Have We Seen the Largest Earthquakes in Eastern North America?

Miguel Merino and Seth Stein, Department of Earth & Planetary Sciences, Northwestern University Evanston, Illinois, USA

Seth Stein, Department of Earth & Planetary Sciences, Northwestern University Evanston, Illinois, USA

John Adams, Geological Survey of Canada, Ottawa, K1A 0Y3 Canada

Corresponding author: M. Merino, Department of Earth and Planetary Sciences Technological Institute, 2145 Sheridan Road, Evanston IL, 60208, USA. (miguel@earth.northwestern.edu)

Abstract

The assumed magnitude of the largest future earthquakes, M_{max}, is crucial in assessing seismic hazard, especially for critical facilities like nuclear power plants. Absent any theoretical basis, estimates are made using various methods and often prove far too low, as for the 2011 Tohoku, Japan, earthquake. Estimating M_{max} is particularly challenging within tectonic plates, where large earthquakes are infrequent compared to the length of the available earthquake history, vary in space and time, and sometimes occur on previously unrecognized faults. For example, it is unclear whether the largest earthquakes along the eastern U.S seaboard and eastern Canada have occurred. We explore this issue by generating synthetic earthquake histories and sampling them over a few hundred years. The maximum magnitudes appearing most often in the simulations are essentially those observed, and smaller than the simulation maxima. Future earthquakes along both coasts may thus be significantly larger than those observed to date.

Introduction

The 2011 Virginia earthquake that shook much of the northeastern U.S. showed that earthquakes large enough to cause significant damage occur in eastern North America [*Wolin et al.,* 2012] (Fig. 1). Assessing the hazard of such earthquakes poses major unresolved issues. Hazard maps, giving the maximum shaking expected in an area with a certain probability in some time period [*Cornell,* 1968], require assuming where and when large earthquakes will occur and how large they will they be. However, the recent Tohoku, Haiti, and Wenchuan earthquakes illustrate that highly destructive earthquakes can occur in areas predicted to be relatively safe [*Geller,* 2011; *Gulkan,* 2013; *Kerr,* 2011; *Peresan and Panza,* 2012; *Stein and Okal,* 2011; *Wyss et al.,* 2012].

This situation arises because parameters required to reliably estimate the hazards are often poorly known [*Stein et al.,* 2012].

A crucial parameter is M_{max}, the magnitude of the largest earthquake expected on a fault or in an area [*Stein et al.,* 2012]. The Tohoku, Haiti, and Wenchuan earthquakes were more damaging than expected because they were much larger than the M_{max} assumed. Unfortunately, no theoretical basis exists to infer M_{max}. Even where we know the long-term rate of motion across a plate boundary fault, or the deformation rate across an intraplate zone, neither predict how strain will be released.

As a result, quite different M_{max} estimates can be made [*Kagan and Jackson,* 2013; *Kijko,* 2004; *U.S Nuclear Regulatory Commission,* 2012]. Because all one can say with certainty is that M_{max} is at least as large as the largest earthquake in the available record, one could use that magnitude or add an increment. However, because catalogs are often short relative to the average recurrence time of large earthquakes [*McGuire,* 1977], larger earthquakes than anticipated often occur. Another approach is to identify faults and use relations between fault length and earthquake magnitude [*Wells and Coppersmith,* 1994] to infer M_{max}. Other approaches extrapolate from current catalogs [*Kijko,* 2004] or combine areas presumed to be geologically similar to sample more large earthquakes [*Kagan and Jackson,* 2013; *U.S Nuclear Regulatory Commission,* 2012].

Estimating M_{max} is challenging at plate boundaries, where known plate motion rates can be compared to earthquake records on known faults to infer the slip in, and thus magnitude of, large earthquakes. The situation is more complicated within plates, where deformation rates are poorly

known, large earthquakes are rarer and variable in space and time, and often occur on previously unrecognized faults [*Camelbeeck et al.,* 2007; *Crone et al.,* 2003; *Liu et al.,* 2011].

We explore this issue for eastern North America. Notable events along the U.S. coast include the 1755 Cape Ann Massachusetts and 1886 Charleston earthquakes. Larger earthquakes are known along the Canadian coast, notably the 1929 Grand Banks and 1933 Baffin Bay events. Thus this passive continental margin has a modest level of seismicity [*Stein et al.,* 1979; *Wolin et al.,* 2012] as do others [*Schulte and Mooney,* 2005; *Stein et al.,* 1989].

A challenge in assessing the earthquakes' hazard is that we know little about their causes, partly because they are relatively rare due to the slow deformation at such margins. They may reflect reactivation of faults created by previous continental collision and breakup, given that passive margins are often reactivated [*Cloetingh et al.,* 2008]. Geodynamic modeling predicts stresses from variations in topography and crustal structure across the margin, combined with sublithospheric mantle flow [*Ghosh et al.,* 2013].

A crucial issue is how much to rely on past large earthquakes, as illustrated by Geological Survey of Canada hazard maps [*Wolin et al.,* 2012]. The 1985 map concentrates hazard at the sites of the Grand Banks and Baffin Bay earthquakes, assuming these are especially hazardous. The 2005 map assumes that similar earthquakes can occur anywhere along the margin.

The observed seismicity being an imperfect sample of more uniform seismicity is plausible geologically and suggested by seismicity between the Grand Banks and Baffin Bay, some of

which may be aftershocks of prehistoric earthquakes [*Wolin et al.,* 2012]. Simulations with short catalogs yield apparent concentrations and gaps that are artifacts of the sampling [*Swafford and Stein,* 2007]. Similarly, although seismicity onshore is patchy, geological observations show evidence of slow long-term deformation [*Pazzaglia et al.,* 2010] and the present seismicity occurs in areas that are not geologically or geomorphologically different from nearby areas that appear aseismic.

A related question is whether the larger earthquakes and higher seismicity along the Canadian coast represent a real difference. The difference could be real, perhaps due to stresses associated with deglaciation, [*Mazzotti et al.,* 2005; *Sella et al.,* 2007; *Wolin et al.,* 2012] or reflect the short earthquake record.

Results and Analysis

Absent reliable ways of assessing M_{max}, we use synthetic earthquake histories to explore what M_{max} values would be observed in a short catalog. We assume earthquakes satisfy a Gutenberg-Richter frequency-magnitude relation, $\log_{10} N = a - bM$, where N is the annual number of earthquakes with magnitude $\geq M$, a defines the seismicity rate, and b is the slope of the line relating the rates of small and large earthquakes. The recurrence interval between events for each magnitude is described by a Gaussian distribution about the predicted mean rate. Thus events with magnitude $\geq M$ have average recurrence time $T_{rM}=10^{-(a-bM)}$ years, with a standard deviation of $0.4T_{rM}$. Each simulation has a specified M_{max} above which no earthquakes occur.

For the U.S. coast we use a=4.24 and b=0.92, calculated from recent seismicity, corresponding to a $M \geq 7$ earthquake on average every ~150 years. We generate four sets of 10,000 histories, assuming that M_{max} is 7.0, 7.2, 7.4, and 7.6. We take a 300-year sample of each history, corresponding to the period over which all large ($M \geq 6$) earthquakes are likely to be known. Samples start 5000 years into the simulation to ensure that 'now' has no significance.

Fig. 2 illustrates three representative histories. These are long enough to adequately represent the rates of small earthquakes, which define a line corresponding to the parent distribution. However, because the 300-year sampling interval corresponds to the recurrence interval of M=7.3 earthquakes, larger events are undersampled.

As a result, two biases can arise. In one, the largest "observed" earthquake is smaller than in the parent distribution, so b is biased upward, underestimating the rate of large earthquakes. In the other, the largest earthquakes are "observed", but their recurrence interval and thus b are underestimated. Hence the short history causes us to either underestimate M_{max} when earthquakes of this size do not appear, or correctly estimate it but conclude that such earthquakes are more common than they really are [*Stein and Newman,* 2004]. These effects can be visualized by considering the parent distribution of recurrence times. Records shorter than the mean recurrence time of the largest events are likely to not contain these events. If they contain such events, these have recurrence times shorter than the mean, implying that they are more common than actually the case.

Fig. 2 also illustrates different ways of inferring the Gutenberg-Richter distribution's parameters. If short sampling causes the rate of large earthquakes to deviate from that extrapolated from small ones, least-squares fitting characterizes larger magnitudes better and thus may not match the rate of the smaller ones well. Maximum likelihood estimation [*Aki,* 1965; *Weichert,* 1980] weights the more populous lower magnitudes heavily and thus can misfit larger magnitudes. Analyses seeking to estimate b typically use maximum likelihood, as discussed in the supplemental information. The two methods give different b values and hence different recurrence time predictions for large events.

Fig. 3 shows M_{max} values "observed" for the U.S. east coast earthquake histories. The most common values vary depending on the assumed M_{max}, but is ~7.3 for simulations with large M_{max}, corresponding to 300-year sampling. Even when M_{max} is significantly larger, it rarely appears. For example, M_{max} =7.6 occurs in only 10% of the simulations. The most common "observed" value is close to that inferred for the Charleston earthquake.

Fig. 3 shows analogous results for eastern Canada for a=4.44 and b=0.94, calculated from recent seismicity, corresponding to an earthquake with M≥7.2 on average every 212 years. Histories for M_{max}=7.4, 7.6, 7.8, and 8.0 are sampled for 100 years, the period over which all large (M≥6) earthquakes are likely to be known. The most common M_{max} values "observed," about 6.8, reflect the sample length even when the parent distribution M_{max} is significantly larger. For example, M_{max} =8.0 appears in only ~2% of the runs. The most common "observed" value is lower than the Grand Banks and Baffin Bay earthquakes, but observing M=7.4 like Baffin Bay's is not extraordinarily rare.

Discussion

Hence future earthquakes along both coasts may be larger than observed to date. Some of the higher Canadian seismicity rate may reflect aftershocks [*Wolin et al.,* 2012] of recent large earthquakes. M_{max} for both coasts may be the same, although if the higher seismicity in Canada is real rather than a sampling artifact, large events would be more common there.

More generally, these simulations demonstrate that M_{max} cannot be reliably estimated from earthquake catalogs. The largest earthquake observed likely reflects the length of the history used, even if larger earthquakes occur (Fig. 4). Although the precise fraction depends on the distribution of recurrence times, a catalog shorter than an earthquake's mean recurrence time is likely to not contain an event of that size.

This effect is significant within plates where large earthquakes are infrequent, as in northwestern Europe [*Leonard et al.,* 2013] or Australia [*Camelbeeck,* 2007]. However, as recent examples illustrate, it also arises at plate boundaries for earthquakes larger than observed in the record used. We can lengthen the record with paleoseismology and use geodesy to infer the size of future events. Even so, although various plausible assumptions about $M_{max,}$ can be made, the only certainty is that it is at least as large as that observed to date, as in the adage "anything that did happen, can happen."

Acknowledgments: We thank the U.S. Geological Survey John Wesley Powell Center for Analysis and Synthesis for hosting a working group that inspired this work, and the group participants for their discussions. We thank the Northern California Earthquake Data Center

(NCEDC), Advanced National Seismic System (ANSS), and Canadian National Earthquake Database (NEDB) for earthquake catalog data. Stein thanks the Alexander von Humboldt Foundation for supporting his stay at the University of Gottingen, where much of the writing was done.

References

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

Camelbeeck, T., K. Vanneste, P. Alexandre, K. Verbeeck, T. Petermans, P. Rosset, M. Everaerts, R. Warnant, and M. Van Camp (2007), Relevance of active faulting and seismicity studies to assessments of long-term earthquake activity and maximum magnitude in intraplate northwest Europe, between the Lower Rhine Embayment and the North Sea, in Continental Intraplate Earthquakes: Science, Hazard, and Policy Issues, edited by S. Stein and S. Mazzotti, pp. 193-224, Geological Soc Amer Inc, Boulder.

Cloetingh, S., F. Beekman, P. A. Ziegler, J. D. Van Wees, and D. Sokoutis (2008), Post-rift compressional reactivation potential of passive margins and extensional basins, in Nature and Origin of Compression in Passive Margins, edited by H. Johnson, T. G. Dore, R. W. Gatliff, R. W. Holdsworth, E. R. Lundin and J. D. Ritchie, pp. 27-70, Geological Soc Publishing House, Bath.

Cornell, C.A. (1968), Engineering seismic risk analysis. *Bull. Seism. Soc. Am*., 58, 1583-1606.

Crone, A. J., P. M. De Martini, M. N. Machette, K. Okumura, and J. R. Prescott (2003), Paleoseismicity of two historically quiescent faults in Australia: Implications for fault behavior in stable continental regions, Bull. Seismol. Soc. Amer., 93(5), 1913-1934.

Geller, R. J. (2011), Shake-up time for Japanese seismology, Nature, 472(7344), 407-409.

Ghosh, A., W. E. Holt, and L. M. Wen (2013), Predicting the lithospheric stress field and plate motions by joint modeling of lithosphere and mantle dynamics, Journal of Geophysical Research-Solid Earth, 118(1), 346-368.

Gulkan, P. (2013), A Dispassionate View of Seismic-Hazard Assessment, Seismological Research Letters, 84(3), 413-416.

Kagan, Y. Y., and D. D. Jackson (2013), Tohoku Earthquake: A Surprise?, Bull. Seismol. Soc. Amer., 103(2B), 1181-1194.

Kerr, R. A. (2011), Seismic Crystal Ball Proving Mostly Cloudy Around the World, Science, 332(6032), 912-913.

Kijko, A. (2004), Estimation of the maximum earthquake magnitude, Mmax, Pure and Applied Geophysics, 161(8), 1655-1681.

Leonard, M. et al. (2013), The challenges of probabilistic seismic hazard assessment in stable continental interiors - an Australian example. *Bull. Seism. Soc. Am.,* subm.

Liu, M., S. Stein, and H. Wang (2011), 2000 years of migrating earthquakes in North China: How earthquakes in midcontinents differ from those at plate boundaries, Lithosphere, 3(2), 128-132.

Mazzotti, S., T. S. James, J. Henton, and J. Adams (2005), GPS crustal strain, postglacial rebound, and seismic hazard in eastern North America: The Saint Lawrence valley example, Journal of Geophysical Research-Solid Earth, 110(B11), 16.

McGuire, R. K. (1977), Effects of uncertainty in seismicity on estimates of seismic hazard for the east coast of the United States, Bull. Seismol. Soc. Amer., 67(3), 827-848.

Pazzaglia, F. et al. (2010) Tectonics and topography of the Cenozoic Appalachians, in *Tectonics of the Susquehanna Piedmont,* eds. Wise D. & Fleeger, G., 111-126 (2010).

Peresan, A., and G. F. Panza (2012), Improving earthquake hazard assessments in Italy: An alternative to “Texas sharpshooting”, Eos, Transactions American Geophysical Union, 93(51), 538-538.

Schulte, S. M., and W. D. Mooney (2005), An updated global earthquake catalogue for stable continental regions: reassessing the correlation with ancient rifts, Geophysical Journal International, 161(3), 707-721.

Sella, G. F., S. Stein, T. H. Dixon, M. Craymer, T. S. James, S. Mazzotti, and R. K. Dokka (2007), Observation of glacial isostatic adjustment in "stable" North America with GPS, Geophys. Res. Lett., 34(2), 6.

Stein, S., and A. Newman (2004), Characteristic and uncharacteristic earthquakes as possible artifacts: Applications to the New Madrid and Wabash seismic zones, Seismological Research Letters, 75(2), 173-187.

Stein, S., and Okal, E. A. (2011), The size of the 2011 Tohoku earthquake needn't have been a surprise. *Eos Trans. AGU* 92, 227-228.

Stein, S., R. J. Geller, and M. Liu (2012), Why earthquake hazard maps often fail and what to do about it, Tectonophysics, 562–563(0), 1-25.

Stein, S., S. Cloetingh, N. H. Sleep, and R. Wortel (1989), Passive margin earthquakes, stresses and rheology, 231-259 pp., Kluwer Academic Publ, Dordrecht.

Stein, S., N. H. Sleep, R. J. Geller, S. C. Wang, and G. C. Kroeger (1979), Earthquakes along the passive margin of eastern Canada, Geophys. Res. Lett., 6(7), 537-540.

Swafford, L., and S. Stein (2007), Limitations of the short earthquake record for seismicity and seismic hazard studies, in Continental Intraplate Earthquakes: Science, Hazard, and Policy Issues, edited by S. Stein and S. Mazzotti, pp. 49-58, Geological Soc Amer Inc, Boulder.

U.S Nuclear Regulatory Commission (2012) Central and Eastern United States Seismic Source Characterization for Nuclear Facilities (NUREG-2115). EPRI, Palo Alto, CA, U.S. DOE, and U.S. NRC, http://www.ceus-ssc.com.

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

Weichert, D. H. (1980), Estimation of the earthquake recurrence parameters for unequal observation periods for different magnitudes, Bull. Seismol. Soc. Amer., 70(4), 1337-1346.

Wolin, E., S. Stein, F. Pazzaglia, A. Meltzer, A. Kafka, and C. Berti (2012), Mineral, Virginia, earthquake illustrates seismicity of a passive-aggressive margin, Geophys. Res. Lett., 39, 7.

Wyss, M., A. Nekrasova, and V. Kossobokov (2012), Errors in expected human losses due to incorrect seismic hazard estimates, Nat. Hazards, 62(3), 927-935.

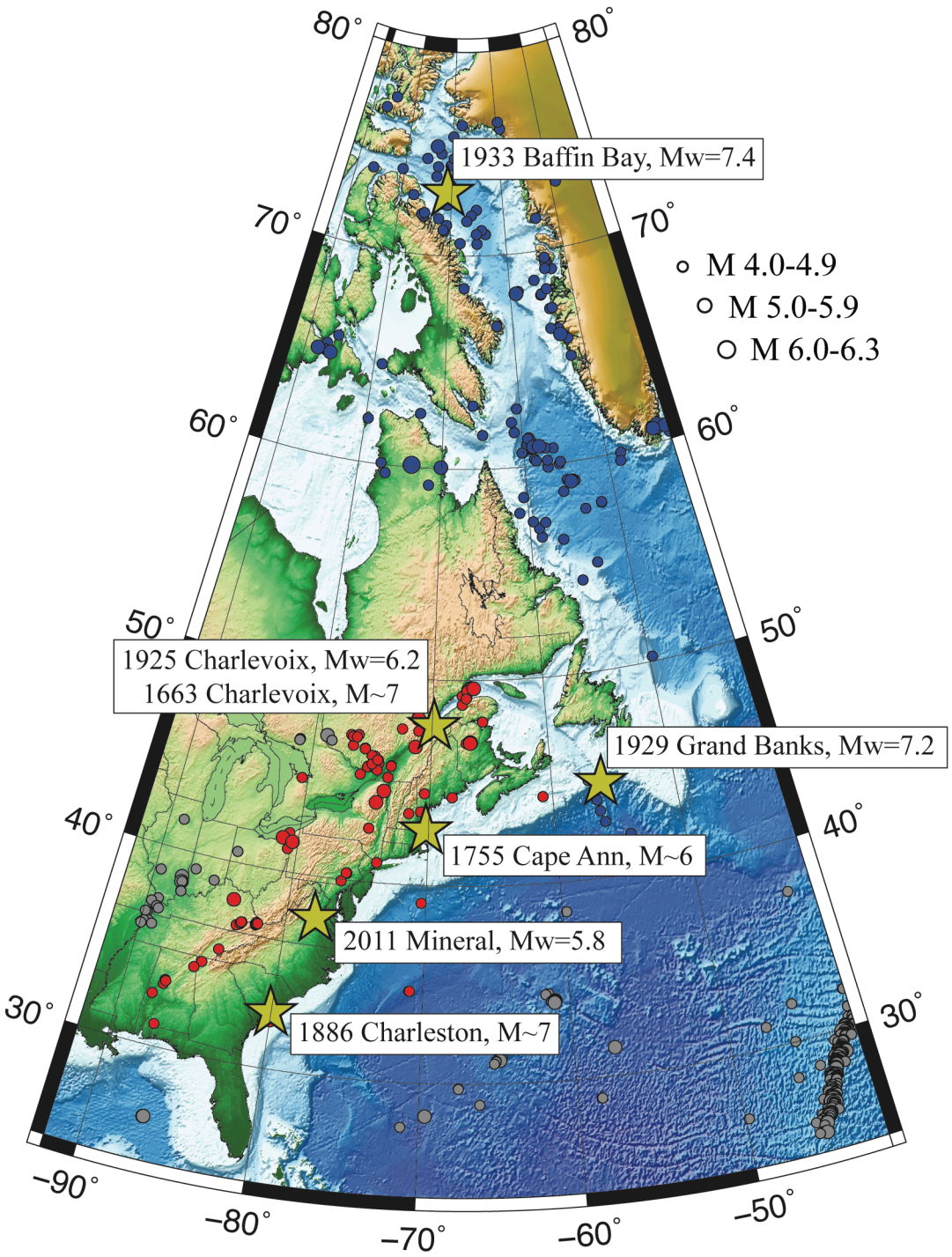
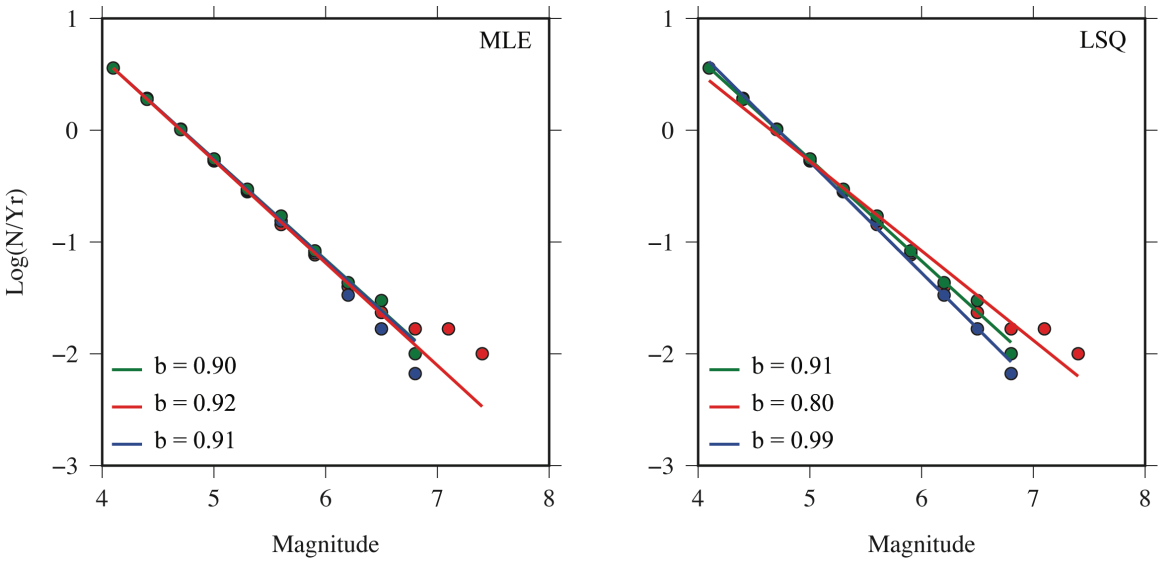


Figure 1: Seismicity of the eastern North American continental margin taken from ANSS and NEDB catalogs. Red and blue dots correspond to the U.S and Canadian margins respectively. Major historical events are also shown. Modified from (8). Earthquake selection is discussed in supplemental information.



297

298

299 Figure 2: Frequency-magnitude results for three of the histories (different colors) generated with
300 the eastern U.S simulations with M_{max}=7.6. These histories are long enough to adequately
301 represent the rates of small earthquakes. However, because the 300-year sampling interval
302 corresponds to the recurrence interval of magnitude ~7.3 earthquakes, larger events with longer
303 recurrence times are undersampled. As a result, maximum likelihood (MLE) and least-squares
304 (LSQ) fits to the data differ.

305

306

307

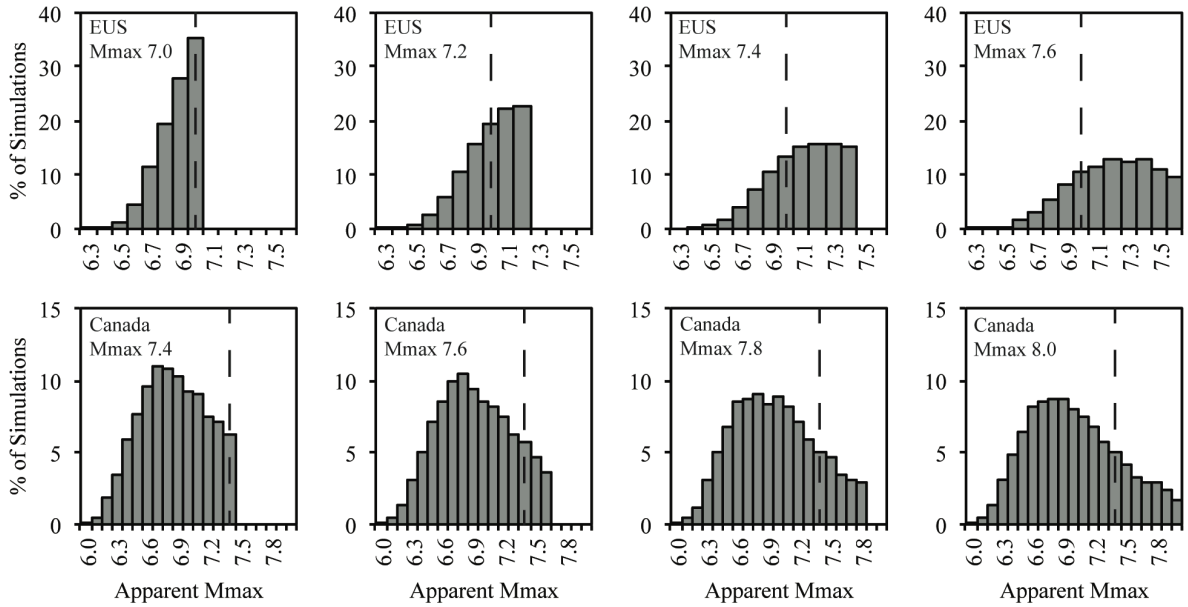


Figure 3: Top: Results for four sets of 10,000 simulations for the eastern U.S coast, each with a different M_{max}. Panels show the percentage of simulations in which a given apparent M_{max} is observed. Dashed line represents the actual observed M_{max} of 7.0, corresponding approximately to the Charleston earthquake. Bottom: Simulations for eastern Canada. Dashed line represents the actual observed M_{max} of 7.4 from the Baffin Bay earthquake.

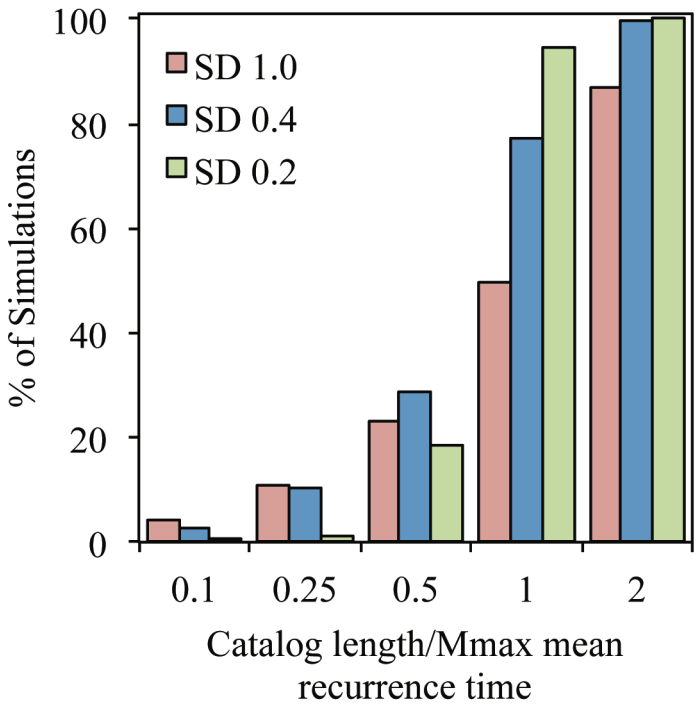


Figure 4: Percentages of simulations that recover the input M_{max} for varying catalog lengths. Catalog lengths are given as a fraction of the mean recurrence time for earthquakes with magnitude M_{max}. Colors show results for Gaussian distributions of recurrence times with standard deviation equal to the indicated fraction of the mean. A catalog shorter than an earthquake's mean recurrence time is likely to not contain an event of that size.

335 ***Supplementary Information***

336 The regional *a* and *b* values were estimated from the Advanced National Seismic System
337 (ANSS) from 1985-2013 for the eastern U.S, and the Canadian National Earthquake Database
338 (NEDB) from 1985-2013 for eastern Canada (Fig. 1). All earthquakes along the eastern
339 Canadian coastline and near Hudson Strait were selected to be eastern Canada passive margin
340 earthquakes. Eastern US earthquakes were selected by taking all earthquakes within 6° of the
341 coast. Earthquakes near the historical Grand Banks earthquake are assigned to the eastern
342 Canada catalog. The regional *a* and *b* values were calculated by the MLE method, which is
343 sensitive to the completeness of the catalog, with a magnitude cutoff of 4.0 for eastern Canada
344 and the eastern U.S (Fig. S1).

345

346 As noted in the text (Fig. 2), a linear Gutenberg-Richter relationship can be fit to a set of
347 frequency-magnitude data using either Maximum Likelihood (MLE) or Least Squares (LSQ)
348 estimates, each with advantages and disadvantages. In general, a least-squares fit characterizes
349 larger magnitudes better, but thus may not match the rate of the smaller ones well. Conversely,
350 MLE weights lower magnitudes heavily and so gives more stable estimates of the distribution's
351 *b*-value (slope). Thus if one expects that the data come from a Gutenberg-Richter distribution,
352 such that the deviations from the linear trend are artifacts of sampling or otherwise, MLE
353 estimation is preferable. As a result, most seismic hazard analyses use MLE. Conversely, if one
354 approaches the data without this expectation, LSQ can be viewed as a better characterization of
355 the data themselves.

356

In our simulations, the "data" come from a population with a Gutenberg-Richter distribution. As a result, MLE recovers the distribution's *b*-value (slope) better because it weights lower magnitudes heavily. Hence for any set of simulations, *b*-values estimated by MLE vary less than those inferred using LSQ.

Fig. S2 illustrates this effect for the set of simulations for the eastern U.S with M_{max} of 7.6. The simulated catalogs are the same for the MLE and LSQ calculations. MLE yields tightly grouped *b* values. In contrast, LSQ has a tradeoff between *b*-value and M_{max} because the line fits the larger magnitudes better. As a result, it is unlikely that it will recover the parameters of the parent distribution (dot) if the imposed M_{max} does occur.

These differences in the *b*-value distributions are shown by the standard deviations in the eastern U.S simulations with an M_{max} of 7.6, which are 0.01 and 0.06 for MLE and LSQ, respectively. The *b*-values calculated by MLE and LSQ both have Gaussian distributions (Fig. S3, S4). Hence the 95% confidence range for a M7.6 earthquake's recurrence interval Tr_M ranges from approximately 350 to 900 years, using for the MLE method. This ~2.5 fold difference in Tr_M would give an uncertainty in the estimated hazard. The combined uncertainties for the eight sets of simulations in Fig. 3 are illustrated in Fig. S5. As expected, *b*-values are recovered well whereas M_{max} is generally underestimated.

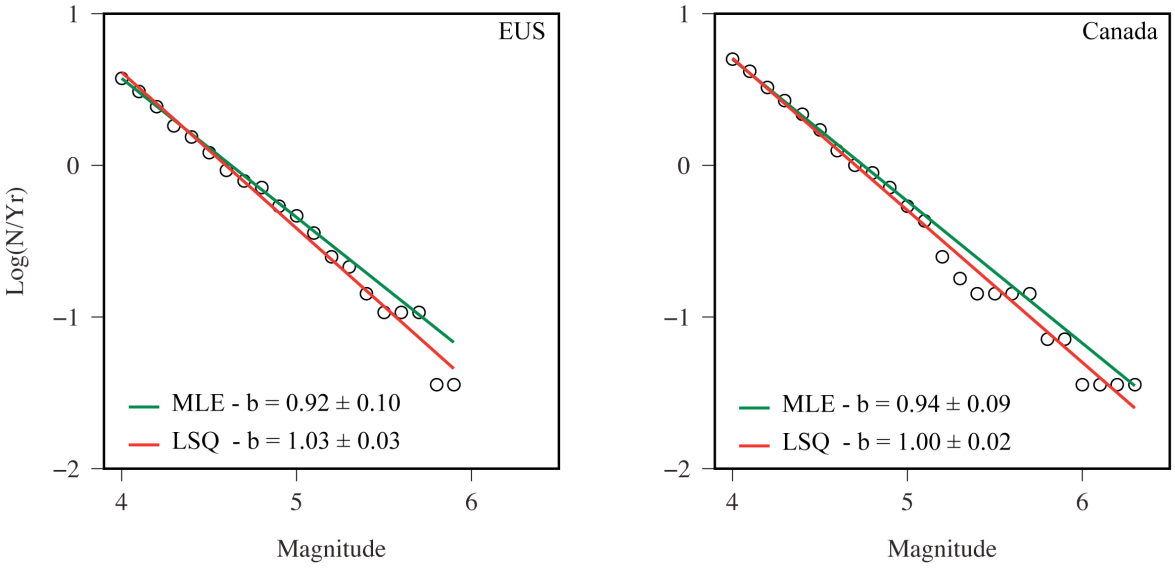


Figure S1: Frequency-magnitude relationships calculated for the eastern U.S (left panel) and eastern Canada (right panel). Green line is for the MLE fit. Red line is for the LSQ fit. The MLE results are used as inputs to the simulations.

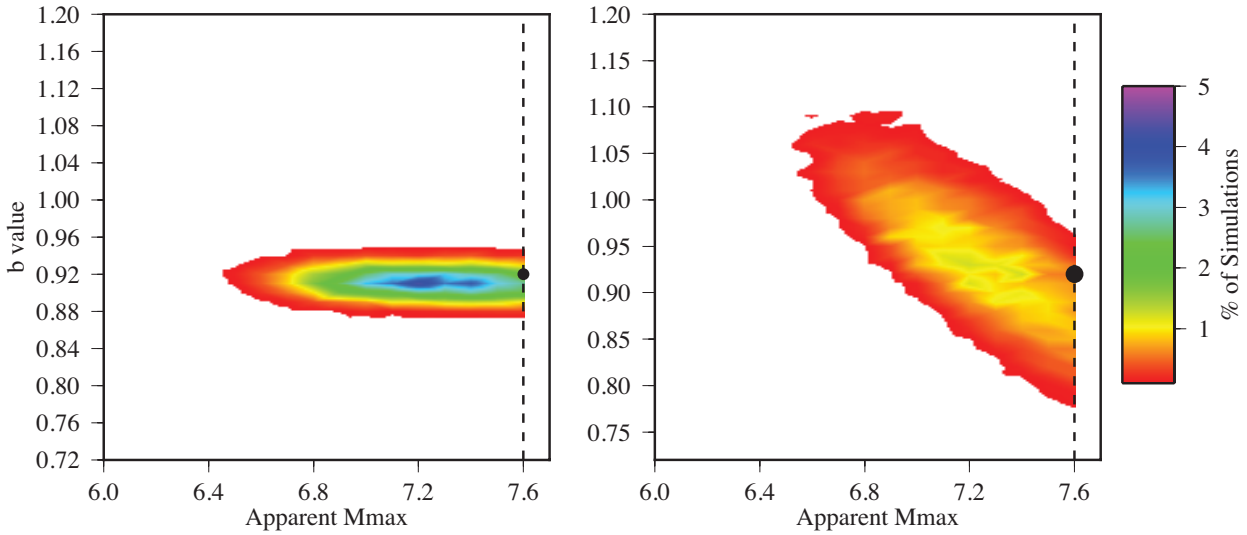
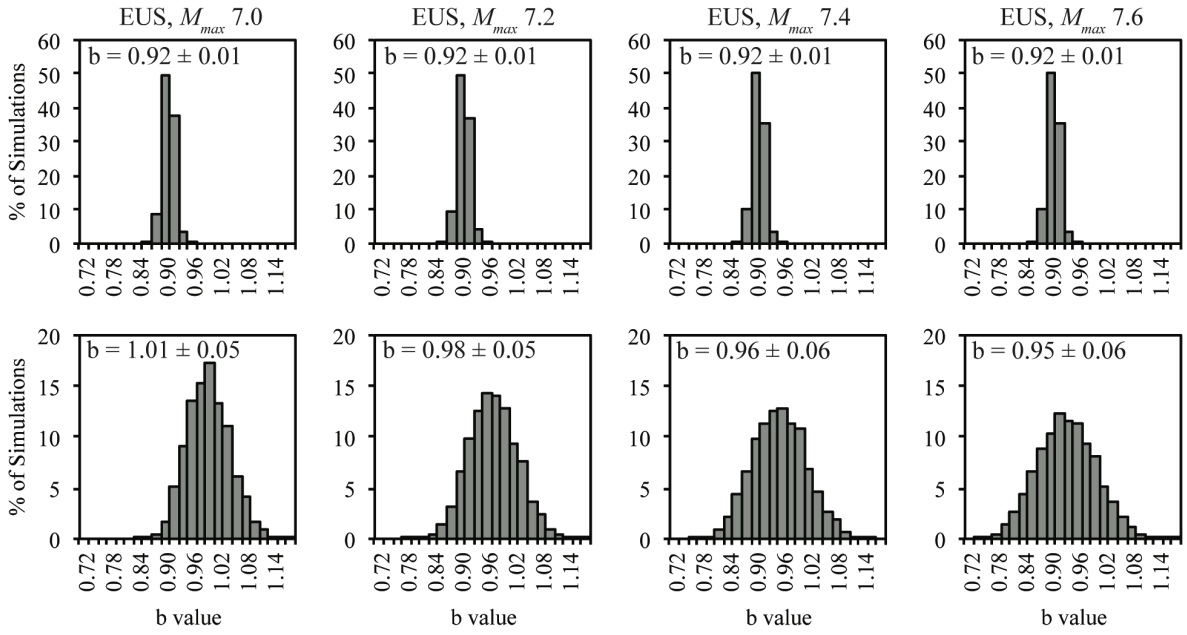
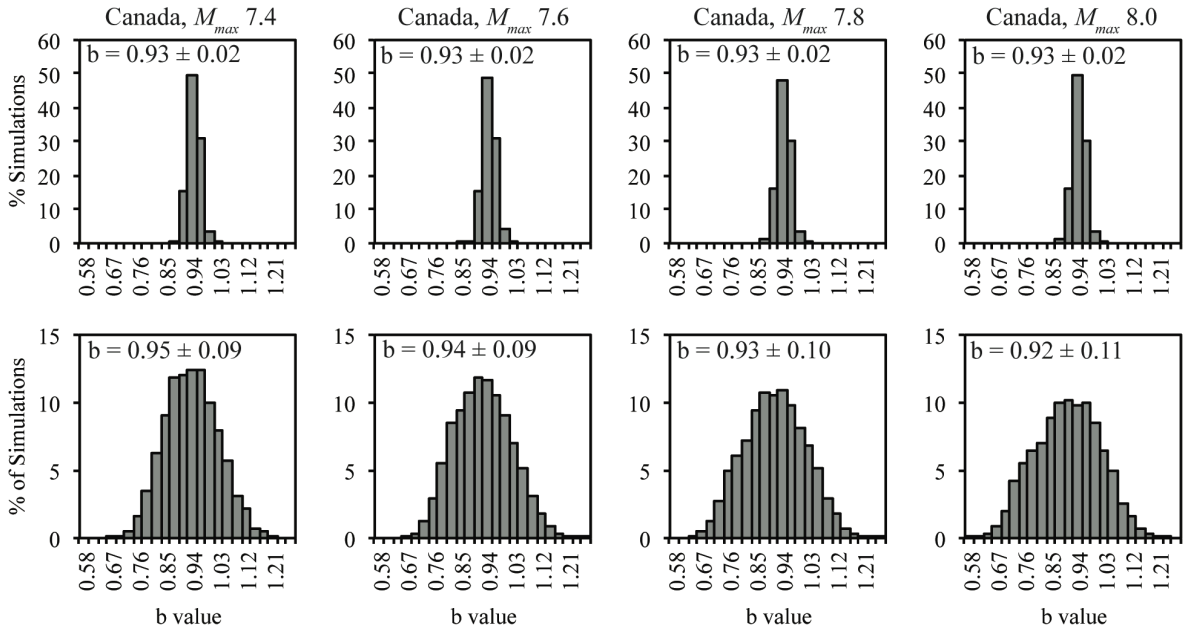


Figure S2: Distribution of apparent M_{max} and b-value results for a set of simulations sampling a parent distribution whose value is shown by the dot.



389

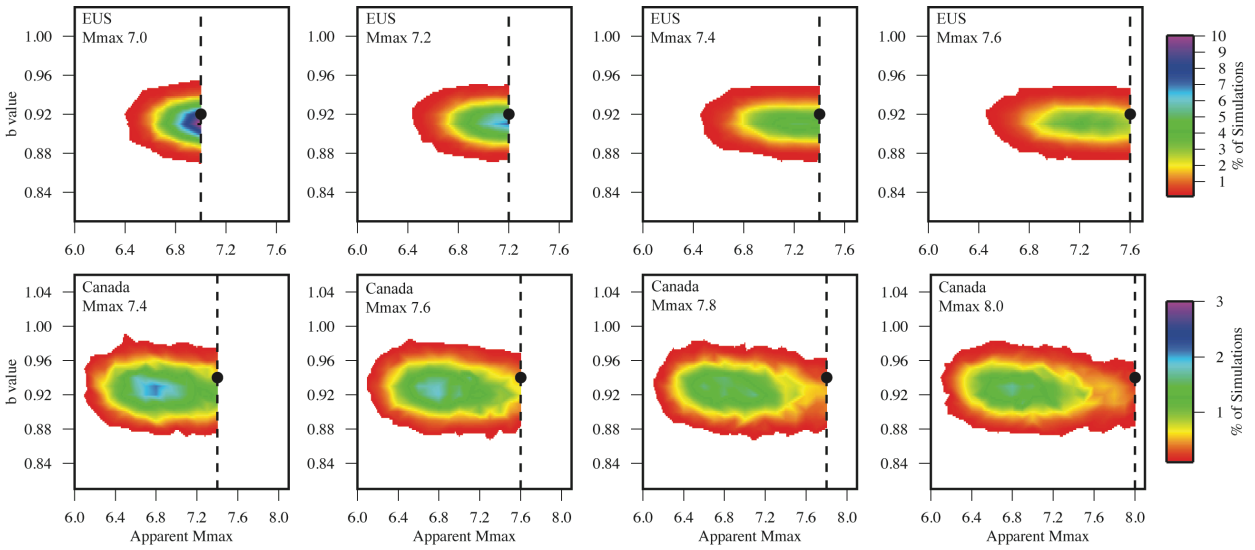
390 Figure S3: Top: *b*-value results for the four sets of 10,000 simulations for the eastern U.S coast,
391 each with a different M_{max}. Panels show the percentage of simulations in which a given *b* value,
392 calculated by the MLE method, is observed. Bottom: Results of the simulations for the eastern
393 U.S coast using the LSQ method to calculate *b* values.



394

395 Figure S4: Same as in figure S3 for eastern Canada.

396



397

398 Figure S5: Distributions of MLE estimates for the eight sets of simulations in Fig. 3. As

399 expected, *b*-values are recovered well whereas M_{max} is generally underestimated.

400

401

402

403

404

405

406