The manuscript by Fornasari et al. is original and intriguing because it faces different issues, among others the quality of the seismic networks and the origin of the seismic noise. Moreover, it shows the effect of the COVID-19 lockdown in Italy on noise, which was expected but never seen in consideration of the unicity of the lockdown period.

Nevertheless, the writing and the general organization of the work are critical. In my opinion this manuscript needs major revisions to be published. What I found to be critical in general is that the writing is really confusing, there are lots of repetitions, refuses, some useless or appended sentences and different jumps in the order of the communication. The result is that the reader cannot understand what the authors want to say. I would encourage the authors in publishing (especially if newbie researchers as at least the main author is) with a point-to-point revision of the text, but sorry I have no time to do that. However, I try to point out and suggest something as follows:

1. the stations used for the analysis are not very well presented. The authors use RAN for the real RAN but also for RAF and ISNet. I suggest to be very clear in the paper or to figure out something that does not actually exist, a sort of Integrated National Accelerometric Network (INAN). I don't know if this latter it's a good strategy: my concern is that we are not considering the INGV accelerometric stations, so we cannot define "National" the integrated network. By the way, why didn't the authors consider the INGV stations?
   The National Accelerometric Network (RAN) is established to monitor strong motions at a national level which is owned and managed by the Italian Civil Protection Department (DPC). The integrated RAN network is the combination with the following networks; i) the Friuli Venezia Giulia Accelerometric Network (RAF, Rete Accelerometrica Friuli Venezia Giulia in Italian) in the North-East Italy, owned and managed by the University of Trieste (UniTS) ii) Irpinia Seismic Network (ISNet) in the South of Italy, owned and managed by Analysis and Monitoring of Environmental Risk society (AMRA). The term "integrated RAN" has been historically used to define the combination of these networks. INGV stations are not included as a part of the study for 2 reasons, i) seismic background noise of INGV stations is studied quite recently by Antonino D'Alessandro (https://doi.org/10.1029/2020EA001579) and ii) our working group has full access to DPC's database. Hence, we would like to present only the data coming from DPC. However, we are glad to collaborate with other Italian seismic networks in the future to create a more complete background noise models for the Italian territory.

2. Also the number of used stations is not very clear. RAN consists of 647 digital stations, RAF of 14 stations and ISNet of 31 stations. Overall, there are 692 accelerometers, no "more of 700" as reported in line 63. The authors say that they used 528 stations because many of them still run in trigger acquisition mode and this number changed over time. So, what should be clear is: how many stations did the authors use in 2019, 2020 and 2022 for the analysis? Figure 2 is difficult to read (by the way, what do the different colors mean in the Figure?) and it is poorly commented.
   We update the information related with the stations so that it is easier to understand the information related with stations. The RAN consists of more than 700 stations of which 532 provided continuous data in the time range that we are interested in. Over the years we have 241, 325, and 526 stations for 2019, 2020, and 2022, respectively. We updated Figure 2 and it can be seen in below. In the updated version diamonds, stars, and circles represent RAN (IT), ISNet (IX), and RAF (RF) networks, respectively. Colorbar presents the completeness of the data. diamonds represent the RAN network regardless of the completeness values. In the zoomed frames, ISNet and RAF networks are given different colors (star for IX and circle for RF) to explain where the stations are located.

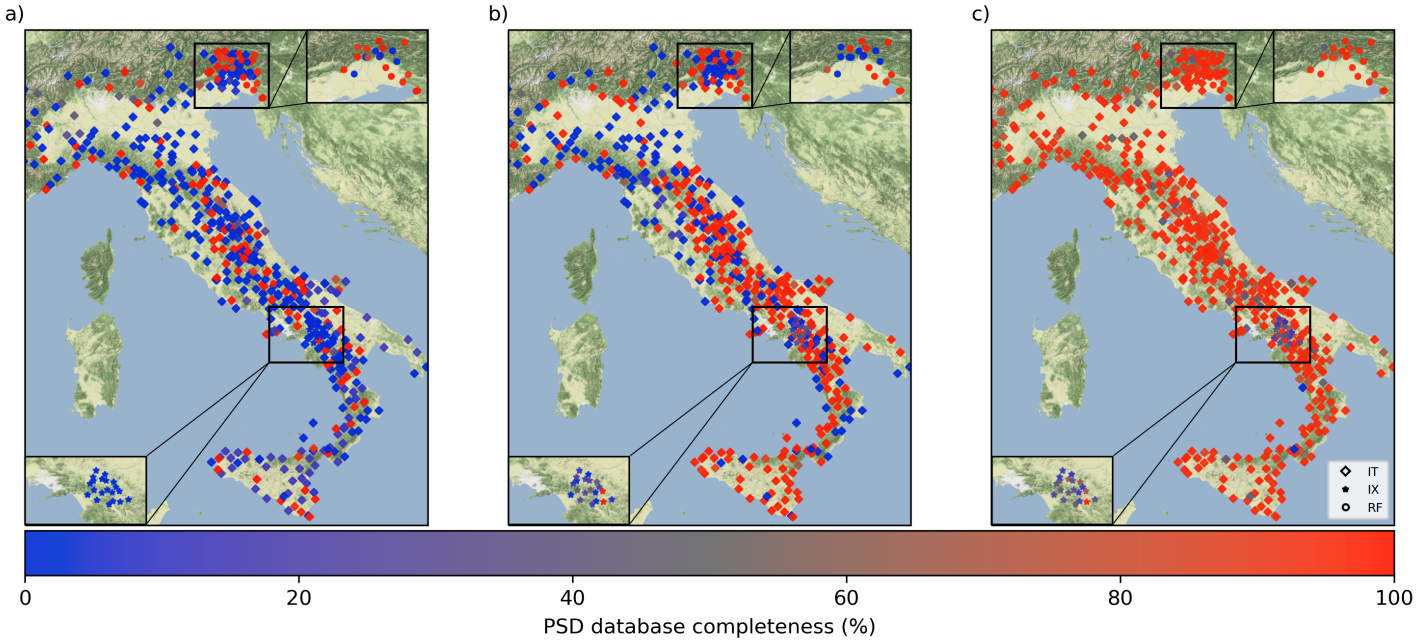

Figure 1: Data availability of the stations in a) 2019, b) lockdown period, and c) 2022. The close up boxes in lower left and upper right highlight ISNet (IX) and RAF (RF), respectively. Basemap data are retrieved from © Stamen Design.

3. Table 1 is also confusing about the stations actually used (715 stations are reported) and in my opinion is useless. A sentence in the text is enough to describe the deployments.
   Number of stations presented in Table 1 are updated limiting the information presented only to the continuous stations, as it can be seen below. We believe that providing this information as a text would be harder to follow. Hence we would like to keep it as it is.

Land Usage	Code	Stations
Settlements	SL	388
Annual Cropland	ACL	48
Permanent Cropland	PCL	12
Grassland	GL	39
Forest	FL	38
Other land	OL	7
Wetland	WL	0
Water	WT	0

Table 1: Land usage at the RAN stations.

4. the paper focuses on the accelerometric stations and the authors show several results for different frequencies (periods). The authors should mention that the accelerometers are not very sensible to low motion, such as noise, especially at low frequencies. Then it's normal that the low-frequency noise recorded by the stations could be out of meaning.
   We agree with the referee. This is why only in Figure 4 and Figure 8 longer periods are briefly mentioned. Periods between 0.1 to 5 seconds are the main interest of the study.

5. the method of computation is merely cited and the differences between the real used method and the standard ones are only reported in Table 2, but not too much discussed. In my opinion the authors should at least present the formula of PSD, the data preprocessing (e.g. instrumental correction, the kind of spectra) but also clearly explain their choices. In other words, the authors should answer these questions: what are the improvements using longer windows and the linear interpolation for gaps? Why don't they use the standard computation for PSD?
   We agree that the method, although standard in this field of study, needs a more in-depth description. We provide a more detailed description of the operations performed. The choice of the analysis window length is within the values commonly used for this kind of study (normally ranging between 1 h to 3 h): as Anthony et al. (2020) noticed, the length of the window became less relevant the shorter the periods of interest are. Our specific choice of using 90 min windows is motivated both by a trade-off between temporal resolution and memory requirements and by the fact that allowed us to perform the computation over the sub-windows without leaving data out (unlike in Anthony et al., 2022).

6. Another point is that there is no word on how the authors took into account the earthquakes or other strong transients that occurred in the time series. Have they been cut off or maintained?
   Transient signal, consisting also of earthquakes, are not removed from the seismic traces since they are low-probability occurrences with respect to ambient seismic noise (McNamara and Buland, 2004, https://doi.org/10.1785/012003001): even though Anthony et al. (2020) showed that earthquakes can affect the noise level significantly for long periods (>10 s), they also concluded that this their effect on shorter periods (i.e., the main focus of our study) is negligible.

7. the Results Section is not very well explained and organized. Figure 3 and then Figure 4 are presented as “the representative noise level” of some (or all) stations. Then there is a (very short) discussion about the day-night, weekend and seasonal variations of noise, variations that can be important. So, I am a little bit confused about what the authors consider “representative”: is it the night level, is the weekend level, or what?
   Figure 3 shows the overall picture of several stations among years. Their medians are presented in Figure 4 (see below). In Figure 3, one can follow the difference between covid lockdown and non-covid time range. Moreover, weekday-weekend differences can be followed. However, it is unlikely to see the day night difference. Day-night, weeekday-weekend, and covid no-covid information are further analysed in multiple figures.
   Day-night variations are presented in Figure 6 and discussed in Discussion section. Weekday-weekend differences are presented in Figure 7. We are only capable of making broad interpretations of the day-night and weekday-weekend variations. For instance, in Figure 6a, there is an overall trend of noisier day. But there are some stations which do not have large day-night differences like the others. Many factors may play role role such as population density, number of residential/industrial places around, car density of the roads nearby etc. Likewise, in the weekday-weekend difference, tourism activities may play role but we do not know the tourism density and their variations over time.
   The word "representative" in line 107 may cause a false interpretation. By representative, we mean for each period that we calculated the value that we get is nothing but noise. As the reviewer mentioned, strong motion recorders may not cover the long period signals properly. This is why in almost all analyses we did, we do the interpretation up to 5 s.

8. Figure 4 is the most important one, and in my opinion deserves more discussion. Unfortunately the figure (and this is the same for all the figures like this) is very hard to be read. In this case, I suggest extracting from Table S1 the first and last 10 stations ordered by the noise level, it could help in the interpretation.
   In Figure 4, the wrong figure was presented by mistake. We changed the figure and it can be seen in the updated version of the paper. We also put the same figure below. Unfortunately, it is really challenging to present all the stations and periods in a single figure and keep the readability high. To overcome this problem, we prepared an HTML version of the figure where zooming is possible. It can be found in the GitHub repository of this study https://github.com/sffornasari/RAN-noise/tree/main/HTML.

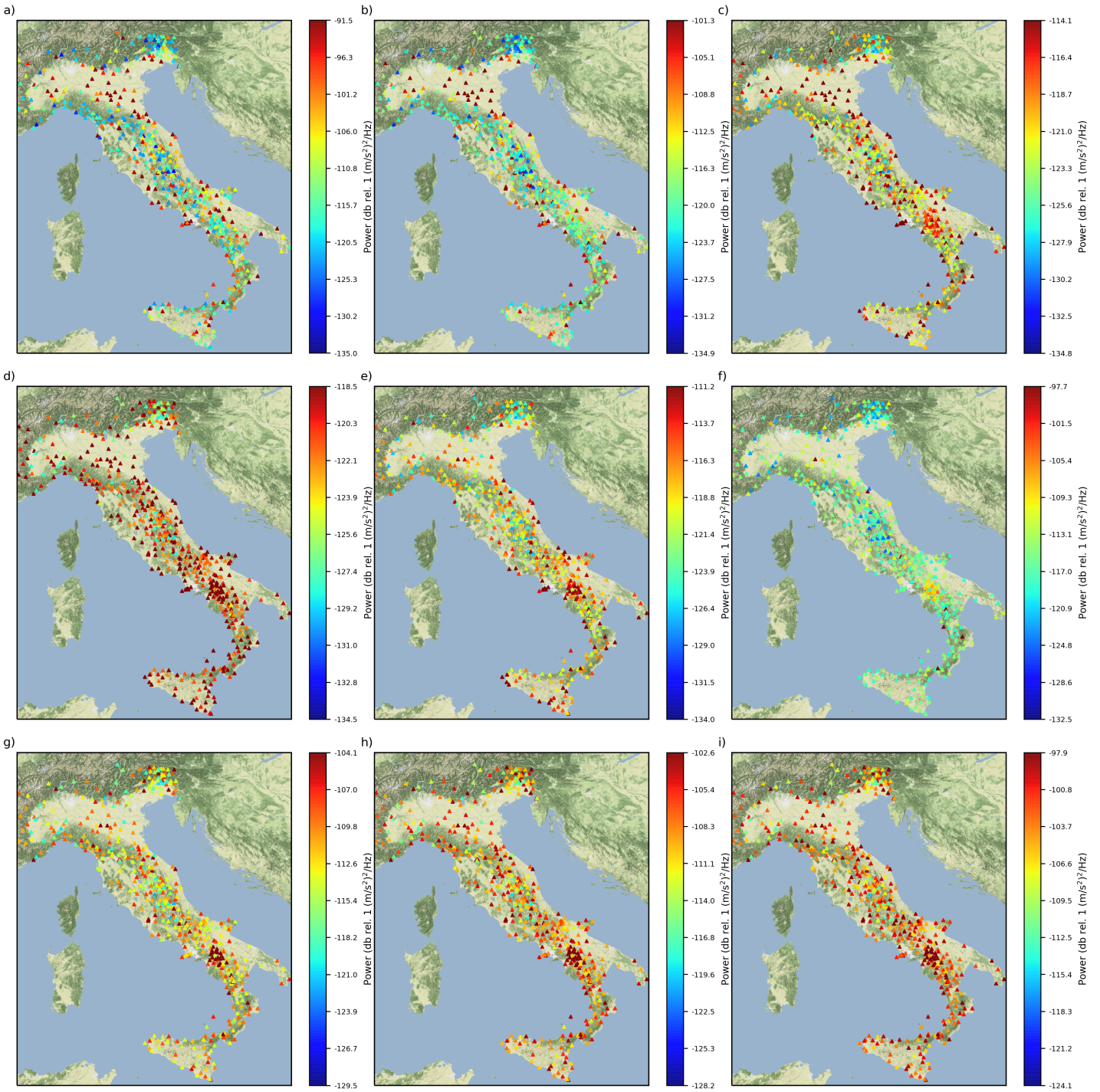

Figure 2: Median vertical component noise maps in one-third octave bands around a-g) 0.1 s, 0.25 s, 0.5 s, 1 s, 2 s, 5 s, 16 s, 32 s, and 80.6 s. Upper and lower limits of the color bar are defined by the model developed by Cauzzi and Clinton 2013. Vertical components are presented in the following figures and Electronic Supplement. Background noise levels of all calculated periods can be found in Figure S1.

9. Another point is Figure 5. What is the meaning of it? In my opinion it's a general overview of the entire "integrated" network but it's not very useful to know that the network is "good" in average.
   PSD probability density function can be considered as a standard procedure for the network operators to understand the quality control of the stations/networks. We would like to present the overall status of the integrated RAN network. This may help readers to compare the integrated RAN with other networks in a 'standard' way.

10. Figure 6 and 7: it's not very clear if the colour scale represents (as for Figure 8) the difference in noise levels between supposed calm (night, weekend) and disturbed (day, weekday, winter) periods.
    In Figure between 6-8 noisier day time, weekday, and winter are represented with red colors. Meaning of the colors are provided in the lower right of Figures 6-10.

11. Moreover, the authors present the PSD analysis for the COVID-19 lockdown period, but it's merely a list of Figures, without comments on them. In general I think that in this section the authors must clearly illustrate and comment on the results.

There is a dedicated section for COVID-19 lockdown period (see Section 5.1 COVID-19 Lockdown). Further analysis related with the temporal and spatial noise level changes related with the lockdown period such as the correlation between the noise level reduction and population and car density, tourism activity and so on but this is not the scope of the study. Hence, we provide a general overview about the effects of the lockdown on background noise. In Figure 9 and 10, background noise difference between lockdown and other dates are quite visible, as expected. To do more detailed analysis, we need many details that we do not have. Hence, we are only able to provide overall results.

12. The Discussion Section (also the others but this in particular) requires a deep English revision.
To overcome some repetitions, we made numerous changes in not only in Discussion but in all paper.

13. The number of analysed “noise levels” is reported as 525(!).
We assume the referee refers the line 146. In this line, it is written that "In 273 stations of 525 noise levels exceed the AHNM developed by Cauzzi and Clinton 2013 (Table 3).". In other words, 273 stations out of 525, i.e. the total number of analysed stations, have at least one period that exceeds the AHNM developed by Cauzzi and Clinton 2013.

14. In the discussion about the sea, swells and/or wind effects on noise, the authors should take into account that the accelerometers are not the best kind of sensor to record these low frequencies (high periods).
In the dedicated paragraph (line 170-175) we mentioned that the long period noises can be associated to the instrumental noise.

15. Lines 173-175 are a good example of what I mean with “jumps in the order of communication” and/or appended sentences. In the long period the accelerometers are completely deaf.
We rephrase the mentioned sentences and they can be seen in below, "As indicated by Cauzzi and Clinton (2013) the main source of long period noises in the case of accelerometric recordings can be associated to the instrumental noise of the RAN stations. As a proof of it, unlike in D’Alessandro et al. (2021) in which the smooth transaction from coastlines to the inlands and mountains are visible, in our study there is no change in noise levels from shores to inland and from high altitudes (Alps, Apenines) to low altitudes (Po valley)."

16. Table 4 refers to a particular station? Or is an average?
It refers to the median noise level changes of all stations for different periods

17. “Changes in the daytime are more significant than the changes in the nighttime between the lockdown and no-lockdown time span.” How do you explain that if the working activity was almost zero during the lockdown?
We interpreted this results as a fact that, being the nighttime already quiet than daytime during ordinary periods, the effect of lockdown measures did not affect the nighttime noise levels as much as the daytime ones, when on the other hand the ordinary activities were severely limited.

18. I think that many comments should be moved to the Results Section.
We assume that these comments are from Discussion section. If we move some of our comments to discussion, this may create a paper harder to read due to separation of the interpretation of our results. If the reviewer can be more specific about the "comments", we can re-evaluate the structure of the text.

19. I don’t really understand the utility of introducing here sub-section 5.2. Maybe, but I am not sure, it could be moved in the Result Section after or together with sub-section 4.1.
As we suggested to the first reviewer, we can re arrange the discussion sections as below,

5. Discussion

5.1 Low periods

5.1.1 Case study: stations located in Trieste

5.1.2 Vehicle noise

5.2 Medium range periods

5.3 Long range periods

5.4 COVID-19 lockdown

20. Also sub-section 5.3 has nothing to do with the Discussion Section, maybe it can be inserted in the Result Section but the authors should introduce the problem.
As we write above, we consider re-organizing the sections of the paper.

21. What is missing is some consideration about the general quality of the sites of deployments of accelerometers. On the basis of the results of this paper, what are the effects on the strong-motion monitoring in Italy? I know that it's a very wide answer, but an effort to answer should be done.
We added a new graph to supplementary material by plotting the 10 most noisiest stations to understand their capabilities on P-wave corner frequencies defined by Brune (1970). Similar critisim is done by the Anonymous Referee 1 and (https://doi.org/10.5194/nhess-2022-258-AC1). This may give an insight of the capabilities of the "worst" stations in terms of background noise. Overall status of the network is explained by Costa et al. (2022, https://doi.org/10.3390/s22155699).