Response to Reviewer #1

We would like to thank the reviewer for their precious time and for the thorough review and the many helpful comments and suggestions made to improve the present work. Please find below the reviewer’s comments and author’s replies to these comments.

Assessment and Further comments

Q1. “[…] I don’t understand the criteria followed for leaving out some important models, especially those models used by different Tsunami Warning Centers along the World […].” “[…] Of course, it’s not possible to include all the developed tsunami models, but as the development of TRITON-G was focused on its operational capabilities, it would be expectable to focus this analysis on models used in this context, moreover if some these models have been developed in the GPU […]”

A1. We based the previous existing work on an historical development and on the models we considered well known and that we were more familiar with, as well as some models mentioned in the National Tsunami Hazard Mitigation Program (NTHMP, 2012). As mentioned by the reviewer is not possible to include all models and we appreciate the important references provided about these other models left out which will be included in the introduction, specially the GPU models suggested since they are relevant to our research.

Section 2

Q2. So, regarding the numerical methods used in this paper: in the propagation stage it’s used the method of characteristics (MOC) in combination with a cubic-polynomial interpolation to find the interpolated values on the “[…] Is the total mass conserved? For instance, what about the mass conservation in the experiment described in p8 line 10?”

A2. On this regards, the discretization used for the run-up method is non-conservative. However, the difference from the initial mass and final mass is negligible and well below the 5% criteria that the NOAA Technical Memorandum OAR PMEL-135 suggests (Synolakis et al., 2007). The long wave typically associated with tsunamis implies a small error in the result and the analytical solution show excellent agreement in the experiment mentioned which implies that no large mass loss is present.

Q3. Regarding the water height positivity (h), is it guaranteed in the propagation stage even considering the cubic polynomial interpolation process?
A3. The propagation stage of the program is not used in the run-up calculation or on land areas. Coasts not flagged as “Inundation” have a wall boundary treatment. In the case of the run-up calculation, the surface gradient method is used. It is constructed as a monotone scheme with flux correction and a slope limiter to preserve height positivity. Dam-break problem tests with thin water layers have been satisfactory with the wave remaining positive even in shallow conditions. In addition, the new run-up benchmark problems (BP) added to the paper contribute as another demonstration of positivity preservation.

Q4. On other hand, what is the convergence order of the numerical models used on each area?

A4. The convergence of the method of characteristic used for propagation and the method for run-up are different. The propagation is 3rd order while the run-up is a 1.5 order. A measurement of this convergence for the parabolic bowl problem is included to illustrate this point (Fig. A).

![Fig A. Parabolic bowl problem cross section with $\epsilon = 10^{-6}$ on left panel. Water depth error for parabolic bowl problem on right panel](image)

Section 3

Q5. As the coordinate system used combines spherical coordinates and cartesian coordinates, how is treated the boundary between the considered domains on each coordinate system?

A5. The Cartesian coordinate system is used exclusively for the blocks set as “Inundation”. These blocks are only found at level 7 (highest level) and only inside focal areas. While the focal areas can be user-defined in any size and located on any coast in the domain, by design the focal areas shall consist of just a few kilometers in length. In this way, the area represented by the inundation areas is extremely small compared to the total domain size. This makes it possible not to use a special boundary treatment between systems since the difference of the incoming wave is almost negligible. Additionally, the
domain used centered in the Indian Ocean does not extend to high latitudes that might introduce large discrepancies in the grid. In Fig. B, the current focal areas used by TRITON-G are shown with the approximate length in kilometers. As noted, they cover just a few kilometers, the largest case being Seychelles, which is just around 27 km by 17 km.

![Fig. B. Focal Areas highlighted in green with the approximate length labeled in kilometers. Top Left: Sri Lanka. Top Right: Comoros. Bottom left: Mozambique. Bottom Right: Seychelles](image)

Q6. Regarding the run-up calculation, in section 3.2 authors propose a technic based on considering \((hu=hv=0)\) when \(h\) is less than a certain small fixed quantity. They confirm that the proposed implementation has been proven to be robust and stable under different benchmarks. Please, include a reference where this numerical technic is used.

A6. Some references about the thin layer technique and its implementation are:


Q7. Finally, in this section it’s included a parabolic bowl problem as validation problem. That’s a good synthetic test but I consider that it’s not enough to define a numerical model as validated when such numerical model that is going to be used with real topobathymetries in real cases, moreover when, as it’s remarked by authors, the model is going to use in the RIMES context. In order to validate this numerical scheme to be used in real cases I suggest, for instance, the use of the inundation benchmark experiments proposed by the National Tsunami Hazard Mitigation Program (NTHMP) where problems with analytical solutions, laboratory experiments and real problems with field measurements are proposed. I would suggest studying the behavior of TRITONG at least in the “mandatory” benchmark problems: BP1, BP4, BP6, BP7 and BP9. Regarding the field measurement experiments, authors show in section 6.1.2 some inundation maps in different locations (Fig. 18 and 19) where the comparisons are made basically against other models. I think that it cannot be considered a numerical model as validated with field measurements by making comparisons basically with other models results. There are available many tsunami field measurements to validate the inundation process. In this sense you can consider cases where there are more available data than in the case studied in section 6.1.2. For instance: BP7, BP9 or many inundation scenarios related to the Tohoku, 2011 event where detailed data are available.

A7. We appreciate the suggestion to validate TRITON-G against the inundation benchmark problems. We have completed the results for benchmarks BP4, BP6, BP7 and BP9. We skipped benchmark problem BP1 considering that it is a one-dimensional analytical problem and we already showed results with good agreement for the parabolic bowl problem, which is a two-dimensional analytical problem. Additionally BP4 is the experimental version of BP1 and results for BP4 are now included. This serves to demonstrate the behavior obtained with TRITON-G under these conditions.

Regarding the map comparison with another model (Fig. 18 and Fig 19) we consider that is an acceptable way to show agreement since the results from these models are peer-reviewed and published (Supparsri et al., 2011). However, in order to produce a more complete validation of our model we agree with the importance of using tsunami field measurements and known benchmark problems. For this reason, we also use BP7 and BP9 as comparison to validate the inundation process.
A new section was added to the manuscript to include results for BP9 (Section 6) while the rest of the benchmarks are located in the appendix.

1. Benchmark problem #4: Solitary wave on a simple beach – Laboratory

Figure C depicts the domain for this test. In this problem, the wave height $H$ is located at a distance $L$ from the beach toe. This test was replicated in a wave tank 31.73-cm-long, 60.96-cm-deep and 39.97-cm-wide at the California Institute of Technology. Several experiments with different water heights were performed. Benchmark Problem 4 (BP4) uses the datasets for $H/d = 0.0185$ non-breaking wave and $H/d = 0.30$ breaking wave for code validation. Results use dimensionless units with the help of parameters like length $d$, velocity scale $U = \sqrt{gd}$ and time scale $T = \sqrt{d/g}$.

![Fig. C Domain sketch for BP4 with slope 1:19.85 (figure taken from benchmark description)](image)

1.1 Problem setup

- **Parameters**: $d = 1$, $g = 9.8$, case A with $H/d = 0.0185$ and case C with $H/d = 0.30$.
- **Friction**: Manning coefficient set to 0.01
- **Computational domain**: the domain along $x$ direction spanned from $x = -20$ to $x = 80$
- **Boundary conditions**: the right side of the computational domain uses a non-reflective boundary condition.
- **Grid resolution**: the numerical results are solved with a resolution of $\Delta x = 0.1$
- **CFL**: 0.9
- **Initial condition**: the initial wave is computed based on the following equations for height ($\eta$) and velocity ($u$) respectively

\[
\eta(x, 0) = H \text{sech}^2[y(x - x_c)/d],
\]

\[
u(x, 0) = -\eta(x, 0) \sqrt{\frac{g}{d}}.
\]

1.2 Tasks to be performed

To accomplish this problem, the following tasks should be performed:
1. Compare numerically calculated surface profiles at $t/T=30:10:70$ for the non-breaking case $H/d = 0.0185$ with the lab data (Case A).

2. Compare numerically calculated surface profiles at $t/T=15:5:30$ for the breaking case $H/d = 0.30$ with the lab data (Case C).

3. Compute maximum run-ups for at least one non-breaking and one breaking wave case.

1.3 Numerical results

We present the numerical results obtained using TRITON-G. Figure D shows the comparison between water surface level measured in the experiment and the modeled numerical results obtained by our model for times 30, 40, 50, 60 and 70 for case A ($H/d = 0.0185$). Our results show good agreement between the numerical simulation and the non-breaking experiment.

Table A shows the errors computed for the normalized root mean square deviation (NRMSD) and for the maximum wave amplitude error (MAX). The error values obtained by the NTHMP workshop models are also included for comparison (taken from Table 1-8, page 41 in (NTHMP, 2012)). These values are divided into two columns, one with results for the non-dispersive models (ND) and the other with results for the non-dispersive and dispersive models together (labeled ALL). Errors obtained from our simulation tend to be similar or smaller than those errors obtained by other ND models, with just slight exception for time 70. Additionally, except for time 70 our errors are smaller than those obtained combining non-dispersive and dispersive mean error value.

Water level comparison for case C ($H/d = 0.30$) at times 15, 20, 25 and 30 is shown in Fig. E. Table B gathers the values for NRMSD and MAX errors for our numerical results and for the NTHMP workshop models. In this case, only the results of models that reported their errors are included (taken from Table 1-8, page 41 in (NTHMP, 2012)).

For case C conditions, the shallow water equations are no longer appropriate for modeling and hydrostatic models tend to produce larger differences than non-hydrostatic ones. Our numerical results in general show good agreement with the experiment. The difference with the steepening of the crest that is noticeable in the results is expected from a hydrostatic model. In spite of that, this steeping in our model is not very large and it can trace the wave front well. Once the wave breaking occurs, our model can simulate reasonably well the run-up. This is also partly reflected in the small NRMSD error estimation obtained by our model after the wave breaking.

Maximum run-up for case A and case C were calculated. For the non-breaking case A, the obtained run-up value is 0.091 and for the breaking case C, the run-up estimated is 0.588. These values are plotted in Fig. F with a yellow and red dot respectively, it can be seen that both values lie well within the experimental results.
Table A. Model surface profile errors with respect to laboratory experiments for case A

\( H/d = 0.0185 \) at times 30, 40, 50, 60, and 70. Results from the NTHMP workshop errors are divided into non-dispersive (ND) models and all models (ALL).

<table>
<thead>
<tr>
<th></th>
<th>TRITON-G</th>
<th>NTHMP</th>
<th>TRITON-G</th>
<th>NTHMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>H = 0.0185</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T 30</td>
<td>8.8</td>
<td>11</td>
<td>4.0</td>
<td>6</td>
</tr>
<tr>
<td>T 40</td>
<td>6.6</td>
<td>9</td>
<td>4.8</td>
<td>3</td>
</tr>
<tr>
<td>T 50</td>
<td>3.5</td>
<td>6</td>
<td>7.4</td>
<td>13</td>
</tr>
<tr>
<td>T 60</td>
<td>3</td>
<td>4</td>
<td>1.4</td>
<td>1</td>
</tr>
<tr>
<td>T 70</td>
<td>11</td>
<td>33</td>
<td>13.5</td>
<td>15</td>
</tr>
</tbody>
</table>

Table B. Modeled surface profile errors with respect to laboratory experiments for case C

\( H/d = 0.30 \) at times 15, 20, 25 and 30. Results from the NTHMP workshop model errors available are shown (ALL).

<table>
<thead>
<tr>
<th></th>
<th>TRITON-G</th>
<th>NTHMP</th>
<th>TRITON-G</th>
<th>NTHMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>H = 0.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T 15</td>
<td>11.3</td>
<td>7</td>
<td>5.4</td>
<td>6</td>
</tr>
<tr>
<td>T 20</td>
<td>5.9</td>
<td>9</td>
<td>23.3</td>
<td>11</td>
</tr>
<tr>
<td>T 25</td>
<td>6.5</td>
<td>6</td>
<td>11.1</td>
<td>10</td>
</tr>
<tr>
<td>T 30</td>
<td>2.9</td>
<td>4</td>
<td>1.4</td>
<td>6</td>
</tr>
</tbody>
</table>
Fig. D. Comparison of numerically calculated free surface profile at different dimensionless times for the non-breaking case $H/d = 0.0185$
Fig. E. Comparison of numerically calculated free surface profile at different dimensionless times for the breaking case $H/d = 0.30$
Fig. F. Scatter plot of non-dimensional maximum run-up from a total of more than 40 experiments conducted by Y. Joseph Zhan (Synolakis, 1987). Orange point indicates TRITON-G result for the breaking case $H/d = 0.30$ and yellow point indicates the result for the non-breaking run-up case $H/d = 0.0185$

2. Benchmark problem #6: Solitary wave on a conical island – Laboratory

The goal of this benchmark is to compare computed model results with laboratory measurements obtained during a physical modeling experiment conducted at the Coastal and Hydraulic Laboratory Engineer Research and Development Center of the U.S. Army Corps of Engineers. The laboratory physical model was constructed as an idealized representation of Babi Island, in the Flores Sea, Indonesia, to compare with Babi Island run-up measured shortly after the 12 December 1992 Flores Island tsunami (Yeh et al., 1994). Figure G shows schematics of the experiment.

2.1 Tasks to be performed

To accomplish this benchmark, it is suggested that, for

- Case A: water depth $d= 32.0$ cm, target $H=0.05$, measured $H=0.045$
- Case B: water depth $d= 32.0$ cm, target $H=0.20$, measured $H=0.096$
- Case C: water depth $d= 32.0$ cm, target $H=0.05$, measured $H=0.181$

model simulations be conducted to address the following:

1. Demonstrate that two wave fronts split in front of the island and collide behind it
2. Compare computed water levels with laboratory data at gauge 6, 9, 16 and 22
3. Compare computed island run-up with laboratory gauge data
2.2 Problem setup

- **Computational domain (in meters):** [-5,23] \( \times \) [0, 28]
- **Boundary condition:** open boundaries
- **Initial condition:** same solitary wave as proposed in BP4 with the correction for two dimensions
- **Grid resolution:** the numerical results presented are solved with a resolution of \( \Delta x = 0.05 \)
- **CFL:** 0.9
- **Friction:** Manning coefficient set to 0.02

Fig. G Basin geometry and coordinate system. Solid lines represent approximate basin and wave maker surfaces. Circles along walls and dashed lines represent wave-absorbing material.
Fig. H Snapshots at several times showing the wave front splitting in front of the island and colliding behind it for BP6 case B.

2.3 Numerical results

We present the numerical results obtained using TRITON-G for the three cases (A, B and C) except for the splitting-colliding item. For this item, Figure H shows the wave front splitting in front of the island and then colliding again behind it for case B (H=0.096), analogue behavior was obtained for the other two cases.
<table>
<thead>
<tr>
<th>Case A</th>
<th>NRMSD</th>
<th>MAX</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TRITON-G</td>
<td>NTHMP</td>
</tr>
<tr>
<td></td>
<td>ND</td>
<td>ALL</td>
</tr>
<tr>
<td>Gauge 6</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>Gauge 9</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Gauge 16</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>Gauge 22</td>
<td>9</td>
<td>8</td>
</tr>
</tbody>
</table>

| Case B |
|--------|-------|-----|
|        | NRMSD | MAX |
|        | TRITON-G | NTHMP | TRITON-G | NTHMP |
|        | ND | ALL | ND | ALL |
| Gauge 6 | 10 | 8 | 8 | 7 | 6 | 6 |
| Gauge 9 | 9 | 8 | 8 | 2 | 7 | 9 |
| Gauge 16 | 9 | 7 | 7 | 14 | 7 | 7 |
| Gauge 22 | 8 | 9 | 9 | 6 | 40 | 27 |

| Case C |
|--------|-------|-----|
|        | NRMSD | MAX |
|        | TRITON-G | NTHMP | TRITON-G | NTHMP |
|        | ND | ALL | ND | ALL |
| Gauge 6 | 13 | 10 | 8 | 3 | 6 | 5 |
| Gauge 9 | 12 | 11 | 11 | 2 | 9 | 13 |
| Gauge 16 | 10 | 9 | 8 | 4 | 3 | 3 |
| Gauge 22 | 9 | 8 | 8 | 10 | 18 | 15 |

Table C. Water level time series TRITON-G model errors with respect to laboratory experiment data for case A, B and C. Mean values obtained for the performing NTHMP models is separated in non-dispersive models (ND) and dispersive and non-dispersive models together (ALL).

Water level comparison uses values for gauges 6, 9, 16 and 22 for each of the 3 cases. Gauge 6 is located at (9.36, 13.80, 31.7), Gauge 9 is located at (10.36, 13.80, 8.2), Gauge 16 is located at (12.96, 11.22, 7.9) and Gauge 22 is located at (15.56, 13.80, 8.3).

Numerical results for Case A, B and C are shown in Figures I, J and K respectively. In the three cases results were stable and in good agreement with the experimental values. The incident wave height and arrival time was captured well for all gauges. Similarly as with BP4, the steepening of the wave with increasing H is expected in a non-hydrodynamic model. After the wave hit the island, some differences between experimental and model wave are noticeable as the initial wave height increased. These oscillations in the experimental data represent the effects of dispersion, which our non-dispersive numerical method is not designed to capture. Despite this, the modeled waves show good agreement with the shape of the experimental waves and the errors estimated tend to be small.
Fig. I Comparison between computed and measured water levels at gauges 6, 9, 16 and 22 for case A (H=0.045)
Fig. J Comparison between computed and measured water levels at gauges 6, 9, 16 and 22 for case B (H=0.096)
Fig. K Comparison between computed and measured water levels at gauges 6, 9, 16 and 22 for case C (H=0.181)
Fig. 1. Comparison between computed and measured run-up around the island for the three cases in BP6.
### Table D. Run-up TRITON-G model errors with respect to laboratory experiment data for case A, B and C. Mean error values obtained by the performing NTHMP models are separated in non-dispersive models (ND) and all models (ALL)

<table>
<thead>
<tr>
<th>Run-up</th>
<th>TRITON-G</th>
<th>NTHMP</th>
<th>MAX</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ND</td>
<td>ALL</td>
</tr>
<tr>
<td>Case A</td>
<td>9</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>Case B</td>
<td>19</td>
<td>21</td>
<td>18</td>
</tr>
<tr>
<td>Case C</td>
<td>20</td>
<td>12</td>
<td>11</td>
</tr>
</tbody>
</table>

Table C gathers the normalize root mean square deviation (NRMSD) error and the maximum wave height (MAX) error. For comparison, mean errors obtained by the participating models in the NTHMP workshop are also included. These are separated in two columns, one for non-dispersive (ND) models and the other for non-disperse and disperse models together (ALL).

NRMSD errors for our model tend to be not very large and in similar range than those of the other non-dispersive models. In the case of the maximum height error (MAX), in almost all cases our model produced smaller error values than the non-dispersive model counterparts. Additionally, in most cases our MAX errors are smaller than those errors of the combined non-dispersive and dispersive mean values.

Figure L shows the comparisons between computed and experimental run-up around the island for the three cases. Case A represent the best agreement with the experimental values. Differences increased with steeper wave cases B and C as several reflections and refraction possibly occur in the basin.

3 Benchmark problem #7: The tsunami run-up onto a complex 3-D beach. Laboratory.

A laboratory experiment using a large-scale tank at the central Research Institute for Electric Power Industry in Abiko, Japan was focused on modelling the run-up of a long wave on a complex beach near the village of Monai (Liu et al., 2008). The beach in the tank was a 1:400-scale model of the bathymetry and topography around a very narrow gully, where extreme run-up was measured.
3.1 Problem setup

The following parameters were used for the computation:

- **Grid resolution**: 393x244 was used with the same resolution 0.014 m as the bathymetry.
- **CFL**: 0.9
- **Initial condition**: water at rest.
- **Friction**: Manning coefficient set to 0.01
- **Boundary conditions**: Solid wall boundary used at the top and bottom. At the left boundary, the given initial wave (shown in Fig. M) was used to specify the condition up to time \( t=22.5 \) s, after that it became a wall boundary condition.

![Input Wave](image)

Fig. M Prescribed input wave for the left boundary condition defined from \( t=0 \) to \( t=22.5 \) s

3.2 Tasks to be performed

To accomplish this benchmark it is suggested to:

1. Model propagation of the incident and reflective wave accordingly to the benchmark-specified boundary condition.
2. Compare the numerical and laboratory-measured water level dynamics at gauges 5, 7 and 9.
3. Show snapshots of the numerically computed water level at the time synchronous with those of the video frames.
4. Compute maximum run-up in the narrow valley.

3.3 Numerical results

This section presents the numerical results for BP7 obtained with TRITON-G to achieve the required tasks.
The comparison with the three requested gauges 5, 7 and 9 is shown in Fig. N from $t = 0$ to $t = 25$ s. For the three cases, good agreement is found between modeled and experimental wave. Values for the normalized Root Mean Square deviation error (NRMSD) and maximum wave amplitude error (MAX) were estimated for the gauge comparison. For gauge 5, the NRMSD error is 10% and MAX is 0.89%. For gauge 7, NRMSD is 10% and MAX is 4.81%. For gauge 9, the NRMSD error is 6.57% and MAX is 2.66%.
Fig. O Comparison between extracted movie frames (left) and TRITON-G simulation (right) for times 15, 15.5, 16, 16.5 and 17 seconds
Comparison with the extracted movie frames is shown in Fig. O. In the left column are the five frames provided from the laboratory recording. These are frames 10, 25, 40, 55 and 70, extracted from the video with a 0.5 s interval. We found good agreement in time and space for times 15 s to 17 s in 0.5 s increments, shown in the right column. The side-by-side comparison shows that the modeled wave follows the experimental wave front well. Additionally, the model captures the rapid run-up/run-down in the narrow gully.

Finally, the data provided by the benchmark workshop include a series of experiment tests for maximum run-up. Its maximum run-up is recorded at $x = 5.1575$ and $y = 1.88$ m with an average value of approximately 0.09 m. In comparison, our numerical result recorded a maximum run-up at around $t = 16.5$ with a height of 0.0936 m at $x = 5.15$ and $y = 1.88$ m.

4 Benchmark Problem #9: Okushiri Island Tsunami - Field

This benchmark problem (BP9) is based on the data collected from the Mw 7.8 Hokkaido-Nansei-Oki tsunami around Okushiri Island in Japan in 1993. The goal is to compare computed model results with the field measurements.

4.1 Problem setup

The following parameters were used for the computation:

- **Bathymetry**: taken from databases provided by (NTHMP, 2012), interpolated where necessary.
- **CFL**: 0.9
- **Simulated time**: 60 minutes
- **Initial condition**: source generated from the database provided by DCRC (Disaster Control Research Center) Japan solution DCRC17a, described in (Takahashi, 1996).
- **Boundary conditions**: open boundaries at the four domain edges.
- **Friction**: Manning coefficient set to 0.02
- **Computational domain**: a mesh refinement is used on the entire domain (shown in Fig. P). Seven levels are used in total. The resolution of base level 1 is 450 m and the resolution of level 7 is approximately 7 m. Dry blocks that did not take part in the computation were removed in the mesh generation process.
4.2 Tasks to be performed

This benchmark requires the following tasks to be performed:

1. Compute run-up around Aonae
2. Compute arrival of the first wave to Aonae
3. Show two waves at Aonae approximately 10 minutes apart; the first wave came from the wet, the second wave came from the east
4. Compute water level at Iwanai and Esashi tide gauges
5. Maximum modeled run-up distribution around Okushiri island
6. Modeled run-up height at Hamatsumae
7. Modeled run-up height at a valley north of Monai.

4.3 Numerical results

In this section, we present the numerical results obtained with TRITON-G for benchmark problem #9.

4.3.1 Run-up around Aonae

The maximum inundation around Aonae peninsula modeled during the simulation is shown in Fig Q. Contours every 4 meters are drawn to show the outline of the topography. Maximum inundation
height computed was nearly 15 meters but the scale used is set to the upper limit of 10 m to highlight the areas where major inundation occurred.

The west side of the peninsula received the impact of the first wave, which produced the largest inundation height. Maximum values of nearly 15 m were obtained in the simulation. Despite a relatively lower inundation height in the east side of the peninsula, deep penetration was found due to the flatter topography in this area. The inundation on the east side was mainly produced by the second wave coming from the east. The south side of the peninsula experienced the impact of both first and second waves and run-up of over 12 m was estimated.

4.3.2 Arrival of first wave to Aonae

The arrival of the first wave at Aonae peninsula is shown in Fig. R. This wave is coming from the west. Snapshots are approximately 5 seconds apart at times 4.9 min and 5.0 min to illustrate the wave arrival. From these snapshots, we estimate that the wave made impact at around 5 minutes after the tsunami generation.
4.3.3 Two waves arriving at Aonae

Fig. S Two waves arriving at Aonae peninsula. Left: first wave coming from the west arrived at around t=5 min. Right: second wave coming from the east arrived at around t=16 min

The two waves arriving at Aonae peninsula are shown in Fig. S. The first one came from the west (Fig. S left) and made impact at around 5.0 min after the tsunami generation. The second major wave to hit the peninsula came from the east and made impact at around 16 min (Fig. S right). Slightly over 10 minutes separated the first and second wave.

4.3.4 Tide gauge comparison at Iwanai and Esashi

Comparison between computed and observed water levels at Iwanai and Esashi tide gauges is presented in Fig. T. The arrival time of the computed wave shows good agreement for Esashi station. The computed wave positive and negative phases also follows rather well the observed values. In the
case of Iwanai station the arrival time is slightly sooner than the observed however the observed wave phase is followed generally well in the computed results. The discrepancies between observed and computed values can be attributed to several reasons. Inaccuracies in the source used for the initial condition can influence greatly the result. Additionally, lack of realistic bathymetry including man-made structures around the area can affect the results as well.

4.3.5 Maximum run-up around Okushiri

The computed maximum run-up distribution around Okushiri Island is shown in Fig. U. Observations were taken from (Kato and Tsuji, 1994). Good agreement is found between observed and computed values around the coast. Most values are within the observed range or within a small difference from the field measurement. The simulation seems to capture well the variations that occurred along the coast.
The model could simulate well the maximum run-up observed around Monai valley within a reasonable 15% error. The major differences are found in the southwest side of the island where run-up values were underestimated with larger difference. The discrepancies could be explained by the use of different grid around the island coast. Additionally, the lack of an accurate high-resolution bathymetry database everywhere can also influence the computed values as well as an inaccurate initial condition.

4.3.6 Run-up height at Hamatsumae

The maximum inundation map for Hamatsumae region is shown in Fig. V. Topography and bathymetry contours are outlined every 4 meters. A grid resolution of approximately 14 m was used for this region. Near the center of the region and to the east, run-ups of nearly 16 meters were computed. Additionally, inundation values ranging from 8 to 10 meters were obtained which match well with field observations.
Fig. V Inundation map of Hamatsumae region with 4-m contours of bathymetry and topography.

4.3.7 Run-up height at a valley north of Monai

The maximum inundation map for the valley north of Monai is shown in Fig. W. Topography and bathymetry contours are outlined every 4 meters. A grid resolution of approximately 7 m was used for this region. Inundation of around 26 m was computed, relatively close to the 30.6 m observed in the field data.

Q8. Regarding the generation process described in section 3.3, authors use the coseismic deformation proposed by Smylie and Manshiha, 1971. I’m surprised at this point because in [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] this work is extended, and it’s provided also an analytical way
to get the “Okada model” coseismic deformation that can easily be extended to CUDA in the GPU context. I would like to know the reason for not considering Okada as model for the generation process.

A8. Perhaps presenting our model as a three-step simulation program might have been misleading. The goal is our model resides on propagation and inundation. Porting to CUDA does not constitute any constraint. The fault generation and theory used is based on RIMES internal decision. Their fault parameters are submitted to our model directly in order to start the simulation.

In future work we will focus on studying and including different fault theories such as the Okada model or dynamic generation.

Section 4

Q9. In section 4, authors use a tree-based mesh refinement similar to the AMR technic used by R.J. LeVeque in GeoClaw. In this case the refinement is not made automatically as in GeoClaw, but it’s customized to be refined in coastal areas and focal areas. It’s an interesting alternative to the use of nested meshes when more detailed information is required in certain coastal areas.

In p14, line 9-10 it is discussed the number of blocks necessary under the considered resolution and according the 65 x 65 node-centered cells if the number of blocks is 230,000 you would have 971,750,000 node-centered cells. How many information is stored on each node-centered cell (in double precision) to represent over 100GB of memory space?

A9. For that test case, each cell stores approximately 112 bytes in double precision. This includes 14 different values to store information about latitude, longitude, the governing variables \( h^t, hu^t, hv^t \), the next time-step variables \( h^{t+1}, hu^{t+1}, hv^{t+1} \), the constant \( H \), the bathymetry \( z \), manning coefficient and three constant values used for optimization.

Later, the introduction of focal areas reduced the domain mesh size and increased the available memory. Using this freed memory, three more constants were stored per cell. Additionally, the inundation output blocks were stored in GPU memory during the simulation for optimization.

Section 5

Q10. I have some comments, but my main concern is that I think that this section is out of the scope of this journal and my recommendation would be to publish it in a journal more related to this field. Anyway, I will make some comments regarding this section. To my knowledge, the overall GPU implementation has been solved in a very efficient way, particularly the implementation of the pipe asynchronous output that is crucial to deal with the GPU-CPU traditional bottleneck.
A10. We appreciate your comment and suggestion however, while we understand that the GPU computing is not the scope of this journal we think that including the GPU implementation is very important since is a key element of our research. Tsunami forecasting requires a fast result and there lies the relevance of implementing our GPU calculation. For this reason, we give a brief description about GPU computing and a general explanation of our CUDA implementation and optimizations.

Q11. In p23, line 8 it is showed that you use CFL 0.8, is not stable the implemented model for CLF nearer to 1? On other way, in p24, line 5 it is used $\Delta t = 1.6$ for blocks with levels over 3. Is this consistent with the CFL condition? Or, can you ensure the stability of the numerical scheme under this assumption?

A11. To answer the first part of the question, in general it is stable to use a CFL condition closer to 1. The Semi-Lagrangian scheme used for the propagation stage allows a large time-step. Additionally, the added inundation benchmark problems (see Question 7) use CFL = 0.9, producing good results while being stable.

CFL values closer to 1 tend to produce stable results if the bathymetry varies smoothly. Based on our experience, some instabilities may arise in cases where the bathymetry presents sudden large gradients or very irregular shapes. The Indian Ocean bathymetry used in our research contain several of these cases. A common solution is to smooth the bathymetry before usage. Clearly, this introduces changes in the measurements and simulation. However, in order to keep the results as realistic as possible, we decided not to smooth the bathymetry and instead trade off a higher CFL value.

About the second part of the question, the CFL condition is consistent when using $\Delta t = 1.6$ and numerical stability is ensured in this situation. In order to illustrate this, Table E contains four columns with the values used for the simulation. The second column shows the maximum $\Delta t$ allowed in each level using CFL = 0.8. In order to speed up the computation, a sub-cycling method was introduced. A global $\Delta t = 1.6$ s is set to calculate the number of cycles that blocks in each level must take. The value of 1.6 is chosen to avoid a sub-cycling overhead since around 80% of the blocks are distributed in levels 1 to 4. The third column in Table E shows the resulting number of sub-cycles per level ($ns$). Finally, the fourth column shows the new CFL values obtained for each level when using sub-cycling ($S.C. CFL$). As it can be seen, in all levels, the CFL condition is less than 1 and stability is ensured.

We included Table E in the manuscript with an explanation to make clearer that stability is ensured when using sub-cycling.
<table>
<thead>
<tr>
<th>Level</th>
<th>Max $\Delta t$ (CFL=0.8)</th>
<th>ns ($\Delta t = 1.6$)</th>
<th>S.C. CFL</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>10.71</td>
<td>1</td>
<td>0.12</td>
</tr>
<tr>
<td>L2</td>
<td>5.13</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>L3</td>
<td>2.37</td>
<td>1</td>
<td>0.54</td>
</tr>
<tr>
<td>L4</td>
<td>1.65</td>
<td>1</td>
<td>0.78</td>
</tr>
<tr>
<td>L5</td>
<td>0.95</td>
<td>2</td>
<td>0.68</td>
</tr>
<tr>
<td>L6</td>
<td>0.55</td>
<td>4</td>
<td>0.59</td>
</tr>
<tr>
<td>L7</td>
<td>0.26</td>
<td>8</td>
<td>0.39</td>
</tr>
</tbody>
</table>

Table E. CFL values used after introducing sub-cycling (S.C. CFL) for each of the seven levels.

The second column shows the maximum $\Delta t$ per level using CFL = 0.8 and the third column shows the number of sub-cycles (ns) required in each level when using $\Delta t = 1.6$.

Q12. In section 5.3.2 it is showed the runtime performance by comparing results obtained with two different GPU architectures: Tesla K80 and Tesla P100. It must be remarked that the configuration of the computation nodes that you are using is different for each architecture. While the Tesla P100 nodes have 4 GPU’s per node, the Tesla K80 nodes have 2 GPU’s per node. Unless your tests are using only two P100 per node, the network communications will reduce the performance between the K80 in front of P100. What kind of network is being used in these clusters?

A12. In the case of Tsubame 3.0, there are four Tesla P100 GPUs per node and the network is Intel Omni-Path HFI 100Gbps. In the case of the K80 machine used, there are four cards in one node (eight GPU in total), connected through PCI-Express 3.0. Confusion might arise from the term “Tesla K80 (12GB×2)”. The values in parenthesis refer to the memory distribution. One K80 card includes two GPU chips, each with 12 GB. We modified the manuscript to describe this machine specification more clearly.

We agree that the performance might be different depending on the GPU/node distribution. The aim on comparing between these machines is to demonstrate the portability of our code from an older architecture to a much newer one. Additionally, it serves to show a more general performance results by using more than one more machine and to highlight that the code can be implemented on different hardware without requiring changes or creating any problem.

Q13. Another point is, are output data being stored into hard disks in these tests or are they related only to pure computation time?

A13. All runtime measurements include the output time storing into hard disks.
Section 6

Q14. Finally, my main comments about section 6 have been included in my recommendations for section 3.

A14. Once again, we thank the reviewer for their important recommendation. We followed the advice given about section 3 and the full answer including the results of the new benchmark problems can be consulted in Question #7 (Q7).

Q15. Anyway, I have some specific comments about this section. It would be nice if the color scale used in Fig. 18 and 19 are the same for each subplot. With these graphics we can compare the inundation extensions, but it’s difficult to compare the inundation height when different scales are used on each graphic. On other hand, and as I pointed before, this test is not enough to say that the model is validated, so I don’t agree with the sentence of p33, lines 8-9.

A15.

Fig X. Inundation comparison for Hambantota, Sri Lanka. Top: RIMES model. Bottom: TRITON-G model.
The subplots in Fig. 18 and 19 have been redone to match better the scales for comparison. The new subplots are shown in Fig. X and Fig. Y.

We have modified the sentences in p33 about the inundation validation using just one test. It is now noted that the additional standard inundation benchmark problems computed produced good results and served as complementary demonstration of TRITON-G’s ability to estimate tsunami inundation.

Fig Y. Kamala (North) and Patong (South) inundation maps comparison. Top: inundation result by (Supparsri et al., 2011). Bottom: TRITON-G inundation result.
References


Corollary

Author’s comment: Additional to all the reviewers’ suggestions, we decided to remove the paragraph about the circular shoal benchmark in page 26, from line 11 to 17. With the introduction of several new benchmark problems (Reviewer #1 Question #7) and the modifications to the original manuscript, it felt unnecessary to keep this reference since the new results covered far more than what this benchmark offered.