Zoller, G., Holschneider, M., Hainzl, S., Zhuang, J. (2014): The Largest Expected Earthquake Magnitudes in Japan: The Statistical Perspective. - *Bulletin of the Seismological Society of America*, *104*, 2, p. 769-779

DOI: http://doi.org/10.1785/0120130103

Bulletin of the Seismological Society of America, Vol. 104, No. 2, pp. 769–779, April 2014, doi: 10.1785/0120130103

The Largest Expected Earthquake Magnitudes in Japan: The Statistical Perspective

by Gert Zöller, Matthias Holschneider, Sebastian Hainzl, and Jiancang Zhuang

Abstract Earthquake catalogs are probably the most informative data source about spatiotemporal seismicity evolution. The catalog quality in one of the most active seismogenic zones in the world, Japan, is excellent, although changes in quality arising, for example, from an evolving network are clearly present. Here, we seek the best estimate for the largest expected earthquake in a given future time interval from a combination of historic and instrumental earthquake catalogs. We extend the technique introduced by Zöller *et al.* (2013) to estimate the maximum magnitude in a time window of length T_f for earthquake catalogs with varying level of completeness. In particular, we consider the case in which two types of catalogs are available: a historic catalog and an instrumental catalog. This leads to competing interests with respect to the estimation of the two parameters from the Gutenberg–Richter law, the b-value and the event rate λ above a given lower-magnitude threshold (the a-value). The b-value is estimated most precisely from the frequently occurring small earthquakes; however, the tendency of small events to cluster in aftershocks, swarms, etc. violates the assumption of a Poisson process that is used for the estimation of λ. We suggest addressing conflict by estimating b solely from instrumental seismicity and using large magnitude events from historic catalogs for the earthquake rate estimation. Applying the method to Japan, there is a probability of about 20% that the maximum expected magnitude during any future time interval of length $T_f = 30$ years is $m \geq 9.0$. Studies of different subregions in Japan indicates high probabilities for M 8 earthquakes along the Tohoku arc and relatively low probabilities in the Tokai, Tonankai, and Nankai region. Finally, for scenarios related to long-time horizons and high-confidence levels, the maximum expected magnitude will be around 10.

Introduction

The devastating Tohoku earthquake occurred on 11 March 2011 and triggered a tsunami leading to the nuclear catastrophe in Fukushima. The economic loss has been estimated between \$250 and \$500 billion (USD). In the aftermath of this event, the question has been raised as to whether or not this event, or in particular the size of this event, happened as a surprise. It is a well-known fact that subduction zones produce large or mega-earthquakes, but the frequency of such events is generally assumed to be low. In fact, the Tohoku earthquake was the largest event since A.D. 684 in Japan, that is, in a time window of 1329 years. The conclusion that the probability of occurrence of such an event is simply $1/1329$ per year fails, of course: Large events are rare events, and uncertainties related to rare events are enormous. Therefore, earthquake occurrence is usually modeled as a random process based on empirical probability distributions. This allows estimation of maximum expected magnitudes in time windows with a straightforward uncertainty management based on a predefined significance level. The significance level, which reflects the error probability that one is willing to accept, is subject to the individual requirements: the higher the loss potential is, the lower the error probability should be. For example, a higher error might be tolerable for individual residences than for nuclear power plants. In this study, we calculate probability distributions of large earthquake magnitudes, which might be useful for decision makers. However, we do not contribute to decision-making issues in terms of economical, social, or political consequences of earthquake occurrence. The trade-off between investments for safety on one hand and potential damage on the other hand may be further quantified by methods of insurance mathematics (Embrechts *et al.*, 2001).

Estimations of the maximum earthquake magnitude based on the classical Gutenberg–Richter model, as carried out in Holschneider *et al.* (2011) and Zöller *et al.* (2013), require complete earthquake catalogs. The estimate of the absolute maximum magnitude M depends predominantly on the largest observed events in the catalog, whereas the maximum expected magnitude in a predefined future time horizon is driven by the parameters of the Gutenberg–Richter law, namely the Richter a- and b-values. In these studies, however, a global level of completeness has been used; that is, for a magnitude of completeness, which changes with time, the highest value has been chosen for the whole time interval. As a consequence, important information that is coded in small events and helps to constrain the b-value is not taken into account. In the present study, we overcome this drawback by explicitly allowing for changes in the magnitude of completeness without dropping earthquakes. As in Zöller *et al.* (2013), we assume that the temporal occurrence of earthquakes follows a Poisson process, with an annual rate of earthquakes λ with $m \geq m_0$ (or the Richter a-value $a = \log_{10} \lambda + bm_0$). The presence of smaller earthquakes now results in a conflict with the b-value estimation: using small magnitude events for accurately estimating b leads to a violation of the Poisson assumption, because small earthquakes are usually subject to strong temporal clustering (e.g., aftershock clusters) and thus to non-Poissonian event statistics.

In many applications, one faces the situation in which the magnitude of completeness varies with time (Woessner and Wiemer, 2005). This is typical for long earthquake histories, including historical and instrumental records. In Japan, knowledge about large earthquakes is available from 684 to the present; however, in some parts of Japan, completeness since 684 at best can only be assumed. The instrumental catalog of the Japan Meteorological Agency (JMA catalog) also shows various changes of the completeness level, beginning in 1923; however, due to the overall high-seismicity level in Japan, it is generally possible to find periods with complete reporting and a reasonable number of events. Holschneider *et al.* (2011) demonstrate that the absolute maximum magnitude M cannot be constrained in terms of confidence intervals, as long as only earthquake catalogs are available. The estimation of M leads to unbound confidence intervals in most cases, that is, the upper bound of the confidence interval is $M = \infty$. In contrast, Zöller *et al.* (2013) show that confidence intervals are well constrained if the maximum expected magnitude in a finite-future time horizon is considered. In the present study, we follow the line of these publications and extend the methodology derived in Zöller *et al.* (2013) to earthquake catalogs that include periods with different levels of completeness. We address the problem that accurate b-value estimation is in conflict with the validity of the Poisson assumption. Next, we perform a case study of the devastating M 9 Tohoku earthquake on 11 March 2011 and estimate the posterior density for the maximum magnitude within $T_{\mathrm{f}} = 30$ years. We discuss the question to which degree the size of this event was expected and elaborate on the consequences for long-time horizons and high levels of confidence. Finally, we compare probabilities for large earthquakes in different subregions of Japan.

Bayesian Estimation of the Maximum Magnitude

Bayesian analysis is a mathematical tool that allows us to infer knowledge (e.g., estimates, confidence intervals) about parameters from available data and a given model. The starting point is a prior probability density function of the parameters to be estimated. This distribution can assimilate prior knowledge of the parameters independently of the given data and the model; in the case that no such knowledge is available, the prior can be chosen as uninformative, or flat. By multiplying the prior density with the likelihood function of the data (i.e., the statistical model), the prior density is updated and the posterior probability density function is obtained and can eventually be used to obtain confidence intervals and other quantities. The fact that the information about the parameters is represented by a full probability function rather than by a point estimate allows for a straightforward uncertainty assessment. In this section, we present the Bayesian posterior distribution for the maximum expected magnitude in a time window of length T_{f}; the derivation of this result is provided in Appendix A.

Here, we consider an earthquake catalog with N events and total duration T. The catalog can be decomposed into k subcatalogs, each defined by the index $i \in \{1, \ldots, k\}$ with individual magnitude of completeness $m_0^{(i)}$, number of earthquakes n_i, duration T_i, and magnitudes $\{m_j^{(i)}\}(j = 1, \ldots, n_i)$. We have

$$N = \sum_{i=1}^{k} n_i \quad \text{and} \quad T = \sum_{i=1}^{k} T_i. \tag{1}$$

In this study, earthquake magnitudes are assumed to be known exactly (e.g., without errors); the Bayesian handling of magnitude errors is left for future studies. First, the statistical model includes the doubly truncated Gutenberg–Richter distribution for earthquake magnitudes:

$$F_\beta(m) = \frac{e^{-\beta m_0} - e^{-\beta m}}{e^{-\beta m_0} - e^{-\beta M}}, \qquad m_0 \leq m \leq M, \tag{2}$$

with probability density function $f_\beta(m) = dF_\beta/dm(m)$, minimum magnitude m_0, absolute maximum magnitude M, and rescaled Richter b-value $\beta = b\ln(10)$. Second, we assume a stationary Poisson process with rate λ, which is related to the Richter a-value by $a = \log_{10} \lambda + bm_0$. The expected number of events in a future time interval T_{f} will be denoted hereinafter as $\Lambda = \lambda T_{\mathrm{f}}$.

In Appendix A, we show that the Bayesian posterior distribution for the maximum magnitude in the future time horizon T_{f} with unknown Gutenberg–Richter values a and b (or Λ and β) depends on two probability density functions $p_1(\beta, \text{catalog}|m)$ and $p_2(\beta|\text{catalog})$, according to

$$\text{posterior}(m|\text{catalog}) \propto \int_0^\infty d\beta\, p_1(\beta, \text{catalog}|m)\, p_2(\beta|\text{catalog})\, p_0(m), \tag{3}$$

in which $p_0(m)$ is the prior distribution of m, which is chosen to be flat. Because in $p_1(\beta, \text{catalog}|m)$ the catalog enters only through the total event number N, the observation times T_i, and the magnitudes of completeness $m_0^{(i)}$ for the subcatalogs $i = 1, \ldots, k$, we can write $p_1(\beta, \text{catalog}|m) = p_1(\beta, N, T_i, m_0^{(i)}|m)$. The density of this extreme value distribution

$$p_1(\beta, N, T_i, m_0^{(i)}|m) = \frac{T_\mathrm{f}(N+1) f_\beta(m) \{\sum_{i=1}^k T_i [1 - F_\beta(m_0^{(i)})]\}^{N+1}}{\{T_\mathrm{f}[1 - F_\beta(m)] + \sum_{i=1}^k T_i [1 - F_\beta(m_0^{(i)})]\}^{N+2}} \quad (4)$$

also is derived in Appendix A (equation A7).

The probability density $p_2(\beta|\text{catalog})$ is the Bayesian posterior density of β with a flat prior distribution for β; here the catalog enters through N, the event numbers n_i, the magnitudes of completeness $m_0^{(i)}$, and sample means $\langle m \rangle_i$ of magnitudes in the subcatalogs $i = 1, \ldots, k$ (see equations A9 and A10):

$$p_2(\beta|\text{catalog}) = p_2(\beta|N, n_i, m_0^{(i)}, \langle m \rangle_i) \propto \beta^N \exp\left[-\beta \sum_{i=1}^k n_i \langle m \rangle_i\right] \prod_{i=1}^k \frac{1}{[e^{-\beta m_0^{(i)}} - e^{-\beta M}]^{n_i}}. \quad (5)$$

Using equation (3), the probability that the maximum magnitude μ in a future time interval of length T_f is $\leq m$ is given by

$$\Pr(\mu \leq m|\text{catalog}) = \int_{-\infty}^{m} \text{posterior}(m'|\text{catalog}) dm'. \quad (6)$$

We note that $\Pr(-\infty < \mu \leq m_0|\text{catalog})$ represents the case in which no earthquake occurs in the time interval.

As described explicitly in Appendix A, we solely use flat Bayesian priors $\propto 1$ for m, Λ, and β. These prior distributions are improper, that is, formally they cannot be considered as (normalized) probability density functions. The choice of prior distributions is always a weak point in Bayesian analysis, because it is to some degree subjective. However, flat priors represent a standard in the case in which no *a priori* information on the variable to be estimated is available. One could also use boxlike priors, for example, by constraining the b-value to $0 \leq b \leq 3$; this will have little influence on the results, because the likelihood function for $b \geq 3$ is essentially zero.

The functions in equations (3)–(5) provide a framework for a straightforward estimation of the maximum expected magnitude in a future time window based on earthquake catalogs with a varying level of completeness. However, the b-value is mainly determined by frequently occurring small earthquakes. Using instrumental seismicity will thus improve the b-value estimation. On the other hand, small events are affected by temporal clustering, especially by aftershock activity leading to a violation of the Poisson assumption. In general, the catalog selection is characterized by a trade-off between the estimation of the b-value and the estimation of the Poisson parameter, which is dominated by large earthquakes and biased by small events because of the violation of the Poisson assumption. In the following, we consider the situation in which a historic catalog as well as an instrumental catalog is available. For convenience, we define the historic catalog as the complete catalog of earthquakes from the first event until the end of the observation period. In particular, it overlaps with the instrumental catalog. Because the Gutenberg–Richter parameters a and b are estimated in different ranges of magnitudes, we assume in this study that the functions $p_1(\beta, \text{catalog}|m)$ and $p_2(\beta|\text{catalog})$ in equation (3) can be approximated reasonably well from the historic and instrumental catalogs such that $p_1(\beta, \text{catalog}|m) \approx p_1(\beta, \text{historic catalog}|m) = p_1(\beta, N^{\text{hist}}, T, m_0^{\text{hist}}|m)$; the function $p_2(\beta|\text{catalog})$ accounts for the uncertainties in the estimation of β and is approximated from the instrumental catalog as $p_2(\beta|\text{catalog}) \approx p_2(\beta|\text{instrumental catalog}) = p_2(\beta|N^{\text{inst}}, m_0^{\text{inst}}, \langle m \rangle^{\text{inst}})$. In particular, although we assume a Poisson process from the beginning, we can take advantage of the small clustered events in the instrumental catalog without directly violating the Poisson assumption. In Figure 1 we illustrate the log-likelihood function,

$$\log[L(b)] = n[\log(b \log(10)) - b \log(10)(\langle m \rangle - m_0)], \quad (7)$$

for the JMA catalog with two lower-magnitude cutoffs: $m_0 = 5$ and $m_0 = 7$. The comparison of both curves illustrates the different degrees of uncertainty in the b-value estimation.

An illustration of the method using synthetically generated earthquake catalogs is given in Appendix B.

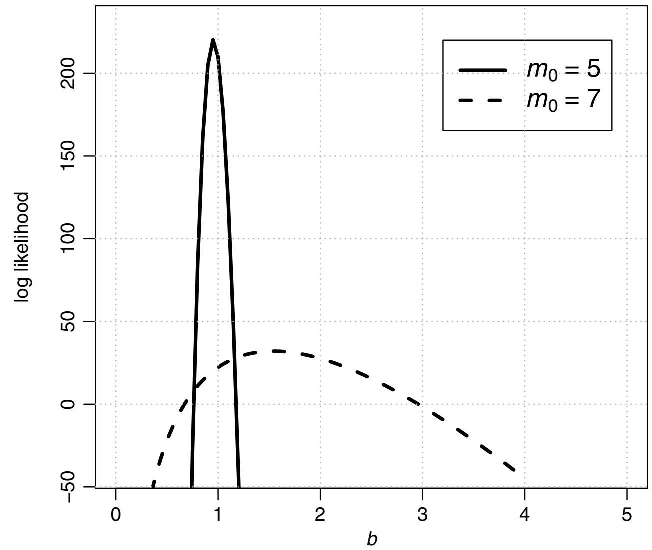

Figure 1. Log likelihood of b-value estimation for the JMA catalog. Solid, $m_0 = 5$; dashed, $m_0 = 7$. The units of the log likelihood are arbitrary because the curves have been shifted to the same domain for a better comparison.

Application and Results

In the previous section, we have demonstrated how different levels of completeness in earthquake catalogs can be accounted for in an effective way to estimate the maximum possible magnitude in a future time interval. One of the most interesting study areas for such estimations is probably Japan, where a historic catalog beginning in A.D. 684 and a high-quality instrumental catalog (the JMA catalog) are available. Both catalogs include numerous large events; the most recent one is the M 9 Tohoku earthquake, which occurred on 11 March 2011 and triggered a devastating tsunami. From the viewpoint of magnitude estimation, the data situation is excellent, because several earthquakes, which are probably close to the absolute maximum magnitude, have been observed. Although the absolute maximum magnitude is unknown, it is likely that it is not much higher than 10, and the difference to the maximum observed earthquake ($m = 9$) is therefore about one unit. Hence, the range of magnitudes, which is not supported by data, is relatively small.

In this section, we will discuss three issues. First, we study which maximum magnitude was expected within 30 years, given the data 30 years before the Tohoku earthquake; this is an update of a result presented in Zöller *et al.* (2013) using the improved magnitude estimation technique. Second, we calculate the maximum expected earthquake magnitude using the most current data. In contrast to the absolute maximum magnitude, which is predominantly affected by the largest observed event, the magnitude of the maximum expected earthquake in a time window is driven by the Gutenberg–Richter parameters a (or Λ) and b. Third, we estimate the maximum expected magnitude in eight subregions of Japan in order to account for spatial variability.

For the following studies, we use the JMA catalog to calculate the Bayesian posterior distribution $p_2(\beta|\text{catalog})$ in equation (3). We note that the magnitude of completeness m_0 is subject to changes in space and time. In the first part of this study, we are interested in time periods with complete reporting for the whole of Japan. As in Nanjo *et al.* (2010), we do not take into account earthquakes before 1970 and define two time periods: we consider seismicity between 1970 and 1981 (30 years before the Tohoku event) for earthquakes with $m \geq 5.0$ and from 2003 until mid-2012 for earthquakes with $m \geq 4.7$. Both periods cover about 10 years with approximately homogeneous reporting. The second period was selected to begin after the new Hi-net stations were installed. During the installation period, the monitoring quality improved continuously. For the lower-magnitude cutoffs, conservative values have been used, because the number of earthquakes remains high: more than 1000 events in period 1 and more than 4000 events in period 2. The parameters entering into the function $p_2(\beta|\text{catalog})$ in equation (3), N and the observational period T, are drawn from the historic catalog beginning in the year 684, which is provided by the National Oceanic and Atmospheric Administration (NOAA) catalog online. Figure 2 shows the distribution of earthquakes with $M \geq 6$ in Japan and the subdivision into eight zones that will be used later in this section. It must be noted that completeness of this catalog strongly depends on space and time. For example, the regions of Hokkaido and Tohoku (regions 1 and 2 in Fig. 2) were especially underpopulated in the early part of the observational period, and region 8 (Ryukyu) is mostly offshore. In light of the small number of events, the application of statistical methods for mapping periods of homogeneous reporting is not feasible. Therefore, for each subregion, we determine a starting point T_0 by the condition of homogeneous earthquake occurrence (constant earthquake rates in time) based on visual inspection. The lower-magnitude cutoff for the NOAA catalog is set to $m_0 = 7$, because the number of earthquakes with $m < 7$ is apparently smaller than a scaling relation would suggest. The calculations for all of Japan are carried out for seismicity after 1901, in which all subcatalogs are assumed to be approximately complete. This involves 92 earthquakes with $m \geq 7$ from 1901 to the present. The use of historic earthquakes introduces additional uncertainties that are difficult or impossible to quantify. Earthquakes may be missed, and magnitude values may be questionable (Hough, 2013). Epicenter locations are also to some extent imprecise, but the locations enter only in the later part of this study, in the assignment of each earthquakes to one of the eight subregions of Japan. For simplicity, we handle historic seismicity in the same way as instrumental seismicity; in particular, it is assumed that all catalog data are exact.

We present the Bayesian posterior distributions based on equation (3) and using data from 1970 to 11 March 1981 (i.e., up to 30 years before the Tohoku earthquake; Fig. 3a) and data until mid-2012 (Fig. 3b). For both cases, we used an uninformative flat prior for the magnitude and the unlimited Gutenberg–Richter law ($M \to \infty$). The probability that no earthquake occurs during $T_f = 30$ years is nonzero and could be illustrated by a δ-like peak left of m_0. Confidence intervals for $1 - \alpha = 0.50$, 0.90, 0.95, and 0.99 are highlighted by color. It is obvious that high levels of confidence (e.g., $1 - \alpha > 0.99$) may result in unrealistically high magnitudes ($m \gg 10$) because the distribution of magnitudes is unlimited. This problem leads again to the crucial question of the absolute maximum magnitude M. For comparison, we have repeated the calculations with the conservative value $M = 10$ (Bird and Kagan, 2004; Kagan and Jackson, 2013); results are given in Figure 4, and values for the upper bound of the $(1 - \alpha)$-confidence interval are provided in Table 1. The results show first that the introduction of the absolute maximum magnitude $M = 10$ has little influence on the results here, because the probability content in this range of magnitudes is small; this will, however, be different for longer time horizons. Second, we find almost identical magnitudes for the two observational periods, indicating overall stability of the results with respect to time. Changes would be expected only if the parameters of the Gutenberg–Richter law were changing or problems with incomplete reporting, other man-made seismicity changes, or low earthquake numbers were present.

Zöller *et al.* (2013) argue that for short-time horizons ($T_f < 1000$ years), the truncation point usually has little influence on the estimated maximum magnitudes, as long as

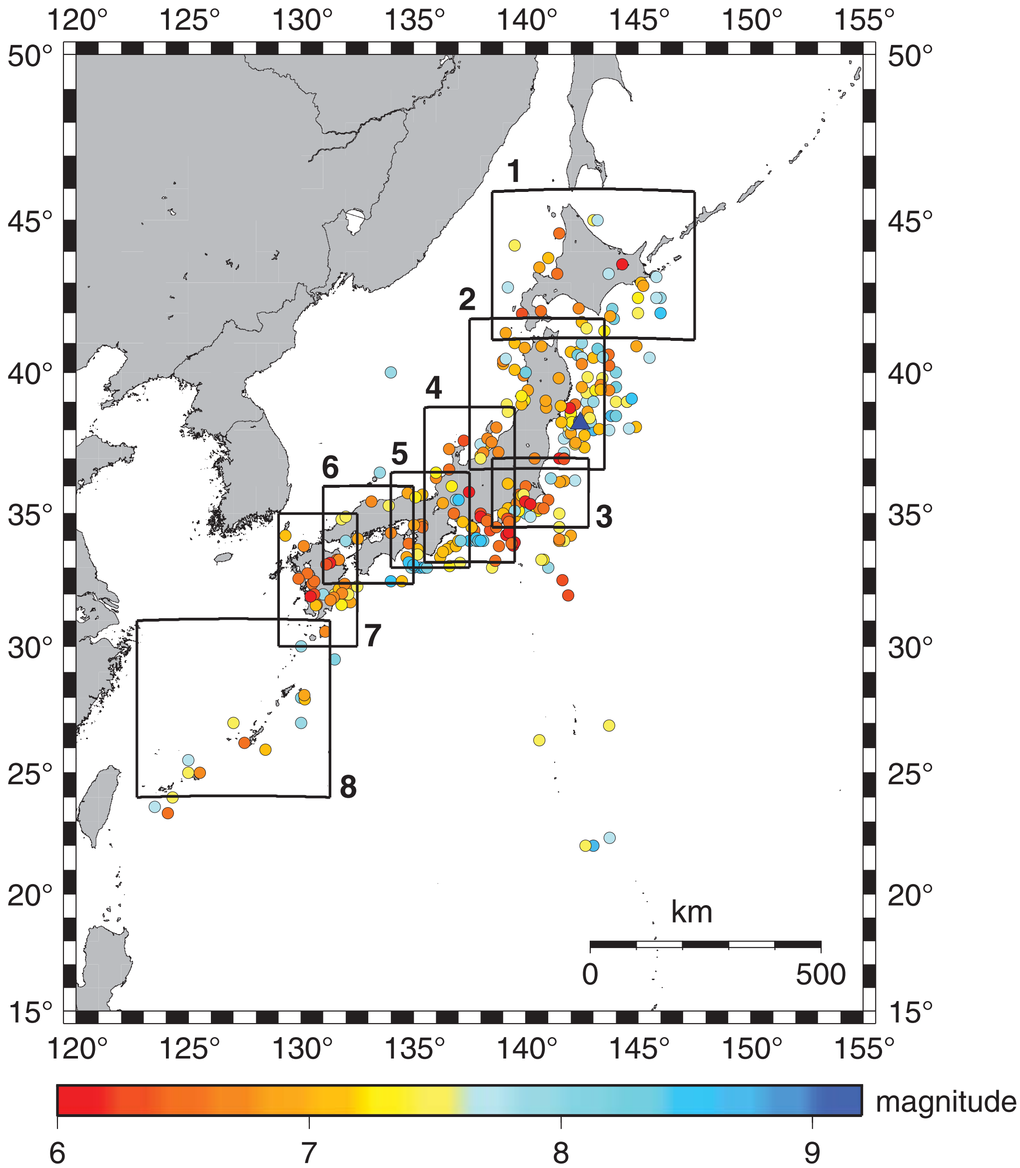

Figure 2. Earthquakes with $m \geq 6$ in Japan between A.D. 648 and 2012 (circles). The blue triangle denotes the magnitude 9 Tohoku earthquake in 2011. The subdivision into eight rectangles is used for the estimation of the maximum expected magnitude: 1, Hokkaido; 2, Tohoku; 3, Kanto; 4, central Japan; 5, Kinki; 6, Chugoku–Shikoku; 7, Kyushu; 8, Ryukyu.

earthquakes close to the truncation magnitude have low probability. In Japan, the situation is different, because earthquakes with $m = 9, \ldots, 10$ are not as unlikely as in other regions. Following Holschneider *et al.* (2011) and Zöller *et al.* (2013), the absolute maximum magnitude cannot be estimated solely from catalogs. Kagan and Jackson (2013) assume the Pareto distribution and use the moment conservation principle in combination with tectonic data as an additional constraint for this parameter; they find corner magnitudes around $m_c = 9.6$ for subduction zones, which, however, involve some uncertainties. For Japan, we therefore suggest using the conservative estimate of $M = 10$. For studies such as ours, we might also use the tapered Pareto distribution with a corner magnitude at around 10 (Bird and Kagan, 2004), which reduces the effect of the unphysical sharp truncation. However, the general inability to estimate the absolute maximum magnitude (or corner magnitude) from catalogs alone also holds for this distribution.

The results presented in Table 1 indicate that for high-confidence levels an earthquake of a size greater than the Tohoku event has to be considered, even within a short-time interval of 30 years, in which the probability of occurrence is around 20%. We may compare this number with a result from the simple extrapolation of the Gutenberg–Richter law. Using equation (B1), the number of earthquakes $\Lambda(m \geq 9; T_f = 30 \text{ years})$ with magnitude ≥ 9 within 30 years can be calculated:

$$\Lambda(m \geq 9; T_f = 30 \text{ years}) = 30 \times \lambda[1 - F_b(9)], \qquad (8)$$

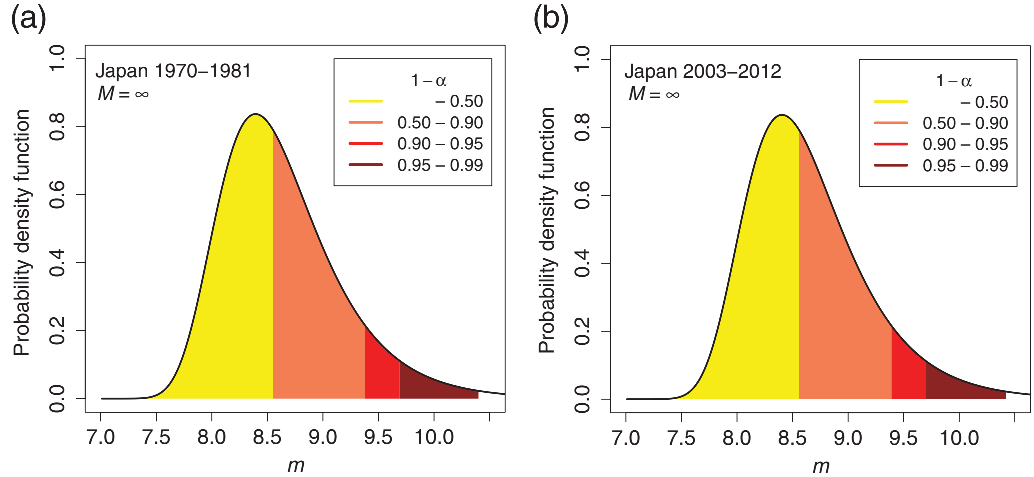

Figure 3. Probability density function of the Bayesian posterior distribution (equation 3) with $M = \infty$ for the maximum expected magnitude in Japan within $T_f = 30$ years based on (a) the earthquake catalog that starts in 1970 and ends 30 years before the Tohoku event, and (b) the earthquake catalog from 2003 to mid-2012. The plots show the magnitude range $m \geq 7$; the probability $\Pr(-\infty < m < m_0|\text{catalog})$ that no earthquake with $m \geq m_0$ occurs within $T_f = 30$ years could be illustrated by a δ-like spike left of $m = 7$.

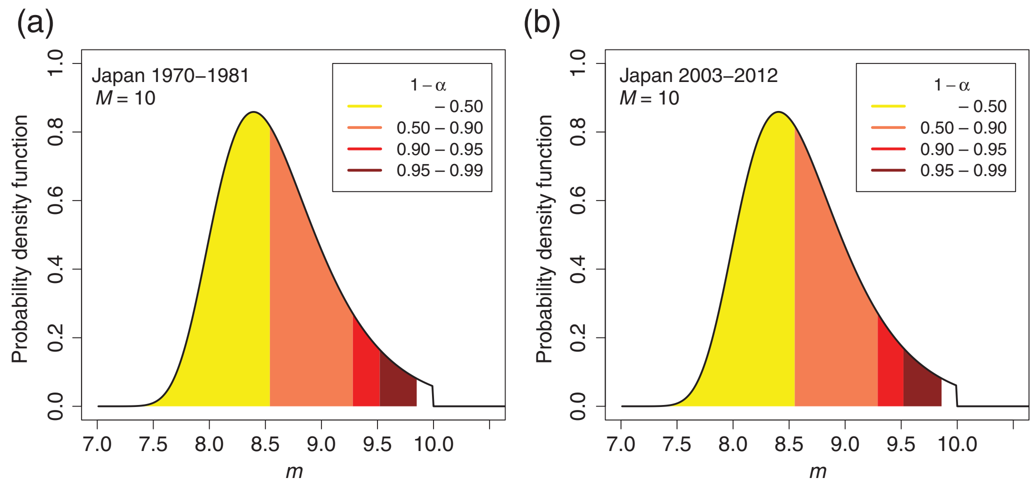

Figure 4. Same as Figure 3 with $M = 10$.

in which $\lambda = 411.9$ year^{-1} is the annual rate of earthquakes above the level of completeness ($m_0 = 4.7$), $b = 0.89$ is the Richter b-value, both estimated for period 2 (2003–2012) of the JMA catalog; $F_b(m)$ is the unlimited Gutenberg–Richter distribution. As a result, we get 1.8 earthquakes with $m \geq 9$ within 30 years—or 82 earthquakes with $m \geq 9$ within 1328 years, the length of the historic catalog that actually includes only one event with $m = 9$. For period 1 (1970–1981), we have $\lambda = 104.6$ year^{-1} for $m_0 = 5.0$ and $b = 0.90$ and get 0.8 earthquakes with $m \geq 9$ within 30 years, corresponding to 35 events in 1328 years. Although the true parameters as well as the number of future earthquakes are unknown, these rates appear to be an overestimation. A main problem related to the simple extrapolation of the Gutenberg–Richter law is related to the fact that the number of future earthquakes is assumed to be known, as soon as a and b are known. Although these parameters are replaced here by their point estimates, our model for the maximum magnitude in a time window is fully probabilistic; that is, it accounts for all possible values of a and b and all resulting scenarios for future seismicity.

We note that all findings in this section are based purely on statistics and on earthquake catalogs; earthquakes are

Table 1
Upper Bounds of $1 - \alpha$ Confidence Intervals for Results in Figures 3 and 4

$1 - \alpha$	1970 until 1981 ($M = \infty / M = 10$)	2003 until Mid-2012 ($M = \infty / M = 10$)
0.50	8.6/8.5	8.6/8.6
0.90	9.4/9.3	9.4/9.3
0.95	9.7/9.5	9.7/9.5
0.99	10.4/9.9	10.4/9.9

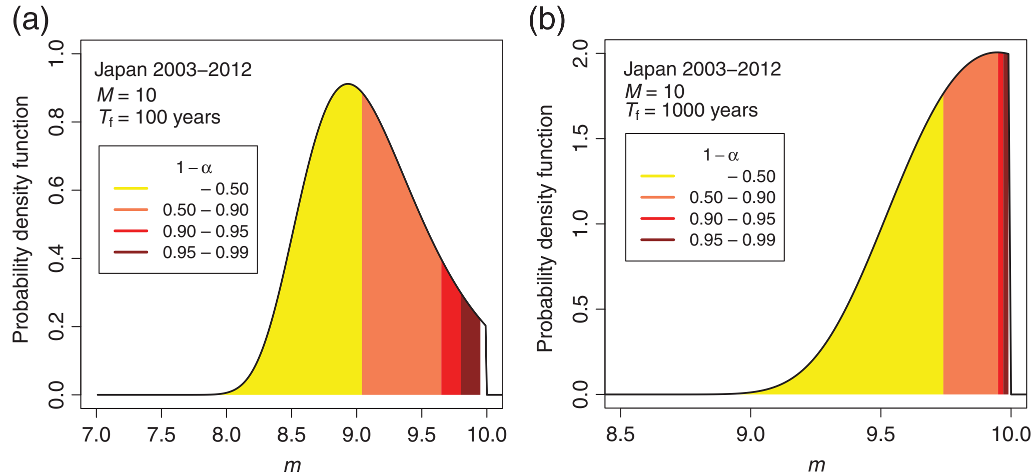

Figure 5. Same as Figure 4b, with (a) $T_f = 100$ years and (b) $T_f = 1000$ years.

assumed to occur independently of each other according to a Poisson process without memory. No physical processes, such as time-dependent evolution of stress, are taken into account here. However, earthquake catalogs probably represent the most precise database for seismic hazard, and the use of basic assumptions on seismicity will keep the uncertainties manageable. Note also that the estimated maximum magnitudes for the two time periods provided in Table 1 are very close.

We emphasize that the relatively short-time horizon, $T_f = 30$ years, already leads to a high probability for large earthquakes. In Figure 5, we show posterior probability densities for $T_f = 100$ and $T_f = 1000$ years, both calculated with $M = 10$. The functions are located around magnitudes between 9 and 9.5, with almost vanishing probability that no earthquake with $m \geq 7$ occurs. For requirements related to specific critical infrastructures, in which high-confidence levels and long-time horizons might be relevant, the maximum expected magnitude will be essentially identical with the absolute maximum magnitude (Fig. 5b). We note that this absolute maximum magnitude will certainly exist due to physical constraints such as energy conservation, although it cannot be estimated from earthquake catalogs alone.

Finally, we focus on different regions in Japan. The subdivision shown in Figure 2 is similar to those used in the monthly reports for the evaluation of seismic activities in Japan by the Headquarters of Earthquake Research Promotion. For each of the eight subregions, we use the JMA catalog after 2003 with individual lower-magnitude thresholds and the NOAA catalog with individual starting points T_0. Although T_0 is set to ensure homogeneous reporting, we introduce two conditions for the choice of m_0 in the JMA catalog: first, m_0 should be at least the magnitude of completeness indicated by scaling behavior in the frequency-size distribution; second, all subcatalogs should have a similar number of earthquakes for the b-value estimation, which is chosen to be around 2500. Because of the overall large number of earthquakes in each subregion, and the resulting small degree of uncertainty with respect to the b-value estimation, the choice of m_0 is of minor influence as long as completeness is given. The absolute maximum magnitude is again set to $M = 10$. Table 2 summarizes the parameters and results for the eight subregions. Maximum expected magnitudes are then estimated for $T_f = 30$ years and $T_f = 100$ years.

The results for the subregions indicate that the Tohoku region has the highest potential for large earthquakes. The magnitude of the 2011 event ($m = 9$) can be expected in a time frame of only 55 years, when $1 - \alpha = 0.95$ is assumed. For the same confidence level and a time horizon of around 10,000 years, the maximum expected magnitude is almost similar to the absolute maximum magnitude $M = 10$. Although the magnitudes for time horizons of 30 and 100 years are overall high, we find smaller values in regions 3–6. Additionally, probabilities for earthquakes with $m \geq 8$ within 30 years are provided. Interestingly, according to the national seismic hazard map of Japan, the probabilities are relatively low in high-hazard regions, for example, as shown in Geller (2011) for the regions of Tokai, Tonankai, and Nankai. On the other hand, we find the highest probabilities for M 8 earthquakes along the Tohoku arc, where overall low hazard is indicated. Our result is similar to recent findings of Zhuang (2012), inferred from the epidemic-type aftershock model (Ogata and Zhuang, 2006). His results also indicate much higher probabilities for large earthquakes at the Tohoku and Kuril arcs than in the Tokai and Nankai region. Our future work will study how sensitively these results depend on the assumed statistical model, for example, on the specific frequency–magnitude distribution and the temporal characteristics of earthquake occurrence.

Conclusions

What can be inferred from earthquake catalogs with respect to the largest earthquake magnitudes? From a statistical point of view, for the absolute maximum magnitude in a truncated frequency-size distribution or the corner frequency

Table 2
Upper Bound $m_{1-\alpha}$ of $1-\alpha$ Confidence Interval and Probability of Occurrence of an Earthquake with $m \geq 8$ within 30 Years

Number	Zone	m_0	T_0	m_{95} (m_{50}) $T_f = 30$ years	m_{95} (m_{50}) $T_f = 100$ years	$p(m \geq 8)$ $T_f = 30$ years
1	Hokkaido	3.7	1839	9.0 (7.8)	9.5 (8.4)	0.35
2	Tohoku	4.3	1611	9.0 (7.8)	9.5 (8.4)	0.38
3	Kanto	4.0	1241	8.5 (7.1)	9.0 (7.7)	0.12
4	Central Japan	3.2	684	8.3 (7.1)	8.8 (7.6)	0.09
5	Kinki	3.0	684	8.1 (<7)	8.6 (7.5)	0.06
6	Chugoku–Shikoku	2.3	1361	8.1 (<7)	8.5 (7.3)	0.06
7	Kyushu	2.8	1662	8.5 (7.3)	9.0 (7.8)	0.14
8	Ryukyu	3.6	1901	8.8 (7.6)	9.2 (8.1)	0.15

Results are given for two values of α and two values of T_f for the eight subregions shown in Figure 2. The lower-magnitude cutoff m_0 is also provided. The results are based on $M = 10$.

in a tapered distribution, the value $M = \infty$ has to be considered (Holschneider *et al.*, 2011). If, however, finite-time horizons are studied and a statistical model for earthquake rates and magnitudes is given, maximum expected magnitudes can be estimated properly (Zöller *et al.*, 2013). In particular, using statistical methods, we can calculate reference scenarios for a seismically active region in a given future time horizon, which might range from days to thousands of years, depending on the scenario under consideration. The focus of this work is the estimation of the maximum magnitude in a future time horizon from earthquake catalogs with different degrees of completeness. This allows assimilation of information from historic seismicity, as well as from instrumental earthquake catalogs of the same region. The underlying seismicity model is based on (1) Gutenberg–Richter distributed magnitudes and (2) a Poisson process in time (Epstein and Lomnitz, 1966). The validity of the Gutenberg–Richter distribution is usually fulfilled for small earthquakes, whereas it is uncertain for large events. The approximate validity of a Poisson process for large enough earthquakes is a subject of discussion (Gardner and Knopoff, 1974; Kagan and Jackson, 1991). Although Lombardi and Marzocchi (2007) suggest a branching model exhibiting long-term clustering for worldwide earthquakes, Michael (2011) shows that worldwide earthquakes since 1900 with magnitudes $m \geq 7$ can be explained within the uncertainties of a memory less Poisson process. In our study, we argue that the Poisson approximation is always reasonable for the long term average of a stationary point process, if the correlation between events in the process is of relatively shorter ranges. For small earthquakes, however, temporal clustering arising from aftershocks and swarms clearly violates the Poisson assumption. In other words, the simultaneous estimation of the Richter b-value and the Poisson parameter Λ (corresponding to the Richter a-value) and their uncertainties produces a trade-off: a precise b estimation requires small earthquakes, whereas the Λ estimation works best in the absence of small events (Poisson assumption).

Here, we take advantage of the fact that the Gutenberg–Richter parameters a and b are estimated essentially from data in different magnitude domains and calculate the posterior probability density function with respect to b from the instrumental catalog, whereas the Poisson parameter is then estimated from the historic catalog. We note that this approximation is related to the additional assumption that aftershocks and other types of clustered seismicity have the same b-value as mainshocks (unclustered seismicity). Otherwise, the presence of aftershocks in the instrumental catalog would bias the b-value. This problem could be addressed by one of the two following methods: first, both historic and instrumental catalogs are used for the estimation of the Gutenberg–Richter parameters, leading to increasing estimation errors as discussed in the Bayesian Estimation of the Maximum Magnitude section; second, the instrumental catalog is declustered (Gardner and Knopoff, 1974; Reasenberg, 1985; Zhuang *et al.*, 2002). However, this is a delicate issue, because new parameters and assumptions are introduced; the results eventually depend on the declustering algorithm. Additional assumptions, which will be addressed in more detail in future work, include the consideration of magnitude errors and the possible effect of missing historic earthquakes (Hough, 2013). Both corrections predominantly will affect the estimation of the a-value from the historic catalog.

Applying the seismicity model described above to seismicity in Japan leads to the conclusion that an earthquake such as the 2011 Tohoku event can be expected with more than about 20% probability within a time window of 30 years length. In the same time window, a magnitude 8.6 or larger earthquake is expected with about 50% probability, which is the probability of success when tossing a coin. We note that the time horizon must not be confused with the return period of such an event; in particular, due to the assumed Poisson process, this result holds for any 30 year time window in the future.

From a purely statistical point of view, we confirm the conclusion of Stein and Okal (2011) and Kagan and Jackson (2013) that the M 9 Tohoku event in 2011 was not a surprise. Although this earthquake is the first magnitude 9 event in 1329 years, the catalogs suggest that the probability of occurrence within only 30 years is around 20% and thus is not negligible. If the Tohoku region is singled out of the JMA catalog, the time horizon where an earthquake like the 2011 event can be ruled out is only 55 years, allowing for 5% error probability. On the other hand, we can address the question of how long do we have to wait until an earthquake with $m \approx 10$ must be considered. For this aim, we fix the significance level $\alpha = 5\%$ and increase T_f until $M_{1-\alpha} \approx 10$. In the Tohoku region, this time horizon is around 10,000 years. The evaluation of space-dependent large earthquake probabilities is in agreement with other recent findings based on statistical modeling (Zhuang, 2012) and identifies Tokai as a region with relatively low potential for an event with $m \geq 8$, although Tokai is a hotspot in the national seismic-hazard map of Japan.

For very long time horizons the situation becomes even more dramatic. In recent studies, we demonstrated that the absolute maximum magnitude M cannot be estimated properly from earthquake catalogs alone (Holschneider *et al.*, 2011). On the contrary, M enters in the frequency–size distribution and must be set to some value (Pisarenko *et al.*, 1996; Kijko, 2004) in order to allow the calculation of the seismic hazard. Although the influence of M for short-time horizons is usually limited, it cannot be neglected for long-time horizons. Although M remains an unknown parameter, it can be assumed for physical reasons that M is not much higher than 10 (Bird and Kagan, 2004; Kagan and Jackson, 2013). Scenarios for specific critical infrastructure can include high levels of confidence and long future time horizons. Whatever the precise value of M is, we have to conclude that for such scenarios, the maximum expected magnitude will be close to the absolute maximum magnitude, that is around 10.

Data and Resources

The instrumental earthquake catalog of Japan has been provided by the Japan Meteorological Agency (JMA). The National Oceanic and Atmospheric Administration earthquake catalog (684–1925) of Japan is available at http://www.ngdc.noaa.gov (last accessed January 2014). Figure 2 was made using Generic Mapping Tools, version 4.2.1 (www.soest.hawaii.edu/gmt, last accessed January 2014; Wessel and Smith, 1991).

Acknowledgments

This work was supported by the Potsdam Research Cluster for Georisk Analysis, Environmental Change and Sustainability (PROGRESS). We thank the Japan Meteorological Agency (JMA) for the JMA catalog. G. Z. is grateful to J. Woessner for a helpful discussion. The manuscript benefitted from valuable comments of Warner Marzocchi and an anonymous reviewer.

References

Bird, P., and Y. Y. Kagan (2004). Plate-tectonic analysis of shallow seismicity: Apparent boundary width, beta, corner magnitude, coupled lithosphere thickness, and coupling in seven tectonic settings, *Bull. Seismol. Soc. Am.* **94,** 2380–2399.

Embrechts, P., C. Klüppelberg, and T. Mikosch (2001). *Modelling Extreme Events: For Insurance and Finance (Stochastic Modelling and Applied Probability)*, Springer, Berlin, Germany, 21–58.

Epstein, B., and C. Lomnitz (1966). A model for the occurrence of large earthquakes, *Nature* **211,** 954–956.

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

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

Holschneider, M., G. Zöller, and S. Hainzl (2011). Estimation of the maximum possible magnitude in the framework of the doubly-truncated Gutenberg–Richter model, *Bull. Seismol. Soc. Am.* **101,** 1649–1659.

Hough, S. E. (2013). Missing great earthquakes, *J. Geophys. Res.* **118,** 1098–1108, doi: 10.1002/jgrb.50083.

Kagan, Y. Y., and D. D. Jackson (1991). Long-term earthquake clustering, *Geophys. J. Int.* **104,** 117–133.

Kagan, Y. Y., and D. D. Jackson (2013). Tohoku earthquake: A surprise? *Bull. Seismol. Soc. Am.* **103,** 1181–1194, doi: 10.1785/0120120110

Kijko, A. (2004). Estimation of the maximum earthquake magnitude $m_{\max}$, *Pure Appl. Geophys.* **161,** 1655–1681.

Lombardi, A. M., and W. Marzocchi (2007). Evidence of clustering and nonstationarity in the time distribution of large worldwide earthquakes uncertainty, *J. Geophys. Res.* **112,** no. B02303, 1271–1282, doi: 10.1029/2006JB004568

Marzocchi, W., and L. Sandri (2003). A review and new insights on the estimation of the *b*-value and its uncertainty, *Ann. Geophys.* **46,** 1271–1282.

Michael, A. J. (2011). Random variability explains apparent global clustering of large earthquakes, *Geophys. Res. Lett.* **38,** L21301, doi: 10.1029/2011GL049443.

Nanjo, K. Z., T. Ishibe, H. Tsuruoka, D. Schorlemmer, Y. Ishigaki, and N. Hirata (2010). Analysis of the completeness magnitude and seismic network coverage of Japan, *Bull. Seismol. Soc. Am.* **100,** 3261–3268.

Ogata, Y., and J. Zhuang (2006). Space–time ETAS models and an improved extension, *Tectonophysics* **413,** 13–23.

Pisarenko, V. F., A. A. Lyubushin, V. B. Lysenko, and T. V. Golubeva (1996). Statistical estimation of seismic hazard parameters: Maximum possible magnitude and related parameters, *Bull. Seismol. Soc. Am.* **86,** 691–700.

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

Stein, S., and E. A. Okal (2011). The size of the 2011 Tohoku earthquake need not have been a surprise, *Eos Trans. AGU* **92,** no. 27, 227–228.

Wessel, P., and W. H. F. Smith (1991). Free software helps map and display data, *Eos Trans. AGU* **72,** 441.

Woessner, J., and S. Wiemer (2005). Assessing the quality of earthquake catalogs: Estimating the magnitude of completeness and its uncertainty, *Bull. Seismol. Soc. Am.* **95,** no. 4, 684–698.

Zhuang, J. (2012). Long-term earthquake forecasts based on the epidemic-type aftershock sequence (ETAS) model for short-term clustering, *Res. Geophys.* **2:e8,** 52–57.

Zhuang, J., Y. Ogata, and D. Vere-Jones (2002). Stochastic declustering of space-time earthquake occurrences, *J. Am. Stat. Assoc.* **97,** 369–380.

Zöller, G., S. Hainzl, and M. Holschneider (2010). Recurrence of large earthquakes: Bayesian inference from catalogs in the presence of magnitude uncertainties, *Pure Appl. Geophys.* **167,** no. 6/7, 845–853, doi: 10.1007/s00024-010-0078-0.

Zöller, G., M. Holschneider, and S. Hainzl (2013). The maximum earthquake magnitude in a time horizon: Theory and case studies, *Bull. Seismol. Soc. Am.* **103,** no. 2a, 860–875, doi: 10.1785/0120120013.

Appendix A

Derivation of the Bayesian Posterior Distribution

With the notation of Bayesian Estimation of the Maximum Magnitudes section, we derive the formula for the

Bayesian posterior density in equation (3), as well as the expressions of the functions p_1 and p_2 in the integrand.

The following assumptions are used for earthquake magnitudes: (1) The distribution of magnitudes follows the Gutenberg–Richter law (equation 2), with a minimum magnitude m_0 that is smaller than the smallest earthquake in the catalog; (2) magnitude values are exact (i.e., they have no error); and (3) the earthquake rate at any magnitude is constant with time. If Λ denotes the Poisson intensity for the future interval of length T_f, the corresponding intensity for the ith subcatalog is

$$\Lambda_i = \frac{\Lambda}{T_f} T_i[1 - F_\beta(m_0^{(i)})], \quad \text{(A1)}$$

with $F_\beta(m)$ from equation (2). For known β and Λ, the probability that the maximum magnitude μ in the future is $\leq m$ becomes

$$\begin{aligned}\Pr(\mu \leq m|\text{catalog}) &= \left[\prod_{i=1}^{k} \frac{\Lambda_i^{n_i}}{n_i!} e^{-\Lambda_i}\right] \exp\{-\Lambda[1 - F_\beta(m)]\} \\ &= \left[\prod_{i=1}^{k} \frac{[T_i(1 - F_\beta(m_0^{(i)}))]^{n_i}}{n_i!}\right] \left(\frac{\Lambda}{T_f}\right)^N \\ &\quad \times \exp\left\{-\Lambda\left[\frac{1}{T_f}\sum_{i=1}^{k}[T_i(1 - F_\beta(m_0^{(i)}))] \right.\right. \\ &\quad \left.\left. + (1 - F_\beta(m))\right]\right\}, \end{aligned} \quad \text{(A2)}$$

in which the catalog is represented by the values N, T_i, and $m_0^{(i)}$ for $i = i, \ldots, k$.

In the next step, we use Bayes theorem in order to update a prior distribution $p_0(m, \beta, \Lambda)$ by means of equation (A2) to a posterior distribution $p(m, \beta, \Lambda|\text{catalog})$:

$$p(m, \beta, \Lambda|\text{catalog}) \propto p(\text{catalog}|\beta, \Lambda, m) p_0(m, \beta, \Lambda). \quad \text{(A3)}$$

Assuming factorizing priors, namely $p_0(m, \Lambda, \beta) = p_0(m)p_0(\Lambda)p_0(\beta)$, and taking into account all possible values of the unknown parameters β and Λ leads to

$$p(m|\text{catalog}) \propto \int_0^\infty d\beta \int_0^\infty d\Lambda p(\text{catalog}|\beta, \Lambda, m) p_0(\Lambda) p_0(\beta) p_0(m). \quad \text{(A4)}$$

Now, we show that the integration with respect to Λ can be carried out analytically, if a flat prior $p_0(\Lambda)$ is assumed. The likelihood $p(\text{catalog}|\beta, \Lambda, m)$ is essentially the derivative of equation (A2) with respect to m. For convenience, we first integrate equation (A2) with respect to Λ:

$$\int_0^\infty \Pr(\mu \leq m) p_0(\Lambda) d\Lambda = \left\{\frac{\sum_{i=1}^k T_i[1 - F_\beta(m_0^{(i)})]}{T_f[1 - F_\beta(m)] + \sum_{i=1}^k T_i[1 - F_\beta(m_0^{(i)})]}\right\}^{N+1}. \quad \text{(A5)}$$

Here, we have used the relation

$$\int_0^\infty \Lambda^N e^{-\Lambda} d\Lambda = \Gamma(N+1) = N!, \quad \text{(A6)}$$

for the $\Gamma-$ function $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$. Then, the derivative of equation (A5) with respect to m is the probability density function p_1 in equation (4):

$$\begin{aligned} p_1(\text{catalog}|\beta, m) &= p_1(N, T_i, m_0^{(i)}|\beta, m) \\ &= \frac{T_f(N+1) f_\beta(m) \{\sum_{i=1}^k T_i[1 - F_\beta(m_0^{(i)})]\}^{N+1}}{\{T_f[1 - F_\beta(m)] + \sum_{i=1}^k T_i[1 - F_\beta(m_0^{(i)})]\}^{N+2}}. \end{aligned} \quad \text{(A7)}$$

Now, equation (A4) is reduced to

$$p(m|\text{catalog}) \propto \int_0^\infty d\beta p_1(\text{catalog}|\beta, m) p_0(\beta) p_0(m). \quad \text{(A8)}$$

Finally, we focus on the function $p_0(\beta)$, which is formally a prior for β in equations (A4) and (A8). For the estimation of β, the information provided by the catalog in terms of the sample likelihood should be taken into account:

$$p_0(\beta) \propto \prod_{i=1}^{k}\left[\prod_{j=1}^{n_i} f_{m_o^{(i)}}(m_j^{(i)})\right] \quad \text{with}$$
$$f_{m_o^{(i)}}(m) = \frac{\beta e^{-\beta m}}{e^{-\beta m_0^{(i)}} - e^{-\beta M}}. \quad \text{(A9)}$$

Although $p_0(\beta)$ acts as a prior distribution in equation (A8), it is at the same time a posterior distribution of an initial estimation of β, arising from an update of a flat prior for β. Therefore, $p_0(\beta)$ in equation (A9) is identical with $p_2(\beta|\text{catalog})$, or $p_2(\beta|N, n_i, m_0^{(i)}, \langle m\rangle_i)$, in equation (5):

$$p_2(\beta|\text{catalog}) = p_0(\beta). \quad \text{(A10)}$$

Combining equations (A8) and (A10), we eventually get the Bayesian posterior distribution in equation (3).

$$\text{posterior}(m|\text{catalog}) \propto \int_0^\infty d\beta p_1(\beta, \text{catalog}|m) p_2(\beta|\text{catalog}) p_0(m). \quad \text{(A11)}$$

Appendix B

Illustration of the Method with Synthetic Data

The method described in the Bayesian Estimation of the Maximum Magnitudes section can be illustrated with synthetic data. For this aim, we generate various synthetic earthquake catalogs based on a Poisson process for earthquake times and a Gutenberg–Richter distribution of magnitudes. The synthetic data resemble instrumental and historic seismicity in the following ways:

1. A 20 year synthetic instrumental catalog with magnitude of completeness $m_0 = 5$ is generated by randomly drawing earthquake recurrence times from an exponential distribution (Poisson process) with an expected values of 100 events per year. Magnitudes are produced randomly using an unlimited Gutenberg–Richter distribution (equation 2) with $M = \infty$ and $b = 1$. The corresponding Richter a-value is $a = 7$.
2. A synthetic historic catalog with two levels of completeness is generated: 1000 years with $m_0^{(1)} = 7.0$ and 100 years with $m_0^{(2)} = 6.0$. The corresponding earthquake rates per year are
$$\Lambda_i = \Lambda \frac{1 - F_\beta(m_0^{(i)})}{1 - F_\beta(m_0)}; \qquad i = 1, 2, \tag{B1}$$
in which $F_\beta(m)$ is the Gutenberg–Richter distribution (equation 2) with $b = 1$.
3. We generate 1000 catalogs, each consisting of an instrumental and a historic subcatalog as described in items 1 and 2. First, we estimate β using the formula
$$\hat{\beta} = \frac{1}{\langle m \rangle^{\text{inst}} - m_0^{\text{inst}} + \Delta m}, \tag{B2}$$
(Marzocchi and Sandri, 2003; Zöller *et al.*, 2010). Second, for $T_{\mathrm{f}} = 50$ years and various values of $\alpha \in (0; 1)$, we calculate the mean value (averaged over 1000 catalogs) of the estimated magnitude $\langle m_{1-\alpha} \rangle$, for which the $(1 - \alpha)$-quantile $m_{1-\alpha}$ is obtained from inverting equation (4).
4. For 1000 future catalogs with $T_{\mathrm{f}} = 30$ years duration and $m_0 = 5$, the $1 - \alpha$ quantile $m_{\text{max obs}}^{1-\alpha}$ of the maximum observed magnitude is calculated and plotted against $\langle m_{1-\alpha} \rangle$ from the previous point.

The quantile plot given in Figure B1 illustrates that the estimated values of $m_{1-\alpha}$ indeed resemble the corresponding results of Monte Carlo simulations; perfect agreement would be given on the diagonal (dashed line).

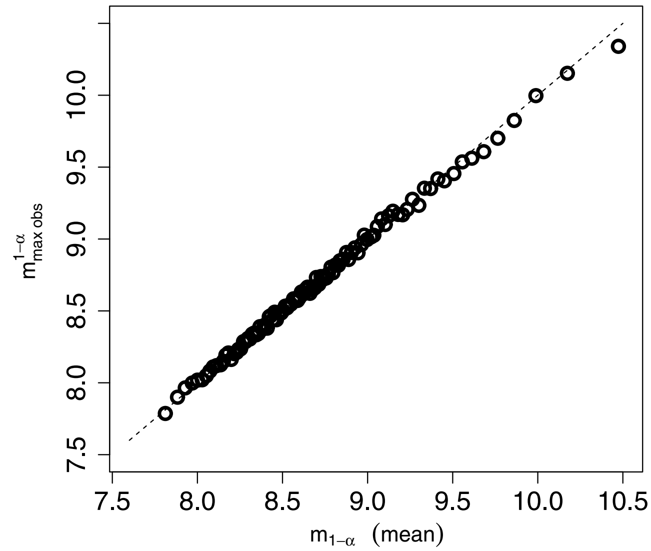

Figure B1. Quantile plot for synthetic seismicity: a comparison of estimated maximum magnitude $m_{1-\alpha}$ (mean value) for $T_{\mathrm{f}} = 50$ years with results $m_{\text{max obs}}^{1-\alpha}$ from Monte Carlo simulations. Each point represents a value of the significance level $\alpha \in (0; 1)$.

Institute of Mathematics
University of Potsdam
Am Neuen Palais 10
14469 Potsdam Germany
zoeller@uni-potsdam.de
(G.Z., M.H.)

GFZ German Research Centre for Geosciences
Telegrafenberg
14473 Potsdam Germany
(S.H.)

Institute of Statistical Mathematics
10-3 Midori-cho
Tachikawa
Tokyo 190-8562 Japan
(J.Z.)

Manuscript received 30 April 2013;
Published Online 18 March 2014