Geophysical Journal International

*Geophys. J. Int.* (2015) **200,** 574–588
GJI Seismology

doi: 10.1093/gji/ggu408

# Probabilistic hazard for seismically induced tsunamis: accuracy and feasibility of inundation maps

S. Lorito, J. Selva, R. Basili, F. Romano, M.M. Tiberti and A. Piatanesi

*Istituto Nazionale di Geofisica e Vulcanologia, via di Vigna Murata* 605, *I*-00143 *Roma, Italy. E-mail: stefano.lorito@ingv.it*

Accepted 2014 October 16. Received 2014 October 16; in original form 2014 April 24

## SUMMARY

Probabilistic tsunami hazard analysis (PTHA) relies on computationally demanding numerical simulations of tsunami generation, propagation, and non-linear inundation on high-resolution topo-bathymetric models. Here we focus on tsunamis generated by co-seismic sea floor displacement, that is, on Seismic PTHA (SPTHA). A very large number of tsunami simulations are typically needed to incorporate in SPTHA the full expected variability of seismic sources (the aleatory uncertainty).

We propose an approach for reducing their number. To this end, we (i) introduce a simplified event tree to achieve an effective and consistent exploration of the seismic source parameter space; (ii) use the computationally inexpensive linear approximation for tsunami propagation to construct a preliminary SPTHA that calculates the probability of maximum offshore tsunami wave height ($HMax$) at a given target site; (iii) apply a two-stage filtering procedure to these 'linear' SPTHA results, for selecting a reduced set of sources and (iv) calculate 'non-linear' probabilistic inundation maps at the target site, using only the selected sources. We find that the selection of the important sources needed for approximating probabilistic inundation maps can be obtained based on the offshore $HMax$ values only. The filtering procedure is semi-automatic and can be easily repeated for any target sites.

We describe and test the performances of our approach with a case study in the Mediterranean that considers potential subduction earthquakes on a section of the Hellenic Arc, three target sites on the coast of eastern Sicily and one site on the coast of southern Crete. The comparison between the filtered SPTHA results and those obtained for the full set of sources indicates that our approach allows for a 75–80 per cent reduction of the number of the numerical simulations needed, while preserving the accuracy of probabilistic inundation maps to a reasonable degree.

**Key words:** Probability distributions; Tsunamis; Computational seismology; Subduction zone processes; Europe.

## 1 INTRODUCTION

In recent years, the tsunami science community has strived to define a set of best practices and standards for probabilistic tsunami hazard analysis (PTHA; Geist & Parsons 2006), mainly based on explicit modelling of seismically-induced tsunamis. Particularly following the 2004 December 26, Indian Ocean tsunami, and with additional motivation following the 2011 Tohoku earthquake, such methods have been progressively employed for different regions and coasts exposed to tsunami risk (Annaka *et al.* 2007; Liu *et al.* 2007; Thio *et al.* 2007, 2010; Burbidge *et al.* 2008; González *et al.* 2009, 2010, 2013; Heidarzadeh & Kijko 2011; Mitsoudis *et al.* 2012; Sørensen *et al.* 2012; Suppasri *et al.* 2012; Lane *et al.* 2013; Omira *et al.* 2013; Horspool *et al.* 2014; Leonard *et al.* 2014). Earthquake-generated tsunamis are indeed the majority of those observed (NGDC/WDS Global Historical Tsunami Database; The Euro-Mediterranean Tsunami Catalogue, Maramai *et al.* 2014). We will refer to the hazard analysis for this kind of events as seismic PTHA (SPTHA).

In any probabilistic analysis of natural hazards there exists a trade-off between the accuracy and precision of the assessment and its practical feasibility. Modelling natural phenomena with sufficient resolution to characterize their full variability is a very demanding computational problem. For example, most probabilistic seismic hazard analysis (PSHA) efforts aim to explore extensively the source parameter space through a zone-based approach (Cornell 1968) by uniformly distributing earthquakes in a given seismogenic surface or volume. Specific simplifications are then adopted to make the hazard assessment practically feasible, for example using empirical ground motion prediction equations instead of performing explicit numerical simulations. However, these choices may decrease the precision and, potentially, the accuracy of hazard estimates.

Sørensen *et al.* (2012) recently proposed a PSHA-reminiscent approach based on earthquake catalogues generated with a Monte Carlo technique to take into account properly the aleatory uncertainty of potentially tsunamigenic earthquakes in the Mediterranean Sea. The exceedingly large size of this kind of simulated earthquake catalogues prevents the calculation of inundation maps, however, and SPTHA end up being expressed in terms of maximum wave height ($HMax$) probability offshore, that is computationally cheaper. Furthermore, this approach makes it difficult to honour the geometric variability of subduction zones, since the probability of earthquakes is assumed homogenous inside a volume.

SPTHA methods have been indeed traditionally based on the use of extended faults and explicit numerical tsunami modelling (e.g. Rikitake & Aida 1988; Annaka *et al.* 2007). This choice is due to the strong tsunami sensitivity to fault geometry and earthquake mechanism (e.g. regarding generation efficiency and tsunami directivity), and to the large anisotropy of tsunami propagation over a complex bathymetry. Exploring the full possible source variability and performing an explicit numerical tsunami simulation for each realization of the source parameters, however, can be extremely demanding from the computational viewpoint. This feasibility issue arises, in particular, when probabilistic inundation maps—and hence very high-resolution non-linear simulations—are needed. For this reason, most of the times a selection including a few major earthquakes—for example, the most likely events for a given average return period (ARP)—is assumed to be sufficient for the assessment (e.g. González *et al.* 2009). This might be an effective approach, provided that: (i) long enough ARPs are considered, when the effect of more frequent smaller earthquakes is expected to be negligible compared with the larger ones and (ii) the analysis is conducted in a relatively simple geophysical context where, for example, tsunami hazard is dominated by large earthquakes occurring in subduction zones, whose geometry is reasonably well constrained. These assumptions help making SPTHA computationally feasible but may decrease its accuracy. In particular, there is a chance that the hazard is severely underestimated in complex and fragmented tectonic environments (e.g. the Caribbean Sea and the Mediterranean Sea) or when shorter ARPs need to be considered. Additionally, even when considering ARPs on the order of several hundreds of years, sampling too coarsely the distributions of source parameter may result in an underestimation of the near-field tsunami hazard.

Recent SPTHA studies for both Australia (Burbidge *et al.* 2008) and California (Thio *et al.* 2010) proposed a method for adequately estimating earthquake source variability while keeping major earthquakes bound to their causative faults. This approach uses SPTHA de-aggregation and offshore $HMax$ matching to select only the earthquakes that are judged to be relevant for subsequent inundation calculations. In the frame of an application for the southeastern Aegean Sea, Mitsoudis *et al.* (2012) exploited some convergence properties of the estimators for the maximum of a given tsunami inundation measure, using a Monte Carlo approach. They used a quasi-deterministic scenario-based method, however, as they consider only the most likely maximum earthquake magnitude (the mode) while randomly varying the epicentres.

In this study, we propose a computationally cheaper approach to treating aleatory uncertainty in SPTHA that would reduce the number of input earthquake scenarios while keeping the results stable and consistent with those obtained with a full scenario set. Our method seeks a balance between avoiding an excessive reduction of the assumed source variability and an excessive simplification of the tsunami simulations. It is worth noting that the selection of sources is based on the resulting tsunami variability at a given site, not on the similarity of causative earthquakes. SPTHA indeed cannot be considered a simple extension of PSHA because the simplifications and the assumptions usually made for PSHA, while suitable for representing earthquake variability, might not represent the full tsunami variability. For example, the 2011 Tohoku catastrophe clearly showed that the sources of the strong ground shaking and of the tsunami may originate at different places on the fault plane (e.g. Lay & Kanamori 2011; Romano *et al.* 2012, 2014; Satake *et al.* 2013); thus, the sources for SPTHA need a special treatment based on how the potentially threatened sites would be impacted. We share this same basic point with some of the above mentioned analyses (Burbidge *et al.* 2008; Thio *et al.* 2010; Mitsoudis *et al.* 2012), along with the attempt to reduce the computational cost associated with inundation maps. Nevertheless, our approach uses a more systematic and quantitative method of selecting the significant sources, to be performed before calculating inundation maps.

To illustrate our method we initially define a quite general framework for the PTHA due to any type of source, and then focus on SPTHA. We approach only the seismic aleatory variability (the variety of possible earthquake realizations) by means of an event tree (ET). Conversely, we do not manage epistemic uncertainties; various strategies exist for doing this, the most commonly used in SPTHA being the logic tree (LT) approach (e.g. Bommer & Scherbaum 2008). Each node of a LT corresponds to a set of modelling assumptions, and a specific SPTHA hazard curve is developed for each branch. Within such a strategy, the ET would be nested into each branch of the LT.

The ET provides a controlled discretisation of the earthquake parameter space that allows us to assess conditional probabilities, which are normally easier to deal with. The discretisation is performed by heuristically oversampling each parameter distribution, in an attempt to preserve all relevant information. The resulting—and likely redundant—set of earthquake parameters is then used to obtain the SPTHA within a Green's functions approach, using *HMax* offshore as the hazard intensity measure. By doing this we obtain a region-wide, homogeneous and computationally inexpensive 'linear' SPTHA.

To set up a feasible local hazard assessment for any selected target area where one may need inundation maps we propose a two-stage filtering procedure. In the first stage we identify and eliminate all the sources that give a negligible contribution at the target location, for example the smallest earthquakes or those directing most of the tsunami energy elsewhere. In the second stage we search within the remaining set of sources for parameter clusters that produce similar $HMax$ patterns offshore of the target. For each magnitude level we thus identify a representative scenario for each cluster; we assign to this scenario the probability of the entire cluster. The subset of sources resulting from the filtering is finally used for calculating 'non-linear' SPTHA, that is, probabilistic inundation maps at the target location.

We illustrate and quantitatively test our approach by means of a SPTHA case study in the Central Mediterranean Sea, using potential subduction earthquakes occurring in the western portion of the Hellenic Arc, and using target sites on the coasts of eastern Sicily, Italy and southern Crete (Greece). The performance of the method is then tested by comparing the inundation maps obtained for the target sites with the reduced (filtered) set of sources against those obtained with the complete set of sources included in the ET.

## 2 OPERATIONAL SPTHA FRAMEWORK

Probabilistic hazard analysis is a methodology that estimates the likelihood that various levels of intensity $\psi$ will be exceeded at a given location $\boldsymbol{x}$ in a given future time period $\Delta T$ (exposure time) due to whatever source. The results of such an analysis are usually expressed as estimated probabilities per year or estimated annual frequencies (e.g. SSHAC 1997; Cornell & Krawinkle 2000; Geist & Parsons 2006), properly extended over $\Delta T$ assuming their stationarity. This probability will be hereinafter indicated as $Pr(\psi \geq \bar{\psi}; \boldsymbol{x}, \Delta T)$, where $\bar{\psi}$ is a threshold value for the specific tsunami metric $\psi$.

Tsunami sources may be of different nature such as earthquakes, landslides, meteorite impacts or even atmospheric phenomena. In principle, all the possible sources should be considered and combined together. In practice, several tsunami sources may be more important than others at a specific location and/or some of them may have a negligible impact for a given ARP. In order to separate the tsunamis due to different sources, assuming their statistical independence, we may write:

$$Pr(\psi \geq \bar{\psi}; \boldsymbol{x}, \Delta T) = 1 - \prod_{s} [1 - Pr(\psi \geq \bar{\psi}; \boldsymbol{x}, \Delta T, s)], \quad (1)$$

where each $Pr(\psi \geq \bar{\psi}; \boldsymbol{x}, \Delta T, s)$ represents the hazard posed by the source $s$. In eq. (1), one of the possible source types is earthquake ($s = E$), but also non-seismic sources can be present ($s = \{nE\}$). Note also that the assumption of independence is common (e.g. Parsons & Geist 2009; Grezio *et al.* 2012) but not granted, since among non-seismic sources there are, for example, seismically induced landslides that cannot be considered statistically independent from earthquakes.

From now on, we will concentrate only on SPTHA, that is, that part of the tsunami hazard due to earthquake sources that directly cause tsunamis, hereafter indicated by $Pr(\psi \geq \bar{\psi}; \boldsymbol{x}, \Delta T, E)$. Notice that SPTHA will coincide with the total PTHA only if other possible sources, including seismically induced landslides, have a negligible occurrence probability and/or an overall negligible effect, so that

$$Pr(\psi \geq \bar{\psi}; \boldsymbol{x}, \Delta T, \{nE\}) \ll Pr(\psi \geq \bar{\psi}; \boldsymbol{x}, \Delta T, E). \quad (1a)$$

The degree at which such an assumption holds may be referred to as 'source type coverage'. If the source type coverage is small, one must consider that SPTHA is just a factor of PTHA, to be combined with the PTHA relative to other significant sources, which is most likely the case at many locations.

If we consider a finite set of independent possible earthquake fault zones ($Z_i, i = 1, 2, \ldots, N_Z$), for which we can model geometry and behaviour, the probability that at least one of them is activated in $\Delta T$, producing a tsunami with $\psi \geq \bar{\psi}$, can be calculated as:

$$Pr(\psi \geq \bar{\psi}; \boldsymbol{x}, \Delta T, E) \approx Pr(\psi \geq \bar{\psi}; \boldsymbol{x}, \Delta T, E \in Z_i) = 1 - \prod_{i=1}^{N_Z} [1 - Pr(\psi \geq \bar{\psi}; \boldsymbol{x}, \Delta T, Z_i)], \quad (2)$$

if

$$Pr(\psi \geq \bar{\psi}; \boldsymbol{x}, \Delta T, E \notin Z_i) \ll Pr(\psi \geq \bar{\psi}; \boldsymbol{x}, \Delta T, E \in Z_i). \quad (2a)$$

The degree at which the pre-defined set of earthquake fault zones effectively represents all possible seismic sources may be referred to as 'seismic source coverage'. If we can assume that the seismic sources in each zone are known to a reasonable degree, each seismic source and the ensuing tsunami may be fully modelled with quasi-deterministic procedures. However, it might be necessary to devise a more complicated approach than the one presented here, with at least a statistical treatment of possibly unknown sources.

Summarizing, the quantification of SPTHA due to the pre-selected earthquake fault zones represents a good estimation of the total PTHA only if both source type coverage and seismic source coverage adequately satisfy the above conditions (1a) and (2a), that is:

$$Pr(\psi \geq \bar{\psi}; \boldsymbol{x}, \Delta T) \approx Pr(\psi \geq \bar{\psi}; \boldsymbol{x}, \Delta T, E) \approx 1 - \prod_{i=1}^{N_Z} [1 - Pr(\psi \geq \bar{\psi}; \boldsymbol{x}, \Delta T, Z_i)]. \quad (3)$$

This assumes that PTHA is solely due to earthquakes constrained to occur on mapped faults, and it dominates over all other non-seismic sources for the considered time window at a given location. If such assumptions are not satisfied, significant bias in the PTHA evaluation may be introduced. For an effective assessment of tsunami hazards it is essential to discuss in detail both source type and seismic source coverage, although this is out of the scope of this paper.

For simplicity, we can assume that the distribution of observed tsunamis at the target location $\boldsymbol{x}$ can be represented as a Poisson arrival time process (e.g. Geist & Parsons 2006; Parsons & Geist 2009), so that the annual probability that at least one tsunami will occur at the given site is

$$Pr(\psi \geq \bar{\psi}; \boldsymbol{x}, \Delta T = 1yr) \approx 1 - \exp\left( - \sum_{i=1}^{N_Z} \lambda_i(\psi \geq \bar{\psi}; \boldsymbol{x}, Z_i) \times 1yr \right), \quad (4)$$

where $\lambda_i$ is the mean annual frequency of $\psi \geq \bar{\psi}$ (annual rate of exceedance of the intensity due to an earthquake in the fault zone $Z_i$). Following eq. (4), SPTHA may be evaluated via the contributions due to each independent fault zone $Z_i$.

In each zone, all of the events that may occur can be operatively discretised in a finite number of possible different earthquakes. Such possible seismic events will be hereinafter indicated as 'seismic scenarios' ($\sigma_j, j = 1, 2, \ldots, N_\sigma$) for the zone. Assuming that this set of scenarios is complete (i.e. $\sum_{j=1}^{N_\sigma} Pr(\sigma_j | Z_i) = 1$), the annual exceedance rate related to the zone can be written as

$$\lambda_i(\psi \geq \bar{\psi}; \boldsymbol{x}, Z_i) = \lambda_{Z_i} \sum_{j=1}^{N_\sigma} Pr(\sigma_j | Z_i) Pr(\psi \geq \bar{\psi} | \sigma_j; \boldsymbol{x}). \quad (5)$$

In eq. (5), there are three terms. The first, $\lambda_{Z_i}$, is the mean annual frequency of earthquakes in the $i$th seismic zone. The second term, $Pr(\sigma_j | Z_i)$, represents the probability that an earthquake occurs in the zone corresponding to the $j$th scenario $\sigma_j$; in other words, it represents the aleatory variability of the seismic source. This is the most complicated factor to be evaluated, and it will be the subject of the next section. The third term, $Pr(\psi \geq \bar{\psi} | \sigma_j; \boldsymbol{x})$, represents the exceedance probability for a given earthquake in the fault zone with scenario $\sigma_j$. This term may be assessed through direct models of tsunami propagation and inundation. Almost all PTHA studies assume that this term is completely deterministic, so that its value will be either 0 or 1 (e.g. Rikitake & Aida 1988). This disregards uncertainties in the generation, propagation and inundation model,

which are sources of epistemic uncertainties. However, we do not deal with this in our study.

We remark that the distinction between aleatory and epistemic uncertainty may sometimes become quite fuzzy. An operational definition might be to consider the aleatory uncertainty as the intrinsic variability of a certain 'model of the world', while that concerning the lack of knowledge regarding the validity of the model itself is classified as epistemic uncertainty (SSHAC 1997). We here try to comply with the distinction made by Geist & Parsons (2006), considering as epistemic the uncertainty that can be reduced by collection of additional data, and as aleatory that inherent in the physical process considered. Note, in fact, that eq. (5) corresponds, for each zone $Z_i$, to the integral form of eq. (7) in section 2.2 of Geist & Parsons (2006), which accounts for the conventional way of dealing with aleatory uncertainty. In both PSHA and PTHA, analytic probability models are usually developed for the uncertainty, and their probability density function (PDF) is integrated into the rate (frequency) term. We here introduce a different, discretised form, for this equation. The corresponding probabilities will also be discretised within the ET framework introduced in the next section.

## 3 THE EVENT TREE (ET) FOR SAMPLING THE ALEATORY VARIABILITY OF THE SEISMIC SOURCE

We explore the aleatory uncertainties for a fault zone through an ET. An ET is a branching graph representation of events in which individual branches are alternative steps from a general prior event, state, or condition, and which evolve into increasingly specific subsequent events. ET analysis is then a logical inductive process used to determine the path from an initiating event to the various consequences and the expected frequency of each consequence, and it has found application in many fields of science (e.g. Clifton & Ericson 2005), including geophysics (e.g. Newhall & Hoblitt 2002). Its different branches represent different possibilities, not 'alternative models' as in LTs (e.g. Bommer & Scherbaum 2008). Here, the ET provides a controlled discretisation of earthquake parameter space, in order to (i) define the set of scenarios $\sigma_j$ that may be originated in the zone and (ii) assess the probability of occurrence $Pr(\sigma_j|Z_i)$ for each $\sigma_j$, leading to the assessment of the overall hazard related to the zone $Z_i$ through eq. (5). The main purposes of a controlled discretisation are to avoid oversimplifications in the treatment of source parameters, and to cover their entire range without ending up with huge synthetic catalogues.

We analyse all of the source parameters at the different ET nodes (or levels) in a pre-defined logical order, according to dependence of each node on the previous one. Although we define the ET for all the parameters, we keep some of them fixed in our example. Thus, while this schematization of the assessment through an ET is general, the following parameterization is intentionally simplified for clarity.

Each parameter range is discretised with a finite number of values, each of them representative of a finite interval. We assume that the discretisation steps are small enough not to influence final results (Bazzurro & Cornell 1999). The probability of occurrence associated with each interval is evaluated at each node. Possible correlations among parameters are considered through conditional probabilities to parameter values at the previous nodes. We here consider six nodes:

(1) *Node 1:* Magnitude $M$ of the event.
(2) *Node 2:* Length $L$ of the seismic rupture.
(3) *Node 3:* Width $W$ of the seismic rupture.
(4) *Node 4:* Geometrical centre $x$ of the rupture area along the strike direction.
(5) *Node 5:* Geometrical centre $y$ of the rupture area along the dip direction.
(6) *Node 6:* Slip $S$ (6a) and rake angle $r$ (6b) distributions inside the rupture area.

The actual procedures adopted for each node, and the evaluation of the associated conditional probabilities, are detailed in what follows. To simplify notations, the distinction between symbols of intervals and their representative values is omitted where possible.

*Node 1:* We define a set of $n_1$ magnitudes $M_k$ ($k = 1, 2, \ldots, n_1$) covering the complete range of expected magnitudes within the source zone. In order to evaluate the probability of $M_k$, we assume that it represents the interval $I_k = [M_k, M_{k+1}[$, being the minimum magnitude in an interval the most probable (the mode). This approximation holds if the interval is chosen small enough. The probability of occurrence of each magnitude interval is then assessed by:

$$P_1(k) \stackrel{\text{def}}{=} Pr\,(M \in I_k) = \Phi\,(m_{k+1}) - \Phi\,(m_k)\,, \tag{6}$$

where $\Phi(m_k) \stackrel{\text{def}}{=} \Pr(m \leq m_k)$ is the cumulative distribution function adopted for the seismic moment $m_k$ corresponding to magnitude $M_k$ (Kanamori & Brodsky 2001). Here, we use the truncated Pareto distribution (Kagan 2002, eq. 9).

*Nodes 2–3:* The geometrical characterization of the rupture area for each magnitude is based on the empirical earthquake scaling law considered appropriate for the source zone. Here, we simplify these two nodes by considering for each $M_k$ only the expected values, $\langle L_k \rangle$ and $\langle W_k \rangle$, and neglecting the uncertainty associated to the chosen scaling laws for $L$ and for $W$. Therefore, the probability at these nodes is equal to 1:

$$\begin{cases} P_2(k) \stackrel{\text{def}}{=} Pr\,(\langle L_k \rangle | M_k) = 1 \\ P_3(k) \stackrel{\text{def}}{=} Pr\,(\langle W_k \rangle | M_k) = 1 \end{cases}. \tag{7}$$

*Node 4:* The length of the fault is divided into equal parts that define $N_4$ possible geometrical centres of the rupture area $x_l$ along the strike. Again, as for the earthquake magnitude intervals, the value of $N_4$ is selected so that the distance between two adjacent points should be small enough to be undetectable in terms of the resulting tsunami at the target coast, for all possible magnitudes. Thus, with increasing magnitude, progressively more edge points can no longer accommodate the associated earthquake size, since their distance from the fault edge is too small (i.e. $< \langle L_k/2 \rangle$); these points thus have a null probability of being the centre of such events. Conversely, for all the other $n_4^k$ remaining points (with $n_4^k \leq N_4$), an equal probability of occurrence is assumed:

$$P_4\,(l, k) \stackrel{\text{def}}{=} Pr\,(x_l | M_k) = 1/n_4^k. \tag{8}$$

*Node 5:* The width of the fault in the position $x_l$ is divided into equal parts that define $N_5$ possible centres $y_m$ along dip. Similarly to node 4, some of the defined points have a null probability of occurrence, since their distance from the fault edge is too small (i.e. $< \langle W_k/2 \rangle$). Also in this case, we assume that the remaining $n_5^k$ points are equally probable:

$$P_5\,(m, k) \stackrel{\text{def}}{=} Pr\,(y_m | M_k) = 1/n_5^k. \tag{9}$$

In setting up the probabilities for nodes 4–5 we implicitly assume that the tectonic rate is constant along fault strike, a roughly constant

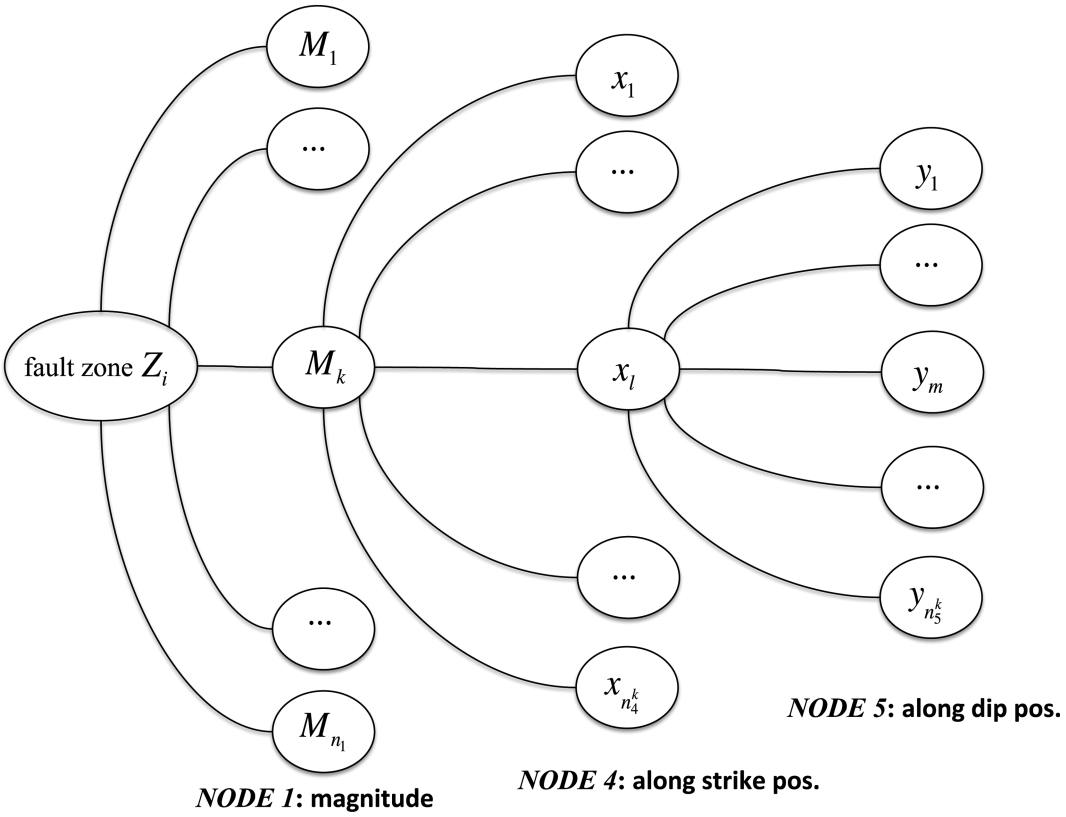


**Figure 1.** Sketch of the simplified event tree (ET). Graphical representation of the simplified ET that includes only NODES 1, 4 and 5, that is, earthquake magnitude and position of the geometrical centre. Symbols are defined in the text. Modelled earthquake magnitude is the minimum (most probable) magnitude, $M_k$, in each $I_k = [M_k, M_{k+1}[$ interval.

seismogenic width, and constant frictional properties over the fault plane; that is, the slip rate for unit area is constant over the fault plane.

*Node 6:* Similarly to nodes 2–3, a further simplifying assumption is made here by taking the slip instantaneous, constant and equal to its average value over the rupture area $\langle S_k \rangle = m_k/(\langle L_k \rangle \times \langle W_k \rangle \times \mu)$, where $\mu$ is the assumed constant rigidity value, and $m_k$ is the seismic moment corresponding to magnitude $M_k$. The rake angle is taken equal to the expected value $r$ from tectonic data for the source zone $Z_i$. Also at this node the probability is thus equal to 1:

$$P_6(l, m, k) \stackrel{\text{def}}{=} Pr\left(\langle S_k \rangle, \langle r \rangle | x_l, y_m, M_k\right) = 1. \tag{10}$$

The probability of $\sigma_j$, for $j = 1, \ldots, N_\sigma$, that is, the second term of eq. (5), can then be evaluated by taking the product of eqs (6)–(10):

$$Pr(\sigma_j | Z_i) = \prod_{n=1}^{6} P_n = \frac{\Phi(m_{k+1}) - \Phi(m_k)}{n_4^k n_5^k}, \quad \text{and}$$
$$N_\sigma = \sum_k n_4^k n_5^k. \tag{11}$$

With the discretisation scheme and the simplifying assumptions adopted in our example, each scenario that has a magnitude $M \in I_k$ is modelled by assigning to it the magnitude $M_k$, its best guess geometrical size, and uniform average slip, so that each earthquake scenario $\sigma_j$ is operatively defined by the set of parameters $[M_k, \langle L_k \rangle, \langle W_k \rangle, x_l, y_m, \langle S_k \rangle, \langle r \rangle]$. Each path within the ET thus characterizes a possible seismic event (the scenario $\sigma_j$) through a controlled discretisation of the parameter space. Each path represents a class of events, that is those included in the same 6-D interval represented by the above parameter set. Actually, only magnitude and position are explored here. The resulting ET is a simplified case for the purpose of illustrating the methodology, and it reduces to the three levels constituted by the nodes 1, 4 and 5 (Fig. 1).

## 4 'LINEAR' SPTHA RESULTS

In this section, the use of the ET is illustrated by means of a case study. A regional SPTHA is achieved in terms of maximum wave height offshore, using linear tsunami modelling. Fig. 2(a) shows the section of the Hellenic Arc used as the source zone for our case study in the Mediterranean Sea. Figs 2(b) and (c) are a zoom on the target sites and the telescopic nested grids of increasing resolution that will be used later on for (non-linear, higher resolution) inundation calculations. All of the details concerning the setup of this case study are reported in the Supporting Information, both regarding the source zone characterization, and the simulation setup for SPTHA calculations.

The Hellenic Arc subduction zone was chosen because it is generally considered capable of hosting great tsunamigenic earthquakes, such as the 365 AD, $M$8+, Crete Earthquake (Guidoboni *et al.* 1994; Papazachos & Papazachou 1997; Shaw *et al.* 2008). This tsunami was observed in at least 11 localities all over the Mediterranean Basin (NGDC/WDS Global Historical Tsunami Database). Its deposits have been found on the eastern Sicily coast and on the Mediterranean Sea bottom (De Martini *et al.* 2010; Polonia *et al.* 2013). This zone has already been considered by many scenario-based tsunami hazard studies including, for example, sites on the

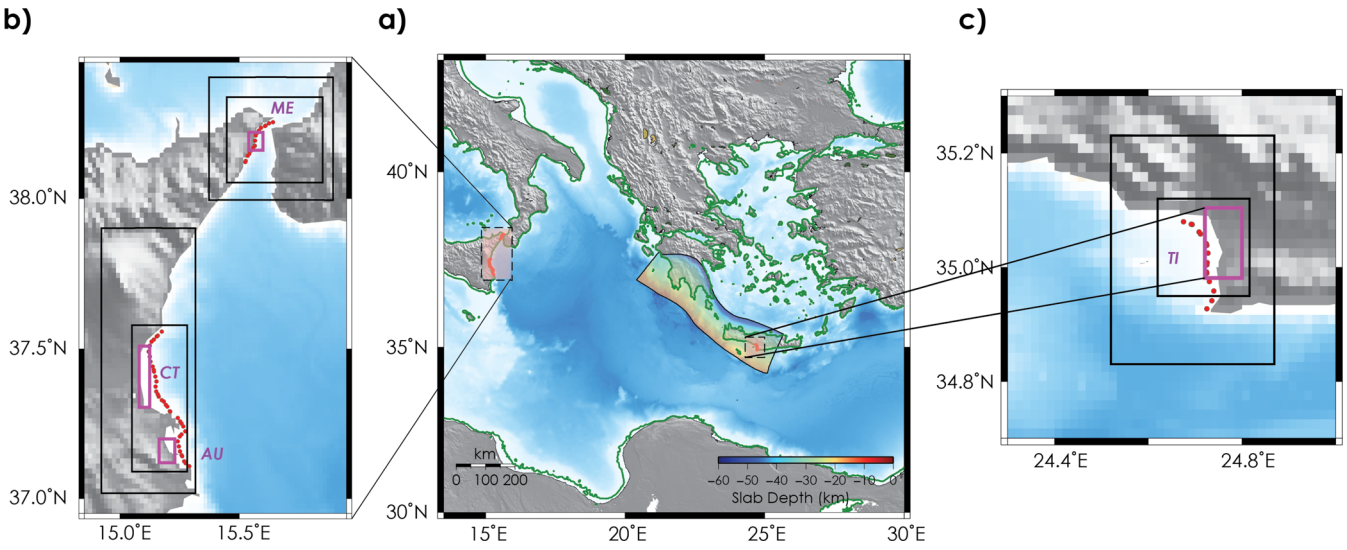


**Figure 2.** Fault zone and target sites for the Hellenic Arc case study. Fig. (a) (in the middle) shows the considered segment of the subduction interface, and the domain used for non-linear inundation calculations. The latter is embedded in the much larger computational domain used for linear calculations (see Fig. 3). Fig. (b) (to the left) is a close-up view showing the telescopic nested grids used for the Messina (ME), Catania (CT) and Augusta (AU) target sites, along the Eastern Sicily coast. Fig. (c) (to the right) is a close-up view around the Timpaki (TI) target site in Crete. For all target sites, the intermediate grids are indicated by black bounding boxes, whereas the highest resolution grids, those used for inundation calculations, are plotted in purple. The control points on the 50 m isobaths offshore of each target site are plotted in red.

Italian coasts (Tinti *et al.* 2005; Lorito *et al.* 2008a; Tiberti *et al.* 2008; Tonini *et al.* 2011; Basili *et al.* 2013).

SPTHA due to this source zone requires calculating all the terms in eq. (5). The annual rate $\lambda_{Z_i}$ is estimated essentially from the seismicity of the Hellenic Arc and from other assumptions such as the maximum magnitude (see Supporting Information). The probability $Pr(\sigma_j|Z_i)$ of each scenario is obtained by eq. (11).

The term $Pr(\psi \geq \bar{\psi}|\sigma_j;\boldsymbol{x})$ can be evaluated through numerical simulation of the tsunami generated by the $N_\sigma$ scenarios $\sigma_j$. The number $N_\sigma$ of modelled independent sources is equal to the number of ET branches with non-zero probability included in eq. (11). Despite the simplifications we made at some nodes in the ET, this number can still be rather high. Therefore, the computational cost of full inundation models for each node, and in case one should include several different source zones, can become a very serious issue. Yet, the computational cost remains reasonable for linear simulations of the maximum wave height $HMax$ offshore, which we set as our intensity measure $\psi$. In fact, relatively coarse grids can be employed. Moreover, instead of performing one tsunami simulation for each of the $N_\sigma$ scenarios, linearity can be exploited using the common approach of Green's functions summation. The $HMax_j$ due to the scenario $\sigma_j$ are here obtained as a linear combination of the tsunamis due to elementary sources that approximate $\sigma_j$. At each location $\boldsymbol{x}$, $HMax_j(\boldsymbol{x})$ is checked against a chosen threshold $\overline{HMax}$ and

$$Pr(\psi \geq \bar{\psi}|\sigma_j;\boldsymbol{x}) = \mathcal{H}\left(HMax_j(\boldsymbol{x}) - \overline{HMax}\right), \tag{12}$$

where $\mathcal{H}(HMax_j(\boldsymbol{x}) - \overline{HMax})$ is the Heaviside function, being equal to 1 for $HMax_j(\boldsymbol{x}) \geq \overline{HMax}$, and 0 otherwise. As stated above, in this way, uncertainties due to numerical modelling, for example, deriving from the use of shallow water approximation, and/or those related to bathymetry models, are completely neglected.

Fig. 3 shows the simulation domain for linear calculations, the chosen 50 m isobaths that are used as target points in the entire Mediterranean basin (Aegean Sea excluded), and the linear SPTHA results expressed as the annual exceedance probability for different thresholds $\overline{HMax}$, ranging from 0.25 to 8 m. Note that $HMax_j$ values are calculated on the 50 m isobaths and then extrapolated to 1 m depth adopting the widely used Green's law approximation. This is a quite crude approximation of the expected tsunami runup, as it ignores focusing/defocusing effects, and it is generally also a quite conservative choice, as attenuating effects such as bottom friction and wave breaking are neglected. This adds up to the use of shallow water equations, which is likely to be a conservative approach as well. Other convenient approaches to parametric run-up estimation have been recently explored in order to avoid computationally expensive simulations of the inundation process (Løvholt *et al.* 2012).

The regional SPTHA from this section of the Hellenic Arc produces local maxima on the Greek coast in the near-field and along the African coast between 20 and 25°E, where the tsunami energy is focused. The strike of the Hellenic Arc poses this African stretch of coast on a bearing roughly perpendicular to the broad side of the modelled earthquake ruptures. Additionally, tsunamis of weak to medium intensity might also affect the entire central and eastern Mediterranean Basin, with an annual frequency of at least 0.001–0.002 (Fig. 3), corresponding to an ARP of 500–1000 yr, or even with a higher frequency at least in the central Mediterranean. According to these calculations, if considering ARPs on the order of 1000 yr or more, catastrophic tsunamis featuring waves of one to several meters, originating in this source zone, may hit almost the entire central and western Mediterranean.

## 5 TWO-STAGE LINEAR SPTHA FILTERING

Our aim now is to reduce the total number of simulations required to obtain inundation maps, by excluding the sources that are not relevant for a specific target site or whose contribution to hazard can be approximated by sources producing similar effects. The selection is based on the analysis of offshore $HMax$ probability.

We focus on the four target sites shown in Fig. 2, namely: Messina (ME), Catania (CT) and Augusta (AU), on the eastern Sicily coast;

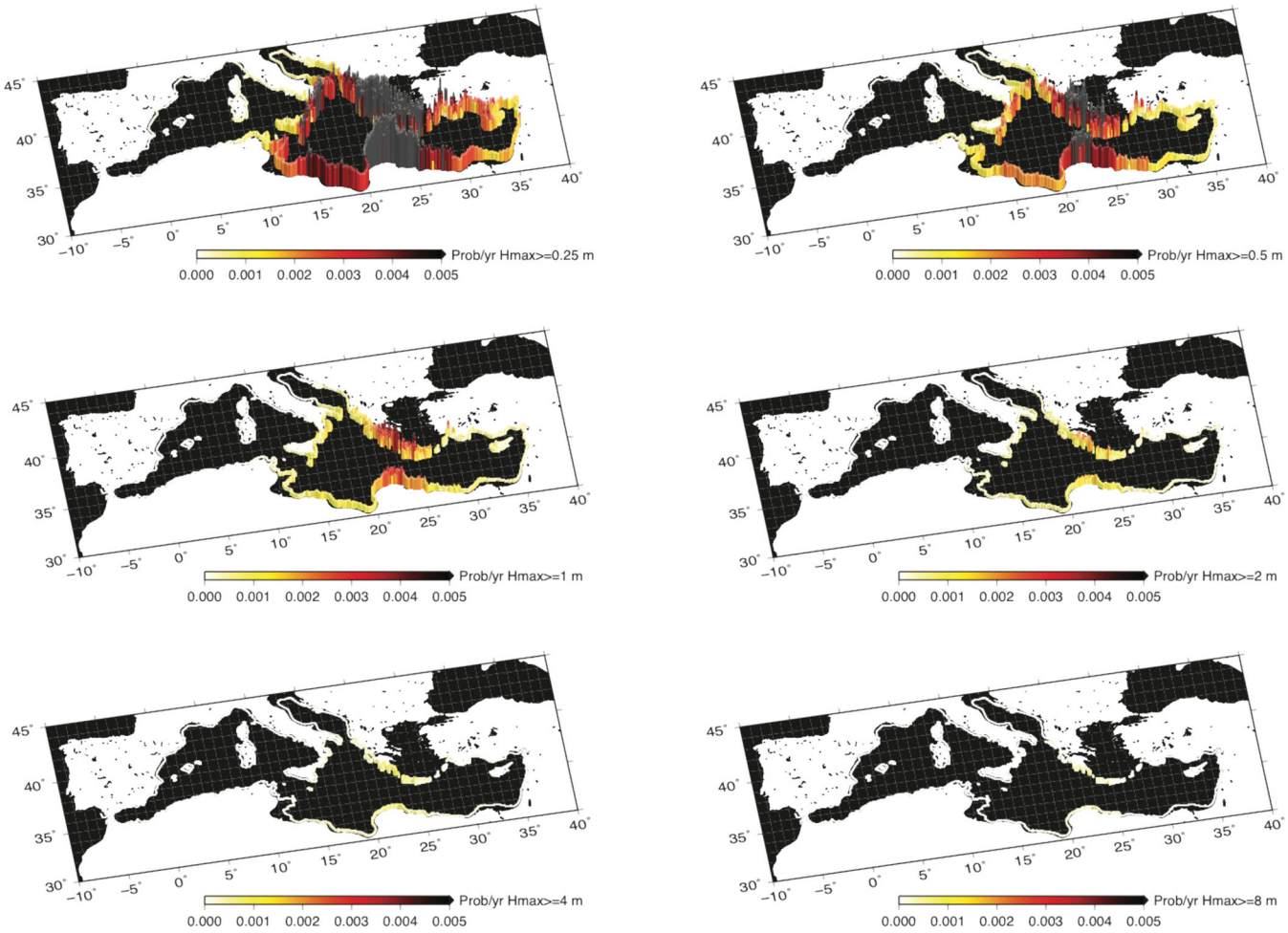


**Figure 3.** Probability of exceeding different offshore $HMax$ thresholds for the considered portion of the Hellenic Arc. The probability values are those obtained with the full set of subduction earthquake scenarios considered in our case study. The values represented by the coloured bars along the 50 m isobaths are those after correction with the Green's law amplification factor. The domain shown is that used for linear calculations of tsunami propagation.

Timpaki (TI) on the southern coast of Crete. These sites were chosen for testing the method under different tsunami propagation conditions associated with different source–target pairs. The relative orientation of the source–target pairs is quite variable among the three sites on the Sicilian coast. Furthermore, tsunamis generated on the Hellenic Arc subduction zone would have diverse paths to the sites, such as through the Messina Straits. The site responses are also expected to be different, because AU and ME harbours are different, and CT is a coastal plain. The Timpaki (TI) site is considered since the source variability in the near-field is expected to have a relatively stronger control on the hazard results with respect to tsunami propagation (Geist 2002).

A target site is operatively defined by selecting a number of $Q$ consecutive 'control' points $\boldsymbol{x} \in X_Q^{\text{target}}$ along the 50 m isobaths, and the corresponding set of $HMax_j(\boldsymbol{x} \in X_Q^{\text{target}})$ values already obtained for each source scenario $\sigma_j$. At each selected site we apply the two-stage filtering procedure described in what follows.

For the first stage, our assumption is that an overall negligible $HMax$ on the set of offshore control points leads to negligible inundation at the nearby coast (Filter 1). We thus filter out all sources that would produce such a negligible $HMax$ profile. To this end we construct an empirical cumulative function of the absolute maxima of the $HMax_j(\boldsymbol{x} \in X_Q^{\text{target}})$ profile due to each scenario at a given test site, and discard those that are below a chosen threshold $\overline{\overline{HMax}}$. Only the subset of sources $\{\sigma_j^1\}$ which 'pass' through Filter 1, are then carried on, that is

$$\{\sigma_j^1\} \stackrel{\text{def}}{=} \left\{\sigma_j : \max_x \left[HMax_j\left(\boldsymbol{x} \in X_Q^{\text{target}}\right)\right] > \overline{\overline{HMax}}\right\}. \qquad (13)$$

For all the discarded scenarios, we just set $Pr(\psi \geq \bar{\psi}|\sigma_j; \boldsymbol{x}) = 0$ in eq. (5). The $HMax_j$ corresponding to selected scenarios $\{\sigma_j^1\}$ will be hereafter indicated as $Hmax_j^1$.

In our example, we set the threshold $\overline{\overline{HMax}} = 0.25$ m (Fig. 4a), thereby obtaining a first significant reduction of the scenarios at each of test site: on the order of 20 per cent for all of the far-field Sicily sites (ME, CT and AU), and on the order of 10 per cent for the TI near-field site on Crete, where larger tsunami amplitudes can be expected.

In the second stage (Filter 2), we perform a hierarchical cluster analysis (HCA) on the set of sources $\{\sigma_j^1\}$ which passed Filter 1. The basic assumption is that similar $HMax_j(\boldsymbol{x} \in X_Q^{\text{target}})$ profiles along a sufficiently extended set of control points in front of the target coast will correspond to a roughly similar inundation pattern. Thus, if clusters can be found among sources $\{\sigma_j^1\}$ that generate similar enough $Hmax_j^1$ profiles, then one single simulation for each cluster

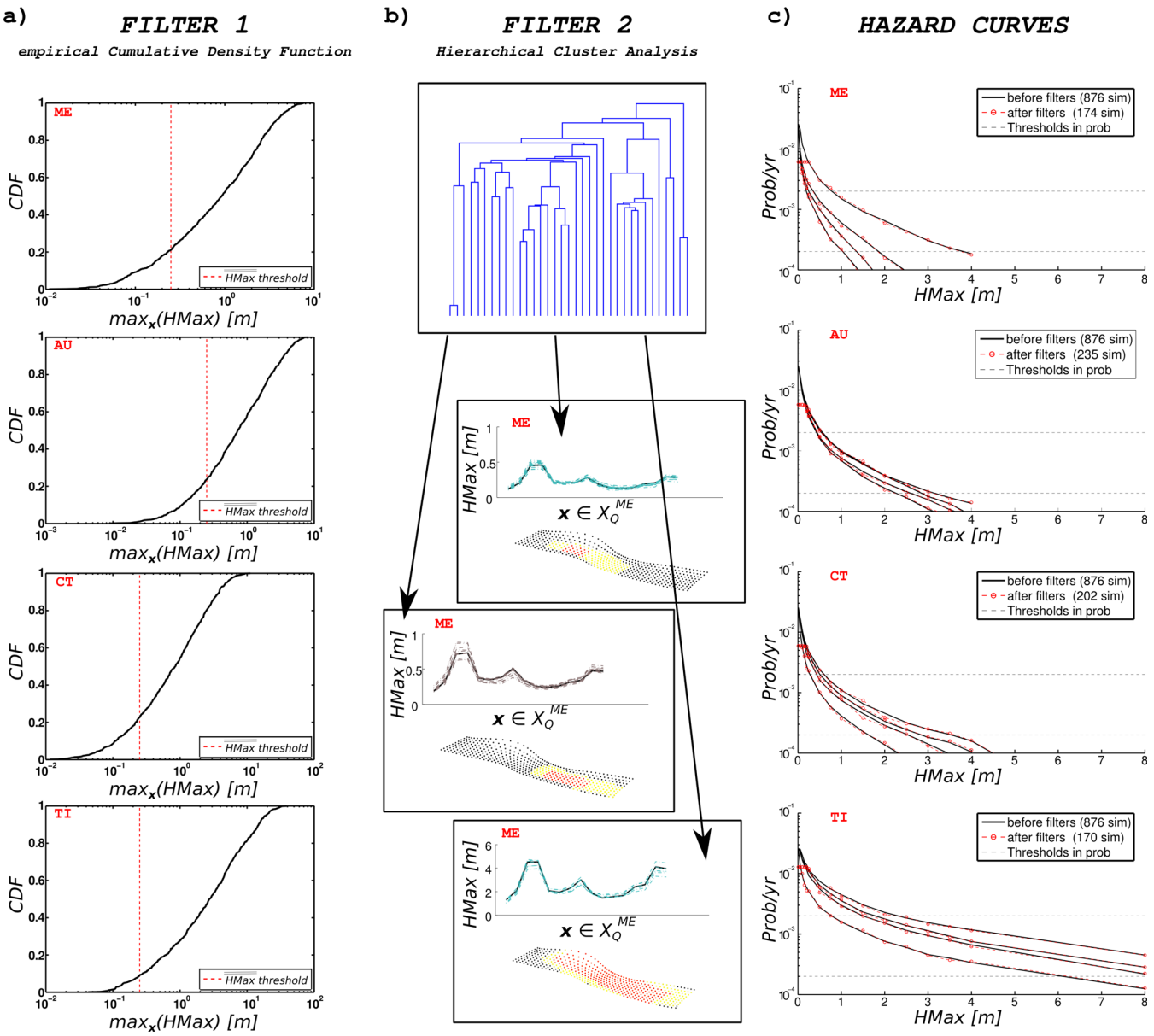


**Figure 4.** Illustration of the filtering procedure for the selected target sites. Fig. (a) shows the empirical cumulative density functions (CDF) at the target sites of Messina (ME), Catania (CT), Augusta (AU) and Timpaki (TI). For each source, the CDF is obtained with the corresponding maximum of the $HMax$ values along the offshore control profiles of Fig. 2. A reduction on the order of 20 per cent is obtained by applying Filter 1 with the chosen threshold of 0.25 m. Fig. (b) shows a typical branching graph resulting from the clustering analysis and a few examples of the $\sigma_j^c$ sources (red dots) representing one cluster $c$ of sources of the same magnitude (area enclosed by yellow dots) after application of Filter 2 (see eq. 15). Fig. (c) compares hazard curves for offshore control points at the target sites before (black curves) and after the two-stage filtering (red dashed curves with dots). The curves obtained with only 20–25 per cent of the full set of earthquakes are very similar to those calculated with the full set. The probability thresholds are those corresponding to ARPs of 500 and 5000 yr used for the inundation maps in Figs 5 and 6.

may be chosen as representative of any source belonging to that cluster. The basics of this cluster analysis method can be found in Anderberg (1973) and Hartigan (1975); below we indicate only a few necessary details of our analysis.

First, we recognize that earthquake magnitude controls $HMax$ to first order. Therefore, separate HCA are performed for each magnitude (one for each $k$ of eq. (6)). A 'cluster distance' needs to be defined among the profiles $Hmax_j^1(\boldsymbol{x} \in X_Q^{\text{target}}, k)$. We choose a distance $d_{CF}$ that has been previously used as a cost function for non-linear inverse problem, and that is known to be sensitive to both amplitude and phase of the profile (Piatanesi & Lorito 2007; Cirella *et al.* 2008; Lorito *et al.* 2008b, 2010, 2011; Romano *et al.* 2010, 2012; Cirella *et al.* 2012):

$$d_{CF}\left(HMax_u^1, HMax_v^1\right) = \left[1 - \frac{2\sum_t HMax_{u,t}^1 HMax_{v,t}^1}{\sum_t (HMax_{u,t}^1)^2 + \sum_t (HMax_{v,t}^1)^2}\right]. \quad (14)$$

Here, $t$ runs over the $\boldsymbol{x} \in X_Q^{\text{target}}$ coordinates of the $HMax^1$ profiles generated by the $u$-th and $v$-th source scenario, which have the same magnitude $M_k$. Sensitivity tests based on linear SPTHA

residuals show that this distance works better in our application than the simple Euclidean distance.

As in all HCA, there does not exist a closed-form method to select the ideal number of clusters (Priestley 1981), such that this number is small enough to avoid overfitting and large enough to avoid grouping profiles that differ significantly from each other. To define this number, we adopt the so-called Beale test. For each potential number of clusters $C^k$ (from 1 to the total number of scenarios with magnitude $M_k$), we evaluate the within-cluster variance (Priestley 1981; Davis 2002). This variance monotonically decreases with increasing $C^k$, but decreases more rapidly with new clusters that 'explain' the data significantly better than when fictitious clusters (essentially due to noise) are present. The optimal number of clusters is then defined as the 'knee' of this curve (Tarassenko 1998). In order to completely automate the selection, we analyse the intracluster variance as a function of the number of clusters with a change point analysis scheme based on the Kolmogorov–Smirnov tests (Mulargia & Tinti 1985; Mulargia *et al.* 1987), and the optimal $C^k$ is then selected as the smallest among the significant change points found at a confidence level $\alpha$ of 0.01.

Once the optimal number of clusters is found, one representative event source for each cluster, $\sigma_j^c$, is defined as that which has the closest profile to the cluster centroid, that is, the one that has the minimum distance $d_{CF}$ from the average profile in the cluster $c$. The set of sources passing through Filter 2 is then:

$$\left\{\sigma_j^2\right\} \stackrel{\text{def}}{=} \left\{\sigma_j^1 : \sigma_j^1 \equiv \sigma_j^c, \forall c\right\}, \tag{15}$$

that is, a scenario $\sigma_j^1$ is promoted to $\sigma_j^2$ if it is the representative of a cluster $c$. This subset contains all the sources for which an explicit tsunami numerical simulation is required. Summarizing, after each filter, the SPTHA can be calculated using a decreasing number of simulations, that is, $\{\sigma_j\} \rightarrow \{\sigma_j^1\} \rightarrow \{\sigma_j^2\}$, and eq. (5) can be approximated as:

$$\begin{aligned}\lambda_i(\psi \geq \bar{\psi}; \boldsymbol{x}, Z_i) &= \lambda_{Z_i} \sum_{j=1}^{N_\sigma} Pr(\sigma_j|Z_i) Pr(\psi \geq \bar{\psi}|\sigma_j; \boldsymbol{x}) \\ &\underbrace{\approx}_{\text{Filter 1}} \lambda_{Z_i} \sum_{j \in \{\sigma_j^1\}} Pr\left(\sigma_j^1|Z_i\right) Pr\left(\psi \geq \bar{\psi}|\sigma_j^1; \boldsymbol{x}\right) \\ &\underbrace{\approx}_{\text{Filter 2}} \lambda_{Z_i} \sum_{j \in \{\sigma_j^2\}} \left[\sum_{\sigma_j^1 \in c} Pr\left(\sigma_j^1|Z_i\right)\right] \\ &\quad \times Pr\left(\psi \geq \bar{\psi}|\sigma_j^2; \boldsymbol{x}\right). \end{aligned} \tag{16}$$

In practice, only one source for each cluster will be used for inundation modelling. The probability of occurrence for the entire set of scenarios belonging to cluster $c$ is then attributed to the representative source event scenario $\sigma_j^c$.

We now compare the linear SPTHA curves obtained with the complete set (no filters) with those obtained with the selected set of scenarios (after Filter 2), for each target location. In this way, we check the sensitivity of the results to our approximation regarding source selection. In the next section, we make the same comparison for inundation maps, to assess the extent at which the offshore $HMax$ probability can be used to guide the subsequent SPTHA based on non-linear inundation simulations, and to assess the order of magnitude and other features of the uncertainty induced by this source selection.

Fig. 4(b) shows a few examples illustrating the cluster analysis. Examples of $HMax$ profiles that were grouped together in a single cluster $c$ are shown, along with the extension of the representative source $\sigma_j^c$, compared to the zone covered by all the sources in the cluster. By applying this method to the case study, we obtain an overall reduction on the order of 75–80 per cent, with the combined action of the two filters, irrespective of the near- or far- field location of the target site. Fig. 4(c) shows the linear SPTHA expressed as the annual probability of exceedance as a function of the wave height, for some control points in front of the target sites. The SPTHA curves obtained with just the selected sources or with the complete set of sources are almost the same despite the substantial reduction of the scenarios behind them. This could lead to consideration of a further reduction by increasing the threshold of 0.25 m used for Filter 1 or by allowing a larger intra-cluster variability (increasing the size of the clusters). However, this cannot be done without first looking at the effect of filtering on the inundation maps in the next section, to see if they are stable and well approximated by the reduced set of tsunami simulations. Nevertheless, we argue that there might be a dependency of the optimal Filter 1 threshold on the ARP to be considered for a specific application. Here, we have fixed this threshold deterministically: because we never consider $HMax$ values lower than that in hazard curves, and because we assume that a 0.25 m offshore $HMax$ does not lead to significant inundation. Therefore, there is no relation here with the ARP to be considered. However, higher earthquake magnitudes have longer ARPs and, at first order, they induce larger tsunamis. Then, choosing a higher threshold for long enough ARPs, might optimize the analysis. Conversely, since the cluster analysis is performed for each magnitude interval, the Filter 2 is less dependent on the ARP of interest. As already stated, ET sampling could instead be a function of the chosen ARP, which could in turn indirectly improve the efficiency of the cluster analysis.

In order to achieve stable results, we had to tune also the length of the control profile $\boldsymbol{x} \in X_Q^{\text{target}}$. Note that the initial size $Q$ of $X_Q^{\text{target}}$—the length of the profile in front of the coast—essentially depends on the extension and morphology of the target site. A compromise has then to be reached, because we empirically found that the longer the ARP of the scenario under scrutiny, the longer is the optimal profile. The relation between the length and the spatial step of the points forming the control profile with the tsunami and/or $HMax$ profile typical wavelengths could be worth of further investigation. However, an overall good approximation was found for all the target sites using a profile slightly longer than the involved coastal stretch.

## 6 INUNDATION MAPS

In this section, we analyse the performance of the filters by comparing the inundation maps obtained with only the selected scenarios with those obtained with the full set. Due to the various simplifying assumptions we made, and to the fact that only one source zone was used, we keep this analysis at a quite basic level, in order to avoid overinterpretation of results. In addition, we used in some cases quite inaccurate digital elevation models, as better data were unavailable to us, with the only exception, perhaps, of the ME target site. Consequently, the inundation maps presented and discussed here are not ready for operational purposes; these results are only a test of the proposed methodology.

Figs 5 and 6 present inundation maps for maximum flow depth and maximum volume flux (current speed times flow depth), respectively, at the Messina (ME) target site before and after the two-stage filtering. They also show the residuals between the two in both map

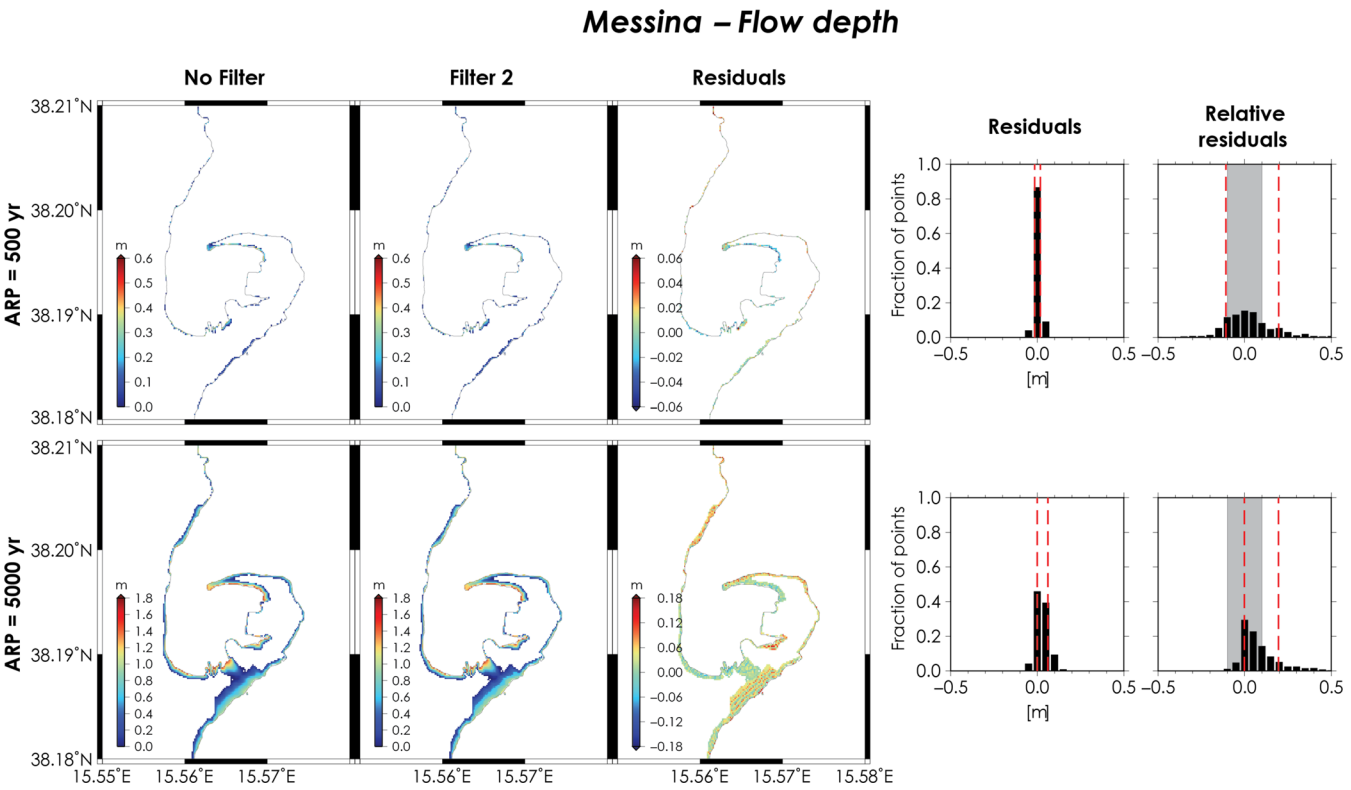


**Figure 5.** Flow depth inundation hazard maps for 500 and 5000 yr ARPs at Messina harbour. The maps in the first column are those corresponding to the full (unfiltered) set of sources, that is the same set of sources used to calculate the linear hazard shown in Fig. 3. The second column shows the maps obtained with the sources selected after filtering, and the third column presents the residuals between the two. The residuals are also shown as histograms of the absolute and relative differences; red dashed lines indicate the 16th and 84th percentiles and grey shaded areas indicate the region where differences are within ±10 per cent. The maps presented are a close-up of the most inundated zone around Messina harbour. The inundation map for the full zone corresponding to the purple rectangle in Fig. 2 is presented in the Supporting Information, along with those for the other target sites.

and histogram format. All the results are plotted for two different ARPs, 500 and 5000 yr; in other words, they are the maximum flow depth values that have an annual exceedance probability of 0.002 and 0.0002, respectively. The residuals are obtained by subtracting the unfiltered hazard intensity from the filtered hazard intensity (Filter 2 minus No Filter), at each 'wet' point on one of the two maps. Thus, a positive residual means that filtered maps overestimate the hazard, and vice versa. The relative residuals, only shown as histograms, are obtained by dividing the residuals by the unfiltered value. Note that Figs 5 and 6 show a close-up view on the most inundated part within the innermost (highest resolution) grid for the ME case, that is around the Messina harbour. The inundation maps for the entire innermost grid of all target sites (ME, CT, AU and TI) are reported in Figs S6–S13.

At first glance, we note that the overall inundation pattern is left unchanged by the filtering. In particular, a visual comparison between maximum inundation lines indicates that they remain stable or change only slightly for all the cases here considered. Also, significant differences between the values are mainly limited to rather isolated points. This is quite satisfactory, given the inherent uncertainty in modelling of inundation maps, whose finest details are seldom used for operational purposes.

It is important to recall that we noted a dependence of the results on the length of the control profile. We found that, after adjusting the control profile in order to obtain a stable approximation for linear SPTHA curves (Fig. 4c), also the corresponding inundation map remains stable. This is not shown here, as too many plots would be needed to illustrate this feature. Therefore, some effort is needed for finding the optimal length for the profile, but then the stability of inundation maps can be reasonably well predicted by only analysing the offshore $HMax$ probability.

The analysis of the histograms of the residuals, and of the relative residuals, indicate that the former are generally less dispersed than the latter, at least for target sites in the far-field of the source. In the far-field, the residuals are more significant for the shortest ARP of 500 yr. The inundation in this case is quite limited in most cases, and this may cause some instability in the histograms, due to the small sample size. Thus, the filtering seems to perform better where more sustained inundation probability is present due to the longer ARP. In the near-field of the source, the situation is quite the opposite, as some extra dispersion of the residuals appears for the longer 5000 yr ARP.

Taken together, the above results indicate that the procedure is able to capture the overall features of the inundation maps, even if they have been obtained with a limited set of scenarios. Some fine-tuning of the filters is needed not only depending on the ARP of interest, but also considering whether the target site is in the near-field or far-field of the causative source.

## 7 DISCUSSION

The procedure proposed in this study focuses on a careful treatment of the aleatory variability of the seismic source; however, we made several simplifying assumptions. Some (or even all) of these simplifications should be avoided for actual hazard assessments.

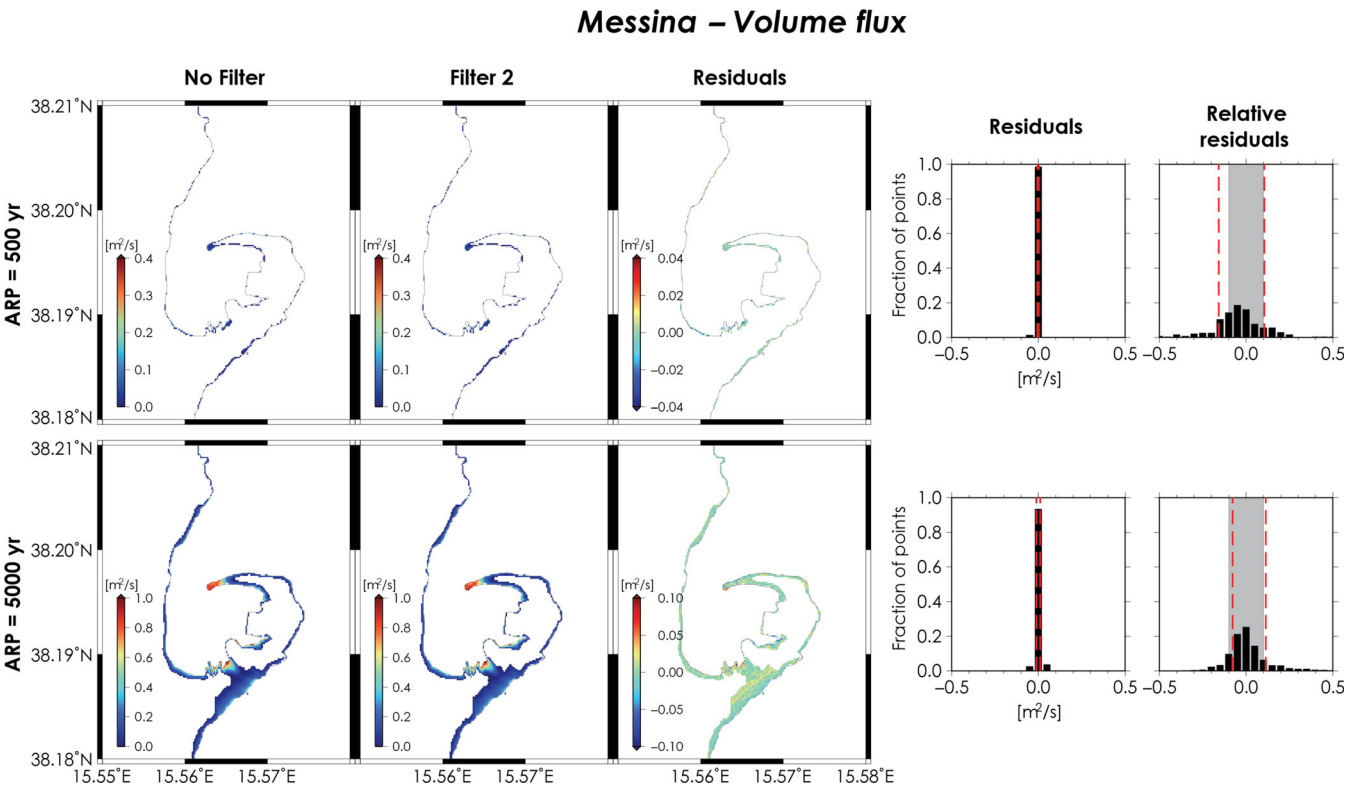


**Figure 6.** Volume flux inundation hazard maps for 500 and 5000 yr ARPs at Messina harbour. Same as Fig. 5, for the volume flux, that is speed times flow depth.

For example, exploring a synthetic catalogue of stochastic heterogeneous slip distributions would likely be necessary when dealing with run-up at coastal sites in the near-field of the seismic source (Geist 2002; McCloskey *et al.* 2007; González *et al.* 2009; Thio *et al.* 2010).

Here, we propose the ET as a tool for the controlled factorization and discretisation of the parameter space. For example, exploring low probability events (e.g. great or mega earthquakes with longer ARPs) with more classical approaches, such as Monte Carlo, may result in exceedingly large synthetic earthquake catalogues. Conversely, with the ET approach, the tails of the assumed PDF can be efficiently explored. This is important because the parameters (e.g. size, peak slip) of 'extreme' tsunamigenic events, such as Sumatra-Andaman 2004 or Tohoku 2011, are likely falling in these tails and, if so, one may think of them as responsible for quite heavy tails of tsunami hazard and risk distributions. The ET nodes may be also variously explored, for example depending on the ARP of interest for a specific application, by adjusting the ranges and the sampling steps as needed. Here we did not fully explore this possibility; instead we chose the source-parameter sampling scheme quite heuristically. As illustrated in the Supporting Information, the only exception is perhaps the uneven sampling of the magnitude, with an increasingly finer sampling at higher magnitudes. Also other parameters that we kept fixed to their expected values without exploring their aleatory variability could be subject to optimization, such as the earthquake geometrical parameters deriving from scaling laws.

A comparison between ETs with different levels of complexity may also provide a systematic assessment of the SPTHA sensitivity to the different source parameters (ET levels), either in general or for a specific source zone-target coast pair. A classic example to illustrate this is that the nodes accounting for the variability of the geometrical parameters, or the degree of heterogeneity of the slip distribution, are expected to have enhanced relevance with respect to earthquake magnitude in the near-field of the source. Such a hierarchization of the source parameters with respect to their influence on the hazard intensity would help optimize the sampling scheme, and a finer sampling could be adopted when hazard sensitivity is higher.

In order to obtain satisfactory results at any specific site by adopting the filtering procedure it is necessary to find an optimal length for the controlling set of offshore points on the 50 m isobath. This optimal length condition appears to be met when the offshore $HMax$ hazard curves are stable with respect to the source selection. The parameters of the filters, in addition to those of the ET, could be likely optimized as a function of the application. For example, the longer the ARP considered, the higher the threshold of Filter 1. Future studies might be oriented at developing a more thorough approach by considering current speed in conjunction with $HMax$ at offshore profiles. We have also noted that an interdependence between the ET and filter optimization might be envisaged.

We demonstrated that our source-filtering procedure preserves the fundamental features of inundation maps reasonably well, both at near- and far-field distances from the source. At this point of the analysis, however, we are not able to say how our modelling of flux might be improved. It may be that the basic shallow water assumption is less appropriate for tsunami currents or, given the higher spatial variability of currents, that a better topographic model and a finer grid are required. Improvement might also be achieved by incorporating (more densely spaced) flux profiles into the source filtering procedure, which currently employs only wave height profiles on coarsely spaced offshore points.

In addition to treating aleatory uncertainty as discussed above, a thorough SPTHA study also requires consideration of epistemic uncertainty (e.g. SSHAC 1997). We thus discuss some major sources of epistemic uncertainty underlying our case study and that should be considered for operational assessments. For example, we ignored epistemic uncertainties related to the parameters of the assumed probability functions for the aleatory variables (e.g. the parameters of the Pareto CDF, or the earthquake scaling law), and the geometrical parameters of the source zone. We also assumed that the convergence rate and the depth limits of the seismogenic zone are constant on the fault plane; that is to say, we assumed a uniform slip probability everywhere. Another issue not addressed here is the likely depth-dependent rigidity of the fault (Bilek & Lay 1999). As noted by Blaser *et al.* (2010), if there is a systematic relationship between fault area and seismic moment, stress drop should be constant, and then slip and rigidity should be inversely proportional.

In addition to these well-recognized sources of epistemic uncertainty there are others sources of uncertainty that sometimes receive little or no attention. A few examples include the uncertainties related to the limited knowledge of potential offshore sources, particularly in complicated tectonic contexts (Basili *et al.* 2013; Matias *et al.* 2013); to modelling of the tsunami generation mechanism (Kervella *et al.* 2007; Nosov & Kolesov 2011; ASCETE project, http://www.ascete.de, accessed 14 November 2014); and to the possibility of activation of secondary structures, such as splay faults (Wendt *et al.* 2009). In particular, the limited knowledge on offshore sources may seriously affect SPTHA, especially in the near-source region. Less constrained offshore structures capable of generating earthquakes may exist. For example, several crustal source zones with hardly predictable geometrical parameters are known to lie in the Mediterranean. Oversimplified assumptions that just ignore such sources would affect the 'seismic source coverage', thus introducing uncontrolled biases in the SPTHA. Also, when dealing with reasonably well-constrained faults, there may still be poorly known parameters that are likely to control the results to a very first order, thus representing a significant challenge even in PSHA studies (Stein *et al.* 2012; Kagan & Jackson 2013). One example is the almost never well-constrained maximum magnitude for a given fault $M_{Z_i}^{\mathrm{MAX}}$ (Holschneider *et al.* 2014; Rong *et al.* 2014), which has first order control on SPTHA estimates. Another important example is the rate of activity or the slip rate, that here we derived quite heuristically from observed seismicity and from geodynamic considerations. To address these potential limitations all operational SPTHA studies should be integrated with a comparison to all available evidence at the target site, such as paleo- and historical tsunami recurrence rates and hazard intensity estimates (González *et al.* 2009; Parsons & Geist 2009).

Finally, we need to mention that we disregarded epistemic uncertainties related to tsunami numerical modelling, such as those in the generation, propagation and inundation model, and those due to the potential inaccuracy of the topo-bathymetric data set.

Various strategies exist for incorporating epistemic uncertainty into hazard studies; for example, the Bayesian approach (Marzocchi *et al.* 2010), the expert elicitation (Selva *et al.* 2012) or the classical LT approach (Geist & Parsons 2006; Annaka *et al.* 2007; Basili *et al.* 2013). In the Bayesian and expert elicitation approaches, epistemic uncertainty evaluation is addressed contextually at each ET node or level. Conversely, in the LT case, each node corresponds to a set of modelling assumptions, and for each node of the LT a complete modelling chain (including a specific ET) should be developed.

## 8 CONCLUSIONS

We have presented a methodology for SPTHA that significantly lowers the computational cost of probabilistic inundation maps by reducing the number of required source scenarios without degrading significantly the quality of the results. The procedure consists of two modules: a controlled factorization and discretisation of source parameters, in a logical order, with a series of conditional probabilities in an ET; and a two-stage filtering scheme that reduces the number of scenarios necessary for the production of inundation maps.

Source selection is based on the expected tsunami effects at target sites rather than on similarities among the sources themselves. That is to say, the criteria for source selection and sampling are intentionally designed for SPTHA purposes instead of being borrowed from PSHA. In our approach the selection of sources is based solely on relatively inexpensive $HMax$ calculations on control offshore profiles in front of the coast. The process of filtering sources is then very cost-effective. It may also provide a tool not only for SPTHA, but also for appropriately selecting worst-case scenarios (e.g. for the definition of evacuation maps) or design scenarios for a specific target site, through de-aggregation of a limited set of inundation maps.

The proposed two-step filtering procedure is semi-automatic and can be easily repeated for different target locations before calculating inundation maps. In order to get satisfactory results at any specific site, however, it is necessary to find an optimal length for the controlling set of offshore points on the chosen isobath. The whole procedure, as regards ET sampling and tuning of the filters, can perhaps be optimized if results are needed for a specific ARP and/or specific source/site pairs.

For describing and testing the performance of our approach we simulated the occurrence of subduction earthquakes on a section of the Hellenic Arc and applied our SPTHA scheme to several locations on the coasts of eastern Sicily and southern Crete. We found that the method reduces the computational effort of inundation modelling by almost 80 per cent.

Further work is needed to compare the uncertainties introduced by this scheme with those inherent to the seismic source treatment and tsunami modelling commonly adopted for SPTHA. The discrepancies introduced by our approximation likely fall within epistemic uncertainties that are not considered here, such as those related to tsunami generation and propagation models, bathymetric and topographic models, or other basic and less constrained unknowns related to earthquake activity rates or slip distribution probability. We recommend that these epistemic uncertainties be thoroughly assessed in order to identify the acceptable level of discrepancy introduced by filtering before starting any operational SPTHA.

## ACKNOWLEDGEMENTS

We acknowledge M. Sørensen, M. Spada and A. Babeyko for fruitful discussions on SPTHA, and thank S. Vinci and F. Doumaz for helping to build some of the topo-bathymetric data sets used in this work. The manuscript was substantially improved thanks to the thoughtful reviews of Frank González and Eric Geist. The authors wish to thank also G. Valensise who assisted in the proof-reading of the manuscript. This work was funded by: the Flagship Project RITMARE—The Italian Research for the Sea—coordinated by the Italian National Research Council and funded by the Italian Ministry of Education, University and Research within the National Research Program 2011–2013; and the EC project ASTARTE—

Assessment, STrategy And Risk Reduction for Tsunamis in Europe. Grant 603839, 7th FP (ENV.2013.6.4–3 ENV.2013.6.4–3).

## REFERENCES


Anderberg, M.R., 1973. *Cluster Analysis for Applications,* Academic Press.

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.

Basili, R., Tiberti, M.M., Kastelic, V., Romano, F., Piatanesi, A., Selva, J. & Lorito, S., 2013. Integrating geologic fault data into tsunami hazard studies, *Nat. Hazards Earth Syst. Sci.,* **13,** 1025–1050.

Bazzurro, P. & Cornell, C.A., 1999. Disaggregation of seismic hazard, *Bull. seism. Soc. Am.,* **89,** 501–520.

Bilek, S.L. & Lay, T., 1999. Rigidity variations with depth along interplate megathrust faults in subduction zones, *Nature,* **400,** 443, doi:10.1038/22739.

Blaser, L., Krüger, F., Ohrnberger, M. & Scherbaum, F., 2010. Scaling relations of earthquake source parameter estimates with special focus on subduction environment, *Bull. seism. Soc. Am.,* **100**(6), 2914–2926.

Bommer, J.J. & Scherbaum, F., 2008. The use and misuse of logic trees in probabilistic seismic hazard analysis, *Earthq. Spectra,* **24**(4), 997–1009.

Burbidge, D., Cummins, P., Mleczko, R. & Thio, H., 2008. A Probabilistic tsunami hazard assessment for Western Australia, *Pure appl. Geophys.,* **165,** 2059–2088.

Cirella, A., Piatanesi, A., Tinti, E., Chini, M. & Cocco, M., 2012. Complexity of the rupture process during the 2009 L'Aquila, Italy, earthquake, *Geophys. J. Int.,* **190,** 1, 607–621.

Cirella, A., Piatanesi, A., Tinti, E. & Cocco, M., 2008. Rupture process of the 2007 Niigata-ken Chuetsu-oki earthquake by non-linear joint inversion of strong motion and GPS data, *Geophys. Res. Lett.,* **35,** L16306, doi:10.1029/2008GL034756.

Clifton, A. & Ericson, II, 2005. *Hazard Analysis Techniques for System Safety,* John Wiley & Sons, Inc., 528 pp.

Cornell, C.A., 1968. Engineering seismic risk analysis, *Bull. seism. Soc. Am.,* **58,** 1583–1606.

Cornell, C. & Krawinkle, H., 2000. Progress and challenges in seismic performance assessment, PEER Center News **3**(2), Available at: http://peer.berkeley.edu/news/2000spring/index.html, last accessed 14 November 2014.

Davis, J.C., 2002. *Statistics and Data Analysis in Geology,* John Wiley and Sons.

De Martini, P.M., Barbano, M.S., Smedile, A., Gerardi, F., Pantosti, D., Del Carlo, P. & Pirrotta, C., 2010. A unique 4000-year long geological record of multiple tsunami inundations in the Augusta Bay (eastern Sicily, Italy), *Mar. Geol.,* **276,** 42–57.

Geist, E.L., 2002. Complex earthquake rupture and local tsunamis, *J. geophys. Res.,* **107**(B5), doi:10.1029/2000JB000139.

Geist, E. & Parsons, T., 2006. Probabilistic analysis of tsunami hazards, *Nat. Hazards,* **37,** 277–314.

González, F.I. *et al.*, 2009. Probabilistic tsunami hazard assessment at Seaside, Oregon, for near- and far-field seismic sources, *J. geophys. Res.,* **114,** C11023, doi:10.1029/2008JC005132.

González, F.I., Le Veque, R.J. & Adams, L.M., 2013. Probabilistic tsunami hazard assessment (PTHA) for Crescent City, CA, Final Report for Phase I, Technical Report.

Gonzalez, M., Medina, R., Olabarrieta, M. & Otero, L., 2010. Tsunami hazard assessment on the Southern Coast of Spain, *Turkish J. Earth Sci.,* **19**(3), 351–366.

Grezio, A., Sandri, L., Marzocchi, W., Argnani, A., Gasparini, P. & Selva, J., 2012. Probabilistic tsunami hazard assessment for Messina Strait Area (Sicily, Italy), *Nat. Hazards,* **64,** 329–358.

Guidoboni, E., Comastri, A. & Traina, G., 1994. *Catalogue of Ancient Earthquakes in the Mediterranean Area up to the 10th Century,* Vol. 1, ING-SGA, 504 pp.

Hartigan, J.A., 1975. *Clustering Algorithms,* John Wiley & Sons, Inc..

Heidarzadeh, M. & Kijko, A., 2011. A probabilistic tsunami hazard assessment for the Makran subduction zone at the Northwestern Indian Ocean, *Nat. Hazards,* **56**(3), 577–593.

Holschneider, M., Zöller, G., Clements, R. & Schorlemmer, D., 2014. Can we test for the maximum possible earthquake magnitude?, *J. geophys. Res.: Solid Earth,* **119**(3), 2019–2028.

Horspool, N. *et al.,* 2014. A probabilistic tsunami hazard assessment for Indonesia, *Nat. Hazards Earth Syst. Sci. Discuss.,* **2,** 3423–3464.

Kagan, Y.Y., 2002. Seismic moment distribution revisited, I. Statistical results, *Geophys. J. Int.,* **148,** 520–541.

Kagan, Y.Y. & Jackson, D.D., 2013. Tohoku earthquake: a surprise?, *Bull. seism. Soc. Am.,* **103**(2B), 1181–1194.

Kanamori, H. & Brodsky, E.E., 2001. The physics of earthquakes, *Phys. Today,* **54**(6), 34–40.

Kervella, Y., Dutykh, D. & Dias, F., 2007. Comparison between three-dimensional linear and nonlinear tsunami generation models, *Theor. Comput. Fluid Dyn.,* **21,** 245–269.

Lane, E.M., Gillibrand, P.A. & Wang, X.A, 2013. Probabilistic tsunami hazard study of the Auckland region, Part II: inundation modelling and hazard assessment, *Pure appl. Geophys.*, **170**(9–10), 1635–1646.

Lay, T. & Kanamori, H., 2011. Insights from the great 2011 Japan earthquake, *Phys. Today,* **64**(12), 33–39.

Leonard, L.J., Rogers, G.C. & Mazzotti, S., 2014. Tsunami hazard assessment of Canada, *Nat. Hazards,* **70**(1), 237–274.

Liu, Y., Santos, A. & Wang, S.M., 2007. Tsunami hazards along Chinese coast from potential earthquakes in South China Sea, *Phys. Earth planet. Int.,* **163**(1–4), 233–244.

Lorito, S., Tiberti, M.M., Basili, R., Piatanesi, A. & Valensise, G., 2008a. Earthquake-generated tsunamis in the Mediterranean Sea: scenarios of potential threats to Southern Italy, *J. geophys. Res.,* **113,** B01301, doi:10.1029/2007JB004943.

Lorito, S., Piatanesi, A. & Lomax, A., 2008b. Rupture process of the 18 April 1906 California earthquake from near-field tsunami waveform inversion, *Bull. seism. Soc. Am.,* **98,** 832–845.

Lorito, S., Piatanesi, A., Cannelli, V., Romano, F. & Melini, D., 2010. Kinematics and source zone properties of the 2004 Sumatra-Andaman earthquake and tsunami: nonlinear joint inversion of tide gauge, satellite altimetry, and GPS data, *J. geophys. Res.,* **115,** B02304, doi:10.1029/2008JB005974.

Lorito, S. *et al.*, 2011. Limited overlap between the seismic gap and the coseismic slip of the great 2010 Chile earthquake, *Nat. Geosci.,* **4,** 173–177.

Løvholt, F., Pedersen, G., Bazin, S., Kühn, D., Bredesen, R.E. & Harbitz, C., 2012. Stochastic analysis of tsunami runup due to heterogeneous coseismic slip and dispersion, *J. geophys. Res.,* **117,** C03047, doi:10.1029/2011JC007616.

Maramai, A., Brizuela, B. & Graziani, L., 2014. The Euro-Mediterranean tsunami catalogue, *Ann. Geophys.,* **57,** S0435, doi:10.4401/ag-6437.

Marzocchi, W., Sandri, L. & Selva, J., 2010. BET_VH: a probabilistic tool for long-term volcanic hazard assessment, *Bull. Volcanol.,* **72**(6):717–733.

Matias, L.M., Cunha, T., Annunziato, A., Baptista, M.A. & Carrilho, F., 2013. Tsunamigenic earthquakes in the Gulf of Cadiz: fault model and recurrence, *Nat. Hazards Earth Syst. Sci.,* **13,** doi:10.5194/nhess-13-1-2013.

McCloskey, J. *et al.*, 2007. Near-field propagation of tsunamis from megathrust earthquakes, *Geophys. Res. Lett.,* **34,** L14316, doi:10.1029/2007GL030494.

Mitsoudis, D.A., Flouri, E.T., Chrysoulakis, N., Kamarianakis, Y., Okal, E. & Synolakis, C.E., 2012. Tsunami hazard in the southeast Aegean Sea, *Coast. Eng.,* **60,** 136–148.

Mulargia, F. & Tinti, S., 1985. Seismic sample areas defined from incomplete catalogs: an application to the Italian territory, *Phys. Earth. planet. Int.,* **40,** 273–300.

Mulargia, F., Gasperini, P. & Tinti, S., 1987. Identifying different regimes in eruptive activity: an application to Etna volcano, *J. Volc. Geotherm. Res.,* **34,** 89–106.

Newhall, C.G. & Hoblitt, R.P., 2002. Constructing event trees for volcanic crises, *Bull. Volcanol.,* **64,** 3–20.
NGDC/WDS, National Geophysical Data Center/World Data System, Global Historical Tsunami Database, National Geophysical Data Center, NOAA, doi:10.7289/V5PN93H7, Available at http://www.ngdc.noaa.gov/hazard/tsu.shtml, last accessed 14 November 2014.
Nosov, M.A. & Kolesov, S.V., 2011. Optimal initial conditions for simulation of seismotectonic tsunamis, *Pure appl. Geophys.,* **168**(6–7), 1223–1237.
Omira, R., Baptista, M.A., Matias, L.M., Miranda, J.M. & Carrilho, F., 2013. Probabilistic tsunami hazard in the North East Atlantic due to seismic sources, implications for NEAMTWS, in *Proceedings of the 2013 AGU Fall Meeting,* NH41B-1710.
Papazachos, B. & Papazachou, C., 1997. *The Earthquakes of Greece,* Ziti Publ., 304 pp.
Parsons, T. & Geist, E.L., 2009. Tsunami probability in the Caribbean region, *Pure appl. Geophys.,* **165,** 2089–2116.
Piatanesi, A. & Lorito, S., 2007. Rupture process of the 2004 Sumatra-Andaman earthquake from tsunami waveform inversion, *Bull. seism. Soc. Am.,* **97**(1A), S223–S231.
Polonia, A., Bonatti, E., Camerlenghi, A., Lucchi, R.G., Panieri, G. & Gasperini, L., 2013. Mediterranean megaturbidite triggered by the AD 365 Crete earthquake and tsunami, *Sci. Rep.,* **3**(1285), doi: 10.1038/srep01285.
Priestley, M.B., 1981. *Spectral Analysis and Time Series,* Elsevier.
Rikitake, T. & Aida, I., 1988. Tsunami hazard probability in Japan, *Bull. seism. Soc. Am.,* **78,** 268–1278.
Romano, F., Piatanesi, A., Lorito, S., D'Agostino, N., Hirata, K., Atzori, S., Yamazaki, Y. & Cocco, M., 2012. Clues form joint inversion of tsunami and geodetic data of the 2011 Tohoku-oki earthquake, *Sci. Rep.,* **385,** doi:10.1038/srep00385.
Romano, F., Piatanesi, A., Lorito, S. & Hirata, K., 2010. Slip distribution of the 2003 Tokachi-oki Mw 8.1 earthquake from joint inversion of tsunami waveforms and geodetic data, *J. geophys. Res.,* **115,** B11313, doi:10.1029/2009JB006665.
Romano, F. *et al.*, 2014. Structural control on the Tohoku earthquake rupture process investigated by 3D FEM, tsunami and geodetic data, *Sci. Rep.,* **4**, doi:10.1038/srep05631.
Rong, Y., Jackson, D.D., Magistrale, H. & Goldfinger, C., 2014. Magnitude limits of subduction zone earthquakes, *Bull. seism. Soc. Am.,* **104**(5), 2359–2377.
Satake, K, Fujii, Y., Harada, T. & Namegaya, Y., 2013. Time and space distribution of coseismic slip of the 2011 Tohoku earthquake as inferred from tsunami waveform data, *Bull. seism. Soc. Am.,* **103,** 1473–1492.
Selva, J., Marzocchi, W., Papale, P. & Sandri, L., 2012. Operational eruption forecasting at high-risk volcanoes: the case of Campi Flegrei, Naples, *J. Appl. Volcanol.,* **1**(5), doi:10.1186/2191-5040-1-5.
Shaw, B. *et al.*, 2008. Eastern Mediterranean tectonics and tsunami hazard inferred from the AD 365 earthquake, *Nat. Geosci.,* **1,** 268–276.
Sørensen, M.B., Spada, M., Babeyko, A., Wiemer, S. & Grünthal, G., 2012. Probabilistic tsunami hazard in the Mediterranean Sea, *J. geophys. Res.,* **117,** B01305, doi:10.1029/2010JB008169.
SSHAC [Senior Seismic Hazard Analysis Committee, R.J. Budnitz, Chairman, G. Apostolakis, D.M. Boore, L.S. Cluff, K.J. Coppersmith, C.A. Cornell & P.A. Morris], 1997. Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts. NUREG/CR6372, US Nuclear Regulatory Commission.
Stein, S., Geller, R.J. & Liu, M., 2012. Why earthquake hazard maps often fail and what to do about it, *Tectonophysics,* **562–563,** 1–25.
Suppasri, A., Imamura, F. & Koshimura, S., 2012. Probabilistic tsunami hazard analysis and risk to coastal populations in Thailand, *J. Earthq. Tsunami,* **6**(2), doi:10.1142/S179343111250011X.
Tarassenko, L., 1998. *A Guide to Neural Computing Applications,* Arnold.
Thio, H.K., Somerville, P. & Ichinose, G., 2007. Probabilistic analysis of strong ground motion and tsunami hazards in Southeast Asia, *J. Earthq. Tsunami,* **1**(2), 119–137.
Thio, H.K., Somerville, P. & Polet, J., 2010. Probabilistic tsunami hazard in California. Pacific Earthquake Engineering Research Center, PEER Report 2010/108, University of California, Berkeley.
Tiberti, M., Lorito, S., Basili, R., Kastelic, V., Piatanesi, A. & Valensise, G., 2008. Scenarios of earthquake-generated tsunamis for the Italian coast of the Adriatic Sea, *Pure appl. Geophys.,* **165,** 2117–2142.
Tinti, S., Armigliato, A., Pagnoni, G. & Zaniboni, F., 2005. Scenarios of giant tsunamis of tectonic origin in the Mediterranean, *ISET J. Earthq. Technol.,* **42,** 171–188.
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), *Nat. Hazards Earth Syst. Sci.,* **11,** 1217–1232.
Wendt, J., Oglesby, D.D. & Geist, E.L., 2009. Tsunamis and splay fault dynamics, *Geophys. Res. Lett.,* **36,** L15303, doi:10.1029/2009GL038295.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

**Table S1.** Subduction parameters for the adopted MFD.
**Table S2.** Parameters of sample ruptures.
**Figure S1.** Upper panel: Map of the Hellenic subduction zone showing colour-coded tessellation of the slab section analysed in this study. Convergence rates normal to mapped boundaries are from block modelling of geodetic data by Vernant *et al.* (2014). Lower panel: Generalized tectonic setting of the Mediterranean Sea. Legend for tectonic plates: EU, Eurasian; AF, African; AE, Aegean; AT, Anatolian; AR, Arabian; AD, Adria microplate. The square indicates the area mapped in the upper panel.
**Figure S2.** MFDs for various maximum magnitude (Mx) and coupling coefficients (C) compared to earthquake distribution from SHEEC. Maximum magnitude range (9.0–9.7) reflects global estimation for subduction zones from Kagan & Jackson (2013). Coupling coefficients are chosen to symmetrically encompass earthquake distribution. The distribution adopted in this study is the one marked by the solid black line with dots representing the magnitude values in Fig. S3 and Table S2.
**Figure S3.** Diagrams showing the relationship among the various parameters of the subfaults used to sample the G-R relationship shown in Fig. S2. Notice how the magnitude separation decreases with increasing magnitude values.
**Figure S4.** Mesh of 480 quadrilaterals subfaults re-projected on the fault surface. Yellow circles represent the positions of possible geometrical centres of the earthquakes defined by the ET NODES 4 and 5 (Fig. 1) on the projection of the subduction interface at the Earth's surface.
**Figure S5.** Diagram (a) showing the total number of subfault combinations found for each magnitude value in Fig. S3, and map views (b) of the subfault centres.
**Figure S6.** Flow depth inundation hazard maps for 500 and 5000 yr ARPs at Messina harbour. The maps in the first column are those corresponding to the full (unfiltered) set of sources, i.e. the same set of sources used to calculate the linear hazard shown in Fig. 3. The second column shows the maps obtained with the sources selected after filtering, and the third column presents the residuals between the two. The residuals are also shown as histograms of the absolute and relative differences; red dashed lines indicate the 16th and 84th percentiles; grey shaded areas indicate the region where differences are within ±10 per cent. The maps presented here correspond to the purple rectangle in Fig. 2.

**Figure S7.** Volume flux inundation hazard maps for 500 and 5000 yr ARPs at Messina harbour.
**Figure S8.** Flow depth inundation hazard maps for 500 and 5000 yr ARPs at Catania.
**Figure S9.** Volume flux inundation hazard maps for 500 and 5000 yr ARPs at Catania.
**Figure S10.** Flow depth inundation hazard maps for 500 and 5000 yr ARPs at Augusta.
**Figure S11.** Volume flux inundation hazard maps for 500 and 5000 yr ARPs at Augusta.
**Figure S12.** Flow depth inundation hazard maps for 500 and 5000 yr ARPs at Timpaki.
**Figure S13.** Volume flux inundation hazard maps for 500 and 5000 yr ARPs at Timpaki.

(http://gji.oxfordjournals.org/lookup/suppl/doi:10.1093/gji/ggu408/-/DC1)

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.