



## Reconstruction of Far-Field Tsunami Amplitude Distributions from Earthquake Sources

ERIC L. GEIST<sup>1</sup> and TOM PARSONS<sup>1</sup>

**Abstract**—The probability distribution of far-field tsunami amplitudes is explained in relation to the distribution of seismic moment at subduction zones. Tsunami amplitude distributions at tide gauge stations follow a similar functional form, well described by a tapered Pareto distribution that is parameterized by a power-law exponent and a corner amplitude. Distribution parameters are first established for eight tide gauge stations in the Pacific, using maximum likelihood estimation. A procedure is then developed to reconstruct the tsunami amplitude distribution that consists of four steps: (1) define the distribution of seismic moment at subduction zones; (2) establish a source-station scaling relation from regression analysis; (3) transform the seismic moment distribution to a tsunami amplitude distribution for each subduction zone; and (4) mix the transformed distribution for all subduction zones to an aggregate tsunami amplitude distribution specific to the tide gauge station. The tsunami amplitude distribution is adequately reconstructed for four tide gauge stations using globally constant seismic moment distribution parameters established in previous studies. In comparisons to empirical tsunami amplitude distributions from maximum likelihood estimation, the reconstructed distributions consistently exhibit higher corner amplitude values, implying that in most cases, the empirical catalogs are too short to include the largest amplitudes. Because the reconstructed distribution is based on a catalog of earthquakes that is much larger than the tsunami catalog, it is less susceptible to the effects of record-breaking events and more indicative of the actual distribution of tsunami amplitudes.

**Key words:** Tsunamis, probability distribution, seismic moment, tsunami amplitude, tide gauge.

### 1. Introduction

The probability distribution of tsunami sizes at a particular location on the coast has been investigated since the late 1960s to establish tsunami hazard curves (see summary in GEIST and PARSONS 2006). Most of the early relationships were based on tsunami

intensity  $i = \log_2 \sqrt{2R_{\text{avg}}}$  that averages tsunami amplitudes or runup over a length of coastline ( $R_{\text{avg}}$ ). More recently, BURROUGHS and TEBBENS (2005) established a power-law relation between the rate of occurrence and maximum-per-event amplitude at tide gauge stations. They indicate that this relation can be used to probabilistically forecast the number and size of future events. The advantage of a purely empirical approach in developing tsunami hazard curves is the avoidance of numerical model-based assumptions and uncertainties, since tsunami amplitudes are directly read from instrumental tide gauge records after detiding. However, undersampling and censoring are major issues facing the estimation of empirical hazard curves (GEIST and PARSONS 2014; GEIST *et al.* 2009). Furthermore, it can be difficult to relate the amplitude distribution at tide gauge stations to the broader distribution of runup along the coast adjacent to the tide gauge station.

Although there are obvious hazard assessment implications, the primary objective of this study is to better understand the physical origin of apparent power-law scaling for tsunami amplitudes. It has been known for a long time (ISHIMOTO and IIDA 1939) that earthquake seismic moment ( $m$ ) over a broad region follows a power-law distribution over many orders of magnitude. The frequency-magnitude form of the distribution is known as the Gutenberg–Richter relationship (1944), where moment magnitude  $M$  is typically referred to in the recent literature:

$$M = \frac{2}{3}(\log_{10} m - 9.05). \quad (1)$$

The origin of a power-law distribution of seismic moment has been ascribed, for example, to a self-organized critical process of stress transfer (e.g., OLAMI *et al.* 1992). For a broad region, magnitude-frequency resolution is such that a mixture of

<sup>1</sup> U.S. Geological Survey, 345 Middlefield Rd., MS 999, Menlo Park, CA 94025, USA. E-mail: egeist@usgs.gov

characteristic faults (e.g., LÓPEZ-RUIZ *et al.* 2004; WESNOUSKY 1994) or Gutenberg–Richter distributed faults with different maximum magnitude cut-offs (e.g., KAGAN 2002a, b) can satisfy the linear power-law trend. Conversely, a finite-sample of power-law distributed earthquake sizes can mimic a characteristic model of earthquake occurrence (PARSONS *et al.* 2012; PARSONS and GEIST 2009).

The power-law distribution of seismic moment is likely the primary cause of how tsunami amplitudes are distributed. However, because tsunamis observed at a far-field station are the aggregations of sources from many different fault zones (Fig. 1a), each with a different corresponding attenuation of amplitude with distance and site response (i.e., propagation characteristics near the recording station), it is not clear how they combine to form the observed empirical distribution. It is also unclear whether the tail of the distribution is related to a physical effect at the source or of the tsunami itself, or is related to undersampling

(BURROUGHS and TEBBENS 2001; GEIST and PARSONS 2014).

In this paper, we attempt to empirically reconstruct the tsunami amplitude distribution at far-field tide gauge stations, starting with the distribution of seismic moment along subduction zones. The reconstruction relies on an estimate of how tsunami amplitude scales with seismic moment for a particular subduction zone and tide gauge station pair. To avoid any number of events recorded since their installation, this  $m \sim A$  scaling relation based on regression analysis of historic catalog data. The seismic moment distribution can then be transformed to a tsunami amplitude distribution from this scaling relation and the transformed distributions from each subduction zone can be aggregated together using a mixture distribution, resulting in the reconstructed distribution for a given tide gauge station.

Below, we start by describing both the tsunami and earthquake data used in the reconstruction. We then perform maximum likelihood estimation (MLE) of parameters that describe a power-law distribution with an exponential taper, termed the tapered Pareto distribution. This distribution spans the basic form of most empirical distributions observed at tide gauge stations. The reconstruction method is then detailed, with results shown for four tide gauge stations in comparison to the empirical and MLE tapered Pareto distributions. These results are then discussed with regard to the sensitivity of the reconstruction method to a record-breaking event.

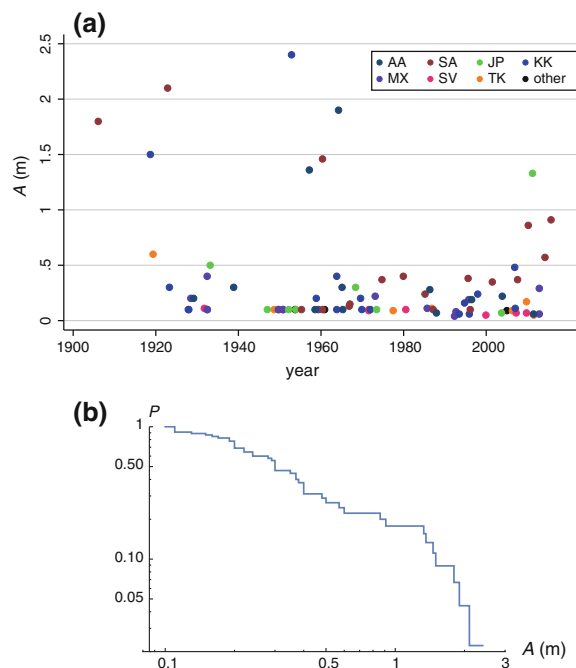


Figure 1

**a** Maximum tsunami amplitude ( $A$  in meters) per event at the Hilo tide gauge station from 1906 to 2015. Key to source zones: AA Alaska-Aleutian, SA South America, JP Japan, KK Kuril-Kamchatka, MX Mexico and Central America, SV Solomon Islands and Vanuatu, TK Tonga-Kermadec. **b** Empirical survivor function ( $P$ ) of the data  $>0.1$  m shown in (a)

## 2. Data

The data used for this study are maximum-per-event amplitudes recorded at tide gauge stations and cataloged in the National Centers for Environmental Information Global Historical Tsunami Database. Runup data are not used to avoid mixed types and locations of measurement. Only far-field events with sources greater than 1000 km from the station are selected because at these distances, tsunami amplitude scales approximately with seismic moment. In the near-field, conversely, tsunami amplitudes are affected by multiple other source parameters, such as slip variability, where the individual scaling relations

are more uncertain. In addition, near-field records of tsunamis are often partial and clipped and the stations themselves are often damaged. Even in examining far-field events we cannot completely avoid problems with the tide gauge records. For example, data gaps occur after the second peak arrival in the record of the 1960 Chile tsunami at Hilo and Nawiliwili. The selection criterion for which tide gauge stations are analyzed is based primarily on the number of events recorded since their installation, and a wide geographic distribution in the north Pacific. A MLE of distribution parameters is performed on data from eight stations: three stations in Hawaii (Hilo, Kahului, and Nawiliwili), two stations in Japan (Kushimoto and Mera), and three stations on the mainland of the U.S. (Crescent City, San Francisco, and San Diego). Reconstruction of the amplitude distribution is performed for four of these stations that have sufficient data: Hilo, Kahului, Kushimoto, and Crescent City.

The raw data from the Hilo tide gauge station are shown in Fig. 1. Most tide gauge stations have a preponderance of 0.1 m readings that are “default” values for small tsunamis although sub-decimeter readings become more common after about 1960 (GEIST and PARSONS 2011). There may be effects on the recorded amplitude related to the response of tide gauges to tsunamis that change over the time period of historical observations (SATAKE *et al.* 1988). Empirical probability distributions are fitted and plotted starting at amplitude thresholds nominally greater than 0.1 m. The regressions used in the reconstruction, however, make use of the entire station catalog shown, for example, in Fig. 1a. The probability distribution typically displayed for earthquakes and tsunamis is the survivor function ( $P$ ), i.e., the complement to the cumulative distribution function (Fig. 1b).

The earthquake magnitudes associated with events listed in the tsunami database are problematic, owing to different magnitude scales used (GEIST 2014). In this study, events listed in the tsunami database from 1900 through 1999 are cross-referenced to the Centennial catalog (ENGDahl and VILLASEÑOR 2002), corrected to a common reference magnitude scale that provides more accurate and consistent magnitude estimates for the analysis. The Centennial catalog is complete to  $M = 7.0$  since

1900 although the precision of magnitudes in the historical period (1900–1963) is roughly 0.25–0.5 magnitude units (ENGDahl and VILLASEÑOR 2002). Earthquake magnitudes for tsunami events occurring from 2000 to 2015 were cross-referenced to the Advanced National Seismic System (ANSS) earthquake catalog. In performing the reconstruction analysis, a clear outlier event with far-field tsunami amplitude much greater than the earthquake magnitude listed in the Centennial catalog is the September 7, 1918 Kuril event. The primary magnitude listed for this event is 7.6 with a deep focal depth of 242.4 km. However, GELLER and KANAMORI (1977) indicate a magnitude of 8.25, which is the magnitude we use in this study, and inverse tsunami travel time analysis indicates that the event occurred in the forearc (HATORI 1971).

### 3. Maximum Likelihood Estimate (MLE) of Distribution Parameters

The probability distribution that we use to model tsunami data is the tapered Pareto (or Kagan) distribution (KAGAN 2002a; VERE-JONES *et al.* 2001). This distribution was previously used to model several different types of natural hazards, including earthquakes, floods, tsunamis, and meteotsunamis (GEIST and PARSONS 2014; GEIST *et al.* 2014). The form of the distribution is the Pareto distribution modified by an exponential tail:

$$P(A) = \left(\frac{A_t}{A}\right)^\beta \exp\left(-\frac{A_t - A}{A_c}\right), \quad \text{for } A_t \leq A \quad (2)$$

where  $\beta$  is the power-law exponent,  $A_c$  is the corner amplitude, and  $A_t$  is the threshold amplitude for tsunami distributions. End-members of this distribution include the pure Pareto distribution ( $A_c \rightarrow \infty$ ) and the exponential distribution ( $\beta \rightarrow 0$ ). Theoretically, in the manner of VERE-JONES *et al.* (2001), the power-law component relates to a scale-invariant physical process whereas the exponential component represents a physical limitation that intervenes to taper the power-law component at the largest sizes. In the case of tsunamis, the power-law component is likely related to the power-law form of the seismic moment distribution that spans many orders of magnitude.

The origin of the exponential component, however, is less clear for tsunamis: it can either be directly related to size limitations for earthquakes or to energy-dissipating processes of the tsunami itself, particularly in shallow-water environments near tide gauge stations.

#### 4. Method

A two-parameter maximum likelihood method is described by KAGAN (2002a) to jointly estimate  $\beta$  and  $A_c$ . The likelihood function associated with the tapered Pareto distribution is given by

$$\ell = n\beta \ln A_t + \frac{1}{A_c} \left( nA_t - \sum_{i=1}^n A_i \right) - \beta \sum_{i=1}^n \ln A_i + \sum_{i=1}^n \ln \left( \frac{\beta}{A_i} + \frac{1}{A_c} \right). \quad (3)$$

where  $n$  is the number of samples. Because the likelihood Eq. (3) includes  $1/A_c$  terms, the parameters estimation is performed on  $\beta$  and  $\eta \equiv 1/A_c$ . Prior to estimating distribution parameters by maximizing the above likelihood equation, the appropriate threshold amplitude  $A_t$  needs to be determined for each station catalog. The method used in this study is described by CLAUSET *et al.* (2009), where the nonparametric Kolmogorov–Smirnov (K–S) statistic is used as a goodness of fit measure. Successive values of  $A_t$  are input to maximum likelihood estimation. The value of  $A_t$  that yields the lowest K–S statistic is the one used for the final maximum likelihood estimate (MLE) for that station.

Figure 2 shows likelihood contours for the Hilo station and the MLE (see GEIST and PARSONS 2014 for optimization method used in determining the maximum of the likelihood equation). Confidence intervals in the parameter estimates are given by the profile likelihood (PAWITAN 2001), holding one parameter at the MLE (light lines in Fig. 2). Hypothesis testing can also be performed on the profile likelihood. However, the confidence contours (95 and 99 % highlighted in Fig. 2) are a more conservative test of hypotheses when there is correlation between the two parameters. This is evident when the likelihood contour ellipses are sloped, for example in Fig. 2 where estimates for the two distribution

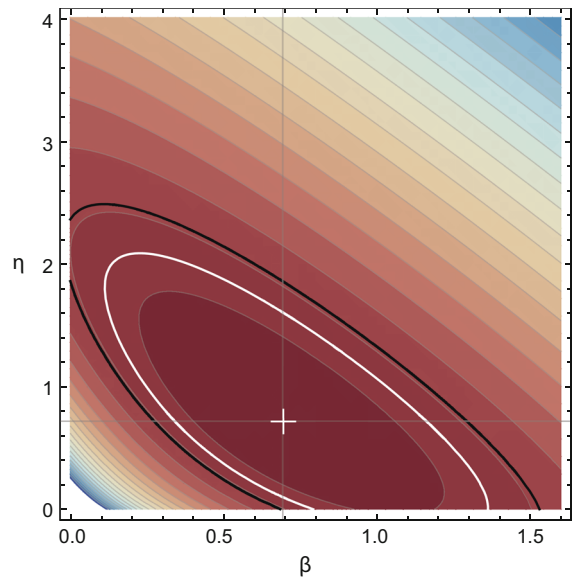


Figure 2

Likelihood map for the joint estimation of  $\beta$  and  $\eta = 1/A_c$  at the Hilo station. MLE indicated by plus. White and black contours indicate 95 and 99 % confidence regions for estimate, respectively

parameters are negatively correlated. To test the null hypothesis that  $H_0: \eta = 0$  (i.e., a pure Pareto distribution), the profile likelihood indicates that  $H_0$  can be rejected at the 95 % confidence level. However, the 95 % confidence contour intersects the  $\eta = 0$  axis away from  $\hat{\beta}$ , indicating the  $H_0$  cannot be rejected. In both cases,  $H_0: \beta = 0$  corresponding to the exponential distribution can be rejected.

#### 5. Results

The above method has been applied to earthquakes in a number of tectonic environments. BIRD and KAGAN (2004) perform a MLE of seismic moment distribution parameters along global subduction zone boundaries (as well as other global tectonic boundaries) using the global centroid moment tensor (gCMT) catalog from 1977 to 2002. They also perform the same analysis on a merged catalog that includes the EKSTRÖM and NETTLES (1997) catalog for the year 1976 and the PACHECO and SYKES catalog (1992) for 1900–1975 and apply a tectonic constraint for moment rate (see also KAGAN 2002b). The MLE's made by BIRD and KAGAN (2004) using

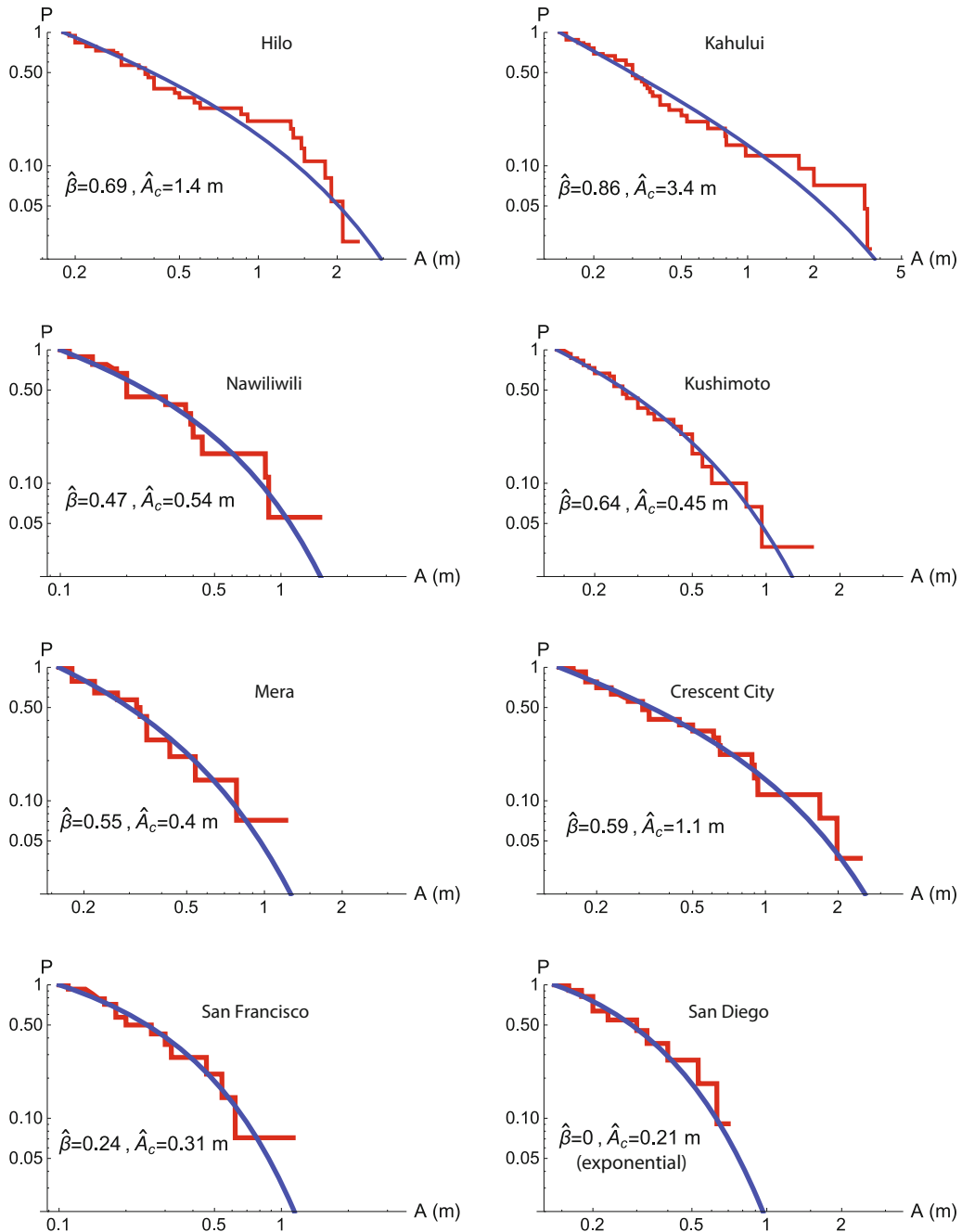


Figure 3

Results of maximum likelihood estimation (MLE) for far-field tsunamis recorded at eight stations. *Red* empirical distribution. *Blue* tapered Pareto distribution

the merged catalog and moment conservation yield  $\beta = 0.64$  and a corner moment magnitude of  $M_c = 9.58$  for all subduction zones. These and other estimates of the seismic moment distribution

parameters will be used in reconstructing the tsunami amplitude distribution described in the next section.

The MLE fit of the tapered Pareto distribution to far-field tsunami amplitudes is shown in Fig. 3 for the

Table 1

Results from maximum likelihood estimation of tapered Pareto distribution at eight stations

Station	$A_t$ (m)	$n$	$\hat{\beta}$	$\hat{A}_c$ (m)	Profile likelihood		2-Parameter likelihood	
					Reject pure Pareto?	Reject exponential?	Reject pure Pareto?	Reject exponential?
Hilo	0.18	37	0.69	1.4	Yes	Yes	No	Yes
Kahului	0.14	42	0.86	3.4	No	Yes	No	Yes
Nawiliwili	0.10	18	0.47	0.54	Yes	Yes	No	No
Kushimoto	0.14	30	0.64	0.45	Yes	Yes	No	No
Mera	0.16	14	0.55	0.40	Yes	No	No	No
Crescent City	0.14	27	0.59	1.1	Yes	Yes	No	No
San Francisco	0.10	14	0.24	0.31	Yes	No	No	No
San Diego	0.14	11	0	0.21	Yes	No	No	No

eight tide gauge stations in the north Pacific. Parameter estimates and results of hypothesis testing are listed in Table 1. Among the eight stations, the value of the corner amplitude  $A_c$  varies significantly, depending on how efficiently tsunamis are transmitted from the source to the tide gauge station. Source effects include variations in both propagation distance and the source-station azimuth relative to the radiation pattern—tsunami amplitude is typically greatest along an azimuth perpendicular to strike (BEN-MENAHEM and ROSENMAN 1972), modified by the refraction and scattering during propagation. Site effects include refractive focusing and defocusing, shelf and harbor resonance, and response related to the excitation of edge waves (e.g., RABINOVICH and THOMSON 2007). The Hawaiian stations of Hilo and Kahului and the U.S. mainland station of Crescent City have the highest values of  $A_c$ . The corner amplitude maybe even higher, given that data gaps of the 1960 tsunami at Hilo and Kahului and the exclusion of the 1964 tsunami at Crescent City (partial record) likely underestimate the largest amplitude occurring at these stations in the historical period. In contrast to the variation in  $A_c$ , the estimates of  $\beta$  among the eight stations are similar. For stations that have a narrow range between  $A_t$  and  $A_c$  (e.g., San Francisco and San Diego), the estimation for  $\beta$  becomes less constrained. In fact, for San Diego it appears that the distribution is dominated by the exponential component of the tapered Pareto function (MLE  $\beta = 0$ ).

For the Hawaii stations (Hilo and Kahului), there are significant fluctuations of the empirical distribution around the tapered Pareto distribution (red and

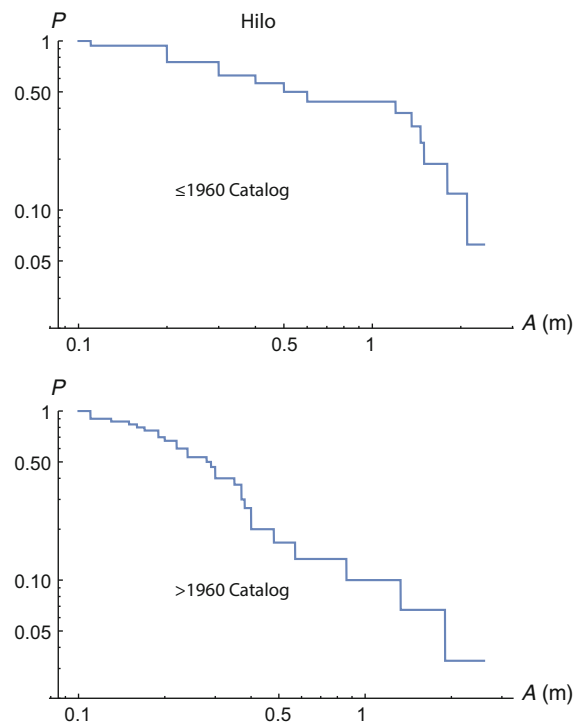


Figure 4  
Empirical amplitude distribution prior to and after 1960

blue lines, respectively, in Fig. 3). It is not immediately clear whether a particular subduction zone is causing these fluctuations, or if there is a temporal signature to the distribution. In successively removing events from individual subduction zones, we noticed no significant effect on the shape of the empirical distribution. However, in examining the distribution before and after 1960, there is a remarkable change (Fig. 4). The greater number of

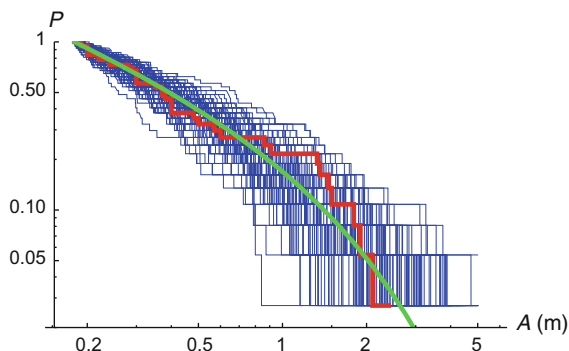


Figure 5

One hundred synthetic distributions (blue lines) sampled from MLE of the tapered Pareto distribution (green line). Empirical distribution shown by red line

M9 earthquakes  $\leq 1960$  are associated with a high number of amplitude readings over 1 m (Fig. 4a). After 1960, the distribution conforms more to a power-law relation (Fig. 4b). It is also possible that small amplitudes are censored in the early parts of the station catalogs. Overall, however, the fluctuations appear to be tied to the apparent clustering of large magnitude earthquakes throughout history (PARSONS and GEIST 2012, 2014).

To determine whether this is simply a consequence of finite sampling, 100 synthetic distributions with a sample size equal to the length of the Hilo station catalog are calculated (Fig. 5). The sampling method is described by KAGAN (2002a) and GEIST and PARSONS (2014). The observed distribution falls within the trumpet-shaped envelope of the synthetic distributions (MAIN *et al.* 2011) although the fluctuations of the empirical record at one point reach the 95th percentile of the envelope defined by the synthetics. In summary, the empirical distribution is temporally unstable, particularly with regard to the corner amplitude parameter (GEIST and PARSONS 2014). Because the reconstructed distributions described below are based on a larger earthquake dataset, they are not as affected by temporal variations in large earthquake occurrence.

## 6. Reconstruction from Earthquake Distributions

To better understand the origin of how tsunami amplitudes are distributed, we reconstruct the observed distributions, starting from seismic moment

distributions defined at subduction zones. Moment distributions from each subduction zone can be transformed to tsunami amplitude distributions, using scaling relations derived from regression analysis. Mixing the tsunami distributions from each subduction zone produces an aggregate distribution that can be compared with the empirical and fitted tapered Pareto distributions. Details of the method are described below.

## 7. Method

Reconstruction of the tsunami distribution at a tide gauge station consists of four steps: (1) specification of the seismic moment distribution for each subduction zone, (2) regression analysis relating maximum-per-event tsunami amplitude at a station to moment magnitude at subduction zones (i.e., for a source-station pair), (3) transformation of the seismic moment distribution to a tsunami distribution based on the regression analysis for each source-station pair, and (4) mixing the tsunami distributions for all significant source-station pairs to obtain a single, aggregate distribution at a tide gauge station.

For Step 1 of the reconstruction, the distribution of seismic moment for each subduction zone is taken from previous studies that use the MLE method described above. Although some studies have estimated distribution parameters based on difference zonation schemes (KAGAN 1997, 1999), these studies and BIRD and KAGAN (2004) argue that all subduction zones have similar values of  $\beta = 0.63$ . BIRD and KAGAN (2004) estimate a global corner moment magnitude for subduction zone earthquakes of  $M_c = 9.58$  ( $m_c = 2.9 \times 10^{23}$  Nm), based on a merged twentieth century earthquake catalog and tectonic moment conservation. In the BIRD and KAGAN (2004) study, a restrictive definition of subduction zone earthquakes is used based on distance and focal mechanism criteria. KAGAN *et al.* (2010) indicate a lower value of  $M_c = 8.75$  using less restrictive earthquake selection criteria at subduction zones; the region where these earthquakes occur is termed “trench zones” as distinct from the focal mechanism specific subduction zones defined by BIRD and KAGAN (2004). KAGAN (2010) indicates that, in

general,  $\beta$  may actually be close to 0.5, owing to the combined effect of a number of factors: systematic and random errors in measuring seismic moment ( $m$ ), clustering of earthquakes, earthquake complexity, and centroid depth distribution. We test the effect that different values of both  $M_c$  and  $\beta$  have on the final result.

Several previous studies have examined scaling between seismic moment and tsunami amplitude at tide gauge stations (ABE 1979, 1989; COMER 1980; GEIST 2012; OKAL 1988; PELAYO and WIENS 1992) (Step 2 of the reconstruction). The observations span several orders of magnitude, confounding regression between these two variables. Under ordinary least squares (OLS) linear regression, there is a disproportionate relative error among samples, biasing the fit preferentially toward large values (SORNETTE 2009). An alternative is to scale the logarithm of tsunami amplitudes with respect to moment magnitude (Eq. 1) as performed by, for example, PELAYO and WIENS (1992) and GEIST (2012). However, log-linear regression using an OLS estimator can also have pitfalls, including an assumption that the residuals of the log-amplitude are uniform and normally distributed. In addition, differences among many small, non-zero values expected from a power-law distributed random variable will be accentuated when the logarithm is applied. Instead, we use Poisson regression with the Huber–White sandwich estimator (HUBER 1967; WHITE 1980) that accounts for heteroskedasticity in the data and that relaxes the Poisson assumption of variance being equal to the mean. Although Poisson regression is commonly applied to count data (e.g., GEIST and PARSONS 2011), it can also be applied to continuous data and solves for coefficients  $\alpha_1$  and  $\alpha_2$  in the equation below:

$$A_j = e^{\alpha_1 + \alpha_2 M_j + \varepsilon_j}, \quad (4)$$

where  $j$  is an index over sample space and  $\varepsilon_j$  is the error term. Hence, amplitude ( $A$ ), rather than log-amplitude is regressed against moment magnitude ( $M$ ). Shown in Fig. 6 is a comparison of the residuals for the South America-Hilo source-station pair. Although, the differences are minor, Poisson regression with the Huber–White estimator is a more robust method to determine the scaling relation between earthquake size and tsunami amplitude.

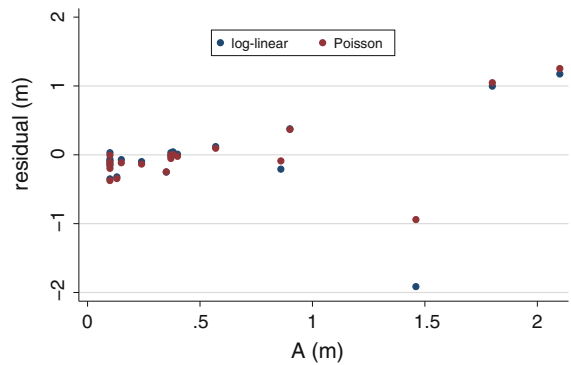


Figure 6  
Comparison of residuals for the South America-Hilo source-station pair using two different regression methods

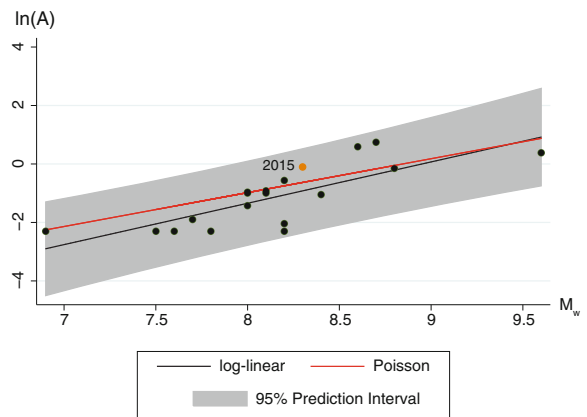


Figure 7  
Results from log-linear regression (black line). 95 % prediction interval shown by shaded region. Regression computed without the 2015 event (orange) for this figure. Poisson regression line shown in red

The exclusion criterion that we use to determine whether the regression is significant for a particular source-station pair is the  $p$  value from the Wald test under the null hypothesis that the regression coefficients are zero. If the coefficients are not significant at the 95 % confidence level, that source-station pair is not used further in the reconstruction.

For demonstration purposes, we show the results of the log-linear OLS regression analysis for the South America-Hilo subduction zone pair in Fig. 7. For comparison, the results of the Poisson regression are shown by the red line, which is not as influenced by small non-zero values as the OLS log-linear regression. As a test, the 95 % prediction interval for



an individual forecast (shaded region), as distinct from the confidence interval of the mean, is calculated without the 2015 Chile event (orange dot). Although the 2015 event generated a larger tsunami than average, it falls within the wide prediction interval. In general, the uncertainty represented by the shaded region includes uncertainty of source parameters other than seismic moment that affect tsunami generation (GEIST 1999).

Step 3 of the reconstruction involves transforming the seismic moment distribution to a tsunami amplitude distribution. In general, for a probability density function  $p(x)$  defined on the interval  $a \leq X \leq b$  and a transformation function  $y(x)$ , the transformed density function  $q(y)$  is given by (KEMPTHORNE and FOLKS 1971)

$$q(y) = p[x(y)] \left| \frac{dx}{dy} \right|, \quad y(a) \leq Y \leq y(b), \quad (5)$$

where  $x(y)$  is the inverse of  $y(x)$ . The above equation is valid only if both  $x(y)$  and  $y(x)$  are single valued and if either  $\frac{dx}{dy} \geq 0$  or  $\frac{dx}{dy} \leq 0$ . Thus,  $y(x)$  must either be monotonically increasing or decreasing. In this study, the transformation takes the functional form of

$$A(m) = \gamma_1 m^{\gamma_2}, \quad (6)$$

where the constants  $\gamma_1$  and  $\gamma_2$  are related to the regression coefficients  $\alpha_1$  and  $\alpha_2$  (Eq. 4), taking into account the definition of moment magnitude (Eq. 1). The value of  $m_t$  used in the transformation is determined from the value of  $A_t$  and the scaling equation for each source-station pair.

Transformed distributions are shown in Fig. 8 for four tide gauge stations. For this example, a globally uniform seismic moment distribution with parameters  $\beta = 0.5$  and  $M_c = 9.58$  is transformed according to Eq. 6, using regression coefficients specific to each source-station pair. The maximum seismic moment may be limited by the length of a subduction zone (MCCAFFREY 2008). In this study, however, all of the subduction zones that produce significant regressions are long enough to support an earthquake greater than the corner moment ( $m_c$ ). Because the same moment distribution is used for each subduction zone, differences in the tsunami amplitude distributions relate to how efficiently tsunamis are transmitted from each subduction zones to the station, owing to propagation

and site response effects. For example, tsunami amplitudes for a given seismic moment are significantly higher at the Kahului station than the nearby Hilo station for each subduction zone, except for the Japan and Tonga-Kermadec subduction zones.

Finally, in Step 4 of the reconstruction, the transformed tsunami distributions from all significant source-station pairs are mixed to form the aggregate tsunami distribution. In general, a mixture density distribution is given by the finite, weighted sum of density functions  $p(x)$ , most often of the same distribution or distribution family:

$$m(x) = \sum_{i=1}^n w_i p_i(x); \quad \sum_{i=1}^n w_i = 1. \quad (7)$$

In this study, weights ( $w_i$ ) in the mixture model are assigned according to the relative rate at which each subduction zone produces tsunamis at the station. The weights are calculated by dividing the number of subduction zone earthquakes generating tsunamis greater than the threshold amplitude  $A_t$  at a station by the total number events used in the reconstruction. In general, the mixed distribution will have subtle variations, depending on the variability of the distribution parameters of the source-station pairs. Often, however, the mixed distribution can be well approximated by the tapered Pareto distribution.

## 8. Results

Different combinations of seismic distribution parameters zones are tested for the reconstruction. In this study, these parameters are prescribed to be the same for all subduction zones, but the method can accommodate regional variations in seismic moment distribution parameters. As for the source-station amplitude distributions shown in Fig. 8, an  $M_c = 9.58$  estimate from BIRD and KAGAN (2004) and a  $\beta = 0.5$  from KAGAN (2010) is first tested. The resulting reconstructed amplitude distribution is shown for the Hilo station in Fig. 9a. For comparison, using the global value of  $\beta = 0.63$  from BIRD and KAGAN (2004) results in a corresponding increase in the equivalent power-law exponent of the amplitude distribution (Fig. 9b). The effect of a lower corner

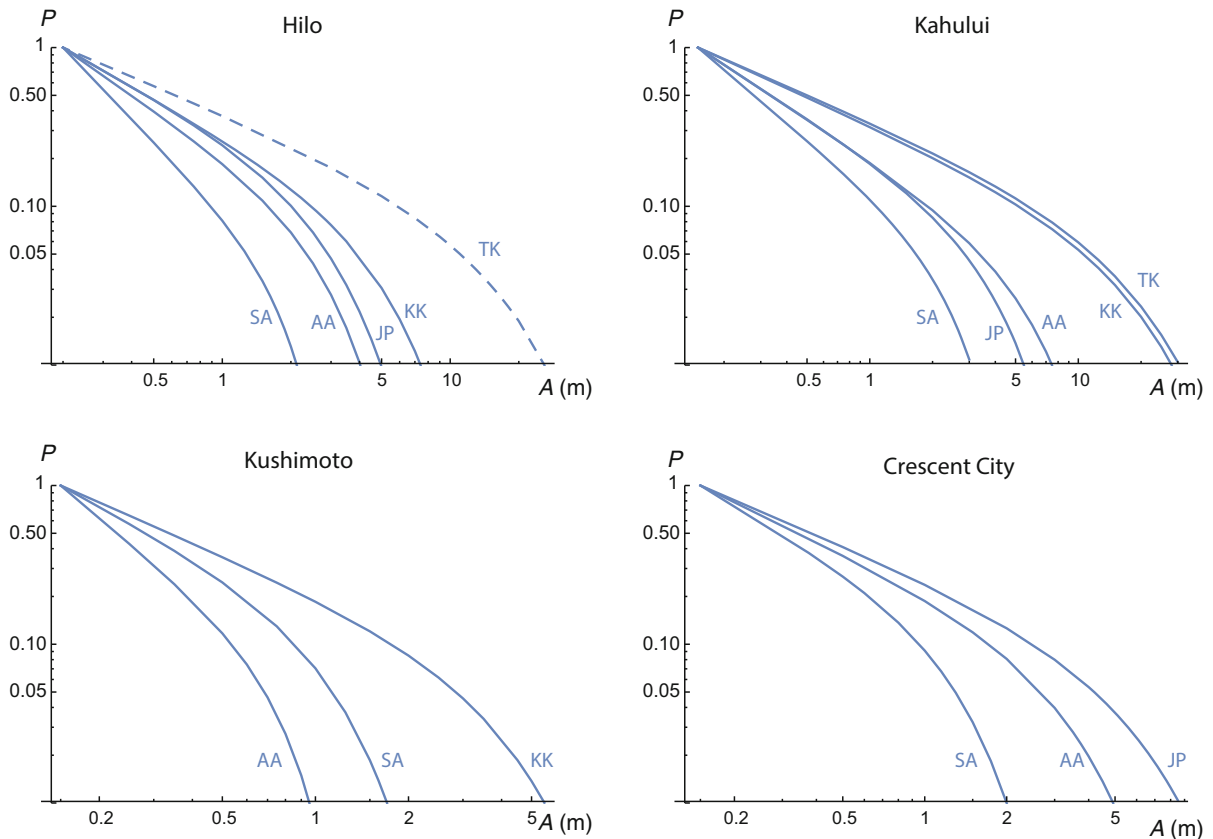


Figure 8

Transformed distributions of tsunami amplitudes for four stations and subduction zones where regression analysis is significant (*solid distributions*). Subduction zones: SA South America, AA Alaska-Aleutian; JP Japan, KK Kuril-Kamchatka, TK Tonga-Kermadec. Note that TK-Hilo regression (*dashed distribution*) is not significant at 95 % confidence, but included for demonstration purposes (see text)

moment magnitude  $M_c = 8.75$  (KAGAN *et al.* 2010), using a broader selection of earthquakes at “trench zones” (KAGAN and JACKSON 2013), is shown in Fig. 9c. There is a corresponding lowering of the equivalent corner amplitude of the tsunami distribution. Using both  $\beta = 0.5$  and  $M_c = 8.75$  (Fig. 9d) results in a tsunami distribution that can be rejected with 95 % confidence, according to the K–S  $p$  value (Fig. 9d).

Results of the reconstruction procedure are shown for four tide gauge stations are shown in Fig. 10, using the optimal combination of moment distribution parameters shown in Fig. 9a:  $M_c = 9.58$  and  $\beta = 0.5$ . In comparison to the MLE distributions (Table 1), the effective power-law exponent and corner amplitude is higher for the reconstructed distributions, except for the Kahului station where  $\beta$  is

approximately the same (Table 2). The reconstructed distributions, therefore, indicate higher probabilities for large amplitudes compared to the MLE distributions.

For the Hilo station (Fig. 10a), the source regions that result in significant regression results are the Alaska-Aleutian, South America, Japan, and Kuril-Kamchatka subduction zones (corresponding amplitude distributions shown in Fig. 8). The Kahului station is 200 km away from the Hilo station and provides a good comparison to determine the effect of site response on the reconstructed distribution. Overall, most subduction zones are more efficient in generating tsunamis at Kahului, resulting in a higher corner amplitude than for Hilo. Interestingly, the regression with respect to the Tonga-Kermadec subduction zone is significant for the Kahului station,

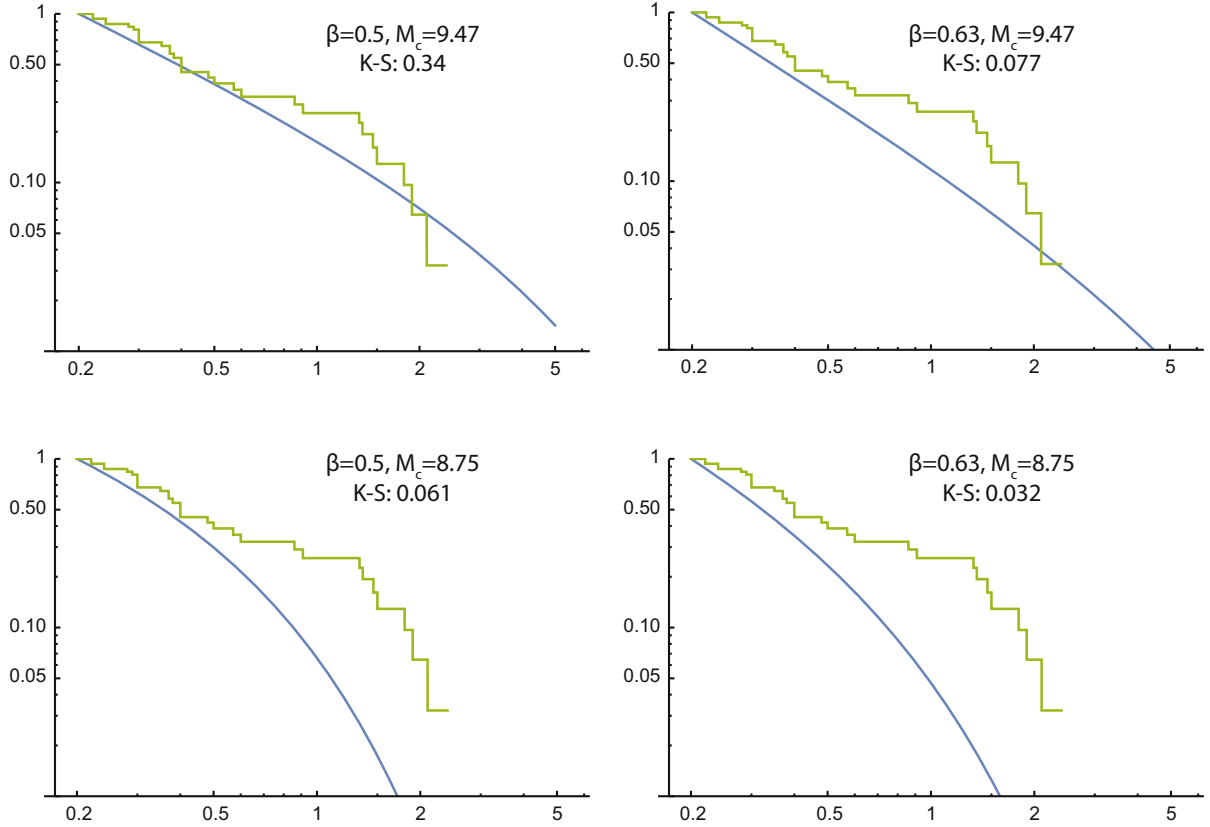


Figure 9

Effect of different combinations of source distribution parameters on reconstructed tsunami amplitude distribution at the Hilo station.  $K-S$  Kolmogorov–Smirnov  $p$  value

whereas it is not for the Hilo station (dashed line in Fig. 8). If we add the Tonga-Kermadec sources to the Hilo reconstruction (relaxing our criteria for accepting source-station regressions to, for example, 80 % confidence), there is only a marginal change to the reconstructed distribution. This is because regressions below the 95 % confidence level often have fewer events relative to the other source zones, thus lowering the weight ( $w_i$ ) for these source regions in the mixing stage (Eq. 7).

Two stations on the margin of the Pacific Basin are also analyzed: Kushimoto and Crescent City. Overall, far-field tsunamis are more efficiently transmitted to the Crescent City station than the Kushimoto station. The heavier tail of the tsunami amplitude distribution associated with Crescent City is likely caused by the regional (site) response near the station (HARRILLO *et al.* 2008), rather than source

effects. It should be reiterated, however, only far-field tsunamis are analyzed; the distribution for the Kushimoto station would look significantly different if near-field sources were included, whereas the Crescent City distribution would look similar to what is displayed in Fig. 10b because there are comparatively few local tsunami sources that have been recorded at Crescent City.

## 9. Discussion

The reconstructed tsunami amplitude distributions presented here are more robust than MLE estimates using just observations from the tsunami catalog because the reconstructions are based on a larger earthquake database. GEIST and PARSONS (2014) show that the MLE of corner amplitude is unstable and can

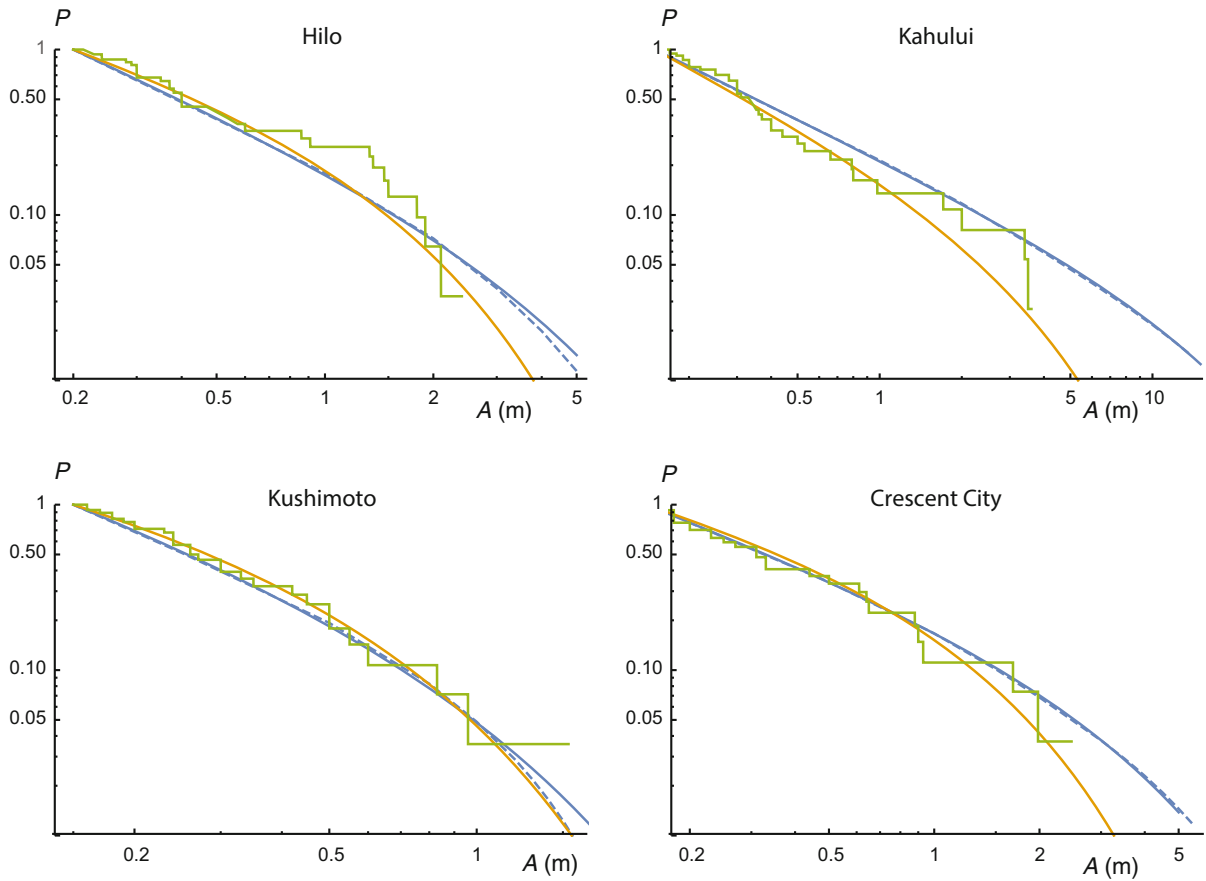


Figure 10

Reconstructed distributions (*blue dashed lines*) and tapered Pareto distribution best fit to reconstruction (*blue solid lines*) for four stations. Empirical and MLE distributions shown by *green and orange lines*, respectively. Note differences in amplitude scales

Table 2

*Results of reconstruction method for four tide gauge stations*

Station	Subduction zones with Significant regressions	Effective $\hat{\beta}$	Effective $\hat{A}_c$ (m)
Hilo	AA, SA, JP, KK	1.25	6.0
Kahului	AA, SA, JP, KK, TK	0.80	21.
Kushimoto	AA, SA, KK	1.05	0.81
Crescent City	AA, SA, JP	0.81	3.4

AA Alaska-Aleutian, SA South America, JP Japan, KK Kuril-Kamchatka, TK Tonga-Kermadec

change significantly with one record-breaking event. In contrast, there is a much smaller effect of a record-breaking event on the reconstructed distributions presented here. Although the tail of the global seismic moment distribution may also be poorly constrained based only on the gCMT catalog (ZÖLLER 2013),

introducing other constraints such as moment conservation linked to plate boundary deformation rates results in a more stable estimate of the global seismic moment distribution (BIRD and KAGAN 2004; KAGAN and JACKSON 2013). In terms of the regression, there can be a slight change in the estimated scaling caused

by a record-breaking event. For example, Fig. 11 shows the log-linear regression fit for the South America-Hilo source-station pair with and without the  $M = 9.6$  1960 Chile earthquakes. This event, along with the  $M = 9.2$  1964 Alaska earthquake, are known to have generated deficient tsunamis relative to their moment magnitude (OKAL 1988) although the maximum amplitude of the 1960 tsunami may not have been accurately recorded at the Hilo and Nawiliwili stations. When the transformed distribution is mixed with those from other subduction zones, the effect of omitting the record-breaking event is minimal. This indicates that the reconstructed distributions are more robust than the MLE distributions, with regard to adding one record-breaking event.

In contrast, strong outliers in the scaling of far-field tsunami amplitude with respect to seismic moment can affect the reconstruction procedure. These outliers are most often caused by measurement error, particularly in the determination of moment magnitude of historic earthquakes where instrumental records are deficient in long-period energy. For example, using a magnitude of 7.5 for the 1918 Kuril event that is listed in the Centennial catalog excludes many regressions involving the Kuril-Kamchatka subduction zone for the four tide gauge stations analyzed. In contrast, using GELLER and KANAMORI'S (1977) magnitude estimate of 8.25 for this event results in significant regressions associated with this subduction zone for all of the stations shown in Fig. 8, except Crescent City. Other data problems

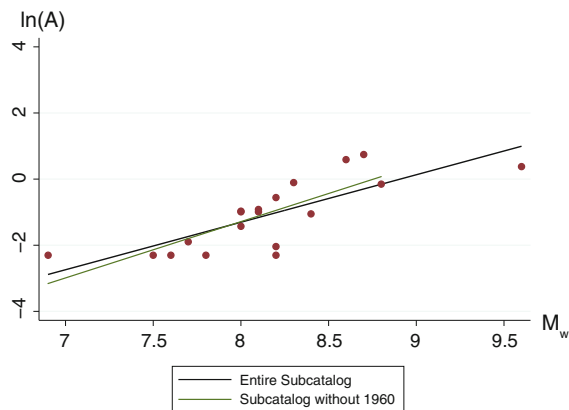


Figure 11

Regression for the Hilo station for South America sources with (black) and without (green) the 1960 event

related to tsunami recording, such as clipping, data gaps, and tide gauge response, can cause more subtle errors that may also affect the results from reconstruction.

There is substantial evidence that the parameters of moment distribution for subduction zones are globally constant (e.g., KAGAN 1997, 1999). Both the MLE and reconstruction analysis indicate that the corner amplitude for tsunami distributions varies significantly, owing to both source (i.e., propagation distance, radiation pattern) and site effects. However, is the power-law exponent  $\beta$  for tsunami amplitude distributions globally constant? Within the bounds uncertainty for both the MLE and reconstructed distributions, we cannot reject the hypothesis that  $\beta$  is constant, although the  $\beta$  predicted by reconstruction is generally higher than that predicted by the MLE. Further work is needed to verify this hypothesis, but if true, a constant power-law exponent may simplify probabilistic hazard analysis for tsunamis.

## 10. Conclusions

We have shown that the probability distribution of tsunami amplitudes is directly related to the distribution of seismic moment at subduction zones. To demonstrate this, we developed a method to reconstruct tsunami amplitude distributions using simple amplitude-moment scaling relations, in lieu of numerical modeling, and aggregating the transformed tsunami amplitude distributions from individual subduction zones to achieve the final reconstructed distribution. In testing different sets of seismic moment distribution parameters that are globally constant, a power-law exponent ( $\beta$ ) of 0.5 (KAGAN 2010) and a corner moment magnitude of ( $M_c$ ) of 9.58 (BIRD and KAGAN 2004) provide the best fit. In comparison to tsunami amplitude distributions estimates from maximum likelihood, the reconstructed distributions consistently exhibit higher effective corner amplitudes ( $A_c$ ) and are less susceptible to changes in  $A_c$  caused by a single record-breaking event. As a result, the hazard associated with exceedance amplitudes near  $A_c$  is higher for the reconstructed distribution compared to the maximum likelihood distribution.

### Acknowledgments

The authors thank Robert Kayen, two anonymous reviewers, and Editor Alexander Rabinovich for their constructive comments on this study and manuscript.

### REFERENCES

- ABE K. (1979), *Size of great earthquake of 1837–1974 inferred from tsunami data*, J. Geophys. Res., *84*, 1561–1568.
- ABE K. (1989), *Quantification of tsunamigenic earthquakes by the Mt scale*, Tectonophysics, *166*, 27–34.
- BEN-MENAHEM A., ROSENMAN M. (1972), *Amplitude patterns of tsunami waves from submarine earthquakes*, J. Geophys. Res., *77*, 3097–3128.
- BIRD P., KAGAN Y.Y. (2004), *Plate-tectonic analysis of shallow seismicity: apparent boundary width, beta-value, corner magnitude, coupled lithosphere thickness, and coupling in 7 tectonic settings*, Bull. Seismol. Soc. Am., *94*, 2380–2399.
- BURROUGHS S.M., TEBBENS S.F. (2001), *Upper-truncated power laws in natural systems*, Pure Appl. Geophys., *158*, 741–757.
- BURROUGHS S.M., TEBBENS S.F. (2005), *Power law scaling and probabilistic forecasting of tsunami runup heights*, Pure Appl. Geophys., *162*, 331–342.
- CLAUSET A., SHALIZI C.R., NEWMAN M.E.J. (2009), *Power-law distributions in empirical data*, SIAM Review, *51*, 661–703.
- COMER R.P. (1980), *Tsunami height and earthquake magnitude: theoretical basis of an empirical relation*, Geophys. Res. Lett., *7*, 445–448.
- EKSTRÖM G., NETTLES M. (1997), *Calibration of the HGLP seismograph network and centroid-moment tensor analysis of significant earthquakes of 1976*, Physics of the Earth and Planetary Interiors, *101*, 221–246.
- ENGDAHL E.R., VILLASEÑOR A. (2002), *Global seismicity: 1900–1999*. In: Lee WHK, Kanamori H, Jennings PC, Kisslinger C (eds), International Handbook of Earthquake and Engineering Seismology, Part A. Academic Press, San Diego, pp. 665–690.
- GEIST E.L. (1999), *Local tsunamis and earthquake source parameters*, Adv. Geophys., *39*, 117–209.
- GEIST E.L. (2012), *Phenomenology of tsunamis II: scaling, Event Statistics, and Inter-Event Triggering*, Adv. Geophys., *53*, 35–92.
- GEIST E.L. (2014), *Explanation of temporal clustering of tsunami sources using the epidemic-type aftershock sequence model*, Bull. Seismol. Soc. Am., *104*, 2091–2103.
- GEIST E.L., PARSONS T. (2006), *Probabilistic analysis of tsunami hazards*, Natural Hazards, *37*, 277–314.
- GEIST E.L., PARSONS T. (2011), *Assessing historical rate changes in global tsunami occurrence*, Geophys. J. Int., *187*, 497–509.
- GEIST E.L., PARSONS T. (2014), *Undersampling power-law size distributions: effect on the assessment of extreme natural hazards*, Natural Hazards, *72*, 565–595. doi:10.1007/s11069-013-1024-0.
- GEIST E.L., PARSONS T., TEN BRINK U.S., LEE H.J. (2009), *Tsunami Probability*. In: Bernard EN, Robinson AR (eds), The Sea, v. 15. Harvard University Press, Cambridge, Massachusetts, pp. 93–135.
- GEIST E.L., TEN BRINK U.S., GOVE M. (2014), *A framework for the probabilistic analysis of meteotsunamis*, Natural Hazards, *74*, 123–142. doi:10.1007/s11069-014-1294-1.
- GELLER R.J., KANAMORI H. (1977), *Magnitudes of great shallow earthquakes from 1904 to 1952*, Bull. Seismol. Soc. Am., *67*, 587–598.
- GUTENBERG B., RICHTER C.F. (1944), *Frequency of earthquakes in California*, Bull. Seismol. Soc. Am., *34*, 185–188.
- HATORI T. (1971), *Tsunami sources in Hokkaido and southern Kuril regions*, Bulletin of the Earthquake Research Institute, *49*, 63–75.
- HORRILLO J., KNIGHT W., KOWALIK Z. (2008), *Kuril Islands tsunami of November 2006: 2. Impact at Crescent City by local enhancement*, J. Geophys. Res., *113*, doi:10.1029/2007JC004404.
- HUBER P.J. (1967), *The behavior of maximum likelihood estimates under nonstandard conditions*. In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, pp. 221–233.
- ISHIMOTO M., IIDA K. (1939), *Observations of earthquakes registered with the microseismograph constructed recently*, Bulletin of the Earthquake Research Institute, *17*, 443–478.
- KAGAN Y.Y. (1997), *Seismic moment-frequency relation for shallow earthquakes: regional comparison*, J. Geophys. Res., *102*, 2835–2852.
- KAGAN Y.Y. (1999), *Universality of the seismic-moment-frequency relation*, Pure Appl. Geophys., *155*, 537–573.
- KAGAN Y.Y. (2002a), *Seismic moment distribution revisited: I. Statistical results*, Geophys. J. Int., *148*, 520–541.
- KAGAN Y.Y. (2002b), *Seismic moment distribution revisited: II. Moment conservation principle*, Geophys. J. Int., *149*, 731–754.
- KAGAN Y.Y. (2010), *Earthquake size distribution: power-law with exponent  $\beta = 1/2$ ?*, Tectonophysics, *490*, 103–114.
- KAGAN Y.Y., BIRD P., JACKSON D.D. (2010), *Earthquake patterns in diverse tectonic zones of the globe*, Pure Appl. Geophys., *167*, 721–741.
- KAGAN Y.Y., JACKSON D.D. (2013), *Tohoku earthquake: a surprise?*, Bull. Seismol. Soc. Am., *103*, 1181–1194.
- KEMPTHORNE O., FOLKS L. (1971), *Probability, statistics, and data analysis*. Iowa State University Press, Ames, Iowa.
- LÓPEZ-RUIZ R., VÁZQUEZ-PRADA M., GÓMEZ J.B., PACHECO A.F. (2004), *A model of characteristic earthquakes and its implications for regional seismicity*, Terra Nova, *16*, 116–120.
- MAIN I., NAYLOR M., GREENHOUGH J., TOUATI S., BELL A.F., MCCLOSKEY J. (2011), *Model selection and uncertainty in earthquake hazard analysis*. In: Faber M, Köhler J, Nishijima K (eds), Applications of Statistics and Probability in Civil Engineering. CRC Press, Leiden, The Netherlands, pp. 735–743.
- MCCAFFREY R. (2008), *Global frequency of magnitude 9 earthquakes*, Geology, *36*, 263–266.
- OKAL E.A. (1988), *Seismic parameters controlling far-field tsunami amplitudes: a review*, Natural Hazards, *1*, 67–96.
- OLAMI Z., FEDER H.J.S., CHRISTENSEN K. (1992), *Self-organized criticality in a continuous, nonconservative cellular automaton modeling earthquakes*, Physical Review Letters, *68*, 1244–1247.
- PACHECO J.F., SYKES L.R. (1992), *Seismic moment catalog of large shallow earthquakes, 1900 to 1989*, Bull. Seismol. Soc. Am., *82*, 1306–1349.
- PARSONS T., CONSOLE R., FALCONE G., MURRU M., YAMASHINA K. (2012), *Comparison of characteristic and Gutenberg-Richter models for time-dependent  $M \geq 7.9$  earthquake probability in the Nankai-Tokai subduction zone, Japan*, Geophys. J. Int., doi:10.1111/j.1365-1246X.2012.05595.x.
- PARSONS T., GEIST E.L. (2009), *Is there a basis for preferring characteristic earthquakes over a Gutenberg-Richter distribution*

- in probabilistic earthquake forecasting?*, Bull. Seismol. Soc. Am., 99, 2012–2019. doi:[10.1785/0120080069](https://doi.org/10.1785/0120080069).
- PARSONS T., GEIST E.L. (2012), *Were global  $M \geq 8.3$  earthquake time intervals random between 1900–2011?*, Bull. Seismol. Soc. Am., 102, doi:[10.1785/0120110282](https://doi.org/10.1785/0120110282).
- PARSONS T., GEIST E.L. (2014), *The 2010–2014.3 global earthquake rate increase*, Geophys. Res. Lett., 41, 4479–4485. doi:[10.1002/2014GL060513](https://doi.org/10.1002/2014GL060513).
- PAWITAN Y. (2001), *In all likelihood: statistical modelling and inference using likelihood*. Oxford University Press, Oxford.
- PELAYO A.M., WIENS D.A. (1992), *Tsunami earthquakes: slow thrust-faulting events in the accretionary wedge*, J. Geophys. Res., 97, 15,321–315,337.
- RABINOVICH A.B., THOMSON R.E. (2007), *The 26 December 2004 Sumatra tsunami: analysis of tide gauge data from the world ocean Part 1. Indian Ocean and South Africa*, Pure Appl. Geophys., 164, 261–308.
- SATAKE K., OKADA M., ABE I. (1988), *Tide gauge response to tsunamis: measurements at 40 tide gauge stations in Japan*, Journal of Marine Research, 46, 557–571.
- SORNETTE D. (2009), *Probability distribution in complex systems*. In: Meyers RA (ed), Encyclopedia of Complexity and Systems Science. Springer, New York, pp. 7009–7024.
- VERE-JONES D., ROBINSON R., YANG W. (2001), *Remarks on the accelerated moment release model: problems of model formulation, simulation and estimation*, Geophys. J. Int., 144, 517–531.
- WESNOUSKY S.G. (1994), *The Gutenberg-Richter or characteristic earthquake distribution, which is it?*, Bull. Seismol. Soc. Am., 84, 1940–1959.
- WHITE H. (1980), *A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity*, Econometrica, 48, 817–838.
- ZÖLLER G. (2013), *Convergence of the frequency-magnitude distribution of global earthquakes: maybe in 200 years*, Geophys. Res. Lett., 40, 3873–3877.

(Received February 1, 2016, revised March 28, 2016, accepted March 30, 2016, Published online April 30, 2016)