

Title: Rapid prediction of alongshore run-up distribution from near-field tsunamis

Authors: Jun-Whan Lee (jwlee89@vt.edu), Jennifer L. Irish, Robert Weiss

Jun-Whan Lee

Department of Civil and Environmental Engineering, Virginia Tech, 750 Drillfield Dr, Blacksburg, VA, 24061, USA, ORCID: 0000-0002-6722-6843 E-mail: jwlee89@vt.edu

Jennifer L. Irish

Department of Civil and Environmental Engineering, Virginia Tech, 750 Drillfield Dr, Blacksburg, VA, 24061, USA, ORCID: 0000-0002-2429-5953

Robert Weiss

Department of Geosciences, Virginia Tech, 926 W Campus Dr, Blacksburg, VA, 24061, USA, ORCID: 0000-0002-7168-5401

This manuscript is a non-peer reviewed preprint submitted to EarthArXiv.

The published version of this preprint is available at Natural Hazards:

<https://doi.org/10.1007/s11069-020-04209-z>

Rapid prediction of alongshore run-up distribution from near-field tsunamis

Jun-Whan Lee · Jennifer L. Irish ·
Robert Weiss

Received: date / Accepted: date

Abstract Rapid prediction of the spatial distribution of the run-up from near-field tsunamis is critically important for tsunami hazard characterization. Even though significant advances have been made over the last decade, physics-based numerical models are still computationally intensive. Here, we present a response surface methodology (RSM)-based model called the tsunami run-up response function (TRRF). Derived from a discrete set of tsunami simulations, TRRF can produce a rapid prediction of a near-field tsunami run-up distribution that takes into account the influence of variable local topographic and bathymetric characteristics in a given region. This new method reduces the number of simulations required to build an RSM model by separately modeling the leading order contribution and the residual part of the tsunami run-up distribution. Using the northern region of Puerto Rico as a case study, we investigated the performance (accuracy, computational time) of the TRRF. The results reveal that the TRRF achieves reliable prediction while reducing the prediction time by six orders of magnitude (computational time: < 1 second per earthquake).

Keywords Tsunami · Run-up · Response surface methodology · Puerto Rico

Jun-Whan Lee
Department of Civil and Environmental Engineering, Virginia Tech, 750 Drillfield Dr,
Blacksburg, VA, 24061, USA, ORCID: 0000-0002-6722-6843
E-mail: jwlee89@vt.edu

Jennifer L. Irish
Department of Civil and Environmental Engineering, Virginia Tech, 750 Drillfield Dr,
Blacksburg, VA, 24061, USA, ORCID: 0000-0002-2429-5953

Robert Weiss
Department of Geosciences, Virginia Tech, 926 W Campus Dr, Blacksburg, VA, 24061, USA,
ORCID: 0000-0002-7168-5401

1 Introduction

Tsunamis are some of the most destructive and costly natural hazards for coastal areas around the world. The 2004 Indian Ocean tsunami and the 2011 Tohoku tsunami are prime examples of how tsunamis can cause extensive damage to coastal communities, especially in near-field areas (Titov et al., 2005; Wei et al., 2013). To mitigate damage and build resilient coastal communities, it is critically important to develop rapid prediction capacities for a near-field tsunami run-up distribution along the coastlines. Physics-based numerical simulation is currently the most accurate method for predicting a tsunami run-up distribution. Though significant advances have been made over the last decade (LeVeque et al., 2011; Lin et al., 2015; Popinet, 2015; Shi et al., 2012), these physics-based numerical models still remain time consuming. For example, robust probabilistic tsunami hazard assessment (PTHA) requires tsunami run-up estimates for a large number of scenarios to allow for accurate quantification of the hazard and related uncertainty (Mori et al., 2018). However, due to the computational burden associated with physics-based numerical simulation, a logic-tree approach is typically employed: it limits the number of scenarios based on historical earthquake characteristics (e.g., magnitude, recurrence interval) used to evaluate uncertainty in tsunami hazard (Annaka et al., 2007; Park and Cox, 2016; Park et al., 2018). The issue with the logic-tree approach is that it relies on expert judgment, which is difficult to justify. Also, the computational burden associated with physics-based numerical simulation—especially for near-field tsunami forecasting—is a major obstacle. For this reason, pre-computed simulation databases are widely used. These databases can provide fast prediction by selecting the best-matched simulation or by interpolating between simulations immediately after the source mechanism is known (Kamigaichi, 2011; Mulia et al., 2018; Setiyono et al., 2017). A problem with the database approach is that it can have substantial errors in real-world scenarios that do not exist in the selected databases.

The response surface methodology (RSM) is an effective statistical-based approach for establishing a relationship between a set of input variables and the output of a system (Box and Wilson, 1951; Myers et al., 2016). Once the RSM model is built, output can be rapidly estimated across the continuum of input spaces. However, because high-dimensional input requires a large number of simulations—which is prohibitively expensive—the RSM has not been used to predict a tsunami run-up distribution. For example, a tsunamigenic-earthquake (the input in an RSM model) is usually represented by nine fault parameters (Fig. 1). A full factorial design is one of the most widely employed designs of experiments (DoE) used to measure the response of every possible combination of independent variables. If we design the synthetic tsunami scenarios using a three-level full factorial approach with nine fault parameters, $19,683 (= 3^9)$ simulations are required. Moreover, if the input/output relationship shows large nonlinearity, a higher level of DoE may be needed, which would necessitate exponentially more simulations.

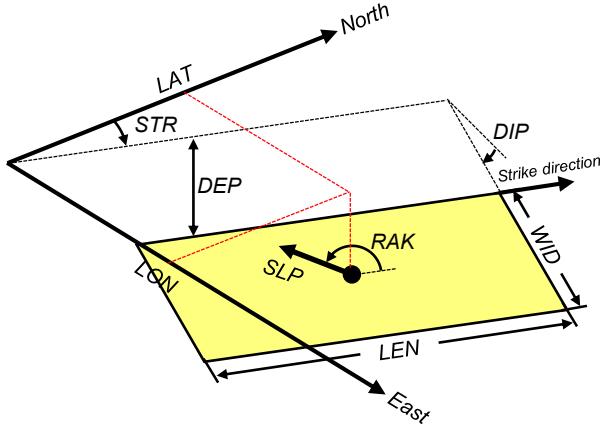


Fig. 1 Schematic sketch of earthquake fault parameters: epicenter latitude (*LAT*), epicenter longitude (*LON*), fault length (*LEN*), fault width (*WID*), top-edge fault depth (*DEP*), strike angle (*STR*), dip angle (*DIP*), rake angle (*RAK*), and slip (*SLP*).

Here, we present a new methodology to rapidly predict the near-field tsunami run-up distribution: the tsunami run-up response function (TRRF). It is based on RSM but requires only 729 ($= 3^6$) simulations through reducing input dimensionality. Input dimensionality is reduced through a decomposition of the leading order tsunami run-up contribution and the residual part of the run-up distribution. We demonstrated the TRRF approach in northern Puerto Rico, where a significant tsunami generated by an earthquake along the Puerto Rico Trench could devastate coastal communities on the northern shore (Grilli et al., 2010; López-Venegas et al., 2015; Reid and Taber, 1919) (Fig. 2).

2 Tsunami run-up response function (TRRF)

The main concept of TRRF is to decompose the tsunami run-up distribution $R(x)$ into source run-up $S(x)$ and topographic run-up $T(x)$ (Fig. 3):

$$R(x) = S(x) + T(x) \quad (1)$$

where the x -axis is parallel to the coastline.

The source run-up $S(x)$ is a leading order contribution that can be represented by Okal and Synolakis (2004)'s empirical formula (hereafter OS formula):

$$S(x) = \frac{b}{\left[\frac{x-c}{a}\right]^2 + 1} \quad (2)$$

where the coefficient a is related to the width of the source run-up, b is the maximum source run-up, and c is the distance from the x -axis origin to the location of the maximum source run-up.

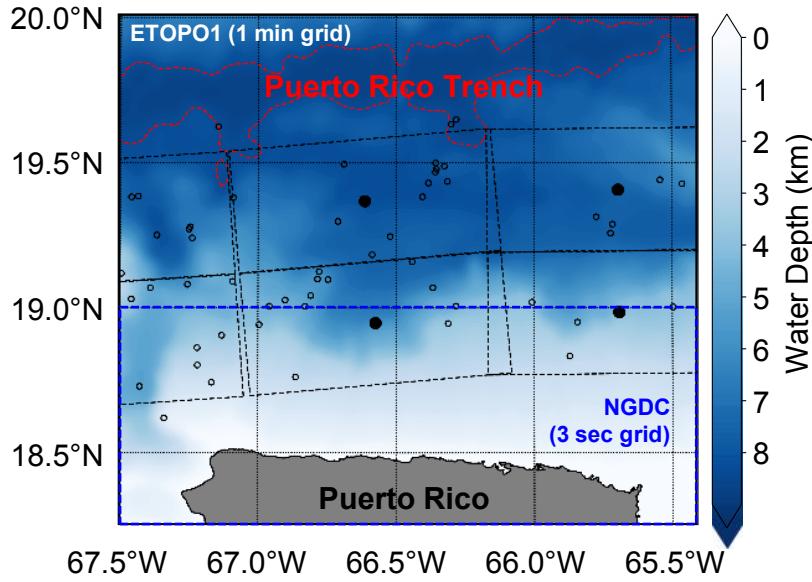


Fig. 2 Map of northern Puerto Rico. Open black circles represent the epicenters of historical earthquakes ($M_w \geq 4.5$, (USGS, 2017)). The filled black circles and dashed black lines represent the epicenters of NOAA's pre-defined unit sources and fault orientation, respectively (Gica, 2008). The dashed red line represents the contour line where the water depth is 8 km. The blue dashed square represents the region where the National Geophysical Data Center (NGDC)'s 3-second topographic grid (NGDC, 2005) is used for numerical simulation.

The topographic run-up $T(x)$ is the residual run-up remaining after subtracting $S(x)$ from $R(x)$. It represents the local (de)amplification of the incoming tsunami wave and the resulting run-up arising from topographic variation. The $T(x)$ can be normalized as follows, hereafter called normalized topographic run-up $NT(x)$.

$$NT(x) = \frac{T(x)}{S(x)} \quad (3)$$

The TRRF can predict a tsunami run-up distribution $R^p(x)$ by putting the source run-up $S^p(x)$ and the normalized topography run-up $NT^p(x)$ (where superscript p represents prediction) to the following equation:

$$R^p(x) = S^p(x) [1 + NT^p(x)] \quad (4)$$

where $NT^p(x)$ is the 50th percentile (or median) of $NT(x)$ among all the simulations used to build the TRRF. The source run-up $S^p(x)$ can be estimated by inputting the OS formula coefficients a , b , and c into Eq. 2. The RSM approach is applied to estimate the OS formula coefficients a and b from six fault parameters (hereafter RSM fault parameters):

$$a = f_a(LAT, LEN, WID, DIP, SLP, DEP) \quad (5)$$

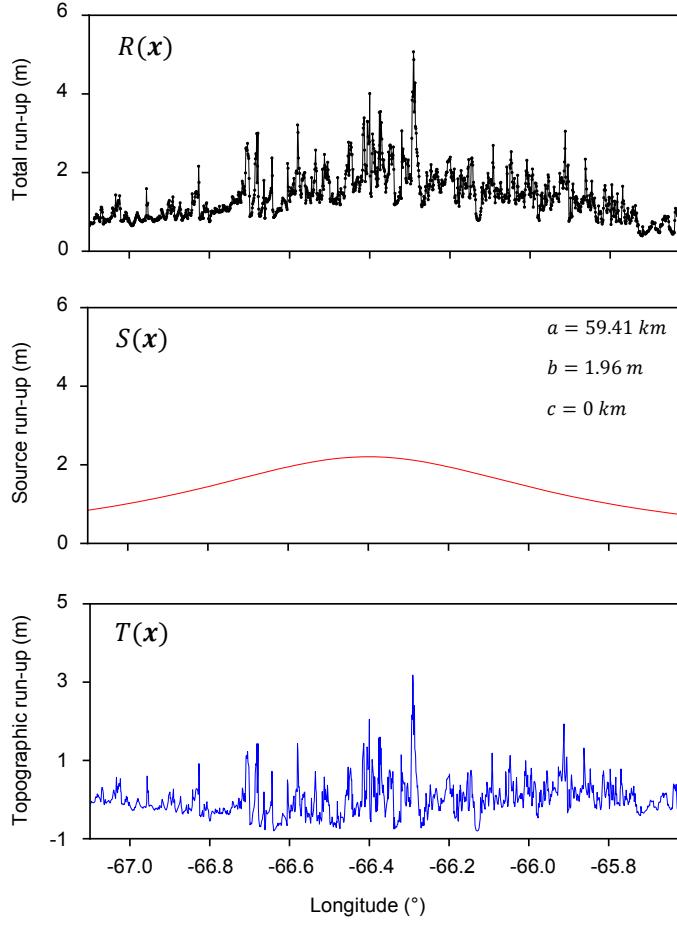


Fig. 3 Example of (a) tsunami run-up distribution, (b) source run-up, and (c) topographic run-up. The fault parameter condition is as follows: $LON = 66.4^{\circ}W$, $LAT = 19.3^{\circ}N$, $STR = 90^{\circ}$, $DIP = 20^{\circ}$, $RAK = 90^{\circ}$, $LEN = 90\text{km}$, $WID = 40\text{km}$, $SLP = 2\text{m}$, $DEP = 30\text{km}$.

$$b = f_b(LAT, LEN, WID, DIP, SLP, DEP) \quad (6)$$

where f_a and f_b are the best-fitting curves to these coefficients; hereafter, these curves are called RSM functions. Since the RSM function inputs consist of six fault parameters, $729 (= 3^6)$ simulations are required to derive RSM functions following a three-level full factorial design (See Appendix A).

Epicenter longitude LON can be considered separately because northern Puerto Rico's coastline runs parallel to the latitude lines. In this condition, the coefficients a and b are independent of epicenter longitude, and the coefficient c can be represented by a function of epicenter longitude:

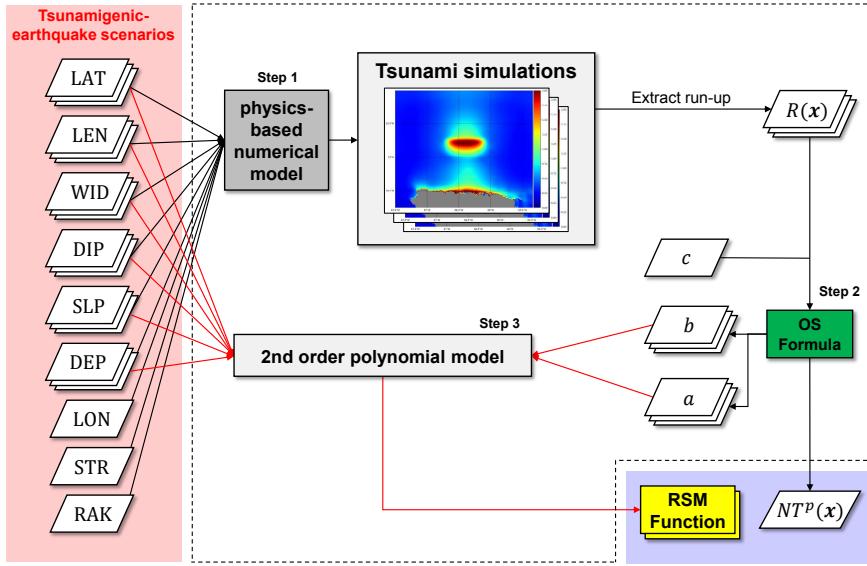


Fig. 4 Computational flow of TRRF development.

$$c = LON - LON_0 \quad (7)$$

where LON_0 is the longitude of the origin.

The strike angle STR and rake angle RAK are also not included in the RSM function inputs because the OS formula (Okal and Synolakis, 2004) is only applicable to an earthquake fault oriented in shore-parallel strike direction ($= 90^\circ$ for northern Puerto Rico) with 90° rake angle. Since this is not the only case that occurs in nature, we developed a method that can represent a fault where strike angle and/or rake angle are not 90° as a series of hypothetical faults where both the strike angle and the rake angle are 90° , hereafter called the angle projection (AP) method.

Section 2.1 will describe the procedures of building a TRRF. Section 2.2 will explain how to apply the AP method to predict a tsunami run-up distribution for the case where strike angle and/or rake angle are not 90° . Lastly, Section 2.3 will describe the procedures of predicting a tsunami run-up distribution once the TRRF is built.

2.1 TRRF development

Fig. 4 shows the procedure of TRRF development. The first step is to simulate 729 tsunamigenic-earthquake scenarios using a physics-based numerical model. The second step is to extract the run-up and apply the OS formula (Eq. 2) to obtain the normalized topographic run-up $NT^p(x)$. The last step is to fit the earthquake fault parameters and the OS formula coefficients (a and b)

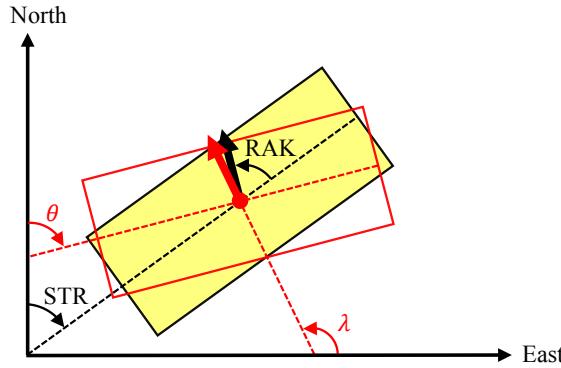


Fig. 5 Schematic sketch of step 1 of AP method. The yellow-filled rectangle is the original fault where STR is strike angle and RAK is rake angle. The red rectangle represents the adjusted fault where θ is the adjusted strike angle and λ is the adjusted rake angle. The arrows represent the slip direction.

to the 2nd order polynomial model to obtain the RSM functions. Once the $NT^p(x)$ and the RSM functions are derived, this procedure does not have to be repeated to predict the tsunami run-up distribution.

2.2 Angle projection (AP) method

The AP method comprises three steps: adjustment of strike angle and rake angle, fault rotation, and decomposition of slip. Note that the units of LON and LAT in this section are in kilometers.

2.2.1 Adjustment of strike angle and rake angle

The direction of near-field tsunami propagation is related to the interaction between the strike angle STR and the rake angle RAK . To consider the interaction between STR and RAK , the first step involves adjusting the STR and the RAK as follows (Fig. 5):

$$\theta = \alpha(STR - 90^\circ) + 90^\circ + \beta(RAK - 90^\circ) \quad (8)$$

$$\lambda = \gamma(RAK - 90^\circ) + 90^\circ \quad (9)$$

where θ is the adjusted strike angle and λ is the adjusted rake angle. The α , β , and γ are the site-specific coefficients (see Section 4).

2.2.2 Fault rotation

If the adjusted strike direction is parallel to the coastline ($\theta = 90^\circ$) and the adjusted rake angle λ is 90° , the maximum source run-up will be located

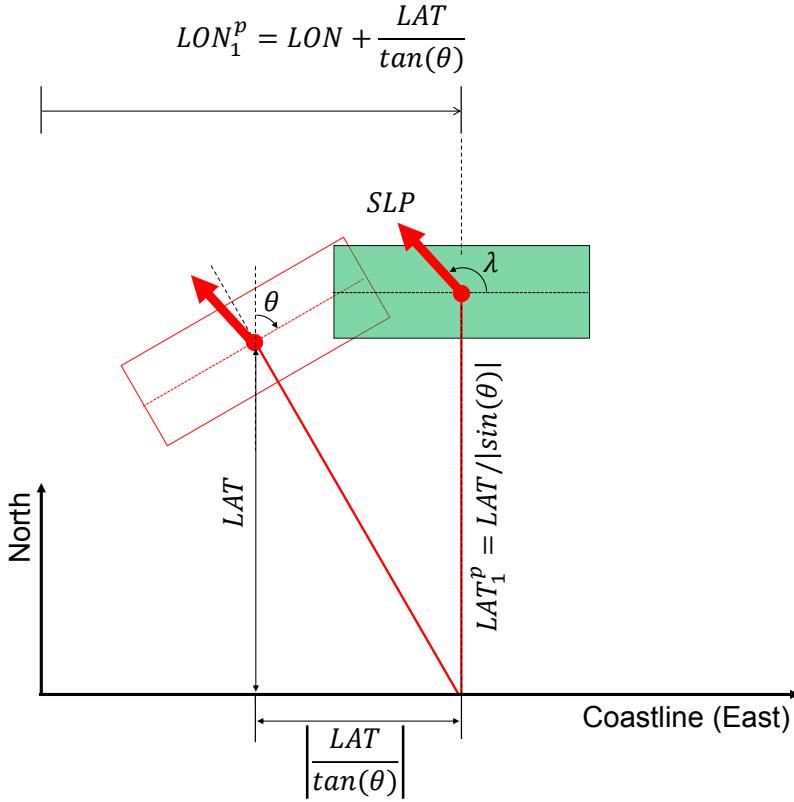


Fig. 6 Schematic sketch of step 2 of AP method. The red rectangle is the adjusted fault where the epicenter is (LAT, LON) . The green rectangle represents the rotated fault where the epicenter is (LAT_1^p, LON_1^p) . The two red lines are of the same length. One line is perpendicular to the strike direction spanning from the epicenter of the adjusted fault to the point where it meets the coastline. The other line is the vertical distance from the epicenter of the rotated fault to the coastline.

at a shore perpendicular from the longitude epicenter LON . On the other hand, if the adjusted strike direction is not parallel to the coastline ($\theta \neq 90^\circ$), the location of the maximum source run-up will be shifted to a direction perpendicular to the adjusted strike direction. To consider the location of the maximum source run-up depending on the adjusted strike angle, the second step involves rotating the adjusted fault ($\theta \neq 90^\circ$) until θ becomes 90° (Fig. 6). The epicenter of the rotated fault (LAT_1^p, LON_1^p) can be calculated as follows:

$$LAT_1^p = \frac{LAT}{|\sin(\theta)|} \quad (10)$$

$$LON_1^p = LON + \frac{LAT}{\tan(\theta)} \quad (11)$$

2.2.3 Decomposition of slip

If the adjusted rake angle λ is not 90° , the run-up will be spread in the slip direction. To consider the spread of run-ups depending on the adjusted rake angle, the third step involves representing the rotated fault ($\lambda \neq 90^\circ$) as a series of hypothetical faults having slips perpendicular to the coastline ($\lambda = 90^\circ$) (Fig. 7). Since the tsunami energy is proportional to SLP , we assume that the source run-up will be spread proportionally to a component of SLP parallel to the coastline. Based on this assumption, while the LEN , WID , DEP , and DIP are identical to the original fault, the epicenter (LAT , LON) and SLP of the i th hypothetical fault ($i = 1, 2 \dots n$) are defined as follows:

$$LAT_i^p = LAT_1^p + \frac{LAT_n^p - LAT_1^p}{n-1}(i-1) \quad (12)$$

$$LON_i^p = LON_1^p + \frac{LON_n^p - LON_1^p}{n-1}(i-1) \quad (13)$$

$$SLP_i^p = SLP_1^p + \frac{SLP_n^p - SLP_1^p}{n-1}(i-1) \quad (14)$$

where n is the total number of hypothetical faults (see Section 4). The SLP_1^p and SLP_n^p are the slips of the first and last hypothetical faults, respectively, defined as follows:

$$SLP_1^p = SLP|\sin(\lambda)| \quad (15)$$

$$SLP_n^p = SLP|\cos(\lambda)| \quad (16)$$

The LAT_n^p and LON_n^p are the epicenter latitude and longitude of the last hypothetical fault and can be calculated based on a geometric setup (see two red lines in Fig. 7):

$$LON_n^p = LON_1^p - \frac{LAT_1^p}{\tan(\lambda)} \quad (17)$$

$$LAT_n^p = \frac{LAT_1^p}{|\sin(\lambda)|} \quad (18)$$

Following this procedure, a fault where strike angle and/or rake angle are not 90° can be converted into a series of hypothetical faults where both strike angle and rake angle are 90° .

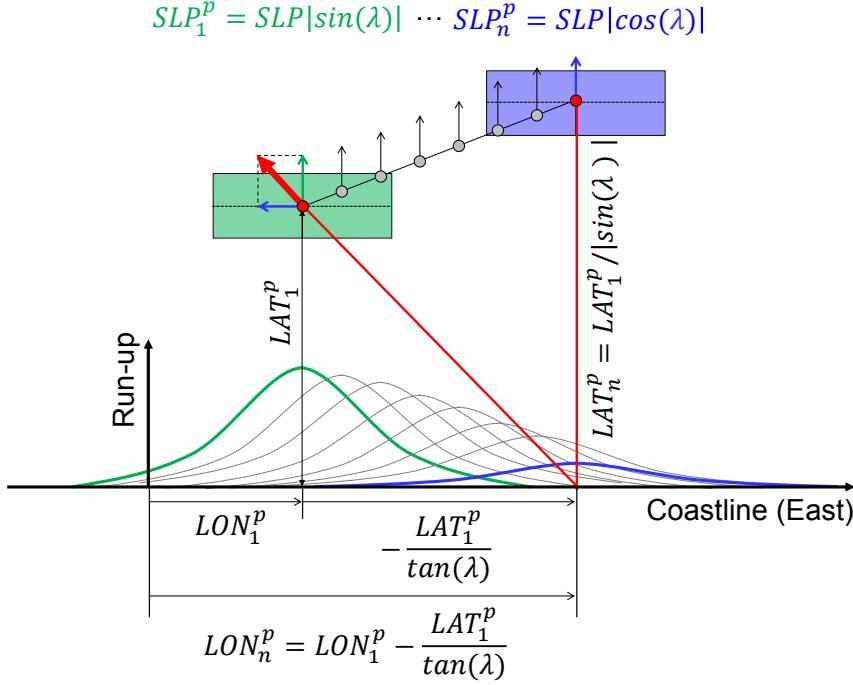


Fig. 7 Schematic sketch of step 3 of AP method. The green rectangle represents the first hypothetical fault where the epicenter is (LAT_1^p, LON_1^p) and the slip is SLP_1^p . The blue rectangle represents the last hypothetical fault among a series of hypothetical faults where the epicenter is (LAT_n^p, LON_n^p) and slip is SLP_n^p . Gray circles and arrows represent the epicenters and the slips of the hypothetical faults, respectively; these are linearly distributed between the first hypothetical fault and the last hypothetical fault. Two red lines are of the same length. One line is parallel to the slip direction spanning from the epicenter of the rotated fault to the point where it meets the coastline. The other line is the vertical distance from the epicenter of the last hypothetical fault to the coastline.

2.3 TRRF application for prediction

Fig. 8 shows the procedure for how the TRRF predicts a tsunami run-up distribution once the TRRF is built. The first step is to convert the earthquake fault into a series of hypothetical faults using the AP method. The second step is to estimate the OS formula coefficients a_i^p , b_i^p ($i = 1, 2 \dots n$) of hypothetical faults using the RSM functions. The third step is to estimate the OS formula coefficient c_i^p ($i = 1, 2 \dots n$) by inputting the epicenter longitudes into Eq. 7.

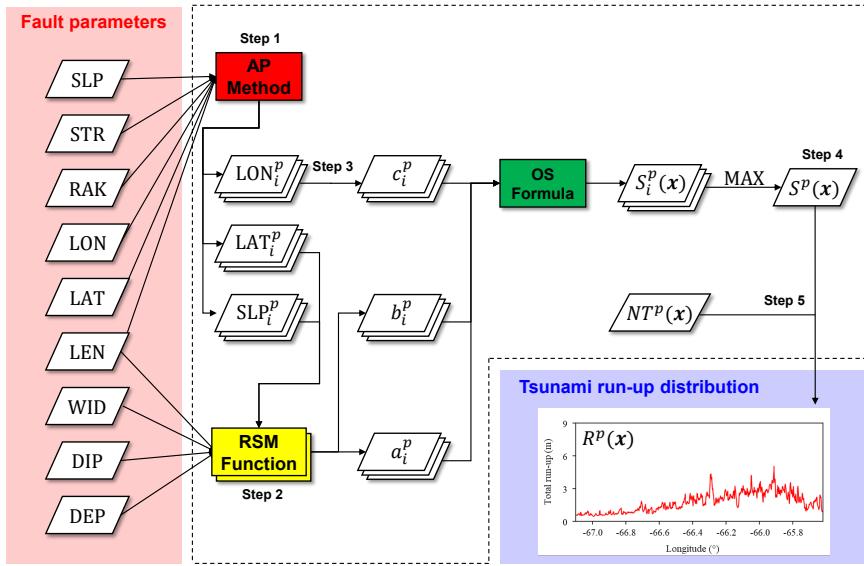


Fig. 8 Computational flow of TRRF application for prediction.

The fourth step is to estimate the final source run-up $S^p(x)$ by inputting the OS formula coefficients into Eq. 2 and taking the maximum values of the estimated source run-ups for all hypothetical faults. Finally, the tsunami run-up distribution $R^p(x)$ can be estimated by inputting the source run-up $S^p(x)$ and the normalized topographic run-up $NT^p(x)$ to Eq. 4.

3 TRRF development for northern Puerto Rico

3.1 Numerical simulation

The 729 tsunamigenic-earthquake scenarios were simulated based on the numerical model Basilisk, which solves the Green-Naghdi equations and employs both Adaptive Mesh Refinement (AMR) and parallelization to facilitate efficient computation. The Basilisk model has not only been validated with several benchmark problems but also been applied to simulate the 2011 Tohoku tsunami (Popinet, 2015). The 729 scenarios were designed as shown in Table 1. The range of the epicenter latitude LAT was determined based on National Oceanic and Atmospheric Administration (NOAA)'s pre-defined unit sources and historical earthquake records in northern Puerto Rico (Fig. 2). The range of the fault length LEN , fault width WID , and slip SLP was set based on the assumption that the moment magnitude (M_w) should be larger than 7.0 for a tsunami to occur. We used the empirical regression of Hanks and Kanamori (1979) and fundamental equation of Aki (1966) to calculate the moment mag-

nitude:

$$M_w = \frac{2}{3} \{ \log(M_0) - 9.05 \} \quad (19)$$

$$M_0 = \mu(LEN \times WID \times SLP) \quad (20)$$

where M_0 is a seismic moment (Nm), μ is rigidity modulus of the Earth's crust (Nm^{-2}), and the units of LEN , WID , and SLP are in meters. We assumed that the rigidity modulus μ is $4.2 \times 10^{10} Nm^{-2}$ in northern Puerto Rico following Grilli et al. (2010). We limited the maximum moment magnitude to 8.0 considering the historical seismic events that led to tsunamis in Puerto Rico (Nealon and Dillon, 2001). We assumed that the LEN should be longer than the WID , and the range of the LEN and WID should follow the scaling laws introduced by Blaser et al. (2010). The range of the dip angle DIP and the depth of the top edge DEP were determined based on the characteristics of a subduction-interface earthquake that usually causes a tsunami. According to Thingbaijam et al. (2017), subduction-interface earthquakes occur between 10° and 30° dip angles and within a slip-centroid depth of 50 km. We assumed that the fault rupture occurred instantaneously, where the initial free surface displacement was calculated using the Okada equations (Okada, 1985). Nearshore bathymetry and onshore topography in the inundation zone were from the 3 arc-second National Geophysical Data Center data set (NGDC, 2005), while the 1 arc-minute ETOPO1 data set (Amante and Eakins, 2009) was used for the entire region (Fig. 2). Considering the grid size, the minimum and maximum AMR levels were set to 5 and 11, respectively. The bottom friction was parameterized using a quadratic drag law in which the bottom drag coefficient C_f was set to 10^{-4} . The numerical model was used to simulate two hours of tsunami propagation to ensure that complete inundation of the onshore areas was captured. The maximum envelope of the water level was interpolated bilinearly onto a regular grid (0.001° interval). We excluded four simulations, which failed to finish the simulations because of instability issue, to build the TRRF. We obtained tsunami run-up distribution $R(x)$ by extracting the maximum inundation height along the coastline ranging from $67.100^\circ W$ to $65.620^\circ W$. The tsunami simulations were conducted in a spherical coordinate system, but the TRRF was defined based on a Cartesian coordinate system. To align the coordinate systems, Vincenty's formulae (Vincenty, 1975) were used to convert the unit of the geometric point from degrees to kilometers. We set the origin at $(18.450^\circ N, 66.400^\circ W)$, and thus the positive $LON(LAT)$ refers to the longitude(latitude) of the epicenter that is farther east (north) than the origin.

3.2 RSM functions and $NT^p(x)$

The RSM functions and the normalized topography run-up $NT^p(x)$ were derived as follows. We calculated the OS formula coefficients a and b by fitting the tsunami run-up distribution $R(x)$ to the OS formula for each simulation (Eq. 2). Here, the OS formula coefficient c was fixed to zero because we set the

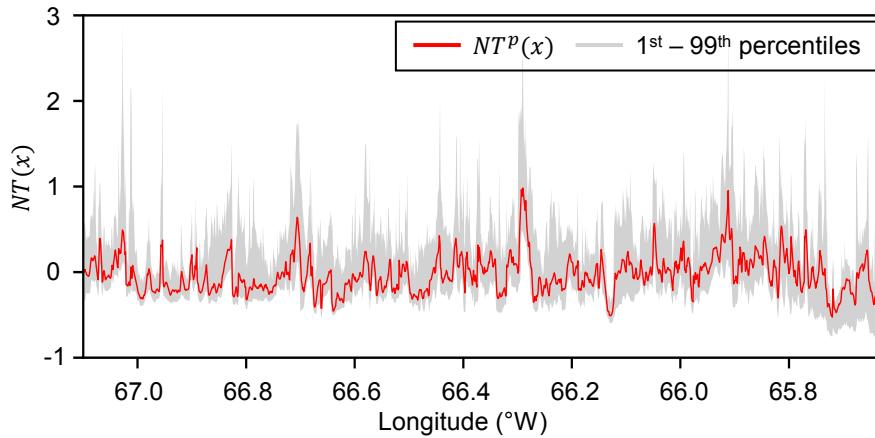


Fig. 9 Normalized topographic run-up $NT^p(x)$ of northern Puerto Rico. The gray represents the range between 1st and 99th percentiles.

longitude of the origin and the epicenter longitude of simulations identically (Eq. 7). We derived the RSM functions by fitting the RSM fault parameters to the OS formula coefficients a and b using second order polynomial models. The normalized topographic run-up $NT(x)$ was calculated for each simulation following Eqs. 1-3. We derived $NT^p(x)$ by selecting the 50th percentile of $NT(x)$ among all simulations (Fig. 9).

3.3 Fault parameter range for TRRF prediction

We set the fault parameter range for TRRF prediction as shown in Table 1. The range of six fault parameters (LAT , LEN , WID , DIP , SLP , DEP) was set to the same range as the tsunamigenic-earthquake scenarios used in the TRRF development. In order to avoid an extrapolation beyond the inference

Table 1 Range of fault parameters used for TRRF development and TRRF prediction

Parameter	Development			Prediction	
	Low	Central	High	Min	Max
$LAT(^{\circ}N)$	19.0	19.3	19.6	19.000	19.600
$DIP(^{\circ})$	10	20	30	10	30
$LEN(km)$	60	90	120	60.000	120.000
$WID(km)$	20	40	60	20.000	60.000
$SLP(m)$	2	3	4	2.00	4.00
$DEP(km)$	10	20	30	10.000	30.000
$LON(^{\circ}W)$		66.4		65.800	67.000
$STR(^{\circ})$		90		50	130
$RAK(^{\circ})$		90		50	130

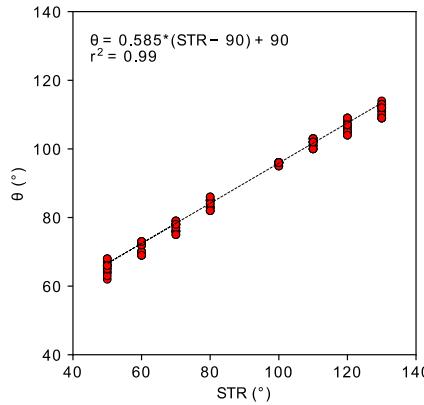


Fig. 10 Best θ values (that show the minimum NRMSE) associated with varying strike angles. Dashed line represents the best-fitting line.

space of the RSM functions, we only considered cases where all *LAT* of the hypothetical faults fell within the range for *LAT*. The range of *LONG* was set to the extent that the fault does not fall outside the region used in the numerical simulation. The strike angle is usually set in the direction tangential to the subduction zone (Gica et al., 2008), and thus we set the range of *STR* to be from 50° to 130° . Even though some tsunamis are generated by strike-slip earthquakes (Heidarzadeh et al., 2017), most tsunamis are caused by thrust earthquakes. Following this characteristic of *RAK*, we set the range of *RAK* to be from 50° to 130° .

4 TRRF calibration

To apply the AP method, (1) the site-specific coefficients (α , β , and γ) of Eqs. 8 and 9 and (2) the number of hypothetical faults (n) must be defined in advance.

To determine the coefficient α , we simulated 80 additional cases (hereafter called STR cases) that were not used in building the TRRF. These additional cases had a fixed longitude of $66.400^\circ W$, where 10 sets varying the RSM fault parameters were randomly selected. For each set, eight different strike angles between 50° and 130° were selected, at 10° intervals, except 90° . The rake angle was fixed to 90° so that θ could be independent of β and λ could be fixed to 90° (see Eqs. 8 and 9). The coefficient α was selected by minimizing TRRF error as represented by normalized root mean square error (*NRMSE*):

$$NRMSE = \frac{\sqrt{\frac{1}{N} \sum_{x=1}^N [R^p(x) - \hat{R}^p(x)]^2}}{\max[\hat{R}^p(x)] - \min[\hat{R}^p(x)]} \times 100 (\%) \quad (21)$$

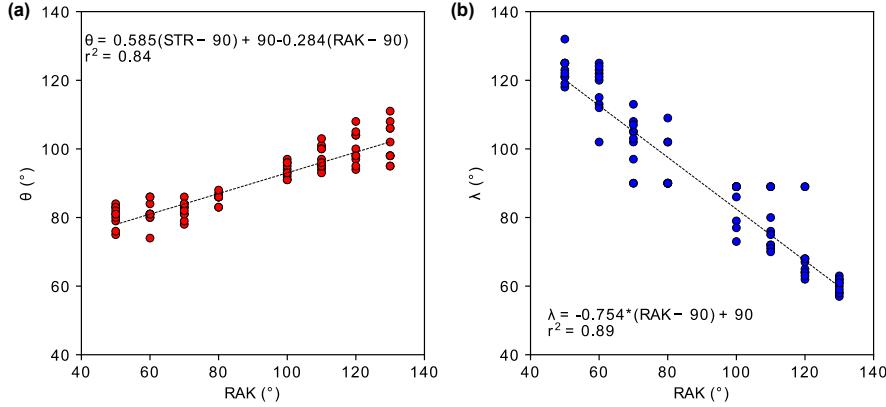


Fig. 11 Best (a) θ and (b) λ values (that show the minimum NRMSE) associated with varying rake angles. Dashed line represents the best-fitting line.

where $R^p(x)$ is the tsunami run-up distribution predicted by the TRRF, $\hat{R}^p(x)$ is the numerically simulated tsunami run-up distribution, and N is the total number of alongshore locations. For each case, we found the θ value that shows the minimum NRMSE in the range of 45° and 135° . We fixed the number of hypothetical faults (n) to 100, which was large enough to provide a convergent prediction. We set the coefficient α to 0.585 by fitting the STR and the θ in Eq. 8 (Fig. 10).

To determine the coefficients (β , and γ), we simulated 80 additional cases (hereafter called RAK cases) where all fault parameters but the rake angle were set in the same way as the STR cases. Unlike the STR cases, the rake angle was set to the same value as the strike angle. For each case, we found the θ value that shows the minimum NRMSE in the range of 45° and 135° . At the same time, we found the λ value that shows the minimum NRMSE in the range of 90° and 179° (if $RAK < 90^\circ$) or the λ value that shows the minimum NRMSE in the range of 1° and 90° (if $RAK \geq 90^\circ$). We set the coefficient β to -0.284 by fitting the RAK and the θ to Eq. 8 (Fig. 11(a)). We set the coefficient γ to -0.754 by fitting the RAK and the λ in Eq. 9 (Fig. 11(b)).

To determine the number of hypothetical faults (n), we revisited the RAK cases. For each case, we decreased the number of hypothetical faults (n) from 100 to 2 (Fig. 12). Then, we found the minimum value needed for convergence since the computational time increases as n increases. In this study, we set the n to 30, which shows less than 0.1% difference in NRMSE.

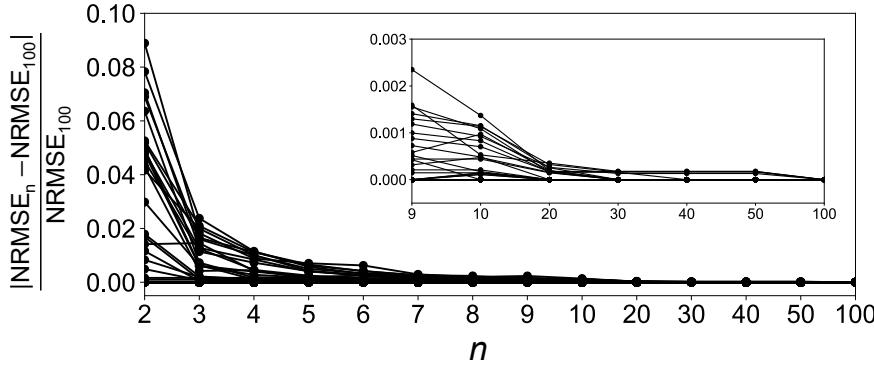


Fig. 12 Relative $NRMSE$ differences as the total number of hypothetical faults (n) increases. $NRMSE_n$ is the $NRMSE$ of the case where n hypothetical faults are considered. $NRMSE_{100}$ is the $NRMSE$ of the case where 100 hypothetical faults are considered.

5 TRRF performance

5.1 Accuracy

The accuracy of the TRRF was investigated by comparing TRRF predictions against the direct numerical simulations. Since the TRRF considers the earthquake fault parameters in three groups ((1) RSM fault parameters, (2) epicenter longitude, and (3) strike angle and rake angle), we systematically tested the accuracy of the TRRF as follows:

- Test 1: RSM Fault Parameters. We simulated 100 additional cases in which the RSM fault parameters were randomly selected, while the epicenter longitude was fixed to $66.400^{\circ}W$ and both the strike and rake angles were fixed to 90° .
- Test 2: Epicenter Longitude. We simulated 100 additional cases in which the fault parameters were selected based on the following conditions. While both the strike and rake angles were fixed to 90° , 10 sets of the RSM fault parameters were randomly selected. For each set, 10 longitudes were selected at a uniform interval in the range of $65.800^{\circ}W$ and $67.000^{\circ}W$.
- Test 3: Strike Angle and Rake angle. We investigated the RAK cases defined in Section 4.
- Test 4: Overall Accuracy. We simulated 100 additional cases in which all fault parameters were randomly selected.

Fig. 13 shows the comparison of the OS formula coefficients based on the 100 cases of Test 1. Note that these 100 cases were never used to derive the RSM functions. The x-axis is the OS formula coefficient obtained by fitting the numerical simulation result to the OS formula. The y-axis is the OS formula coefficient obtained by putting the six fault parameters (LAT , SLP , LEN ,

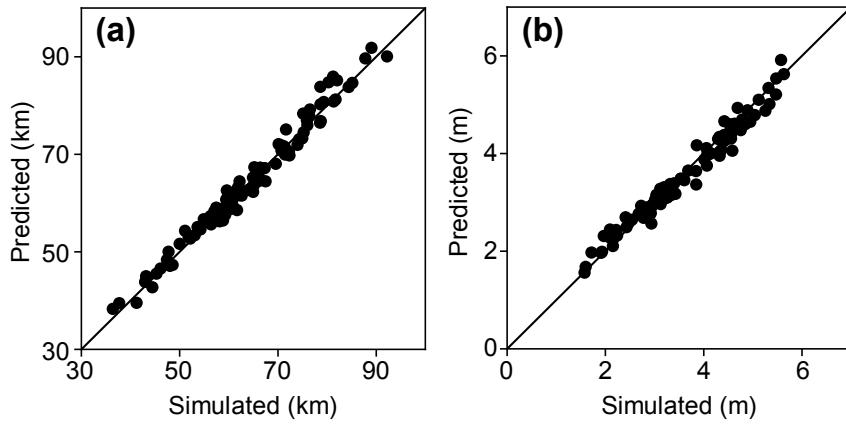


Fig. 13 Comparison of OS formula coefficients between Basilisk (simulated) and TRRF (predicted). (a) OS formula coefficient a . (b) OS formula coefficient b .

WID , DIP , DEP) to the RSM functions. The high-correlated results confirm that the RSM functions can predict the OS formula coefficients well.

Fig. 14 shows the selected alongshore tsunami run-up predictions for each test. Comprehensive summary statistics of test results are presented in Table 2. Fig. 14(a) and (b) show the best case (minimum $NRMSE$) and the worst case (maximum $NRMSE$) of Test 1, respectively. In both cases, the TRRF prediction followed the overall trend of the numerical simulation result well. However, there are a few localities where the TRRF did not predict the run-up well such as the run-ups near $66.2^\circ W$ in the worst case. Fig. 14(c) and (d) display the Test 2 results where all the fault parameter conditions were the same except the epicenter longitude. The results show that the TRRF can effectively capture the influence of the epicenter longitude. Summary statistics, which have only increased slightly compared to Test 1, confirm that Eq. 7 is valid (Table 2). Fig. 14(e) and (f) present the Test 3 results in which all fault parameter conditions are the same except for the strike and rake angles. Note that the TRRF predictions strongly align with the numerical simulation results while capturing the asymmetrical shape of the tsunami run-up distribution. Summary statistics, which increased only slightly from Test 1, confirm the

Table 2 Normalized root mean square error (%) of TRRF with respect to the numerical model (Basilisk)

	Test 1	Test 2	Test 3	Test 4
Mean	5.77	6.18	6.03	6.47
Std	1.18	1.18	1.47	1.36
Max	8.80	9.21	10.02	10.11
Min	3.37	3.40	3.45	3.32

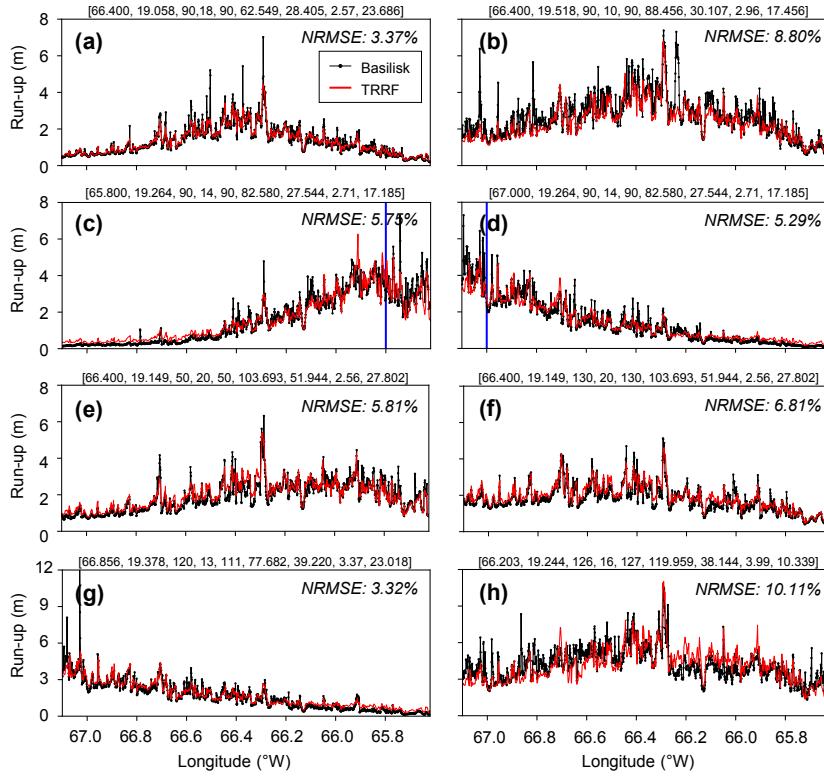


Fig. 14 Selected examples of tests: (a and b) Test 1. (c and d) Test 2. (e and f) Test 3. (g and h) Test 4. Black line and red line are the tsunami run-up distributions predicted by Basilisk and TRRF, respectively. The number inside the bracket above each pane represents the fault parameters in this sequence: $LONG(^{\circ}W)$, $LAT(^{\circ}N)$, $STR(^{\circ})$, $DIP(^{\circ})$, $RAK(^{\circ})$, $LEN(km)$, $WID(km)$, $SLP(m)$, $DEP(km)$. The blue line in (c) and (d) represents the location of epicenter longitude.

performance of the AP method (Table 2). Fig. 14(g) and (h) show the best case (minimum $NRMSE$) and worst case (maximum $NRMSE$) of Test 4, respectively. Both examples have a few localities where the TRRF did not predict the run-ups well, but the overall trend of the TRRF predictions agrees well with the numerical simulation results.

5.2 Computational time

The efficiency of the TRRF was investigated by comparing the computational time between the physics-based numerical model and the TRRF. When the physics-based numerical model (Basilisk) was used to predict the tsunami run-up distribution, the computational time was about one hour (24 CPU hours) on average for each scenario (24 cores, OpenMP, Intel Xeon E5-2680v3). On

the other hand, when the TRRF was used to predict the tsunami run-up distribution, the computational time was only 0.01 CPU second per scenario (desktop computer with one core, Intel I7-7700). The TRRF's CPU time is nearly 9 million times shorter than that of numerical simulation. The difference in computational time between the TRRF and the numerical model would be even greater given higher resolution grids and larger geographic areas than those used in this study.

6 Discussion

The performance of TRRF was investigated based on total 380 additional simulations in Section 5. When the TRRF predictions are compared against the direct numerical simulations, it is clear that the TRRF can produce reliable run-up predictions over real topography, given the computational time. However, as shown in Fig. 14, even though the TRRF predicts the leading order of tsunami run-up distribution well, there are a few localities where the difference of the run-up is more than two-fold. We found that these localities are correlated to the uncertainty (or the range of percentiles) of $NT^p(x)$ (Fig. 9). For example, a large uncertainty was commonly found in places with complex topography, such as areas surrounded by mountains (e.g., $65.735^\circ W$), areas containing a river (e.g., $66.955^\circ W$), steep cliffs (e.g., $66.444^\circ W$), and coastal dunes (e.g., $66.239^\circ W$). Even though it is difficult to fully interpret the physics behind the normalized topographic run-up $NT^p(x)$, we think that this high uncertainty may be attributed to the nonlinear behavior of the tsunami wave as it propagates and inundates complex topography. In its present form, the TRRF does not directly consider potential nonlinearities between the source and topographic run-up components in the hypothesis that the tsunami run-up distribution can be expressed as a sum of the source and topographic run-ups (Eq. 1). Future studies should investigate ways to account for the uncertainty to improve the accuracy of the TRRF approach.

Moreover, future studies should expand the applicability of the TRRF by considering the following limitations. One is that the TRRF is only applicable to uniform slip distribution. Several studies have shown that tsunami prediction can vary depending on heterogeneous slip models even when the earthquake magnitude is the same (Geist, 2002; Li et al., 2016; Ruiz et al., 2015). The other limitation of the TRRF is that it is only applicable to tsunamis generated by seafloor displacements associated with earthquakes. After earthquakes, landslides are the second most common cause of tsunamis (Harbitz et al., 2014). Moderate earthquakes do not always cause tsunamis themselves, but they can, in some instances, trigger large landslides that result in tsunamis (Uri et al., 2009). Though landslide-generated tsunamis are rare, a single occurrence can cause substantial damage and loss of life. For example, in 2017, a landslide-generated tsunami off the western coast of Greenland flooded several villages and resulted in casualties (Paris et al., 2019). A recent study also revealed that the 2018 Indonesian tsunami, which claimed more than 2,000 lives

and severely damaged coastal communities, was caused by the combination of an earthquake and a landslide (Sassa and Takagawa, 2019). Several other key elements would merit attention in future studies. For example, the arrival time and inundation distance are as important to consider as the run-up. A high tide could enhance tsunami inundation, while a receding tide could dissipate tsunami energy (Tolkova et al., 2015; Zhang et al., 2011). Likewise, a modest amount of sea-level rise could dramatically impact the tsunami run-up distribution (Li et al., 2018).

7 Conclusions

In the present study, we presented a new methodology, called TRRF, that can predict the alongshore run-up distribution from a near-field tsunami. We adopted the OS formula and developed what we call the AP method to reduce the number of simulations to build the TRRF. The tsunami run-up distribution was decomposed into source run-up and topographic run-up, that source run-up can be modeled by earthquake fault parameters, and that normalized topographic run-up is associated with local topographic characteristics. Using the northern region of Puerto Rico as a case study, the performance of the TRRF was investigated based on total 380 additional simulations. The results showed that the TRRF can produce rapid near-field tsunami run-up predictions over real topography. We expect that future applications of the TRRF will have the potential to save lives and promote resiliency of coastal communities.

Acknowledgements This material is based upon work supported by the National Science Foundation under Grant Nos. 1630099 and 1735139. The authors acknowledge Advanced Research Computing at Virginia Tech for providing computational resources and technical support that have contributed to the results reported within this paper. URL: <http://www.arc.vt.edu>

Appendix A: Design of experiments

The tsunamigenic-earthquake scenarios were designed in a three-level full factorial. The level of the design of experiments (DoE) was determined based on preliminary simulations where 60 cases were considered as follows. We set a reference case where the fault parameters are as follows: $LON = 66.4^\circ W$, $LAT = 19.3^\circ N$, $STR = 90^\circ$, $DIP = 20^\circ$, $RAK = 90^\circ$, $LEN = 90km$, $WID = 40km$, $SLP = 3m$, $DEP = 20km$, which are the same as the central level values used in Table 1. Based on the reference case, we performed 60 simulations, varying each of the six fault parameters (LAT , LEN , WID , DIP , SLP , DEP) one at a time through a uniformly distributed array of 10 values within the parameter's range shown in Table 1. The 60 simulations show that the 2nd order polynomial model, which requires at least three-level

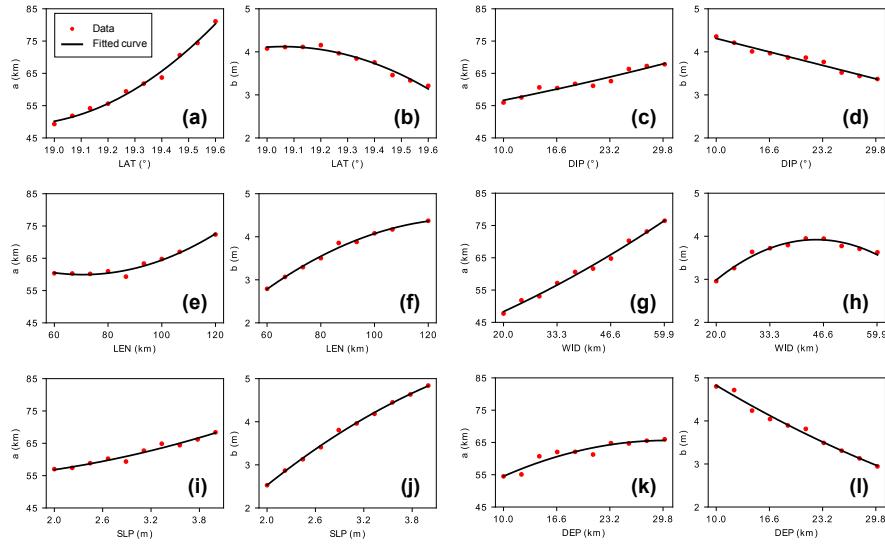


Fig. 15 The OS formula coefficients (a and b) variation in terms of RSM fault parameters: (a,b) epicenter latitude, (c,d) dip angle, (e,f) fault length, (g,h) fault width, (i,j) slip, (k,l) top-edge fault depth. The red dots represent the simulated coefficients and the black line represents the best-fitting curve based on the 2nd-order polynomial model.

data, is enough to fit the fault parameters to the OS formula's coefficients a and b (Fig. 15).

References

Aki K (1966) Generation and propagation of g waves from the niigata earthquake of june 16, 1964. part 2. estimation of earthquake movement, released energy, and stress-strain drop from the g wave spectrum. *Bulletin of the Earthquake Research Institute* 44:73–88

Amante C, Eakins BW (2009) Data from "etopo1 1 arc-minute global relief model" national geophysical data center, noaa. available at <https://dx.doi.org/10.7289/v5c8276m>. deposited 1 january 2009.

Annaka T, Satake K, Sakakiyama T, Yanagisawa K, Shuto N (2007) Logic-tree approach for probabilistic tsunami hazard analysis and its applications to the japanese coasts. *Pure & Applied Geophysics* 164

Blaser L, Krüger F, Ohrnberger M, Scherbaum F (2010) Scaling relations of earthquake source parameter estimates with special focus on subduction environment. *Bulletin of the Seismological Society of America* 100(6):2914–2926, DOI 10.1785/0120100111

Box GE, Wilson KB (1951) On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society: Series B (Methodological)* 13(1):1–38

Geist EL (2002) Complex earthquake rupture and local tsunamis. *Journal of Geophysical Research: Solid Earth* 107(B5):ESE 2–1–ESE 2–15, DOI 10.1029/2000JB000139

Gica E, Spillane MC, Titov VV, Chamberlin CD, Newman JC (2008) Development of the forecast propagation database for noaa's short-term inundation forecast for tsunamis (sift). NOAA Technical Memorandum OAR PMEL 139

Grilli ST, Dubosq S, Pophet N, Pérignon Y, Kirby JT, Shi F (2010) Numerical simulation and first-order hazard analysis of large co-seismic tsunamis generated in the Puerto Rico trench: Near-field impact on the North shore of Puerto Rico and far-field impact on the US East Coast. *Natural Hazards and Earth System Science* 10(10):2109–2125, DOI 10.5194/nhess-10-2109-2010

Hanks TC, Kanamori H (1979) A moment magnitude scale. *Journal of Geophysical Research: Solid Earth* 84(B5):2348–2350

Harbitz CB, Løvholt F, Bungum H (2014) Submarine landslide tsunamis: how extreme and how likely? *Natural Hazards* 72(3):1341–1374

Heidarzadeh M, Harada T, Satake K, Ishibe T, Takagawa T (2017) Tsunamis from strike-slip earthquakes in the wharton basin, northeast indian ocean: March 2016 m w7. 8 event and its relationship with the april 2012 m w 8.6 event. *Geophysical Journal International* 211(3):1601–1612

Kamigaichi O (2011) *Tsunami Forecasting and Warning*, Springer New York, New York, NY, pp 982–1007

LeVeque RJ, George DL, Berger MJ (2011) Tsunami modelling with adaptively refined finite volume methods. *Acta Numerica* 20:211–289

Li L, Switzer AD, Chan CH, Wang Y, Weiss R, Qiu Q (2016) How heterogeneous coseismic slip affects regional probabilistic tsunami hazard assessment: A case study in the south china sea. *Journal of Geophysical Research: Solid*

Earth 121(8):6250–6272

Li L, Switzer AD, Wang Y, Chan CH, Qiu Q, Weiss R (2018) A modest 0.5-m rise in sea level will double the tsunami hazard in macau. *Science advances* 4(8):eaat1180

Lin SC, Wu TR, Yen E, Chen HY, Hsu J, Tsai YL, Lee CJ, Philip LFL (2015) Development of a tsunami early warning system for the south china sea. *Ocean Engineering* 100:1 – 18, DOI <https://doi.org/10.1016/j.oceaneng.2015.02.003>

López-Venegas AM, Horrillo J, Pampell-Manis A, Huérano V, Mercado A (2015) Advanced tsunami numerical simulations and energy considerations by use of 3d–2d coupled models: The october 11, 1918, mona passage tsunami. *Pure and Applied Geophysics* 172(6):1679–1698

Mori N, Goda K, Cox D (2018) Recent process in probabilistic tsunami hazard analysis (ptha) for mega thrust subduction earthquakes. In: *The 2011 Japan Earthquake and Tsunami: Reconstruction and Restoration*, Springer, pp 469–485

Mulia IE, Gusman AR, Satake K (2018) Alternative to non-linear model for simulating tsunami inundation in real-time. *Geophysical Journal International* 214(3):2002–2013, DOI 10.1093/gji/ggy238

Myers RH, Montgomery DC, Anderson-Cook CM (2016) Response surface methodology: process and product optimization using designed experiments. John Wiley & Sons

Nealon JW, Dillon WP (2001) Earthquakes and tsunamis in puerto rico and the us virgin islands. Tech. rep., US Geological Survey

NGDC (2005) Data from "u.s. coastal relief model - puerto rico." national geophysical data center, noaa. available at <https://dx.doi.org/10.7289/v57h1ggw>. deposited 1 january 2005. URL <https://www.ngdc.noaa.gov/metadata/geoportal/rest/metadata/item/gov.noaa.ngdc.mgg.dem:290/html>

Okada Y (1985) Surface deformation due to shear and tensile faults in a half-space. *Bulletin of the seismological society of America* 75(4):1135–1154

Okal EA, Synolakis CE (2004) Source discriminants for near-field tsunamis. *Geophysical Journal International* 158(3):899–912, DOI 10.1111/j.1365-246X.2004.02347.x

Paris A, Okal EA, Guérin C, Heinrich P, Schindelé F, Hébert H (2019) Numerical modeling of the june 17, 2017 landslide and tsunami events in karrat fjord, west greenland. *Pure and Applied Geophysics* pp 1–23

Park H, Cox DT (2016) Probabilistic assessment of near-field tsunami hazards: Inundation depth, velocity, momentum flux, arrival time, and duration applied to seaside, oregon. *Coastal Engineering* 117:79–96

Park H, Cox DT, Barbosa AR (2018) Probabilistic tsunami hazard assessment (ptha) for resilience assessment of a coastal community. *Natural Hazards* 94(3):1117–1139

Popinet S (2015) A quadtree-adaptive multigrid solver for the serre–green–naghdi equations. *Journal of Computational Physics* 302:336–358

Reid HF, Taber S (1919) The porto rico earthquakes of october-november, 1918. *Bulletin of the Seismological Society of America* 9(4):95–127

Ruiz JA, Fuentes M, Riquelme S, Campos J, Cisternas A (2015) Numerical simulation of tsunami runup in northern Chile based on non-uniform k2slip distributions. *Natural Hazards* 79(2):1177–1198, DOI 10.1007/s11069-015-1901-9

Sassa S, Takagawa T (2019) Liquefied gravity flow-induced tsunami: first evidence and comparison from the 2018 indonesia sulawesi earthquake and tsunami disasters. *Landslides* 16(1):195–200

Setiyono U, Gusman AR, Satake K, Fujii Y (2017) Pre-computed tsunami inundation database and forecast simulation in pelabuhan ratu, indonesia. *Pure and Applied Geophysics* 174(8):3219–3235, DOI 10.1007/s00024-017-1633-8

Shi F, Kirby JT, Harris JC, Geiman JD, Grilli ST (2012) A high-order adaptive time-stepping tvd solver for boussinesq modeling of breaking waves and coastal inundation. *Ocean Modelling* 43:36–51

Thingbaijam KKS, Mai PM, Goda K (2017) New empirical earthquake source-scaling laws. *Bulletin of the Seismological Society of America* 107(5):2225–2246, DOI 10.1785/0120170017

Titov V, Rabinovich AB, Mofjeld HO, Thomson RE, González FI (2005) The global reach of the 26 december 2004 sumatra tsunami. *Science* 309(5743):2045–2048

Tolkova E, Tanaka H, Roh M (2015) Tsunami observations in rivers from a perspective of tsunami interaction with tide and riverine flow. *Pure and Applied Geophysics* 172(3-4):953–968

Uri S, Lee HJ, Geist EL, Twichell D (2009) Assessment of tsunami hazard to the us east coast using relationships between submarine landslides and earthquakes. *Marine Geology* 264(1-2):65–73

USGS (2017) Data from "advanced national seismic system (anss) comprehensive catalog of earthquake events and products." earthquake hazards program, u.s. geological survey. available at <https://doi.org/10.5066/f7ms3qzh>. deposited 1 january 2017. URL <https://earthquake.usgs.gov/earthquakes/search/>

Vincenty T (1975) Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey review* 23(176):88–93

Wei Y, Chamberlin C, Titov VV, Tang L, Bernard EN (2013) Modeling of the 2011 japan tsunami: Lessons for near-field forecast. *Pure and Applied Geophysics* 170(6-8):1309–1331

Zhang YJ, Witter RC, Priest GR (2011) Tsunami–tide interaction in 1964 prince william sound tsunami. *Ocean Modelling* 40(3-4):246–259