Response to the reviewers' comments to authors (second revision)

We thank the editor in charge for the opportunity to submit a new revision of our manuscript. We also thank the reviewers for providing feedback for improvement. We have responded in detail to each comment in the following pages, and we think that these amendments have significantly strengthened our paper.

We believe that the new version of our manuscript is a significantly improved one and more readable for broader audiences.

<u>**Reviewer # 1**</u>

We thank the reviewer for his/her comments. It follows detailed replies to each of them.

- Most of my concerns have been addressed in the revised manuscript. In the previous round, I suggested to check the validity of the results by conducting statistical analysis with different resolutions; however, the revised manuscript had not addressed this comment. For this response, authors explained that they could not conduct additional analysis due to their software specification, though, in my understanding, coarser values for additional analyses can be generated from current finer results. Since the comment concerns the effect of resolution to collect statistics from the simulations, in my view, the response regarding the validity of their simulation model cannot help to address this concern. Additional analysis with different resolutions is better to be included to ensure the validity of the obtained results. At least, discussion regarding the potential effect of the model resolutions is better to be presented.

During the first review, the reviewer pointed out that

- I understand that this resolution "4x4 m" is based on the finest resolution of the tsunami simulation; however, this resolution might too fine for counting tsunami casualties in agent-based simulations. The investigation with different resolution is needed to ensure the validity of the result. If consistent important features are found in different resolution, it supports the validity of the analysys method. Reliable coarser values can be generated by integrating finer values.

  A: To answer the reviewer's comment, we included a new SHAP values analysis with the same source information but combined in different resolutions, according to a geometric sequence with a factor of 2 between the size of each examined unit. To do this, we developed an algorithm that, starting from the 4x4 m cell (resolution x1) with the largest southern latitude and western longitude in each case study (i.e. located at the bottom left corner of the study area), grouped these basic units into five successive spatial partitions, each of them covering the complete evacuation move boundary. These partitions comprised cells 8x8 (resolution x2), 16x16 (resolution x4), 32x32 (resolution x8), 64x64 (resolution x16), and 128x128 m (resolution x32) wide, respectively. For each of these larger cells, the algorithm calculated the value of every independent variable as the average of the combined 4x4 m units. Then, we ran again the SHAP values analysis at these coarser resolutions. The results (summarized in the new Figure 8) show that the overall importance hierarchy of the independent variables remains unchanged through the different resolutions of analysis. Complementarily,

noticeable and disparate changes can be seen in the amount of their impacts on the predicted death ratio, if we compare the more and less accurate resolutions. We also discuss these methods and findings in the "Discussion" section. Page 18, line 371, page 23, line 441, page 26, line 530, and page 27, line 563.

Reviewer # 2

We thank the reviewer for his/her comments. It follows detailed replies to each of them.

- In line 232, it says that "Tsunamis were simulated for 45 min...". However, figure 3 shows that tsunamis in most of the study sites were simulated for 60 min, except Valparaiso-Vina del Mar. it should be coherent.

  A: This contradiction appeared in the manuscript because we first ran the flood models for 60 minutes (except in Valparaíso-Viña del Mar) and afterwards we used them as inputs for the agent-based simulations, which we ran for only 45 minutes, as this was sufficient to encompass the total evacuation time of each case study (which also allowed us to save computing time). We made this point clear in the new version of the manuscript. Page 10, line 204, and page 13, line 218.

- Figure 2. The scale of the vertical displacement is confusing. In one scenario, the vertical displacement of 7m is high, but in the other 4 m is high. Instead of using a colour scale with high or low, just use numbers. In addition, it could be convenient to use the same scale for all scenarios, thus they can be compared.

  A: We re-worked the figure accordingly. The new figure 2 (page 11) uses the same colour scale for all scenarios. Additionally, this scale includes numerical intervals.