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 δS2S,

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.

1. Reviewer Comments natural hazards and earth system sciences

The authors have tested the geomorphological sediment thickness (GST) and classical local soil-conditions proxies such as Vs30 in response to this question: Is GST a relevant global site proxy for PSHA analysis. To do this, they used site-to-site residuals issued from Kotha FAS-Model.

The selected study areas and proposed methodology are of scientific and engineering interest, aligned with the scope of the natural hazards and earth system sciences. However, the claim that the developed model is based on site-to-site residuals using linear regression is questioned, the organization of the different sections can be improved, and the manuscript lacks adequate explanation in a few areas for a reviewer to fairly assess the technical quality of the study:

1. General remarks

- In my opinion, the use of delta_S2S residuals (which is considered here as an epistemic uncertainty) is strongly linked to the Kotha model. To validate the method, several GMMs need to be considered and it needs to be proven that the delta_S2S values do not depend too much on the GMM used.
    - It is true that the $\delta S2S_s$ residuals are dependent on the GMM used and represent the systematic deviation of recorded ground motions from that GMM median predictions related to a site *s*. Because there is no proxy-based site term included in the GMM, the $\delta S2S_s$ residuals captures all the site-specific response, and can therefore be used to evaluate the ability site proxies to predict site amplification for that GMM. Section 2 have been renamed site-to-site term and partly rewritten to make this clearer (lines 88-96 and 117-121).

- For me, the simplest way is to deduce the amplification from the FAS ratio and to find a correlation between this ratio and the site proxies. e.g. if this ratio = 2 its interpretation is simple, but with a delta_S2S=-1.5 how can we interpret it especially this study targets ESHM20 by proposing a regional proxy (GST).?
    - As described above, $\delta S2S_s$ is relative to the mean of all sites and not to a rock reference, meaning that $\delta S2S_s$ with a positive value are amplified compared to the mean, and a

negative value signifies de-amplification relative to the mean of all sites. Line 121 have been slightly rephrased to make this clearer.

- With this study we remark that Delta_S2S=f(Proxy) is an heteroscedastic model (Khota model). This Heteroscedasticity lead to biased estimates of the standard errors of the regression coefficients. This can make it difficult to determine the significance of the coefficients and can lead to incorrect conclusions about the relationship between the proxy in the model. Additionally, heteroscedasticity can also affect the efficiency of the parameter estimates,leading to less precise estimates than would be obtained with a homoscedastic model. I know you are looking for this heteroscedasticity, since the GMM (Kotha) model does not contain any proxy representing site effect. However, I'm wondering if, can you add a site proxy (e.g.GST) to the GMM model to ensure that the delta_S2S follows a homoscedastic model (goal is to validate GMM model).
    - This comment can be understood in two ways. Either it is indicated that heteroscedasticity may affect the results because the $\delta S2S_s$ may be magnitude dependent and include nonlinear amplification, or that the un-even distribution of the site proxies may bias the resulting site amplification models.
    - Regarding the first issue, because $\delta S2S_s$ can be derived on a limited number of events (minimum 3), magnitude bias and nonlinearity may affect the $\delta S2S_s$. However, as also argued by Weatherill et al., 2023, $\delta S2S_s$ is generally assumed to be linear because most dataset contains mainly lower magnitude intensity ground motions, which is also the case for the ESM dataset which has few records from large earthquakes on soft soils stations, this has also been observed by Guéguen et al. (2019)
    - For the second issue, the following fitted (predicted) model vs residuals figure show that although the residuals of the site amplification models have a large scatter, they are more or less equally distributed along the fitted values (x-axes) range, except for a few outliers and a slight downgoing trend at high predicted values based on slope. We therefore assume that site amplification models are not demonstratively heteroscedastic:

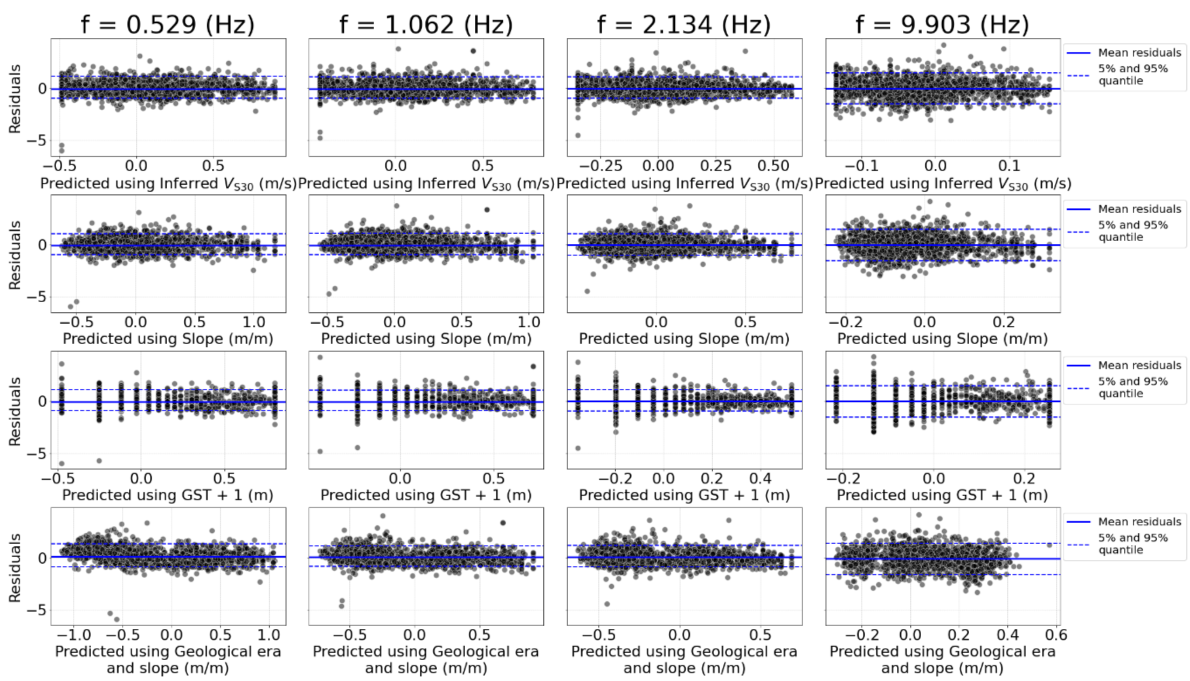

Figure 1 fitted (predicted) model vs residuals between $\delta S2S_s$ and the site amplification models based on inferred V_{S30} (top row) slope (second top row), geomorphological sedimentary thickness (second bottom row) and geological era and slope (bottom row), for f = 0.529 Hz (left), f = 1.062 Hz and f = 9.903 Hz.

- Why did you use linear regression to have Site amplification factor ?. Site behavior is so complicated that a simple linear model is insufficient to consider the underlying physics. Or use small strain conditions to remove nonlinear effect. In addition, delta_S2S. In this sense, it must be shown that the model developed does not suffer from underfitting.
    - It is indeed true that site behavior is too complex for a linear relation to fully capture, but because the main object of this study is to compare the ability of the proxies to predict site amplifications, and not to find the best possible relation between the site amplification and the proxies, we use linear regression for the simplicity. This has been clarified in lines 272-274.

- And I wonder how to obtain the amplification value with delta_S2S.
    - As described above, because no proxy-based site term is included in the GMM used to derive the $\delta S2S_s$., $\delta S2S_s$. captures all the site-specific response, and can therefore be used as an empirical site-amplification function describing the local amplification, or de-amplification, of each station with respect to the median of all sites. Section 2 have been renamed site-to-site term and have been partly rewritten to make this clearer (lines 88-96 and 117-121).

- Finally, if this is a sensitivity study, it should be mentioned in the title. Like : Using site-to-site residuals to testing the Relevance of geomorphological sediment thickness as a regional site proxy. Application to Europe and Eastern Turkey
    - 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 lines 78-80 have been rephrased to make it clearer that we are using site-to-site terms.

2. Specific Remarks

- Eq 1: Why did you not use the site term (e.g. Vs30 inferred) as an explanatory variable for the site effect (fixed effect), this helps reduce random variability; and the delats2 are used to consider uncertainties that are not taken into account by Vs30 (i.e. For example, if VS30 is used for site classification, two sites with the same VS30 can still have significantly different site profiles and therefore have different site amplifications).
    - As described above, using no proxy-based, like VS30, site term in GMM makes sure the resulting $\delta S2S_s$ residuals capture all the site-specific response, this has been better emphasized in lines 116-119.

- Eq 1 is a non-linear functional form; why did you use linear mixed effect model rather be non-linear model like INLA ?.
    - We use the robust linear mixed-effects regression to derive the random effects while down-weighting the outliers, and to stay consistent with previous work. However, using INLA, which uses a Bayesian framework, might be a possibility in future work.

- Line 114 : rationalization ? do you mean regionalization.
    - Yes, thank you for pointing out this error, it has been corrected.

- The use of a GMM model such as the one used in this study may complicate the interpretation of the results. I suggest you try a site amplification model:

Amp(FAS_sur/FAS_Downhole) This way, you would only have delta_S2S and delta_Amp (describes the record-to-record variability of the amplification at site s for earthquake e).

- Because we are looking at Europe-wide amplification, deriving the amplification from the spectral ratio with a nearby rock reference or borehole station is not an alternative, a paragraph in lines 89-96 have been added to make this clearer.

- To make sure there's a correlation between amplification factor (AF) and delta_S2_S, I'd like to have a figure that gives Amp vs exp(delta_S2S). Amp=FAS_soft_soil/FAS_Rock. You can use the EC8 classification.
    - As described above, using spectral ratio with nearby rock reference is not an alternative when wanting to obtain amplification factors for such a large area as Europe, lines 89-96 have been added to make this clearer. However, $\delta S2S_s$ can be used as an empirical site-amplification function describing the local amplification, or de-amplification, of each station with respect to the median of all sites (Kotha et al., 2018). In fact, previous studies have compared $\delta S2S_s$ to and show strong similarities with amplification factors derived using a generalized inversion technique (GIT, e.g. Bindi et al., 2017; Wang et al., 2023), this is also described in line 122:

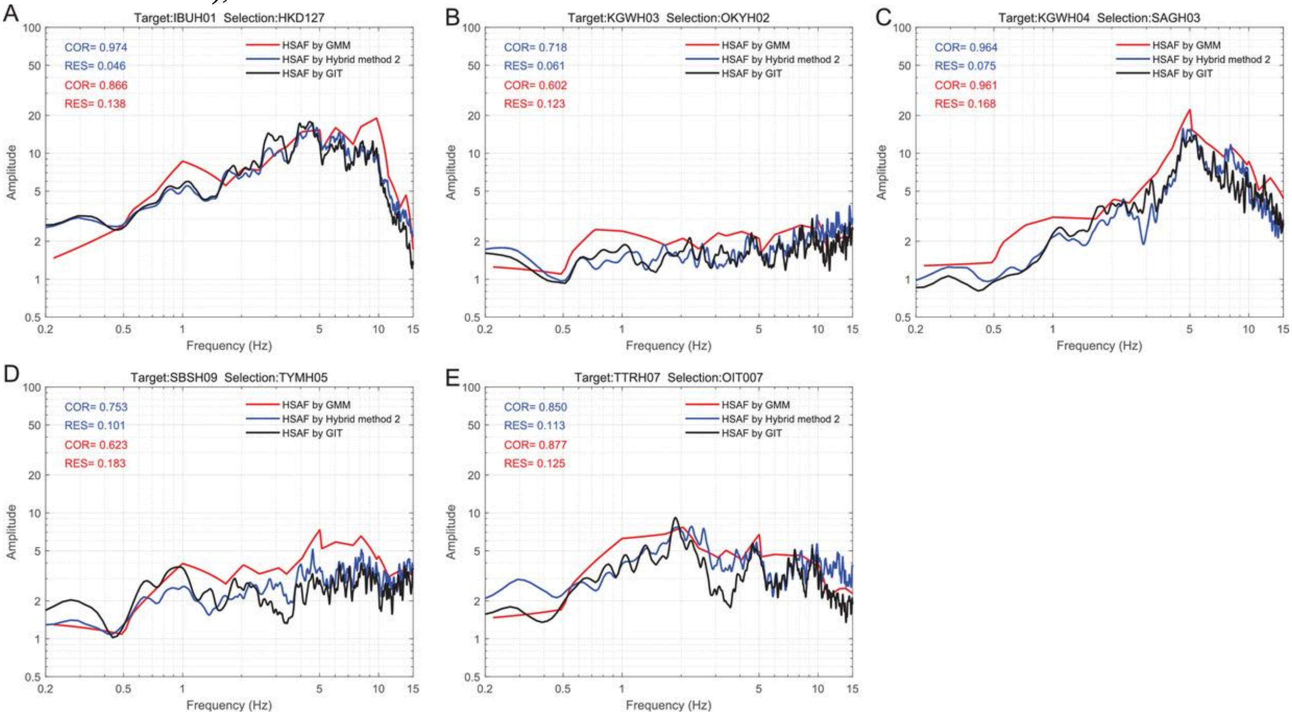

Figure2: from Wang et al. (2022), comparing horizontal site amplification factors (HSAF) estimated using GMM for the Fourier spectra performed by Loviknes et al. (2021) Hybrid method 2, and GIT at the Japanese Kik-net stations (A) IBUH01, (B) KGWH03, (C) KGWH04, (D) SBSH09, and (E) TTRH07.

- In figure 3, I wonder if the (non-Gaussian) distribution of geomorphological sediment thickness (e.g. more than 800 sites have H=0-2 m)... is this non-uniform distribution taken into account when building the model?.
    - Yes, as discussed in lines, the non-Gaussian distribution of GST is caused by the extreme/end-values 0 and 50 m. To deal with this issue we have attempted censored regression, removing the extreme values and are finally using 10-fold validation to ensure that the linear regression is not to depended on the station distribution. Line 243 have been rephrased to make this clearer.

- In eq 6, log is ‘t log10 or Ln ?. Also does Ys(f,Proxy) represent delta_S2S ?.
    - Yes it should be ln, thank you for pointing this out. Ys(f,Proxy) represent the predicted $\delta S2S_s$ based on the proxy-based site amplification model derived from the linear regression, is has been better clarified in line 211.

- **I'm not convinced that delta_S2S can represent amplification itself.** But rather the epistemic uncertainty of the site effect part (closely linked to the GMM used). In addition, amplification is normally unitless. However, here, delta_s2s takes unit of the FAS. Thank you for explaining this to me.
    - As described above, $\delta S2S_s$ residuals represent the amplification, or de-amplification, of each station with reference to the median GMM prediction for all the sites. δS2S is unitless and exp($\delta S2S_s$) is the amplification factor. However δS2S is derived for FAS (Fourier amplitude spectra) and not for response spectra (which is the common practice).

- Line 214 : extreme values or outlier value ?.
    - Extreme values here refer to the values on each end of the range, so for GST, 0 and 50 m, to avoid confusion the word extreme has been changed to end-value in line 259.

- Line 217 : What do you mean by "fold cross validation test". Give us some explanations. And why “10” fold ?.
    - The explanation of the method has been improved in lines 264-267.

- Figure 5, usually with classical GMMs, we have Vref (e.g Vs30 = 760 m/s), here we don’t see this threshold why ?.
    - As described above, no VS30-term is included in the GMM and a refernce VS30 is therefore not used. Instead $\delta S2S_s$ is referenced to the median of all the sites, not to rock sites.

- Eq 7 is nothing more and nothing less than the residual equation in equation 6. Why named correction term?. I would like to see a figure that gives the delta_S2Scor vs proxies for some frequencies, it gives us an idea on the presence or not of bias.
    - The correction term $\delta S2S_{s,cor}(f)$ in equation 7 and 8 represents the remaining site amplification that is not captured by the proxy based amplification prediction $\delta S2S_s(f,Proxy)$ and $\Phi_{s2s,\,cor}(f,Proxy)$ is the site-to-site variability of $\delta S2S_{s,cor}(f)$. If the site amplification model where able to perfectly predict and capture the full range of the site amplification at a specific site, $\varphi_{s2s,\,cor}(f,Proxy)$ would be reduced to zero. However, such an ideal case is not realistic and conventional site amplification models can only aim to reduce the $\Phi_{s2s,\,cor}(f,Proxy)$ as much as possible. A lower $\Phi_{s2s,\,cor}(f,Proxy)$, therefore indicates that the proxy are able to capture the site amplification. This has been clarified in lines 322-316. The new figure A3 showing $\delta S2S_{s,cor}(f)$ with the proxies are included in the appendix, and show that there are obvious biases as $\delta S2S_{s,cor}(f)$ is evenly distributed around zero.

- In line 304, you wrote: “The object of this study is to predict regional site amplification over a large area using regionally or globally available site proxies”. In my view, this sentence must be in “introduction part”.
    - This sentence has been rephrased to “The object of this study is to test regionally or globally available site proxies as predictors for regional site amplification over a large

area” (now line 369) and parts of the introduction have been rephrased to make the object clearer.

- The comparison between Figure 8 (training phase) and Figure 3 (testing phase) are not consistent. In fact, you have to validate with a smaller interval (like 350-400 m/s) and add soft sites comparison.
    - Thank you for pointing this out, we have changed Figure 8 (now Fig. 11) to follow the Eurocode 8 classes instead of the Vs30 intervals to better justify the selected ranges and follow the distribution in Figure 3.

References

Bindi, D., Spallarossa, D., and Pacor, F.: Between-event and between-station variability observed in the Fourier and response spectra domains: comparison with seismological models, Geophysical Journal International, 210, 1092–1104, https://doi.org/10.1093/gji/ggx217, 2017.

Guégen P, Bonilla F, Douglas J: Comparison of soil nonlinearity (In Situ Stress-Strain Relation and G/Gmax Reduction) observed in strong-motion databases and modeled in ground-motion prediction equations. Bull Seismol Soc Am 109(1):178–186, 2019.

Kotha, S. R., Cotton, F., and Bindi, D.: A new approach to site classification: mixed-effects ground motion prediction equation with spectral clustering of site amplification functions, Soil Dynamics and Earthquake Engineering, 110, 318–329, https://doi.org/10.1016/j.soildyn.2018.01.051, 2018.

Loviknes, K., Kotha, S. R., Cotton, F., and Schorlemmer, D.: Testing nonlinear amplification factors of ground-motion models, Bulletin of the Seismological Society of America, 111, 2121–2137, https://doi.org/10.1785/0120200386, 2021.

Wang, Z., Nakano, K., Ito, E., Kawase, H., and Matsushima, S.: A hybrid approach for deriving horizontal site amplification factors considering both the similarity of HVSRe and the vertical amplification correction function, Earthquake Engineering & Structural Dynamics, 52, 128–146, https://doi.org/10.1007/s10518-023-01712-z, 2023.

Weatherill, G., Crowley, H., Roullé, A., Tourlière, B., Lemoine, A., Gracianne, C., Kotha, S. R., and Cotton, F.: Modelling site response at regional scale for the 2020 European Seismic Risk Model (ESRM20), Bulletin of Earthquake Engineering, 21, 665–714, https://doi.org/10.1007/s10518-022-01526-5, 2023.