL, Harmsen SC, Cramer CH, Perkins DM and Rukstales KS (2002) Documentation for the 2002 update of the National Seismic Hazard Maps. Open-File Report 02-420. U.S. Department of the Interior, U.S. Geological Survey. Available at: https://doi.org/10.3133/ofr02420 (accessed January 2021).

Frohlich C and Davis SD (1990) Single-link cluster analysis as a method to evaluate spatial and temporal properties of earthquake catalogues. Geophysical Journal International 100(1): 19–32.

Gardner JK and Knopoff L (1974) Is the sequence of earthquakes in southern California, with aftershocks removed, Poissonian? Bulletin of the Seismological Society of America 64: 1363–1367.

Grünthal G and Wahlström R (2012) The European-Mediterranean earthquake catalogue (EMEC) for the last millennium. Journal of Seismology 16(3): 535–570.

Grünthal G, Stromeyer D, Bosse C, Cotton F and Bindi D (2018) The probabilistic seismic hazard assessment of Germany—Version 2016, considering the range of epistemic uncertainties and aleatory variability. Bulletin of Earthquake Engineering 16(10): 4339–4395.

Hoel PG (1971) Introduction to Mathematical Statistics. 4th ed. New York: John Wiley & Sons.

Hough SE, Seeber L and Armbruster JG (2003) Intraplate triggered earthquakes: Observations and interpretation. Bulletin of the Seismological Society of America 93: 2212–2221.

Iervolino I (2019) Generalized earthquake counting processes for sequence-based hazard. Bulletin of the Seismological Society of America 109(4): 1435–1450.

Iervolino I, Giorgio M and Polidoro B (2014) Sequence-based probabilistic seismic hazard analysis. Bulletin of the Seismological Society of America 104(2): 1006–1012.

Jiang T, Peng Z, Wang W and Chen Q-F (2010) Remotely triggered seismicity in continental China following the 2008 Mw 7.9 Wenchuan earthquake. Bulletin of the Seismological Society of America 100(5B): 2574–2589.
Lilliefors HW (1967) On the Kolmogorov-Smirnov test for normality with mean and variance unknown. Journal of the American Statistical Association 62(318): 399–402.
Lilliefors HW (1969) On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown. Journal of the American Statistical Association 64(325): 387–389.
Luen B and Stark PB (2012) Poisson tests of declustered catalogues. Geophysical Journal International 189(1): 691–700.
McGuire RK (2004) ''Seismic Hazard and Risk Analysis,'' Engineering Monographs, MNO-10. Oakland, CA: Earthquake Engineering Research Institute.
Marsan D and Lengline O (2008) Extending earthquakes' reach through cascading. Science 319: 1076–1079.
Marzocchi W and Jordan TH (2014) Testing for ontological errors in probabilistic forecasting models of natural systems. Proceedings of the National Academy of Sciences of the United States of America 111(33): 11973–11978.
Marzocchi W and Jordan TH (2017) A unified probabilistic framework for seismic hazard analysis. Bulletin of the Seismological Society of America 107(6): 2738–2744.
Marzocchi W and Taroni M (2014) Some thoughts on declustering in probabilistic seismic-hazard analysis. Bulletin of the Seismological Society of America 104(4): 1838–1845.
Meltzner AJ and Wald DJ (2003) Aftershocks and triggered events of the great 1906 California earthquake. Bulletin of the Seismological Society of America 93(5): 2160–2186.
Molchan GM and Dmitrieva OE (1992) Aftershock identification: Methods and new approaches. Geophysical Journal International 109: 501–516.
Ogata Y (1988) Statistical models for earthquake occurrences and residual analysis for point processes. Journal of the American Statistical Association 83(401): 9–27.
Ogata Y (1989) Statistical models for standard seismicity and detection of anomalies by residual analysis. Tectonophysics 169: 159–174.
Ogata Y (1999) Seismicity analysis through point-process modeling: A review. Pure and Applied Geophysics 155: 471–507.
Papadopoulos GA (2002) The Athens, Greece, earthquake (Ms 5.9) of 7 September 1999: An event triggered by the Izmit, Turkey, 17 August 1999 earthquake? Bulletin of the Seismological Society of America 92(1): 312–321.
Petersen MD, Frankel AD, Harmsen SC, Mueller CS, Haller KM, Wheeler RL, Wesson RL, Zeng Y, Boyd OS, Perkins DM, Luco N, Field EH, Wills CJ and Rukstales KS (2008) Documentation for the 2008 update of the United States National Seismic Hazard Maps. Open-File Report 2008-1128, January. Reston, VA: US Geological Survey.
Petersen MD, Moschetti MP, Powers PM, Mueller CS, Haller KM, Frankel AD, Zeng Y, Rezaeian S, Harmsen SC, Boyd OS, Field N, Chen R, Rukstales KS, Luco N, Wheeler RL, Williams RA and Olsen AH (2014) Documentation for the 2014 Update of the United States National Seismic Hazard Maps. Open File Report 2014-1091. US Geological Survey.
Prejean SG, Hill DP, Brodsky EE, Hough SE, Johnston MJS, Malone SD, Oppenheimer DH, Pitt AM and Richards-Dinger KB (2004) Remotely triggered seismicity on the United States west coast following the Mw 7.9 Denali Fault earthquake. Bulletin of the Seismological Society of America 94(6B): S348–S359.
Pizzi A, Di Domenica A, Gallovicˇ F, et al (2017) Fault segmentation as constraint to the occurrence of the main shocks of the 2016 Central Italy seismic sequence. Tectonics 36: 2370–2387.
Prozorov AG and Dziewonski AM(1982) A method of studying variations in the clustering property of earthquakes: Application to the analysis of global seismicity. Journal of Geophysical Research 87(B4): 2829–2839.
Reasenberg P (1985) Second-order moment of central California seismicity, 1969-1982. Journal of Geophysical Research 90(B7): 5479–5495.
Richter CF (1958) Elementary Seismology. San Francisco, CA: W.H. Freeman and Company.
Savage WU (1972) Microearthquake clustering near Fairview Peak, Nevada and in the Nevada seismic zone. Journal of Geophysical Research 77: 7049–7056.
Shumway A (2019) Data release for the 2014 National Seismic Hazard Model for the conterminous U.S. US Geological Survey. Available at: https://doi.org/10.5066/P9P77LGZ (accessed January 2021).
Sieh K, Jones L, Hauksson E, Hudnut K, Eberhart-Phillips D, Heaton T, Hough S, Hutton K, Kanamori H, Lilje A, Lindvall S, McGill SF, Mori J, Rubin C, Spotila JA, Stock J, Thio HK, Treiman J, Wernicke B and Zachariasen J (1993) Near-field investigations of the Landers earthquake sequence, April to July 1992. Science 260: 171–176.
Wesnousky SG (1986) Earthquakes, quaternary faults, and seismic hazard in California. Journal of Geophysical Research 91(B12): 12587–12631.
Yeo GL and Cornell CA (2009) A probabilistic framework for quantification of aftershock ground-motion hazard in California: Methodology and parametric study. Earthquake Engineering and Structural Dynamics 38: 45–60.
Zaliapin I and Ben-Zion Y (2011) Asymmetric distribution of aftershocks on large faults in California. Geophysical Journal International 185: 1288–1304.
Zaliapin I and Ben-Zion Y (2020) Earthquake declustering using the nearest-neighbor approach in space-time-magnitude domain. Journal of Geophysical Research: Solid Earth 125: e2018JB017120. DOI: 10.1029/2018JB017120.
Zaliapin I, Gabrielov A, Keilis-Borok V and Wong H (2008) Clustering analysis of seismicity and aftershock identification. Physical Review Letters 101: 018501.

Appendix 1

Critical values of the Kolmogorov–Smirnov Test for a Poisson distribution when the mean occurrence rate is estimated from the data

It is common to test whether the occurrence times of earthquakes in a catalog, often with foreshocks and aftershocks removed, are random and uncorrelated, that is, Poissonian. If so, the distribution of time intervals between the earthquakes should be consistent with an exponential distribution. One test is based on the Kolmogorov–Smirnov statistic. If n time intervals in a catalog are ordered such that $\tau_1 \leq \tau_2 \leq \tau_n$, the empirical cumulative distribution of the times between earthquakes is:

$$S_n(\tau) = \begin{cases} 0 & \tau < \tau_1 \\ \frac{i}{n} & \tau_i \leq \tau < \tau_{i+1}, i = 1, 2, \ldots n-1 \\ 1 & \tau \geq \tau_n \end{cases} \quad (1)$$

The corresponding empirical cumulative exponential distribution is:

$$F(\tau) = 1 - \exp\left(-\frac{\tau}{\bar{\tau}}\right) \quad (2)$$

where $\bar{\tau}$, the maximum likelihood estimate of the inverse rate parameter of the exponential distribution, is the mean of $\{\tau_i\}$. The Kolmogorov–Smirnov statistic D_n is the maximum difference between S_n and F:

$$D_n = \sup_{-\infty < \tau < \infty} |F(\tau) - S_n(\tau)| \quad (3)$$

Lilliefors (1967, 1969) points out that the standard tables of critical values for the Kolmogorov–Smirnov test (e.g. Hoel, 1971) are not valid when parameters of the distribution are estimated from the data itself, as is the case in Equation 2. Lilliefors (1969) used a Monte Carlo calculation to determine the appropriate critical values.

We have reproduced the Lilliefors (1969) calculations. As in his paper, we used Monte Carlo realizations of earthquake occurrence times to find critical values of the Kolmogorov–Smirnov statistic for a selection of sample sizes. The critical values found here, D_n^α, are given in Table 2 and plotted in Figure 9. The calculations were performed using the same MATLAB code as is used to measure the Kolmogorov–Smirnov statistic in this article. If $D_n > D_n^\alpha$, then one rejects the hypothesis that the distribution $S_n(\tau)$ is Poissonian with confidence of $1-\alpha$.

Table 2 is similar to, but more accurate than, the table given by Lilliefors (1969). For each value of n, Lilliefors estimated critical values based on 5000 Monte Carlo simulations, while Table 2 uses 2,500,000. Figure 9 also projects the critical values determined for $n \geq 5000$ back to smaller values of n. Critical values for n = 50 are about 2% smaller than predicted by this asymptotic value. As can be seen in Figure 9, values of D_n^α for n <1000 are nearly linear on log–log axes, so they can be accurately interpolated in log–log space between the values given in Table 2.

Table 2. Kolmogorov–Smirnov critical values for an exponential distribution when the mean rate is estimated from the data.

n	α = 0.2	α = 0.1	α = 0.05	α = 0.01
3	0.451	0.511	0.551	0.601
4	0.401	0.444	0.484	0.557
5	0.360	0.404	0.442	0.513
10	0.263	0.295	0.324	0.381
15	0.217	0.244	0.269	0.317
20	0.189	0.213	0.234	0.277
25	0.170	0.192	0.211	0.249
30	0.156	0.176	0.193	0.229
35	0.145	0.163	0.179	0.213
40	0.136	0.153	0.168	0.199
50	0.122	0.137	0.151	0.179
100	0.0868	0.0977	0.108	0.127
200	0.0617	0.0695	0.0764	0.0905
500	0.0392	0.0442	0.0486	0.0575
1000	0.0278	0.0313	0.0344	0.0407
2000	0.0197	0.0222	0.0244	0.0288
5000	0.0125	0.0140	0.0154	0.0183
$n > 5000$[a]	$\frac{0.882}{\sqrt{n}}$	$\frac{0.993}{\sqrt{n}}$	$\frac{1.091}{\sqrt{n}}$	$\frac{1.291}{\sqrt{n}}$

Source: Recalculated after (Lilliefors, 1969).
Each value based on the distribution of the Kolmogorov–Smirnov statistic was determined using 2,500,000 Monte Carlo simulations. These values are the average of values obtained in 25 runs of 100,000 simulations each. The standard error of all values (standard deviation of the 25 samples divided by their respective mean) is less than 0.3%, and the corresponding uncertainties in the mean are nearly all smaller than 10^{-4}.
[a]Standard deviations of the numerators for these asymptotic values, based on variation of the results from the multiple runs, are as follows: α = 0.2, 0.882±0.001; α = 0.1, 0.993±0.001; α = 0.05, 1.091±0.002; and α = 0.01, 1.291±0.003. For two places beyond the decimal (0.88, 0.99, 1.09, 1.29), these asymptotic values are valid for $n > 500$.

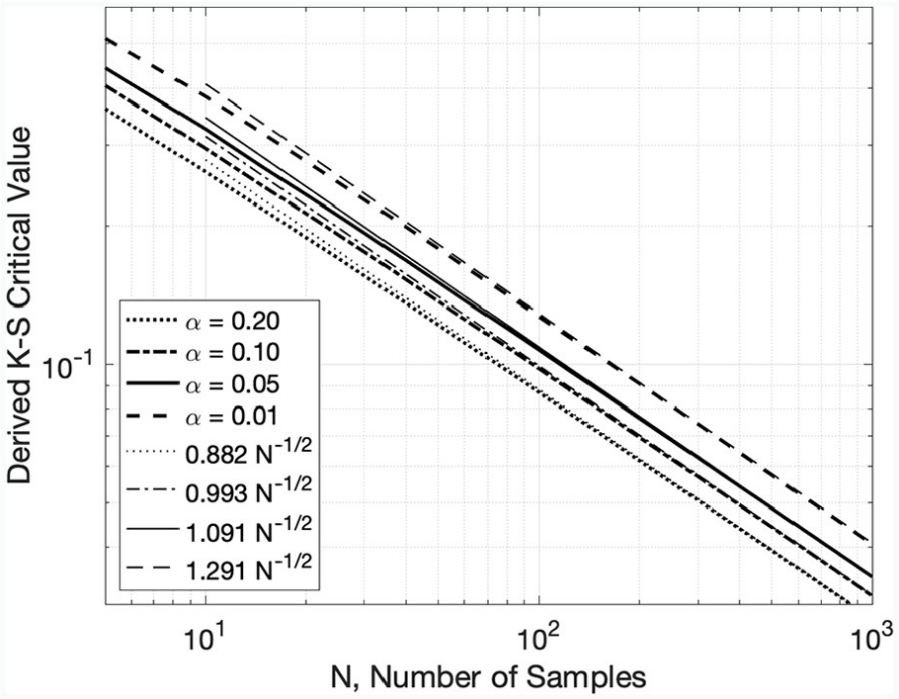

Figure 10. *Critical values for a Poisson (exponential) distribution. Thin straight lines of the same style extrapolate the asymptotic values for N . 5000 back to smaller numbers of realizations.*

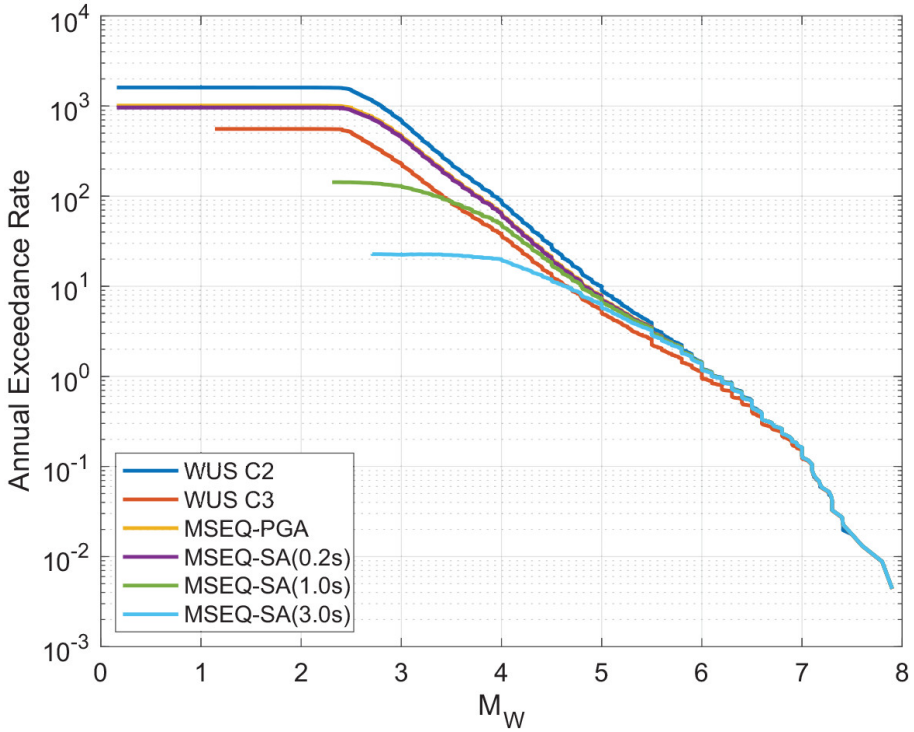

Figure 10. *Magnitude–frequency distributions for the USGS Western US C2 catalog for 2018 and for some subset catalogs derived from this catalog. The C3 catalog was declustered by the USGS and used for input to the 2018 National Seismic Hazard Model. The MSEQ catalogs were derived as discussed in the main text.*

Appendix 2

Sample hazard estimate using the Maximum Shaking Earthquake catalog

The purpose of this appendix is to show one comparison of a hazard calculation using the Maximum Shaking Earthquake (MSEQ) catalog with the 2014 U. S. Geological Survey (USGS) hazard model (Petersen et al., 2014) for the same location. This uses a strategically selected site, and it is premature to suggest that these results might be general.

Figure 10 shows the magnitude–frequency distributions (MFDs) of six catalogs for the Western United States. The first is the USGS C2 catalog, which includes earthquakes in or near the Western United States above magnitude 2.5 (see ‘‘Data and

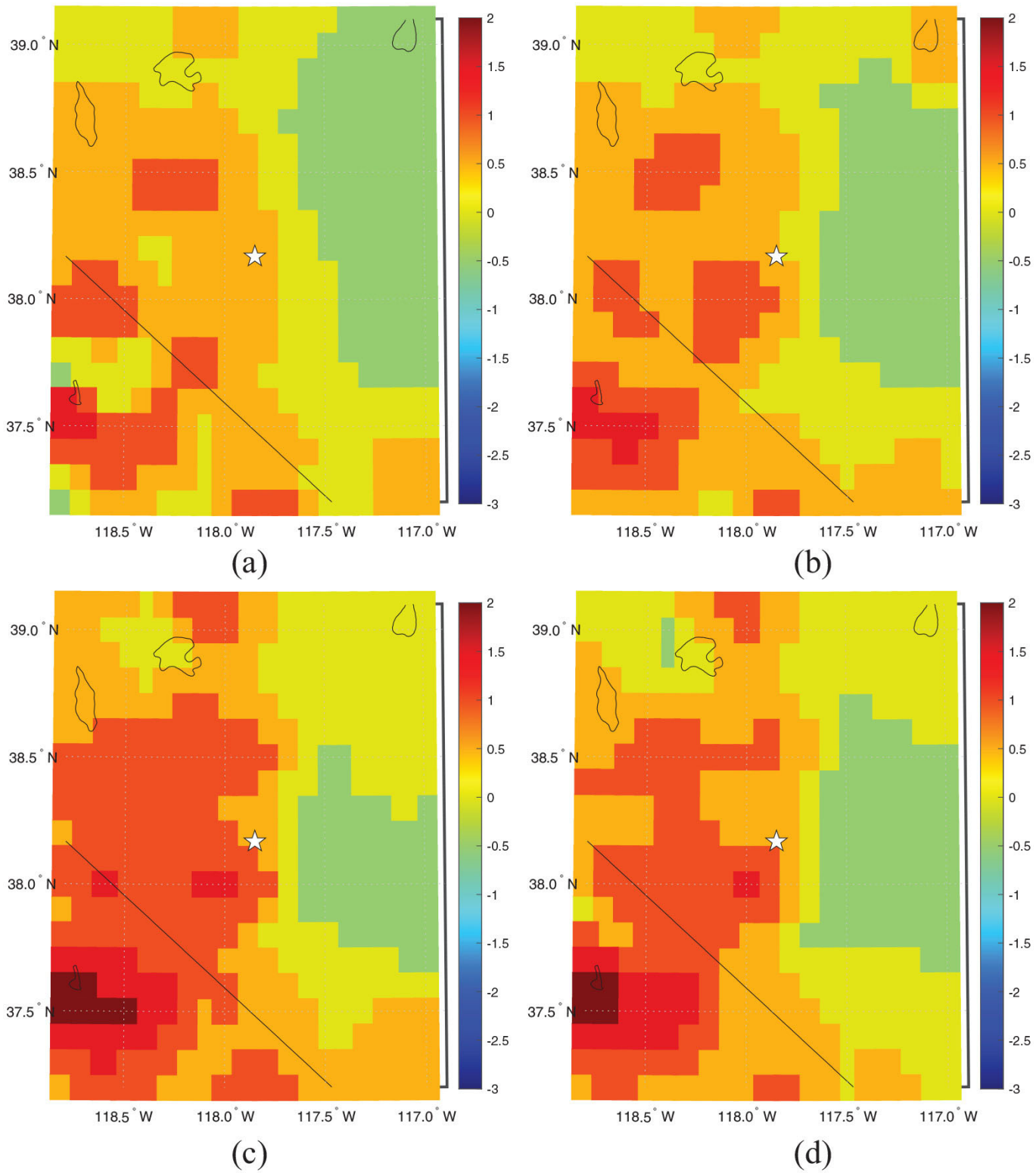

Figure 11. *Adaptive smoothing of three seismicity catalogs in the vicinity of the Monte Cristo Range earthquake. (a) USGS a-grid. (b) Our approximation of the a-grid using the USGS C3 catalog with a simplified model, as described in the text. (c) The a-grid obtained from the MSEQ-PGA catalog using the approximations used for part b. (d) The a-grid obtained from the MSEQ-SA (1.0 s) catalog using the approximations used for part b.*

resources'' section). The second is the USGS C3 catalog (see ''Data and resources'' section), which is developed by the USGS by declustering the C2 catalog using the Gardner and Knopoff (1974) method. The other four are the MSEQ catalogs based on the Bindi et al. (2017) ground motion prediction equations for peak acceleration and for SA at 0.2, 1.0, and 3.0 s. All of these MFDs were developed as described in the Online Supplement (Finding MFDs). The MFDs for PGA and for SA (0.2 s) are very similar, with the PGA model retaining about 5% to 6% more small earthquakes. The total number of earthquakes in these two MSEQ seismicity models is about half way between the complete catalog and the C3 catalog. The MFDs for the MSEQ catalogs for SA (1.0 s) and SA (3.0 s) are considerably depleted in small magnitudes by comparison, but converge to the MFD of the C2 catalog for magnitudes greater than about 5.5. The MFD of the C3 catalog only converges with that of the C2 catalog for magnitudes

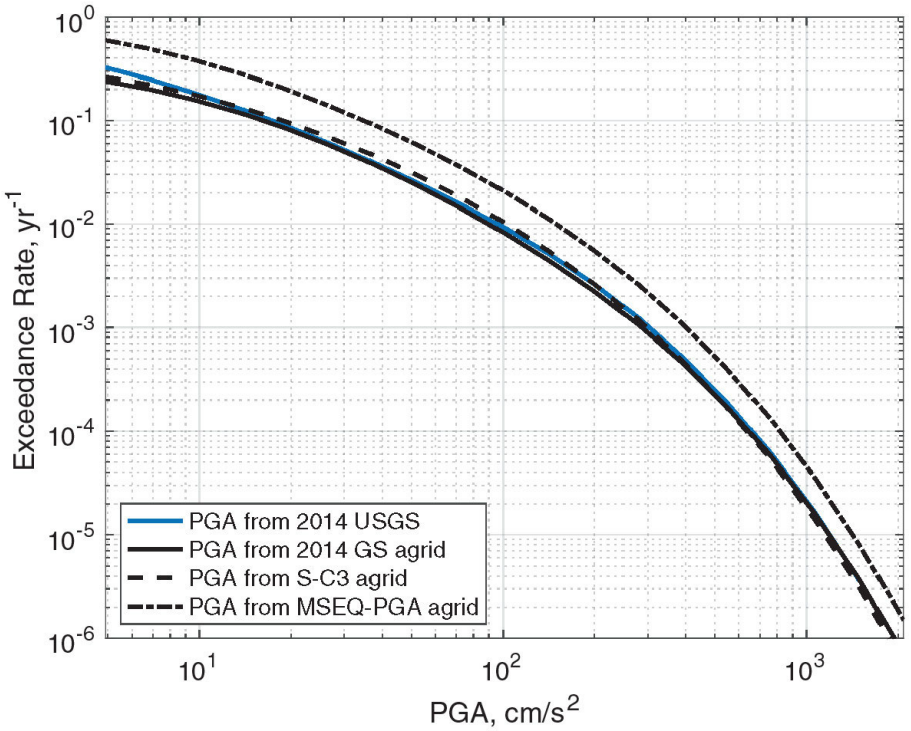

Figure 13. *Hazard estimates for SA (1.0 s) for the location of the Monte Cristo Range earthquake. All hazard curves are computed using only the ASK14 GMPE.*

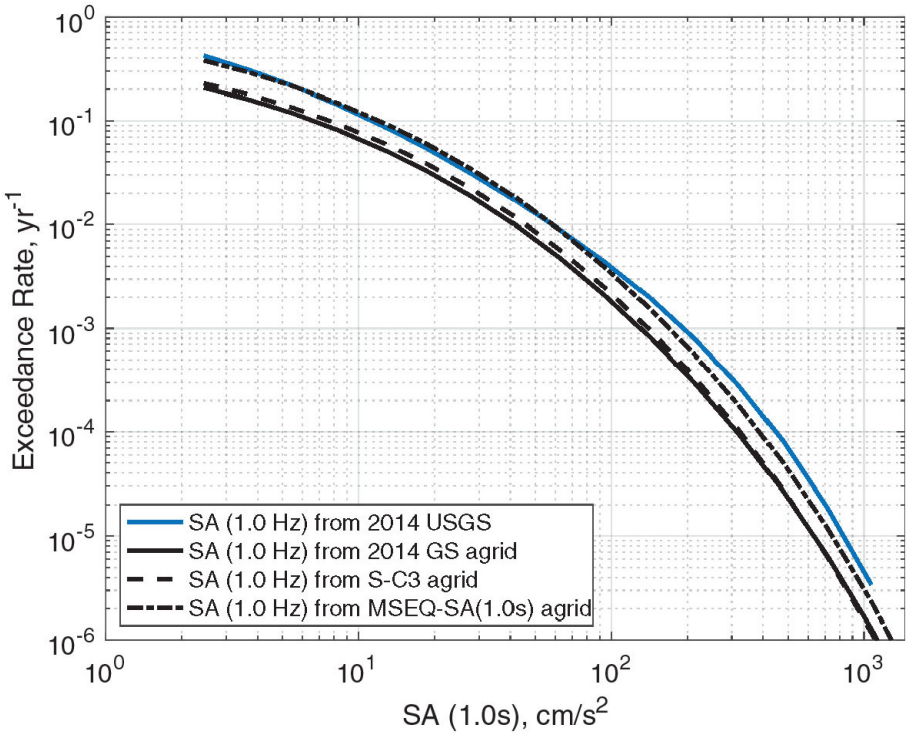

Figure 13. *Hazard estimates for SA (1.0 s) for the location of the Monte Cristo Range earthquake. All hazard curves are computed using only the ASK14 GMPE.*

greater than about 7.0. The difference is because the MSEQ catalogs include some foreshocks, and aftershocks that cause ground motions that are locally stronger than the main shock ground motions. Consequently, MFDs of all of the MSEQ catalogs exceed that for the C3 catalog for magnitudes greater than about 4.7. The b values are estimated by least squares over the domain $4 \leq M \leq 7$. For catalog C2, b = 0.91, while the b value of catalog C3 is found to be 0.80, consistent with the USGS findings. The b values of the MSEQ-PGA and MSEQ-SA (0.2 s) catalogs by the same method are 0.88, in between these two values (see Figures 11 to 13).

From these catalogs, the hazard is estimated for the location of the Mw6.5 Monte Cristo Range earthquake of May 15, 2020. The epicenter is in central Nevada (38.169°N, 117.850°W). This location is chosen because the causative fault is not in

USGS National Seismic Hazard Model (NSHM), so the hazard estimate at this location is dominated by the background seismicity. The procedure in Petersen et al. (2014) to calculate the hazard from the declustered catalogs is actually very complex. This sample calculation uses several simplifications. To explain these simplifications, it is helpful to briefly summarize the NSHM approach. The NSHM has both fault and area sources. The area sources are inserted in a logic tree and developed from the seismicity model by two different smoothing models, the ''fixed'' and ''adaptive'' models. This smoothed seismicity is supplemented by additional, overlapping area source zones drawn for considered geophysical reasons. The combined area source contributions are expressed as an ''agrid,'' that is, the value of a in an incremental Gutenberg–Richter model, $\log n\,(M) = a - bM$, where $n(M)$ is the number of earthquakes in a magnitude range of width ΔM = 0.1. Each element of the ''a-grid'' applies to the Gutenberg-Richter model in a rectangular area that is 0.1 degrees longitude and 0.1 degrees latitude. The hazard is calculated from this seismicity model using five different ground motion prediction equations, and then averaged.

The NSHM ''a-grid'' for region around the selected location is shown in Figure 11a. The NSHM uses the b value of the C3 catalog, b = 0.8, so that same value is used for every smoothing in this exercise. Our first simplification is to calculate an a-grid using only the adaptive smoothing model (as described in Petersen et al., 2014) and omitting the additional area source contributions. The result is shown in Figure 11b. This simplified algorithm was also applied to the MSEQ catalogs, and the resulting a-grids for the MSEQPGA and MSEQ-SA (1.0 s) catalogs are shown in Figure 11c and d. There are some differences between the a-grids in Figure 11a and b due to the simplifications that are used. The a-grids associated with the MSEQ catalogs tend to be higher, especially in the southwest part of these maps.

Hazard curves for the exceedance rate of peak acceleration are shown in Figure 12. The blue line gives the hazard curve obtained from USGS for the 2014 NSHM. The others are approximations using simplified calculations. First, the solid line shows the hazard for PGA using the NSHM a-grid in Figure 11a, but using only one of the GMPEs (Abrahamson et al., 2014). This result is close to the complete NSHM model. This is evidence that the PGA hazard is dominated by the background hazard estimate, and that using only one GMPE does not greatly distort the hazard estimate. The dashed curve is obtained from the smoothed C3 catalog shown in Figure 11b. Again, the similarity of this to the USGS hazard is evidence that this simplified approach to estimating the a-grid gives a reasonable approximation. Finally, the dash–dot line obtained from the MSEQ-PGA catalog is higher, consistent with the difference in the level of seismicity in the corresponding MFD. Similar calculations using the fixed distance smoothing kernel give results very similar to these, so they are not shown.

This process was repeated for SA (1.0 s). As noted previously, Figure 11d shows a somewhat reduced seismicity rate compared to Figure 11c, but the overall impression is that the two models are similar. Figure 13, on the contrary, shows some significant changes. The first thing to note is that the hazard from the USGS a-grid is nearly the same as the hazard from the a-grid derived in the simplified approach in Figure 11b. Thus, the simplifications in developing the a-grid are again having little effect on the result. However, both of these hazard curves lie below the USGS hazard curve. The cause is our omission of the fault sources. While these sources did not contribute much to the hazard for PGA in Figure 12, they make an important contribution to the hazard at longer periods in Figure 13. These sources can have larger earthquakes, and the long period motions do not decrease as rapidly with distance as does PGA.

The curve from the MSEQ-SA (1.0 s) catalog actually agrees better with the USGS hazard curve than do the curves for the background seismicity based on the C3 catalog. But this agreement is caused by significant offsetting differences. The MSEQ-SA (1.0 s) curve is biased low due to lack of including the regional faults, but is higher than the comparable background seismicity curves in the NSHM, apparently due to the inclusion of more large events.

Data and resources

The US Geological Survey earthquake catalogs are available from https://www.sciencebase.gov/catalog/item/5db9be62e4b06957974eb5ca, last accessed March 17, 2020.