Comments of reviewer 2 - our replies are in red

Reviewer report for 'Inundation and evacuation of shoreline populations during landslide-triggered tsunami: An integrated numerical and statistical hazard assessment' by Emmie M. Bonilauri et al.

General Comments

This paper studies a very challenging problem, that of whether and how residents and visitors to Stromboli can be evacuated to safety in time after a landslide on the flanks of the volcano causes a tsunami. On the whole this is a good-quality paper, presenting modelling of landslide-caused tsunamis at the volcano, methods for their detection and quantification, and estimates of which of the areas at risk can potentially be evacuated in time.

The biggest problem I found in reading this paper was with understanding Section 3.4 on Evacuation Capacity, where it was difficult for me to interpret as some of the key metrics were not clearly enough defined, which made it difficult to understand some of the figures and to confirm some of the stated implications. I will put more specific comments below, but I would very much like to see this section improved as it is important for the paper as a whole.

I also feel that some of the assumptions that have gone in to the evacuation modelling are rather optimistic, for example the assumption of no reaction time. The discussion section makes it clear that the authors are aware of many of these issues, but some further elaboration may be useful.

Thanks for these relevant comments. We have tried to improve our explanations and added some further thoughts on the issue of the evacuation time at the end of the discussion (Section 4.3, lines 470 – 473), as detailed below in our replies to your specific comments.

Specific Comments

Landslide modelling

A challenge with landslide modelling is the wide range of possible initial parameters. The paper considers variations in three landslide parameters: the position upslope, the volume, and the density. The authors mention that off-axis landslide position and rheology are parameters for future study.

However lines 333-334 could perhaps be revised to be more clear that the volume and position were found to be the most important parameters of those considered *in this study*. As it is not clear from this study alone that all other parameters are 'of second order'.

We clarified the text as "In this study, landslide position and volume are considered the key parameters in determining the hazard score (Cerminara et al., 2024)". This has been added in Section 4.1, lines 399 – 400, with adding the reference of the INGV database recently published.

The result that it is landslides from the lowest subaerial position (position 3) caused the highest impact is interesting and a bit surprising. The authors assert that this is related to the landslide not having time to deform before reaching the sea surface. This is a plausible explanation to me, although it would be nice to see this demonstrated with a figure or reference. Since the rate of deformation is related to the rheology, it also makes me question whether the effect of rheology is truly 'second order' (see previous paragraph).

Following the comments made for Reviewer 1, who also raised this interesting point, this is now addressed by adding a short discussion (Section 4.1, lines 390 – 398) and two new appendices (Appendix D and E).

Analysis of Signals

The use of LASSO penalised linear regression to improve the prediction of inundation is interesting and could be a useful tool in many circumstances. The explanation of the method in the main text and the appendix was quite brief, and the supplied reference Giraud (2021) also rather hard to follow, so any additional explanation of how the method was applied to this problem would be welcome.

In a real case, the volume, density or position of the landslide is not immediately known, thus we only have the signals of the two gauges. As a result, the aim of LASSO is to find out whether there is any chance of detecting the impact of tsunamis on coastlines both quickly and accurately (without counterproductive false alarms) before the tsunami arrives and without knowledge of the landslide characteristics. We thus determine how we can use simulated wave signals to correctly determine the impact score. Consequently, our aim is not to analyse the shape of the waveforms as a function of the physical characteristics of the landslide.

To set the impact score, we used the signals from the 2 beacons, i.e., 40 variables (1 wave height every 2 seconds between 0 and 40 seconds for 2 beacons). This is now explained in the text.

The classic method for this type of problem (that is, to find an unknown value from several knowns) is linear regression. This approach, however, lacks robustness when the number of explanatory variables (here, 40) is too large compared with the number of individuals (here, the 156 simulations). LASSO linear regression is thus a modification of traditional linear regression that identifies a subset of explanatory variables (in this case, times of interest for measuring wave heights) of sufficiently small size for the results to be robust. This has been clarified through addition of text to the methods (Section 2.2, lines 133 – 143) and results sections (Section 3.3, lines 296 – 300), as also recommended by reviewer 1.

My main concern here is how robust the LASSO regression algorithm would be in scenarios that differ in one way or another from those on which it has been trained. For example: landslides that occur off axis, or have different rheologies to those assumed; or how the algorithm would work if there was not a singular landslide but one quickly followed by another (as in 2002 but with a shorter gap) such that there was a superposition of waves.

Our LASSO method is robust in terms of ability to adapt to landslides with volumes of between 5 and 30 million m^3, densities of between 1.7 and 2.5 kg/m^3 and landslide source positions of between 500 m and - 584 m. That is, within the limits of our simulations. This has been added in Section 3.3, lines 267 – 270.

As argued in a recent paper by Ripepe and Lacanna (2024) the waveform did not change with the slip volume and that it was possible to determine the inundation extent/run-up from the tsunami amplitude, which we also show here. See Section 3.3, lines 270 – 279, for more details.

Another question is how to determine time t=0 instrumentally from the water-level data, and hence when to extract data to use in the algorithm. There is also a potentially longer wait to collect all of the datapoints which may limit the use in some near-source cases. To study all of these things would require another paper, for now I just suggest the authors consider a bit more discussion.

Currently LASSO needs 40 seconds to recognize and classify the waveform. This time was a result of a payoff between precision in inundation area and time needed to classify the waveform to an acceptable degree of accuracy. We could reduce the time, but that would make output decreasingly accurate. This means that for extreme proximal locations output will be delivered after wave arrival, but they will have been alerted by the siren. Discussed now in Section 4.3, lines 459 – 462.

In Appendix A, I was confused for a while by the way that 'X' was used both for the inundation of cells and for the tsunami detector time series. Maybe use a different letter for the inundations and clarify how 'Y' is calculated from it?

X is a vector, and Y is calculated as a function of the various vectors. The impact score is first calculated from the inundation models, and then we find the impact score for the signal associated with each inundated pixel with vector X. Thus, X is the same for both the inundation of cells and the detector time series.

Evacuation Capacity

I found Section 3.4 quite hard to follow, mostly because the concept of 'warning time' (both 'real' and 'needed') was introduced without really clear definitions. With some effort I could establish what I think these are, but I think it would be better to spell this out more explicitly. Similarly it would be good to be fully clear about what the maximum and minimum of these times were calculated.

Defined and clarified in the beginning of Section 3.4, lines 308 – 314. We define "warning time needed" and "real warning time". The former is the time needed to move from any given point to a safe point, this being the exit from the inundation zone Refuge Area Entry Point (RAEP). Instead, the "real time", is the time available for escape prior to wave arrival. Also, instead of using "evacuable" or "non-evacuable" we use "inside threshold time" and "outside threshold time" to avoid appearing to make any statement regarding evacuation.

Based on my interpretation (which could be wrong) I found figure 11 a bit difficult to understand. As by my understanding, the cells that need the most warning time (hollow bars) are unrelated (or even inversely related) to those which need the most warning time (solid bars), yet because they appear next to each other in the figures I initially thought that they would be the same locations.

The figure is designed to portray the cases where we lack the time to evacuate a point as a function of the time of the day and/or zone. This is now clarified in the figure caption (Section 3.4, lines 359 – 360).

Pedestrian Evacuation Model

The evacuation model used is relatively simple, and the most important thing for this paper is to make sure that the simplifications and approximations are well documented and explained. Simplifications I'm aware of (some of which are mentioned in the text) include:

- Assumption of no reaction time
- Assumption of no warning dissemination time
- Assumption of no variation in walking speed
- Assumption of no congestion

While some of these could be approximated by extending the current model, ultimately this problem is really calling out for a full agent-based modelling approach (though I'm not suggesting that the authors need to do that for this paper).

We are currently testing such problems using on-site escape tests using multiple agents faced with differing traffic flow scenario's, starting positions (beach, water, bed), and degree of preparation (footware, reaction time). This will enable us to better calibrate our escape times, and to produce a distribution of potential times depending on traffic and route conditions. We now address this issue at end of the discussion (Section 4.3, lines 470 – 473) and as also following the advice of reviewer 1.

Summary

Although I have made some critical comments above, in all I find this to be a valuable and well-written multi-disciplinary paper, certainly worthy of publication after some minor adjustments.