The Authors would like to thank the anonymous referees whose comments made it possible to improve the paper considerably.

In the document, the texts in light blue are the authors' answers.

Text in italics are parts of the modified or newly added text.

For further clarity on this note, the authors have completely rewritten abstracts and modified the workflow of Fig. 1.

Ref 2

This paper studies the variability of Vs profiles within seismic microzones and finds that the Vs scatter is larger when the microzones are defined on purely geographical properties (GEOGR) than when they are assessed by taking into account also geological and geophysical data (GEOL_GEOP). This should probably be an expected result, as the geophysical data, used to define the microzones in the GEOL_GEOPH case, should already include Vs profiles (am I correct?). I think that the paper needs to be rethought and rewritten in some parts and that the calculations performed by the authors are not enough to support their final assessments.

MAIN ISSUES

As a general remark, I think that the use of the expression “Vs uncertainty” throughout the paper is misleading. By “Vs uncertainty” I understand the whole set of experimental/analysis/interpretation errors included in a Vs profile definition. Here, it seems to me that by “Vs uncertainty” the authors mean only Vs variability at a specific depth among different Vs profiles provided by different authors. If this is the case, I suggest that the authors remove the word “uncertainty” from the entire paper and use only “variability” (starting from the title, that I find very misleading). The abstract need to be entirely rewritten. It is hard to understand and very ambiguous (please see the comment in the attached text).

General answer

1) The referee asks us to describe the measurement uncertainties of the Vs profiles. This request cannot be accepted by the authors because the Vs profiles analysed are part of the dataset of a project in which various researchers and earth science professionals from all areas of Italy have produced data over the last 15 years active in Italy. The process by which these surveys are available to the authors involves:
   - Professionals and researchers carry out the surveys in the local area.
   - These surveys are checked from a scientific-technical point of view by a team of experts (research institutes, universities, state technicians) at regional level.
   - The dataset of the surveys is then forwarded to the State technical office (Civil Protection Department), which carries out a further congruity check.
   - The surveys are entered into a final dataset, which is the starting dataset used for this work.

As can be understood, it is impossible for the authors to distinguish (epistemic) measurement uncertainties from random variability. In this sense, the sigma lnvs that the authors calculated is total uncertainty and/or variability.

2) In this sense, the authors agree with the definitions of uncertainty and variability contained in https://www.epa.gov/expobox/uncertainty-and-variability, the key elements of which we quote:

“Refers to the inherent heterogeneity or diversity of data in an assessment. It is "a quantitative description of the range or spread of a set of values" and is often expressed through statistical metrics such as variance, standard deviation, and interquartile ranges that reflect the variability of the data. Refers to a lack of data or an incomplete understanding of the context of the risk assessment decision. It can be either qualitative or quantitative. Variability cannot be reduced, but it can be better characterized. Uncertainty can be reduced or eliminated with more or better data.”

In the text, we will therefore always refer to the calculated sigma with the two meanings variability and uncertainties (see also Passeri F., Foti S. , Marek A. R., A new geostatistical model for shear wave velocity profiles, Soil Dynamics and Earthquake Engineering, Volume 136, 2020, 106247, ISSN 0267-7261, https://doi.org/10.1016/j.soildyn.2020.106247.)

For the topic of ‘variability’, the authors have partially integrated the title and parts of the paper.

3) It is true that the investigations concerning the SM clusters were also used to construct the seismic microzones as claimed by the referee. However the authors are not interested in the seismic microzones construction process but are interested in quantifying the level of variability and uncertainty within the microzones.

For further clarity on this note, the authors have completely rewritten abstracts and modified the workflow of Fig. 1.

In the abstract, the referee states: ‘As far as I know, SM does not attempt to provide a <<accurate>> prediction of seismic response.

In order to respond to this remark, some definitions need to be clarified:

Accuracy refers to the ability of an instrument to indicate the true value of a measurement.

Precision refers to the ability to repeatedly indicate the same measurement value, whether or not it corresponds to the true value.

If the meaning of the observation respects these 2 terms, then the results of seismic microzonation studies are accurate because, as the SM is a seismic risk mitigation tool, it does not aim to predict the exact value of the shaking (precision) of an event (this would be a scenario), but aims to predict a range of shaking around an average value (accuracy). These results are obtained by using at least 7 events as seismic inputs and randomizing some ground parameters (including Vs profiles, the subject of this paper).

Dataset

In line 58 I think the authors should properly address the real authors of that large amount of Vs profiles made available.

As explained earlier in the general note, this is not possible.

This sentence “As a by-product, this work also provides for the first time the largest database of Vs profiles” suggests that the authors themselves compiled this entire Vs database while it seems to me that they are using data collected and processed by other people (the Mori et al. 2024 link does not say anything explicit about this).

See previous replies.

The sentence has been rewritten “*As a by-product, this work also provides for the first time the largest database of Vs profiles, collected by earth science researchers and practitioners since 2009*”.

Second, I am quite concerned about the presented dataset. On line 62 the authors state “we build a robust and reliable dataset of Vs profiles by removing any errors or duplicates (section 2.1);” but it seems to me that they did not check any data, analysis or interpretation quality. It seems to me that they just checked for double entries or similar. Is this the case? If so, this does not make the database neither “robust” nor “reliable”.

The analysed dataset is already the result of validations by regional authorities that check the the seismic microzonation from the technical and scientific point of view under the supervision of university professors who also control the agreement with H/V and other geological and geotechnical surveys. So, it is reasonable thinking that the here considered Vs profile are robust and reliable since already checked by local authorities. Thus the authors, starting from this official dataset, clean it of outliers, filter it with criteria derived from geostatistical analysis, analyze it to derive the standard deviation lnvs and then publish it in the ‘Zenodo’ link.

I plotted the provided Vs profiles vs depth and got the picture below: a large number of Vs profiles reaches depths much larger than 30 m, even 120 m and it seems that the wide majority of these very deep Vs profiles were obtained by means of MASW or ReMi surveys. Honestly, in my 25 years experience I have never been able to achieve depths much larger than 20-30 m with MASW or ReMi and definitely not by using them alone (without combining them with H/V or other surveys). Just in a few lucky cases of soils without large impedance contrasts it is possible to go deeper than 20 m using conventional MASW/ReMi approaches. I really need to be convinced that these Vs profiles reaching depths larger than 30-40 m are reliable. Please provide us with examples and convince us that such data are robust.

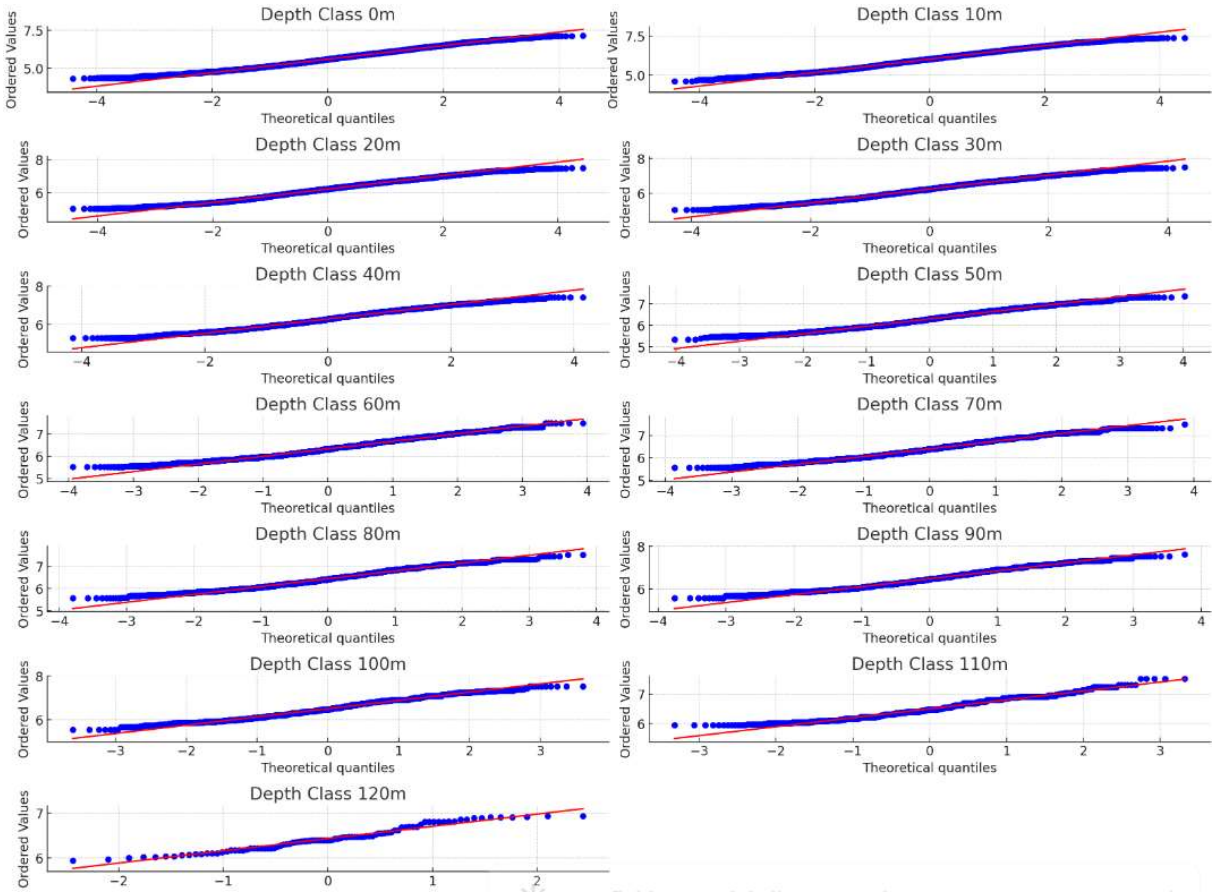
In theory, the authors agree with these observations but trust the statistical robustness of the dataset. It is still emphasized that the analyzed dataset is already the result of validations by local authorities that control the seismic microzonation at regional level under the supervision of university professors who also control the agreement with H/V and other geological and geotechnical surveys.

Statistics. Parametric statistics based on means and standard deviations cannot be applied to datasets with only 3 values, which is the minimum threshold set by the authors. The authors claim that the Vs distribution at a specific depth for a specific zone is log-normal but they should show this at least for the cases where they have a large enough number of Vs values (i.e. more than 30 values). Fig. 3 shows the distribution of Vs30 values, which is something different from my present request.

This hypothesis of lognormality of Vs with depth was originally verified by Toro (1995) on 557 Vs profiles and confirmed by many subsequent papers.

We produce QQ plots to verify the lognormal hypothesis as well. Here are the QQ plots for different 10-metre thickness classes (0-10m, 10-20m,...), considering all aggregated Vs profiles.

Each plot represents an interval of thicknesses and shows the quantiles of the log-transformed Vs data with respect to a normal distribution. If the points roughly follow a straight line, this would suggest that the original Vs distribution might be lognormal for that depth interval.



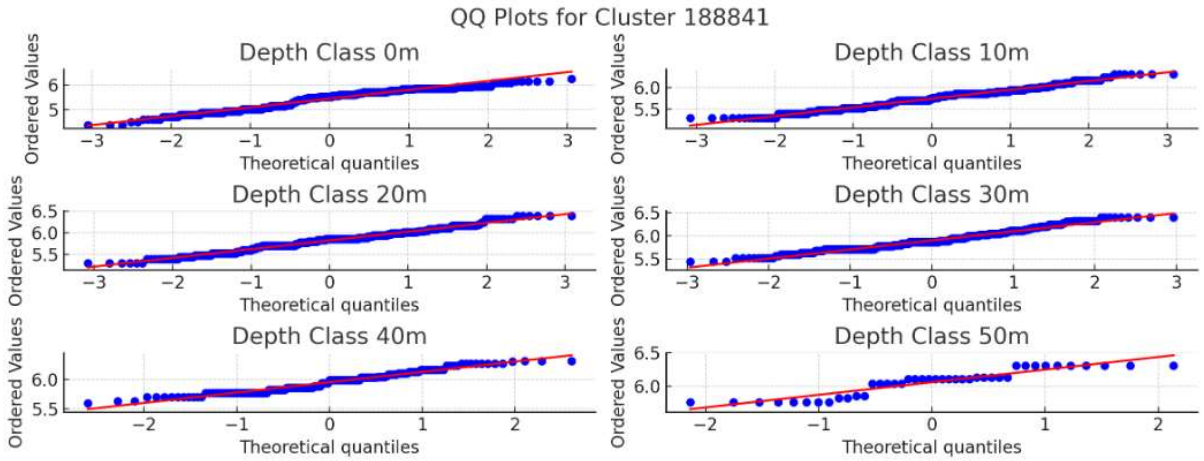
The idea is that, for large datasets, it is normal for there to be some deviations from the straight line, especially at the extreme values (the tails of the distribution). These deviations do not necessarily invalidate the assumption that the data follow a lognormal distribution, especially if:

1. Most of the points align well with the central straight line, indicating that the central data (where most of the observations are located) follow the theoretical distribution well.

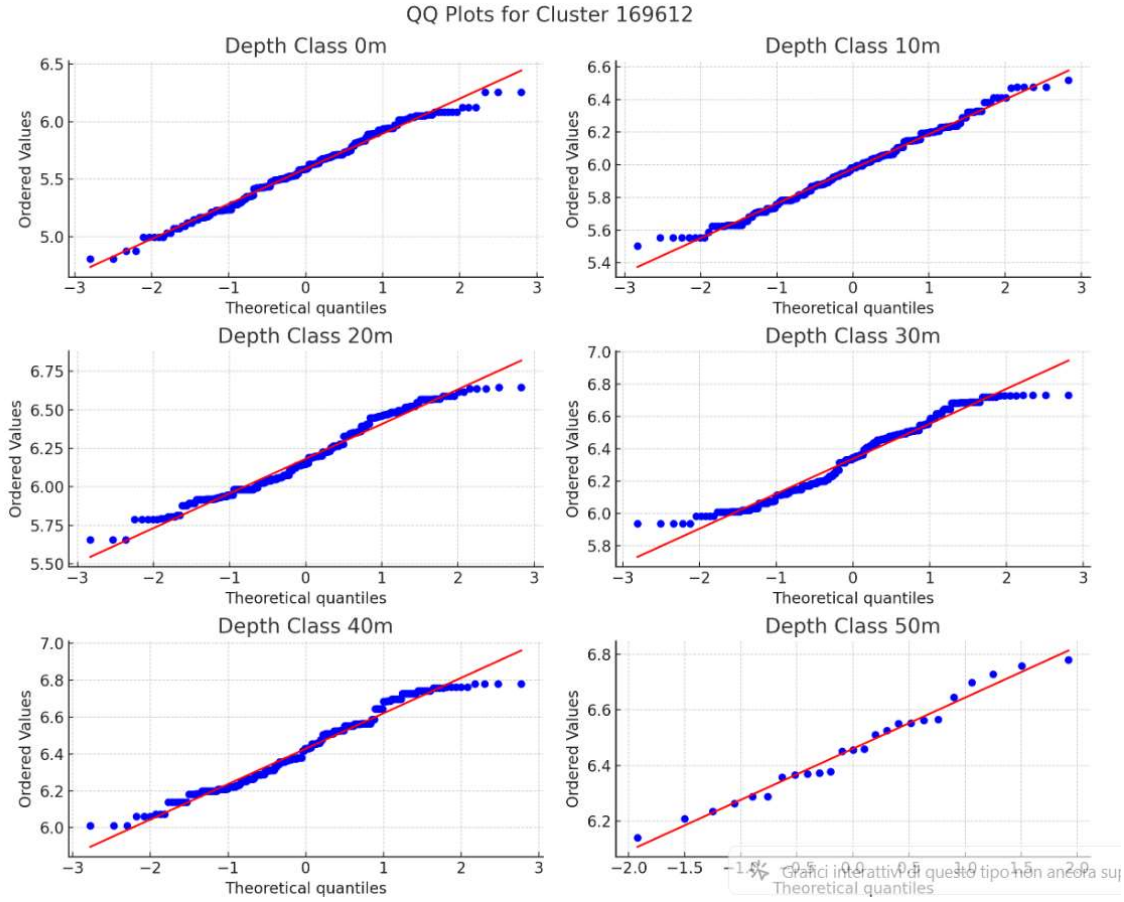
2. Deviations are small and unsystematic: Some degree of scatter at the edges of the plot is normal, especially if the dataset is large and includes noise or natural variability.

In summary, while the QQ plot for each depth interval showed some deviations from perfect linearity, we could still conclude that the Vs distribution is lognormal as the deviations are small and non-systematic.

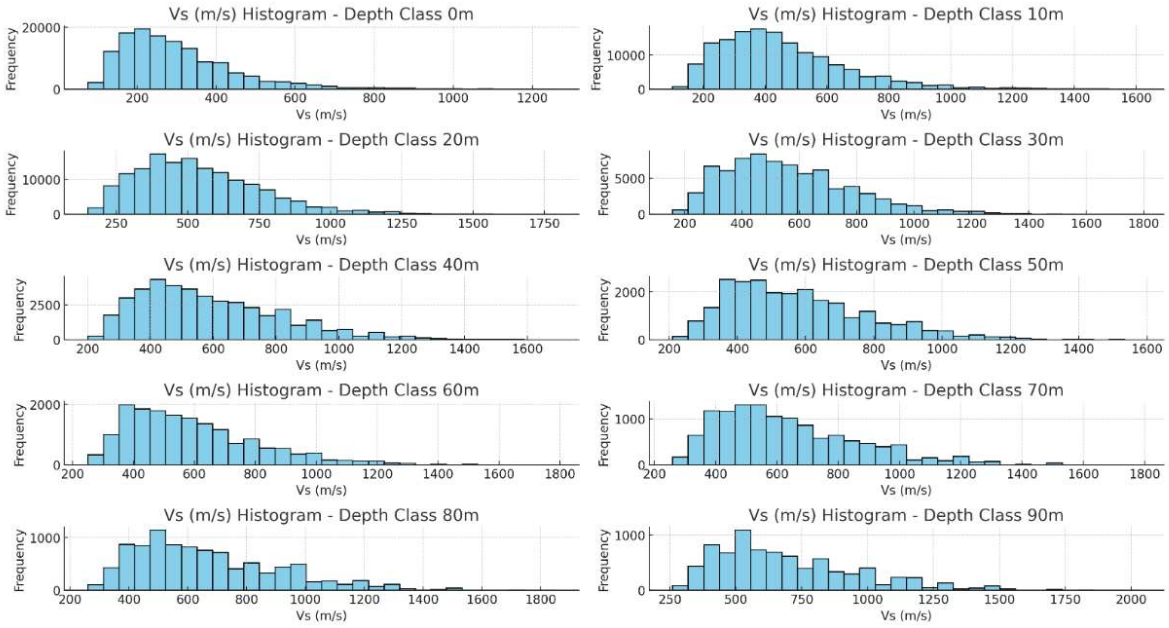
Here are the QQ plots for the cluster with the largest number of surveys. These plots also confirm what is described above.



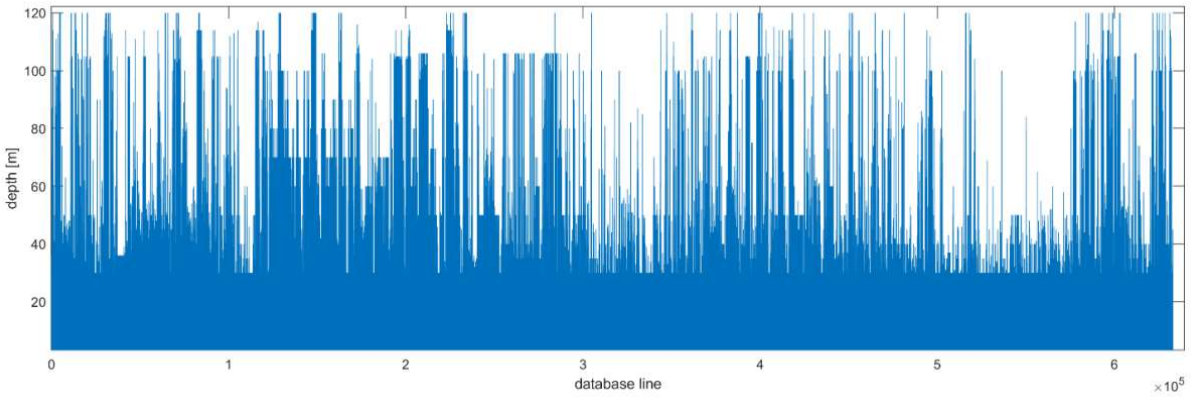
Here instead is the graph for a cluster with 30 surveys.



The next figure shows Vs histograms for the same depth classes.

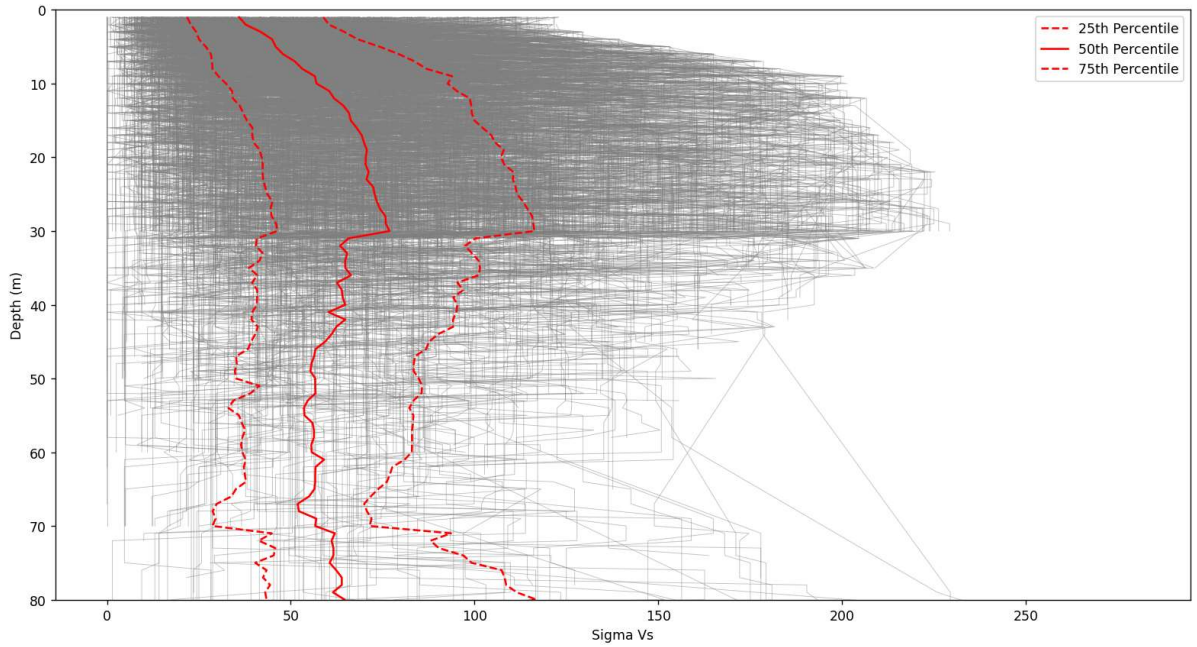


The statistics applied by the authors does not make sense on database with less than 20-30 values. Please provide us with more information and maybe apply a non parametric statistics.



Starting from the assumption that even the reference literature predicts the use of a lognormal distribution in the presence of few surveys (see Table 2 in Toro, 2022), we specify that only 0.37% of the clusters have more than 20 surveys.

The sigma Vs is given without considering the assumption of lognormality of the distribution (non-parametric statistics) as shown in the next figure.



$\sigma(ln\,Vs)$ = 0.15 (am I properly understanding the parentheses?) means Vs deviations lower than ±15 m/s (am I correct?). Is this real? Such precision cannot be reached even in the field at those depths (> 30 m), with experimental surveys. If the deviation of the mean of different Vs profiles for a given depth within a specific microzone is lower that 15 m/s (or something like that), than the experimental error in Vs assessment (which is certainly larger and is not considered in this paper) should also be taken into account.

As confirmed by the non-parametric graph provided above, around 30 meters the median of the non-parametric sigma is about 75 m/s, and up to 80 meters it is about 60 m/s.

The authors start discussing Vs profiles and Vs30. Then, at some point late in the paper, they introduce Vs10 and Vs20. Do we really need the latter two in this discussion?

This is because these are the same parameters analyzed in Zhou et al., 2023 for China and we wanted to compare and validate our results with those provided by Zhou et al., 2023.

With reference to line 268: An important point to address is understanding whether the Vs profiles collected for a same seismic microzone were produced by the same person (or few people) because this would mean that data are not fully independent and this generate a bias in their distribution. Please consider this point in the discussion. You are probably not dealing with truly independent datasets (at least from the interpretation point of view).

Database surveys were carried out for a variety of different types of projects: public and private building construction, environmental analysis, and natural hazard assessments. Finally, the surveys were conducted over a rather large period of time (about 15 years). These two considerations assure the authors of the statistical independence of the data, although it is not possible to exclude that in some cases it was a single professional who carried out multiple surveys

MINOR ISSUES

I found a little bit annoying the use of so many adjectives/nouns like “crucial”, “significantly”, “precision”, “effectiveness”, “discoveries”, “unparalleled”, “robust and reliable”, “detailed”, “valuable”, particularly when it is not truly demonstrated in the text that such performances have been truly achieved. You can be convincing also without using all those adjectives.

We have amended the text accordingly.

Line 38. ‘It has played a *significant* role’.

Line 46. See general answer

Line 49. See general answer

Line 56. See general answer

Line 58 and 62. See general answer

Line 64. The Moran Index is a statistical measure used to assess the spatial autocorrelation of a geographic dataset. In other words, it helps determine whether the values of a variable are spatially distributed (patterned) randomly, clustered (aggregated) or dispersed. This concept can be found in Figure 7.

Line 69. The aim of the work is to evaluate the sigma of the Vs profiles in depth in two independent clusters: 1. the Seismic Microzonation clusters, already defined by other researchers with geological, geomorphological, geotechnical and geophysical data; 2. the geographical clusters constructed from authors with the HDBSCAN technique.

So in general, there is no optimisation to find clusters but simply a comparison.

Line 73. ‘The results demonstrate the *impact* of the analysis performed.’

Line 83. OK, removed phase

Line 86. OK, change geometry in the table corrected: geophone deployment/array geometry

Line 97. OK, remove detailed corrected

Line 100. The highest values of Vs30 are distributed in the mountain ranges, while the lowest values are concentrated in the central-eastern region of the Po Valley, *as expected*.

Line 143. Figure 4: “for details….” removed

Line 147. Sentence reformulated: ‘Valuable insights into the spatial structure and variability of regionalized variables can be gained by analysing the shape and parameters of the experimental variogram.’

Line 154. Yes

Line 156. GIS language. In the context of ArcGIS, ‘attributes’ refer to the data or information associated with geographic features. Each feature in a GIS (Geographic Information System) dataset, such as a point, line, or polygon, has a set of attributes that describe its characteristics. These attributes are typically stored in a table, where each row corresponds to a feature and each column represents a different attribute.

Figure 6. Add marks

Lin 211.removed synthetic corrected

Line 211-212 There is no physical reason. See general answer

Figure 7. Moran's Index is a statistical measure used to assess the spatial autocorrelation of a geographic dataset. In other words, it helps determine whether the values of a variable are spatially distributed (patterned) randomly, clustered (aggregated) or dispersed.

Line 237. See general answer

Line 263. Database surveys were carried out for a variety of different types of projects: public and private building construction, environmental analysis, and natural hazard assessments. Finally, the surveys were conducted over a rather large period of time (about 15 years). These two considerations assure the authors of the statistical independence of the data, although it is not possible to exclude that in some cases it was a single professional who carried out multiple surveys

Line 317. Removed synthetic

Line 344. Replaced crucial with important

CITATIONS TO BE ADDED IN THE TEXT: EC8 and NTC18 texts were removed from the paper. NEHRP added