Title: "Ground motion prediction maps using seismic microzonation data and machine learning" by Federico Mori et al., Nat. Hazards Earth Syst. Sci. Discuss.,

Editor comment

Dear authors,

this manuscript can now be accepted - but, please, implement still some minor corrections suggested by the reviewers before we provide the version to be processed for in press.

Sincerely yours

Hans-Balder Havenith, assoc. editor

Dear Editor,

we thank you and reviewers for their thorough assessment of our paper. We have carefully addressed the reviewers' comments and made corresponding changes to the manuscript. We have modified and added some figures. We have carefully revised the ambiguous statement in the article. The response to reviewers' comments has been reported below.

Referee #1 comment

The authors would like to thank the reviewer for his comments.

The manuscript has been improved. This paper involves lots of work, it would be a plus to provide some insights. One point mentioned by both reviewers is about the strength of correlation between each input and output, answering this question will give us some knowledge of the complex pattern within the data to render the ML models somewhat interpretable. Could the authors provide a plot on the feature importance from which we could see which predictor variable is necessary or unnecessary in the model. This is also to validate the findings of Zhou et al (2020) whether the pattern reported in China on topographic effect modeling is still applicable in Italy.

Reply:

We have inserted a sentence at the end of Section 3.1 and referred back to the appendix for the plot.

Referee #3 comment

The authors would like to thank the reviewer for his comments.

The paper shows the approach to derive seismic parameters through ML approach. I read the manuscript and checked the modifications carried out following past reviewers comments and I found the Authors' effort satisfactory.

Referee #4 comment

The authors would like to thank the reviewer for his comments.

This paper aims to assess the advantages / limitations of machine learning techniques - with respect to more conventional approaches ie. GMM - to produce ground motion prediction maps for large areas considering both topographic and stratigraphic local conditions. To do so, the authors identified key-parameters on which they develop their methodologies. In addition, they provide a comparison of their ground motion prediction maps to the results of advanced numerical simulations based on detailed sub-soil models.

This reviewer acknowledges that the authors took into consideration some of the guidelines of the reviewers of the first step of the review process, which is good. Answering the remarks / questions below might help to further improve the quality of this paper which, in the present state, still requires minor revisions according to this reviewer.

Scientific comments:

Introduction:

- Line 26: « regarding the spatial correlation … » the meaning of the spatial correlation in this context is not clear to this reviewer

Reply:

The following statement was added: "where the spatial correlation is the spatial characteristics of the ground motion arising from similarities in the seismic wave paths and local-site effects".

- Line 81: as already pointed out by reviewers in the first step of the review process, this paragraph should be erased because it is not commented in the text and it does not help to better understand the work.

Reply:

These lines were moved to the section "Discussion and conclusions" since they are briefly presented in the appendix and are outlined as future research.

- Line 85: this paragraph could be part of the discussion

Reply:

These lines were moved to the section "Discussion and conclusions".

Data:

- Line 96: « some data distributions seem to be imbalanced »: the meaning of this sentence is not clear to this reviewer

Reply:

As already written in line 94: An imbalanced training input dataset is characterised by an unequal distribution of values.

Method:

- Line 61: is Vs30 parameter enough to represent sub-soil conditions and take into account possible geological site effects

Reply:

"V_{S30}, the fundamental frequency of the deposit (f_0), and the depth to the engineering bedrock (H_{800}) are the key-parameters which well gauge the effect of local sub-soil conditions on the seismic wave propagation (i.e., lithostratigraphic effect)" was already reported in the manuscript. The following statement was added consistent to what already reported in "Introduction": "The only V_{S30} was used in the adopted ML approach since the Italian V_{S30} map was provided by Mori et al. 2020a while national f_0 and H_{800} maps are not currently available".

- Line 194: the references should be moved to Line 184 according to this reviewer.

Reply:

Lines were moved accordingly.

Results :

- Line 277: this statement of a good agreement between geophysical data and ground motion prediction maps is not convincing to this reviewer. In addition, it would be worth showing also the geomorphological maps to allow the authors to draw such conclusions. As previously mentionned by reviewers in the first step of the review process, it is very difficult to visually determine if both maps are coherent (see Accumoli or Amatrice for instance and compare their values to Petrana). Maybe a different definition of the colorbar scales in both Figures could ease the comparison between plots.

Reply:

The required maps have been added to the appendix.

- Line 295 : according to this reviewer, the comparison with numerical modeling results, although worthy of investigation, is not described into sufficient details to allow the readers to compare ML results to more advanced numerical simulations results. Some criticisms were already made by reviewers during the first step of the review process. Because this section of the comparison could be an original / novel contribution of the paper, to this reviewer it deserves further explanations : methodology used in the numerical modeling, materials behaviour and properties, mesh resolution, seismic inputs parameters, etc.

Besides, the correspondance between numerical results and this study should be quantified to this reviewer: looking at the shape of the curves, the fit is not so obvious. Although 0.3s might be of interest for this area, what about the response of both methods at other spectral periods ?

Reply:

We apologize since further explanation of the numerical modelling cannot be reported in the manuscript since they have been performed by different research groups not including all the authors of this manuscript. The research papers were already cited in the manuscript Gaudiosi et al. (2021), Giallini et al. (2020), and Grelle et al. (2020). The response of numerical method at other spectral periods was not included in the manuscript since not available to us (i.e., not provided to us by the authors of the mentioned works).

Correspondence between numerical results and this study was considered valuable in terms of trend rather than by a quantitative point of view as already stated in lines “It should be noted that the trend of the values of our study reproduces that of the numerical simulations”. We would emphasize that the qualitatively good results of our work are the peak values of Sa trends due to the presence of a crest of the topography and of a buried valley as in the case of Amatrice (section AA’ in Figure 10) and Arquata del Tronto (section CC’ in Figure 10), respectively. On the contrary, the ShakeMap cannot reproduce the ground motion variation induced by the local variation of site conditions.

Finally, the following table and statement was added to the manuscript since the real records of OSS site and AMT station were considered the benchmark for the validation of our results. Future validation will be considered since we are selecting denser station array than the currently available.

“The difference between different methodology and the recorded values were quantified according to the following Equation and are provided in **Table 1**.

$$\varepsilon_{Sa} = \frac{Sa_{0.3\ \text{estimated}} - Sa_{0.3\ \text{recorded}}}{Sa_{0.3\ \text{recorded}}} \cdot 100 \qquad (1)$$ ”.

Table 1. Difference ε in percentage between $Sa_{0.3}$ determined by means of different methodologies and recorded values for Central Italy earthquake occurred on August 24, 2016.

	Section BB'	Section CC'
	ε_{Sa} (%)	ε_{Sa} (%)
this study	-7	-14
Numerical	-43	-44
ShakeMap	-42	-40

Discussion / conclusion

- Line 371 : « it is possible to predict the losses in the area struck by the earthquake in near real time » : the aim of this paper was to produce ground-motion prediction maps over large areas based on ML technique. As mentioned in the text, the link between ground motions and damage is not straightforward. Therefore the content of this sentence calls for some nuance.

Reply:

We have deleted damage and inserted "shaking scenarios".

Technical corrections :

- Line 11: what is a « burried morphology » ? Example ?

Reply:

“(e.g., irregular sub-interface between soft and stiff soils)” was added to the text.

- Line 56: « capable of properly representing … »

Reply:

Modified accordingly.

- Line 69: « in terms of »

Reply:

Modified accordingly.

- Line 104 « it seems a hard task to improve the training dataset » ; line 105 « anyway » : to this reviewer, this is more spoken language

Reply:

Line was modified as following: “Hence, the training dataset cannot actually be improved.”.

“Anyway” was deleted.

Figures / tables:

- Table 1: hxx refers to the « second order partial derivative of dxx » ? Same for hyy ? Please check.

Reply:

Modified accordingly.

- Figures 2 and 3: replace ‘°’ by ‘th’ in the percentiles definitions (25 and 75). Add units of the parameters (25 th, median and 75 th) when relevant.

Reply:

‘°’ was replaced by ‘th’. The units were not added to (25 th, median and 75 th) since it is already reported on the x-axis title.

- Figure 4: erase « performance performance »

Reply:

Modified accordingly.