We are grateful to the editor and the reviewers for their comments and suggestions and have made four main improvements to the manuscript:

1. We have elaborated on the description of the site-to-site term $\delta S2S$,

2. We have clarified that the resulting site amplification predictions should be interpreted in reference to the median prediction of the associated GMM,

3. We have simplified the section on Eastern Türkiye and included a comparison of the within-event residuals obtained using the GMM predictions and the four proxy-based site amplification models,

4. We have addressed all the minor comments and issues from the reviewers.

In the following we have addressed the comments and are describing the changes that have been done. The reviewer's comments are given in blue, and our replies in black.

2. Review comments for the paper entitled "Introducing inferred geomorphological sediment thickness as a new site proxy to predict ground-shaking amplification at regional scale. Application to Europe and Eastern Turkey » by Karina Loviknes et al. Natural Hazards and Earth System Sciences

The authors of this paper develops a new model based on the geomorphological sediment thickness (GST) derived from Pelletier et al. (2016) to predict site amplification at continental or regional scale. This new model is compared to three known models based respectively on Vs30 proxy, slope, and geological era. Then, the authors apply the four models at the border region between Turkey and Syria and, more locally, for three main cities of the region. The proposed methodology is worth publishing, aligned with the scope of the natural hazards and earth system sciences.

However, the objectives of the paper are not clear enough for the readership to endorse the authors's hypothesis. To my point of view, the rationale should be more detailed and the conclusions should emphasize the use limitations of the proposed model. Rephrasing should concern the points given below.

- **The title of the paper** does not fully reflect the contents of the paper. In a large part of the paper, the authors compare the new model with three existing models. This comparison does not appear in the title.
    - We have changed the title to "Exploring inferred geomorphological sediment thickness as a new site proxy to predict ground-shaking amplification at regional scale. Application to Europe and Eastern Turkey", and also stressed the comparison more in the introduction (e.g. line 65)

- In the introduction (chapter 1), it is not clear why a new prediction model for site amplification is needed. A full discussion on the limits of the existing models (e.g. quality of input data, resolution, limits of application, etc…) and on the field of applications of this new model would probably better explain the choice of the authors. Would the new proposed model be applicable for a large number of GMM or only for GMM based on Kotha et al. Method?

- The resulting site amplification predictions and maps should only be interpreted in reference to the associated GMM used to derive $\delta S2S_s$, this has been clarified in the caption of Figures 12, 14, A5 and A6 and in the text (e.g. lines 275-276 and 442)
- The aim of this study, and of deriving the site amplification models, is to test whether GST add any valuable information to site amplification prediction and can be used as a site proxy. To answer this question, we derive site amplification model using both new and common proxies and compare the ability of the site proxies to predict the observed amplification. The main aim is therefore not to launch a new site amplification model, but to examine the effects of using different site proxies as alternatives to VS30. The limitations of VS30 are stated in the introduction. In order to clarify this, we have rephrased some of the descriptions of the object of the study, for example in line 72.

- In chapters 2 to 5, some points should be clarified to make the text easy to understand to a diversified audience (including for non-specialists in GMM development):
  - all the terms of the equations should be explained (for example Rref, a, g, Mh, etc…),
    - Thank you for pointing this out, we have added a sentence explaining the reference values in lines 113-114.

  - the authors should explained the frequency values chosen for the tests (f=0.529, 1.062 and 9.903 Hz) yet the interpretation of the prediction results is frequency dependant and should be altered by the limitations of the selected model,
    - The three frequency values are selected to show the amplification at three different frequency values, the test is however performed at all frequencies between $f = 0.460 - 9.903$ Hz, this has been clarified in line 138.

  - l. 155: the authors explain that they have used a different processing from previous works: what is the impact of this choice?
    - The difference from previous work is not directly the processing applied, but the value used. The Pelletier et al. (2016) provided several different grids describing different aspects of regolith, soil and sediment thickness. In this study we use the combination of the two soil and sediment thickness grids (for hillslope and upland/lowland) in order to access the value for a broader area. Several clarifications have been made in lines 170-180.

  - l. 158: the GST data does not extend beyond 50 m depth. What is the impact of this limitation on the prediction model, especially at low frequency?
    - At low frequency the performance of the site amplification model based on GST is slightly reduced, this is now discussed in lines 252-255.

  - Paragraph 3.3: a discussion on the resolution of the geological era model would be necessary and its impact at high frequency,
    - The resolution of the geological era (1:1,500,000) have been added in lines 188 and 194 and is also attributed as passible factor to explain the poor performance of the models at high frequencies.

- Paragraph 4:
  - which arguments could the authors present to demonstrate that the relation between site amplification and measured V30 is log-linear?

    - We have added a figure (new Fig. 4) showing that the linear regression between ds2s and ln(VS30) have a higher correlation coefficient than the linear regression between ds2s and VS30 in linear scale, this is also described in lines 216-222.

  - Could the authors give estimators of the goodness of fit for all linear regressions?
    - Yes, thank you for pointing out that this was missing. We have added a figure (new Fig. 6) showing the coefficient of determination for the different regressions shown in Figure 5, and a figure in the Appendix (Fig. A2) showing the coefficient of determination for the linear regressions shown in Figure 5 and 6.

  - Could the authors improve their figures in terms of legibility (in particular dash lines are a poor graphical choice),
    - We have altered some of the Fig 4 (new Fig 5) and color scales used for the lines, but in some cases dashed lines are still used in order to keep the figures colorblind friendly.

  - l.236: high sediment thicknesses induce lower frequency site amplifications but not necessary higher site amplification than low sediment thicknesses,
    - Yes, but the sentence in line 236 (now line 289), only refers to the general trends observed in Fig. 7 and are not trying to make any conclusions about site-specific amplification.

  - the geological era is inferred from a low resolution model: how this could impact
    - As described above, the resolution of the geological era could be one of the causes for the poor performance of the models at high frequencies, this is mentioned in line 337.

- Paragraph 5:
  - Could the authors re-explicit the way the indicator works?
    - Yes, we have added lines 311-316 to better explain how the indicators work.

  - Could the authors harmonize the symbology used for equations 7 and 8 and the symbology used for figure 7?
    - Yes, thank you for pointing this out, style of phi has been changed.

  - The authors show that none of the amplification models are not distinguishable for frequencies above 3 Hz. What are the consequences of this statement? Does this mean that they cannot be used for frequencies above 3 Hz?
    - Yes, and a similar result was also found by Bergamo et al. (2022) who also used higher resolution site proxies. Bergamo et al. (2022) found that low resolution proxies worked best at low frequencies, higher resolution proxies worked well at intermediate frequencies, while direct proxies work at a wider frequency range (up to 5Hz). Lines 338-341 have been added to emphasize this.

- In chapter 6, the authors apply the new GST based model to Europe and to the Turkey-Syria border region. They compare the four models for three different soil classes based on Vs30 measured values (175 m/s, 375 m/s and 775 m/s). The main points to be discussed are the following:
  - How does they choose those values? Are they representative of the distribution shown in Figure 3? In terms of site effects, the first and last class (175 and 775 m/s) correspond to

very specific configurations (very soft soils and rocky sites): testing the model in more "regular" configurations should improve the robustness of the conclusions.

  - Thank you for this comment, the three ranges (soft soil, soil and rock) were chosen to show how the model perform for very different soil types, but it is true that these ranges do not represent well the distribution in figure 3. We have therefore changed figure 8 (now Fig. 11) to follow the Eurocode 8 (EC8) classes instead of Vs30 intervals, both to better justify the choice of values and to better follow the distribution in Figure 3.

- l.310-314: I do not agree with the hypothesis associating soft soils to high GST values (and stiff soils with low GST values). This hypothesis does not take into account the Vs of the sedimentary layer. In the deep Tertiary basin for example, one could have a stiff soil (in terms of VS30) with high thicknesses and high site amplification. This point should be discussed in details and the impact of such hypothesis should be emphasize.
  - As described in the comment above, Figure 8 (now Fig. 11) have been changed to follow the EC8 classes instead of Vs30 and the GST values are now chosen based on the new draft for EC8 following Paolucci et al. (2021)

- Table 1: how many data are available in each site condition classes (soft soil/stiff soil/rock)?
  - The value has been added and Table 1 have been changed to match the EC8 classifications as shown in new Fig 11.

- Figure 9: is it adequate to consider GST values inferior to 5 m (this configuration is considered as rock site in EC8 classification)? This point should be discussed.
  - As described in the comments above, the GST values are now selected based on the new draft for EC8 following Paolucci et al. (2021)

- l 345-l. 351: the authors apply the four selected prediction models to Europe but they give neither spatial nor numerical indicator to compare the site amplification values of each model (for example through the plot of their respective distribution) and their respective impact on risk assesment. A thorough discussion on the results is necessary to emphasize the pros and cons of the proposed model. Though the aim of the paper is not to re-create the exact site- specific amplification, it should be also interesting to test whether the models are in adequation with the regional site amplification maps inferred from other site effects proxy (for example the Italian VS30 map of Mori et al).
  - As described above, the object of this study, and of deriving the site amplification models, is to assess the ability of the proxies to predict site amplification, not to develop and propose new site amplification model. Several sentences have been rephrased to make this clearer (e.g. line 367-368 and 525).
  - The results of this study have shown that using different site proxies to predict site amplification gives notable different results and captures the epistemic uncertainty associated with modelling site amplification when using inferred proxies. This epistemic uncertainty should be incorporated into the final risk calculation, however, performing risk and loss calculations using the proxy-based site amplification models is beyond the scope of this work. A discussion on the impact the study has on risk assessment is added in paragraph in lines 535-544.

- Figure 11: Where are the Taurus mountains? Please complete the maps with the countries names. At the Turkey-Syria border, the GST model presents a possible artefact and the geological era model does not provide data on the Syrian side. In those conditions, why did the authors choose this area for their test? Would not it have been more appropriate to use a region with a better coverage in terms of calibration data
    - The area was chosen because the recent earthquakes demonstrated the need for cross-boundary site amplification maps. In addition, a comparison using the ground motions recorded during the recent events has been added (lines 491-510).

- Figure 12: Please complete with the February 2023 epicentres.
    - This has been done in Fig 11 (now Fig. 13) and the new Figure 15 showing the distribution of the new dataset.

- l. 400 and more: As said by the authors, the proposed models are regional. In this context, is it correct to test them at city scale? If so, why did they choose the Antakya, Aleppo and Gaziantep cities since they made no comparison with the observed site amplification or damage during the February 2023 events. What is the aim of this test?
    - While it is true that the site proxies and models are only meant for regional scale, the city comparison serves to emphasize how different the predictions are and the importance of capturing epistemic uncertainty when adopting these approaches in seismic risk analysis or even in rapid post-event impact assessment. However, as also stated in the manuscript, this comparison is not entirely appropriate, and we realize that it might also add to the confusion about of the aim of the models. We have therefore removed this part and instead added a comparison of the models to the within-event residuals from the recent events (lines 491-510).

- In the conclusion, the first point is that the authors do not conclusively demonstrate the value of the proposed new model since it gives equivalent or weaker results than the other ones. **The authors should give stronger arguments to show the interest of using their model in future hazard and risk studies at regional scale.** A second point to address is that the application of the four models at European scale show a great variability. **If the authors wish to continue applying these models, it is necessary to quantify the impact of such broad epistemic uncertainties for site amplification prediction on risk assessment.** Third, the authors should not avoid discussing about the limitations of the proposed model (pros and cons) and should propose a plan of actions to improve it and, consequently, decrease the related uncertainties.
    - As described above, the object of this study is to evaluate the ability of the proxies to predict site amplification, the models are therefore exploratory and not developed with the aim of capturing the best possible relation between the proxies and empirical amplification. Several sentences have been rephrased to make this clearer (e.g. line 368-369 and 524).
    - Furthermore, the results of this study have shown that using different site proxies to predict site amplification gives significantly different results. This emphasizes the importance of capturing the epistemic uncertainty associated with modelling site amplification when using inferred proxies. The epistemic uncertainty needs to be incorporated into the final risk calculation and to fully assess the impact of this epistemic uncertainty, risk and loss calculations should be performed using the different site amplification models, however, this is beyond the scope of this work. This is discussed in the added paragraph in lines 536-544 and emphasized more in the conclusion.

References

Bergamo, P., Hammer, C., and Fäh, D.: Correspondence between Site Amplification and Topographical, Geological Parameters: Collation of Data from Swiss and Japanese Stations, and Neural Networks-Based Prediction of Local Response, Bulletin of the Seismological Society of America, 112, 1008–1030, https://doi.org/10.1785/0120210225, 2022.

Pelletier, J. D., Broxton, P. D., Hazenberg, P., Zeng, X., Troch, P. A., Niu, G.-Y., Williams, Z., Brunke, M. A., and Gochis, D.: A gridded global data set of soil, intact regolith, and sedimentary deposit thicknesses for regional and global land surface modeling, Journal of Advances in Modeling Earth Systems, 8, 41–65, https://doi.org/10.1002/2015MS000526, 2016.

Paolucci, R., Aimar, M., Ciancimino, A., Dotti, M., Foti, S., Lanzano, G., Mattevi, P., Pacor, F., and Vanini, M.: Checking the site categorization criteria and amplification factors of the 2021 draft of Eurocode 8 Part 1–1, Bulletin of Earthquake Engineering, 19, 4199–4234, https://doi.org/10.1007/s10518-021-01118-9, 2021.