# Probabilistic Tsunami Hazard Assessment from Incomplete and Uncertain Historical Catalogues with Application to Tsunamigenic Regions in the Pacific Ocean

Ansie Smit[1], Andrzej Kijko[2], Alfred Stein[3]

(1) University of Pretoria Natural Hazard Centre, Department of Geology, Department of Statistics, University of Pretoria, Private Bag X20, Hatfield, Pretoria,0028, South Africa, ansie.smit@up.ac.za

(2) University of Pretoria Natural Hazard Centre, Department of Geology, University of Pretoria, Private Bag X20, Hatfield, Pretoria,0028, South Africa, andrzej.kijko@up.ac.za

(3) Faculty of Geo-information Science & Earth Observation (ITC), University of Twente, Enschede, The Netherlands, Department of Statistics, University of Pretoria, South Africa, a.stein@utwente.nl

Corresponding author: Ansie Smit; ansie.smit@up.ac.za; Tel: +27 12 420 2282; ORCID: 0000-0003-0315-0875

**Key Words**: probabilistic tsunami hazard assessment; incomplete and uncertain data; uncertain tsunami occurrence model; Pacific Ocean

**Abstract**

The paper presents a new method for empirical assessment of tsunami recurrence parameters namely, the mean tsunami activity rate $\lambda_T$, the Soloviev-Imamura frequency-magnitude power law $b_T$-value, and the coastline-characteristic, maximum possible tsunami intensity $i_{\max}$. The three coastline-characteristic recurrence parameters are estimated locally by maximum likelihood techniques using only tsunami event catalogues. The method provides for incompleteness of the tsunami catalogue, uncertainty in the tsunami intensity determination, and uncertainty associated with the parameters in the applied tsunami occurrence models. Aleatory and epistemic uncertainty is introduced in the tsunami models by means of the use of mixture distributions. Both the mean tsunami activity rate $\lambda_T$ of the Poisson occurrence model, and the $b_T$-value of the Soloviev-Imamura frequency-intensity power law are random variables. The proposed procedure was applied to estimate the probabilities of exceedance and return periods for tsunamis in the tsunamigenic regions of Japan, Kuril-Kamchatka and South America.

## 1. Introduction

Since 2004, four devastating tsunamis following major earthquakes have caused unprecedented destruction and loss of life. The events in Sumatra (2004), Japan (2011), and Chile (2010, 2015) have re-emphasised the need for comprehensive and accurate hazard assessment, taking into account both earthquakes and tsunamis. Earthquake hazard assessment procedures, both probabilistic and deterministic, are well-defined and applied; however, until quite recently, the same could not be said about tsunami hazard assessment.

Tsunamis can be generated by earthquakes, volcanogenic processes, submarine landslides, and meteorites. Among those, earthquakes with frequent occurrences generate the most destructive tsunamis, whereas the other processes have a much smaller probability of occurrence. Therefore, most current research focusses primarily on assessing the tsunami hazard generated by earthquakes.

In essence, the tsunami hazard for a specified area can be assessed in two ways, namely, by deterministic and probabilistic procedures. Deterministic methodologies focus on a single event, usually the worst-case scenario, by modelling the resultant wave run-up height (or tsunami intensity). Little or no emphasis is placed on the probability of such an event occurring (e.g. Tinti and Armigliato, 2003; Hébert *et al*., 2005; Paulatto *et al*., 2007; Lorito *et al*., 2008; Shaw *et al*., 2008; Wijetunge, 2014). Such an approach provides valuable information for disaster management agencies charged with response planning to minimize the potential effects. Recently researchers increasingly focus on assessing the deterministic tsunami hazard with the help of the worst-case credible tsunami scenario method (Tonini *et al*, 2011; Grilli *et al*., 2011 Harbitz *et al*., 2012). This involves the development of an aggregated scenario based on several, independently assessed deterministic scenarios. Probabilistic tsunami hazard assessment (PTHA), in turn, considers all possible earthquake scenarios (both locally and regionally) that could generate tsunamis of various intensities. Accordingly, this approach provides a comprehensive estimate of the total hazard that confronts a specific area (e.g. Geist and Parsons, 2006; Power *et al*., 2007; Thio *et al*., 2007, Sørensen *et al*., 2012; Brizuela *et al*., 2014). Determining the probability of occurrences and return periods is crucial for the risk mitigation processes of disaster management, engineering, and insurance companies. The disadvantage of PTHA through probabilistically assessing the occurrence of potential sources, is that it requires extensive research and modelling to identify all the potential sources, like for example earthquakes, and to assess the hazard and the resultant wave propagation. Moreover, these models have to consider aleatory and epistemic uncertainties to prevent the underestimation of the tsunami hazard. Aleatory uncertainty refers to the natural stochastic nature, which is characteristic of the physical system, whereas epistemic uncertainty is attributed to the incompleteness of the information on the system.

PTHA can be conducted using three distinct procedures. In the first procedure PTHA comprises of computational analyses, which take into account the known and available information on the earthquake source parameters as well as the propagation of waves, the recurrence rate, and the underlying uncertainties. This procedure generally follows the same route as standard probabilistic seismic hazard analysis (PSHA), as developed by Cornell (1968) and McGuire (1976). Examples of the application of this procedure are available in Lin and Tung (1982), Rikitake and Aida (1988), Ward and Asphaug (2000), Downes and Stirling (2001), Ward (2001, 2003), Geist (2005), Geist and Parsons (2006), Liu *et al*. (2007), Burbidge *et al*. (2008), González *et al*. (2009), Power *et al*. (2007), Sørensen *et al*. (2012) and Brizuela *et al*., (2014). The steps followed in computational PTHA are the identification and characterization of tsunamigenic source recurrences parameters. Such characterization includes the estimation of the coastline-characteristic maximum possible earthquake magnitude, the creation of tsunamigenic scenario earthquake event catalogues based on identified sources, numerical propagation modelling and the estimation of potential inundation. The results of all possible scenario events are then combined to create tsunami hazard curves and maps. The techniques used to provide explicit provision for epistemic uncertainty in the PTHA includes the use of a logic-tree approach, similar to that used for PSHA (Petersen *et al*., 2002; Annaka *et al*., 2007).

A second procedure is based on Bayesian inferences (Grezio *et al.*, 2010; Grezio *et al.*, 2012; Yadav *et al.*, 2013; Tatsumi *et al.*, 2014; Anita *et al.*, 2015). This procedure allows incorporation of all significant, independent prior information pertaining to the tsunami hazard for a specified region. In this way, the knowledge of the physical nature of the process is combined with the likelihood of the process, based on historical event information.

The third procedure focuses entirely on the empirical analysis of historical tsunami event records (Burroughs and Tebbens, 2005; Tinti *et al.*, 2005; Burbidge *et al.*, 2008). Accordingly, the assessment of the hazard can proceed without prior knowledge of the type or location of the process that triggered the tsunami. If sufficient data are available from tsunami catalogues, site-specific hazard curves can be developed or alternatively be used as regional priori information relating to far-field tsunami-generating sources (Geist and Parson, 2006). Using this procedure, the PTHA for an area can be assessed by means of stochastic modelling of conditional probabilities of tsunami recurrence times (Orfanogiannaki and Papadopoulos, 2007) or by means of frequency-magnitude relationships (Soloviev, 1970; Rikitake and Aida, 1988; Tinti, 1993). Empirical methods, however, are highly dependent on the completeness and quality of the tsunami catalogue.

The major problem with the empirical modelling of tsunami intensity data is that it requires adequate historical and instrumental information for the tsunami run-up heights or intensity values. This problem is discussed in various studies, e.g. Burroughs and Tebbens (2005) and Anita *et al.* (2015). Moreover, small datasets can introduce high levels of uncertainty and significant bias to the results. The proposed procedure of the current study aims to solve this problem by applying methodology specifically designed to assess recurrence parameters for highly incomplete and uncertain datasets.

The methodology proposed in this study is focused on the empirical assessment of the tsunami recurrence parameters for probabilistic tsunami hazard assessment, based on a similar methodology as described by Kijko *et al.* (2016) for earthquakes. These recurrence parameters are the mean tsunami activity rate $\lambda_T$, the Soloviev frequency-magnitude power law $b_T$-value, and the coastline-characteristic, maximum possible tsunami intensity $i_{\max}$. The three tsunami recurrence parameters are derived from tsunami catalogues by taking into consideration the incompleteness of the catalogue, uncertainty in the tsunami intensity determination, as well as the uncertainty associated with the applied tsunami occurrence models. The uncertainty in the tsunami occurrence models is introduced by assuming that both the mean tsunami activity rate $\lambda_T$ and the frequency-magnitude power law $b_T$-value are random variables, each described by the secondary mixing distribution such as the gamma distribution. This approach results in the extension of the classic frequency–intensity Soloviev-Imamura relation and the Poisson distribution of the number of tsunamis with their mixture distribution counterparts (Benjamin, 1968; Campbell, 1982, 1983). These mixture distributions are often referred to as Bayesian distributions (Campbell, 1982) or sometimes as compound (Benjamin and Cornell, 1970) distributions. This procedure is adapted for application to a tsunami event catalogue to assess the probabilities of exceedance for a specified tsunami intensity $i$, and the relevant return periods. In the next section, the methodology by Kijko *et al.* (2016) is defined in terms of tsunamis and, subsequently, applied to the tsunamigenic zones of Japan, Kuril-Kamchatka, and South America.

## 2. Methodology

The proposed PTHA procedure requires knowledge of three parameters, namely, (1) the mean tsunami activity rate $\lambda_T$ representing the number of tsunamis expected to occur within a specified stretch of coastline, with intensities equal or larger than $i_{min}$ within a unit of time, usually one year; (2) the parameter $b_T$ describing the ratio between weak and strong tsunami intensities; and (3) the coastline-characteristic maximum possible tsunami intensity $i_{max}$. In PTHA, the parameter $b_T$ plays the same role as the $b$-value in the frequency-magnitude Gutenberg-Richter relationship for earthquakes. In addition, the coastline-characteristic maximum possible tsunami intensity $i_{max}$ is assessed in a similar manner as that of the area-characteristic maximum possible earthquake magnitude $m_{max}$.

In this section, the applied models and procedures for the assessment of the three key tsunami recurrence parameters are outlined through the introduction of the tsunami occurrence distribution. The parameters are the activity rate $\lambda_T$, the intensity distribution model for $b_T$, which describes the relation between number of events and the intensities, as well as a method to estimate the coastal-characteristic maximum possible intensity $i_{max}$. In the rest of the text, the parameters $\lambda_T$ and $b_T$ will be denoted as $\lambda$ and $b$, respectively, unless indicated otherwise.

### 2.1. Nature of Input Data

Similar to earthquake catalogues, tsunami records are frequently incomplete, thus complicating the statistical analysis of tsunami hazard (Burroughs and Tebbens, 2005; Anita *et al.*, 2015). Fig. 1 depicts the proposed approach to assess tsunami hazard by making provision for the tsunami event catalogue to contain the following two types of data, namely, (1) the strongest historic tsunamis that occurred over the last few hundred years, and (2) complete recent data observed over a recent and relatively short period of time. Often, the complete part of the catalogue can be divided into sub-catalogues, each of which is complete for tsunami events above a given threshold intensity $i_{min}^{(j)}$, and which occurred in a certain period of time $t_j$, where $j = 1, \dots, s$ and $s$ is the number of complete tsunami sub-catalogues. The different levels of completeness $i_{min}^{(j)}$ often follow a monotonically decreasing trend over time, with the most recent sub-catalogues having the lowest level of completeness. One advantage of viewing the data in this form is that provision can be made explicitly for time gaps $(T_g)$ in the event catalogue when no event information is available or has been collected.

**[FIGURE 1]**

**Fig. 1**. Illustration of data that can be used to obtain recurrence parameters of the tsunami. The applied approach allows combining the largest (historic tsunamis), and complete data having a variable level of completeness. This approach accepts 'gaps' $(T_g)$ where records are missing. (Modified after Kijko and Sellevoll, 1992)

The desired coastline-characteristic tsunami hazard parameters $\boldsymbol{\theta} = (\bar{\lambda}, \bar{\beta}, i_{\max})$ are estimated using the maximum likelihood estimation (MLE). Likelihood functions are defined for the historic part $L_H(\boldsymbol{\theta})$, as well as for complete catalogue $L_C(\boldsymbol{\theta})$ which consists of the likelihood function of each of the $s$ sub-catalogues. Utilizing the additive property of the likelihood function $L(\boldsymbol{\theta})$, (Rao, 1973), the MLE function for the entire catalogue is of the form

$$L(\boldsymbol{\theta}) = L_H(\boldsymbol{\theta}) \times L_C(\boldsymbol{\theta}) . \tag{1}$$

Each of the likelihood functions are defined in terms of the tsunami occurrence and intensity distribution models.

### 2.2. Tsunami occurrence model

First, the tsunami occurrence model is defined. It relies on the assumption that the temporal distribution of the tsunami events observed along a stretch of coastline can be modelled by a Poisson process (e.g. Geist and Parsons, 2006). This corresponds with a similar assumption for the temporal distributions of tsunami-generating processes such as earthquakes (Kramer, 1996; Cornell, 1968; Lomnitz, 1973; Gardner and Knopoff, 1974; Cao and Gao, 2002), submarine landslides (Urlaub *et al.*, 2013.), volcanogenic processes (Jones *et al.*, 1999; Dzierma and Wehrmann, 2010), and meteorites (Ward and Asphaug, 2000). Approximately 75% of the global recorded historical tsunamis are attributed to earthquakes (Gusiakov, 2009).

Following the Poisson assumption, the probability that along a certain stretch of coastline $n$ tsunami events will be observed within a specified time interval $t$ is described as

$$P_N(n|\lambda, t) \equiv P(N = n|\lambda, t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad n = 0,1,2 \dots \tag{2}$$

where $\lambda \equiv \lambda(i_{\min})$ refers to the parameter of the Poisson distribution and describes the mean (annual) activity rate of tsunamis with an intensity greater than or equal to $i_{\min}$.

### 2.3. Intensity distribution model

Similar to earthquakes, the number of tsunami events observed are related to the intensity (size) of the events. Studies by Soloviev (1970), Houston *et al.* (1977, including references listed within), Horikawa and Shuto (1983), Burroughs and Tebbens, (2005), and Geist and Parsons (2006) indicate that tsunami intensity can be described by a power-law distribution in the format

5

$$n(i) = a10^{-bi} \tag{3}$$

where $i$ is the Soloviev-Imamura tsunami intensity, $n(i)$ is the number of tsunami events per annum, and $a$ and $b$ are defined as constants. Defining $\beta = b\ln(10)$ and $\alpha = \ln a$, equation (3) can be rewritten as

$$n(i) = e^{\alpha - \beta i}. \tag{4}$$

Empirical studies show that the tsunami intensity $i$ follows the same power-law distribution as the earthquake magnitude $m$ in the case where the applied scale of tsunami strength is one of the scales, as introduced by Sieberg (1927), Soloviev (1970), Ambraseys (1962), Papadopoulos and Imamura (2001), and Geist and Parsons (2006). Equation (4) is therefore equivalent to the classic Gutenberg-Richter frequency-magnitude relation $\log n(m) = a - bm$ (Gutenberg and Richter, 1942; 1956), with $m$ indicating the earthquake magnitude. The power law defined in equation (3) is not the only model used to describe the distribution of tsunami intensity $i$. A comprehensive review and discussion of alternative tsunami size distributions were presented by Burroughs and Tebbens (2005) and Geist and Parsons (2006). The Soloviev (1970) scale is used in the current study as it is the standard intensity scale used in most tsunami event datasets.

Regardless of the applied tsunami intensity scale, the intensity value $i$ is linked with the average tsunami run-up $h$ along a coastline. For example, when the intensity $i$ is expressed in the units of the Soloviev scale (Soloviev, 1970), the value of $i$ is calculated from the average tsunami run-up height $h$ (in meters) along a coastline as

$$i = \log_2\left(\sqrt{2h}\right). \tag{5}$$

We assume that the tsunami intensities $i$ occurring along a stretch of coastline are continuous, independent random values distributed according to the power-law (equation (3)). Following, for example, Burroughs and Tebbens (2001; 2005) the probability distribution function (PDF) and the cumulative distribution function (CDF) of tsunami intensity takes the same functional form as the CDF for earthquake magnitude (Page, 1968; Cosentino $et~al.$, 1977). The PDF and CDF is therefore defined for intensity $i$, for a minimum threshold value $i_{min}$, and is truncated from the top by the coastline-characteristic maximum possible tsunami intensity $i_{\max}$ , such that

$$f_{\mathrm{I}}(i) = \begin{cases} 0 & i_{\min} < i \\ \dfrac{\beta \exp[-\beta(i - i_{\min})]}{1 - \exp[-\beta(i_{\max} - i_{\min})]} & \text{for} \quad i_{\min} \leq i \leq i_{\max}, \\ 0 & i > i_{\max} \end{cases} \tag{6}$$

6

and

$$F_{\mathrm{I}}(i) = \begin{cases} 0 & i_{\min} \leq i \\ \dfrac{1 - \exp[-\beta(i - i_{\min})]}{1 - \exp[-\beta(i_{\max} - i_{\min})]} & \text{for} \quad i_{\min} \leq i \leq i_{\max} \,. \\ 1 & i > i_{\max}, \end{cases} \tag{7}$$

for $\beta = b\ln(10)$.

### 2.4. Account for uncertainty in occurrence models

The types of uncertainties which are accommodated in the proposed procedure are the uncertainty associated with the determination of tsunami event intensities and the uncertainty associated with the parameters of the applied occurrence models.

Most earthquake event catalogues indicate that seismogenic processes are subjected to variations ranging from temporal trends to pure random fluctuations. A wide range of opinions on spatial and temporal fluctuations in seismicity patterns, as well as an extensive reference list can be found e.g. in the publications by Ogata and Abe (1991), Mora *et al.* (2001), Karakaisis *et al*. (2002) and, more recently, Talbi and Yamazaki (2009). As the majority of tsunami phenomena are generated by earthquakes (Gusiakov, 2009), it is assumed that the fluctuation in the earthquake activity rate gives rise to similar fluctuations in their tsunami counterparts.

In the proposed approach, the potential temporal and spatial fluctuations in the tsunami recurrence parameters are accommodated by using mixture distributions. In utilizing mixing distributions, the parameters of the Poisson occurrence model ($\lambda$) and the exponential intensity distribution ($\beta$) are treated as random variables that can be described by a specific distribution. The Poisson and exponential distributions are defined as the original distribution, and the random variable parameter distributions as the mixing distribution. A convenient choice for a mixing distribution in many research fields is the two parameter gamma PDF. This distribution is flexible and able to take on several different forms through the shape ($p$) and scale ($q$) parameters and provide solvable solutions for the mixture distributions. The mixture distributions can also be described in terms of empirical Bayesian theory where the gamma mixing distribution is considered as the priori distribution, the resultant mixture distribution as the predictive posterior distribution and the parameters $p$ and $q$ as the hyperparameters. Yadav *et al*. (2013) applied the Bayesian formalism in terms of a uniform priori distribution for all three recurrence parameters. The application of the classic Bayesian theory requires that an assumption be made regarding the uncertainty in the data through an a priori distribution before the analysis of the data. With mixture distributions, the uncertainty in the data follows directly from the observed data itself. The mixture distributions defined in this paper provides an alternative formalism to the classic Bayesian method while relying purely on the empirical data.

The tsunami occurrence distribution (equation (2)) which makes provision for uncertainty in parameter $\lambda$, can be defined as a Poisson-gamma distribution of the form

$$P_n(\bar{\lambda}, t, v_\lambda) = \int_0^\infty P_n(\lambda, t) f_\Lambda(\lambda) d\lambda = \frac{\Gamma(n + q_\lambda)}{n! \, \Gamma(q_\lambda)} \left(\frac{p_\lambda}{t + p_\lambda}\right)^{q_\lambda} \left(\frac{t}{t + p_\lambda}\right)^n \qquad (8)$$

where $\Gamma(q)$ is the gamma function, the parameters $p_\lambda$ and $q_\lambda$ represent the parameters of the gamma function such that $p_\lambda = \bar{\lambda}/\sigma_\lambda^2$ and $q_\lambda \equiv v_\lambda^{-2} = \bar{\lambda}^2/\sigma_\lambda^2$, and $\bar{\lambda}$ denotes the mean value of the activity rate $\lambda$. Similarly, the tsunami intensity distribution (equation (7)) can be redefined to include the uncertainty through the mixture exponential-gamma distribution defined as

$$F_I(i|v_\beta, i_{\min}) = \left[1 - \left(\frac{p_\beta}{p_\beta + i_{\max} - i_{\min}}\right)^{q_\beta}\right]^{-1} \left[1 - \left(\frac{p_\beta}{p_\beta + i - i_{\min}}\right)^{q_\beta}\right], \qquad (9)$$

with $p_\beta = \bar{\beta}/\sigma_\beta^2$ and $q_\beta \equiv v_\beta^{-2} = \bar{\beta}^2/\sigma_\beta^2$., and $\bar{\beta}$ and $\sigma_\beta$, respectively, denote the mean and the standard deviation of $\beta$. Both equations (8) and (9) can be rewritten in terms of $q_\lambda$ and $q_\beta$.

### 2.5. Accounting for Uncertainty in Tsunami Intensity Determination

By definition, it is impossible to assess the tsunami intensity without error. The scaling coefficients in the frequency-intensity power law depend on various factors, such as the length of the tsunami catalogue, the measurement range, and the level of completeness of the catalogue (Geist and Parsons, 2006). Historical tsunami catalogues can contain inaccuracies attributable to the incomplete nature of the available information. Errors that can appear in the catalogue consist of the misclassification of other meteorological or hydrological phenomena and the misinterpretation of the available data descriptions (Gusiakov, 2009).

To account for the possibility of aleatory and epistemic errors in the estimation of tsunami intensities, two types of models are considered, namely, the Hard Bounds Model and the Soft Bounds Model. In the Hard Bounds Model (Kijko and Sellevoll, 1992), the real, unknown tsunami intensity is assumed to occur within fixed boundary limits. The Soft Bounds Model, in turn, assumes that the uncertainty over the real, unknown magnitude can be described by a Gaussian distribution. For the purpose of this article, a Soft Bound Model is applied with a zero mean and a standard deviation $\sigma_I$, such that the CDF (10) of the observed tsunami intensity becomes (Tinti and Mulargia, 1985)

$$G_I(i) = \int_{i_{\min}}^{i_{\max}} f_I(\zeta) \left[\int_{-\infty}^{i-\zeta} \frac{1}{\sigma_I \sqrt{2\pi}} \exp\left(-\frac{\epsilon^2}{2\sigma_I^2}\right) d\varepsilon\right] d\zeta, \qquad (10)$$

in which the PDF $f_I(\zeta)$ is the mixture exponential-gamma intensity distribution (equation (8)) and has the closed-form expression (Kijko *et al.*, 2016)

8

$$G_I(i) = \frac{C_\beta \bar{\beta} q_\beta^{q_\beta+1}}{2\sigma_I} \{A + B\}$$

$$A = \frac{(r_1 + r_2\alpha)^{-q_\beta}}{r_2 q_\beta}\Bigg|_{\frac{i-i_{\min}}{\sigma_I}}^{\frac{i-i_{\max}}{\sigma_I}}$$

(11)

$$B = \left(\frac{2}{\pi}\right)^{1/2} \sum_{h=0}^{\infty} \frac{(-1)^h}{2^h h! (2h+1)} \frac{1}{b^{w+1}} \sum_{j=0}^{2h+1} \frac{w! (-r_1)^j (r_1 + r_2\alpha)^{w-q_\beta-j}}{(w-j)! j! (w - q_\beta - j)}\Bigg|_{\frac{i-i_{\min}}{\sigma_I}}^{\frac{i-i_{\max}}{\sigma_I}}$$

with $a = q_\beta + \bar{\beta}(i - i_{\min})$, $b = -\bar{\beta}\sigma_I$, $C_\beta = \left[1 - \left(\frac{q_\beta}{q_\beta + \beta(i_{max}-i_{min})}\right)^{q_\beta}\right]^{-1}$, $r_1 = q_\beta + \beta(i - i_{min})$, $r_2 = -\bar{\beta}\sigma_i$

and $w = 2h + 1$. The derivation of equation (12) is available in the appendix of Kijko *et al.*, (2016).

### 3. Estimation of Recurrence Parameters

The desired coastline-characteristic tsunami hazard parameters $\boldsymbol{\theta} = \left(\bar{\lambda}, \bar{\beta}, i_{\max}\right)$ are estimated by maximising the likelihood function (equation (1)). The likelihood functions are derived separately for historical and complete tsunami catalogues.

#### 3.1. Historical tsunami catalogues

As historical, pre-instrumental tsunami catalogues are considered to contain only large events, the likelihood function can be constructed by using an extreme distribution function. In the historical catalogue it is assumed that the time span $t_0$, contains $n_0$ largest tsunami events, each with an intensity equal to or exceeding a certain intensity threshold value $i_0$, such that $i_0$ is larger than or equal to the overall minimum intensity of interest for the entire catalogue $i_{\min}$. This overall minimum intensity should not exceed any of the other threshold intensities in the overall catalogue. The time span $t_0$ is further defined as the sum of the time intervals $t_{0k}$ between the historical events for $k = 1, \dots, n_0$.

The probability that in an arbitrary time interval $t$, either no tsunami occurs or all the occurring events have intensities not exceeding $i_0$ could be expressed as the CDF

$$F_I^{max}(i|v_\beta, i_0, t) = \sum_{n=0}^{\infty} P_k(\bar{\lambda}, t, v_\lambda)[F_I(i|v_\beta, i_0)]^n = \left[\frac{q_\lambda}{q_\lambda + \bar{\lambda}_0 t_0[1 - F_I(i|v_\beta,\ i_0)]}\right]^{q_\lambda}$$

(12)

with the corresponding PDF for the largest historic tsunami intensity event in time period $t$ defined as

$$f_I^{\max}(i|v_\beta, i_0, t) = \frac{\bar{\lambda}_0\, t q_\lambda f_I(i|v_\beta, i_0) F_I^{\max}(i|v_\beta, i_0, t)}{q_\lambda + \bar{\lambda}_0 t[1 - F_I(i|v_\beta, i_0)]},$$

(13)

where $\bar{\lambda}_0 = \bar{\lambda}\left[1 - F_I^{max}(i|v_\beta, i_0)\right]$ denotes the mean (usually annual) activity rate for tsunami events with intensities $i \geq i_0$, and $\bar{\lambda} \equiv \bar{\lambda}(m_{\min})$ is the mean activity rate corresponding to the intensity value $i_{\min}$.

9

The likelihood function for the tsunami intensity $i_0$ within a period $t_0$, for the unknown vector parameter $\boldsymbol{\theta}$, is described in terms of the $(n_0 \times 1)$ vectors $\boldsymbol{i}_0$ and $\boldsymbol{t}_0$ in the form

$$L_H(\boldsymbol{\theta}|\boldsymbol{i}_0, \boldsymbol{t}_0, \boldsymbol{v}) \equiv L_H(\boldsymbol{\theta}) = \prod_{k=1}^{n_0} f_{\mathrm{I}}^{\max}(i_{0k}|v_\beta, i_0, t_{0k}). \tag{14}$$

The vectors consist of the $n_0$ largest tsunami intensities $i_{0k}$, associated with time intervals $t_{0k}$, when $k = 1, \dots, n_0$. The vector $\boldsymbol{v} = (v_\lambda, v_\beta)^T$ consists of the coefficients of variation of the unknown $\bar{\lambda}$ and $\bar{\beta}$.

### 3.2. Complete catalogues at different levels of completeness

The complete part of the tsunami event catalogue can be divided into $s$ sub-catalogues, each with a time span $t_j$ ($j = 1, \dots, s$), and completeness starting from the known tsunami intensity $i_{\min}^{(j)}$. The function $L_{\mathrm{j}}(\boldsymbol{\theta}|\boldsymbol{i}_{\mathrm{j}})$ is the likelihood function of the unknown vectors of the $j$th sub-catalogue i.e.. $\boldsymbol{\theta} = (\bar{\lambda}, \bar{\beta}, i_{\max})$ and $\boldsymbol{i}_j$ (tsunami intensities with thresholds such that $i_{jk} \geq i_{\min}^{(j)}$, for $k = 1,2,\dots, n_j$). Furthermore, by assuming that the tsunami intensities are independent of each other, the likelihood function $L_{\mathrm{j}}(\boldsymbol{\theta}|\boldsymbol{i}_{\mathrm{j}}, \boldsymbol{t}_j)$ is defined for each complete sub-catalogue $j$ and is the product of the two functions, $L_{\mathrm{j}}(\bar{\lambda}|n_j, t_j)$, the likelihood function of $\bar{\lambda}$, as well as $L_{\mathrm{j}}(\bar{\beta}|\boldsymbol{i}_{\mathrm{j}})$, the likelihood function of $\bar{\beta}$. The likelihood function $L_{\mathrm{j}}(\bar{\lambda}|n_{\mathrm{j}}, t_{\mathrm{j}})$ is subsequently defined on the basis that the occurrence model of tsunamis follows the mixture Poisson-gamma distribution (equation (8)), such that

$$L_j(\bar{\lambda}|n_j, t_j) = (\bar{\lambda}^{(j)} t_j + q_\lambda)^{-q_\lambda} \left( \frac{\bar{\lambda}^{(j)} t_j}{\bar{\lambda}^{(j)} t_j + q_\lambda} \right)^{n_j}, \tag{15}$$

in which $\bar{\lambda}^{(j)}$ is the mean (annual) activity rate, corresponding to threshold intensity level of completeness $i_{\min}^{(j)}$ and is given by

$$\bar{\lambda}^{(j)} = \bar{\lambda}[1 - F_{\mathrm{I}}(i_{\min}^{(j)}|v_\beta, i_{\min})]. \tag{16}$$

with $F_M(\cdot \,|\, \cdot)$ as defined in equation (9). Similarly, the likelihood function $L_{\mathrm{j}}(\bar{\beta}|\boldsymbol{i}_{\mathrm{j}})$, of parameter $\bar{\beta}$ based on $j$th sub-catalogue of the complete part of tsunami catalogue, is defined as

$$L_j(\bar{\beta}|\boldsymbol{i}_j) = \left[ \bar{\beta} \left[ 1 - \left( \frac{p_\beta}{p_\beta + i_{max} - i_{min}} \right)^{q_\beta} \right]^{-1} \right]^{n_j} \prod_{k=1}^{n_j} \left[ 1 + \frac{\bar{\beta}}{q_\beta} (i_{jk} - i_{\min}^{(j)}) \right]^{-(q_\beta+1)}. \tag{17}$$

The likelihood function, based on all $s$ complete sub-catalogues takes the form

$$L_C(\boldsymbol{\theta}) = \prod_{j=1}^{s} L_j(\bar{\lambda}|n_j, t_j)\, L_j(\bar{\beta}|\boldsymbol{i}_j). \tag{18}$$

### 3.3. Estimation of $i_{max}$

The coastal-characteristic maximum possible intensity $i_{\max}$ can not be calculated using the sample likelihood function $L(\boldsymbol{\theta})$, as defined in (18), as it violates the conditions of regularity (Eadie *et al.,* 1971; Davison, 2003). The procedure by Cooke (1979) is therefore considered, where the possible maximum intensity is determined as the observed maximum intensity of the entire catalogue, plus some additional information. This leads to an estimated value of the largest expected tsunami intensity $\boldsymbol{E}[i_{\max}^{obs}; t]$ as discussed by the authors Kijko (2004) and Kijko and Singh (2011)

$$i_{\max} = i_{\max}^{obs} + \int_{i_{\min}}^{i_{\max}} [F_I(\zeta)]^n d\zeta. \tag{19}$$

Here, $F_I(\zeta)$ denotes the mixture CDF of tsunami intensity (equation (9)). The integral in equation (19) only has an analytical solution through the application of the approximation by Cramér (1961). This leads to an equation for the coastal-characteristic maximum possible intensity in the form

$$i_{\max} = i_{\max}^{obs} + \frac{\delta^{\frac{1}{q_\beta}} exp\left[\frac{nr^{q_\beta}}{(1-r^{q_\beta})}\right]}{\bar{\beta}}\left[\Gamma\left(-\frac{1}{q_\beta}, \delta r^{q_\beta}\right) - \Gamma\left(-\frac{1}{q_\beta}, \delta\right)\right] \tag{20}$$

in which $r = \frac{p_\beta}{p_\beta + i_{\max} - i_{\min}}$, $\delta = n\left[1 - \left(\frac{p_\beta}{p_\beta + i_{max} - i_{min}}\right)^{q_\beta}\right]^{-1}$ and $\Gamma(\cdot,\cdot)$ is the complementary incomplete gamma function (Abramowitz and Stegun, 1970), which is solved iteratively. The approximate variance of $i_{\max}$ is

$$Var(i_{\max}) = \sigma_I^2 + \left[\frac{\delta^{\frac{1}{q_\beta}} exp\left[\frac{nr^{q_\beta}}{(1-r^{q_\beta})}\right]}{\bar{\beta}}\left[\Gamma\left(-\frac{1}{q_\beta}, \delta r^{q_\beta}\right) - \Gamma\left(-\frac{1}{q_\beta}, \delta\right)\right]\right]^2, \tag{21}$$

in which $\sigma_I$ denotes the standard error in the determination of the largest observed tsunami intensity $i_{\max}^{obs}$.

One disadvantage of using the procedure in the assessment of $i_{\max}$ is the potential underestimation of the value, especially when the tsunami event catalogue is short (Kijko, 2004). Kijko (2012) discussed potential improvements of the procedure; however, it is not implemented in this article.

The $i_{max}$ in equation (20) is defined through the truncation of the frequency-intensity power law and is known as a sharp cut-off intensity value. It means that no tsunami of intensity exceeding $i_{max}$ is considered to be possible. One has to note that this is only one of the many possible models of assessing a sharp cut-off value according to the

11

frequency-intensity distributions. An alternative model for the assessment of the upper limit for the frequency-intensity can also be considered namely the soft cut-off model as described in earthquake studies. This model assumes that after a certain magnitude value, the prevailing distribution decays at a faster rate than the assumed power-law. Authors such as Kagan (1991, 1997, 2002 a, b), Main (1996), Sornette and Sornette (1999), Vere-Jones et al. (2001), Pisarenko and Sornette (2001), Kagan and Schoenberg (2001), Bird and Kagan (2004), Kagan (2010) and Pisarenko et al. (2008, 2014) discuss the finiteness of earthquake magnitude in terms of the dissipative nature of physical dynamic systems. Several potential 'decay' distributions are discussed including the exponential and gamma distribution. For the purpose of this paper, the hard cut-off method was used but the applied methodology is open to any model of the frequency-intensity distribution.

The desired coastline-characteristic tsunami likelihood function $L(\boldsymbol{\theta})$ as defined in equation (1) is obtained by combining the likelihood functions for historical (equation (14)) and complete catalogues (equation (18)) maximising the likelihood function (equation (1)). By maximizing $L(\boldsymbol{\theta})$ through the MLE procedure, the tsunami parameters $\boldsymbol{\theta} = \left(\bar{\lambda}, \bar{\beta}, i_{\max}\right)$ are obtained.

## 4. Special Instances

Similar to earthquake catalogues, tsunami event catalogues can at times consist of only historical or only complete data (Fig. 1). The tsunami recurrence parameters for an instance where only historical data are available is defined as

$$\begin{cases} \dfrac{1}{\lambda} = \dfrac{\bar{\iota}_0 - \overline{t_0 A}}{A_2 - A_1} \\[4mm] \dfrac{1}{\beta} = \dfrac{\overline{t_0 \iota_0 A} - \overline{t_0} A_2 i_{\max}}{\overline{t_0 A} - \overline{t_0} A_2} \end{cases} \tag{22}$$

where $\boldsymbol{A}$ is a column vector, with elements $A(i_{0k}) = \exp(-\bar{\beta} i_{0k})$, $k = 1, \dots, n_0$, and

- $\bar{\iota}_{0k} = \sum_{k=1}^{n_0} i_{0k}/n_0$
- $\overline{t_{0k}} = \sum_{k=1}^{n_0} t_{0k}/n_0$
- $\overline{t_{0k} A} = \sum_{k=1}^{n_0} t_{0k} exp(-\bar{\beta} i_{0k})/n_0$
- $\overline{t_{0k} \iota_{0k} A} = \sum_{k=1}^{n_0} t_{0k} i_{0k} exp(-\bar{\beta} i_{0k})/n_0$
- $A_1 = \exp(-\bar{\beta} i_{\min})$.
- $A_2 = \exp\left(-\bar{\beta} i_{\max}\right)$

The maximum likelihood estimates for the recurrence parameters $\lambda$ and $\beta$ for equation (22) are obtained for the specified $i_{\max}$.

Similarly, the estimated recurrence parameters are obtained in an instance when only complete tsunami catalogues are available. By applying the additive property of likelihood functions (Rao, 1973), the maximum likelihood estimate of the $\bar{\lambda}$ and $\bar{\beta}$-value are derived as in Kijko and Smit (2012)

$$\begin{cases} \hat{\lambda}(i_{\min}) = \dfrac{n}{\sum_{j=1}^{s} t_j \exp[-\hat{\beta}\left(i_{\min}^{(j)} - i_{\min}\right)]} \\[2em] \hat{\beta} = \left(\dfrac{u_1}{\hat{\beta}^{(1)}} + \dfrac{u_2}{\hat{\beta}^{(2)}} + \cdots + \dfrac{u_s}{\hat{\beta}^{(s)}}\right)^{-1} \end{cases} \tag{23}$$

where $u_j = n_j/s$; $n = \sum_{j=1}^{s} n_j$ is the total number of tsunamis in a complete catalogue, with intensities equal to or exceeding the relevant level of completeness $i_{\min}^{(j)}$, and $\hat{\beta}^{(j)}$ is defined as

$$\hat{\beta}^{(j)} = \frac{1}{\bar{\iota}_j - i_{\min}^{(j)}} . \tag{24}$$

Here, $\bar{\iota}_j = \sum_{k=1}^{n_j} i_{jk}/n_j$ denotes the mean value of tsunami intensities observed within the complete part of catalogue $j$.

## 5.  Example of Application

We now apply the proposed methodology to three tsunamigenic regions in the Pacific Ocean, namely Japan (JAP), Kuril-Kamchatka (K-K), and the western coast of South America (SAM). The international tsunami database used in this study was provided by Dr. V. K. Gusiakov of the Novosibirsk Tsunami Laboratory of the Institute of Computational Mathematics and Mathematical Geophysics (NTL/ICMMG) SDRAS, Novosibirsk, Russia (HTDB/WLD, 2013). This database hosts worldwide historical and instrumental tsunami observations for the period 47 BC to 2015, with the main intensity measurement being the Soloviev–Imamura scale, ranging from -5 to 5. The boundaries for the identified tsunamigenic regions as provided in the NTL/ICMMG database were used. These boundaries, depicted in Fig 2, follow the zoning used in the NGDC map by Lockridge and Smith (1984) and Gusiakov (2005).

[FIGURE 2]

**Fig. 2.** Positions of the three investigated tsunamigenic regions in the Pacific Ocean (modified after Gusiakov, 2005), being the source locations of the tsunamis for Japan (JAP), Kuril-Kamchatka (K-K), and South America (SAM).

The GITEC catalogue criteria (Tinti and Maramai, 1996; Tinti *et al*., 2001) were used to ensure that the data employed represented the best available information. This criterion assigns a validity index to each observation in the database, such that index value 0 indicates that the event is considered extremely improbable (probability near 0%), 1 is considered as improbable (probability approximately 25%), 2 as questionable (probability approximately 50%), 3 as probable (probability approximately 75%), and 4 as a definite tsunami (probability near 100%). The final catalogues for the three identified regions contain only tsunami events with a validity index of 3 or higher. Events with no coordinates (latitude or longitude) or no intensity values were excluded. The final catalogues for Japan, Kuril-Kamchatka, and South America contained 214, 73, and 79 events, respectively.

In the first step, the historical and complete sub-catalogues were identified for each of the three regions. Each sub-catalogue and level of completeness were identified with the help of the visual cumulative method (Mulargia and Tinti, 1985). Provision was made for intensity errors of 0.5 and 0.1 units, respectively, for the historical and the complete catalogues in each of the regions. Table 1 provides the input parameters, as well as the estimated recurrence parameters based on the proposed procedure for probabilistic tsunami hazard assessment.

It was assumed that the uncertainty for the tsunami-occurrence model parameters was equal to an adjustable, arbitrary value of 25%, meaning that the space-time variation in the estimated recurrence parameters of $\lambda$ and $\beta$ (or equivalently the $b$-value) was 0.25. This 25% uncertainty follows from extensive investigations into the analogous parameters of earthquake occurrence models based on tectonic datasets across the globe. Fig. 3 depicts the estimated mean return period, Fig. 4 the annual probability of exceedance, and Fig. 5 the probabilities of exceedance for 5, 10, and 25 years for various tsunami intensities. Figures 3 and 4 also provide the one standard deviation confidence intervals for the mean return periods and probabilities of exceedance. The values are calculated using the mean and standard deviations from the variance-covariance matrix. From a physical point of view, these confidence intervals reflect the uncertainty of the input data as well as the uncertainty associated with the parameters in the applied occurrence models.

Table 2 provides examples of the return period and probabilities of exceedance for each of the regions for intensities larger than 1.5 and 2.5 and for the time periods 1, 10, and 25 years.

**[TABLE 1]**

**Table 1.** Probabilistic tsunami hazard assessment input parameters and estimated recurrence parameters. The value $m_{min}$ represents the level of completeness and SE represents the standard error (assumed uncertainty) in the intensity estimation.

**[FIGURES 3a-c; 4a-c; 5a-c]**

(a)
(b)
(c)
**Fig. 3**. The annual mean return period (in years) and its associated uncertainty for various tsunami intensities for a) Japan, b) Kuril-Kamchatka, and c) South America. The shaded area represents the relevant one standard deviation confidence interval for the calculated values.

(a)
(b)
(c)

**Fig. 4**. The annual probability of exceedance and its associated uncertainty for various tsunami intensities for a) Japan, b) Kuril-Kamchatka, and c) South America. The shaded area represents the one standard deviation relevant confidence interval for the calculated values.

(a)

(b)

(c)

**Fig. 5**. The probability of exceedance for 5, 10, and 25 years for various tsunami intensities for a) Japan, b) Kuril-Kamchatka, and c) South America

[TABLE 2]

**Table 2**. Return period and probabilities of exceedance for Japan, Kuril-Kamchatka, and South America for intensities larger than 1.5, 2.0, and 2.5 and for time periods 1, 10, and 25 years.

The results obtained were compared with the respectable $b$-values obtained by Orfanogiannaki and Papadopoulos (2007), who also investigated the regions of Japan, Kuril-Kamchatka, and South America by developing stochastic approaches to utilise observed earthquakes in order to assess the potential for tsunami generation. These authors investigated the probability of tsunami occurrence, firstly, by using conditional probabilities for tsunami occurrence as a function of time and, secondly, by obtaining the product of the conditional probabilities with the ratio of the number of earthquakes generating tsunamis and the total number of earthquakes by applying the total probability theorem. Orfanogiannaki and Papadopoulos (2007) utilised the database, as described in Gusiakov (2001), for the time periods 1900 to 2000. The $b$-values, obtained with the help of linear regression for the three regions of interest, are 0.34 for Japan, 0.56 for Kuril-Kamchatka, and 0.34 for South America. The associated estimated mean annual rates of activity $\lambda$ for the three regions are 1.02 for Japan, 2.26 for Kuril-Kamchatka, and 1.64 for South America. The results obtained for Japan are comparable to those obtained by Orfanogiannaki and Papadopoulos (2007), but not to those for Kuril-Kamchatka and South America. In addition, Orfanogiannaki and Papadopoulos (2007) and Koravos *et al*., (2015) investigated the tsunami hazard for these areas by calculating the conditional probabilities of occurrence for specified tsunami intensities. These values are not comparable to the probabilities of exceedance calculated in the present study.

A question that immediately arises is what the difference will be between the results of the PTHA using the mixture distributions compared to the results provided by the same model when no uncertainty in the parameters of the occurrence models was taken into consideration. As a test, the data for Japan was applied to the defined mixture distributions (Poisson-gamma and exponential-gamma) as well as to the simple occurrence distributions of Poisson and exponential. Japan was chosen as it is the dataset with the most observations. The results are provided in Table 3 below. A significant increase can be seen in the estimated parameters and probabilities of exceedance as well as

15

significant decrease in the return periods. The estimated hazard therefore increased when uncertainty regarding the occurrence model was taken into consideration.

[TABLE 3]

**Table 3.** Comparison of the model for Japan with and without the application of mixture distributions to account for uncertainty in the occurrence parameters.

## 6.    Discussion and Conclusions

Over many years, considerable effort has been made to assess the tsunami hazard at vulnerable coastal areas; yet, significant uncertainty is still associated with the applied methodologies and research results. This uncertainty is attributable in part to the different generating mechanisms, ranging from near- and far-field earthquakes to submarine slumps and volcanoes. In addition, the bathymetry of the affected local coastline influences the intensity and height of the tsunami wave, either increasing or reducing the force of the incoming wave. Ultimately, the small number of events in any tsunami catalogue significantly increases the number of assumptions that have to be made in any tsunami hazard assessment process.

This paper presented a methodology to assess the coastline-characteristic tsunami hazard based on empirical catalogues. The methodology, originally defined in terms of probabilistic earthquake hazard assessment, is highly flexible, allowing for the transition to PTHA. It takes into account only the observed tsunami intensity and makes no assumptions about the source of the tsunami. The formalisms focus on the assessment of tsunami recurrence parameters, the mean tsunami activity rate $\lambda$ and the Soloviev-Imamura frequency-intensity power law $b$-value based on incomplete catalogues, the uncertainty in the tsunami intensity determination, as well as the uncertainty associated with the parameters of the applied tsunami occurrence models. The procedure explicitly allows for both aleatory and epistemic uncertainty by introducing mixture distributions where the recurrence parameters are assumed to be random variables that can fluctuate. For this purpose the gamma distribution was introduced as the mixing distribution in both the tsunami occurrence and intensity distribution models. The ability to view the recurrence parameters as random variables allows the user to define models that closer echo reality. As seen in Table 3, the exclusion of parameter uncertainty has a significant impact on the estimated hazard indicating that the simpler models may lead to the underestimation of the tsunami hazard for a region.

Although the methodology is described in terms of historical and complete catalogues, pre-historic tsunami information can be included in the analysis as well. For a more detailed discussion on the derivation of the parameters, refer to Kijko and Sellevoll (1992), Kijko and Smit (2012), and Kijko *et al*. (2016). Furthermore, the methodology allows for the application of any frequency-intensity power law describing the tsunami intensity (with the relevant adjustments to the intensity distributions).

It is safe to assume that the other parameters associated with tsunamis, such as the direction of the wave propagation and the local bathymetry, would affect the hazard estimates, resulting in adjusted site effects locally. The accuracy of the estimated recurrence parameters can be enhanced by incorporating these site effects by means of Bayesian statistics, similar to the procedure followed by Grezio *et al*. (2010, 2015), and Yadav *et al*. (2013). A posterior

16

probability is created by combining physical knowledge of the tsunami process, as a priori probability, with the likelihood function $L(\boldsymbol{\theta})$.

Several factors have to be taken into consideration when this proposed methodology is used. Firstly, the formalisms are dependent on the choice of level of completeness (LoC). This concern could potentially be addressed through the application of similar formalisms by Kijko and Smit (2017) who provide two new equations for the estimation of seismic recurrence parameters without using LoC. In addition, the uncertainty associated with the intensity error influences the derived recurrence parameters, as does the choice of uncertainty associated with the occurrence model. The choice of frequency-intensity distribution could also have an impact on the results since these formalisms do not take into account the local bathymetry conditions. Great care has to be taken to ensure that a model is presented that is as objective as possible and, in addition, can accommodate the various uncertainties without severely over- or underestimating the hazard.

Future research relating to the discussed methodology includes an investigation into the effect the inclusion of parameter uncertainty has on the final hazard estimates, the testing of the validity of the selected model compared to the observed data as well as other potential models with the use of the e,g. Akaike information criterion (AIC) as well as how an alternative method of $i_{max}$ estimation (soft cut-off method) will affect the estimated hazard for a region.

**Acknowledgements**

**References**

Abramowitz, M., & Stegun, I. A. (1970). Handbook of mathematical functions, 9th ed., Dover, New York.

Ambraseys, N. (1962). Data for the investigation of the seismic sea-waves in the Eastern Mediterranean, Bull. Seism. Soc. Am. Vol. 52, pp. 895-913

Anita, G., Tonini, R., Sandri, L., Pierdominici, S., & Selva, J. (2015). A Methodology for a comprehensive probabilistic tsunami hazard assessment: Multiple sources and short-term interactions. Journal of Marine Science and Engineering, 3(1), 23–51.

Annaka, T., Satake, K., Sakakiyama, T., Yanagisawa, K., & Shuto N. (2007). Logic-tree approach for probabilistic tsunami hazard analysis and its applications to the Japanese coasts. Pure. Appl. Geophys. 164, 577-592.

Benjamin, J. R. (1968). Probabilistic models for seismic forces design. J. Struct. Div., ASCE 94, (ST5), pp. 1175-1196.

Benjamin, J. R., & Cornell, C.A. (1970). *Probability, Statistics and Decision for Civil Engineers*, McGraw-Hill, New York.

Bird, P. & Kagan, Y.Y. (2004). Plate-tectonic analysis of shallow seismicity: Apparent boundary width, beta, corner magnitude, coupled lithosphere thickness, and coupling in seven tectonic settings. Bulletin of the Seismological Society of America, 94(6), pp.2380-2399.

Brizuela, B., Armigliato, A. & Tinti, S., (2014). Assessment of tsunami hazards for the Central American Pacific coast from southern Mexico to northern Peru. Natural Hazards and Earth System Sciences, 14(7), p.1889-1903.

Burbidge, D., Cummins, P. R., Mleczko R., & Thio, H. K. (2008). A probabilistic tsunami hazard assessment for Western Australia. Pure Appl. Geophys. 2008 Dec 1; 165(11-12):2059-88.

Burroughs, S. M., & Tebbens, S. F. (2001). Upper-truncated power laws in natural systems. Pure Appl. Geophys. 158, pp. 741–757.

Burroughs, S. M., & Tebbens, S. F. (2005). Power law scaling and probabilistic forecasting of tsunami run-up heights. Pure Appl. Geophys. 162, pp. 331–342.

Campbell, K. W. (1982). Bayesian analysis of extreme earthquake occurrences. Part I. Probabilistic hazard model. Bull. Seism. Soc. Am., 72, pp. 1689-1705.

Campbell, K. W. (1983). Bayesian analysis of extreme earthquake occurrences. Part II. Application to the San Jacinto Fault zone of southern California. Bull. Seism. Soc. Am., 73, pp. 1099-1115.

Cao, A. & Gao, S. S., (2002). Temporal variation of seismic b-values beneath northeastern Japan island arc. Geophys. Res. Lett., 29(9), pp. 48-1 – 48-3.

Cooke, P. (1979). Statistical inference for bounds of random variables, Biometrika 66, 2, 367-374.

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

Cosentino, P., Ficara, V., & Luzio, D. (1977). Truncated exponential frequency – magnitude relationship in the earthquake statistics. Bull. Seism. Am. 67, pp. 1615-1623.

Cramér, H. (1961). Mathematical Methods of Statistics, Princeton University Press, Princeton, New Jersey.

Davison, A. C. (2003). Statistical models, Cambridge. Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge.

Downes, G. L. & Stirling, M. W. (2001). Groundwork for development of a probabilistic tsunami hazard model for New Zealand, International Tsunami Symposium 2001, Seattle, Washington, pp. 293–301.

Dzierma. Y., & Wehrmann, H. (2010). Eruption time series statistically examined: Probabilities of future eruptions at Villarrica and Llaima Volcanoes, Southern Volcanic Zone, Chile. Journal of Volcanology and Geothermal Research, 193(1), 82-92.

Eadie, W. T., Drijard, D., James, F. E., Roos, M., & Sadoulet, B. (1971). Statistical methods in experimental physics. Second reprint 1982. North-Holland, Amsterdam.

Gardner, J. K. & Knopoff, L. (1974). Is the sequence of earthquakes in southern California, with aftershocks removed, Poissonian? Bull. Seism. Soc. Am. 64, pp. 1363-1368.

Geist, E.L. & Parsons, T. (2006). Probabilistic analysis of tsunami hazards. Natural Hazards, 37, pp. 277-314.

Geist, E. L. (2005). Local tsunami hazards in the Pacific Northwest from Cascadia subduction zone earthquakes. U.S. Geological Survey Professional Paper 1661-B, 17 pp.

González, F. I., Geist, E. L., Jaffe, B., Kânoğlu, U., Mofjeld, H., Synolakis, C. E., et al. (2009). Probabilistic tsunami hazard assessment at seaside, Oregon, for near- and far-field seismic sources. J Geophys Res Oceans 114(C11), pp. 1978–2012. doi:10.1029/2008JC005132

Grezio, A., Marzocchi, W., Sandri, L., & Gasparini, P. (2010). A Bayesian procedure for probabilistic tsunami hazard assessment. Nat. Haz., 53, pp. 159–174.

Grezio, A., Sandri, L., Marzocchi, W., Argnani, A., & Gasparini, P. (2012). Probabilistic tsunami hazard assessment for Messina Strait area (Sicily-Italy). Nat. Haz., 64, pp. 329–358.

Grilli, S. T., Harris, J., & Tajalli Bakhsh, T. (2011) Literature Review of Tsunami Sources Affecting Tsunami Hazard Along the US East Coast, NTHMP Progress report, Res. Rept. CACR-11-08, p.60 Center for Applied Coastal Research, University of Delaware, Newark, 2011.

Gusiakov, V. K. (2001). Basic Pacific tsunami catalogs and database, 47 BC–2000 AD: results of the first stage of the project. Proceedings of the International Tsunami Symposium, August 7–9, 2001, Seattle, USA, PMEL/NOAA, pp. 263– 272.

Gusiakov, V. K. (2005). Tsunami generation potential of different tsunamigenic regions in the Pacific. Marine Geology, 215(1), pp.3-9.

Gusiakov, V. K. (2009). Tsunami history: Recorded. In the sea; Bernard, E.N., Robinson, A.R., Eds.; Harvard University Press: Boston, MA, USA; pp. 1-15.

Gutenberg, B., & Richter, C. F. (1942). Earthquake magnitude, intensity, energy, and acceleration, Bull. Seismol. Soc. Am. 32, pp. 163–191.

Gutenberg, B., & Richter, C. F. (1956). Earthquake magnitude, intensity, energy, and acceleration: (Second paper), Bull. Seismol. Soc. Am. 46, pp. 105–145.

Harbitz, C. B., Glimsdal, S., Bazin, S., Zamora, N., Løvholt, F., Bungum, H., Smebye, H., Gauer, P. and Kjekstad, O., (2012). Tsunami hazard in the Caribbean: regional exposure derived from credible worst case scenarios. Continental Shelf Research, 38, pp.1-23.

Hébert, H., Schindelé, F., Altinok, Y., Alpar, B., & Gazioglu, C. (2005). Tsunami hazard in the Marmara Sea (Turkey): A numerical approach to discuss active faulting and impact on the Istanbul coastal areas, Mar. Geol., 215, pp. 23–43, doi:10.1016/j.margeo.2004.11.006.

Horikawa, K., & Shuto, N. (1983). Tsunami disasters and protection measures in Japan, In: K. Iida and T. Iwasaki (eds), Tsunamis — their science and engineering, Tokyo, Terra Scientific Publishing Company, pp. 9–22.

Houston, J. R., Carver, R. D., & Markle, D. G. (1977). Tsunami-wave elevation frequency of occurrence for the Hawaiian Islands. Technical Report H-77-16, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS, 66 pp.

Jones, G., Chester, D.K., & Shooshtarian, F. (1999). Statistical analysis of the frequency of eruptions at Furnas Volcano, Sao Miguel, Azores. J. Volc. Geotherm. Res., 92(1), pp. 31-38.

Kagan, Y.Y. (1991). Seismic moment distribution. Geophysical Journal International, 106(1), pp.123-134.

Kagan, Y.Y. (1997). Seismic moment-frequency relation for shallow earthquakes: Regional comparison. Journal of Geophysical Research: Solid Earth, 102(B2), pp.2835-2852.

Kagan, Y.Y. (2002a). Seismic moment distribution revisited: I. Statistical results. Geophysical Journal International, 148(3), pp.520-541.

Kagan, Y.Y. (2002b). Seismic moment distribution revisited: II. Moment conservation principle. Geophysical Journal International, 149(3), pp.731-754.

Kagan, Y.Y. (2010). Statistical distributions of earthquake numbers: consequence of branching process. Geophysical Journal International, 180(3), pp.1313-1328.

Kagan, Y.Y., & Schoenberg, F. (2001). Estimation of the upper cutoff parameter for the tapered Pareto distribution. Journal of Applied Probability, 38(A), pp.158-175.

Karakaisis, G. F., Papazachos, C. B., Scordilis, E.M., Papaioannou, Ch. A., & Papazachos, B. C. (2002). Time variation of seismicity and seismic hazard in Hellenic Arc trench system. XXXVIII General Assembly of ESC, 1-6 September 2002, Genoa, Italy, pp. 1-6.

Kijko, A. (2004). Estimation of the maximum earthquake magnitude mmax, Pure Appl. Geophys, 161, pp. 1-27.

Kijko, A. (2012). On Bayesian procedure for maximum earthquake magnitude estimation. Research in Geophysics, 2, Number 1. pp. 46-51. http://www.pagepress.org /journals/index.php/rg.

Kijko, A., & Sellevoll, M.A. (1992). Estimation of earthquake hazard parameters from incomplete data files, Part II, Incorporation of magnitude heterogeneity. Bull. Seism. Soc. Am. 82, pp. 120-134.

Kijko, A., & Singh, M. (2011). Statistical tools for maximum possible earthquake magnitude estimation. Acta Geophysica, 59, pp. 674-700

Kijko, A., & Smit, A. (2012). Extension of the Aki-Utsu b-value estimator for incomplete catalogs. Bull. Seism. Soc. Am, June 2012, Vol 102, 3. pp. 1283-1287.

Kijko, A., & Smit A. (2017) Estimation of the frequency-magnitude Gutenberg-Richter b-value without making assumptions on the level of completeness. Seismological Research Letters. Online publication 25 January 2017, pp.1-8. doi: 10.1785/0220160177

Kijko, A., Smit, A., & Sellevoll, M. A. (2016). Estimation of earthquake hazard parameters from incomplete data Files. Part III. Incorporation of uncertainty of earthquake-occurrence model. Bull. Seism. Soc. Am., 106(3), pp1210-1222. doi: 10.1785/0120150252

Koravos, G. C., Yadav, R. B. S. & Tsapanos, T. M., (2015). Evaluation of tsunami potential based on conditional probability for specific zones of the Pacific tsunamigenic rim. Tectonophysics, 658, pp.159-168.

Kramer, S. L. (1996). Geotechnical earthquake engineering. Pearson Education India; Sep 1.

Lin, I. -C., & Tung, C. C. (1982). A preliminary investigation of tsunami hazard, Bull. Seism. Soc. Am., 72, pp. 2323–2337.

Liu, Y., Santos, A., Wang, S. M., Shi, Y., Liu, H., & Yuen, D. A. (2007). Tsunami hazards along Chinese coast from potential earthquakes in South China Sea, Phys. Earth Planet. Int. 163, pp. 233–244.

Lockridge, P. A., & Smith, R. H., (1984). Tsunamis in the Pacific Basin, 1900–1983 (map). Boulder, CO, NOAA/NGDC.

Lomnitz, C. (1973). Poisson processes in earthquake studies. Bull. Seism. Soc. Am. 63, pp. 735-758.

Lorito, S., Tiberti, M. M., Basili, R., Piatanesi, A., & Valensise, G. (2008). Earthquake-generated tsunamis in the Mediterranean Sea: Scenarios of potential threats to Southern Italy, J. Geophys. Res., 113, B01301, pp1-14. doi:10.1029/2007JB004943.

Main, I. (1996). Statistical physics, seismogenesis, and seismic hazard. Reviews of Geophysics, 34(4), pp.433-462.

McGuire, R. K. (1976). Fortran computer program for seismic risk analysis. USGS. Open File Rep. 76-67.

Mora, P. R. (2001). Pageoph topical volumes microscopic and macroscopic simulation: Towards predictive modelling of the Earthquake process.

Mulargia, F., & Tinti, S. (1985). Seismic sample areas defined from incomplete catalogs: An application to the Italian territory. Physics of the Earth and Planetary Interiors, 40 (4), pp. 273–300.

Ogata, Y., & Abe, K. (1991). Some statistical features of the long-term variation of the global and regional seismic activity. Int. Stat. Rev. 59, pp. 139-161.

Orfanogiannaki, K. & Papadopoulos, G. A., (2007). Conditional probability approach of the assessment of tsunami potential: Application in three tsunamigenic regions of the Pacific Ocean. Pure Appl. Geophys.,164 (2-3), pp.593-603.

Page, R. (1968). Aftershocks and microaftershocks. Bull. Seism. Soc. Am., 58, pp. 1131-1168.

Papadopoulos, G. A., & Imamura, F. (2001). A proposal for a new tsunami intensity scale. Proc. Int. Tsunami Symposium 2001. pp. 569–577.

Paulatto, M., Pinat, T., & Romanelli, F. (2007). Tsunami hazard scenarios in the Adriatic Sea domain, Nat. Hazards Earth Syst. Sci., 7, pp. 309–325, doi:10.5194/nhess-7-309-2007.

Petersen, M. D., Cramer, C. H., & Frankel, A. D. (2002). Simulations of seismic hazard for the Pacific Northwest of the United States from earthquakes associated with the Cascadia subduction zone. In Earthquake Processes: Physical Modelling, Numerical Simulation and Data Analysis Part I (pp. 2147–2168). Basel: Birkhäuser.

Pisarenko, V.F. & Sornette, D. (2003). Characterization of the frequency of extreme earthquake events by the generalized Pareto distribution. Pure and Applied Geophysics, 160(12), pp.2343-2364.

Pisarenko, V.F., Sornette, A., Sornette, D. & Rodkin, M.V. (2008). New approach to the characterization of M max and of the tail of the distribution of earthquake magnitudes. Pure and Applied Geophysics, 165(5), pp.847-888.

Pisarenko, V.F., Sornette, A., Sornette, D. & Rodkin, M.V. (2014). Characterization of the tail of the distribution of earthquake magnitudes by combining the GEV and GPD descriptions of extreme value theory. Pure and Applied Geophysics, 171(8), pp.1599-1624.

Power, W., Downes, G., & Stirling, M. (2007). Estimation of tsunami hazard in New Zealand due to South American earthquakes, Pure Appl. Geophys., 164, pp. 547–564, doi:10.1007/s00024-006-0166-3.

Rao, C. R. (1973). Linear statistical inference and its application, Edition 2. John Willey and Sons, New York, pp.625.

Rikitake, T. & Aida, I. (1988). Tsunami hazard probability in Japan, Bull. Seismol. Soc. Am. 78, pp. 1268–1278.

Shaw, B., Ambraseys, N. N., England, P. C., Floyd, M. A., Gorman, G. J., Higham, T. F. G., Jackson, J. A., Nocquet, J.-M., Pain, C. C., & Piggott, M. D. (2008). Eastern Mediterranean tectonics and tsunami hazard inferred from the AD 365 earthquake, Nat. Geosci., 1, pp. 268–276, doi:10.1038/ngeo151

Sieberg, A. (1927). Geologische, physikalische und angewandte Erdbebenkund. Verslag von Gustav Fisher, Jena.

Soloviev, S. L. (1970). Recurrence of tsunamis in the Pacific. In Tsunamis in the Pacific Ocean, edited by W.M. Adams, East-West Centre Press, Honolulu, pp. 149-163

Sørensen, M. B., Spada, M., Babeyko, A., Wiemer, S., & Grünthal G. (2012). Probabilistic tsunami hazard in the Mediterranean Sea. J. Geophys. Res., 117, B01305, pp. 1-15. doi:10.1029/2010JB008169

Sornette, D. & Sornette, A. (1999). General theory of the modified Gutenberg-Richter law for large seismic moments. Bulletin of the Seismological Society of America, 89(4), pp.1121-1130.

Talbi, A., & Yamazaki, F. (2009). A mixed model for earthquake inter-event times. J. Seismol. 10, pp. 289-307.

Tatsumi, D., Calder, C.A., & Tomita, T. (2014). Bayesian near-field tsunami forecasting with uncertainty estimates. J. Geophys. Res. Oceans, 119, pp. 2201–2211.

Thio, H. K., Somerville, P., & Ichinose, G. (2007). Probabilistic analysis of strong ground motion and tsunami hazard in Southeast Asia, J. Earthquake Tsunami, 1(2), pp. 119–137, doi:10.1142/S1793431107000080.

Tinti, S. (1993). Evaluation of tsunami hazard in Calabria and eastern Sicily, Italy, Tsunamis in the world. In (ed. Tinti, S.) Advances in Natural and Technological Hazards Research (Kluwer Academic Publishers, Dordrecht), pp. 141–157.

Tinti, S., & Armigliato, A. (2003). The use of scenarios to evaluate tsunami impact in southern Italy, Mar. Geol., 199, pp. 221–243, doi:10.1016/ S0025-3227(03)00192-0.

Tinti, S., Armigliato, A., Tonini, R., Maramai, A., & Graziani, L. (2005). Assessing the hazard related to tsunamis of tectonic origin: a hybrid statistical-deterministic method applied to southern Italy coasts. ISET J Earthquake Tech, 42, pp.189-201.

Tinti, S., & Maramai, A. (1996). Catalogue of tsunamis generated in Italy and in Cote d' Azur, France: a step towards a unified catalogue of tsunamis in Europe. Annali di Geofisica 39, pp. 1253–1299.

Tinti, S., Maramai, A., & Graziani, L. (2001). A new version of the European tsunami catalogue: Updating and revision. Natural Hazards and Earth System Science 1, pp. 255–262.

Tinti, S., & Mulargia, F. (1985). Effects of magnitude uncertainties in the Gutenberg-Richter frequency-magnitude law. Bull. Seism. Soc. Am. 75, pp. 1681-1697.

Tonini, R., Armigliato, A., Pagnoni, G., Zaniboni, F. & Tinti, S., (2011). Tsunami hazard for the city of Catania, eastern Sicily, Italy, assessed by means of Worst-case Credible Tsunami Scenario Analysis (WCTSA). Natural Hazards and Earth System Sciences, 11(5), p.1217-1232.

Urlaub, M., Talling, P. J., & Masson, D. G. (2013). Timing and frequency of large submarine landslides: Implications for understanding triggers and future geohazard. Quaternary Science Reviews, 72,63-82

Vere-Jones, D., Robinson, R., & Yang, W. (2001). Remarks on the accelerated moment release model: problems of model formulation, simulation and estimation. Geophysical Journal International, 144(3), pp.517-531.

Ward, S. N. (2001). Landslide tsunami, J. Geophys. Res. 106, pp. 11201–11215.

Ward, S. N. (2003). Tsunamis, In: R. A. Meyers (ed.), The Encyclopedia of physical science and technology, Academic Press, pp. 175–191.

Ward, S. N. & Asphaug, E. (2000). Asteroid impact tsunami: a probabilistic hazard assessment, Icarus 145, 64–78.

Wijetunge, J. J., (2014). A deterministic analysis of tsunami hazard and risk for the southwest coast of Sri Lanka. Continental Shelf Research, 79, pp.23-35

Yadav, R. B. S., Tsapanos, T. M., Tripathi, J. N. & Chopra, S. (2013). An evaluation of tsunami hazard using Bayesian approach in the Indian Ocean. Tectonophysics, 593, pp. 172-182.

**Tables and Figures**

**Table 1.** Probabilistic tsunami hazard assessment input parameters and estimated recurrence parameters. The value $m_{min}$ represents the level of completeness (LoC) and SE represents the standard error (assumed uncertainty) in the intensity estimation.

| | | Japan (JAP) | Kuril-Kamchatka (K-K) | South America (SAM) |
|---|---|---|---|---|
| **Catalogue information** | | | | |
| Historical | Period | 684 AD to 1960 | 1737–1959 | 1513–1959 |
| | Number of events | 79 | 6 | 8 |
| | LoC ($m_{min}^H$) | 1.00 | 3.0 | 3.50 |
| | Intensity SE | 0.5 | 0.5 | 0.5 |
| Complete | Period | 1961–2011 | 1960–2009 | 1960–2015 |
| | Number of events | 40 | 21 | 8 |
| | LoC ($m_{min}^C$) | −2.00 | −0.5 | 2.0 |
| | Intensity SE | 0.1 | 0.1 | 0.1 |
| Observed maximum intensity | | $4.2 \pm 0.1$ | $4.0 \pm 0.5$ | $4.1 \pm 0.1$ |
| **Estimated recurrence parameters** | | | | |
| Mean annual rate of activity ($\lambda$) | | $1.5 \pm 0.4$ | $0.3 \pm 0.1$ | $0.2 \pm 0.1$ |
| Frequency-magnitude $b$-value | | $0.4 \pm 0.04$ | $0.2 \pm 0.1$ | $0.5 \pm 0.2$ |
| Coastline-characteristic maximum intensity $i_{max}$ | | $4.3 \pm 0.2$ | $4.2 \pm 0.5$ | $4.2 \pm 0.2$ |

**Table 2.** Return period and probabilities of exceedance for Japan, Kuril-Kamchatka, and South America for intensities larger than 1.5, 2.0, and 2.5 and for time periods 1, 10, and 25 years.

| Tsunami Intensity | Recurrence variable | JAP | K-K | SAM |
|---|---|---|---|---|
| $I \geq 1.5$ | Return Period (yrs) | 15 | 8 | - |
| | Prob. Exceedance (%): 1 year | 7 | 11 | - |
| | Prob. Exceedance (%): 10 years | 49 | 69 | - |
| | Prob. Exceedance (%): 25 years | 80 | 94 | - |
| $I \geq 2.0$ | Return Period (yrs) | 23 | 12 | 6 |
| | Prob. Exceedance (%): 1 year | 4 | 8 | 15 |
| | Prob. Exceedance (%): 10 years | 35 | 57 | 79 |
| | Prob. Exceedance (%): 25 years | 65 | 87 | 97 |
| $I \geq 2.5$ | Return Period (yrs) | 37 | 17 | 12 |
| | Prob. Exceedance (%): 1 year | 3 | 6 | 8 |
| | Prob. Exceedance (%): 10 years | 24 | 43 | 55 |
| | Prob. Exceedance (%): 25 years | 48 | 75 | 86 |

**Table 3**. Comparison of the model for Japan with and without the application of mixture distributions to account for uncertainty in the occurrence parameters.

| Estimates | With mixture distributions (WMD) | Without mixture distribution (WOMD) | Percentage increase (WMD-WOMD)/WOMD |
|---|---|---|---|
| Mean annual rate of activity ($\lambda$) | $1.5 \pm 0.4$ | $0.9 \pm 0.1$ | 64% |
| Frequency-magnitude $b$-value | $0.4 \pm 0.04$ | $0.34 \pm 0.02$ | 18% |
| Coastline-characteristic maximum intensity $i_{max}$ | $4.35 \pm 0.17$ | $4.35 \pm 0.18$ | 0% |
| Return Periods | | | |
| $I \geq 1.5$ | 15 yrs | 19 yrs | -24% |
| $I \geq 2.0$ | 23 yrs | 30 yrs | -25% |
| $I \geq 2.5$ | 37 yrs | 50 yrs | -25% |
| Probability of Exceedance (1 year) | | | |
| $I \geq 1.5$ | 7% | 5% | 24% |
| $I \geq 2.0$ | 4% | 3% | 24% |
| $I \geq 2.5$ | 3% | 2% | 25% |
| Probability of Exceedance (10 years) | | | |
| $I \geq 1.5$ | 49% | 40% | 17% |
| $I \geq 2.0$ | 35% | 28% | 20% |
| $I \geq 2.5$ | 24% | 18% | 22% |
| Probability of Exceedance (25 years) | | | |
| $I \geq 1.5$ | 80% | 73% | 10% |
| $I \geq 2.0$ | 65% | 56% | 14% |
| $I \geq 2.5$ | 48% | 40% | 18% |
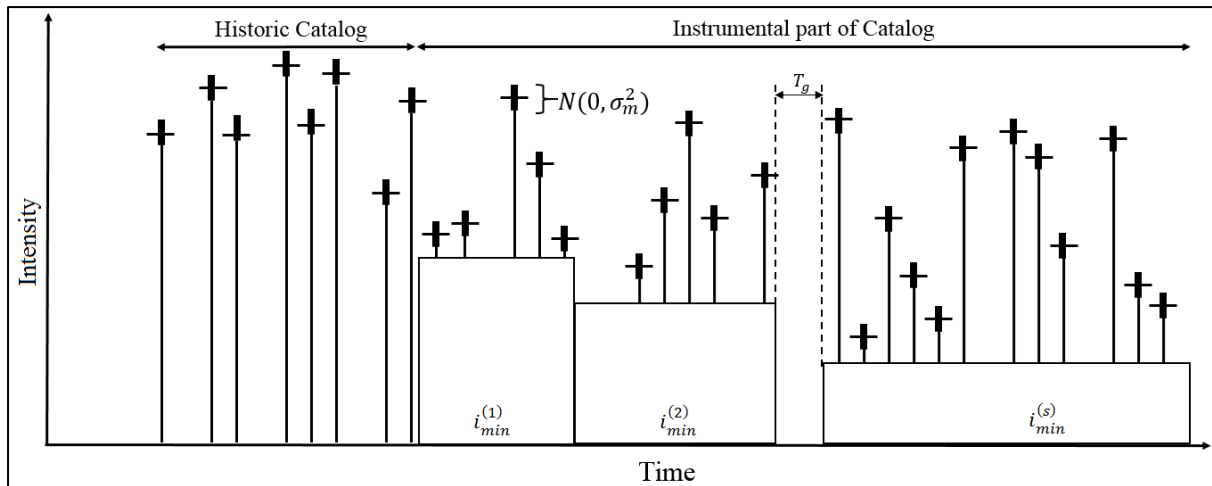
**Fig.1**. Illustration of data that can be used to obtain recurrence parameters of the tsunami. The applied approach allows combining the largest (historic tsunamis), and complete data having a variable level of completeness. This approach accepts 'gaps' ($T_g$) where records are missing. (Modified after Kijko and Sellevoll, 1992)
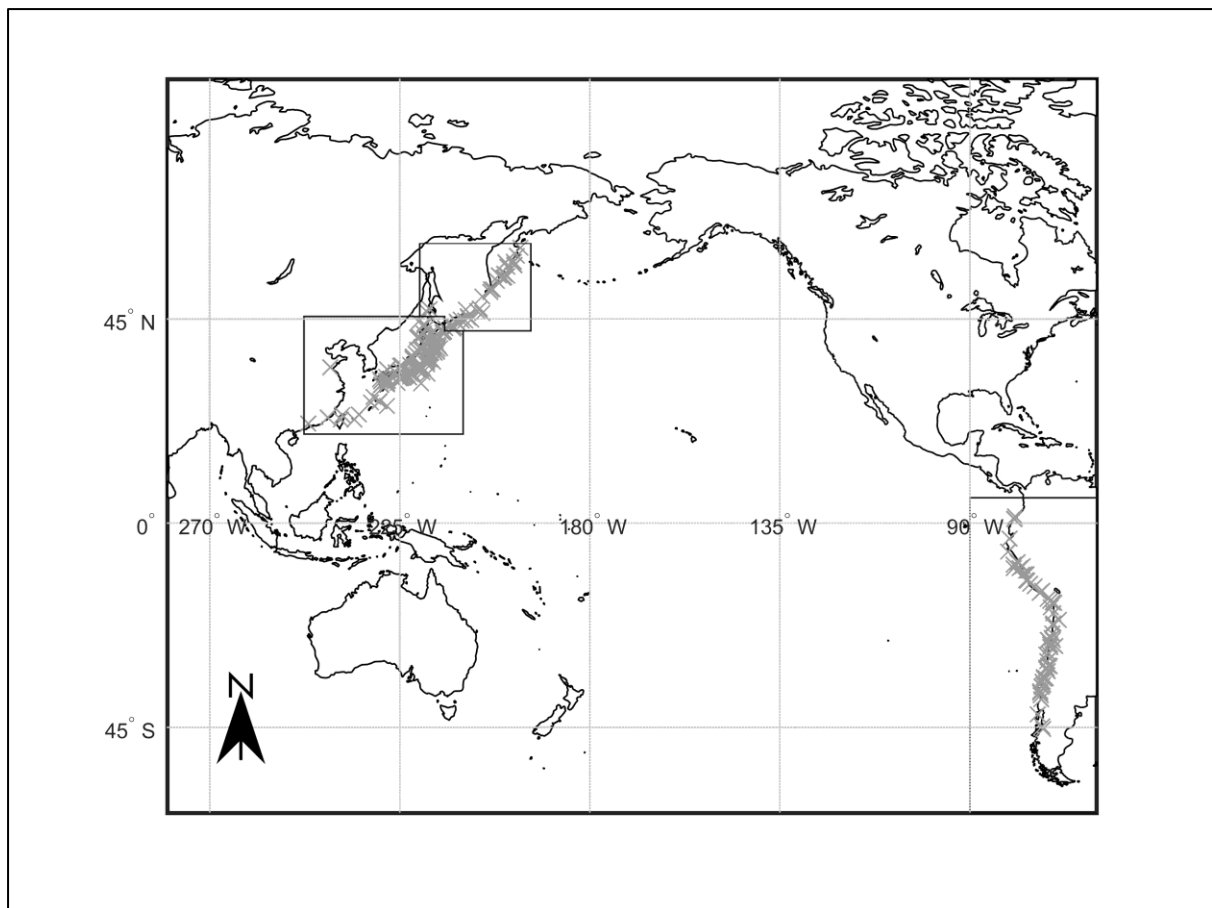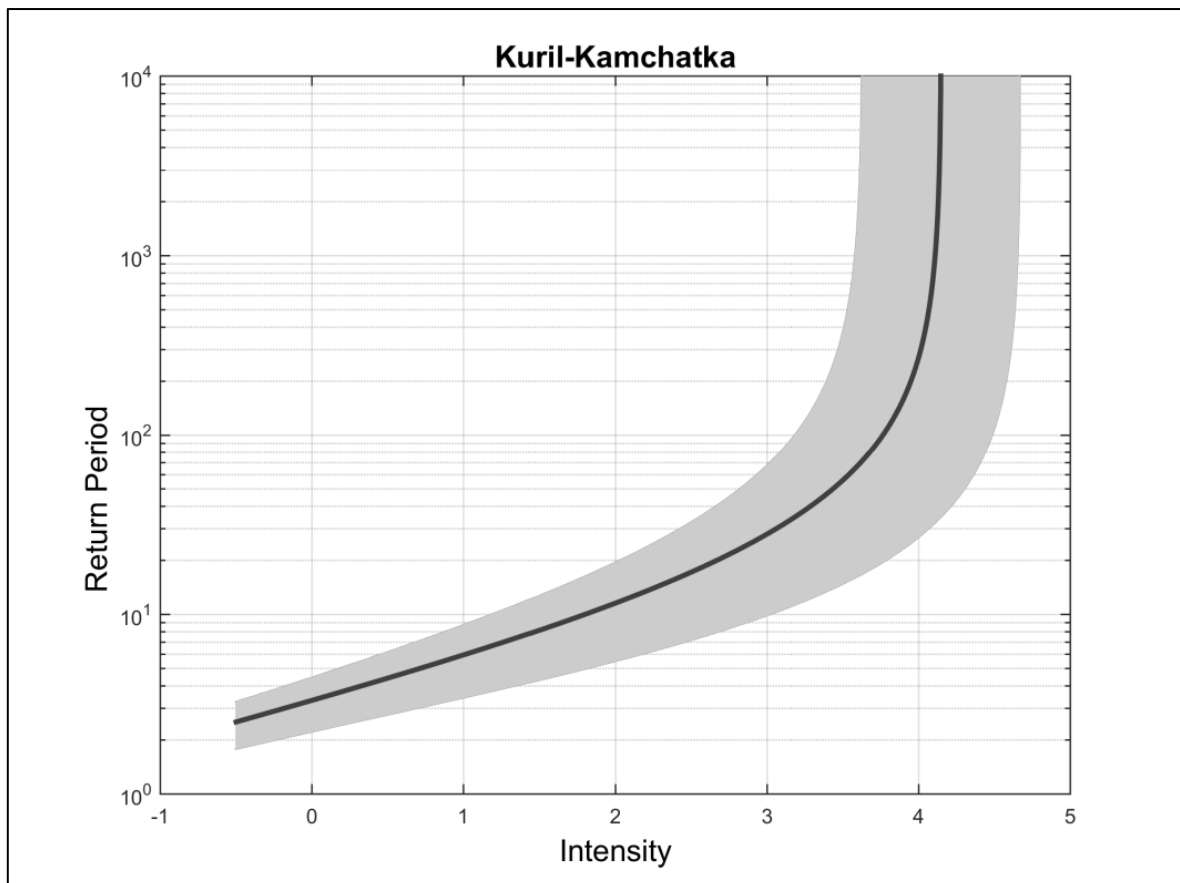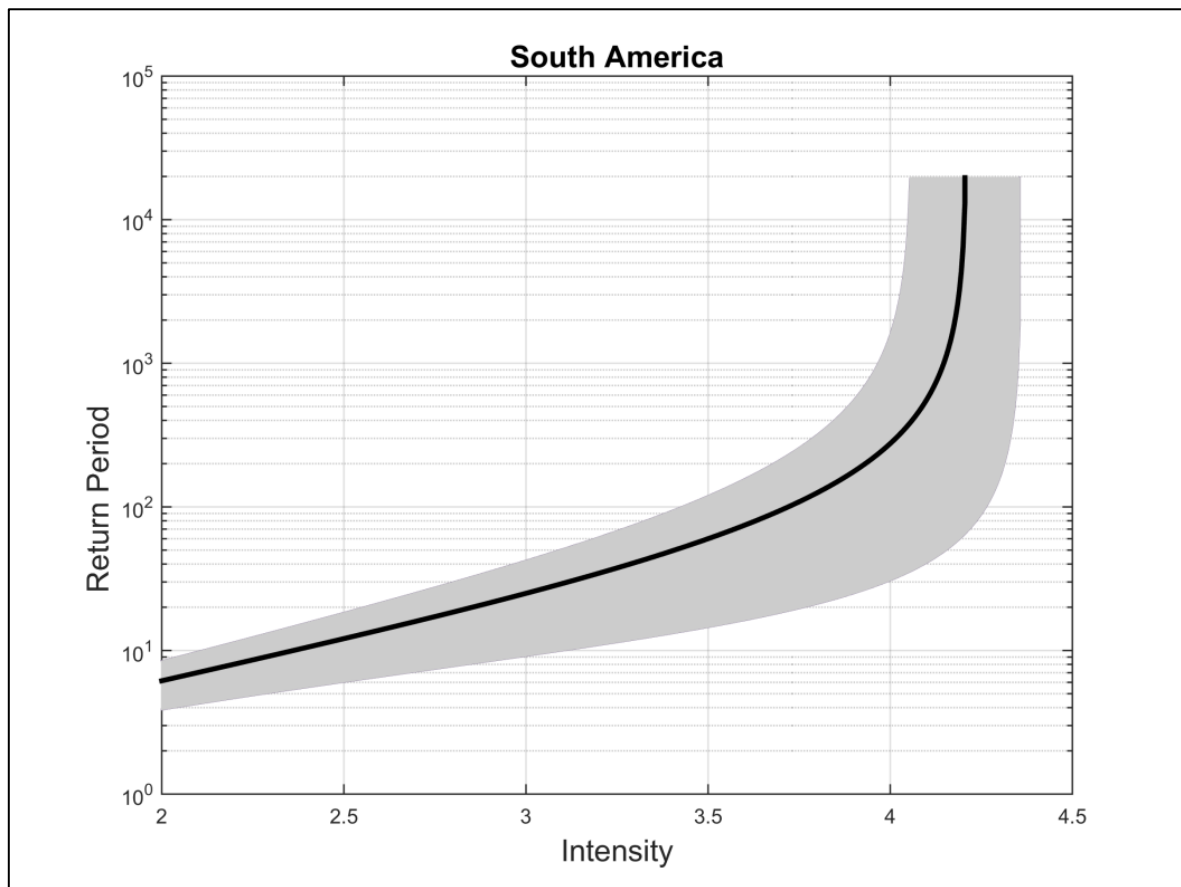


**Fig. 2.** Positions of the three investigated tsunamigenic regions in the Pacific Ocean (modified after Gusiakov, 2005), being the source locations of the tsunamis for Japan (JAP), Kuril-Kamchatka (K-K), and South America (SAM).

(a)

(b)



(c)

**Fig. 3**. The annual mean return period (in years) and its associated uncertainty for various tsunami intensities for a) Japan, b) Kuril-Kamchatka, and c) South America. The shaded area represents the relevant one standard deviation confidence interval for the calculated values.
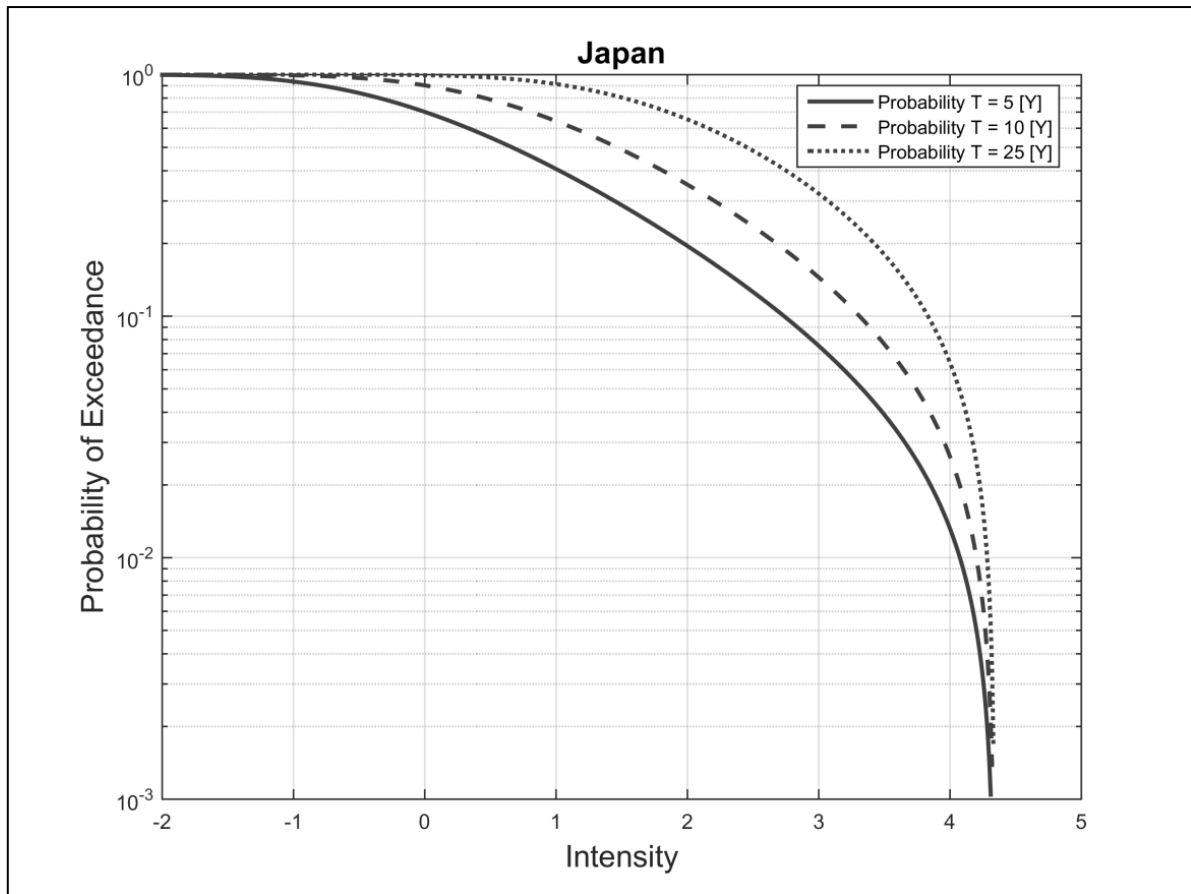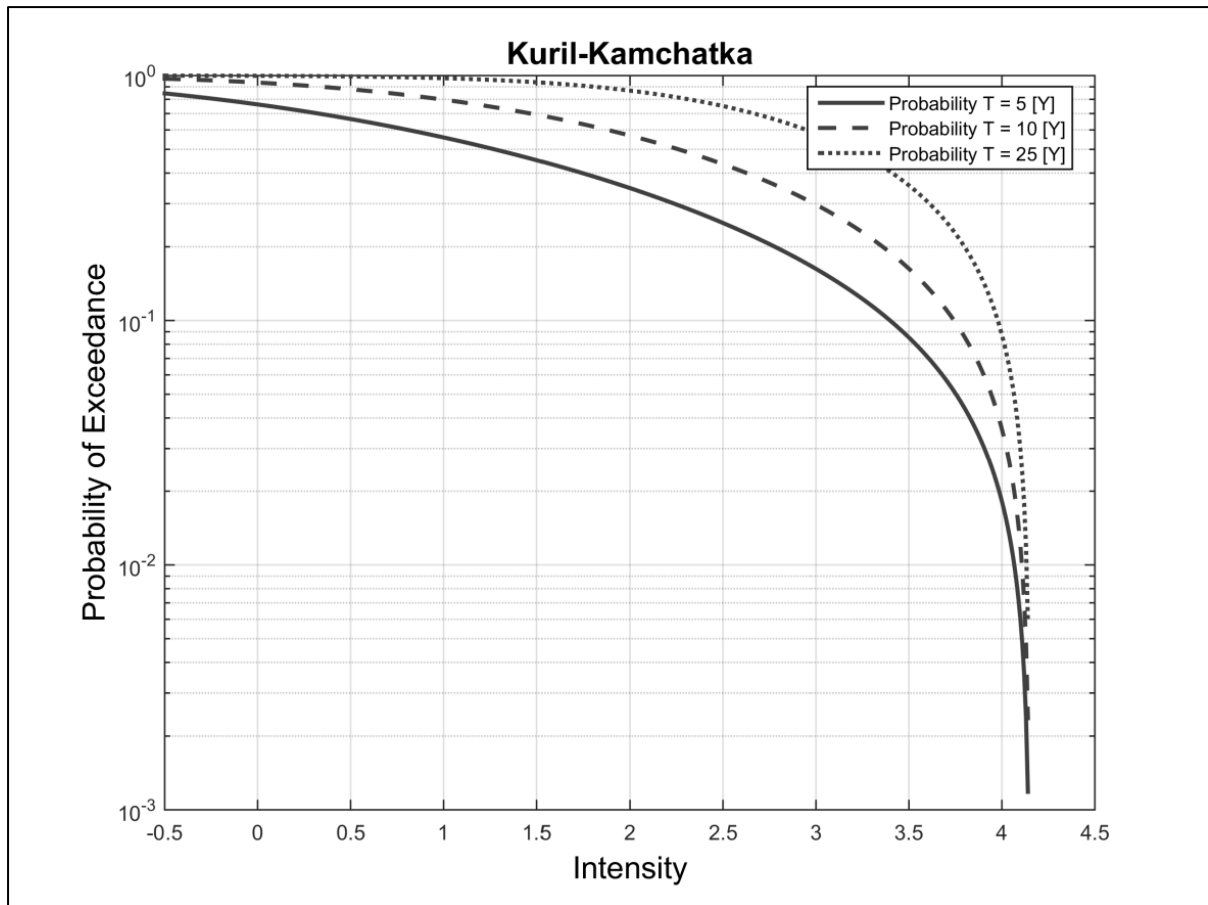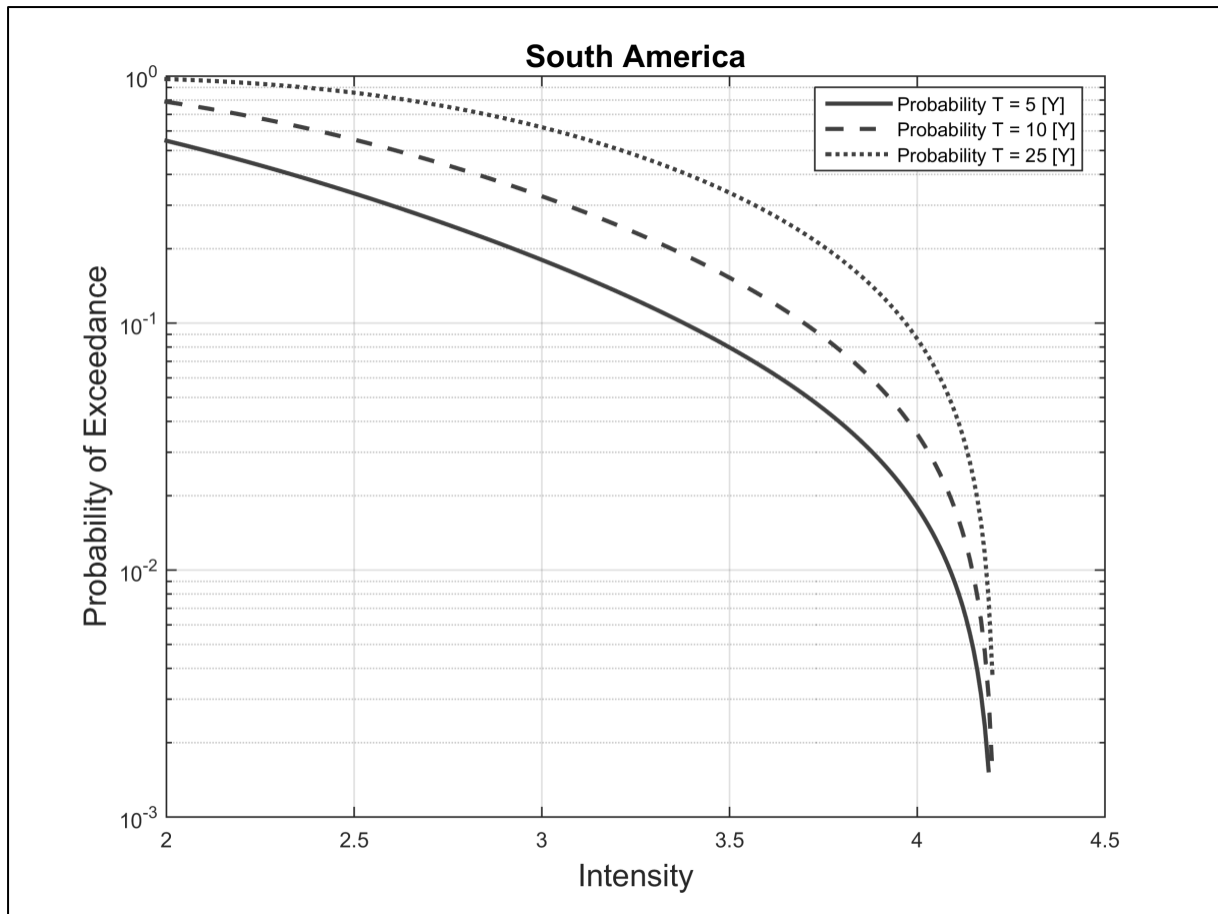
(a)

(b)

(c)

**Fig. 4**. The annual probability of exceedance and its associated uncertainty for various tsunami intensities for a) Japan, b) Kuril-Kamchatka, and c) South America. The shaded area represents the relevant one standard deviation confidence interval for the calculated values.

(a)

(b)

(c)

**Fig. 5**. The probability of exceedance for 5, 10, and 25 years for various tsunami intensities for a) Japan, b) Kuril-Kamchatka, and c) South America