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

Manuscript Number: egusphere-2023-1370

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.

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?

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…),
- 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,
- l. 155: the authors explain that they have used a different processing from previous works: what is the impact of this choice?
- 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?
- Paragraph 3.3: a discussion on the resolution of the geological era model would be necessary and its impact at high frequency,
- Paragraph 4:

- which arguments could the authors present to demonstrate that the relation between site amplification and measured V30 is log-linear?
    - Could the authors give estimators of the goodness of fit for all linear regressions?
    - Could the authors improve their figures in terms of legibility (in particular dash lines are a poor graphical choice),
    - l.236: high sediment thicknesses induce lower frequency site amplifications but not necessary higher site amplification than low sediment thicknesses,
    - the geological era is inferred from a low resolution model: how this could impact
- Paragraph 5:
    - Could the authors re-explicit the way the indicator works?
    - Could the authors harmonize the symbology used for equations 7 and 8 and the symbology used for figure 7?
    - 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?

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.
- 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.
- Table 1: how many data are available in each site condition classes (soft soil/stiff soil/rock)?
- 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.
- 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).
- 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?
- Figure 12: Please complete with the February 2023 epicentres.
- 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?

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.

<u>Review conclusions:</u>

I recommend major revisions to strengthen the work done by the authors before publication.