A Bayesian procedure for Probabilistic Tsunami Hazard Assessment

Anita Grezio · Warner Marzocchi · Laura Sandri · Paolo Gasparini

Abstract In this paper, a Bayesian procedure is implemented for the Probability Tsunami Hazard Assessment (PTHA). The approach is general and modular incorporating all significant information relevant for the hazard assessment, such as theoretical and empirical background, analytical or numerical models, instrumental and historical data. The procedure provides the *posterior* probability distribution that integrates the prior probability distribution based on the physical knowledge of the process and the likelihood based on the historical data. Also, the method deals with aleatory and epistemic uncertainties incorporating in a formal way all sources of relevant uncertainty, from the tsunami generation process to the wave propagation and impact on the coasts. The modular structure of the procedure is flexible and easy to modify and/or update as long as new models and/or information are available. Finally, the procedure is applied to an hypothetical region, *Neverland*, to clarify the PTHA evaluation in a realistic case.

1 Introduction

The enormous death toll of the recent Sumatra tsunami 2004 brought out suddenly the necessity to establish (a priori) practical measures devoted to mitigate the risk caused by

A. Grezio · P. Gasparini
AMRA Scarl, Via Nuova Agnano 12, 80012 Naples, Italy

A. Grezio (✉) · L. Sandri
Istituto Nazionale di Geofisica e Vulcanologia, Via Creti 12, 40128 Bologna, Italy
e-mail: grezio@bo.ingv.it

W. Marzocchi
Istituto Nazionale di Geofisica e Vulcanologia, Via di Vigna Murata 605, 00143 Rome, Italy

P. Gasparini
Department of Physical Sciences, University of Naples, "Federico II", Naples, Italy

such calamitous events. The primary scientific component to achieve this goal is the quantification of the *tsunami hazard*. In English, hazard has the generic meaning "potential source of danger". Yet, for more than 30 years, hazard has been also used in seismology and volcanology in a more quantitative way, that is, as "the probability of a certain hazardous event in a specific time–space window" (Fournier d'Albe 1979). For this reason, and in agreement with previous studies (Geist and Parsons 2006; Annaka et al. 2007; Liu et al. 2007) and with seismic (Senior Seismic Hazard Analysis Committee (SSHAC) 1997) and volcanic hazard analysis (Marzocchi et al. 2007), we assert that the quantitative tsunami hazard estimation is the "Probabilistic Tsunami Hazard Assessment" (PTHA). The probabilistic nature of the tsunami hazard comes out from the complexity, nonlinearities, limited knowledge, and large number of degrees of freedom involved in the processes describing the arrival of tsunami waves on a coast. All of these components make a deterministic prediction impossible (Geist and Parsons 2006).

After the Sumatra–Andaman tsunami, many efforts have been made to achieve a sound strategy for PTHA. All attempts proposed so far can be gathered in two broad categories. In one category, PTHA is estimated by using tsunami catalogues (Burroughs and Tebbens 2005; Tinti et al. 2005; Orfanogiannaki and Papadopoulos 2007), while in the other case, different "scenario-based" PTHAs are suggested (Geist and Parsons 2006; Farreras et al. 2007; Liu et al. 2007; Power et al. 2007; Yanagisawa et al. 2007, Burbidge et al.2008). The difference between them has important implications. In the first case, a real PTHA is achieved, but the few available data make it applicable only on sporadic specific sites, and the few information cannot give the whole range of possible phenomena (in probabilistic terms, the sample can hardly be considered a representative sample of the real distribution). On the other hand, the use of deterministic "scenario-based" models does not provide directly a reliable PTHA. As a matter of fact, PTHA requires the simultaneous evaluation and integration of different scenarios, each one of them weighted with its probability of occurrence. This is not usually done because of two main reasons: (1) often, it is not easy to assign the probability of occurrence to each scenario; (2) scenarios are usually created with sophisticated codes that require an enormous demand of CPU time, and therefore they do not allow the whole range of possible source uncertainties to be taken into account.

The great importance of PTHA is due to its practical implications for society. In this perspective, it is fundamental that PTHA is *accurate* (i.e. without significant biases), because a biased estimation would be misleading. On the other hand, PTHA may have a low *precision* (i.e. a large uncertainty) that could reflect our scarce knowledge of some physical processes involved, from the source of the tsunami to its impact on the coastline. An accurate PTHA can be realistically achieved only incorporating all sources of uncertainty. Nonetheless, many efforts on PTHA have still a clear deterministic nature (Sato et al. 2003; Farreras et al. 2007; Power et al. 2007; Yanagisawa et al. 2007; Horsburgh et al. 2008) focusing more on describing in detail some aspects of the whole tsunami process instead of incorporating all the sources of uncertainty. It is argued that the use of sophisticated deterministic models certainly increases the precision, but, if all uncertainties are not taken into account, they may introduce a significant bias making the estimation highly inaccurate.

Here, we propose a formal probabilistic framework for PTHA based on Bayesian inference. The method starts from a physical knowledge of the process (prior distribution) that usually derives from theoretical or empirical models, and then it shapes the final (*posterior*) probability distribution of the tsunami event accounting for the available data. Notably, the method has also the paramount feature to deal with aleatory and epistemic

uncertainties in a proper way. In order to make the structure more flexible and easy to update and implement, the procedure has been developed through a modular structure.

The next section of the paper describes the Bayesian methodology applied to the tsunami probabilistic assessment. Then, the practical implementation of the PTHA is delineated through a modular scheme. Finally, the strategy is applied to a hypothetical region called *Neverland*.

2 Bayesian methodology for PTHA

The PTHA proposed in this paper is based on the Bayesian inference applied to the probability Θ. Bayesian inference is the process of fitting a probability model to a set of data y, and summarizing the result by a probability distribution of the parameters of the model. Here, the parameter of the model is Θ, which represents the probability of a destructive event per unit time τ. In mathematical terms, a destructive event can be defined as the overcoming of a specific threshold z_t of a selected physical parameter Z such as wave height, run-up, energy fluxes, wave propagation in land, etc. The probability Θ does not vary in the time interval τ with important implications in the PTHA. The variable Θ may be obtained by stationary or nonstationary models as well, but the probability is assumed to be constant in the forecasting time interval τ. This is not a problem for the stationary models or for the time-dependent models that have a characteristic time much larger than τ. On the contrary, if τ is comparable or larger than the characteristic time of the time-dependent model, it may be recommended to run the code frequently inside the time interval τ in order to take into account all events occurring in τ. For example, the daily ($\tau = 24$ h) earthquake forecast in California (Gerstenberger et al. 2005) is updated every 30 min because in that case the time-dependent model may have large fluctuations if some high magnitude earthquake occurs in the forecasting time interval τ. Nonetheless, the selection of the time interval τ and the frequency of the "model updating" strongly depend on the final use of the PTHA. For instance, it is not appropriate for updating the PTHA every 30 min, if the assessment is used for land use planning (i.e. τ of some decades as for seismic hazard).

Hereinafter, we use capital letters to indicate generic random variables, while a specific value of each variable is not. In order to make probability statements about Θ given Y, we must use a probabilistic model to link the variables, i.e. the Bayes' rule.

The Bayes' rule states that:

$$[\Theta|Y] = \frac{[\Theta, Y]}{[Y]} = \frac{[\Theta][Y|\Theta]}{[Y]} \tag{1}$$

where $[\cdot]$ is a generic Probability Density Function (PDF), and $[\cdot|\cdot]$ is a generic conditional probability density. Regarding the terms in Eq. 1, $[\Theta]$ is the prior distribution containing all our knowledge on the parameter based on models and/or theoretical beliefs; $[Y | \Theta]$ is the sampling distribution (the so-called likelihood function), that is the PDF of observing the data y given a specific value of Θ; $[\Theta | Y]$ is the posterior distribution given a generic set of data Y; $[Y]$ is a normalization factor accounting for the total probability of observing the data y, i.e. $\int_\theta [\Theta][Y|\Theta]$.

Bayes' rule is widely and commonly used in many classical and Bayesian probability calculations. Compared to the most usual applications where the probabilities of Eq. 1 are single values, here the statistical distributions of the probabilities are used in order to

account for the uncertainties on the probability estimates (epistemic and aleatory uncertainties, Marzocchi et al. 2004, 2008). The core of the Bayesian inference is represented by the choice of the functional forms of the prior distribution and the likelihood function, that require some physical and statistical assumptions on the process. In the next subsections, the basic components of Bayesian inference are described and embedded in a quantitative structure suitable for PTHA.

2.1 The prior model

The prior model has to summarize the knowledge of the process without considering possible observations (that enter in the likelihood distribution). For PTHA, the prior model can be quantified through numerical/mathematical models.

The prior distribution for Θ is

$$\Theta \sim Beta(\alpha, \beta) \tag{2}$$

that is a *Beta* distribution with parameters α and β, whose density function is

$$[\Theta] = \frac{1}{B(\alpha, \beta)} \Theta^{\alpha-1}(1 - \Theta)^{\beta-1}, \quad 0 < \Theta < 1, \ \alpha > 0, \ \beta > 0, \tag{3}$$

where $B(\alpha, \beta)$ denotes the *Beta* function

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1 - x)^{\beta-1} \mathrm{d}x. \tag{4}$$

The expected value for the *Beta* distribution is

$$E(\Theta) = \frac{\alpha}{\alpha + \beta} \tag{5}$$

and the variance is

$$V(\Theta) = \frac{E(\Theta)(1 - E(\Theta))}{\alpha + \beta + 1}. \tag{6}$$

We chose the *Beta* distribution because it is an unimodal distribution of a random variable defined in the range [0, 1], and it is the conjugate prior distribution in the binomial model (Gelman et al. 1995). This choice is subjective and other distributions can be adopted, such as the Gauss distribution for the logit transformation of the probability. In practice, the differences associated with the use of reasonable distributions are not usually significant. For this reason, the *Beta* distribution is the most used in practical applications (Savage 1994; Gelman et al. 1995).

For example, in the present tsunami study, the average of the *Beta* distribution (Eq. 5) is set equal to the weighted percentage of simulated tsunamigenic sources producing $Z > z_t$ in a set of computed run-ups. These weights are the probability of occurrence of each tsunamigenic source. In practice, the parameters α and β of the prior distribution are constrained by inverting Eqs. 5 and 6 after having set the sample mean and the variance of the prior distribution.

The variance of the prior distribution denotes the *confidence degree* of our prior information, i.e. the dispersion around the central value represents an evaluation of the

epistemic uncertainties due to our limited knowledge of the process. Here, the confidence degree is set up writing the variance in terms of *equivalent number of data* Λ, as defined by Marzocchi et al. (2004, 2008)

$$\Lambda = \alpha + \beta - 1. \tag{7}$$

Then, Eq. 6 can be rewritten as

$$V(\Theta) = \frac{E(\Theta)(1 - E(\Theta))}{\Lambda + 2}. \tag{8}$$

The parameter Λ has been introduced as a more friendly measure than V of the confidence on the prior distribution, or in other terms, of the epistemic uncertainty. In practice, the parameter Λ weights the prior model in terms of "number of data"; this means that many more than Λ real data are necessary to change the prior distribution (Marzocchi et al. 2008). In general, a high Λ corresponds to a large confidence on the reliability of the prior model, so that the prior distribution needs a large number of past data in the likelihood to be modified significantly. On the contrary, Λ must be small if the prior model is only a very rough first order approximation, so that even a small number of past data in the likelihood can drastically modify the prior distribution. The minimum possible Λ value is 1, representing the maximum possible epistemic uncertainty or maximum of ignorance. As Λ increases, the *Beta* function becomes more and more spiked around the given mean. In the limit case, it is a Dirac's δ, in other words the epistemic uncertainty becomes negligible when a large amount of information are available.

In summary, the input parameters to set up the prior distribution are: (1) the average of the prior *Beta* distribution (Eq. 5) that may be estimated by the weighted percentage of times in which the run-up Z is larger than the selected threshold z_t in our simulations; (2) the equivalent number of data Λ (Eqs. 7, 8) that represents the epistemic uncertainty of our model, or, in other terms, our confidence degree on the model used.

2.2 The likelihood

The likelihood takes into account the historical and/or instrumental past data relative to the variable Z. Let Y represents the variable that counts the number of past time windows in which $Z > z_t$ is a set of n inspections. This n is equal to the total number of nonoverlapping time windows investigated.

Under the assumption of independence of events, Y is defined by

$$Y|(\Theta = \theta) \sim Bin(n, \theta) \tag{9}$$

where *Bin* stands for the following binomial PDF

$$[Y|(\Theta = \theta)] = \binom{n}{Y} \theta^Y (1 - \theta)^{n-Y}. \tag{10}$$

2.3 The posterior model

Through Bayes' theorem and adopting the results of the conjugate families for the binomial model, the posterior distribution for Θ is

$$\Theta|(Y=y) \sim Beta(\alpha+y, \beta+n-y). \tag{11}$$

In this step, we integrate the prior model with the potential information from the past (Eq. 1). Equation 11 indicates the relative weight between the prior distribution and likelihood. Since Λ is function of the sum of α and β (Eq. 7) and n is the number of real data, the ratio of Λ to n controls the relative importance between the prior distribution and the likelihood in the posterior distribution.

3 The modular structure of the Bayesian PTHA

A convenient way to set up a Bayesian PTHA is through a modular structure. In this way, each part of the tsunami process can be modelled separately, from the source and the wave generation, through the propagation, up to the impact. This procedure also permits to single out the uncertainties of each step of the tsunami process, making easier any PTHA improvement, as long as new information/knowledge reducing such uncertainties become available. The modular procedure intends to be similar to the deaggregation practice of the PSHA (Probabilistic Seismic Hazard Assessment) (Cornell 1968; McGuire 1995).

The basic modules M of the structure derive from the Bayesian procedure described in the previous section (Fig. 1).

- *PRIOR*-M: The prior model codifies the information from the simulated tsunami events into a probability distribution (Eq. 2);
- *LIKEL*-M: The likelihood transforms the historical knowledge based on the tsunami catalogues into n total number of data and y number of successes (Eqs. 9, 10);
- *POSTER*-M: The posterior model integrates n and y into the prior distribution to estimate the final posterior distribution (Eq. 11).

In PTHA studies, usually there are no, or very few, data available. On the other hand, there is a plethora of tsunami models, with quite different degree of mathematical complexity (Shuto et al. 1991; Titov and Gonzalez 1997; Ward 2002; Choi et al. 2003; Sato
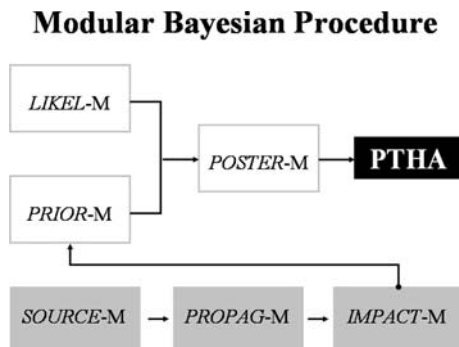

Fig. 1 Scheme of Probability Tsunami Hazard Assessment (*PTHA*) with the modular Bayesian procedure. The identified tsunami sources (in module *SOURCE*-M) generate the initial tsunami waves, which propagates (in module *PROPAG*-M) up to the impact area (in module *IMPACT*-M). The tsunami effects at the key sites are evaluated in terms of exceeding a selected threshold of a selected parameter. The last information set up the prior model (in module *PRIOR*-M), and the historical data determine the likelihood (in module *LIKEL*-M). Their combination in the posterior model (in module *POSTER*-M) estimates the PTHA

et al. 2003; Tinti and Armigliato 2003; Geist 2005; Annaka et al. 2007; Kowalik et al. 2007; Liu et al. 2007; Power et al. 2007; Yanagisawa et al. 2007; Horsburgh et al. 2008; Kietpawpan et al. 2008). In the Bayesian perspective, the PTHA based only on model simulations implies that the posterior distribution is almost set by the prior distribution (i.e. there is no consideration of the historical information, the likelihood) and is supposed to may have a good reliability (i.e. Λ is considered large).

Figure 1 illustrates how the prior distribution can be set through the tsunami modelling. In particular, three modular components of this stage are distinguished:

- *SOURCE*-M: This module identifies the sources and determines the generation of the tsunami waves. The definition of the potential source locations and their relative intensities allow the tsunami initial waves generation to be properly modelled.
- *PROPAG*-M: This module describes the tsunami wave propagation. Most of the propagation models are nonlinear shallow-water models.
- *IMPACT*-M: This module evaluates the tsunami effects at the impact area. The tsunami wave path and intensity can be modified by topography and coastal features. Shoaling waves reduce their velocity, shorten their wavelengths and grow their heights approaching the coastline causing different impacts on different key-sites.

All of these components of tsunami modelling rise uncertainty into PTHA. Part of uncertainties is introduced by epistemic uncertainties due to the incomplete knowledge of all components of the whole physical process. Other components of the uncertainties are due to the aleatory uncertainties, i.e. regardless of our knowledge of the physical processes involved in the tsunami generation. Some aspects remain intrinsically unpredictable, at least in a deterministic sense. An example is the randomness of tsunamigenic events, like earthquakes, submarine landslides and so on. Notably, while the former kind of uncertainty can be potentially reduced by the increase of knowledge and data, the latter cannot be avoided or diminished.

In order to account for all relevant sources of uncertainty, it is necessary to simulate many possible realizations of the tsunami generation, propagation and impact model. The use of simulations is mandatory if we want to incorporate into PTHA all relevant uncertainties to get an accurate estimation. For example, the use of a large number of investigations allows all potential tsunamigenic sources to be accounted for. In general, the need of a statistically significant set of simulations pushes towards the use of models that can run in a reasonable CPU time. However, most of numerical models have an enormous demand of CPU time and a massive storage of output data that make them almost impossible to use for a great number of simulations, therefore for a full PTHA.

We argue that uncertainties on the tsunami sources have usually a larger importance than the uncertainties on the propagation and impact of the tsunami waves (which are basically linked to the use of different codes with different resolutions and parameterizations). In other words, the whole range of possible tsunami sources leads to a greater PTHA (aleatory) variability compared to the (epistemic) variability introduced by the use of models having different degrees of complexity and running for one or very few selected possible sources. Unlike logic-tree schemes, that include only epistemic uncertainties of a well-defined tsunami source (Geist and Parsons 2006; Annaka et al. 2007), the modular Bayesian procedure intends to consider both epistemic and aleatory uncertainties mainly related to the tsunami source. The larger variability associated with the tsunami source suggests that the use of the most sophisticated models is not always the best option to adopt. In fact, sophisticated models may be not able to explore the great variability of the parameters related to the tsunamigenic sources. In this case, the use of simpler physical

models is strongly recommended by the possibility to examine a broader range of possible tsunamigenic sources and many other causes of uncertainty.

A full PTHA requires a high number of computations of the variable Z. This variable, and its relative threshold z_t, are usually defined by the End-Users who know which parameter is more relevant to investigate for the specific site under study. The high number of Z computations are necessary to estimate the weighted percentage of times in which $Z > z_t$ in order to set the average of the prior distribution. For example, a reasonable value for the threshold z_t is a run-up of 0.5 m, because such a wave can hit and hurt people, and some damages to ports, harbours or tourist infrastructures can already happen with these levels of run-up. Then, a set of 1,000 computations of Z usually provides a reasonable global picture of the phenomena under study; in general, the higher the number of sources of uncertainty, the higher the number of simulations.

4 A case study: *Neverland* region

In order to illustrate the application of the Bayesian PTHA, we consider a hypothetical region called *Neverland* that is characterized by imaginary tsunamigenic seismicity. Note that the main goal of this exercise is to illustrate the application of the whole procedure. Some modules probably do not represent the best and most accurate models that can be used in simulations. At the same time, some assumptions adopted for this exercise may appear very simple. Anyway, once the application of the procedure is clarified, the replacement of some modules and/or the use of more reliable assumptions for real applications are straightforward.

We concentrate only on submarine seismic sources, assuming that other tsunamigenic sources, such as submarine slides, landslides and meteorite impacts, can be neglected. Note that, despite *Neverland* is not a real region, the method described for this area can be similarly applied to any region. The description reported below follows the general modular structure schematized in Fig. 1.

4.1 *SOURCE*-M

This module identifies the tsunamigenic seismic sources, describes the fault characteristics and generates the initial tsunami waves in the hypothetical region *Neverland* illustrated in Fig. 2. We assume that a seismic catalogue is available for that region and complete for the last 200 years. We focus the PTHA on 3 key sites selected on the basis of the coastal characteristics: key-site 1 presents a very gentle slope, key-site 2 is a peninsula with a steep bathymetry and key-site 3 is a cape with coastal characteristics occurring between the two previous sites. Figure 2 also shows the seismic activity reported by the seismic catalogue of the region.

At first, 1,000 synthetic seismic potentially tsunanigenic sources are simulated. The epicentral position of each earthquake is randomly selected from the epicentres reported in the *Neverland* earthquake catalogue. In this way, we assume that the seismic catalogue is a statistically representative sample of the whole spatial distribution of tsunamigenic sources. The depth is randomly sampled from a uniform distribution [0,15] km.

A probability of occurrence that is Poisson in time is assigned to each earthquake. The choice of a Poisson distribution in time implies that each event is considered independent from the previous ones. Obviously, this is not strictly true (also large seismic events are usually clustered in time and space; e.g. Kagan and Jackson 2000; Lombardi and
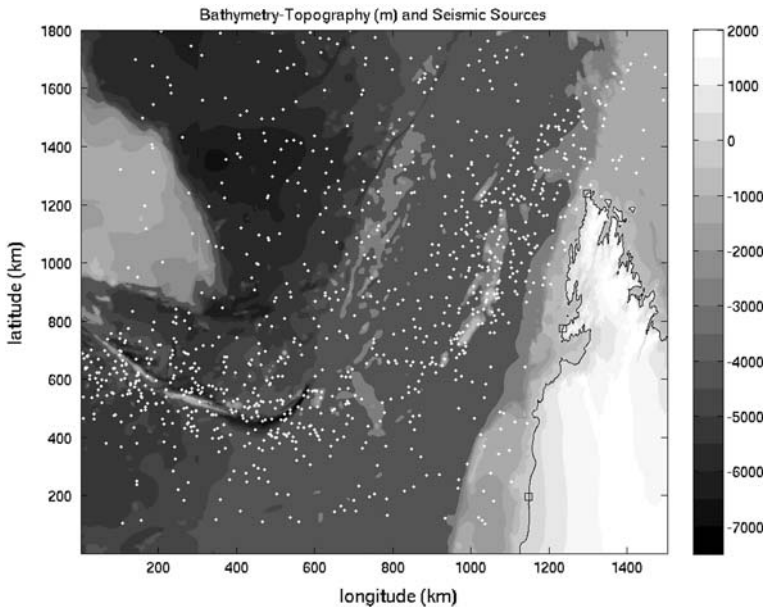
Fig. 2 Bathymetry and topography in the *Neverland* region. The *white dots* indicate the potential submarine seismic sources and the three *black squares* at the coastline represent the key sites (from south to north they are reported, respectively, as key sites 1, 2 and 3)

Marzocchi 2007; Faenza et al. 2008; Marzocchi and Lombardi 2008), but here a Poisson distribution is assumed to be a good first order approximation for M 5.5+, as implicitly made for all applications of the Cornell method (Cornell 1968) in seismic hazard assessment (see Gruppo di Lavoro 2004, for the case of the Italian seismic hazard map). Note that, the proposed method can handle also different earthquake occurrence models, such as space and time clustering models or recurrence models.

The range of moment magnitude spans from 5.5 to 8.1. The lowest moment magnitude (5.5) represents the smallest event that can produce a tsunami; the upper limit is set by assuming that the tectonics of the region is not capable to produce earthquakes larger than 8.1 moment magnitude. In order to efficiently sample a relevant number of the highest moment magnitude that are the most tsunamigenic ones, the moment magnitude of each simulated event is sampled from a uniform distribution [5.5–8.1]. The direct sampling from a Gutenberg–Richter distribution would result in a very few events of high magnitude. However, in order to account for properly the effects of large magnitude events that are obviously less likely to occur compared to events of smaller magnitude, we weigh the probability with a (truncated) Gutenberg–Richter law (Gutenberg and Richter 1954) with a *b*-value estimated from the seismic catalogue of *Neverland*.

The geometric fault parameters (length, width, slip) are calculated by the magnitude using the empirical relationships provided by Wells and Coppersmith 1994. Here, we do not consider the uncertainties related to these empirical rules, because we assume that their effects will be negligible compared to the aleatory uncertainty linked to the space-time-magnitude distribution of the sources.

The seismo-tectonics of the area is described by the moment tensors (Fig. 3) calculated on the base of a number of events statistically significant within each box. The box
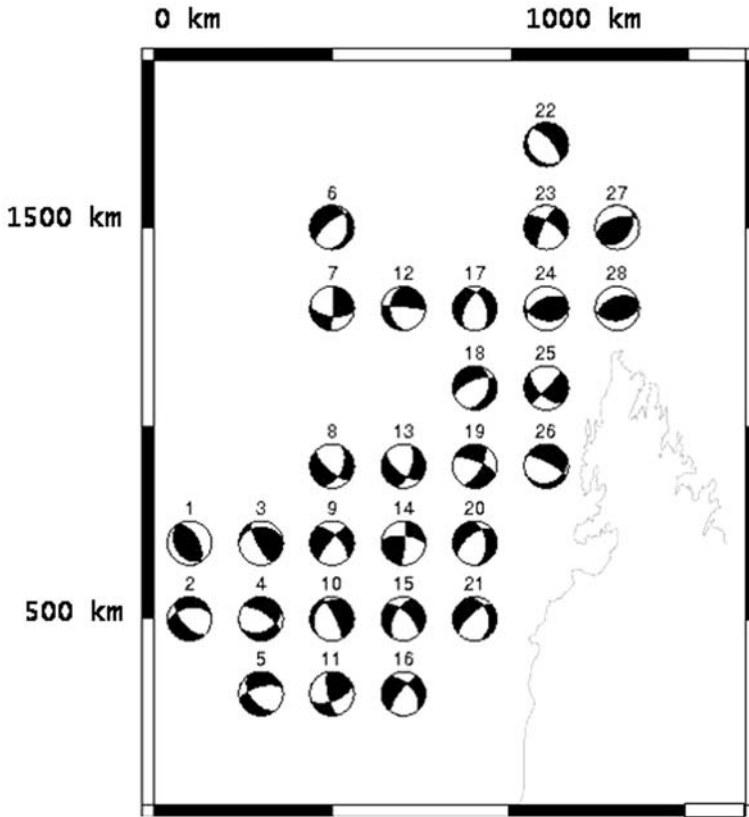
Fig. 3 Moment tensors showing the seism-tectonics of the *Neverland* region

resolution must allow the adequate representation of the strongest earthquakes associated with the longest faults. The Strike, Dip and Rake angles of a box (Fig. 4) are associated with each seismic event that occurs within the same box. If in a box the focal mechanism is not determined, then the seismic events inside that box have the three angles of the nearest source.

Each submarine fault lies on an inclined plane where the dip angle and the Depth D set, respectively, the plane inclination and the fault base. The Depths D are uniformly distributed in the ocean crust that is considered with a maximum thickness of 15 km. We assume that: (1) if DD (=Width $\times$ sin(Dip)) is lower than 15 km, then the correspondent fault base Depth D spans randomly from DD down to 15 km; (2) if DD is greater than 15 km then D is simply equal to DD and the top of the fault reaches the ocean floor. These constraints ensure that each fault is in the upper part of the lithosphere where the rock behaviour is mostly brittle.

Each earthquake generates an ocean floor deformation that is calculated applying the analytical formulas of Okada 1992. The deformation induces the sea surface deformation and the initial tsunami wave. In the cases where the initial tsunami wave was considered unrealistically high, the elevation of the initial tsunami wave was set equal to 2 m. This maximum tsunami initial wave height was based on the satellite measurements. Fujii and Satake (2007) observed from satellite data that the amplitudes of the initial tsunami waves
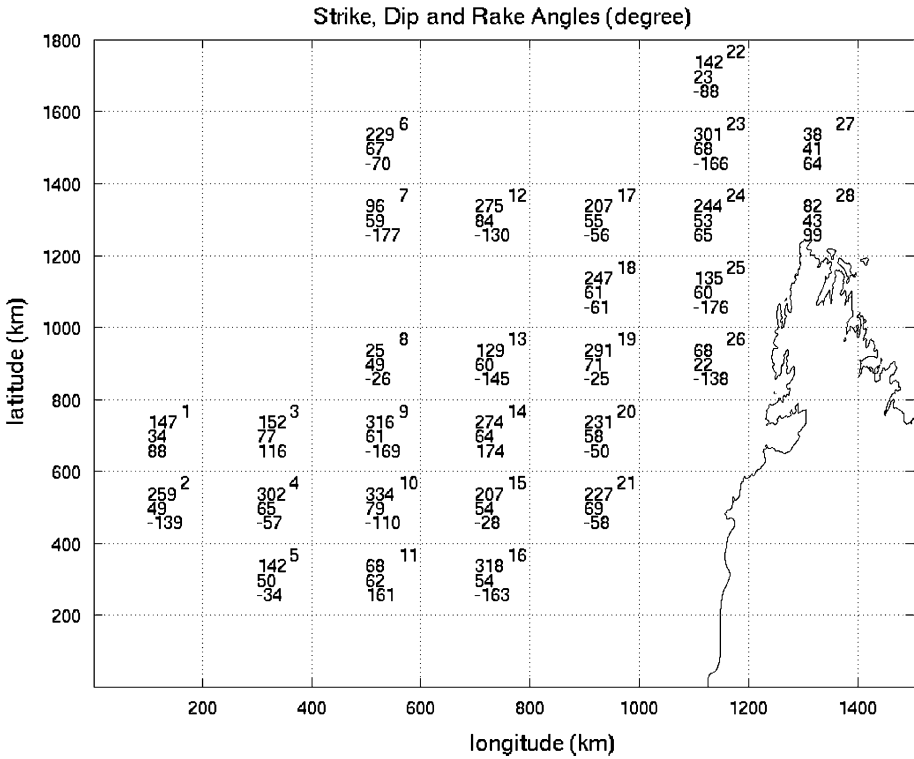
Fig. 4 Strike, Dip and Rake angles of the 1,000 seismic events. The *number in the upper right corner* represents the box number associated to the moment tensors of Fig. 3

ranged from several tens of centimetres to 2 m in the 2004 Sumatra–Andaman event. We argue that this approximation is not critical in a seismic application, because we estimate the probability of a run-up overcoming 0.5 m instead of the absolute run-up.

4.2 *PROPAG*-M and *IMPACT*-M

In the present application, the prior model uses an empirical amplification law (Synolakis 1987) that relates tsunami coastal run-up Z to the initial wave height η and the water depth d at the source. It takes into account the coastal sea-bottom slope φ

$$z = 2.831(\cot \varphi)^{1/2} \eta^{5/4} d^{-1/4}. \tag{12}$$

The slope φ is an average value calculated in an opportune area in front of each key site. This law is a very simple model for the computation of the maximum run-up of a solitary wave climbing up a sloping beach. Owing to its simplicity, the law allows 1,000 coastal run-ups to be evaluated in a very short time. De facto, despite Eq. 12 necessarily contains information about wave propagation, it skips the *PROPAG*-M module, creating a direct link between the *SOURCE*-M and *IMPACT*-M modules.

Any other reliable model (empirical, analytical, numerical or theoretical) or a sequence of models (directly connected or nested) can simulate one or more phases of the tsunami process (potential source, wave propagation, impact effect). The modular structure simply

allows the replacement or update of each module (*SOURCE*-M, *PROPAG*-M and *IMPACT*-M) at any stage keeping the information on the process coming from the previous modules.

The empirical Synolakis' law does not consider the nonlinear processes (such as dispersion and wave breaking) and the geometric effects (such as refraction, spreading and directionality of the tsunami radiation pattern). In order to test the reliability of our simple model, we consider the real case of the Sumatra–Andaman event (26th December 2004). Two key sites were identified for testing: Yala (Sri Lanka) and Banda Aceh (Sumatra). Liu et al. 2005 showed the measured tsunami run-ups in Sri Lanka spanning from 4 m south of Colombo to 10 m near Yala. More detailed studies are published by Papadopoulos et al. (2006) that measured run-ups spanning between 5.5 m in Maratuwa and c.a. 11 m in Hambantota, respectively, in the eastern coast and in the South-Western coast of Sri Lanka. Borrero 2005 reported that in Banda Aceh (Sumatra), the run-up exceeded 8–9 m, and in Lhoknga the run-up was higher than 10–15 m. Using our procedure, the run-ups calculated in Sri Lanka and in Sumatra were, respectively, 7.48 and 9.96 m. We conclude that the present procedure for the run-up computations can match the run-up measurements at the first order, which is usually adequate for the probabilistic study. In other words, as mentioned in section 3, we assume that the epistemic uncertainty introduced by the simple Synolakis' model is significantly lower than the aleatory uncertainty produced by the realistic spatial-size distribution of the tsunamigenic sources.

4.3 PTHA for *Neverland*: *PRIOR*-M, *LIKEL*-M and *POSTER*-M

At each key site, the sample mean of the prior distribution is obtained as

$$\hat{E}(\Theta) = \sum_{i=1}^{1000} p_i H(z_i > 0.5) \tag{13}$$

where p_i is the probability of occurrence of the *i-th* tsunamigenic source, and H is the Heaviside function, i.e. it is 1 when the model run-up z_i is higher than the 0.5 m threshold and 0 otherwise. In this application, we assign the reliability to the prior model by setting $\Lambda = 10$ on the base of practical and expertise considerations. This means that more than ten real data can change significantly our guess based on the prior model.

The prior *Beta* distribution obtained from Eqs. 5 and 7 is showed in Fig. 5 for the 3 key sites. At this stage, the prior model incorporates the hypothetical information from the recent past through the likelihood function (Eq. 10). Historical events in the last 500 years are listed in the hypothetical catalogue reported in Table 1. For example, at key site 1 the catalogue indicates that 6 years out of 500 experienced a run-up larger than the selected threshold of 0.5 m; this means that $n = 500$ and $y = 6$. These quantities in Eq. 11 set the posterior distributions, and the probabilities are computed $\times$ year^{-1}. Figure 6 shows a much smaller variance compared to the prior distributions, mostly due to the large number of years in the time window (500). In Table 2, the mean and the variance of the prior and the posterior distributions are reported for each key site of *Neverland*.

5 Final remarks

In the present study, we have implemented a new general method for PTHA based on a Bayesian approach. The approach has many pivotal features.
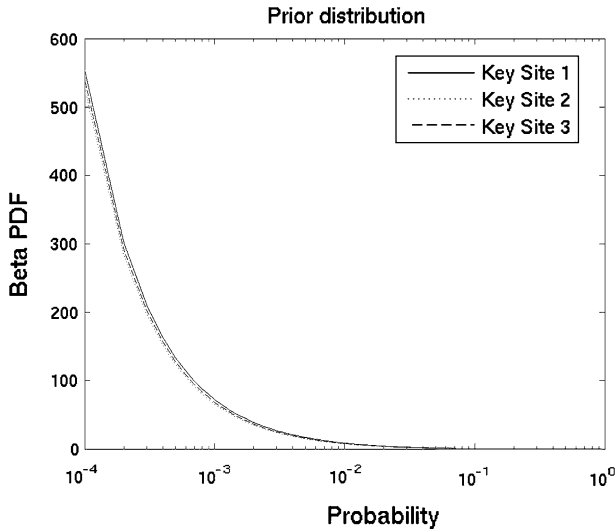
Fig. 5 Prior *Beta* distribution for the 3 key sites at *Neverland*

Table 1 Hypothetical tsunami catalogue of the historical runups in *Neverland* recorded at key sites 1, 2 and 3 in the last 500 years

Date (year/month/day)	Location (km)	Moment magnitude (M_w)	Key site 1 (m)	Key site 2 (m)	Key site 3 (m)
1542-12-04	157.45–891.24	7.9	11	5	6
1612-05-31	1002.56–349.78	7.8	7	a	4
1697-09-01	456.34–900.12	8.1	14	6	8
1876-07-21	289.78–568.09	7.6	4	a	a
1905-08-18	789.51–490.48	7.4	0.7	0.3	0.4
2003-02-06	1116.26–1567.81	7.7	2.3	0.9	1.2

[a] This symbol means no data available for that event

(1) The different information coming from mathematical models, empirical laws, expert beliefs, instrumental measurements and historical data, is accounted for and integrated in a homogeneous way.

(2) The most relevant aleatory and epistemic uncertainties are formally incorporated. This is a key point for a full and reliable PTHA. The PTHA estimated by scenario-based models does not provide a full PTHA, but a probability estimation conditioned to the occurrence of a specific tsunamigenic source, modelling the related epistemic uncertainty. In the procedure described here, it is possible to incorporate the aleatory uncertainty associated to the tsunami generation process, whose importance may overcome other causes of uncertainty.

(3) The modular structure allows an easy update of the procedure as long as new models and/or additional data are accessible to reduce epistemic uncertainty. As a consequence, each module can be replaced as soon as more advanced and reliable information relative to model and/or data are available. The prior and the posterior distributions can be more accurate/precise if specific parts of the tsunami process are better understood and
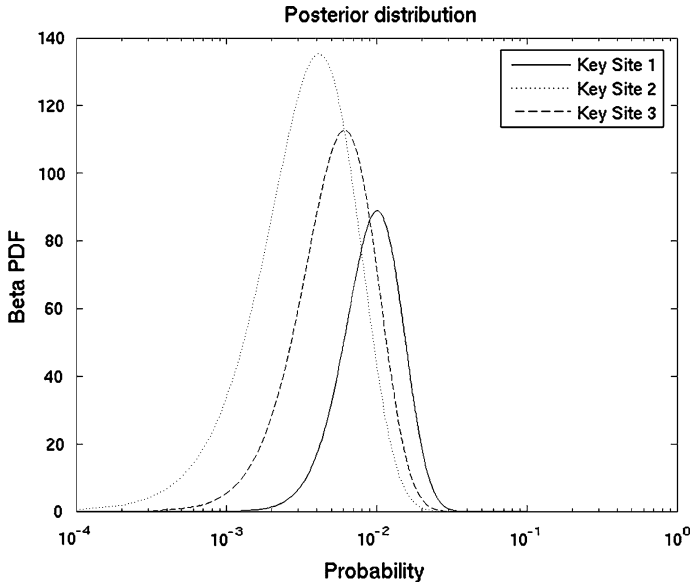
Fig. 6 Posterior *Beta* distribution for the 3 key sites at *Neverland*

Table 2 Mean and variance of the prior and posterior distributions ($\times$year^{-1})

	Key site 1		Key site 2		Key site 3	
	Mean	Variance	Mean	Variance	Mean	Variance
Prior distribution	1.08×10^{-2}	8.9×10^{-4}	0.91×10^{-2}	7.5×10^{-4}	0.97×10^{-2}	8.0×10^{-4}
Posterior distribution	1.20×10^{-2}	2.3×10^{-5}	0.61×10^{-2}	1.2×10^{-5}	0.80×10^{-2}	1.6×10^{-5}

more effectively represented and/or if instrumental records and historical references undergone to further developments and increased documentations.

As example, such a procedure has been applied to the hypothetical case of *Neverland*. Even though this region clearly does not exist in the world, it is realistic, and it represents a well possible situation where PTHA can be estimated. The demonstrative case highlights the need to use tsunami models that can run thousand simulations. Despite the fact that a very complex and accurate physical model is always desirable, in practice, in PTHA, it is much more useful to adopt simpler models that can simulate thousand events accounting for the complex aleatory distribution of the tsunamigenic potential sources. Under this perspective, more sophisticated models that require a huge computational effort are mostly effective to study the detailed impact of a single or few events given the initial conditions of the scenarios possibly reducing the epistemic uncertainties of the propagation and the impact phases.

References

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

Borrero JC (2005) Field data and satellite imagery of tsunami effects in Banda Aceh. Science 308:156

Burbidge D, Cummins PR, Mleczko R, Thio HK (2008) A probabilistic tsunami hazard assessment for Western Australia. Pure Appl Geophys. doi:10.1007/s00024-008-0421-x

Burroughs SM, Tebbens SF (2005) Power-law scaling and probabilistic forecasting of tsunami runup heights. Pure Appl Geophys 162:331–342

Choi BH, Pelinovsky E, Hong SJ, Woo SB (2003) Computation of tsunamis in the East (Japan) Sea using dynamically interfaced nested model. Pure Appl Geophys 160:1383–1414

Cornell CA (1968) Engineering seismic risk analysis. Bull Seismol Soc Am 58:1583–1606

Faenza L, Marzocchi W, Serretti P, Boschi E (2008) On the spatio-temporal distribution of M 7.0+ worldwide seismicity with a non-parametric statistics. Tectonophysics 449:97–104. doi:10.1016/j.tecto.2007.11.066

Farreras S, Ortiz M, Gonzalez JI (2007) Steps towards the implementation of a tsunami detection, warning, mitigation and preparedness program for Southwestern coastal areas of Mexico. Pure Appl Geophys 164:605–616

Fournier d'Albe EM (1979) Objectives of volcanic monitoring and prediction. J Geol Soc 136:321–326

Fujii Y, Satake K (2007) Tsunami source of the 2004 Sumatra-Andaman earthquake inferred from tide gauge and satellite data. Bull Seismol Soc Am 97:192–207

Geist EL (2005) Rapid tsunami models and earthquake source parameters: far-field and local applications. ISET J Earthquake Tech 42:127–136

Geist EL, Parsons T (2006) Probabilistic analysis of tsunami hazards. Nat Haz 37:277–314

Gelman A, Carlin JB, Stern HS, Rubin DB (1995) Bayesian data analysis. Chapman and Hall, New York

Gerstenberger MC, Wiemer S, Jones LM, Reasenberg PA (2005) Real-time forecasts of tomorrow's earthquakes in California. Nature 435:328–331

Gruppo di lavoro (2004) Redazione della mappa di pericolosità sismica prevista dall'Ordinanza PCM 3274 del 20 marzo 2003. Rapporto conclusivo per il Dipartimento della Protezione Civile, INGV, Milano-Roma, aprile 2004, 65 pp + 5 appendici

Gutenberg B, Richter C (1954) Seismicity of the earth and associated phenomena, 2nd edn. Princeton University Press, New Jersey

Horsburgh KJ, Wilson C, Baptie BJ, Cooper A, Cresswell D, Musson RMW, Ottemoller L, Richardson S, Sargeant SL (2008) Impact of a Lisbon-type tsunami on the U.K. coastline and the implications for tsunami propagation over broad continental shelves. J Geophys Res 113:C04007. doi:10.1029/2007JC004425

Kagan YY, Jackson DD (2000) Probabilistic forecasting of earthquakes. Geophys J Int 143:438–453

Kietpawpan M, Visuthismajarn P, Tanavud C, Robson MG (2008) Method of calculating tsunami travel times in the Andaman Sea region. Nat Haz 46:89–106

Kowalik Z, Knight W, Logan T, Whitmore P (2007) The tsunami of 26 December, 2004: numerical modeling and energy considerations. Pure Appl Geophys 164:379–393

Liu PL-F, Lynett P, Fernando H, Jaffe BE, Fritz H, Higman B, Morton R, Goff J, Synolakis C (2005) Observation by international tsunami survey team in Sri Lanka. Science 308:1595

Liu Y, Santos A, Wang SM, Shi Y, Liu H, Yuen DA (2007) Tsunami hazards along Chinese coast from potential earthquakes in South China Sea. Phys Earth Plan Int 163:233–244

Lombardi AM, Marzocchi W (2007) Evidence of clustering and nonstationarity in the time distribution of large worldwide earthquakes. J Geophys Res 112:B02303. doi:10.1029/2006JB004568

Marzocchi W, Lombardi AM (2008) A double branching model for earthquake occurrence. J Geophys Res 113:B08317. doi:10.1029/2007JB005472

Marzocchi W, Sandri L, Gasparini P, Newhall C, Boschi E (2004) Quantifying probabilities of volcanic events: the example of volcanic hazard at Mount Vesuvius. J Geophys Res 109:B11201. doi:10.1029/2004JB003155

Marzocchi W, Neri A, Newhall CG, Papale P (2007) Probabilistic volcanic hazard and risk assessment. EOS, Tran AGU 88(32):318

Marzocchi W, Sandri L, Selva J (2008) BET_EF: a probabilistic tool for long- and short-term eruption forecasting. Bull Volcanol 70:623–632. doi:10.1007/s00445-007-0157-y

McGuire R (1995) Probabilistic seismic hazard analysis and design earthquakes: closing the loop. Bull Seismol Soc Am 85:1275–1284

Okada Y (1992) Internal deformation due to shear and tensile faults in a half-space. Bull Seismol Soc Am 82:1018–1040

Orfanogiannaki K, Papadopoulos GA (2007) Conditional probability approach of the assessment of tsunami potential: application in three tsunamigenic regions of the Pacific Ocean. Pure Appl Geophys 164:593–603

Papadopoulos GA, Caputo R, McAdoo B, Pavlides S, Karastathis V, Fokaefs A, Orfanogiannaki K, Valkaniotis S (2006) The large tsunami of 26 December 2004: field observations and eyewitnesses accounts from Sri Lanka, Maldives Is and Thailand. Earth Planets Space 58:233–241

Power W, Downes G, Stirling M (2007) Estimation of tsunami hazard in New Zealand due to South American earthquakes. Pure Appl Geophys 164:547–564

Sato H, Murakami H, Kozuki Y, Yamamoto N (2003) Study of a simplified method of tsunami risk assessment. Nat Haz 29:325–340

Savage JC (1994) Empirical earthquakes probabilities from observed recurrence intervals. Bull Seismol Soc Am 84:219–221

Senior Seismic Hazard Analysis Committee (SSHAC) (1997), Recommendation for probabilistic seismic hazard analysis: guidance on uncertainty and use of experts, Report NUREG/CR-6372, Washington DC

Shuto N, Goto C, Imamura F (1991) Numerical simulation as a means of warning for near field tsunamis. Coast Eng Jpn 33(2):173–193

Synolakis CE (1987) The runup of solitary waves. J Fluid Mech 185:523–545

Tinti S, Armigliato A (2003) The use of scenarios to evaluate the tsunami impact in southern Italy. Mar Geol 199:221–243

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 42:189–201

Titov V, Gonzalez FI (1997) Implementation and testing of the Method of Splitting Tsunami (MOST) model, Technical Report NOAA Tech. Memo. ERL PMEL-112 (PB98-122773), NOAA/Pacific Marine Environmental Laboratory, Seattle, WA

Ward SN (2002) Tsunamis. In: Meyers RA (ed) Encyclopedia of physical science and technology. Academic Press, New York

Wells DL, Coppersmith KJ (1994) New empirical relationships among magnitude, rupture length, rupture width, rupture area and surface displacement. Bull Seismol Soc Am 84:974–1002

Yanagisawa K, Imamura F, Sakakiyama T, Annaka T, Takeda T, Shuto N (2007) Tsunami assessment for risk management at nuclear power facilities in Japan. Pure Appl Geophys 164:565–576