



Regression analysis of macroscopic tsunami intensity and offshore wave height in the Mediterranean Sea

Jacopo Selva^{1,2}, Giovanni Zanaboni³, Laura Graziani⁴

¹ University of Naples “Federico II”, Dipartimento di Scienze della Terra, delle Risorse, e dell’Ambiente (DiSTAR), Naples, Italy

² Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Vesuviano, Naples, Italy

³ University of Naples of Bologna, Dipartimento di Fisica e Astronomia, Bologna, Italy

⁴ Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Nazionale Terremoti, Rome Italy

Correspondence to: Jacopo Selva (jacopo.selva@unina.it)

Abstract. Quantifying the physical dimensions of historical tsunamis is fundamental for validating hazard models and informing coastal risk management. However, pre-instrumental records predominantly consist of qualitative descriptions from which macroscopic intensity are defined to quantify the size of the tsunami. In this study, we present the first empirical, statistically robust, and fully invertible relationship linking local macroscopic tsunami intensity to shoreline wave height. By coupling a simplified 1D onshore hydraulic propagation model with high-resolution coastal transects and local roughness constraints, we homogenized observation metrics (run-up and inundation distance) into a consistent wave height reference for three benchmark Mediterranean events: the 1908 Messina Strait, the 2002 Stromboli, and the 2020 Aegean Sea tsunamis. Applying Orthogonal Distance Regression (ODR) to account for uncertainties in both dependent and independent variables, we defined a linear scaling law applicable to any existing historical observation. The relationship is validated against existing observations and subsequently applied to the Italian Tsunami Effects Database (ITED), which collects all the tsunami effects observed along the Italian coasts since 79 AD, reconstructing the tsunami history of Italian coasts and establishing an empirical bridge between historical qualitative archives and modern tsunami modelling and hazard assessments.

1 Introduction

A primary limitation in quantitative tsunami hazard assessments and computational models—such as the regional NEAMTHM18 hazard model for the North-Eastern Atlantic, the Mediterranean, and Connected Seas (Basili et al., 2021), which underpins European tsunami hazard and national warning protocols (DPC, 2018; Selva et al., 2021; Tonini et al., 2021)—is the absence of an explicit empirical control based on past observation of tsunamis, due to the extreme limitation of such data. While quantitative hazard models target physical metrics like coastal wave height, run-up, or flow depth to evaluate risk and design mitigation strategies (Grezio et al., 2017; Behrens et al., 2021; Tonini et al., 2021; Goda and De Risi, 2023), observational historical catalogues rely heavily on qualitative macroseismic-like descriptions (Ambraseys, 1962; Baptista and Miranda, 2009; Maramai et al., 2019a, 2019b; Papadopoulos and Imamura, 2001; Papadopoulos et al., 2007,



2020; Sieberg, 1923, 1927; Soloviev et al., 2000). For this reason, PTHA models are heavily based on the source modeling, looking to merge all existing sources of information about potential tsunami sources (e.g. Selva et al. 2016), but the final results can only qualitatively be compared with the record of past tsunamis. This is mainly due to the lack of a solid relationship bridging calculated physical tsunami parameters to historical macroscopic observations.

- 35 Tsunamis are events rare. Detailed knowledge of their impact is limited to a few events, unevenly distributed over time. Since 2000 BC, the Global Historical Tsunami Database (NCEI/WDS) reports less than 3000 events. Unfortunately, only about 35% have the highest reliability value (4, Tsunami Defined) and about half of these are concentrated since the 1960s and constitute the majority of the record. The small number of available observations is the reason for the high uncertainty in the description of the tsunami sources (Behrens and Dias, 2015).
- 40 To extend over time observations, tsunami catalogues collected all existing information about older historical tsunamis, as described in historical documents (Papadopoulos et al., 2014 and reference therein). In the Euro-Mediterranean region, catalogues like the EMTC (Euro-Mediterranean Tsunami Catalogue) and ITED (Italian tsunami effects database) summarize highly heterogeneous historical descriptions using macroscopic intensity grades (Maramai et al., 2019a, 2019b). For example, the EMTC contains observations for almost 300 tsunamis from 6150 BC to present. The information about the
- 45 physical intensity of the tsunami in the different sites is not always directly available, and when available is provided with a variety of metrics (e.g. inundation length, run up, etc.) and referred to coastal configuration which may differ from its present state. Consequently, metrics such as the 12-degree Papadopoulos and Imamura (2001) scale were explicitly engineered to be independent of physical parameters, focusing instead on discrete effects on humans, objects, vessels, and structures (in line with EMS-98; Grünthal, 1998).
- 50 Quantitative hazard frameworks cannot be directly benchmarked or tested against centuries of macroscopic observational data because no functional conversion exists between physical hydrodynamics and macroscopic grades. Unlike modern seismic hazard methodologies—which routinely employ empirical conversions between ground motion parameters and macroseismic intensity to test national hazard models (Faccioli and Cauzzi, 2006; Gomez-Capera et al., 2007, 2020; Faenza and Michelini, 2010; Meletti et al., 2021; Visini et al., 2021)—tsunami hazard modeling lacks a parallel bridging function.
- 55 Developing such an empirical relationship is therefore essential to converting historical macroscopic intensities into physical parameters, expanding observational datasets, and systematically evaluating past and present PTHA models like NEAMTHM18.

To address this gap, we defined a strategy to develop an empirical relationship bridging calculated physical tsunami parameters to historical macroscopic intensity scales in the Mediterranean Sea, which may be adapted and applied also in

60 other areas. In particular, Section 2 selects a subset of three historical events in the Mediterranean Sea with co-located physical measurements and macroscopic intensity estimates, homogenizing observations to coastal wave height while accounting for local coastal morphology. Section 3 derives and tests the target empirical regression between physical wave height and macroscopic intensity. Finally, Section 4 tests this relationship with all existing data and applies it across the



Italian national catalogue (ITED) to reconstruct physical wave height estimates for all documented historical tsunamis
65 impacting the Italian coastlines.

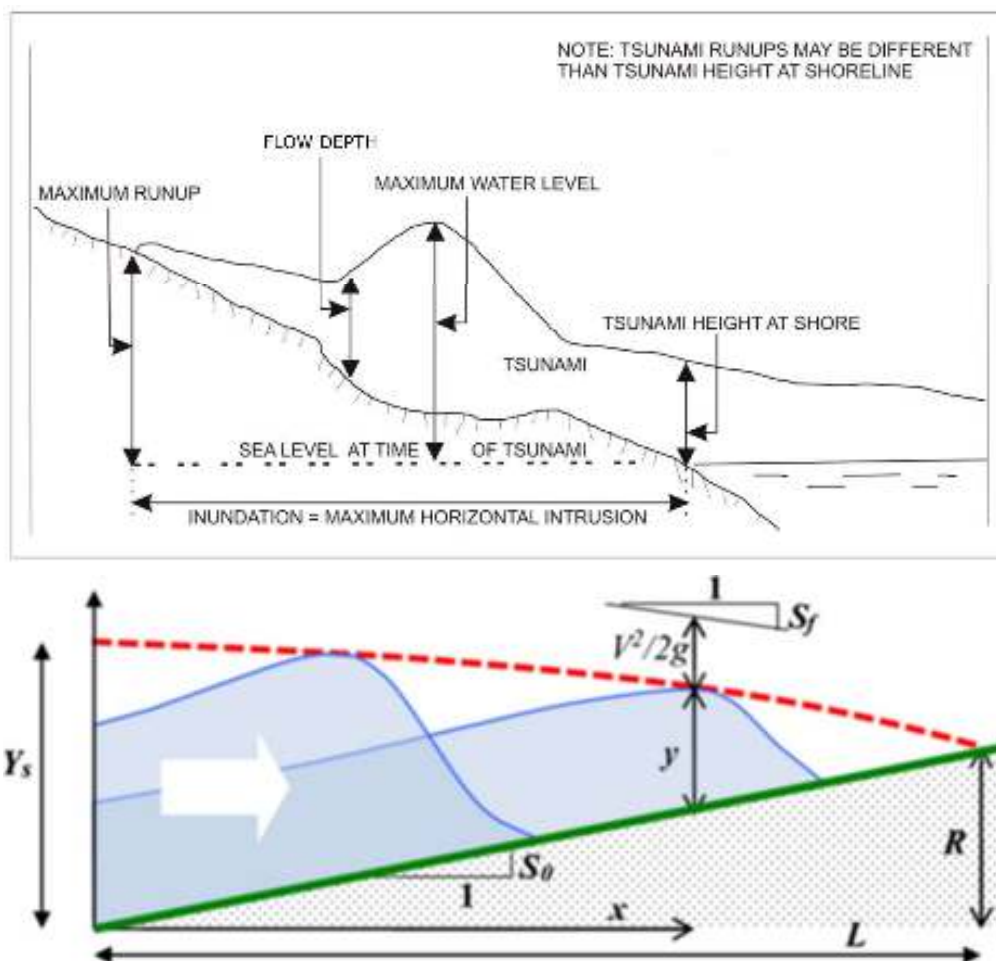
2. Building a homogeneous dataset of observations

To build a homogeneous and unbiased dataset for the regression analysis, we selected a few important events with multiple observations, on which to dive for building a homogeneous starting dataset of observations.

The Messina and Reggio Calabria earthquake and tsunami of 1908 AD is the event with most observations in the Euro-
70 Mediterranean tsunami catalogue (EMTC, Maramai et al., 2019a, 2019b). This event was generated by a magnitude 7.1 earthquake, which probably generated an earthquake-induced landslide, with both sources generating a significant tsunami that caused, along with the earthquake, more than 80,000 fatalities (Favalli et al. 2009, Guidoboni et al. 2018), and that impacted a large area with a wide variety from large to small tsunami intensity, up to more than 10 m of wave height (Favalli et al. 2009). A rich documentation was produced at the time of the event, allowing a detailed reconstruction of the event in
75 many target areas (Platania, 1909a; 1909b; Omori 1909; Baratta, 1910).

To complement this event, we selected two additional more recent Mediterranean events: the tsunami following a volcanic landslide at Stromboli in 2002; and the 2020 Aegean Sea earthquake and tsunami, which affected the coasts of Greece and Turkey. This allows extending source typology, including one event with a purely non-seismic source (Stromboli 2002), one with a purely seismic source (Aegean Sea 2020). This is important for a wider application of the relationship, as existing
80 datasets of macroscopic tsunami intensity mix the different sources and the actual source is sometimes even unknown for some historical event (Maramai et al., 2019a, 2019b).

More information about the selected events and the related source data can be found in Appendix A. Available observations include wave height along the shoreline, tsunami elevation, flow depth, run-up, inundation/withdrawn length (Figure 1a), or simply a qualitative description of the observed phenomena. Existing databases provide data acquired from all available
85 sources, without proposing any quantification of physical measurements other than those reported in historical documents or in the scientific literature, based on observations or, mainly for recent events, direct measurements.



90 Figure 1: a) Physical measures of tsunamis (modified from Triantafyllou et al., 2021). b) Wave impact on a constant slope S_0 land profile, reporting wave-crest levels (Y_s), onshore inundation distance (L), run-up (R), total head friction gradient (S_f), water depth (y) and the profile of wave-crest heights (dashed red line; modified from Smart et al. 2016).



The most common parameters for quantitatively measure tsunamis are run-up and inundation length, as they can be directly
95 derived by the marks left by sediments and water. For this reason, databases report almost exclusive these two types of
measure.

Run-up and inundation length can be related as:

$$R = LS_0 \quad (1)$$

where R is the run-up, L the inundation length, and S the mean slope (Fig. 1a). This very simple and intuitive relationship is
100 possible as run-up and ingression both indicate the point of maximum onshore penetration, but in two different directions:
horizontal and vertical. Eq. 1 shows that to relate run-up and inundation length, we need to have a slope value that
adequately characterises the affected location.

On the other hand, run-up and inundation length implicitly contain a great variability as a function of topography and
hydrography, whose exploration or analysis always require a significant effort in retrieving specific local topographic data at
105 the time of the event. In addition, such measures may reflect a very local situation, more than the general impact in an area
that is commonly referred to when macroscopic tsunami intensity is assigned. In addition, local conditions may significantly
change through time, due to both natural (e.g. coastal dynamic) and/or artificial (e.g. new infrastructure development)
reasons.

To systematically correlate the macroscopic tsunami intensity with physical parameters, it is necessary to make reference to
110 a homogeneous, stable, and standardized type of measures. To this end, we prefer to select a third tsunami measure, that is
the wave height along the coastline. Despite the fact that wave height along the coastline is more difficult to measure
directly, this measure is more stable along the coast and through time, and thus more suitable to be representative of an entire
area, rather than a single point. In addition, wave height is consistent with the tsunami parameter typically evaluated when
the offshore tsunami propagation is modelled without an explicit evaluation of the local inundation, which is the typical case
115 of national and regional hazard assessments (Grezio et al. 2017, Behrens et al. 2021).

Consequently, to build a homogeneous input dataset for the regression, we first establish a reference method to generate an
approximate estimation of the wave height at the coastline from any of the other physical intensity measures (Section 2.1).
Then, we estimate the model's parameters for each observation point to apply the method in the selected three case studies
(Section 2.2). The results are then tested to verify the consistency of the obtained estimates with existing information, when
120 available (Section 2.3).

2.1 Approximate model for wave heights

The simplest relationship between the main physical parameters describing tsunamis can be obtained neglecting all local
effects. Such 'rule-of-thumb' approach can be used anywhere to link the offshore wave height to the potential maximum
125 inland effects simply approximating friction and elevation changes. An example of this type of approach is the one proposed
by Leonard et al. (2008), which has been adopted for developing evacuation plans along the coasts of New Zealand and for
the Italian Coasts (DPC, 2018), translating the regional tsunami hazard model results (e.g., NEAMTHM18, Basili et al.,



2021) into evacuation zones (Tonini et al. 2021). This empirical relationship is based on a simple wave reduction law that only differentiates whether inundation occurs along a river or not, and considers 1 m of maximum run-up decay for every 200 m of tsunami travel inland, or 1 m for every 400 m in significant rivers, until a topographic obstacle is met (Basili et al., 2021; Leonard et al., 2008; Smart et al., 2016; Tonini et al., 2021).

This type of approximations may be useful for national-wide applications, not requiring accurate information about the morphology of the target area. When the average slope of the terrain and its effective friction are known, a better approximation can be obtained using simplified 1D physical onshore propagation model (e.g. Smart et al. 2016; Glimsdal et al. 2019).

A further step up can be obtained with more sophisticated approaches, accounting for the energy conservation principle and modelling the energy dissipation along potential inundation profiles with a modelling approach known as Energy Grade Line Analysis. Energy Grade Line has been adopted in different contexts providing good results for engineering applications (e.g. Wiebe, 2013; Smart et al., 2016; Kimpton et al., 2024), and has been adopted in the American Society of Civil Engineers (ASCE) design standard for tsunami-exposed structures (Chock, 2016; ASCE/SEI7-16, 2016; ASCE/SEI7-22, 2022). Different implementations of this method have been proposed, typically requiring a detailed topographic model of the coastal area and a fine-tuning calibration of the model's parameters that strongly depend on the target area (Estrada et al., 2025).

Here, we follow an intermediate level approach, implementing the simplified 1D model proposed by Smart et al. (2016), which enables the estimation of run-up from wave height based on the concept of the roughness aperture a . This method is selected considering that historical topographic maps are often difficult to reconstruct, and that it should be uniformly applied to multiple target areas.

The Smart et al. (2016) model considers a tsunami wave, propagating onshore with a given velocity over a uniform ground slope, from a one-dimensional perspective. The parameters involved are summarised in the diagram in Figure 1b. Ground permeability is ignored and the wave is considered quasi-steady (Smart et al., 2016). In particular, differentiating along the direction of flow the Bernoulli's equation for free-flowing currents, and assuming that the wave velocity is near constant throughout a vertical section (shallow-water approximation), and accounting for the flow resistance through the Darcy's friction factor f , the maximum wave height at the coast can be derived from inundation distance and run-up, that is

$$Y_s = aS_0 \left(e^{\frac{2L}{3a}} - 1 \right) = aS_0 \left(e^{\frac{2R}{3aS_0}} - 1 \right) \quad (2)$$

where a is the roughness aperture, which is a function of the local Darcy's friction factor f and the distance between the flow protrusions such as trees or houses d :

$$a = 2 \frac{d}{f} . \quad (3)$$

In Eq. 2, the three main physical parameters for describing the local tsunami wave (wave height Y_s , the run-up R , and the inundation length L) are related to each other through two local unknown parameters: the mean characteristic slope S_0 , expressed as percentage, and ground cover roughness aperture a , expressed in meters.



The use of a mean slope by using a constant criterion for all observations (e.g. the mean slope up to 100 m from the coast) might be limiting for the purpose of an accurate description with regard to specific locations. Instead, it is necessary to take into account the size of the local tsunami, so that the slope of an area of similar size to that actually flooded by water during that event could be assessed. If, for example, a run-up value of 3.6 m is observed in a given observation point, it is appropriate to assess the slope of a hillside up to an altitude above sea level of the same order of magnitude (e.g. 5 m). On the other hand, if a run-up value of 8.2 m is observed, it is more appropriate to assess the slope up to an altitude above sea level of at least 10 m. Similarly, where an ingress of 20 m is observed, the slope must be calculated along a consistent planimetric length (e.g. approximately 50 m). On the contrary, using a common criterion for all observations would consider areas certainly not inundated by water, leading to potential biases.

To apply the relationship, also the average friction in the inundated area should be evaluated. Friction should account for both the ground and what stays above it, as both may significantly affect the flowing water. In scientific literature, the reference usually used for high-resolution tsunami simulations is the Manning's coefficient, an empirically derived coefficient used in the Gauckler-Manning-Strickler equation for the description of friction losses (Gayer et al., 2010; Gibbons et al., 2022; Kaiser et al., 2011). This coefficient is tabulated for most materials and soils because the equation in which it appears describes the flow of liquids in free-flowing channels and it is usually used in engineering for civil and environmental applications. However, it is virtually impossible to define the appropriate coefficient in all points for historical events, since land-use changes significantly through time.

The roughness aperture proposed by Smart et al. (2016) is not meant to describe the roughness of a surface, but rather to indicate how water is obstructed in its flow, and its estimation is necessary for each observation. With reference to Eq. (3), it is reasonable to use a typical value for f of 0.05 for fully turbulent flow in rough conduits (Smart et al., 2016). What influences most is the distance between the obstacles, i.e. how dense the obstacles are on the ground that do not allow water to penetrate evenly. Smart et al. (2016) proposed a set of values for roughness aperture characterizing land cover types for the case study they considered (2009 South Pacific tsunami that mainly affected Samoa and South Java), as reported in Table 1. In table, we added a value which can be considered appropriate to characterise built-up areas rich in obstacles for the passage of water, assuming a distance between the flow protrusions $d = 1.25$ m. This value provide a roughness aperture $a = 50$ m that is intermediate between light buildings (which are less stable and dense than standard built-up areas) and dense vegetation (which is denser than a standard built-up area).

Using Table 1, the onshore roughness condition may be roughly estimated by knowing the local climate and the extension of villages at the time of the event, even if the land-use at historical times cannot be always well characterised.

Table 1: Reference values for the roughness aperture parameter a , modified from Smart et al. (2016).

Onshore condition	Aperture parameter a (m)
Flat or smooth open ground, beach	200



Undulating open ground	100
Light buildings, plantations	80
Residential area (buildings and other obstacles)	50
Dense vegetation	10

2.2 Parameter estimation

The evaluation of the mean characteristic slope and the roughness aperture at each observation point (OP) has been made by defining a set of transects with variable lengths and heights in the area of the reported observation. Transects are drawn perpendicularly to the local coastline, starting from an elevation of approximately 0 m and extending inland for a variable length depending on the observation, as shown in Figure 2a.

Run-up and ingression individually describe the effect at one precise point, which generally refers to the maximum values documented in a larger area: run-up data generally refer to a built-up area, while ingression is measured along a less steep path, such as a rice field, valleys or streams. However, especially in the case of historical events, it is necessary to consider several sections for the characterization of the affected area, since there is not always a precise benchmark and the OPs in the catalogues refer to entire localities or districts, mentioning several specific locations in a variable narrative adopted to describe the overall intensity in the area.

Transects were retrieved through the geographic information system (GIS) Google Earth Pro (https://www.google.it/intl/en/earth/) The images, distance and elevation data stack several sources together, allowing for greater coverage. The accuracy of the altimetry data can be considered sufficiently precise for the purpose of an approximate assessment of the slope. Despite Google Earth's lack of accurate data description and errors, many studies evaluated the accuracy of the data, demonstrating a sufficient accuracy for many engineering design purposes, with associated root mean square error usually smaller than 2 m (e.g. Benker et al. (2011), with 1.63 m; El-Ashmawy (2016), with 1.85 m; Wang et al. (2017), with 1.32 m.



Figure 2: a) Transects drawn between Briga Marina (San Paolo district) and Malati district, in Sicily. The central transect extends
 215 over the bed of the Briga stream, with a smaller slope than the residential areas. b) Example of transects plotted from Google
 Earth in Santa Tecla (Sicily). For one of the transects, the elevation profile from which the mean characteristic slope was
 calculated is displayed. Maps data: © Google Earth 2022; images: © Terrametrics, SIO, NOAA, U.S. Navy, NGA, GEBCO.



220 Transects with terrain elevation profiles were drawn at points where observations were available (Figure 2). Considering the variety of Observation Points (OPs), many specific issues emerged. For example, in Figure 2b we show four transects plotted for the OP of Santa Tecla (Sicily), a location severely affected by the tsunami following the 1908 Messina tsunami. This OP provides a very useful example of the required corrections necessary in many OPs. In Santa Tecla, the first two transects (north) were drawn in the centre of the locality, a place naturally protected by a steep cliff. Following consultation of the report by Baratta (1910) (visualised through CFTILab - Advanced Laboratory of Historical Seismology, Tarabusi et al., 2020), it was possible to identify the location of the run-up measured by Platania (Platania, 1909a; 1909b; Baratta, 1910), clearly indicated as the 'southern beach' (Baratta, 1910) and that 'damages were particularly in the southern part of the village' (Platania, 1909b) (from ITED, Maramai et al., 2019b). Whenever the precise location was not identified, a flag was assigned to the relative observations, as potentially affected by large uncertainty. This occurs only for the historical events, like the 1908 Messina earthquake. All the additional interpretations required are detailed in Table S1. In addition, satellite image quality and the correspondence between elevation data and images (coastline elevation 0 m) may vary depending on the specific geographical area considered. In some cases, when the value at 0 m was too far from the coastline, it was not possible to assess the slope. In such cases, the transect(s) of the Observation Point were not considered.

The roughness aperture was instead estimated by comparing the descriptions in Table 1 with the existing or reconstructed characteristics of the area, based on the best knowledge it was possible to reconstruct for the local settings at the OP at the time of the event, assigning one of the available values in the table.

The estimation was extended to all OPs where at least one observation was available. The details about the location of all the used transects as well as the retrieved average slope and the roughness aperture can be found in the supplementary files.

240 **2.3 Estimation of quantitative tsunami parameters**

By evaluating mean characteristic slope and roughness apertures in all OPs and associated transects, we could complement all observation records with at least one numerical quantitative observation with an explicit estimation of all the following three physical parameters: run-up, ingression and wave height on the shoreline (Fig. 1a).

To evaluate the accuracy of the adopted strategy, we focus on the evaluation of the model accuracy at the OPs where both run-up and ingression data are present. Unfortunately, too few data regarding wave height are available, to make a similar comparison for the entire model. Thanks to the punctual evaluation of the mean characteristic slope of each site (transect) hit by a tsunami wave, it was possible to derive run-ups from the ingression data and vice versa, using Eq. (1). Thus, such OPs provide the opportunity to independently test the accuracy of the procedure and of the quantified average slope.

The results are reported in Figure 3, including the three case studies. Plots for the individual events are reported as supplementary Figures S7, S8 and S9. The results demonstrate a good consistency between observed and estimated values, with a general good correlation for all the events considered, demonstrating that the slope obtained from transects is sufficiently precise for most of the observation points.



Referring to the 1908 event (green and red in Figure 3, reported separately in Figure S7), it can be seen that the OPs labelled as uncertain are the most dispersed, demonstrating that the precise location of the site is a key factor for the accuracy of the estimation. The results for the 2002 Stromboli tsunami (yellow in Figure 3, reported separately in Figure S8) show that estimations are very well correlated with observations, with a relatively small dispersion. This is probably due to the relatively simple coastal configuration at Stromboli, with a high angle and approximately homogeneous slope at most of the observation points. The results for the 2020 Aegean Sea tsunami (blue in Figure 3, reported separately in Figure S9) display a larger dispersion. This is probably due to the fact that Google data are more inaccurate in Greece and in Turkey than in Italy. Also, the available field observations are strongly spatially clustered, a feature that cannot be fully reproduced by the applied simplified 1D and averaged-along-the-transect model and that may increment the impact of some local deviation. For example, as it can be observed in Figure S9, there are four points that seem to follow a different trend from the one proposed by the relationship. These points are maximum inundation values along the course of rivers or ditches, where the slope is almost zero and it was not possible to accurately retrieve such a small value (slopes less than 0.70 %). Despite this inaccuracy in the slope, the wave heights computed at the shoreline (using roughness $a = 200$ m inside equation (4)) provided values comparable to the tsunami elevation (TE) values close to the shoreline reported in those locations by Dogan et al. (2021).

Overall, Figure 3 demonstrates that the computational model provides results generally consistent with observations, even if some specificities due to local peculiarities may be missed. However, since our general goal is to relate macroscopic intensity (defined over extended areas) with averaged physical parameters, this level of accuracy is judged sufficient, and potential uncertainty and inaccuracies simply adding uncertainty on the statistical regression.

All the estimated parameters can be retrieved in the supplementary files. When both inundation length and run-up were available, wave height was estimated from both and the maximum among the two was selected as preferred value for that OP.

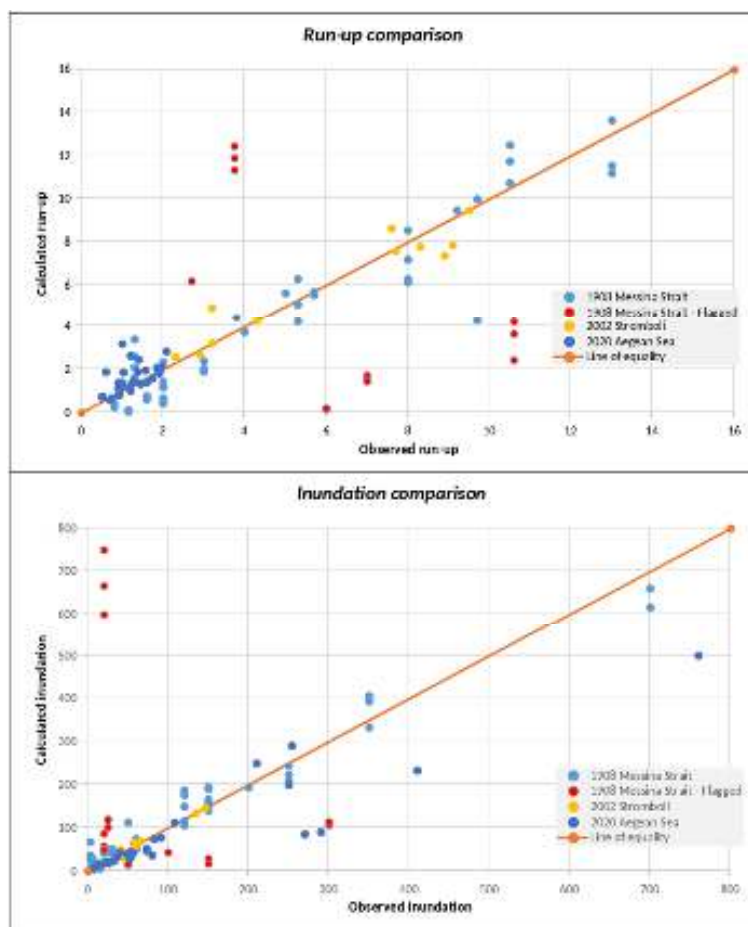


Figure 3: Run-up (panel A) and inundation length (panel B) measured and calculated for all the events. Points in red correspond to places where it was not possible to conduct an effective measurement of the elevations. Plots reporting separately the comparison for the three events are shown in Figures S7, S8, and S9.

280

3 Regression analysis

Even if the relationship between tsunami parameters and tsunami intensity does not exist, a wide literature is available for the analogue problem in seismology, where the field of macroseismology analyze the macroscopic earthquake effects in old



and modern events and their link to the earthquake source. Seismological research has long been oriented towards an attempt to correlate seismic intensities with physical parameters of earthquakes, such as local ground shaking accelerations or earthquake magnitudes. Such associations are not included in the concept of seismic intensity by definition (Papadopoulos & Imamura, 2001), but are fundamental to enable the comparison between modern models to the observations in large past events, as well as observations in areas with few or no instruments.

Apart from a few cases (e.g. Gomez-Capera et al., 2020), all regressions in seismological studies adopt the same functional form: a linear regression between the macroseismic intensity and the logarithm of the peak ground motion (Faenza & Michelini 2010; Gomez-Capera et al. 2007; Cataldi et al., 2021). The purpose of correlation proposals in seismology is to provide a tool for immediate estimation of intensities from ground shaking data (Faccioli & Cauzzi, 2006; Faenza & Michelini, 2010), but the opposite procedure is also a fundamental tool, especially for associating a physical quantity with a past intensity value. These relationships are usually quite noisy. Consequently, an explicit estimation of uncertainty is required. In addition, when they are obtained by means of an ordinary least squares relationship, they do not consider the error on both variables and cannot be inverted. Therefore, an approach allowing for the consideration of uncertainty in both variables is fundamental.

In the following, we first describe the method adopted for estimating the regression (Section 3.1), and then we apply the same rationale to tsunami data (Section 3.2).

3.1 Regression method

Ordinary least-square (OLS) is the most common method for estimating parameters of a regression model. To use the OLS, some conditions must be fulfilled, including that the uncertainties in the independent variable are negligible with respect to the ones on the dependent variable. Being the x position assumed certain, this corresponds to minimizing the residual along the y -axis, that is in vertical, as represented in Figure 4a. This assumption is possible whenever a clear prevalence of the uncertainty can be established among the two considered variables.

When both the independent and dependent variables are subject to normally and independently distributed measurement errors (errors-in-variables), the estimation of the regression parameters may be based on the so-called generalised orthogonal regression (GOR), also called orthogonal distance regression or Deming regression (Boggs et al., 1988). This statistical approach minimises the weighted distance between the fitting line and each data point along a line that is inclined proportionally to the ratio between the variances of the two variables. It is more complex than OLS and requires to set the ratio between the variances of the measurement errors on the two variables, usually indicated with λ . This approach is adopted in many fields, including geophysics, especially seismology (Castellaro et al., 2006; Castellaro & Bormann, 2007; Gomez-Capera et al., 2007; Faenza & Michelini, 2010; Lolli & Gasperini, 2012; Das et al., 2013; Cataldi et al., 2021; Pallavi et al, 2022). A particular case occurs when λ is set to 1 (Castellaro et al., 2006), which is sometimes called simple orthogonal regression (SOR) and the orthogonal distance between the line and the points is minimised (Fig 4b).

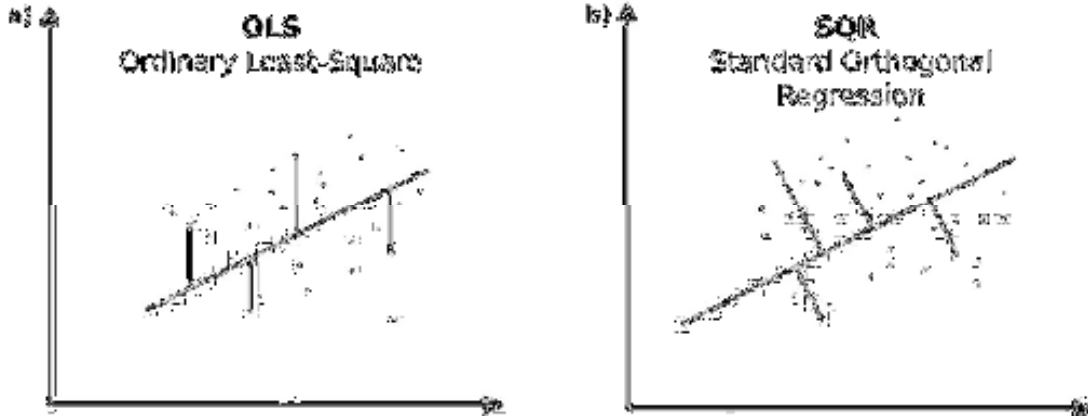


Figure 4: a) Ordinary least-square (OLS) where variable x is assumed without associated error and residual minimized vertically, and b) simple orthogonal regression (SOR), where variables are assumed with comparable error distribution and residuals minimized orthogonality to the regression line (modified from Leng et al. 2007).

When both variables present uncertainties, a fitting using ODR is more appropriate (Fuller, 1987). In general, being ODR a generalization of OLS, ODR never performs worse than OLS and often it performs better; in addition, orthogonal regression allows least-squares data fitting to be extended to problems with independent variables that are both not known exactly (Boggs et al., 1988). This is common in geophysics, thus it has been shown that orthogonal regression leads to a better fitting, even its simpler formulation, that is SOR, which is equivalent to GOR with $\lambda = 1$ (Castellaro et al., 2006; Castellaro & Bormann, 2007; Gomez-Capera et al., 2007; Faenza & Michelini, 2010; Lolli & Gasperini, 2012). Notably, without making assumptions about the negligibility of the uncertainty of one variable, the relation becomes also invertible (Gomez-Capera et al., 2007).

To apply ODR to the tsunami datasets of Section 2, we adopted the specific function for Deming regression provided by Hall (2022), in the Matlab environment, which is based on the formulation of Jensen (2007). In short, in the linear relationship

$$y = \alpha + \beta x \quad (4)$$

of variables x and y and parameters α and β , it is assumed that observations are subject to additive random variation:

$$X_i = x_i + u_i$$

$$Y_i = y_i + e_i = \alpha + \beta x_i + e_i \quad (5)$$

where X_i and Y_i are the observations; x_i and y_i are the (unknown) true values of respectively the dependent and independent variable; u_i and e_i denote the measurement errors. Measurement errors are assumed independent and normally distributed, with mean zero and with their own variance, that is $u_i \sim N(0, \sigma_u^2)$ and $e_i \sim N(0, \sigma_e^2)$. The initial assumption for Deming



regression or general orthogonal regression (GOR) is that λ , the ratio between the variances of the errors in measurements, is known:

$$\lambda = \frac{\sigma_e^2}{\sigma_u^2} \quad (6)$$

The differentiation of the log-likelihood function provides an explicit formulation for estimating the parameters α and β of the linear regression in Eq. (4). Two expressions may be obtained also for estimating the variance of the model, either maximizing the log-likelihood function (Jensen, 2007) or following the formulation discussed in Fuller (1987). In the following, we quantify the variance with both methods, always obtaining consistent results. For a quick reference, the details of the derivation are reported in Appendix B.

3.2 Application to tsunami intensity scales

In analogy with the approach followed for the seismological parameters by Faenza and Michelini (2010), the linear model is built between intensity and the (base-10) logarithm of the wave height. The input dataset consists of all OPs for which an estimate of wave height is available, that is for all the OPs in which a physical measure for the tsunami intensity (of whatever kind) is available. We consider separately the cases including or excluding the OPs labelled as uncertain, leading to a total of 249 and 272 OPs, respectively. Given that flagged observations are connected with potential biases due to the lack of clear information in the historical record, the estimation obtained removing flagged observations is preferred.

At first, we focus on the 12-grade intensity scale (P-I) proposed by Papadopoulos & Imamura (2001). To check the effective correlation between the two parameters, in Table 2 we report the Pearson correlation coefficient R . We statistically test the null hypothesis of no correlation (Fisher 1990; MathWorks, 2022), adopting a significance level of 0.05. A strong correlation was found in the entire dataset, as well as in the individual events, with high R and low p -values. The hypothesis of no correlation is rejected for both the Messina and Aegean events, while it cannot be rejected only in the case of Stromboli, probably due to the very limited range of intensity covered in this event. Considering all data combining the events, both including or excluding the highly uncertainty flagged observations, very high correlation coefficients are found, and the hypothesis of no correlation is coherently rejected.

To implement the regression, we have first to set λ , that is the ratio between the variances of the measurement errors on the two variables. For tsunami intensity as well as for the tsunami physical measures, especially for old events, a precise estimation of the errors on the observations is impossible. However it is reasonable to assume that the ratio between them is roughly constant, as proportional to the time elapsed since the event occurred for both variables. Indeed, for historical tsunamis, intensities are assigned a posteriori, sometimes many years or even centuries after the event, relying only on historical documents and photographic evidence or witnesses accounts. Usually, the only available information is description of the effects on the human environment, as reported in the sources. Measures of the physical parameters are extremely rare for historical tsunamis. It is useful to use a cross-referenced approach, when possible, but it may happen that sources mix eye-witness accounts with actual measurements. For recent tsunamis, intensities can be assigned by reviewing a large



amount of material, including reports, video materials or photographs, thing that reduces significantly uncertainty. Also
 375 specific physical measurements are usually collected with specific campaigns, evaluated and discussed by researchers
 adopting a scientific approach. Thus, uncertainty decreases for both variables. It should also be taken into account that the
 resolving power of intensity scales, related to the number of degrees in the scale, allows for overlapping intervals between
 two adjacent degrees, this results in balanced intensity ratings. Therefore, it is possible to assume, as first approximation, that
 the ratio λ may be roughly constant for both historical and more recent events.

380 **Table 2: Values of R, the Pearson correlation coefficient and the p-value.**

	R	p-value	Ho
1908 Messina Strait tsunami	0.75	< 0.001	Rejected
2002 Stromboli tsunami	0.32	0.22	Not rejected
2020 Aegean Sea tsunami	0.86	< 0.001	Rejected
All data (including flagged data)	0.77	< 0.001	Rejected
All data, without flagged data	0.82	< 0.001	Rejected

Lacking of more specific information, we decided to adopt the simplest formulation, that we set λ is set to 1 adopting a
 simple orthogonal regression - SOR (Castellaro et al., 2006). The stability of the results is then tested for other radically
 different values of λ , that is $\lambda=0.1$ and $\lambda=10$. The regression lines obtained, specifically for each tsunami event separately,
 385 are reported in Figure 5, which compare SOR with OLS. To provide a reference to the uncertainty on the model, we also
 report the lines corresponding to $\pm \sigma$, where distance is evaluated orthogonally to the regression line. The estimated
 parameters for the regressions are reported in Table 3.

Notably, the results for the recent 2020 Aegean sea tsunami (Figure 5c) show a small dispersion, demonstrating that the
 model adopted is indeed able to catch the average physical intensity at the target, which well correlates with the macroscopic
 390 tsunami intensity. The 1908 Messina Strait event (Figure 5a) shows a very good correlation too, simply with observations
 that are more dispersed than for the previous case. For this event, there is a point accumulation value for Intensity 4. The
 data labelled as uncertain are generally at the extremes of the distribution. For the 2002 Stromboli event, the intensity
 variability is lower (only four intensity levels), but the regression results are adequate. No significant differences are evident
 between OLS and SOR.

395

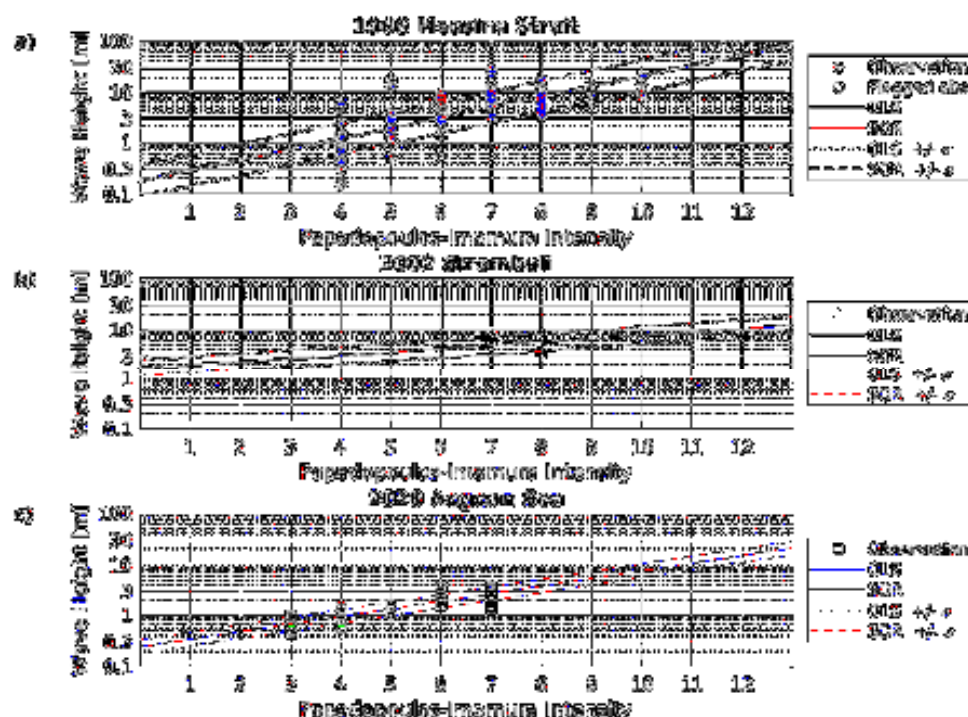


Figure 5: Papadopoulos-Imamura Intensity and wave height regression for a) the 1980 Messina strait tsunami, b) the 2002 Stromboli tsunami, and c) the 2020 Aegean Sea tsunami. The same plots for the Sieberg-Ambraseys scale are reported in figure S10.

400

In Figures 6a, we report the relationship obtained by merging all the three datasets. In particular, we present the results not-considering the points labelled as uncertain. The equivalent plot considering all observations (including the flagged ones) is reported in Figure S11a. The dataset covers a wide range of intensities, from 3 to 10. We can note that the differences between ordinary OLS and orthogonal SOR are not large. Also, there are no clear trends related to a specific event with respect to the others. The only exception is that, at least for intensity 7, the 2020 Aegean Sea observations seem slightly smaller than those of the Messina Strait tsunami. However, this may be related to the fact that this intensity is the largest for the Aegean event, and thus a sort of saturation effect may be observed.

405

To produce a more quantitative comparison, in Table 3, we report all the numerical values for the regression parameters for all the considered cases. Values are very similar between the two types of regressions. Uncertainty on the parameters are



410 generally larger for SOR, correctly representing the fact that only SOR considers uncertainty on both variables. The largest relative errors occur for the Stromboli tsunami, where less observations are available and the intensity range is smaller. Assuming the existence of a unique relationship, the regression parameters should not differ significantly within uncertainty. To this end, again SOR estimates appear more consistent, as all confidence intervals for both parameters are very close or cross to each other.

415 For the combined datasets, the dominant event for the regression is the 1908 Messina Strait event, which is connected to more observations. Notably, the regression lines with and without flags are almost identical, but the associated error is reduced when uncertain labeled observations are excluded. Based on this analysis, we consider the orthogonal regression on the total dataset deprived of flagged data as the best choice, thus the Simple Orthogonal Regression (SOR) takes the form:

420
$$\log_{10}H = -0.85 (\pm 0.07) + 0.21 (\pm 0.01)I_{PI} \quad (7)$$

with $\sigma = 0.25$.

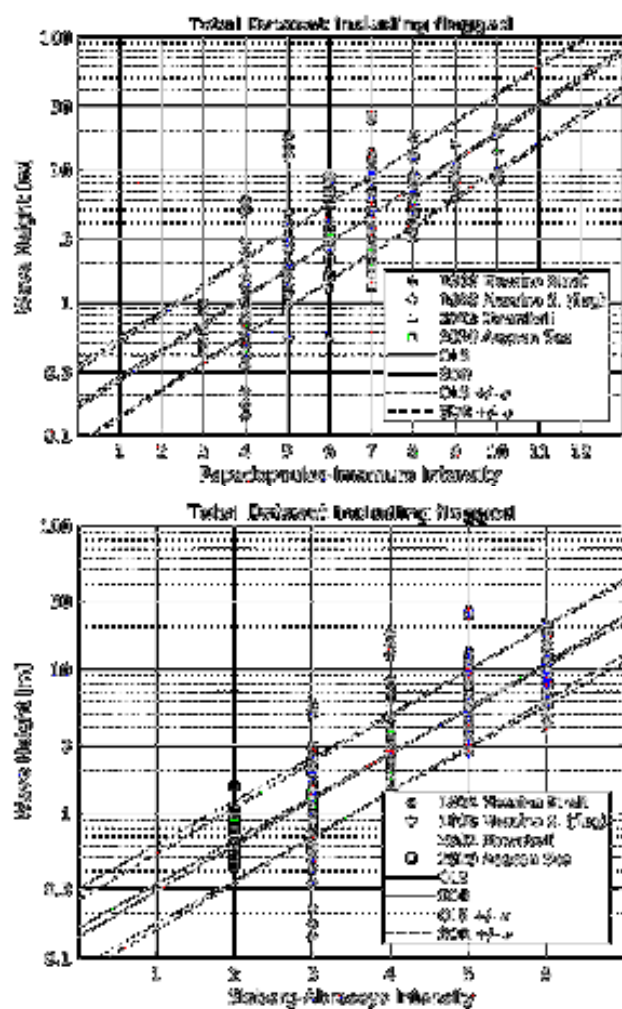
The distribution of the residuals (evaluated orthogonally to the linear trend) is reported in Figure 7, showing that this relationship is nicely centred on the data, with residuals that are well distributed around 0. Considering the residuals as a function of the intensity, no specific trends are observed, showing that the linear regression looks adequate to fit
425 observations. Notably, the statistical regression proves to be stable also with respect to the choice of the parameter λ (Fig. S10).

The same analysis was repeated adopting the intensity scale (S-A) proposed by Sieberg (1927) as modified by Ambresys (1962). The results are reported in Figures 7b, S11, S12, S13, and Tables S2. The obtained relationship takes the form:

430
$$\log_{10}H = -0.83 (\pm 0.07) + 0.31 (\pm 0.01)I_{SA} \quad (8)$$

with $\sigma = 0.24$.

In this case, the scale has only 6 degrees, and the observation points cover the interval from degree 2 to degree 6. In the case of Stromboli, only one intensity value is different from degree 5, making the results for that single event inconsistent. However, collectively the results are consistent also adopting the SA intensity scale, as the reference regression considered
435 (SOR for the total dataset without flags) is nicely centred on the observations with residuals are well distributed around 0.



440 Figure 6: a) Regression between Papadopoulos-Imamura intensity and wave height using the entire dataset, excluding flagged observation (high uncertainty). b) Same of a), for the Sieberg-Ambraseys intensity. The same plots including flagged events are reported in Figure S12.



445

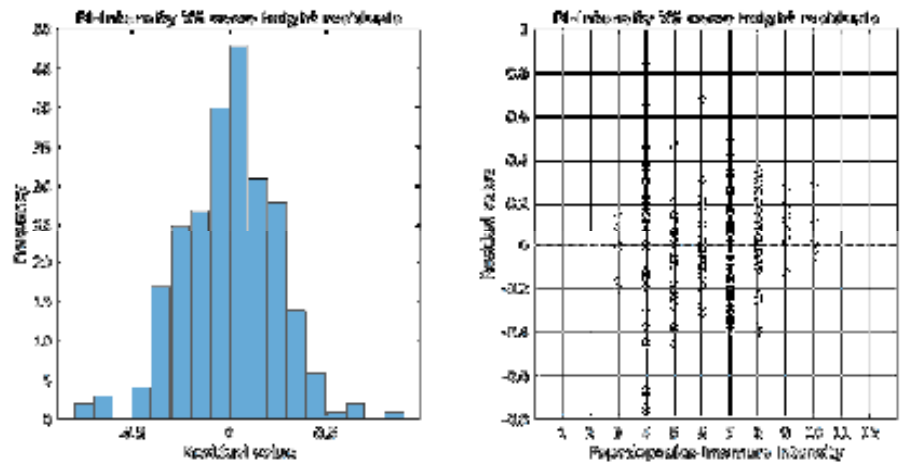


Figure 7 - Analysis of the orthogonal residuals for the Papadopoulos-Imamura intensity: a) distribution of the orthogonal residual.
b) Distribution of the orthogonal residual as a function of the intensity.

450



Table 3: Summary of parameters (α and β) estimated from ordinary least squares regression (OLS) and orthogonal regression (SOR) models for the Papadopoulos and Imamura scale, with associated uncertainties, for intercept, slope, and for the entire model (σ). The estimation of this uncertainty is found consistent with the one proposed by Fuller (1987).

Dataset	Method	Parameter α	Parameter β	σ
1908 Messina	OLS	-0.72 ± 0.08	0.20 ± 0.01	0.31
Strait Tsunami	SOR	-0.76 ± 0.10	0.21 ± 0.01	0.31
2002 Stromboli	OLS	0.23 ± 0.37	0.06 ± 0.05	0.17
Tsunami	SOR	0.22 ± 0.84	0.07 ± 0.11	0.17
2020 Aegean	OLS	-0.61 ± 0.07	0.15 ± 0.01	0.13
Sea Tsunami	SOR	-0.62 ± 0.07	0.15 ± 0.01	0.13
All observations (including flagged observations)	OLS	-0.76 ± 0.07	0.20 ± 0.01	0.29
	SOR	-0.79 ± 0.07	0.21 ± 0.01	0.29
High-quality observations (excluding flagged observations)	OLS	-0.82 ± 0.06	0.21 ± 0.01	0.25
	SOR	-0.85 ± 0.07	0.21 ± 0.01	0.25



4. Application to the Italian Tsunami Effect Database

Thanks to the historical catalogues, it is possible to significantly extend our knowledge about how many tsunamis occurred and where and how strong they impacted, and to help improving the knowledge of tsunamigenic source areas, hazard studies, and risk management. Using the empirical relationship developed in Section 3, as reported in Eqs. 7 and 8, it is possible to estimate the wave height H for all existing observations of tsunamis whenever the available information (from measures to simple descriptions) is sufficient to establish a macroscopic intensity. On the one hand, this allows for checking the consistency of the relationship, whenever it is applied to observations points where a direct measure of the wave height already exist. On the other hand, the relationship may be applied also where only general descriptions exist, providing a significantly extended picture about the tsunami history in the different regions. what was are available. These two topics are the goals of the next two subsections 4.1 and 4.2, respectively.

4.1 Sanity check of the regression

A first sanity check of the proposed regression can be performed restricting the analysis to those OPs where both the macroscopic intensity and the wave height values are available (ITED's field `ited_wave_h`). Considering all the data in ITED with these characteristics, we have 7 observations, 4 of which belong to the 1908 Messina Strait tsunami. To these ITED data, we can add other 3 points available for the 2020 Aegean Sea event, leading to a total of 10 direct observations of wave height. To further extend the test, we can include also the observations of sea rise available in ITED (ITED's field `ited_sea_rise`). These observations are less specific than the ones discussed above, but more frequent for past tsunamis, being easier to record from non-instrumental observations. In ITED, we have 21 observations (10 from the 1908 Messina Strait tsunami), starting from 1646 AD and covering multiple types of source. In this way, we reach a total of 31 observations for the sanity check, with 17 observations already used in input for the development of the intensity wave-height relationship, and 14 completely independent. All the specific data are reported in Table 4.

The results are reported in Figure 8, showing the overall consistency between the observations and the proposed relationship. Interestingly, one of the largest deviation from the proposed regression corresponds to the Gioia Tauro observation associated with the 1783 tsunami, for which a wave height of 20 m is reported for a Papadopoulos-Imamura intensity of 7. Given that the event occurred more than two centuries ago, the historical record may be affected by significant uncertainties. Another hypothesis is that the observation was erroneously attributed to the tsunami generated by the main earthquake, whereas it may instead refer to the devastating landslide-induced tsunami that occurred the following day, which produced much stronger local effects and is associated intensities as high as grade 9 and 10 of the Papadopoulos-Imamura intensity (Graziani et al., 2006). Following this hypothetical attribution and considering an intensity 10, also this observation becomes very well correlated with the empirical relationship derived (Figure 8).



Table 4: Summary of the data used for the sanity check. More details in Section 4.1.

Type	Observation point	PI intensity	Observation, m	Source event	Observation source
Wave height	Gioia Tauro	7 / VII	20	1783 Calabrian Earthquake	ITED: FID 50, field ited_wave_h
Wave height	Catona	8 / VIII	4	1908 Messina Strait	ITED: FID 172, field ited_wave_h
Wave height	Cannitello	5 / V	1	1908 Messina Strait	ITED: FID 175, field ited_wave_h
Wave height	Messina	8 / VIII	6	1908 Messina Strait	ITED: FID 209, field ited_wave_h
Wave height	Torre del Faro	5 / V	2	1908 Messina Strait	ITED: FID 261, field ited_wave_h
Wave height	Alassio	4 / IV	3	1968 Liguria-Cote d'Azur	ITED: FID 288, field ited_wave_h
Wave height	Porto di Levante	5 / V	2	1988 Vulcano earthquake	ITED: FID 289, field ited_wave_h
Wave height	Demircili	5 / V	1	2020 Aegean Sea earthquake	Dogan et al. (2021)
Wave height	Sigacik_36	7 / VII	2.3	2020 Samos earthquake	Dogan et al. (2021)
Wave height	Akarca_2	7 / VII	2.5	2020 Samos earthquake	Dogan et al. (2021)
Sea rise	Livorno	4 / IV	2.5	1646 Livorno Earthquake	ITED: FID 24, field ited_sea_rise
Sea rise	Augusta	9 / IX	15	1693 Val di Noto Earthquake	ITED: FID 29, field ited_sea_rise
Sea rise	Siracusa	5 / V	2.5	1693 Val di Noto Earthquake	ITED: FID 35, field ited_sea_rise
Sea rise	Messina	4 / IV	2	1783 Messina Strait	ITED: FID 62, field ited_sea_rise
Sea rise	Gaeta	4 / IV	2.5	1805 Molise Earthquake	ITED: FID 85, field ited_sea_rise
Sea rise	Sorrento	4 / IV	2.5	1805 Molise Earthquake	ITED: FID 86, field ited_sea_rise
Sea rise	La Spezia	4 / IV	1	1809 Liguria	ITED: FID 87, field ited_sea_rise
Sea rise	Sant'Angelo	7 / VII	30	1836 Calabrian Earthquake	ITED: FID 106, field ited_sea_rise
Sea rise	Livorno	4 / IV	0.9	1846 Tuscany earthquake	ITED: FID 107, field ited_sea_rise
Sea rise	Milazzo	4 / IV	0.8	1905 Calabrian earthquake	ITED: FID 137, field ited_sea_rise
Sea rise	Cozzo Spadaro	4 / IV	1.5	1908 Messina Strait	ITED: FID 185, field ited_sea_rise
Sea rise	Crotone	6 / VI	2	1908 Messina Strait	ITED: FID 186, field ited_sea_rise
Sea rise	Porto Empedocle	4 / IV	0.5	1908 Messina Strait	ITED: FID 227, field ited_sea_rise
Sea rise	Pozzallo	4 / IV	1	1908 Messina Strait	ITED: FID 230, field ited_sea_rise
Sea rise	Siracusa	4 / IV	2	1908 Messina Strait	ITED: FID 251, field ited_sea_rise
Sea rise	Siracusa S. Lucia	4 / IV	2	1908 Messina Strait	ITED: FID 252, field ited_sea_rise
Sea rise	Termini Imerese	3 / III	0.8	1908 Messina Strait	ITED: FID 258, field ited_sea_rise
Sea rise	Tropea	4 / IV	2.5	1908 Messina Strait	ITED: FID 263, field ited_sea_rise



Sea rise	Monteleone (Porto Santa Venere)	4 / IV	1.5	1908 Messina Strait	ITED: FID 264, field ited_sea_rise
Sea rise	Stromboli-Piscit�	5 / V	10	1916 Stromboli	ITED: FID 270, field ited_sea_rise
Sea rise	Stromboli-San Vincenzo	8 / VIII	2.5	1930 Stromboli	ITED: FID 280, field ited_sea_rise

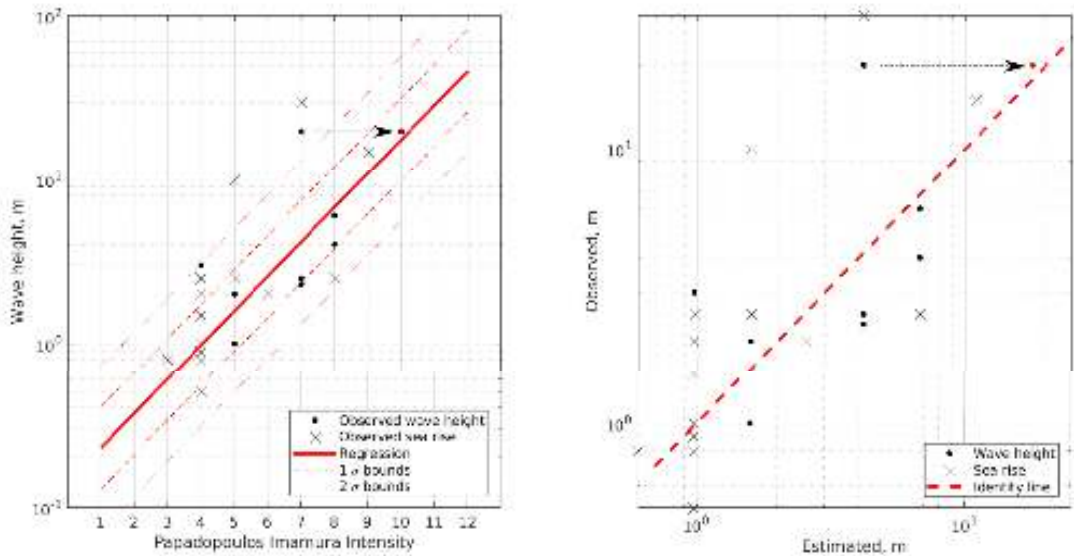


Figure 8: Sanity check of the empirical relationship, comparing a) measured wave height with the empirical relationship and b) estimated VS measured wave height. The 1783 observation at Giogia Tauro is reported both in its original position (black dot and red cross), and in its potentially corrected position (red dot). See text for more details.

4.2 Tsunami-history of the Italian coasts

Extending the analysis to all the OPs of ITED allows to make a first estimation of all tsunami wave height observed in the Italy, allowing to complement the tsunami-history of the entire Italian coast. In Figure 9, we report all historical tsunami impact observation points on the Italian coast, considering location and intensity provided by the dataset ITED (Maramai et al., 2019b). Here, the existing record is complemented with the estimated wave height on the coast. The numerical values of

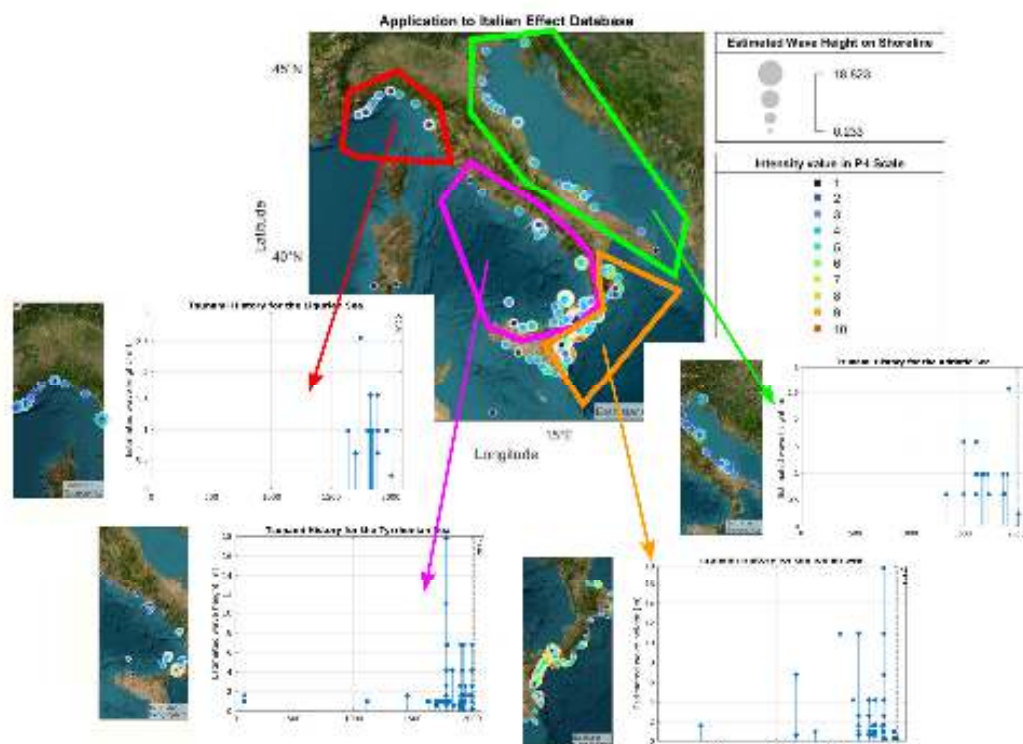


such estimates along with their uncertainty, are provided as a supplementary file, by adding 6 columns to the original ITED dataset, that is the wave height estimation from both intensities and the 1σ confidence interval. The uncertainty is evaluated using the estimated standard error of regression σ , that is assuming that the uncertainty on the regression model is negligible with respect to the statistical variability of variables.

505 The distribution of the points is not homogeneous in time and space, in fact most of them are concentrated along the coasts of Sicily and Calabria, also because many are referred to the 1908 Messina Strait event. Multiple estimates for the same OPs are infrequent, since it is rare that multiple tsunami observations at the same location have been sufficiently documented to be catalogued. A useful analogy can be drawn again with the seismological analogue, that is with the Italian Macroseismic Database (DBMI15), where macroseismic intensity data are used to reconstruct the seismic history of thousands of
510 individual localities (Locati et al., 2022). Similarly, ITED provides the historical record of tsunami effects along the Italian coast.

Using the empirical relationship developed in Section 3, we can complement this information by adding a homogeneous estimate of shoreline wave height to each observation point, thus enriching the tsunami record with a quantitative physical parameter. Given that the tsunami history cannot be reconstructed at individual observation points, here we stack the
515 information in the traditional Sea areas surrounding Italy, as indicated in Figure 9: Ligurian, Tyrrhenian, Ionian and Adriatic Sea. Individual seas have been used for this aggregation as they can be considered, in many cases, approximately isolated to each other, a part for tsunamis occurring at the boundary between them. In any case, any other representation is possible.

In Figure 9, we report the resulting tsunami-history. It can be seen that the historical distribution of tsunamis in the various Italian seas is approximately homogeneous from 1600 to the present day, whereas before that only a few events are present
520 in the Tyrrhenian and Ionian Seas. The largest estimated wave heights are smaller than 5 m for all observations for the Ligurian and Adriatic Seas, while for the Tyrrhenian and Ionian Seas, values larger than 10 m are also observed, with the largest wave heights reaching almost 20 m. These time histories were produced using the Papadopoulos-Imamura intensity and, consequently, the relationship of Eq. 7.



525

Figure 9: Map displaying the application of empirical relationship of Eq. 7 to the entire ITED dataset, converting all the Papadopoulos-Imamura intensity to wave height values. In the subplots, we report the tsunami history in the different domains. Map produced in Matlab, with data Esri, HERE, Garmin, GeoTechnologies, Inc., USGS/Powered by Esri.

5. Final Remarks

530 In this study, we derive and test the first empirical relationship between tsunami physical parameters and macroscopic tsunami intensity. To this end, we first built up a reference dataset composed by three events: the 1908 Messina Strait tsunami generated by an earthquake and a simultaneous seismic-induced landslide, the 2002 Stromboli tsunami generated by a volcanic landslide, and the 2020 Aegean Sea tsunami generated by an earthquake.

As first step, the starting dataset was homogenized both in terms of macroscopic intensity and physical tsunami measures. A
535 macroscopic tsunami intensity has been estimated from field survey reports whenever an explicit estimation of intensity was not already available. When a physical measure of the tsunami was available, we applied a simple model to homogenize the



record. We selected the wave height at the coastline as a reference parameter since it is independent of local onshore topographic effects and applicable for larger stretches of coastline than any inland parameters. Starting from official databases or post-event reports, each tsunami observation was analysed individually, assigning a slope value (where possible) and an average water flow resistance value to characterise the roughness of the inundated area, quantified through the roughness aperture. These two parameters were then used to derive the missing physical tsunami measures (e.g. run-up and wave height, if inundation length was available). The selected approach was tested where multiple observations of different types were available, and it proved to be effective.

To derive a robust and invertible relationship between macroscopic intensity and wave height, we adopted an orthogonal regression model. Orthogonal regression technique allows accounting for uncertainties on both variables, which is fundamental, considering that both wave height and tsunami intensity are affected by large uncertainties. This choice follows the analogue issue in the seismological field, and proved to be efficient also for tsunami data. For simplicity, we selected a ratio between the two variances equal to 1, and we tested the stability of the obtained regression line for alternative choices. The analysis of residuals shows no significant trends, with an overall distribution centred and distributed around 0, meaning that a linear regression is sufficient to satisfactorily model all the input data. The parameters of the regression are stable also considering individual events, standing for the generalization of the found regression beyond the input case studies.

The relationship obtained allows for explicitly estimate wave height at coastline whenever information on the impact is sufficient to assign a macroscopic tsunami intensity. The estimates proved to be consistent with measured wave heights for the (few) cases in which observations were available.

The application of the relationship to the entire Italian catalogue at all the observation points made possible to obtain a specific assessment of the 'tsunami history' of individual seas, which represents an important reference for understanding tsunami risk in the different areas. Since estimates are provided in terms of wave height, they may provide an important check point for both tsunami models of individual events, and tsunami hazard analyses. Indeed, all models, both local and regional, allow for an explicit estimation of this parameter. Therefore, this dataset provides the unique opportunity for quantitatively testing hazard models, as well as reference values for evaluating civil protection plans. Notably, the dataset can be also seen as a reference for each individual event, providing also in this case an important reference for source and impact studies regarding each individual event.

This study has several limitations. Compared to earthquake databases, tsunami catalogues have a reduced number of records. For Italy, most of the observations refer to events that occurred more than a century ago and are therefore associated with measurement/witness errors in the physical data and ambiguities in the current assignment of intensities, leading to large uncertainties. Furthermore, only a few events in the Italian database contain multiple observation points. However, this limitation is also a motivation for the study, as it may contribute to significantly extending the record for both individual events and specific target areas.

Also the used range of intensity is reduced. Making reference to the Papadopoulos-Imamura intensities, the range of available data covers the interval between grade 3 to grade 10. The largest intensities (grades 11 and 12) are not present in



the Italian catalogue and nor were associated with the 2020 Aegean Sea event. Therefore intensity value 10 represents the theoretical upper limit of applicability of the proposed relationship. The larger intensities are associated with very high wave height values, but there is no criterion to establish whether they are reasonable or not. This limitation is analogous to that of twelve-degree macroseismic scales, e.g. EMS-98, in which degree 12 represents the maximum conceivable level of effects, which cannot necessarily be observed in an earthquake. Thus the "working range" of all these scales tends usually to be from intensity 3 to intensity 10 (Grunthal, 1998).

The relationship was derived for two macroscopic tsunami intensity scales, the Papadopoulos-Imamura and the Sieberg-Ambraseys ones. The Papadopoulos-Imamura scale has more grades (12 instead of 10), and this allowed for a more consistent application also in cases with a more limited range of variability, as the 2002 Stromboli tsunami. Therefore, we consider the relationship obtained with the Papadopoulos-Imamura intensity the one with more robust results, to be considered for reference. However, both estimates have been produced, in order to allow for future scrutiny.

Finally, we note that coastal erosion and other natural and artificial coastal dynamics impacted the possibility to efficiently define topographic profiles, especially for historical events. For example, the 1908 event in the Messina Strait, especially the Calabria coast, experienced this in several locations, sometimes requiring the reconstruction of the configuration of the coast at the time of the event. Modern topographic profiles were derived from Google Earth Pro. Google Earth was a very powerful tool because it quickly provides sufficiently accurate DTM (Digital Terrain Model) data of many geographical areas together with constantly updated satellite photographs. This combination was decisive for the choice of the GIS system over other available sources of information like GEBCO. However, some areas have discrepancies between the terrain model and the image, making it difficult or impossible to interpret the elevations for slope calculation. This problem occurred especially along the coasts of Greece and Turkey for the analysis of the 2020 Aegean Sea tsunami. Overall, all these issues generated difficulties in reconstructing the average coastal slope at the time of the tsunami event, and this estimation can be certainly improved in the future with more specific local studies. Similarly, a better definition of the roughness aperture will be also possible in the future. Here, we used typical values at discrete levels based on the averaged land cover of the area. In the future, a more careful analysis may lead to a more extensive tabulation of its values depending on land use and cover at the time of the event. To allow future scrutiny and give ground for improvements, all the transects used and the estimated roughness at all observation points are provided as supplementary files.

Code and data availability

All the data used and the results are provided as supplementary files.



Author contributions

600 JS and GZ conceived the analysis. GZ developed all the computational codes, under the coordination of JS. GZ and LG reviewed all the historical records and implemented all the related analyses. JS prepared the manuscript with contributions from all co-authors. All authors reviewed and approved the manuscript.

Competing interests

Authors declare no competing interest.

605 Acknowledgements

This manuscript is adapted and integrated from the Master's thesis of G. Zanaboni (Università di Bologna, 2022), titled 'Development of a relationship between physical parameters of tsunami waves and tsunami intensity based on macroscopic effects and its application to the Italian tsunami database', supervised by J. Selva and L. Graziani.

610 Financial support

None

References

- Ambraseys, N. N.: Data for the investigation of the seismic sea-waves in the Eastern Mediterranean. Bulletin of the Seismological Society of America, 52(4), 895-913, Available at <https://pubs.geoscienceworld.org/ssa/bssa/article-abstract/52/4/895/116131/Data-for-the-investigation-of-the-seismic-sea?redirectedFrom=fulltext> (last access: Sept 2026) 615 1962.
- American Society of Civil Engineers: Minimum Design Loads and Associated Criteria for Buildings and other Structures. American Society of Civil Engineers, ASCE/SEI7-16, Reston, Virginia, USA, <https://doi.org/10.1061/9780784414248>, 2017.
- 620 American Society of Civil Engineers: Minimum Design Loads and Associated Criteria for Buildings and other Structures (ASCE/SEI7-22). American Society of Civil Engineers, Reston, Virginia, USA, <https://doi.org/10.1061/9780784415788>, 2022.
- Baptista, M. A., & Miranda, J. M.: Revision of the Portuguese catalog of tsunamis. Natural Hazards and Earth System Sciences, 9, 25-42, <https://doi.org/10.5194/nhess-9-25-2009>, 2009.



- 625 Baratta, M.: La catastrofe sismica calabro messinese (28 dicembre 1908). Società Geografica Italiana, Roma, 1910.
Available at
https://cfti.ingv.it/collections/multimedia/DVD_1908/Fonti%20scientifiche%20e%20istituzionali/La%20catastrofe%20sismi%20calabro%20messinese%20di%20M.Baratta%20-%201910/Testo/index.html (last access: Sep 2026)
- Basili, R., Brizuela, B., Herrero, A., Iqbal, S., Lorito, S., Maesano, F. E., Murphy, S., Perfetti, P., Romano, F., Scala, A.,
630 Selva, J., Taroni, M., Thio, H. K., Tiberti, M. M., Tonini, R., Volpe, M., Glimsdal, S., Harbitz, C. B., Løvholt, F., Baptista,
M. A., Carrilho, Matias, F., L. M., Omira, R., Babeyko, A., Hoechner, A., Gurbuz, M., & Pekcan, O.: The Making of the
NEAM Tsunami Hazard Model 2018 (NEAMTHM18). *Frontiers in Earth Science*, 8, 616594,
<https://doi.org/10.3389/feart.2020.616594>, 2021.
- Behrens, J., & Dias, F.: New computational methods in tsunami science. *Philosophical Transactions of the Royal Society A:*
635 *Mathematical, Physical and Engineering Sciences*, 373, 20140382, <https://doi.org/10.1098/rsta.2014.0382>, 2015.
- Behrens, J., Løvholt, F., Jalayer, F., Lorito, A., Salgado-Gálvez, M. A., Sørensen, M., Abadie, S., Aguirre-Ayerbe, I., Aniel-
Quiroga, I., Babeyko, A., Baiguera, M., Basili, R., Belliazzi, S., Grezio, A., Johnson, K., Murphy, S., Paris, R., Rafliana, I.,
De Risi, R., Rossetto, T., Selva, J., Taroni, M., Del Zoppo, M., Armigliato, A., Bureš, V., Cech, P., Cecioni, C.,
Christodoulides, P., Davies, G., Dias, F., Bayraktar, H. B., González, M., Gritsevich, M., Guillas, S., Harbitz, C. B.,
640 Kanoglu, U., Macías, J., Papadopoulos, G. A., Polet, J., Romano, F., Salamon, A., Scala, A., Stepinac, M., Tappin, D., Thio,
H. K., Tonini, R., Triantafyllou, I., Ulrich, T., Varini, E., Volpe, M., & Vyhmeister, E.: Probabilistic Tsunami Hazard and
Risk Analysis - A Review of Research Gaps. *Frontiers in Earth Science*, 9, 628772,
<https://doi.org/10.3389/feart.2021.628772>, 2021.
- Benker, S. C., Langford, R. P., & Pavlis, T. L.: Positional accuracy of the Google Earth terrain model derived from
645 stratigraphic unconformities in the Big Bend region, Texas, USA. *Geocarto International*, 26, 291-303,
<https://doi.org/10.1080/10106049.2011.568125>, 2011.
- Billi, A., Funicello, R., Minelli, L., Faccenna, C., Neri, G., Orecchio, B., Presti, D.: On the cause of the 1908 Messina
tsunami, southern Italy. *Geophys. Res. Lett.* 35, L06301. <https://doi.org/10.1029/2008GL033251>, 2008.
- Boggs, P. T., Spiegelman, C. H., Donaldson, J. R., & Schnabel, R. B.: A computational examination of orthogonal distance
650 regression. *Journal of Econometrics*, 38, 169-201, [https://doi.org/10.1016/0304-4076\(88\)90032-2](https://doi.org/10.1016/0304-4076(88)90032-2), 1988.
- Boschi, E., Gasperini, P., Smriglio, G., Valensise, G.: The new “Catalogue of Strong Italian Earthquakes.” *Annals of*
Geophysics 38, 20. <https://doi.org/10.4401/ag-4072>, 1995.
- Castellaro, S., & Bormann, P.: Performance of Different Regression Procedures on the Magnitude Conversion Problem.
Bulletin of the Seismological Society of America, 97, 1167-1175, <https://doi.org/10.1785/0120060102>, 2007.
- 655 Castellaro, S., Mulargia, F., Kagan, Y.Y.: Regression problems for magnitudes *Geophysical Journal International* 165, 913–
930. <https://doi.org/10.1111/j.1365-246X.2006.02955.x>, 2006.



- Cataldi, L., Tiberi, L., and Costa, G.: Estimation of MCS intensity for Italy from high quality accelerometric data, using GMICEs and Gaussian Naïve Bayes Classifiers. *Bulletin of Earthquake Engineering*, 19, 2325-2342, <https://doi.org/10.1007/s10518-021-01064-6>, 2021.
- 660 Chiocci, F.L., Romagnoli, C., Tommasi, P., Bosman, A.: The Stromboli 2002 tsunamigenic submarine slide: Characteristics and possible failure mechanisms. *J. Geophys. Res.* 113, B10102, <https://doi.org/10.1029/2007JB005172>, 2008.
- Chock, G. Y. K.: Design for tsunami loads and effects in the ASCE 7-16 standard. *Journal of Structural Engineering*, 142, 04016093, [https://doi.org/10.1061/\(ASCE\)ST.1943-541X.0001565](https://doi.org/10.1061/(ASCE)ST.1943-541X.0001565), 2016.
- Das, R., Wason, H. R., & Sharma, M. L.: General Orthogonal Regression Relations between Body-Wave and Moment
- 665 Magnitudes. *Seismological Research Letters*, 84, 219-224, <https://doi.org/10.1785/0220120125>, 2013.
- Dogan, G. G., Yalciner, A. C., Yuksel, Y., Ulutaş, E., Polat, O., Güler, I., Şahin, C., Tarih, A., and Kanoğlu, U.: The 30 October 2020 Aegean Sea Tsunami: Post-Event Field Survey Along Turkish Coast. *Pure and Applied Geophysics*, 178, 785-812, <https://doi.org/10.1007/s00024-021-02693-3>, 2021.
- Dipartimento della Protezione Civile (DPC): Indicazioni alle Componenti ed alle Strutture operative del Servizio nazionale di protezione civile per l'aggiornamento delle pianificazioni di protezione civile per il rischio maremoto. URL: <https://www.protezionecivile.gov.it/it/normativa/indicazioni-alle-componenti-ed-alle-strutture-operative-del-servizio-nazionale-di-protezione-civile-per-l-aggiornamento-delle-pianificazioni-di-protezione-civile-per-il-rischio-maremoto> (last access: Sept 2026), 2018.
- El-Ashmawy, K. L. A.: Investigation of the Accuracy of Google Earth Elevation Data. *Artificial Satellites*, 51, 89-97, <https://doi.org/10.1515/arsa-2016-0008>, 2016.
- 675 Estrada, M. F., Cienfuegos, R., Urrutia, A., Catalán, P. A., and Winckler, P.: Forward energy grade line analysis for tsunami inundation mapping. *Coastal Engineering*, 197, 104673, <https://doi.org/10.1016/j.coastaleng.2024.104673>, 2025.
- Faccioli, E., & Cauzzi, C.: Macroseismic intensities for seismic scenarios estimated from instrumentally based correlations, <https://doi.org/10.13140/RG.2.1.3984.2641>, 2006.
- Faenza, L., & Michellini, A.: Regression analysis of MCS intensity and ground motion parameters in Italy and its application
- 680 in ShakeMap. *Geophysical Journal International*, 180, 1138-1152, <https://doi.org/10.1111/j.1365-246X.2009.04467.x>, 2010.
- Favalli, M., Boschi, E., Mazzarini, F., and Pareschi, M. T.: Seismic and landslide source of the 1908 Straits of Messina tsunami (Sicily, Italy). *Geophysical Research Letters*, 36, L16304, <https://doi.org/10.1029/2009GL039135>, 2009.
- Fisher, R.A.: *Statistical Methods for Research Workers*, Oxford Science Publication, <https://doi.org/10.1093/oso/9780198522294.002.0003>, 1990.
- 685 Fuller, W. A.: *Measurement error models*. John Wiley & Sons. <https://doi.org/10.1002/9780470316665>, 1987.



- Gayer, G., Leschka, S., Nöhren, I., Larsen, O., and Günther, H.: Tsunami inundation modelling based on detailed roughness maps of densely populated areas. *Natural Hazards and Earth System Sciences*, 10, 1679-1687, <https://doi.org/10.5194/nhess-10-1679-2010>, 2010.
- Gibbons, S. J., Lorito, S., de la Asunción, M. M., Volpe, M., Selva, J., Macías, J., Sánchez-Linares, C., Brizuela, B., Vöge, M., Tonini, R., Lanucara, P., Glimsdal, S., Romano, F., Meyer, J. C., and Løvholt, F.: The Sensitivity of Tsunami Impact to Earthquake Source Parameters and Manning Friction in High-Resolution Inundation Simulations. *Frontiers in Earth Science*, 9, 757618, <https://doi.org/10.3389/feart.2021.757618>, 2022.
- Glimsdal S, Løvholt F, Harbitz C, Romano F, Lorito S, Orefice S, Brizuela B, Selva J, Hoechner A, Volpe M, Babeyko A, Tonini R, Wronna M, Omira R (2019), A new approximate method for quantifying tsunami maximum inundation height probability, *Pure and Applied Geophysics*, <https://doi.org/10.1007/s00024-019-02091-w>
- Goda, K., and De Risi, R.: Future perspectives of earthquake-tsunami catastrophe modelling: From single hazards to cascading and compounding multi-hazards. *Frontiers in Built Environment*, 8, 1022736, <https://doi.org/10.3389/fbuil.2022.1022736>, 2023.
- Gomez-Capera, A. A., Albarello, D., and Gasperini, P.: Aggiornamento relazioni fra l'intensità macrosismica e PGA. Technical report, Convenzione INGV-DPC 2004-2006, <http://es1.mi.ingv.it/d11.html> (last access: Sept 2026), 2007.
- Gomez-Capera, A. A., D'Amico, M., Lanzano, G., Locati, M., and Santulin, M.: Relationships between ground motion parameters and macroseismic intensity for Italy. *Bulletin of Earthquake Engineering*, 18, 5143-5164, <https://doi.org/10.1007/s10518-020-00905-0>, 2020.
- Graziani, L., Maramai, A., and Tinti, S.: A revision of the 1783-1784 Calabrian (southern Italy) tsunamis. *Natural Hazards and Earth System Sciences*, 6, 1053-1060, <https://doi.org/10.5194/nhess-6-1053-2006>, 2006.
- Grezio, A., Babeyko, A., Baptista, M. A., Behrens, J., Costa, A., Davies, G., Geist, E., Glimsdal, S., Gonzales, F. I., Griffin, J., Harbitz, C., LeVeque, R.J., Lorito, S., Løvholt, F., Omira, R., Mueller, C., Paris, R., Parsons, T., Polet, J., Power, W., Selva, J., Sorensen, M. B., and Thio, H. K.: Probabilistic Tsunami Hazard Analysis: Multiple Sources and Global Applications. *Reviews of Geophysics*, 55, <https://doi.org/10.1002/2017RG000579>, 2017.
- Grünthal, G. : European macroseismic scale 1998. European Seismological Commission (ESC), <https://doi.org/10.2312/EMS-98.full.en>, 1998.
- Guidoboni, E., and Mariotti, D.: Il terremoto e il maremoto del 28 dicembre 1908: effetti e parametri sismici. In G. Bertolaso, E. Boschi, E. Guidoboni, & G. Valensise (Editors), *Il terremoto e il maremoto del 28 dicembre 1908: analisi sismologica, impatto, prospettive* (pp. 17–136). Roma–Bologna: Dipartimento della Protezione Civile (DPC) – Istituto Nazionale di Geofisica e Vulcanologia (INGV), Available at <https://cfti.ingv.it/collections/index.php/it/articoli/67-guidoboni-e-mariotti-d-2008-il-terremoto-e-il-maremoto-del-1908-effetti-e-parametri-sismici> (last access: Sept 2026), 2008.



- Guidoboni, E., Ferrari, G., Mariotti, D., Comastri, A., Tarabusi, G., Sgattoni, G., and Valensise, G.: CFTI5Med, Catalogo dei Forti Terremoti in Italia (461 a.C.-1997) e nell'area Mediterranea (760 a.C.-1500). Istituto Nazionale di Geofisica e Vulcanologia (INGV), <https://doi.org/10.6092/ingv.it-cfti5>, 2018.
- 720 Guidoboni, E., Ferrari, G., Tarabusi, G., Sgattoni, G., Comastri, A., Mariotti, D., Ciuccarelli, C., Bianchi, M.G., and Valensise, G.: CFTI5Med, the new release of the catalogue of strong earthquakes in Italy and in the Mediterranean area. *Sci. Data* 6, 80. <https://doi.org/10.1038/s41597-019-0091-9>, 2019.
- Hall J.: Linear Deming Regression, MATLAB Central File Exchange, 2022. Available at <https://www.mathworks.com/matlabcentral/fileexchange/33484-linear-deming-regression>
- 725 Hébert, H., & Schindelé, F.: Tsunami Impact Computed from Offshore Modeling and Coastal Amplification Laws: Insights from the 2004 Indian Ocean Tsunami. *Pure and Applied Geophysics*, 172, 3385–3407, <https://doi.org/10.1007/s00024-015-1136-4>, 2015.
- Jensen, A.C.: Deming regression [Technical report, MethComp package]. Steno Diabetes Center, 2007. Available at <https://www.scribd.com/document/476651504/Deming>
- 730 Kalligeris, N., Skanavis, V., Charalampakis, M., Melis, N.S., Voukouvalas, E., Annunziato, A., and Synolakis, C.E.: Field survey of the 30 October 2020 Samos (Aegean Sea) tsunami in the Greek islands. *Bull. Earthq. Eng.* 20, 7873–7905. <https://doi.org/10.1007/s10518-021-01250-6>, 2022.
- Kaiser, G., Scheele, L., Kortenhaus, A., Løvholt, F., Römer, H., and S. Leschka: The influence of land cover roughness on the results of high resolution tsunami inundation modeling. *Natural Hazards and Earth System Sciences*, 11, 2521–2540, <https://doi.org/10.5194/nhess-11-2521-2011>, 2011.
- 735 Kimpton, T., Higuera, P., Whittaker, C., Wotherspoon, L., and Zorn, C.: A rapid simplified method for determining tsunami inundation extent based on energy conservation. *Computer-Aided Civil and Infrastructure Engineering*, <https://doi.org/10.1111/mice.13168>, 2024.
- Leonard, G.S., Power, W., Lukovic, B., Smith, W., Johnston, D., and Downes, G.: Interim Tsunami Evacuation Mapping for the Wellington Region. GNS Science report 30, 22p, 2008. Available at <https://www.gw.govt.nz/document/23825/interim-tsunami-evacuation-mapping-for-the-wellington-region/>
- Locati, M., Camassi, R., Rovida, A., Ercolani, E., Bernardini, F., Castelli, V., Caracciolo, C. H., Tertulliani, A., Rossi, A., Azzaro, R., D'Amico, S., and Antonucci, A.: Italian Macroseismic Database (DBMI15), Version 4.0. Istituto Nazionale di Geofisica e Vulcanologia (INGV). <https://doi.org/10.13127/DBMI/DBMI15.4>, 2022.
- 745 Lolli, B., and Gasperini, P.: A comparison among general orthogonal regression methods applied to earthquake magnitude conversions. *Geophysical Journal International*, 190, 1135–1151, <https://doi.org/10.1111/j.1365-246X.2012.05530.x>, 2012.



- Maramai, A., Graziani, L., Alessio, G., Burrato, P., Colini, L., Cucci, L., Nappi, R., Nardi, A., and Vilardo, G.: Near- and far-field survey report of the 30 December 2002 Stromboli (Southern Italy) tsunami. *Mar. Geol.* 215, 93–106. <https://doi.org/10.1016/j.margeo.2004.11.009>, 2005.
- 750 Maramai, A., Graziani, L., and Brizuela, B.: Euro-Mediterranean Tsunami Catalogue (EMTC), version 2.0. Istituto Nazionale di Geofisica e Vulcanologia (INGV), <https://doi.org/10.13127/tsunami/emtc.2.0>, 2019a.
- Maramai A., Graziani L., Brizuela B. Italian Tsunami Effects Database (ITED) (Version 1.0) [Data set]. Istituto Nazionale di Geofisica e Vulcanologia (INGV). <https://doi.org/10.13127/tsunami/ited.1.0>, 2019b
- Mariotti, D.: I terremoti del ‘900: La “catastrofe sismica” del 28 dicembre 1908 – INGVterremoti [WWW Document], 2015.
- 755 URL <https://ingvterremoti.com/2015/12/28/i-terremoti-del-900-la-catastrofe-sismica-del-28-dicembre-1908/> (accessed 28.7.26).
- MathWorks: MATLAB (Version 9.12/R2022a) [Software]. Natick, Massachusetts: The MathWorks Inc, URL: <https://www.mathworks.com>, 2022.
- Meletti, C., Marzocchi, W., D’Amico, V., Lanzano, G., Luzi, L., Martinelli, F., Pace, B., Rovida, A., Taroni, M., Visini, F.,
760 and the MPS19 Working Group. The new Italian seismic hazard model (MPS19). *Annals of Geophysics*, 64, 6, <https://doi.org/10.4401/ag-8579>, 2021.
- Mercalli, G.: Contributo allo studio del terremoto calabro-messinese del 28 dicembre 1908. *Atti del Regio Istituto d’Incoraggiamento di Napoli, serie VI, vol. 7*, pp. 249–292 (also published separately as pp. 1–46). <https://doi.org/10.3931/e-rara-80840>, 1909.
- 765 NCEI/WDS - National Geophysical Data Center / World Data Service: NCEI/WDS Global Historical Tsunami Database. NOAA National Centers for Environmental Information. <https://doi.org/10.7289/V5PN93H7>, 2022.
- Omori, F.: Preliminary Report on the Messina-Reggio Earthquake of Dec. 28, 1908. *Bulletin of the Imperial Earthquake Investigation Committee* 3, 37–45, 1909..
- Pallavi, S.J., Dilbag Singh, M.K. and Lee, H.-N.: Comprehensive Review of Orthogonal Regression and its Applications in
770 Different Domains. *Archives of Computational Methods in Engineering*, <https://doi.org/10.1007/s11831-022-09728-5>, 2022 .
- Papadopoulos, G., and Imamura, F.: Proposal for a new tsunami intensity scale. *ITS 2001 Proceedings*, 5, 5-1, available at https://www.researchgate.net/publication/228781850_Proposal_for_a_new_tsunami_intensity_scale (last access: Sept 2026) 2001.
- Papadopoulos, G. A., Daskalaki, E., Fokaefs, A. and Giraleas, N.: Tsunami hazards in the Eastern Mediterranean: strong
775 earthquakes and tsunamis in the East Hellenic Arc and Trench system. *Natural Hazards and Earth System Sciences*, 7, 57-64, <https://doi.org/10.5194/nhess-7-57-2007>, 2007.



- Papadopoulos, G. A., Gràcia, E., Urgeles, R., Sallares, V., De Martini, P.M., Pantosti, D., González, M., Yalciner, A.C., Mascle, J., Sakellariou, D., Salamon, A., Tinti, S., Karastathis, V., Fokaefs, A., Camerlenghi, A., Novikova, T., and Papageorgiou, A.: Historical and pre-historical tsunamis in the Mediterranean and its connected seas: Geological signatures, generation mechanisms and coastal impacts. *Marine Geology*, 354, 81-109, <https://doi.org/10.1016/j.margeo.2014.04.014>, 2014.
- Papadopoulos, G. A., Imamura, F., Nosov, M., Charalampakis, M.: Chapter 3 - Tsunami magnitude scales. In: Engel, M., Pilarczyk, J., May, S. M., Brill, D., Garrett, E. (Eds.), *Geological Records of Tsunamis and Other Extreme Waves*. Elsevier, pp. 33-46, <https://doi.org/10.1016/B978-0-12-815686-5.00003-1>, 2020.
- Piatanesi, A., Lorito, S., and Romano, F.: Il grande maremoto del 1908: analisi e modellazione. In G. Bertolaso, E. Boschi, E. Guidoboni, and G. Valensise (Editors), *Il terremoto e il maremoto del 28 dicembre 1908: analisi sismologica, impatto, prospettive* (pp. 183–196). Roma–Bologna: Dipartimento della Protezione Civile (DPC) – Istituto Nazionale di Geofisica e Vulcanologia (INGV), 2008. Available at <https://www.earth-prints.org/server/api/core/bitstreams/cecc9b71-ef23-4746-9f01-76a95dde03cf/content>
- Platania, G.: I fenomeni marittimi che accompagnarono il terremoto di Messina del 28 dicembre 1908. *Rivista Geografica Italiana*, 16, 154-161, 1909a.
- Platania, G.: Il maremoto dello stretto di Messina del 28 dicembre 1908. *Bollettino della Società Sismologica Italiana*, 1909b.
- Rovida, A., Locati, M., Camassi, R., Lolli, B., Gasperini, P., and Antonucci, A.: Catalogo Parametrico dei Terremoti Italiani (CPTI15), versione 4.0. Istituto Nazionale di Geofisica e Vulcanologia (INGV). <https://doi.org/10.13127/cpti/cpti15.4>, 2022.
- Schambach, L., Grilli, S.T., Tappin, D.R., Gangemi, M.D., and Barbaro, G.: New simulations and understanding of the 1908 Messina tsunami for a dual seismic and deep submarine mass failure source. *Mar. Geol.* 421, 106093. <https://doi.org/10.1016/j.margeo.2019.106093>, 2020.
- Selva, J., Amato A, Armigliato A, Basili R, Bernardi F, Brizuela B, Cerminara M, de' Micheli Vitturi M, Di Bucci D, Di Manna P, Esposti Ongaro T, Lacanna G, Lorito S, Løvholt F, Mangione D, Panunzi E, Piatanesi A, Ricciardi A, Ripepe M, Romano F, Santini M, Scalzo A, Tonini R, Volpe M, Zaniboni F.: Tsunami risk management for crustal earthquakes and non-seismic sources in Italy. *La Rivista del Nuovo Cimento*, 44, 69-144, <https://doi.org/10.1007/s40766-021-00016-9>, 2021.
- Sieberg, A.H.: *Geologische, physikalische und angewandte Erdbebenkunde*. G. Fischer, 1927.
- Smart, G., Crowley, K., and Lane, E.: Estimating tsunami run-up. *Natural Hazards*, 80, <https://doi.org/10.1007/s11069-015-2052-8>, 2016.



- Soloviev, S. L., Solovieva, O.N., Go, C.N., Kim, K.S., and Shchetnikov, N.A.: Tsunamis in the Mediterranean Sea 2000 B.C.-2000 A.D., *Advances in Natural and Technological Hazards Research*. Springer Netherlands, Dordrecht, <https://doi.org/10.1007/978-94-015-9510-0>, 2000.
- 810 Tarabusi, G., Ferrari G., Ciuccarelli C., Bianchi M. G., Sgattoni G., Comastri A., Mariotti D., Valensise G., and Guidoboni E.: CFTILab - Laboratorio Avanzato di Sismologia Storica. Istituto Nazionale di Geofisica e Vulcanologia (INGV), <https://doi.org/10.13127/CFTI/CFTILAB>, 2020.
- Tinti, S., and Armigliato, A.: The use of scenarios to evaluate the tsunami impact in southern Italy. *Marine Geology* 199, 221–243. [https://doi.org/10.1016/S0025-3227\(03\)00192-0](https://doi.org/10.1016/S0025-3227(03)00192-0), 2003.
- 815 Tinti, S., Pagnoni, G., and Zaniboni, F.: The landslides and tsunamis of the 30th of December 2002 in Strombolian analysed through numerical simulations. *Bull. Volcanol.* 68, 462–479. <https://doi.org/10.1007/s00445-005-0022-9>, 2006.
- Tonini, R., Di Manna, P., Lorito, S., Selva, J., and Volpe, M.: Testing Tsunami Inundation Maps for Evacuation Planning in Italy. *Frontiers in Earth Science*, 9, 628061, <https://doi.org/10.3389/feart.2021.628061>, 2021.
- Triantafyllou, I., Gogou, M., Mavroulis, S., Lekkas, E., Papadopoulos, G.A., and Thravalos, M.: The Tsunami Caused by the
820 30 October 2020 Samos (Aegean Sea) Mw7.0 Earthquake: Hydrodynamic Features, Source Properties and Impact Assessment from Post-Event Field Survey and Video Records. *J. Mar. Sci. Eng.* 9, 68. <https://doi.org/10.3390/jmse9010068>, 2021.
- Visini, F., Pace, B., Meletti, C., Marzocchi, W., Akinci, A., Azzaro, R., Barani, S., Barberi, G., Barreca, G., Basili, R., Bird, P., Bonini, M., Burrato, P., Busetti, M., Carafa, M.M.C., Cocina, O., Console, R., Corti, G., D’Agostino, N., D’Amico, S.,
825 D’Amico, V., Dal Cin, M., Falcone, G., Fracassi, U., Gee, R., Kastelic, V., Lai, C.G., Langer, H., Maesano, F.E., Marchesini, A., Martelli, L., Monaco, C., Murru, M., Peruzza, L., Poli, M.E., Pondrelli, S., Rebez, A., Rotondi, R., Rovida, A., Sani, F., Santulin, M., Scafidi, D., Selva, J., Slejko, D., Spallarossa, D., Tamaro, A., Tarabusi, G., Taroni, M., Tiberti, M.M., Tusa, G., Tuvè, T., Valensise, G., Vannoli, P., Varini, E., Zanferrari, A., and Zuccolo, E.: Earthquake Rupture Forecasts for the MPS19 Seismic Hazard Model of Italy. *Ann. Geophys.* 64 (2), SE220. <https://doi.org/10.4401/ag-8608>,
830 2021.
- Wang, Y., Zou, Y., Henrickson, K., Wang, Y., Tang, J., Park, B-J.: Google Earth elevation data extraction and accuracy assessment for transportation applications. *PLOS ONE*, 12, e0175756, <https://doi.org/10.1371/journal.pone.0175756>, 2017.
- Wiebe, D.: Tsunami Inundation: Estimating Damage and Predicting Flow Properties. Oregon State University, URL: <https://ir.library.oregonstate.edu/downloads/wh246w11m?locale=en> (last access: Sept 2026), 2013.

835



Appendix A – Selected events

Section A.1 The 1908 Messina and Reggio Calabria tsunami

The earthquake that struck the Strait of Messina on December 28, 1908, remains a prominent seismic event in Italian history, registering a moment magnitude of $M_w = 7.1$ in regional earthquake catalogs (Mariotti, 2015; Guidoboni et al., 2018, 2019; Rovida et al., 2022). The combined devastation of structural collapses and the incoming sea waves caused approximately 80,000 deaths, with the tsunami alone estimated to have claimed between several hundred and 2,000 lives (Boschi et al., 1995; Mariotti, 2015; NCEI/WDS, 2022). Post-event field investigations carried out by contemporary researchers generated an extensive observational record (Mercalli, 1909; Omori, 1909; Platania, 1909a, 1909b; Baratta, 1910). Along eastern Sicily, vertical run-up measurements generally ranged from 6 to 9.5 m, peaking at 11.7 m near Sant'Alessio, while coastal impacts diminished sharply north of Messina and attenuated gradually south past Catania (Platania, 1909a, 1909b; Baratta, 1910). Along the Calabrian shore, run-up levels reached 6–11 m, with a maximum elevation of 13 m documented near Pellaro, though inundation remained minor along northern Tyrrhenian Calabrian shores and decreased steadily southward along the Ionian coast, except for questionable measurements at Cirò Marina and Roccella Ionica (Platania, 1909a, 1909b; Baratta, 1910; Guidoboni & Mariotti, 2008). Reconstructing the precise mechanism of the 1908 tsunami has prompted sustained scientific debate. Early reconnaissance noted discrepancies between localized wave trajectory markers and primary seismic dislocation models (Omori, 1909). Modern hydrodynamic modeling indicates that tectonic fault dislocation alone fails to fully reproduce the extreme run-up heights documented along both sides of the strait (Tinti & Armigliato, 2003). Consequently, numerous studies suggest a hybrid mechanism involving a seismically triggered submarine mass movement (Piatanesi et al., 2008; Billi et al., 2008; Favalli et al., 2009; Schambach et al., 2020; NCEI/WDS, 2022), a hypothesis supported by variable eyewitness reports regarding initial sea-level drawdown and wave arrival times. Data for this event were extracted from the Italian Tsunami Effects Database (ITED; Maramai et al., 2019b). While all 119 Observation Points (OPs) for this event in ITED include macroscopic intensity values, physical wave parameters are available for a subset: 43 OPs report run-up only, 15 report inundation distance only, 35 contain both run-up and inundation, and 4 provide explicit wave heights. Points lacking physical data or containing only initial sea withdrawal observations were omitted. To evaluate local slope profiles accurately, generic municipal center coordinates provided in ITED were refined by re-examining primary historical field reports (Mercalli, 1909; Baratta, 1910; Tarabusi et al., 2020) alongside local planning maps and archived photographic records. Out of 208 constructed spatial transects, 23 were assigned uncertainty flags due to persistent geographical ambiguities or historical inconsistencies, enabling sensitivity testing with and without these flagged records



Section A.2 The 2002 Stromboli tsunami

The earthquake that struck the Strait of Messina on December 28, 1908, remains a prominent seismic event in Italian history, registering a moment magnitude of $M_w = 7.1$ in regional earthquake catalogs (Mariotti, 2015; Guidoboni et al., 2018, 2019; Rovida et al., 2022). The combined devastation of structural collapses and the incoming sea waves caused approximately 80,000 deaths, with the tsunami alone estimated to have claimed between several hundred and 2,000 lives (Boschi et al., 1995; Mariotti, 2015; NCEI/WDS, 2022). Post-event field investigations carried out by contemporary researchers generated an extensive observational record (Mercalli, 1909; Omori, 1909; Platania, 1909a, 1909b; Baratta, 1910). Along eastern Sicily, vertical run-up measurements generally ranged from 6 to 9.5 m, peaking at 11.7 m near Sant'Alessio, while coastal impacts diminished sharply north of Messina and attenuated gradually south past Catania (Platania, 1909a, 1909b; Baratta, 1910). Along the Calabrian shore, run-up levels reached 6–11 m, with a maximum elevation of 13 m documented near Pellarò, though inundation remained minor along northern Tyrrhenian Calabrian shores and decreased steadily southward along the Ionian coast, except for questionable measurements at Cirò Marina and Roccella Ionica (Platania, 1909a, 1909b; Baratta, 1910; Guidoboni & Mariotti, 2008). Reconstructing the precise mechanism of the 1908 tsunami has prompted sustained scientific debate. Early reconnaissance noted discrepancies between localized wave trajectory markers and primary seismic dislocation models (Omori, 1909). Modern hydrodynamic modeling indicates that tectonic fault dislocation alone fails to fully reproduce the extreme run-up heights documented along both sides of the strait (Tinti & Armigliato, 2003). Consequently, numerous studies suggest a hybrid mechanism involving a seismically triggered submarine mass movement (Piatanesi et al., 2008; Billi et al., 2008; Favalli et al., 2009; Schambach et al., 2020; NCEI/WDS, 2022), a hypothesis supported by variable eyewitness reports regarding initial sea-level drawdown and wave arrival times. Data for this event were extracted from the Italian Tsunami Effects Database (ITED; Maramai et al., 2019b). While all 119 Observation Points (OPs) for this event in ITED include macroscopic intensity values, physical wave parameters are available for a subset: 43 OPs report run-up only, 15 report inundation distance only, 35 contain both run-up and inundation, and 4 provide explicit wave heights. Points lacking physical data or containing only initial sea withdrawal observations were omitted. To evaluate local slope profiles accurately, generic municipal center coordinates provided in ITED were refined by re-examining primary historical field reports (Mercalli, 1909; Baratta, 1910; Tarabusi et al., 2020) alongside local planning maps and archived photographic records. Out of 208 constructed spatial transects, 23 were assigned uncertainty flags due to persistent geographical ambiguities or historical inconsistencies, enabling sensitivity testing with and without these flagged records

Section A.3 The 2020 Aegean Sea tsunami

On October 30, 2020, an $M_w = 7.0$ earthquake struck north of Samos Island, generating a tsunami that affected the surrounding Greek islands and the western coast of Turkey (Dogan et al., 2021; Kalligeris et al., 2021; Triantafyllou et al., 2021). The Turkish coast suffered significant damage, particularly around Akarca where run-up heights reached 3.8 m, resulting in one fatality and several injuries (Dogan et al., 2021; Triantafyllou et al., 2021). Impact along Greek shorelines



was comparatively moderate, restricted to low-lying areas below 2 m elevation without causing casualties, assisted in part by automated early warning broadcasts issued approximately 10 minutes post-earthquake (Kalligeris et al., 2021; Triantafyllou et al., 2021). Because existing catalogs lacked formal intensity ratings for this event, local intensities were systematically assigned to survey sites using post-event reconnaissance literature (Dogan et al., 2021; Kalligeris et al., 2021; Triantafyllou et al., 2021), following the dual-scale classification scheme adopted by EMTC and ITED (Sieberg–Ambraseys and Papadopoulos–Imamura scales; Ambraseys, 1962; Papadopoulos & Imamura, 2001; Maramai et al., 2019a, 2019b). Although reduced GIS elevation quality required discarding several OPs (particularly on Samos where 0 m sea-level references mismatched coastline imagery), 48 transects met the precision criteria required for analysis.

On October 30, 2020, an $M_w = 7.0$ earthquake struck north of Samos Island, generating a tsunami that affected the surrounding Greek islands and the western coast of Turkey (Dogan et al., 2021; Kalligeris et al., 2021; Triantafyllou et al., 2021). The Turkish coast suffered significant damage, particularly around Akarca where run-up heights reached 3.8 m, resulting in one fatality and several injuries (Dogan et al., 2021; Triantafyllou et al., 2021). Impact along Greek shorelines was comparatively moderate, restricted to low-lying areas below 2 m elevation without causing casualties, assisted in part by automated early warning broadcasts issued approximately 10 minutes post-earthquake (Kalligeris et al., 2021; Triantafyllou et al., 2021). Because existing catalogs lacked formal intensity ratings for this event, local intensities were systematically assigned to survey sites using post-event reconnaissance literature (Dogan et al., 2021; Kalligeris et al., 2021; Triantafyllou et al., 2021), following the dual-scale classification scheme adopted by EMTC and ITED (Sieberg–Ambraseys and Papadopoulos–Imamura scales; Ambraseys, 1962; Papadopoulos & Imamura, 2001; Maramai et al., 2019a, 2019b). Although reduced GIS elevation quality required discarding several OPs (particularly on Samos where 0 m sea-level references mismatched coastline imagery), 48 transects met the precision criteria required for analysis.

Appendix B - Regression analysis

The differentiation of the log-likelihood function provides (e.g. Fuller 1987):

$$\hat{\beta} = \frac{m_{yy} - \lambda m_{xx} + \sqrt{(m_{yy} - \lambda m_{xx})^2 + 4\lambda m_{xy}^2}}{2m_{xy}} \quad (B1)$$

$$\hat{\alpha} = \bar{Y} - \bar{X} \hat{\beta} \quad (B2)$$

$$\hat{x}_i = \frac{\lambda x_i + \hat{\beta}(y_i - \hat{\alpha})}{\lambda + \hat{\beta}^2} \quad (B3)$$

where $\bar{X}$ and $\bar{Y}$ are the mean of the observations; m_{xx} and m_{yy} are the sample variance of x and y , and m_{xy} is the sample covariance between x , y and they are the elements of the sample covariance matrix, that is

$$m_{xx} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad m_{yy} = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad m_{xy} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \quad (B4)$$

The variance of the model obtained maximising the log-likelihood function has the form (Jensen, 2007):



$$\hat{\sigma} = \frac{\lambda \sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta} \bar{X}_i)^2}{2\lambda(n-2)} \quad (\text{B5})$$

930 where $(n-2)$ is used for unbiased estimate since there are $(n+2)$ parameters to consider, hence the degrees of freedom are $2n - (n+2) = n-2$.

The variance of the model can be also calculated following the formulation discussed in Fuller (1987):

$$\hat{\sigma}_F = \frac{\sum_{i=1}^n (Y_i - \bar{Y} - \hat{\beta} (X_i - \bar{X}))^2}{(n-2)} \quad (\text{B6})$$