Response point-by-point to Anonymous Referee #1

The point-by-point answers are in blue color, below each Reviewer's comment (reported in *Italic*).

 # # Overall comments #

This looks to be a good paper about an important topic that is of clear interest to NHESS readers. It is mostly well written, and describes innovative ideas which are likely to be of broad utility in tsunami hazard assessment.

- ➢ *My only 'significant' concern is that the authors do not provide a 'conceptual justification' for the differences in the results of the two filtering methods they apply, for the case of high H_{max}. As it stands, as a reader I don't know why this happens. Intuitively the reasons are not obvious, and as it relies on some rather delicate calculations (which we know are sensitive to choices of coefficients in filters, etc). At the moment I cannot be confident that the results are 'stable' enough to justify the conclusion that "it is important to distinguish near and far-field sources in the filtering approach". If the authors can provide some 'conceptual backing' to support these results, then in my judgement the paper should clearly be accepted for publication in NHESS. In saying this, please note that I accept the fact that some aspects of these filtering approaches cannot be completely stable (e.g. in the authors example, results with H_{max} < 1m are not meaningful). This is expected, and not a problem. However, they need to provide more justification that the results at higher return periods are stable enough to justify the key conclusion.*

(the answer below is the same for a similar question from Reviewer 2)

The conceptual explanation traces back to the fact that the two procedures are not equivalent from a physical point of view and we could roughly say that one is in principle "correct" and the other one is "wrong". Maybe in saying "it is important to distinguish near and far-field sources in the filtering approach" we were not clear enough. What we wanted to stress is that a blind filtering procedure based on offshore tsunami amplitudes produces a non representative selection of the important scenarios, as it could aggregate or even remove important local scenarios.

We try to explain it better below.

In the original procedure by Lorito et al., offshore tsunami amplitudes are supposed to be representative of the coastal inundation, regardless of the source location with respect to the coast. That was reasonable, since it considered either far field scenarios with respect to the coast of Sicily, or scenarios which deformed the coast of Crete Island always in the same direction, since they were all subduction earthquake on the neary Hellenic Arc.

Indeed, offshore tsunami profiles could be strongly misleading when coseismic deformation of the coast occurs, either as coastal uplift or subsidence depending on the causative earthquake. The coseismic displacement induced by local earthquakes can modify the actual onshore tsunami intensity corresponding to the same offshore wave. Hence, near field scenarios must be separately treated, and clustered considering the source similarities, including the co-seismic coastal displacement, rather than the offshore tsunami wave similarity.

We will try to report these “conceptual” arguments in the revised manuscript as concisely as possible.

The tuning of the thresholds in the filtering procedure is a different task, but we note that the same thresholds have been used with and without the correction for near field, so that the differences we found in the results obtained from the two procedures are not in our opinion imputable to those choices.

On the other hand, we can now support such conceptual justification providing the physical explanation of the specific results, based on the new quantity MU (mean uplift) we calculated and described in our introductive general remarks. This also answers to one of your specific comments below.

In general, lower ‘corrected’ hazard means that the predominant effect by local sources contributing to a specific point on the hazard curve - that is to the probability of exceedance for a given intensity threshold - is represented by coastal uplift, which in turn decreases tsunami hazard. In other words, there is a prevalence of clusters represented by scenarios causing uplift. Conversely, higher hazard would correspond to coastal subsidence.

As we said, we investigated this aspect, computing, for different intensity thresholds above 1m, the MU on a random point along the coastline of the inner grid, produced by near field representative scenarios contributing to the hazard at that threshold, weighted by the occurrence probability associated to each scenario (corresponding to the probability of the entire cluster it represents) and normalized to the probability of all of the scenarios contributing to the same intensity threshold.

The obtained positive values, although not representative of the real coastal displacement as averaged on all the scenarios (including that ones which do not produce appreciable coseismic local deformation), indicate that the dominant contribution to the coseismic deformation is an uplift of the coast, in agreement with the percentage differences retrieved between the two approaches.

We hope to have answered in this way to the “significant” concern expressed by the Reviewer. We must acknowledge that this comment made us deepen the analysis and consider our results much more carefully - and indeed we found a bug.

 # # Specific comments #

- ➢ *P7, around line 10: I think you mean that you neglect a bunch of 'other' important sources of uncertainty, but, you do comprehensively test the filtering procedure (??right?? – actually upon reading the full paper I'm still uncertain). At the moment the paragraph doesn't make it clear if your example is actually a 'strong' evaluation of the filtering procedure, given the idealized assumptions on the source. Please make this clearer.*

The test site illustrative application is not a real hazard assessment, as it is based on a quite rough probability model as well as on some strong assumptions regarding the filtering thresholds and the source modeling.

Nevertheless, although relatively simplified, our source model is still quite complex, and includes even epistemic uncertainties on many source parameters, e.g concerning the seismic rates, the shape of the magnitude-frequency distribution, even the seismogenic depth for the two considered subduction zones, and several others. It also includes ensemble uncertainty modeling. We now include a new Figure in the Supplementary Materials (Figure S2), which should make clearer that the model deals with epistemic uncertainty, as it shows the comparison between the mean offshore hazard curves at selected points along the 50m isobath (see Figure 2a of the manuscript), as well as the comparison between some quantiles of the epistemic uncertainty, for the filtered and original set of scenarios. Please refer to Selva et al. 2016 for further details on the adopted source model.

So, we consider the model as fully suitable to test and describe the procedure. We anyway restate that the aim of the application is to highlight that inaccurate (biased) evaluation of site-specific tsunami hazard would be obtained if scenarios located in the near field of the target area are not properly taken into account, irrespectively of the completeness and consequent complexity of the hazard assessment. A "real" application would be just more complicated and more computationally demanding.

- ➢ P8, top of page – it would be good to report on some sensitivity analysis of this to give the reader a 'feel' for how severe these approximations are (e.g. you could halve the number of clusters, so you don't have to do more simulations).

The filtering procedure surely introduces some approximations and ideally the goal should be to reduce the computational cost of PTHA while keeping the error with respect to the whole set of sources as small as possible. In the present work, considering the illustrative nature of the case study, we enlarged the accepted error to further reduce the number of explicit numerical simulations.

First of all, the most severe approximation was made during the filtering on tsunami amplitude: it goes without saying that a threshold of 1m might be not acceptable in case of a real hazard assessment, while it is an acceptable threshold for illustrative purposes. It is worth stressing that this filter, independently from the threshold value, does not affect subsequent steps of the procedure, as it represents a rigid cut-off of the number of scenarios we are accounting for.

Another strong assumption was made regarding the cluster analysis: the k-medoids partitioning algorithm is based on the minimization of the sum of the intra-cluster distances, i.e. the distances between each element of a cluster and the cluster centroid. Strong constraints on the distances result in a more accurate partitioning, in terms of similarity between the elements of each cluster, but lead to a great number of clusters. Instead, larger ranges of acceptability increase the efficiency of the algorithm, in terms of number of resulting clusters, to the detriment of the accuracy.

As an example, we provide here a sensitivity analysis on the threshold imposed on the intra-cluster variance (step (3a)): the new Figure S3 shows the relative differences in absolute value between the offshore (i.e. at the control points along the 50m isobath) hazard curves computed from the complete initial set of sources and the filtered set (at the end of the cluster analysis). The red box corresponds to the threshold value we chose (0.2): it appears evident that a smaller value would have allowed a stronger constraint on the error introduced by the cluster analysis, while considerably increasing the number of resulting clusters. Vice versa, higher thresholds produce a smaller number of clusters, but fail in reproducing the hazard (error up to 40%). Our choice in our opinion represented the best trade off for our purposes. Again, in case of a real hazard assessment, the lower threshold would be likely better.

- *P8, bottom of page – 'it is worth noting that results at H_{max} < 1m ...' – OK, but because those results are not meaningful, can you please 'clip' your figure limits so that they do not include H_{max} < 1m. That will help the reader focus on parts of the curve that you do consider meaningful, and ease the interpretation of the figures.*

We apologise as we have misspoken: what we intended is that the hazard curves below 1m can be (negatively) biased since they are depleted from the scenarios removed by Filter H. Following your suggestion, we rephrased and shadowed that part of the plots in Figure 3. This depletion is also clearly observed in new Figure S2 for low amplitudes.

- *P9, paragraph around lines 10-15 – It's not evident to me why method 3a should 'over-estimate' rates for H_{max}>3 (or indeed why the difference is reversed at lower H_{max}). Can you give a heuristic explanation of why this could happen? Without some idea of this, my thinking is 'maybe a calculation/convergence type error' (!). Or is it that, for large enough H_{max}, the associated local sources have a greater tendency to be filtered than the distant ones, for some reason – and the converse for smaller H_{max}? Definitely not obvious to me – please discuss it.*

As discussed before, a lower hazard at a certain point of the hazard curve, due to the near field correction, means that the coseismic field from local sources that dominate the hazard produces a coastal uplift. We agree this was not sufficiently proved before, but the new Figure (3c), with the MU superimposed to the percentage differences between the two approaches now should better illustrate that the overestimation is correlated to the dominant coastal uplift.

- *P9, paragraph around line 5 – I agree that you've shown that a 'blind' cluster analysis might produce quite different results from the 2-stage approach proposed in the paper. However, I'm less confident about the stability of either procedure. Can you really say that the 2-stage approach is better, based on the results presented here? Consider the following "devil's advocate" theory – from what you've presented, I hypothesis that "Both of your approaches are strongly affected by the details of the filtering coefficients, and equally big differences could be expected from merely adjusting those in reasonable ranges". In other words, how can readers be confident that the results are not just 'noise'? Probably you can justify this, but I don't see it from the current text. So please add in some discussion that explains 'why' these results happen, and why you expect them to be 'basically robust' {notwithstanding that you have to make some severe approximations for low events – that's ok – but at least for high events, we need some conceptual explanation of the results}.*

As it should be clear now from our previous answers, involving the new Figure 3 and MU, there are firm conceptual reasons supporting the need of a "2-stage approach". You are indeed right that the results are stronger for larger amplitudes. This is clearer with the new results.

The simplest explanation remains though the same: the original assumption that offshore tsunami amplitudes are representative of the coastal inundation may fail if local sources producing appreciable coseismic deformation of the coast - of conflicting sign, i.e. some uplift and some subsidence depending on the source - are involved.

Hence, one (the only?) way we can take this into account is to separate near and far field scenarios and treat local sources removing the approximation introduced by the cluster analysis on the tsunami amplitudes, since offshore profile can not be considered reliable for such sources. This considerations hold irrespectively of the stability of the results in our example. However, our results "behave" as expected, being dependent on the approach used.

Indeed, the near field treatment is still an approximation, as we reduced the number of numerical simulations with respect to the "exact" case, by performing a cluster analysis based on the coseismic fields. However, it should be a better approximation with respect to aggregate local and remote scenarios on the basis of the offshore tsunami amplitudes.

Moreover, our application is also aimed to investigate if such procedure is really needed from the point of view of results, that is if, apart from the physical meaning of the procedure, results are actually affected by the near field correction.

The fact that the contribution from the near field turned out to be significant, even investigating a target site with relatively low near-field tsunamigenic seismicity, was not straightforward.

We finally stress that the two approaches (with or without the correction for near field) only differ in the way local sources are treated: the filtering coefficients are basically consistent. In other words, the different results can not be related to the filtering thresholds.

We repeat, the “conceptual explanation”, should now be there with the new results and the new analysis presented. And we must acknowledge that this comment was really useful.

- P11, line 6 – as mentioned above, please provide more ’conceptual explanation’ as to why this happens.

As extensively discussed in the previous answers, this is related to the coseismic deformation induced by local sources, which, if properly accounted, modify the effective tsunami hazard.

Detailed comments

- P3, L31 – suggest changing ’is released’ to ’is not used’.
- P5, L5: – suggest changing ’will produce as well similar inundation patterns’ to ’will also produce similar inundation patterns’.
- P5, lines 6-7 – Please provide the equation for the cost function. I looked up the 2010 paper, but it appears to refer to time-series comparisons rather than H-max comparisons. Better to make it very obvious to the reader.
- P6, around lines 10-11 – It’s not clear to me how you use the co-seismic deformation as a metric for source-proximity in the cluster analysis. Ahh, I see you do this below around lines 25. Give that, please add "(details below)" at the end of the sentence that finishes on line 11.
- P7, line 20 – there is a number with multiple ’.’ inside – this is not familiar notation to me, do you intend to use some other separator?

We thank for your suggestions, which will be all addressed in the revised manuscript. In particular, we will clarify that the cost function equation firstly introduced in the 2010 papers to compare time series while solving inverse problems was modified by Lorito et. al 2015 for Hmax comparison.