We thank the reviewer for the constructive comments. Here we list point by point the Reviewer's comments (*in italic*) and our reply.

R#1

1. The organization of the introduction section is somehow confusing, the two long paragraphs are rather tedious, making it hard to follow. I'd rather split into short paragraphs, and each short paragraph discusses just one main idea. For instance, the general introduction of 'Moho', the current research status, the main research interest, the choice of method (its advantages compared to CSS and RF) could be separate paragraphs.

-- We have shortened the introduction, and split the text according to the topics, as suggested by both reviewers.

2. line 34, "Anyways ...": this sentence does not seem to connect with the context.

-- We modified the text to connect this sentence with the surrounding text.

3. line 41, "The wide-angle ...": recent seismic tomography studies give pretty reliable estimates of the Moho depth (such as Lu et al. 2018; Lu et al. 2020; Qorbani et al. 2020). I think it would be good complementary info, at least should be mentioned, in spite of the relative weak sensitivity of seismic wave traveltime to interfaces.

-- We thank the reviewer for the suggestion, and added the references in the introduction, together with one more reference from Molinari et al "3D crustal structure of the Eastern Alpine region from ambient noise tomography" Results in Geophysical Sciences, 2020. Anyways the ambient noise tomography used in these works is not a good tool for inferring the presence of impedance contrasts at depth (as the RF and GloPSI are), rather it is good for identifying lateral variations.

4. lines 85-105: could you further clarify the motivations of using GloPSI? I have difficulty in understanding why the GloPSI could provide new info beyond the RF analysis, as for instance, the influence of a complex crustal structure would affect the two methods in imaging the Moho in a similar fashion.

-- In this work our intention is to both: 1-provide new and additional information on the depth of the Moho, and 2- test which similarities and difference the two techniques (RF and GloPSI) retain. Both of them use transmitted waves and are sensitive to the presence of changes in acoustic impedance at depth. There is an important difference: RF only show something when the waves make some angle of incidence with respect to the reflector. GloPSI, on the other hand, only retrieves a reflection when the angles of incidence with respect to the interface are close to zero. Hence, some reflectors would be seen with GloPSI and not with RF and vice versa.

And as we show in the end in this work, both the techniques have poorer resolution when the crustal structure is complex. We thank the reviewer for this point, and we add this in the introduction [lines 105-109]

5. line 116, "Our ...": could you clarify the reasons for using the time range -10 to 80 s around the P-wave onset? This might involve two subquestions: i) why not use S-waves? I would also expect a clear cross-term between S and S reflection phase from auto-correlations, combined with that of P-wave, could help to interpret the final results; ii) why use a long lag time until 80 s after the P-wave onset? Does this mean that the long P-coda also contributes to the recovered cross-term between P and P reflection phase?

--A sentence has been added to clarify that we want to include all the (receiver-side) scattering following the direct P-wave. For most events, until about 80 seconds still reverberations come in, which all contribute to the receiver-side illumination.
Mixing P and S waves is not desired as an P-wave velocity model is later used for migration.
An S-wave implementation of GloPSI is discussed in Frank et al. (2014). Constructing an S-wave reflectivity image would also be useful for EASI, e.g., to image possible presence of melt. However, this is outside of the scope of the current manuscript.

6. line 124, "For ... ray parameter 0 to 0.06 s/km": I do not see the reasons for using such a ray parameter range. I roughly calculated the arrival time for different phases of the receiver side, and it seems to me that the chosen ray parameter range would not help to cancel out the "spurious arrivals", such as cross-terms between P-waves and its later reverberations (depth phases).

--You are correct, the ray parameter range determines the aperture of the virtual source that is constructed. It does not say anything about spurious terms being stacked out or not. In the same section, we add another sentence to clarify that stacking over events with different depths is needed in order to suppress spurious cross terms due depth phases.

7. line 126, "After ...": could you clarify the choice of frequency range? I think the low frequency content will not contribute to the final results since it is less sensitive to the interface due to finite-frequency effects. Moreover, I am afraid it will result in artefacts in the later processing procedures.

-- You are right here. This is an important detail we did not describe well. Additional text has now been added in the Method section. Before autocorrelation we retain a quite wide frequency band (0.04-0.8 Hz). The lower part of this frequency band has poor signal-to-noise ratio and a limited information content on the receiver-side structure. However, these lower frequencies are still effective in obtaining a sharper delta pulse at t=0 after autocorrelation. We can then use a shorter time window to remove the delta pulse and therewith we retain a larger part of the reflections. These lower frequencies are subsequently removed with a high-pass filter with a cut-off frequency at 0.2 Hz.

8. line 127, "spectral balancing ...": could you further explain the motivation of applying spectral balancing? This might also recap comment 7 on the usage of low-frequency content.

-- We added additional explanation on this point. Seismic interferometry is the evaluation of an integral equation. Each contribution in the integrand should have a similar frequency content. If not, the stationary-phase process of enhancing reflections at the physical travel time, fails. Individual earthquakes have largely varying spectral content due to different source properties (corner frequencies) and different propagation effects (elastic and anelastic attenuation). Hence, some spectral balancing needs to be applied prior to seismic interferometry.

9. line 148, "We ...": this might be a fundamental concern: by checking fig. S1 to S8 in the supplementary materials, I have the feeling that the result highly depends on the choice of the pool of events used for imaging. In this way, the results will be more subjective and less convincing.
[now line 172]

-- Indeed the results do depend on the pool of events used, as in any other technique in geophysics. With showing the figure S1 to S8 we wanted to point out the outcomes when using an unbalanced number of events from the two sides (from north and south in this case) of the transect. When we choose events mainly from the south or from the north as in figure S4 and S5, there is no ray crossing beneath the profile, and the spurious phases (generated from the same backazimuthal directions) are not cancelling, rather summing up.
With a high number of events, balanced from both sides of the transect, the receiver side-signals are stacking “positively” and the real features are emerging, as in figures 2 and 4 in the main text.
Since we got this remark from R1, we understand that these concepts are not clear in the main text, and therefore modify the text in lines 174-175 and 184-175.

10. line 165: the use of "clearly visible" is somehow overrated.

-- We understand this remark and change *"clearly visible"* with “visible as blue-red-blue triplet”

11. lines 165-173: It is not clear to me the reasons behind these observations. I guess the difference between crustal features (positive, red) and mantle features (negative, blue) in the BAR image is coming from the low-frequency content in the autocorrelations, as the low-frequency representation of the reflection response. The removal of the low-frequency content leads to the change from a single impulse to bluered- blue phase alternation in DPR image.
[now lines

--We have removed this observations, since they are confusing the reader and since the interpretation needs to be done on the migrated image only

12. lines 180, "We also": I have difficulty in understanding the absolute values of std in Fig 3. If it is std of the amplitude, I would suggest having an additional assessment of the depth uncertainty of this cross-term between the P and P reflection phase associated with the Moho interface.

--The standard deviation is now expressed in % with respect to the maximum amplitudes in the respective panel in figure 2.
--We have added in “Table 2” the maximum and minimum Moho depths estimated by the bootstrapped images, for the distances between 0 and 270 km along profile, that correspond to the reliable Moho depths inferred by this study.

13. lines 194, "This is ...": why the SSR signals are still visible seeing that they are much less constructive than the RSR signals?

--We rewrote the part on SSR, as also replying to the comments of R2.3. There is no physical reason why some SSR should cancel out in the northern part when stacking in 64 events instead of 27 and not in the southern part of the profile.

14. lines 215, "Unfortunately": I would recall comment 3, the results from recent tomographic studies.

--We have added some references to those studies too

15. lines 222, "The suggested ...": I think the GloPSI method has difficulties in imaging the Moho interface (in spite of its geometry) in the presence of a complex crustal structure (see also comment 3). As a consequence, it is hard to conclude that there exists a complex Moho topography. In other words, it is simply not imagined. This might concern the interpretation throughout the MS.

We have modified the text by excluding the Moho topography and referring to the internal crustal structure.