ORIGINAL PAPER

Multi-level emulation of tsunami simulations over Cilacap, South Java, Indonesia

Ayao Ehara[1] · Dimitra M. Salmanidou[1] · Mohammad Heidarzadeh[2] · Serge Guillas[1] (iD)

Abstract

Carrying out a Probabilistic Tsunami Hazard Assessment (PTHA) requires a large number of simulations done at a high resolution. Statistical emulation builds a surrogate to replace the simulator and thus reduces computational costs when propagating uncertainties from the earthquake sources to the tsunami inundations. To reduce further these costs, we propose here to build emulators that exploit multiple levels of resolution and a sequential design of computer experiments. By running a few tsunami simulations at high resolution and many more simulations at lower resolutions we are able to provide realistic assessments whereas, for the same budget, using only the high resolution tsunami simulations do not provide a satisfactory outcome. As a result, PTHA can be considered with higher precision using the highest spatial resolutions, and for impacts over larger regions. We provide an illustration to the city of Cilacap in Indonesia that demonstrates the benefit of our approach.

Keywords Gaussian process · Multi-fidelity · Emulation · Experimental design · Uncertainty propagation · Hazard assessment

1 Introduction

A tsunami is an unpredictable geophysical phenomenon and has the potential to cause massive destruction along the coasts and leave large death tolls. The unpredictability of tsunami phenomena poses a challenge to tsunami researchers. Tsunami waves are primarily generated by seabed displacements such as underwater earthquakes and submarine landslides. After generation, the tsunami waves propagate rapidly over the deep ocean, get amplified when the water depth decreases, and run up the coast, provoking potentially severe damage. Recent tsunamis such as the Palu (September 2018, more than 2000 deaths) and Anak Krakatau (December 2018, more than 450 deaths) tsunamis, both in Indonesia [1, 2], have been particularly devastating, as their characteristics were somehow unexpected due to a lack of adequate probabilistic scenario-based assessment of the source mechanism. Notably, the 2004 Indian Ocean tsunami killed roughly 228,000 people across the Indian Ocean, especially from vulnerable communities, and the 2011 Tohoku tsunami killed approximately 19,000 people in Japan [3, 4]. In particular, there is a significant need for Government to assess the generic risk to buildings, and the concrete impact on the full range of assets of households. In response to the urgent necessity, the first end-to-end practical modelling that interconnects tsunami risk and in-depth economic analysis of the losses in the affected area has recently been developed [5]. Since uncertainties in the size of future tsunamis and their local impacts are key to the estimation of the risk, the significance of sophisticated ways of dealing with more precise uncertainty quantification in tsunami hazard is all the more increasing.

Probabilistic scenario-based tsunami hazard assessments [6–8] deliver a priori critical data to tsunami disaster planning and practice. There exist variants in probabilistic scenario-based assessments as discussed in recent in-depth reviews [9, 10]. However, apart from the difficulties in

✉ Ayao Ehara
  ayao.ehara.18@ucl.ac.uk

✉ Serge Guillas
  s.guillas@ucl.ac.uk

[1] Department of Statistical Science, University College London, Gower Street, London, WC1E 6BT, UK

[2] Department of Architecture and Civil Engineering, University of Bath, Claverton Down Road, Bath, BA2 7AY, UK

assigning probabilities to scenarios, the computational burden needed for simulating each scenario prohibits an exhaustive sweep over the entire range of plausible scenarios. Therefore, as mentioned e.g. in Gopinathan et al. [11], statistical emulation, one of the methods in probabilistic scenario-based assessment, is an essential tool for the end-to-end physical and numerical modelling of local tsunami impact, i.e., from the earthquake source to tsunami velocities and heights, in order to surmount the prohibitive computational cost of running a large number of simulations. The outputs from a statistical emulator constitute the predictions of future earthquake scenarios and allow to quantify and assess high risk but low probability hazard thresholds.

Statistical emulation in tsunami risk assessment is a relatively unexplored field although Giles et al. [12] and Gopinathan et al. [11] have established an insightful framework in Gaussian Process (GP) tsunami emulation. Alternative approaches e.g., polynomial chaos and neural networks have been proposed [13–15]. More examples of GP emulation in tsunami simulation can be found in recent publications [16–20]. The drawbacks of this prior research are that (i) computational costs for generating training data can still be too expensive and (ii) most of these approaches lack flexible experimental design, although Salmanidou et al. [19] uses a sequential design strategy, Mutual Information for Computer Experiments (MICE) [21]. The flexibility in the design of experiments is crucial in an expensive simulation setting since only a small number of runs are allowed under a realistic budget of computational time. The Latin Hypercube design is a common design strategy. However, MICE is considered a more efficient way of sampling in this circumstance, as it maximises the gain over the input parameter space. As for reducing the computational time for simulation, the multi-fidelity approach is rarely applied, with only one example to our knowledge, the sparse-grid interpolation [22], which has not been tested using an efficient design of experiments.

Our new framework of a GP emulation, called Multilevel Adaptive Sequential Design of Computer Experiments (MLASCE) [23], is intended to overcome these two difficulties by both reducing the number of costly simulations and rendering the sampling much more efficient, through the combination of a multi-fidelity approach and the adaptive design algorithm MICE. To our knowledge, this is the first time that multi-fidelity GP emulation is implemented for future earthquake-generated tsunami simulations. In the following, we show how our statistical surrogate effectively replaces the numerical tsunami model and provides fast and precise uncertainty quantification of future tsunami hazard over the city of Cilacap in West Java, Indonesia.

2 Tsunami simulations

For tsunami simulations, we apply the well-validated and popular numerical package JAGURS [24–27]. The code solves linear and nonlinear shallow water as well as the Boussinesq equations and has been efficiently used in the past to model earthquake- [24, 25] and landslide-generated tsunamis [28, 29]. The computation of the crustal displacement is integrated in the code following the Okada dislocation model [30]. The bathymetry and elevation data is derived from the General Bathymetric Chart of the Oceans (GEBCO_2019 Grid product at a resolution of 15 arc-seconds) and the Geospatial Information Agency of Indonesia (National Bathymetric dataset, BATNAS, at a resolution of 6 arc-seconds and digital elevation dataset, DEMNAS, at a spatial resolution of 0.27 arc-seconds) and were compiled as in Salmanidou et al. [5].

The utilisation of nested grids allows us to focus on the areas of interest to study the impact of large earthquake scenarios. In total, we use 3 nested grids spanning across three different resolution layers with a grid ratio of 1 : 4. The first layer is the background grid SD01 (level 1 grid) with longitudes of 103°E-119°E and latitudes of 05°S-13°S and with a spatial resolution of $\Delta x = \Delta y = 0.004°$. The intermediary grid SD02 (level 2 grid) has a geographical domain spanning from longitudes of 107.02°E-112.48°E and latitudes of 08.98°S-07.02°S with $\Delta x = \Delta y = 0.001°$. The third grid layer SD03 (level 3 grid) focuses on the area with longitudes of 108.52°E-109.28°E and latitudes of 07.8°S-07.62°S with $\Delta x = \Delta y = 0.00025°$ and includes the city of Cilacap, Indonesia. Figure 1 shows the visualisation of these grid files.

3 Specifying source parameters

The Eastern Sunda Arc zone spanning from Sunda Strait, in the northwest, to Flores Island, in the southeast, is one of the Indonesia's six tsunamigenic zones. In the last century, circa 50 $M_w > 6.5$ earthquakes have occurred in this seismically-active region (US Geological Survey catalogue[1]). The seismicity in the Java trench is originated from the subduction of the Indo-Australian Plate beneath the Sunda Plate. We therefore focus on a range of tsunamigenic earthquakes at the subduction zone of the Java trench to study the maximum run-up heights of a potential tsunami on the shoreline of Cilacap. The numerical tsunami model, JAGURS, is run over the region (Fig. 1) and the scenarios are determined by the source parameters reflecting the characteristics of this subduction zone. We define these events below. The area for our emulation is given in Fig. 2.

---

[1] https://earthquake.usgs.gov/earthquakes/search/
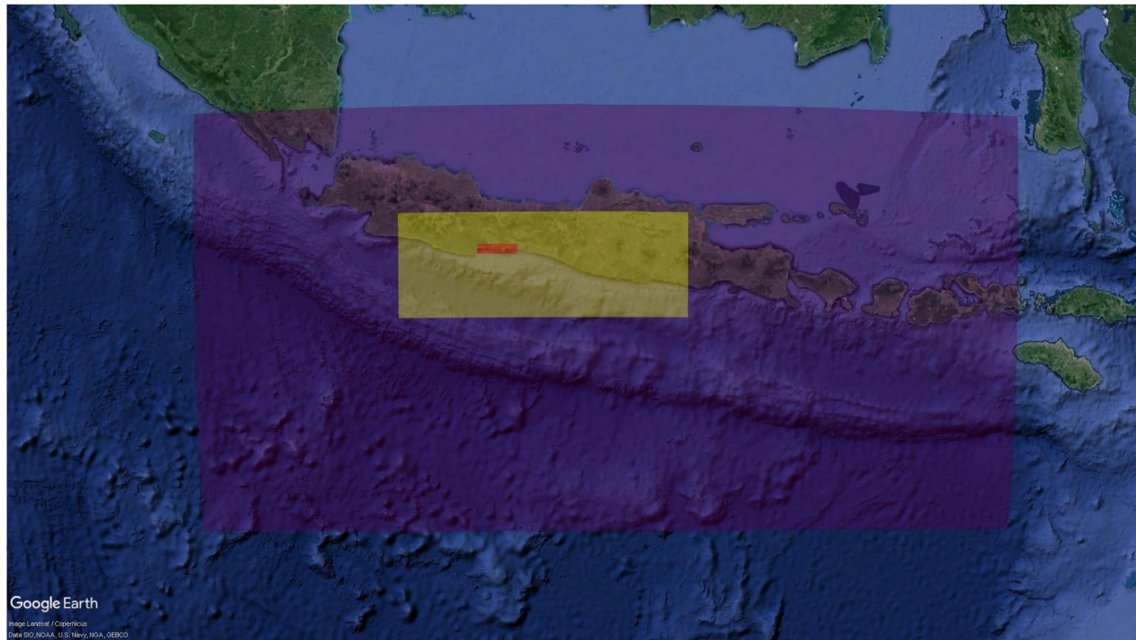
Fig. 1 The nested grids. The blue domain denotes the level 1 grid file (SD01), the yellow shows the level 2 grid (SD02) and the red corresponds to the level 3 grid (SD03)

We define $h_{\max}(\phi, \theta, x)$ as the maximum wave height at a specific gauge in the time period of $t \in [0, T]$ for a specific set of initial source parameters x:

$$h_{\max}(\phi, \theta, x) = \max_{0 \leq t \leq T} h(t, \phi, \theta, x) \tag{1}$$

where ϕ, θ denote respectively longitude and latitude of the gauge, $h(t, \phi, \theta, x)$ is the wave elevation at time t for a location ϕ, θ and tsunami source parameters x, and T is the end of the simulation time period.

These source parameters dictate the initial conditions of the equations based on the Okada dislocation model [30] and are listed in Table 1. Therefore, we establish the input-output structure by specifying the dependency of the output value $h_{\max}(\phi, \theta, x)$ on these source parameters x as the inputs, for each fixed value of the location of the gauge ϕ, θ.

To implement the designs of the scenarios, the ranges and empirical relationships for these parameters are determined. The free parameters are latitude, longitude, and slip. Dip, strike, and rake are fixed to 10.0°, 272.0°, and 90.0°, respectively, throughout the simulations. The values of the other source parameters, depth, length, and width depend on the free parameters. The longitude and latitude of the fault are assumed to be in the area between the Java trench and the contour line with depth = 60 km (specified in green and red lines in Fig. 3). More specifically, Longitude values range from 103°E to 119°E and for each longitude, latitude values are constrained to be within the corresponding range from the southern (the trench) and northern (the contour line with depth = 60 km) limits described in Fig. 3).

The data from Slab2 [31] are used to determine the depth parameter. Slab2 is the most recent model describing the 3D geometries of all seismically active subduction zones worldwide. Fault contour depths are available so that a polynomial approximation with third order is used to interpolate the depth of the fault depending on its location. The farther the scenarios from the trench, the deeper the depth of the earthquake becomes.

The slip takes its value between 2 to 12 m assuming earthquakes of larger magnitudes than the 2006 Java earthquake as described in Fujii & Satake [32], which investigates the details of 2006 Java earthquake and estimates the slip parameter. In order to decide the other parameters (length and width), the empirical formula based on Allen et al. [33] is used. While Wells & Coppersmith [34] is well known for providing the versatile empirical equations, Allen et al. [33] offers the updated equations reflecting more recent studies. In addition, the equations by Allen et al. [33] are developed based on more large earthquakes in their database, which makes them more suitable for this study because this work involves large-magnitude scenarios. The length and width are determined by the following equations from Allen et al. [33].

$$\log_{10}(\text{length}) = -3.03 + 0.63(\text{magnitude}) \tag{2}$$

$$\log_{10}(\text{width}) = -1.01 + 0.35(\text{magnitude}) \tag{3}$$

$$\text{magnitude} = \frac{\log_{10}(\text{slip}) + 4.73}{0.71}. \tag{4}$$

We run for $T = 14400$ s. We validate this choice by running two simulations with the faults at the east and west
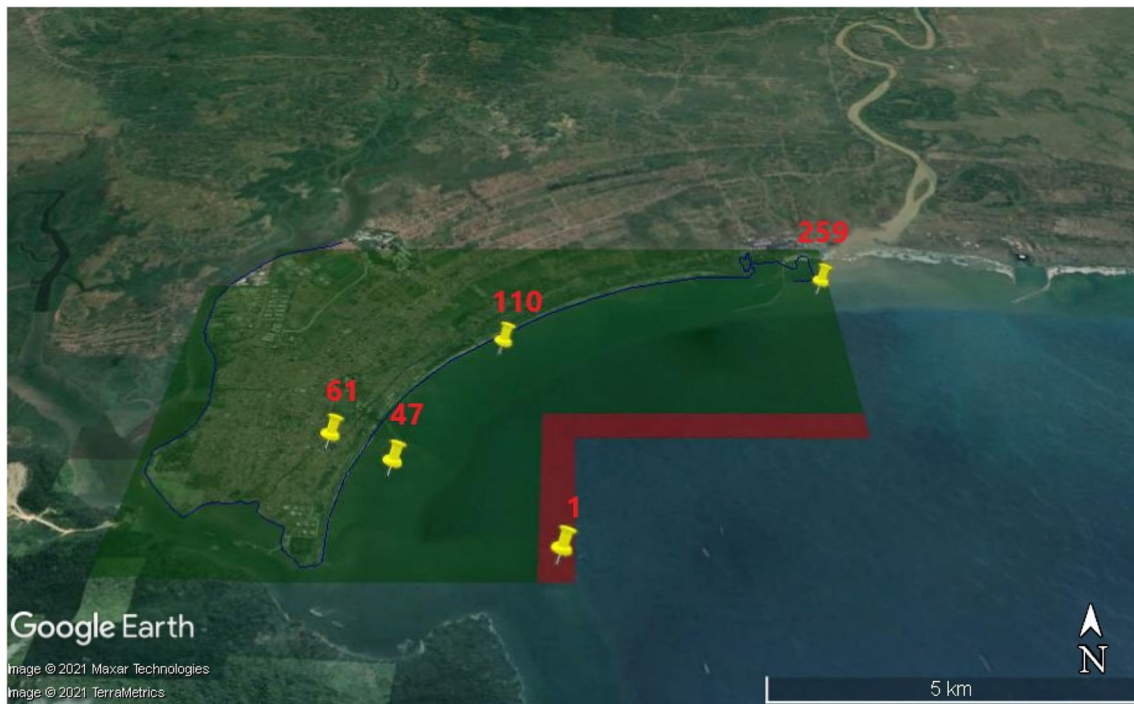
Fig. 2 The area for emulation. It covers the sub-districts of Cilacap. The area is divided into 259 squares with the side length = 0.005° (approximately 550m). Every gauge is located at the centre of the square cell. The yellow pins are specific gauge locations (whose number is specified above the pin) used in the sequel. The 15 gauges located at the centre points at the red squares will be averaged as our quantity of interest in the output employed in the design of computer experiments

ends of the domain (Fig. 4) which demonstrate that we can capture the maximum elevation of the tsunami.

Our earthquake sample is comprised of 44 events in the magnitude range of M7.0 - M8.2 and a slip range of 2.0 m - 12 m. The slip of 2 m corresponds to an event with M7.0 and the slip of 12 m correlates with an M8.2 event (Fig. 5b). We note that the largest recorded events in the region are M7.7 (2006) and M7.8 (1994) [5]; therefore, the upper range of our earthquake scenarios in this research (M8.2) is considered larger than the recorded maximum event (M7.8) in order to account for potential future large earthquakes. As shown in Fig. 5b, the earthquake events are approximately equally distributed in the magnitude range of M7.0 - M8.2. For each scenario, a uniform slip distribution on the fault surface is considered with no heterogeneity.

Table 1 Source parameters in JAGURS

Latitude	Longitude	Depth (km)	Length (km)	Width (km)
◯	◯	×	×	×
Dip (°)	Strike (°)	Rake (°)	Slip (m)	
10.0°	272.0°	90.0°	◯	

In the bottom row, ◯ means the corresponding source parameter is a free parameter and × indicates that the source parameter is dependent on the other parameters via the empirical formulas (2), (3) and (4)

4 Multi-fidelity GP emulator

The GP emulator is a versatile statistical surrogate that mimics the input-output relationship of the tsunami simulator. It is trained over a set of input points, called a design, whose size is much reduced compared to the number of predictions. The source parameters (points in the space of input parameters) for training are chosen with the specific purpose of capturing the input-output relationship. This can be done for instance via one shot designs such as Latin Hypercubes, or the sequential design algorithm MICE (mutual information for computer experiments) that adaptively selects the input values where the computer simulator is evaluated to maximize the expected information gain (mutual information) over the input space [21].

Despite the computational efficiency of JAGURS using multiple nested grids (only running the finest resolution locally around Cilacap), the computational cost per run still makes the simulations for many inputs very challenging. Therefore, we employ MLACSE [23], a multi-fidelity GP emulator that exploits MICE to emulate the most sophisticated version of JAGURS up to the finest level. In this approach, we only run JAGURS up to the finest resolution over a small subset of the runs, thereby saving large amounts of computing time. The level 1 version uses only a single grid file SD01. The two nested grid files SD01
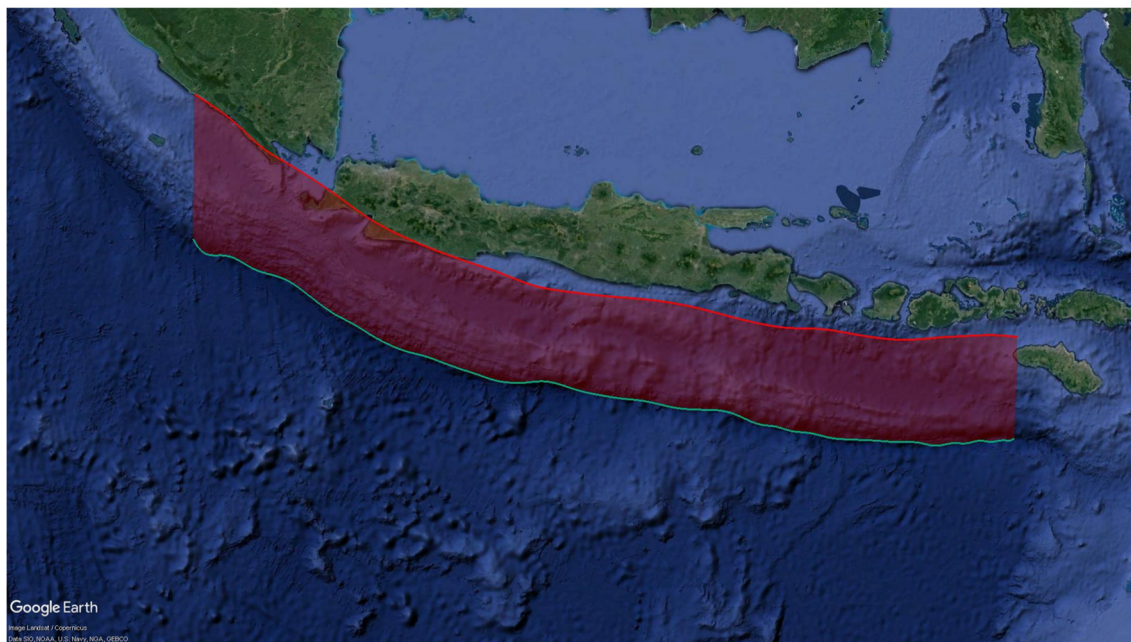
Fig. 3 The area for the experimental design, longitude and latitude of the fault location. The northern red line is the contour line with the depth 60km (Slab2 [31]) and the southern green line denotes the Java trench. The location of the fault is assumed to fall in the area filled in red

and SD02 are used for the level 2 version. The level 3 version is the most expensive and it requires three nested grids, SD01, SD02, and SD03. As the level of refinement increases, the accuracy of the simulation improves. Thus, the level 3 version is the most accurate whereas the level 1 provides the least accurate representation. We also assign the level l simulation to the level l version of the grid ($l = 1, 2, 3$). In this paper, the configurations of JAGURS are the same across every level, except for the number

of the grid files. In other words, the various versions of JAGURS' simulations differ only in the number of the nested bathymetry files. The time step in the finite difference method is 0.2 seconds for every version of JAGURS. This time step is determined so that we can avoid the numerical instability of Finite Differences calculations according to the Courant Number criterion (https://github.com/jagurs-admin/jagurs).

Let x denote the three-dimensional input vector composed of the slip, longitude and latitude. The ranges of these variables are given in the previous section ($[2, 12] \times [103, 119] \times [0, 1]$). The latitude is first expressed as the proportion from the southern boundary to northern taking its values in $[0, 1]$ and then converted into the actual value depending on the longitude. Furthermore, for a fixed value of the location of the gauge ϕ, θ corresponding to a gauge j ($j = 1, \ldots, 259$, see Fig. 2), we define $h_{l,j}(x)$ ($l = 1, 2, 3$) as the maximum wave elevation, i.e. $h_{\max}(\phi, \theta, x)$ for the source parameters x, over the 4 hours of simulation given by the level l version of the grid set-up. We also define the difference of the outputs from two successive grid levels as follows and the telescopic sum of differences adds up to the highest resolution simulation.
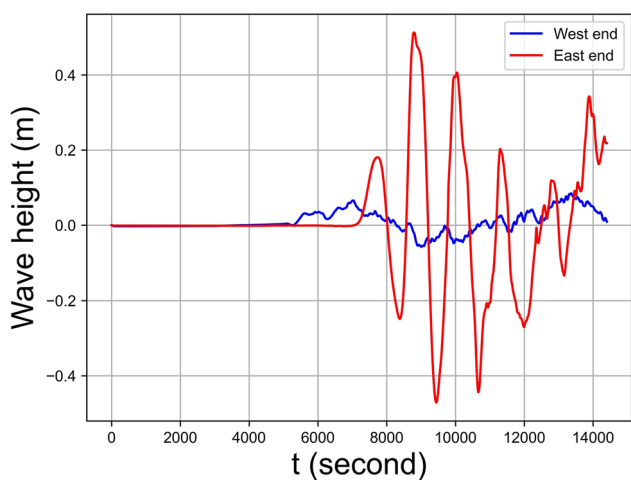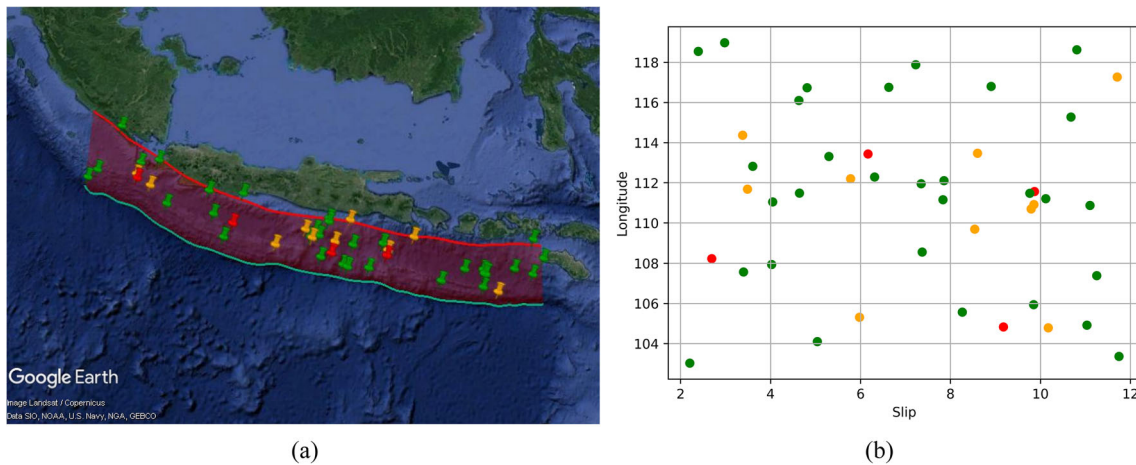

Fig. 4 Time series of wave height at gauge 1 (Fig. 2, longitude= 109.0525°E, latitude = 7.7525°S). The red and blue lines correspond to the scenarios where the rupture occurs at the east and west end. The location of the west end is 103°E and 7.23°S and that of the east end is 119°E, 10.31°S. Slip is 12m in both cases

$$\delta_{1,j}(x) = h_{1,j}(x) \tag{5}$$

$$\delta_{l,j}(x) = h_{l,j}(x) - h_{l-1,j}(x) \quad (l = 2, 3) \tag{6}$$

$$h_{3,j}(x) = \sum_{l=1}^{3} \delta_{l,j}(x) \tag{7}$$

(a)   (b)

Fig. 5 Experimental design obtained by MLASCE. The green pins denote the sample points of $\delta_1(x)$, orange pins $\delta_2(x)$ and red pins $\delta_3(x)$. Design points shown (a): across longitude and latitude of $\delta_1(x)$, $\delta_2(x)$ and $\delta_3(x)$. (b) across slip and longitude of $\delta_1(x)$, $\delta_2(x)$ and $\delta_3(x)$

For the purpose of choosing the design sequentially and across levels, we need to provide an output quantity of interest that is used to determine the number of samples (and their actual locations at each level) as shown in Table 2. We have employed in the past one representative output location [19] to create the sequential design of computer experiments. Indeed one point is enough to generate a diverse design that reflects the input-output relationship: there might be some different dependencies elsewhere in the outputs but these may not affect much the design's ability to pick up the major modes of variation effectively. However, in our new context of multiple resolutions, to make sure that we are not providing a representative point in the outputs that might skew the design towards a particular region of the input space due to some spurious edge effect of cells across levels, we decided to increase the robustness of the quantity of interest. This quantity is now the average of maximum wave heights at 15 gauges (the centre points at the red squares in Fig. 2). It covers a region that is near shore, over a region that covers multiple cells at all resolutions, to pick up sensitivity near the locations of interest (onshore for inundation), but not onshore itself where inundation might result many zero values and where the topography is having a very different impact across levels, possibly degrading the design. The choice of the L shape for these 15 gauges is

rather arbitrary as it suits these constraints well. Note that any reasonable set of gauges, with any geometry would be suitable, as long as the input-output relationship enables the sequential design across level to pick up variations and allocate points in the input space providing the most information for the design to be most efficient. Note that we computed these arbitrary average across cells for all levels, but these 15 cells used the center as a representative, in order to simplify the computation but an alternative could be to compute the matching averages over cells for all resolutions. The mismatch is not consequential in these area as we also average over 15 cells to dampen any very local effect. The key point is to generate a reasonably efficient design with points located in the input space where the action is. This average over 15 cells is used in implementing MLASCE: $h_1(x)$, $h_2(x)$, and $h_3(x)$ are the average of the maximum wave heights at these locations from the level 1, level 2, and level 3 grids used in the JAGURS simulations.

$$h_l(x) := \sum_{j=1}^{15} h_{l,j}(x)/15 \quad (l = 1, 2, 3)$$

$h_3(x)$ is decomposed into the sum of the increments in the same way

$$\delta_1(x) = h_1(x) \tag{8}$$
$$\delta_i(x) = h_i(x) - h_{i-1}(x) \quad (i = 2, 3) \tag{9}$$
$$h_3(x) = \sum_{i=1}^{3} \delta_i(x). \tag{10}$$

By decomposing $h_3(x)$ into the incremental functions $\delta_i(x)$, we can avoid the huge cost of running $h_3(x)$ numerous times and construct a good approximation of $h_3(x)$ by utilising the emulation of the cheaper simulations $h_1(x)$ and $h_2(x)$ and add the incremental effect of $\delta_3(x)$ where necessary. We can

Table 2 The typical cost (computational time) for each simulation

	δ_1	δ_2	δ_3
Time (s)	1885(s)	7739(s)	13612(s)
No of Samples	30	10	4

The bottom row shows the numbers of samples obtained from implementing MLASCE under the total computational budget 54 hours. 8 cores were used for implementing every simulation

record the maximum wave height at every gauge (Fig. 2) per run of $h_1(x)$ given a certain x. Similarly, by running SD02 and SD03, the difference of these maximum wave heights between the two grid versions in JAGURS is also stored. During the process of MICE, we construct mutually independent GPs $\eta_l(x) \sim GP(0, K_l(x, x'))$ $(l = 1, 2, 3)$ to emulate $\delta_l(x)$. We employ the Matérn 5/2 kernel for the choice of $K_l(x, x')$. This kernel is smooth enough to avoid the GP becoming too rough whilst not being excessively smooth and it is a common choice for modelling physical relationships [11, 12].

The total budget available for our computation is 54 hours. We treat the computational time of the simulations as the cost. The costs of these 3 simulations are given in Table 2. After implementing MLASCE, we obtain the experimental designs $X_{l,N_l} = (x_{l,1}, \ldots, x_{l,N_l})^\top$ and $\delta_{l,j}(X_{l,N_l}) = (\delta_{l,j}(x_{l,1}), \ldots, \delta_{l,j}(x_{l,N_l}))^\top$ for $l = 1, 2, 3$ where N_l is the number of the design points at level l and $x_{l,n}$ $(n = 1, \ldots N_l)$ denotes the three dimensional input vector chosen by MLASCE. We construct the multi-fidelity GP emulators based on the record of $\{X_{l,N_l}, \delta_{l,j}(X_{l,N_l})\}$ by fitting the mutually independent GPs $\eta_{l,j} \sim GP(0, K_{l,j}(x, x'))$ to each $\delta_l(x)$ for every gauge j where $K_{l,j}(x, x')$ is the Matérn 5/2 kernel for every l and j. We use the posterior mean function $m_L^{\delta_L,j}(x) = \sum_{l=1}^L m_{l,j,N_l}^{\delta_l,j}(x)$ $(L = 1, 2, 3)$ for the level L emulation of the maximum wave height at gauge j given a new input x. The expression of $m_{l,j,N_l}^{\delta_l,j}(x)$ is given by the usual kriging equation:

$$m_{l,j,N_l}^{\delta_l,j}(x) = K_{l,j}\left(x, X_{l,N_l}\right)^\top K_{l,j}(X_{l,N_l}, X_{l,N_l})^{-1} \delta_{l,j}(X_{l,N_l})$$

5 Emulation results

5.1 Experimental design and emulated wave heights

By implementing the above-mentioned procedure, MLASCE, we collected 30 evaluations for the level 1 version of JAGURS'simulations, 10 for the level 2 and 4 for the level 3, see Table 2. These source parameters are shown in Fig. 5a and b.

In order to understand how the maximum wave height is emulated, we choose three representative gauges 259, 110, and 47 on the shoreline (see Fig. 2). Here, we also create another emulator, the emulator based on the samples only from the finest version of the grid set-up, denoted $h_{3,j}(x)$ (under the same budget), and coined the single level emulator. The comparison with the single level emulator provides an insight into the usefulness of the multi-fidelity emulators. Figure 6a – c (longitude-slip coordinate) and 7a – c (longitude-latitude coordinate) show the two-dimensional plots of the multi-level emulators $m_L^{\delta_L,259}(x)$, $m_L^{\delta_L,110}(x)$ and

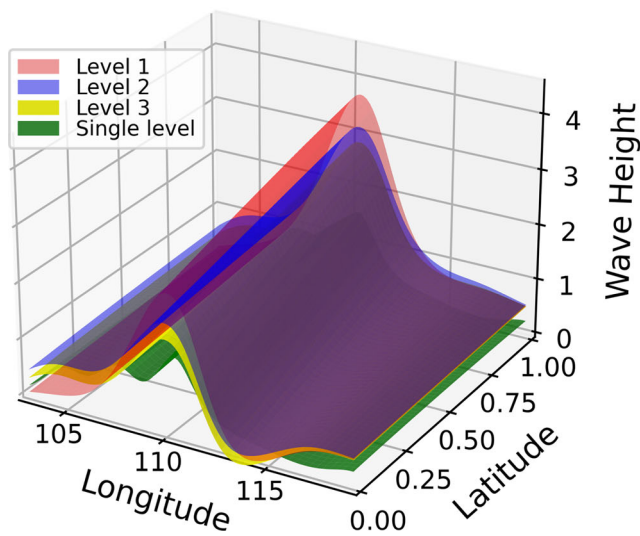(a) The emulators at gauge 259.

(b) The emulators at gauge 47.
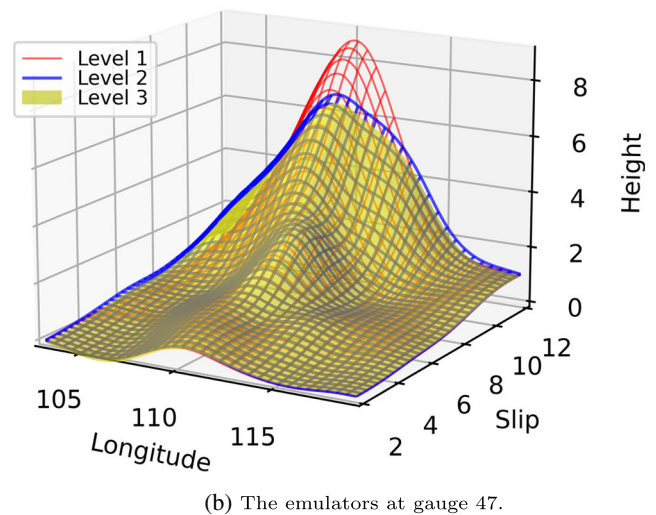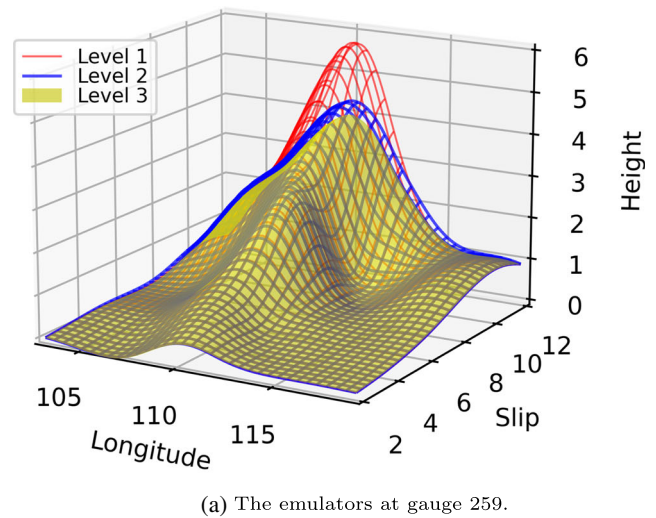
Fig. 6 Multi-level emulators of wave height simulated by JAGURS at three gauges against longitude and slip (latitude fixed corresponding to a depth of 40 km). Red wireframe: level 1. Blue wireframe: level 2. Yellow surface: level 1. The Z-axis denotes the maximum wave height (meters), the X-axis and Y-axis denote the longitude (°) and slip (m)

$m_L^{\delta_L,47}(x)$ $(L = 1, 2, 3)$. In Figs. 6a – 7b, the clear pattern of earthquake scenarios are presented. Overall, the maximum wave height is expected to be higher as the value of slip increases while the location of the fault exhibits a different pattern.

Since the longitude of Cilacap is approximately 109°E, the emulators give the higher maximum wave height as the fault location nears 109°E. Therefore, we have to pay more attention to the earthquakes occurring in the area close to the longitude 109°E. The level 3 emulator modifies the pattern given by the level 1 and 2 emulators and the level 2 emulator provides the larger corrections to the level 1 emulator at this gauge location. In this specific case, level 1 simulations tend to give higher wave elevations than the higher level SD02 and SD03 simulations do. This is
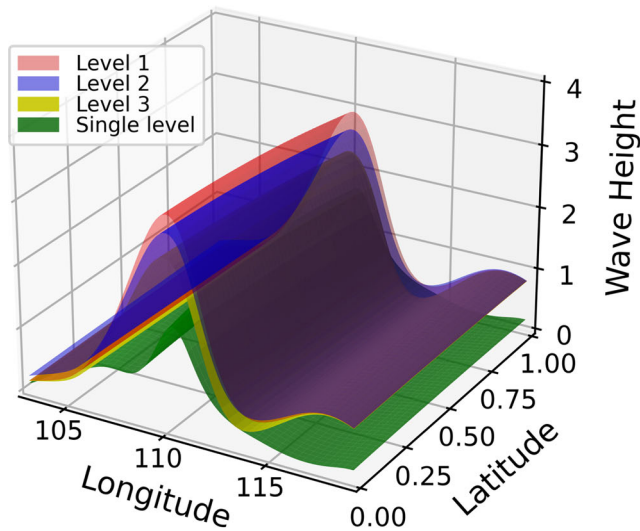
(a) The emulators at gauge 259.

(b) The emulators at gauge 47.

Fig. 7 Multi-level emulators of wave height simulated by JAGURS at three gauges against longitude and latitude (slip fixed at 8 m). Red surface: level 1. Blue surface: level 2. Yellow surface: level 3. The single level emulator is. represented by a green surface. The Z-axis denotes the maximum wave height (meters), the X-axis and Y-axis denote the longitude and latitude

partially explained by the fact that the level 1 grid (SD01) has the coarser spatial resolution, thereby ignores the details of the geographical characteristics, such as the obstacles impeding the tidal wave. Thus, the higher level simulations modify the tsunami propagation by accounting for the higher resolution topography and coastline. Figure 8a (for the level 2 increment) and b (for the level 3 increment) show how the predictions by the level 1 emulator are corrected as the uncertain inputs change. The value of the latitude of the rupture point is relatively non-influential as we can

see very small variations in Fig. 7a – c. As noted earlier, the longitude influences the maximum tsunami elevations to a large extent and the amount of slip directly affects the predicted values. A similar pattern occurs in other locations, as for example in gauges 47 and 110. In particular, Fig. 6b shows the sharp modifications to the maximum wave height given by the level 2 and level 3 emulators. The level 1 simulation tends to overpredict the maximum wave height at this location. The complexity of the bathymetry is only captured by the computationally expensive fine mesh grids and the multi-fidelity method is beneficial in incorporating this information where necessary.

We can see the vivid contrast between the multi-fidelity emulator by MLASCE and the single level 3 emulator in Fig. 9a – c. We compare the multi-fidelity emulator with one built from an ensemble of simulations at nested level 3 only. Only 24 simulations can be created using MICE for the single level 3 v. a set of 30, 10, and 4 simulations respectively across levels 1, 2, and 3 for the multilevel emulator, with the same budget. The multilevel emulator at level 3 is substantially superior to the single level emulator at level 3 (as can be seen in the next section). The differences are large since the shape is complex and the single level emulator cannot be built properly with less than 10 runs per parameter (i.e. 30 runs for the three parameters), the usual rule of thumb for GP fitting. cOur multi-fidelity emulators show the clear dependence on the longitude of the rupture location while the latitude does not seem significant. The same pattern of overestimation at level SD01 and correction at levels SD02 and SD03 occurs at other gauges.

5.2 Validating the result

In the previous section we concluded that the quality of the physical representation of the emulated input-output relationship of the multi-fidelity was different from the one obtained by a single level GP emulator at level SD03 (using the same training budget). Although the multi-fidelity GP emulators give plausible results, validating the reliability of these outputs is an indispensable task.

First, we validate the reliability of the emulations of the increments $(\delta_{l,j}(x))$ and employ Leave-One-Out (L-O-O) diagnostics for this purpose, which is a common way of validation in the community [11, 12]. Our three sets of training data of the three levels, which are obtained from evaluating the level 1, 2, and 3 simulators, respectively consist of 30, 10, and 4 pairs of input–output quantities of $\delta_{l,j}(x)$ (l: level, j: gauge). In L-O-O, a reduced training set (dropping one sample from the original data set) is employed to build an emulator, which is then used to predict the output at inputs of the one sample that was left out. The predicted output is compared with the actual output of this sample. We drop one sample point from the training sample

Fig. 8 The emulations of the increments at gauge 259 are represented by the green (a) and red (b) surface. The Z-axis denotes the maximum wave height (meters), while the X-axis and Y-axis denote the longitude and slip. The depth is 30 km
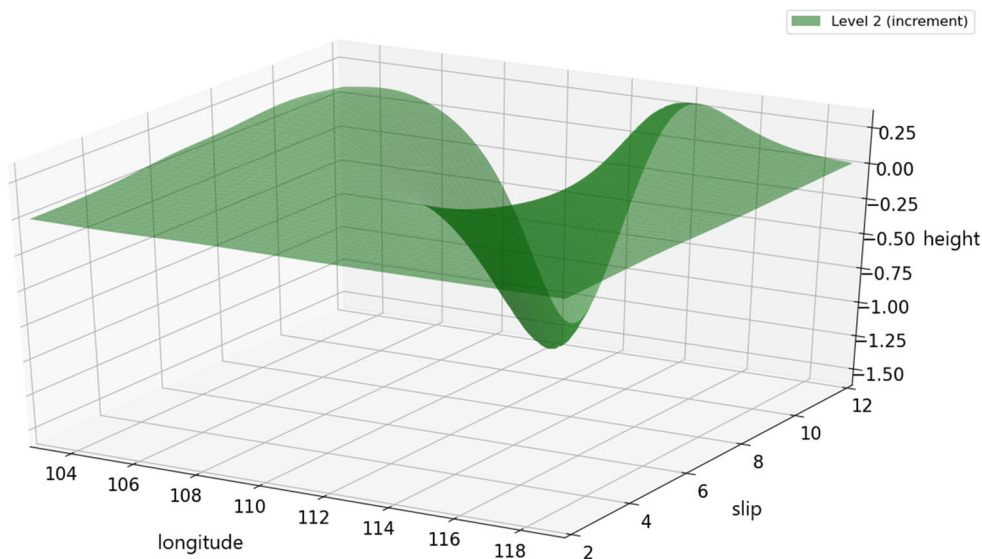

(a) Level 2 increment $\delta_{2,259}(x)$.
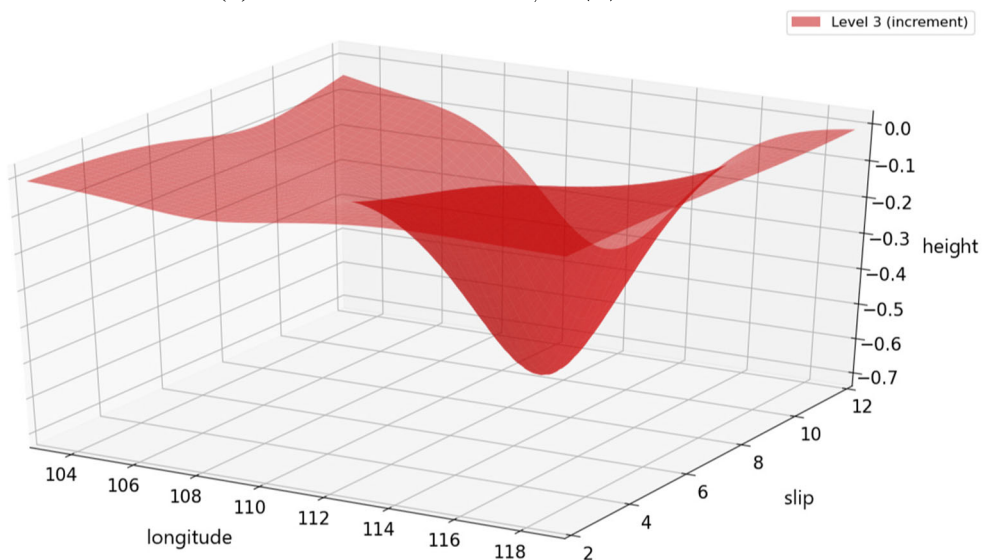
(b) Level 3 increment $\delta_{3,259}(x)$.

for validation, so it is not used in training. To show that our multi-fidelity emulators perform well regardless of the specific gauges, we compute the average of the root mean square errors between the predicted and true values at all gauges. As the prediction variance is seen as the standard tool of measuring the reliability of the prediction values in the community (e.g. [11]), we also calculate the average value of the prediction variances accordingly. Table 3 shows the result of this L-O-O for the emulations of the increments $\delta_{l,j}(x)$ for every level at all gauges. The errors are low for level 1 considering the values (RMSE of 0.42 for up to 12 m wave heights, but only acceptable for levels 2 and 3 as these differences take values over a range of roughly 1.5 and 0.7 m respectively for errors of 0.92 m and 0.29 m. The

level 2 shows a slightly higher variance proportionally when compared to the other levels, which is attributed to the small number of samples compared to its variations.

Second, we validate the results of the performance of the multi-fidelity emulators, comparing the predictions to the true output given by the level 3 simulation. The input-output data of the level 3 simulation is the same as the training data set used for building the single level emulator SD03. None of these data points are included in the experimental designs for our multi-fidelity emulators. We compute the average of the root mean square errors between the true output value and the predictions by the multi-fidelity emulators (level 1, 2, and 3) evaluated at the corresponding inputs. The average of the prediction variances is also calculated. Table 4 shows

Fig. 9 The yellow surface represents the level 3 (multi-fidelity) emulator. The green surface represents the single level emulator (level 3 simulations only). The depth is 40 km. The Z-axis denotes the maximum wave height (meters), while the X-axis and Y-axis denote the longitude and slip
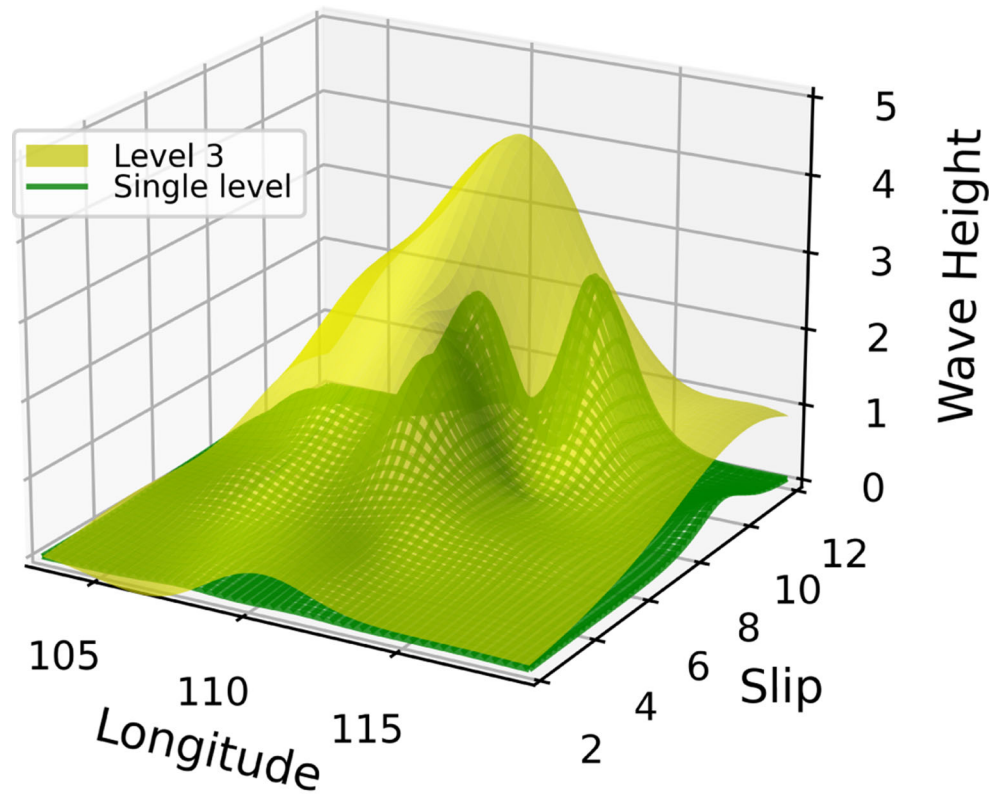
(a) The emulators at gauge 259.
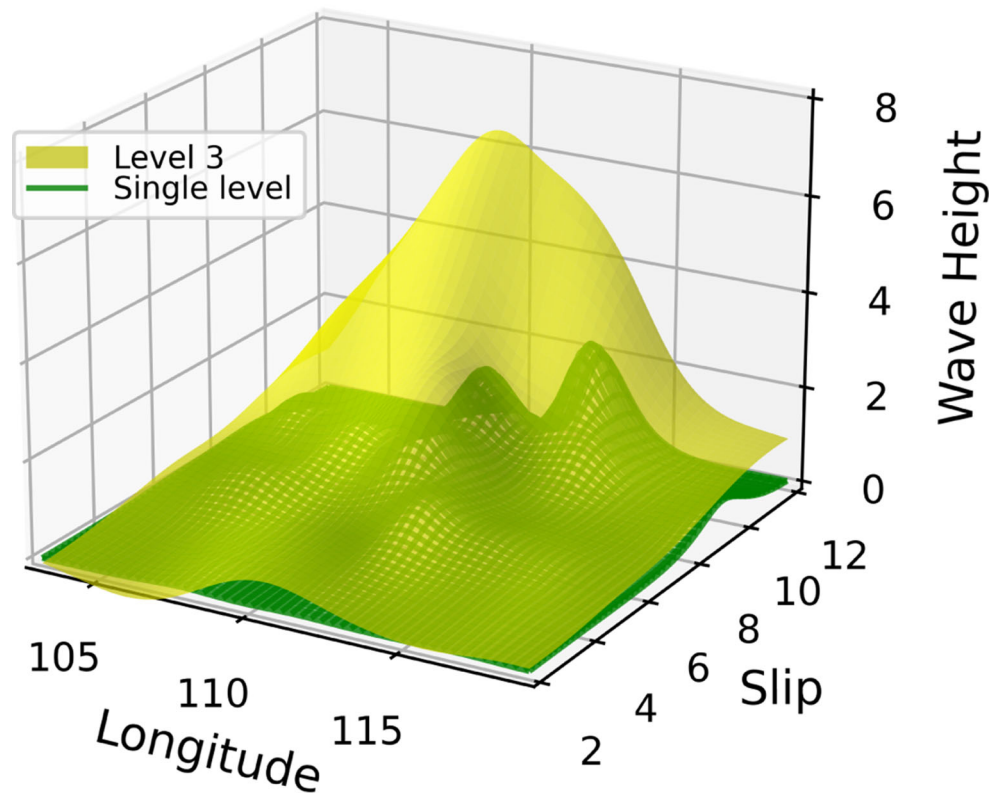
(b) The emulators at gauge 47.

Table 3 L-O-O for the emulations of the increments $\delta_{l,j}(x)$.: Root mean square error (RMSE in meters) and average of predictive variances (σ_p)

Increment	Level 1	Level 2	Level 3
RMSE (m)	0.41975	0.92617	0.28566
σ_p	0.3000	0.5093	0.1360

the performances provided by the multi-fidelity emulators. As the level increases, the prediction error tends to decrease although the difference between the level 1 and 2 emulators is small, which is partially explained by observing that the number of the data points for constructing the emulation of the increment $\delta_{2,j}$ is relatively small compared to the values of $\delta_{1,j}$. The prediction variance increases as the level increases since we simply add up the predictive variances of the emulation of each increment $\delta_{l,j}$. In other words, the uncertainty of emulating $\delta_{l,j}$ of each level is incorporated into the higher levels of the multi-fidelity emulators.

Finally, we show the comparison between the predicted and true maximum wave height at all of the gauges 1-259. Here, the true maximum wave height means the evaluation given by the level 3 grid set-up. The source parameters are fixed in order to evaluate a catastrophic scenario, which is represented by the longitude 109°E, latitude 8.554°S (corresponding to the depth 40 km) and slip 9m, which is not quite close to the training samples for every emulator. Figure 10b shows the result given by our multi-fidelity emulator and Fig. 10c is the true output from the level 3 simulation. The maximum wave height on the land is approximated by our emulator, and both our emulator and the simulation present similar maximum water elevation patterns. However, our emulator tends to give seemingly erroneous higher wave height (albeit by a tiny amount only by less than 20 cm on land). The reason is quite similar to that explained before. Since the level 1 grid set-up is unable to consider the detailed information of the geographical obstacles in its computation, it is likely to produce an overestimated result in this case. On the other hand, the single level emulator based on only the samples from the level 3 simulation seems inadequate in predicting the inundation in Cilacap (Fig. 10c). As explained, the predictions by this single level emulation are heavily skewed towards around 0 meter hence the

Table 4 Performance by the multi-fidelity emulators using 1 (one fidelity), 2 or 3 levels: Root mean square error (RMSE in meters) and average of predictive variances (σ_p)

| Increment	Level 1	Level 2	Level 3
RMSE (m)	0.87750	0.87241	0.80777
σ_p	0.2414	0.6571	0.7613

predicted wave heights show the similar pattern. Under the limited budget of computational resources (time), the multi-fidelity emulations perform much better than the single level emulation. Note that the squares in Fig. 10b–c (the same squares in Fig. 2) are not the grids of any of the bathymetry files.

5.3 Uncertainty propagation

Once the emulators are built, the maximum tsunami elevation can be predicted for any input scenario. The prediction involves the utilisation of the built emulator with a given set of inputs to calculate the mean predictions and their uncertainty on the outputs. Uncertainties are fully propagated to display sometimes complex distributions of outputs such as skewed distributions (as in the case below): variance would not be enough to describe such uncertainties. As the papers investigating tsunami emulations demonstrate [16, 18, 19], emulation provides a complete description of uncertainties through the full pdf not just mean and variance as in traditional analyses. The uncertainties in the inputs can be quantified by their distributions, which can be modified to investigate different scenarios. A beta distribution and uniform distribution are the common tools and employed for each parameter, from which 2,000 scenarios are randomly selected.

The shape parameters of the distributions can be utilised to express the scientific knowledge about the source. There are three free input (source) parameters, longitude, latitude, and slip, and we prepare the corresponding beta distributions for latitude and slip and uniform distribution for longitude. Unfortunately, little is known about the characteristics of these parameters. Therefore, we set the shape parameters for the slip so that the mode is 10, which corresponds to the upper range of earthquake scenarios that we wish to explore here. Due to the wide range of the longitude, it is unlikely that the location of the rupture with respect to the longitude will be known in advance, a uniform distribution is therefore used to express this uncertainty. We set the mode of the latitude as 0.5, which is the central value of the domain of experiment, and the shape parameter of this beta distribution is determined accordingly. Note that due to the complex shape of the boundary of the domain of experiment, the design of the latitude takes its value in [0, 1], which is the proportion from the southern to the northern boundary, then transformed into the actual value.

The distributions of these three source parameters are shown in Fig. 11a – c. The shapes of these histograms match the original theoretical distributions we have chosen. We selected two representative points, one offshore and one on land (Fig. 2), to describe the distributions of the maximum wave elevations resulting from these source parameters as inputs. Figure 11d shows the result. As demonstrated in
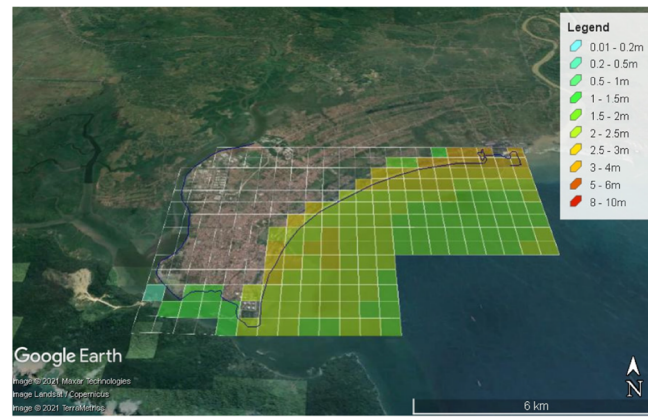
(a)
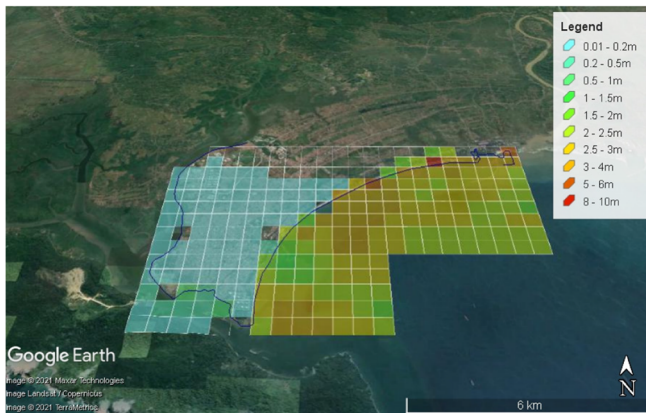
(b)
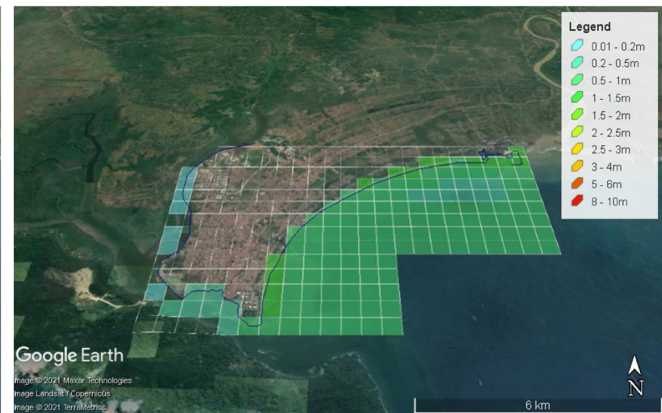
(c)

Fig. 10 Comparison of emulated and true wave heights (meters) with the same source parameters: (a) true maximum wave heights computed by the level 3 simulation at the gauges. (b) multi-level emulation of the maximum wave heights at the gauges. (c) single level emulation of the maximum wave heights at the gauges. Slip 9m, longitude 109°E, and latitude (8.554°S) that corresponds to the depth 40 km

Section 5.1, the maximum wave elevations crucially depend on the rupture location even when the value of the slip is close to its maximum. Figure 11d corroborates this observation from a different angle. The two distributions are skewed toward the minimum value, 0 - 1 meters, since the tsunami elevations drastically increase only when the longitude of the rupture place is close to Cilacap. However, if the rupture occurs at around longitude 109°E, higher wave heights are predicted and these phenomena cannot be ignored since the distributions are relatively fat-tailed.

We further offer 50th and 90th percentiles of the tsunami heights at all gauges (Fig. 12a and b). Figure 12a describes that relatively high tsunami elevations (1-2 meters) can occur even on the land close to the shoreline, which implies that the tsunami risk is high, with high probability. Figure 12b provides a disastrous scenario. The predicted wave height on the shoreline and nearby land is more than 3 meters and this result is consistent with the similar scenario (slip 9 m and longitude 109°E) in Fig. 10a. We note again

that the squares in Fig. 12a–b (the same squares in Fig. 2) are not the grids of any of the bathymetry files.

It should be emphasized that only a few milliseconds are necessary to obtain the entire predictions per scenario with our multi-fidelity emulator while the expensive simulation needs more than hours. Therefore, our methodology provides fast and exhaustive scenario assessments with good prediction accuracy.

6 Conclusions

In this paper, a multi-fidelity GP emulator was employed to efficiently perform many computational experiments to evaluate the earthquake-generated tsunami hazard in Cilacap in Java, Indonesia. This approach provided an informative, innovative construction of the statistical emulators, which overcomes the problem of the limited amount of computational time. It forms the first study of its kind, to the authors' knowledge, which involves the application of a

(a) Longitude

(b) Latitude

(c) Slip

(d) Maximum wave elevations
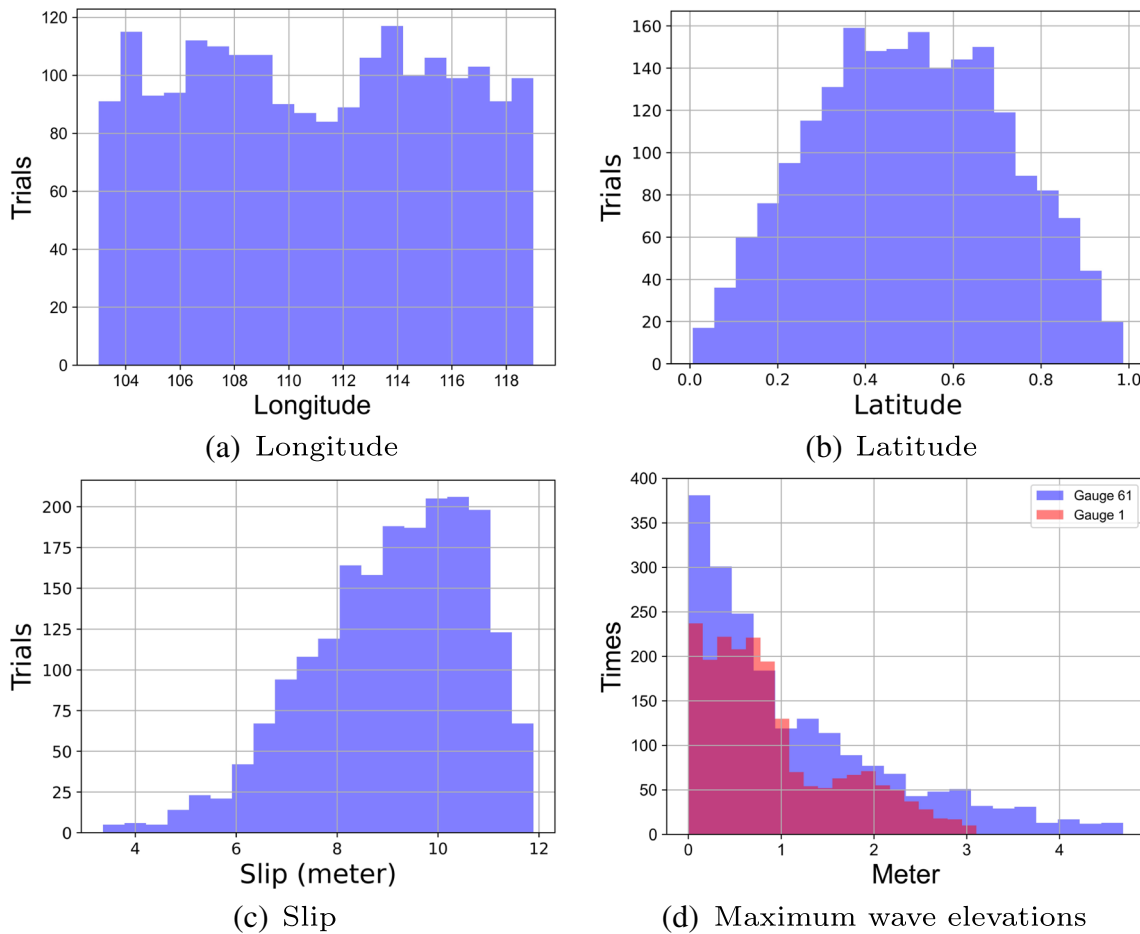
Fig. 11 Hypothetical histograms, for the set of future events considered here, of the three source parameters (11a)-(11c) used as inputs. The resulting histograms (obtained by emulation of the tsunami simulator) of the maximum wave elevations outputs at gauge 1 (red) and 61 (blue) are shown in (11d)

multi-fidelity statistical emulator with a sequential design algorithm for real-case tsunami hazard prediction. Using three levels of computations, from low to high-resolution, in Cilacap, the maximum tsunami wave heights were predicted at 259 offshore and onshore locations with the utilisation of the emulators. Following the successful construction of the emulators, 2,000 earthquake scenarios were predicted using distributions on the values of slip, and the location of the faults. These distributions express the current reflection on the possible rupture scenarios in the region but can be further refined in the future, to address added knowledge on the source characteristics. The predictions showed a high dependence of the maximum wave heights on the longitude of the rupture location as well as the amount of the slip. The tsunami elevation becomes threatening suddenly when the longitude of the rupture location becomes closer to that of Cilacap with a high amount of slip. The distributions of the maximum wave heights are heavily skewed toward 0 - 0.5 meters; nevertheless, these distributions are fat-tailed, which implies that a high tsunami risk is not negligible. In fact,

more than 1-meter maximum tsunami heights are predicted on the shoreline of Cilacap as the 50th percentiles show. Furthermore, the 90th percentiles are more severe and result in more than 2 meters high for the tsunami heights onshore.

There are some aspects that need to be considered in future work to further refine the probabilistic outputs. First, the empirical equations of the source parameters can be made more realistic. Indeed, there may be a correlation between the amount of slip and the depth of the rupture place, which are currently independent of each other. Second, the distributions of the source parameters for future predictions should reflect the specific details of the Indonesian subduction zone. The present versions of these distributions are simply independent of each other, and the shape parameters would be determined to incorporate prior knowledge. Third, the number of levels can be modified. The increment from level 2 to level 3 is relatively small compared to that from level 1 to level 2. Taking into account the massive computational cost for the level 3 simulation, one could use a more balanced combination of the number

Fig. 12 The 50th (a) and 90th
(b) percentiles of the tsunami
elevations at all gauges
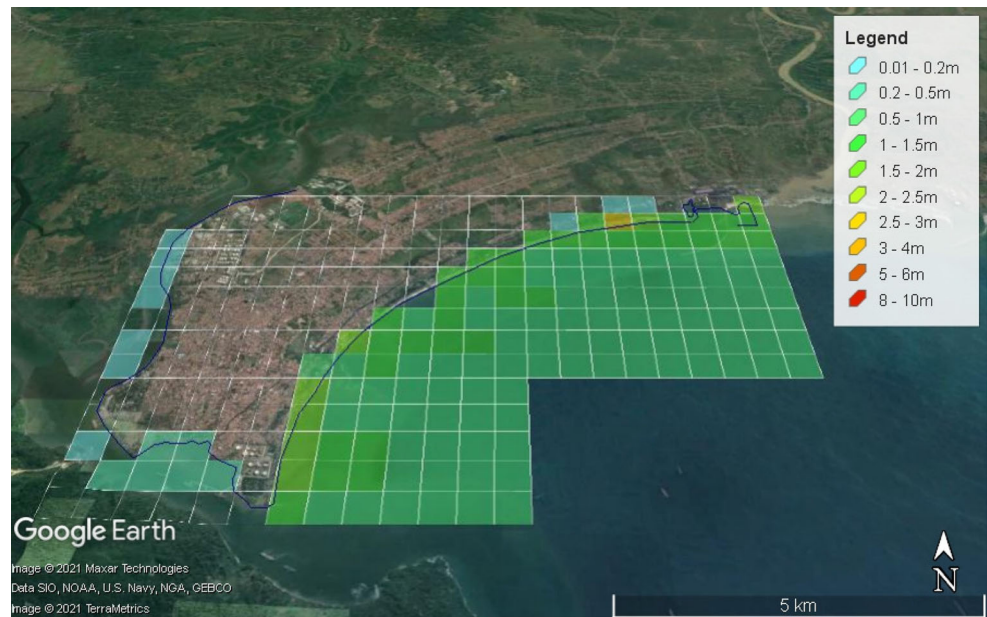
(a)
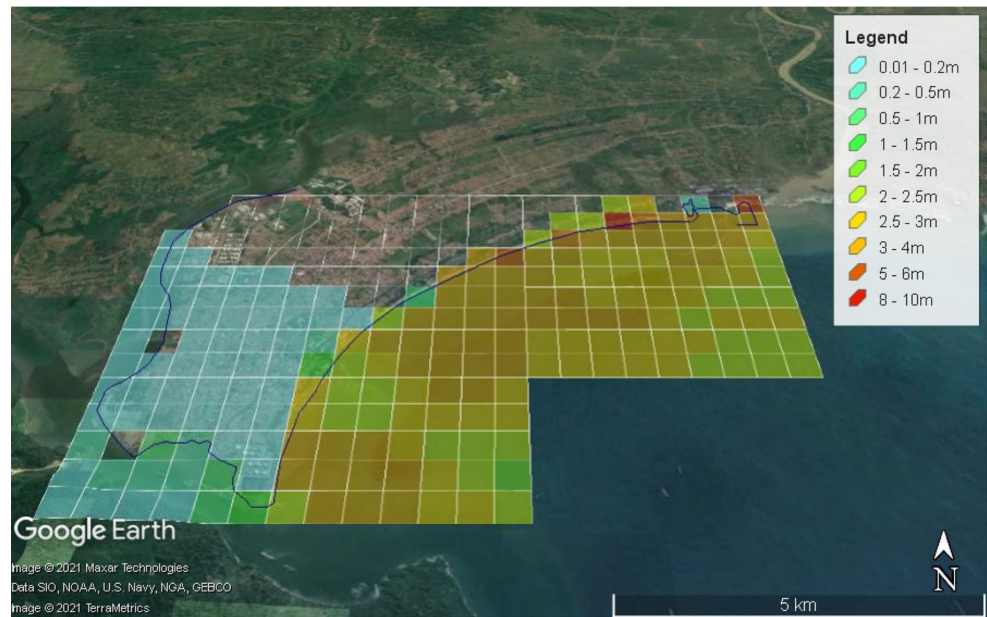
(b)

of the simulations and magnitude of the increments. Finally,
producing high resolution predictions of tsunami elevations
in the entire island of Java would be more beneficial for
stakeholders and hazard practitioners. A potential obstacle
of this attempt would be computational limitations in terms
of time and memory management.

Declarations

Conflict of Interests The authors have no financial or proprietary
interests in any material discussed in this article.

References

1. Heidarzadeh, M., Ishibe, T., Sandanbata, O., Muhari, A., Wijanarto, A.B.: Numerical modeling of the subaerial landslide source of the 22 december 2018 Anak Krakatoa Volcanic Tsunami, Indonesia. Ocean Eng. **106733**, 195 (2020)

2. Heidarzadeh, M., Muhari, A., Wijanarto, A.B.: Insights on the source of the 28 september 2018 Sulawesi Tsunami, Indonesia based on spectral analyses and numerical simulations. Pure Appl. Geophys. **176**(1), 25–43 (2019)

3. National Geophysical Data Center / World Data Service: NCEI/WDS Global Historical Tsunami Database. NOAA National Centers for Environmental Information. Accessed in 2020 at https://www.ngdc.noaa.gov/hazard/tsu_db.shtml. https://doi.org/10.7289/V5PN93H7

4. Tsuji, Y., Satake, K., Ishibe, T., Kusumoto, S., Harada, T., Nishiyama, A., Heidarzadeh, M.: Field surveys of tsunami heights from the 2011 off the pacific coast of Tohoku Japan earthquake (2011)

5. Salmanidou, D.M., Ehara, A., Himaz, R., Heidarzadeh, M., Guillas, S.: Impact of future tsunamis from the java trench on household welfare: merging geophysics and economics through catastrophe modelling. International Journal of Disaster Risk Reduction 102291 (2021)

6. Lorito, S., Selva, J., Basili, R., Romano, F., Tiberti, M., Piatanesi, A.: Probabilistic hazard for seismically induced tsunamis: Accuracy and feasibility of inundation maps. Geophys. J. Int. **200**(1), 574–588 (2015)

7. Sepulveda, I., Liu, P.L.-F., Grigoriu, M., Pritchard, M.: Tsunami hazard assessments with consideration of uncertain earthquake slip distribution and location. J. Geophys. Res.: Solid Earth **122**(9), 7252–7271 (2017)

8. Fukutani, Y., Moriguchi, S., Terada, K., Otake, Y.: Time-dependent probabilistic tsunami inundation assessment using mode decomposition to assess uncertainty for an earthquake scenario. J. Geophys. Res.: Oceans **126**(7), 2021–017250 (2021)

9. Behrens, J., Løvholt, F., Jalayer, F., Lorito, S., Salgado-Gálvez, M.A., Sørensen, M., Abadie, S., Aguirre-Ayerbe, I., Aniel-Quiroga, I., Babeyko, A., et al: Probabilistic tsunami hazard and risk analysis: A review of research gaps. Front. Earth Sci. **9**, 114 (2021)

10. Grezio, A., Babeyko, A., Baptista, M.A., Behrens, J., Costa, A., Davies, G., Geist, E.L., Glimsdal, S., González, F.I., Griffin, J., et al: Probabilistic tsunami hazard analysis: Multiple sources and global applications. Rev. Geophys. **55**(4), 1158–1198 (2017)

11. Gopinathan, D., Heidarzadeh, M., Guillas, S.: Probabilistic quantification of tsunami current hazard using statistical emulation. Proc. R. Soc. A **477**(2250), 20210180 (2021)

12. Giles, D., Gopinathan, D., Guillas, S., Dias, F.: Faster than real time tsunami warning with associated hazard uncertainties. Frontiers in Earth Science, 8 (2021)

13. Giraldi, L., Le Maître, O.P., Mandli, K.T., Dawson, C.N., Hoteit, I., Knio, O.M.: Bayesian inference of earthquake parameters from buoy data using a polynomial chaos-based surrogate. Comput. Geosci. **21**(4), 683–699 (2017)

14. Rodríguez, J.F., Macías, J., Castro, M.J., de la Asunción, M., Sánchez-Linares, C.: Use of neural networks for tsunami maximum height and arrival time predictions. GeoHazards **3**(2), 323–344 (2022)

15. Núñez, J., Catalán, P.A., Valle, C., Zamora, N., Valderrama, A.: Discriminating the occurrence of inundation in tsunami early warning with one-dimensional convolutional neural networks. Sci. Rep. **12**(1), 1–20 (2022)

16. Guillas, S., Sarri, A., Day, S.J., Liu, X., Dias, F., et al: Functional emulation of high resolution tsunami modelling over cascadia. Ann. Appl. Stat **12**(4), 2023–2053 (2018)

17. Gopinathan, D., Heidarzadeh, M., Guillas, S.: Probabilistic quantification of tsunami currents in karachi port, makran subduction zone using statistical emulation (2020)

18. Salmanidou, D.M., Heidarzadeh, M., Guillas, S.: Probabilistic landslide-generated tsunamis in the indus canyon, nw indian ocean, using statistical emulation. Pure Appl. Geophys. **176**(7), 3099–3114 (2019)

19. Salmanidou, D.M., Beck, J., Guillas, S.: Probabilistic, high-resolution tsunami predictions in north cascadia by exploiting sequential design for efficient emulation. Natural Hazards and Earth System Sciences Discussions, 1–24 (2021)

20. Snelling, B., Neethling, S., Horsburgh, K., Collins, G., Piggott, M.: Uncertainty quantification of landslide generated waves using gaussian process emulation and variance-based sensitivity analysis. Water **12**(2), 416 (2020)

21. Beck, J., Guillas, S.: Sequential design with mutual information for computer experiments (mice): Emulation of a tsunami model. SIAM/ASA J. Uncert. Quantif. **4**(1), 739–766 (2016)

22. de Baar, J.H., Roberts, S.G.: Multifidelity sparse-grid-based uncertainty quantification for the hokkaido nansei-oki tsunami. Pure Appl. Geophys. **174**(8), 3107–3121 (2017)

23. Ehara, A., Guillas, S.: An adaptive strategy for sequential designs of multilevel computer experiments. arXiv:2104.02037, International Journal for Uncertainty Quantification, in press (2021)

24. Baba, T., Takahashi, N., Kaneda, Y., Ando, K., Matsuoka, D., Kato, T.: Parallel implementation of dispersive tsunami wave modeling with a nesting algorithm for the 2011 Tohoku Tsunami. Pure Appl. Geophys. **172**(12), 3455–3472 (2015). https://doi.org/10.1007/s00024-015-1049-2

25. Baba, T., Ando, K., Matsuoka, D., Hyodo, M., Hori, T., Takahashi, N., Obayashi, R., Imato, Y., Kitamura, D., Uehara, H., Kato, T., Saka, R.: Large-scale, high-speed tsunami prediction for the Great Nankai Trough Earthquake on the K computer. Int. J. High Perform. Comput. Applic. **30**(1), 71–84 (2016). https://doi.org/10.1177/1094342015584090

26. Satake, K., Heidarzadeh, M.: A review of source models of the 2015 illapel, Chile earthquake and insights from tsunami data. Pure Appl. Geophys. **174**(1), 1–9 (2017)

27. Heidarzadeh, M., Murotani, S., Satake, K., Takagawa, T., Saito, T.: Fault size and depth extent of the Ecuador earthquake (mw 7.8) of 16 april 2016 from teleseismic and tsunami data. Geophys. Res. Lett. **44**(5), 2211–2219 (2017)

28. Baba, T., Gon, Y., Imai, K., Yamashita, K., Matsuno, T., Hayashi, M., Ichihara, H.: Modeling of a dispersive tsunami caused by a submarine landslide based on detailed bathymetry of the

continental slope in the nankai trough, southwest japan. Tectonophysics **768**, 228182 (2019). https://doi.org/10.1016/j.tecto.2019.228182

29. Salmanidou, D.M., Heidarzadeh, M., Guillas, S.: Probabilistic landslide-generated tsunamis in the Indus Canyon, NW Indian Ocean, using statistical emulation. Pure Appl. Geophys. **176**(7), 3099–3114 (2019). https://doi.org/10.1007/s00024-019-02187-3

30. Okada, Y.: Surface deformation due to shear and tensile faults in a half-space. Bull. Seismol. Soc. Am. **75**(4), 1135–1154 (1985)

31. Hayes, G.P., Moore, G.L., Portner, D.E., Hearne, M., Flamme, H., Furtney, M., Smoczyk, G.M.: Slab2, a comprehensive subduction zone geometry model. Science **362**(6410), 58–61 (2018)

32. Fujii, Y., Satake, K.: Source of the july 2006 west java tsunami estimated from tide gauge records. Geophys. Res. Lett. **33**(24). https://doi.org/10.1029/2006GL028049 (2006)

33. Allen, T.I., Hayes, G.P.: Alternative rupture-scaling relationships for subduction interface and other offshore environments. Bull. Seismol. Soc. Am. **107**(3), 1240–1253 (2017)

34. Wells, D.L., Coppersmith, K.J.: New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. Bull. Seismol. Soc. Amer. **84**(4), 974–1002 (1994)