Reviewer 1

We are very grateful to the reviewer for his/her constructive critiques and comments. In the following, we state the referee's comments (in blue) followed by the response and actions taken (in black).

Add brief summaries to figure captions to clarify key observations, especially in Figures 8–9 and 16–17.

The captions are extended to include a summary of observations in Figures 8-9 for Tohoku 2011, 13-14 for Alaska 2018 and 16-17 for Tateyama 2009 events.

Four test cases are presented, highlighting different strengths of the methodology. While satisfactory agreement is observed for many DART buoy observations, some cases show larger deviations. How do the authors explain variations in model performance across different test cases? For example, were there consistent factors (like earthquake depth, distance from hydrophone) that influenced prediction accuracy? Could the authors include a summary table comparing key performance metrics (RMSE, computational time) to provide a clearer picture of strengths and limitations?

The following description has been added, along with two new tables and an extended table, to illustrate the model's sensitivity to the source and its variations across different locations.

Among the four case studies discussed in the paper, Sumatra was triggered by a large oblique-slip earthquake with a significant vertical component and prolonged duration, whereas Tohoku and Tateyama involved thrust fault movements. Tohoku was a high-magnitude, long-duration bottom-shaking event, while Tateyama was weaker and shorter in duration. In contrast, the Alaska case was characterised by a strike-slip fault, dominated by horizontal motion and moderately shorter duration compared to Sumatra and Tohoku. Despite its large magnitude, the horizontal motion in Alaska resulted in only a minor tsunami. The vertical ground motion played a critical role in tsunami generation for Sumatra, Tohoku, and Tateyama, whereas the horizontal motion in Alaska limited tsunami generation. Consequently, model performance depends heavily on earthquake magnitude and vertical motion, as defined by the dip angle, with better results observed for large, vertically dominant ground motions. Furthermore, the accuracy of model predictions improves when the gauges are closer to the hydrophones. The reason is that AGWs are less dissipated due to interactions with the seafloor geometry, allowing the *inverse model* to better capture and estimate the fault geometry. (see Table 1).
From an observational perspective, ground-truth data for the Sumatra case are limited to a few selected locations, as summarized in Table 2, while DART buoy observations were available for the Tateyama, Tohoku, and Alaska cases, as outlined in Table 3. The accuracy of the model at observation locations is further influenced by two key factors. The first is the ratio of the shortest distance to the direct distance (SD/DD) between the epicentre and the observation points; a ratio closer to 1 indicates wave propagation over relatively consistent depths, aligning well with the assumptions of the *direct model*. The second is the proximity of the observations to the source, as observations closer to the epicentre, reflected

Table 1: Summary table for 4 case studies Ekström et al., 2012).

Case	**Sumatra**	**Tateyama**	**Tohoku**	**Alaska**
Date	26/12/2004	12/08/2009	11/03/2011	23/01/2018
Time (GMT)	01:01:09	22:48:55	05:47:32	09:32:04
Lon	94.26	140.68	143.05	-149.12
Lat	3.09	32.74	37.52	56.22
Moment Magnitude (Mw)	9	6.6	9.1	7.9
Depth [km]	28.6	55.2	20	33.6
Half Duration [s]	95	4.8	70	22.3
Strike [°]	329	55	203	257
Dip [°]	8	18	10	80
Slip [°]	110	130	88	4
Type	Oblique-slip	Thrust	Thrust	Strike-slip
Hydrophone	H08S1	H11N1	H11N1	H11N1
Lon	71.01	166.89	166.89	166.89
Lat	-6.34	19.71	19.71	19.71
Distance [km]	2786	3005	3039	5427
Acoustic Travel Time [s]	1856	2003	2026	3485

Table 2: Direct Distance (DD), ration between Shortest Distance to Direct Distance (SD/DD) and Travel Time (TT) for Sumatra 2004.

Location	**Lat**	**Lon**	**DD [km]**	**SD/DD**	**TT [hr]**
Madras Bandar	13.14	80.45	1885	1.08	3.0
Batticaloa	7.71	81.69	1483	1.03	2.2
S Maldives	-0.74	73.20	2379	1.06	3.5
Phuket	7.88	98.40	702	1.24	2.1
Banda Aceh	5.55	95.32	298	1.85	1.1

in shorter travel times, tend to show higher model accuracy.

Reference

Ekström, Göran, Meredith Nettles, and A. M. Dziewoński. "The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes." Physics of the Earth and Planetary Interiors 200 (2012): 1-9.

Table 3: DART buoy stations legend: Direct Distance (DD), ration between Shortest Distance to Direct Distance (SD/DD) and Travel Time (TT) for Tateyama 2009, Tohoku 2011 and Alaska 2018 events.

				Tateyama			**Tohoku**			**Alaska**		
Index	**DART**	**Lat**	**Lon**	**DD [km]**	**SD/DD**	**TT [hr]**	**DD [km]**	**SD/DD**	**TT [hr]**	**DD [km]**	**SD/DD**	**TT [hr]**
1	21418	38.71	148.67	992	1.11	1.5	509	1.08	0.7	4867	1.07	6.2
2	21413	30.55	152.12	1137	1.09	1.7	1139	1.07	1.4	5317	1.07	6.7
3	52404	20.94	132.31	1544	1.01	2.4	2115	1.10	2.9	7385	1.06	9.4
4	21419	44.46	155.74	1851	1.08	2.6	1312	1.05	1.7	4003	1.09	5.2
5	52401	19.29	155.77	2145	1.10	3.0	2373	1.06	3.0	6097	1.05	7.7
6	21416	48.04	163.49	2567	1.04	3.4	2027	1.05	2.5	3287	1.13	4.4
7	52405	12.88	132.33	2364	1.06	3.4	2939	1.07	3.9	8110	1.06	10.3
8	52402	11.58	154.59	2771	1.04	3.7	3106	1.02	3.8	6885	1.06	8.7
9	21415	50.17	171.84	3214	1.05	4.3	2678	1.07	3.4	2645	1.13	3.6
10	21414	48.94	178.27	3602	1.05	4.7	3088	1.04	3.8	2315	1.11	3.1
11	52403	4.03	145.60	3247	1.13	4.8	3733	1.09	5.0	8120	1.06	10.5
12	46413	48.67	-174.59	4107	1.04	5.3	3602	1.04	4.5	1896	1.10	2.5
13	46408	49.63	-169.87	4460	1.05	5.8	3950	1.05	4.9	1554	1.06	2.1
14	46402	50.44	-165.02	4812	1.05	6.2	4297	1.04	5.3	1219	1.07	1.7
15	46403	52.65	-156.93	5374	1.06	6.9	4844	1.05	6.1	626	1.04	0.9
16	52406	-5.33	165.08	4979	1.07	7.0	5283	1.05	7.0	7990	1.04	10.5
17	46409	55.30	-148.50	5904	1.07	7.7	5359	1.07	6.8	89	1.05	0.1
18	46410	57.50	-144.00	6148	1.09	8.2	5594	1.09	7.4	356	1.20	0.6
19	51407	19.63	-156.51	6371	1.05	8.3	6119	1.06	7.9	4091	1.09	5.6
20	55012	-15.80	158.50	5739	1.07	8.4	6145	1.07	8.6	9350	1.08	12.7
21	51425	-9.50	-176.25	6574	1.08	8.7	6726	1.05	8.4	7701	1.04	9.9
22	46419	48.76	-129.62	7332	1.04	9.8	6801	1.03	8.9	1544	1.02	2.4
23	46404	45.86	-128.78	7521	1.04	10.0	7000	1.04	9.1	1809	1.02	2.7
24	46407	42.60	-128.90	7666	1.05	10.1	7161	1.04	9.3	2078	1.06	3.1
25	46411	39.35	-127.01	7971	1.06	10.5	7477	1.04	9.7	2464	1.04	3.6
26	46412	32.25	-120.70	8860	1.05	11.5	8385	1.05	10.7	3435	1.05	5.0
27	55023	-14.80	153.59	5478	1.08	8.0	5921	1.07	8.2	9493	1.10	13.3
28	56003	-15.02	117.99	5829	1.06	8.9	6403	1.07	9.4	11563	1.06	15.9

Reviewer Comments #2, and responses

1 General Summary:

This study presents GREAT v1.0, a new tsunami early warning system that utilizes the analysis of acoustic signals generated by earthquakes under the ocean. The approach considers the fact that acoustic waves travel much faster than tsunami waves, allowing instantaneous assessment of tsunami hazard. The system integrates several state-of-the-art models, spanning wave path modeling to machine learning, direct tsunami amplitude inference, and inverse problem solution, to make rapid and reliable forecasts.

The study is very interesting, well structured and provides concise insight into the building blocks of the model, validation procedures, and potential applications. Certain areas may need minor clarification to further advance the paper and ease the transition to operational utilization.

We are very grateful to the reviewer for his/her constructive critiques and comments. In the following, we state the referee's comments (in blue) followed by the response and actions taken (in black).

2 Minor Comments

2.1 Machine Learning Dataset Expansion

The authors acknowledge the current limitations of the dataset. It would be helpful to learn more about their plans to expand it, e.g., if they anticipate adding data from GPS buoys or regional seismic-acoustic networks. Mentioning such details could reflect both the feasibility and timeline for expanding the dataset.

Response to comment

Each component of the GREAT software has a different sensitivity to additional data. For example, incorporating new tsunami measurements, such as GPS buoys, tide gauges, or satellite altimeters, into the current version of the software—where they are primarily used for validation—can enhance confidence in model reliability across various geographical locations (offshore, nearshore, and at varying distances from the tsunami source).
In the next version of the model, the machine learning (ML) component will

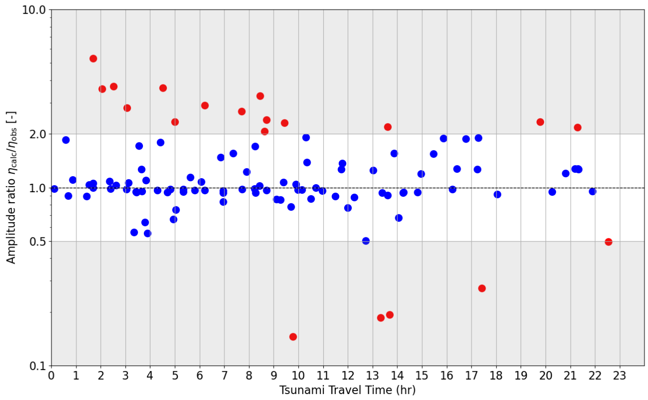

Figure 1: Amplitude ratio against tsunami travel time for Tateyama 2009, Tohoku 2011 and Alaska 2018 study cases at various DART buoy locations with tsunami travel time up to 24hr.

be expanded to utilize these data as training datasets. This shift would alter their role from validation datasets to critical inputs, improving the model's predictive capabilities.

Regarding acoustic datasets beyond the sparse CTBTO hydrophone data, there are two key considerations. First, increasing the number of datasets would enhance response time for faster warnings and provide multiple datasets per event, improving confidence in detection and analysis. Second, we are currently testing alternative sources, such as ONC hydrophones, which introduce challenges related to variations in data format, accuracy, and frequency range. Addressing these differences requires careful consideration to ensure proper integration and account for potential observational errors.

2.2 Far-Field and Land-Separated Prediction Differences

The model seems to be most accurate near the earthquake epicenter but less so at distant locations or at locations separated by land masses. Would the refinement of bathymetric data or the inclusion of more sophisticated coastal models enhance these discrepancies?

Response to comment

The accuracy of the model is assessed using DART data. A challenge with DART buoys is their hybrid sampling rate, which is too low [$\Delta t = 15$ min] under normal conditions and only increases [$\Delta t = 1$ min, 15 s] if triggered above a certain threshold. Typically, at these locations, the DART buoys are not triggered, resulting in a low sampling rate and data dominated by irrelevant noise.

Another factor is that when amplitudes are too small, the uncertainty is too high. However, in such cases, a tsunami threat does not exist, making it less relevant for real-time analysis. We have revised the results section to include a threshold of 0.05 m to minimize noise and exclude excessively small amplitudes (see Figure 1, which we added in the discussion section).

2.3 Minimum Hydrophone Density for Effective Detection

An order-of-magnitude estimate of the minimum hydrophone station density that would be needed to reliably detect and characterize near-field tsunamis in high-risk areas would be beneficial. This would guide sensor deployment planning in the future.

Response to comment

One significant challenge facing this emerging technology is the limited number of available hydroacoustic stations. Specifically, the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) operates six hydrophone stations worldwide, from which we have access to four stations. Moreover, the geographic distribution of these hydrophones limits the technology's applicability to specific regions. For seismic source tsunamis, the technology is most effective within a 1,000 km radius of each station - which allows an end-to-end assessment within an average of less than six minutes. Employing these figures as an indicator for an optimised global hydrophone station density, would require roughly 30 hydrophone stations.

2.4 CTBTO Hydrophone Network Configuration

As the system relies heavily on the CTBTO network (initially designed for nuclear monitoring), have the authors addressed whether its current density and position are ideally suited for tsunami detection? Would the supplementation of sensors in high-risk regions enhance performance, especially for smaller or maybe more remote events?

Response to comment

See response above.

2.5 Operational Reliability and Everyday Use

Whereas computational efficiency is commendable, greater insight into actual-world performance beneath operating conditions will be helpful. This might include discussion of potential hardware limitations, data transmission time delays, or even sensor failure, and the way these are addressed.

Response to comment

Since its deployment at the Instituto Português do Mar e da Atmosfera (IPMA) in June 2024, our tsunami warning technology has been subjected to real-time operational testing. This phase aims to assess the system's performance under actual operating conditions, addressing challenges such as hardware limitations, data transmission delays, and potential sensor failures. A comprehensive analysis of these factors is underway, with findings to be published upon the study's conclusion.

2.6 Model Integration and Error Propagation

GREAT v1.0 is made up of a number of sub-models (fault geometry estimation, wave speed calculation, etc.). How do the authors think that the tiny errors in one component may or may not be magnified and lead to erroneous tsunami predictions in another? Have they performed an uncertainty analysis to quantify and minimize these risks? The addition of surrogates of some components might be useful in carrying out a sensitivity or uncertainty analysis economically. This would allow the investigation of situations of error propagation without excessive computational cost.

Response to comment

- Given that the analytical solutions are linear, small changes in input properties do not result in large deviations, making error magnification unlikely;

- The machine learning (ML) model operates independently from the analytical model. Thus, a strong match between the two models increases confidence in the assessment. Since they are complementary and independent, they can be treated as ensemble members for probabilistic analysis;

- we have DART buoys integrated in the system which provides another independent way to assess the analysis (in the case the data is available in real-time) - integrating more real-time data can provide an additional layer of validation of the results;

- In the worst case scenario where there is no convergence among the models, or the results don't seem to be reasonable, since the technology is complementary, at that stage the traditional (conservative) approach can be employed.

Overall, error propagation between components can be mitigated by introducing limiters and thresholds to prevent unrealistic estimates.

GREAT v1.0 is an organized, valuable and promising tsunami warning system. Dataset increase, spacing of the sensors, accuracy of far-field forecasts, and reliability of operation—and uncertainty analysis—are minor issues that will further establish its practical usefulness. Transparency on these aspects will allow easier transition to operational use.

Citation: https://doi.org/10.5194/gmd-2024-139-RC2