Dear Editor,
we have addressed the points raised by the two reviewers; the main changes with respect to the initial version are the following:

- the introduction has been shortened and re-arranged

- the Method section has been re-written

- within the Results the interpretation of the intermediate processing images has been deleted

- the interpretation of the results has been done on the image obtained by 64 events rather than 27 events, and the Moho has been picked on the maximum amplitude rather than on the lower zero crossing (within the Moho triplet)

-all figures have been modified according to the changes in the text

---

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.

R#2

Introduction
R2.1 Main comment: This section is unnecessarily long and needs to be significantly shortened.
It should also be better organised by clearly separating the presentation of the geodynamic context, the past studies (mostly from CSS and receiver functions) and the questions pending in the area of interest. The first part related to the Moho and Pn/Pg/PmP phases + development of CSS methods worldwide is not necessary. Later on, there are numerous back-and-forth between the presentation of the different seismic profiles, their main results in terms of Moho depths and the geodynamic implications making it very difficult to follow. The focus should be mostly on the available data and previous results in the area of interest, naming along the EASI line.

--The text of the introduction is now re-arranged by separating the geodynamics parts, which are now in the initial paragraphs, and the previous studies, which are following. The digression about the methods and their resolution is cut out, and is partially moved to the Discussion section.

Other comments:

(l19) A proper reference to Mohorovicic (1910) is lacking
--Added

(l19-20) “in seismic records from intra-crustal earthquake” => lacks a verb
--we modified with “produces in seismic records”

(l27) replace “lithosphere” by “domains”
--replaced

(l30) CSS should englobe both refraction and reflection methods
--we cancelled the repetition

(l34) remove “for the Moho topography”
--removed

(l34-35) Why “anyways”? Why unravelling the Moho beneath the Alps is a challenging task (compared to other regions)?
--This sentence is deleted

(l45) Replace “well known” by “typical”
--replaced

(l48 and elsewhere) replace “W” by “west”
--replaced (and elsewhere too)

(l46-50) The sentence is too long and should be simplified
--This sentence has been cut out

(l70) Which passive methods are you referring to? Why do you say that it is more challenging to differentiate the Moho from other interfaces from passive methods?
--This sentence has been cut out

(l71) Why such a sentence about the fact that sources and receivers are at the surface in CSS?
--We cancel this sentence since it is unnecessary

(l71-79) This discussion on migration and 3D effects is unnecessary long, unclear, and I don’t see the link with the present study.
--We cancel this sentence from the introduction

(l80-81) Where should be this “Moho triple junction” or “Moho gap”?
--We mark this in Figure 1

(l86) Define the acronym “RF” used for receiver functions
--Defined

(l90) “The Moho is not imaged”: Add an adjective like “well”
--Added

(l91-92) “inconclusive converted signals by RF” does not mean anything
--changed by weak

(l95-96) I don’t understand what you mean by “turning passive measurements” (namely earthquake signals) “into deterministic seismic responses”.
--The sentence is cancelled

(l96) Which “principle” are you talking about?
--The principle of generating new seismic responses by virtual sources; but the sentence is deleted now from the text.

(l94-98) Unnecessary long
--we have shortened by deleting the sentences outlined above.

(Figure 2): What are the units of the color scales? + Add orientation on top of Fig 2A (and Fig 3A)
-- The amplitude is adimentional; the raw signals are generated by the autorcorrelation of the earthquakes and the final signal for each station is retrieved by stacking all autocorrelation functions.
--We added the orientation on fig 2A and 3A

Data and method
R2.2 Main comment: The presentation of the GloPSI technique should be improved / simplified / better organised. On the one hand, there are unnecessary repetition making it difficult to follow and understand the various processing steps. On the other hand, it lacks more precise information on these various processing steps, how they are implemented and their respective role. For example, the author should remind what is the muting of the delta-pulse and why it is required. Same comment applies for the multiple suppression. Based on Figure 2b and 2c it is unclear to me what is the influence of this processing step (I just barely only see a reduction of the amplitudes between Fig 2b and 2c). Also, the values chosen at each step should be given (for example which filter is applied to remove the “delta pulse”) and the effect of modifying these values on the resulting images can be discussed. For example, it would be interesting to test/show the effect of spectral balancing and be more precise about the way it has been implemented (for reproductivity of the study).

Other comments:
(l113-114) Unnecessary repetition of the fact that you also use direct P waves
(l114) Be more precise about the distance range around 150 you exclude
(l115) Replace P by PKP
(l116) What do you mean by “The 64 events display a high station coverage”? (rephrase)
(l121) replace “result” by “part” + explain what you mean by “muting the delta pulse” (the explanation appears later in the text but should be improved)
(l123) There is no moveout correction performed before the stacking?
(l124) What are the “spurious phases created”? (probably refers to SSR)
(l127-128) The explanation for the “spectral balancing” is unclear. To me the objective is to get closer to the spectrum of a delta-like function
(l130) If the autocorrelation is only applied on the phase spectrum then the spectral balancing (which is performed on the amplitude spectrum I imagine) is unnecessary
(l134) Be more precise about the static correction you applied and explain the technique used to eliminate the surface-related multiples

--Thanks for your detailed suggestions. We largely rewrote this section, adding explanation and details on the processing to allow reproducibility. Also we improved the structure to make it easier to read.

(l139-140) You already presented the Alp01 profile in the introduction
--we have cancelled the description of the profile

Results
Main comments:
R2.3 - The selection of the 27 events out of the 64 available ones is still unclear to me. First, I don’t see major differences between Figure 2 (with 27 events) and figure S8 (with all the 64 events). Secondly, the authors states in the text (l193- l197), supported by the interpolated reflectivity images (Fig. 4a and 4b), that using the 64 events tends to reduce source-side reverberations in the reflectivity images. But later they favour the results based on the selected 27 events (l201 : “Consequently, we decide to use image obtained with 27 events : : :”). Third, it is unclear to me why some source-side reverberations (SSR) should cancel out in one part of the profile (northern part) and not on the other part (southern part). Is there a physical reason for that?

--The fact that there are not large differences between using a pool of 27 or 64 events is a good result. It means that the pool of events chosen is well balanced, both the images are equivalent in terms of results. We have now decided to use the image produced by the pool of 64 events, where the suppression of SSRs is better.
We rewrote the part on SSR. There is indeed 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.

R2.4 - Crust/Moho signature: I don't understand why in the BAR images the crust should (physically) correspond to positive (red) features and the mantle to negative (blue) ones. Is it an effect of the high pass filter applied to the initial data? Why (physically) the Moho would appear as a blue-red-blue triplet after muting the delta pulse? Moreover, such triplet is not always well seen (or with a symmetric shape) like for stations 15 to 25 on Fig. 2.

--Good point. We cannot think of a sensible physical explanation. In the BAR images, the crust appears to correspond to the red part. In the northern part this holds true. However, further to the south where the crust is thicker, it does not hold true. The response at the very low frequencies is largely spurious. We have added the explanation of why we first include these frequencies (to obtain a narrow delta pulse that can be removed without muting much of the reflections response) and later remove them again (because these low frequencies have very limited information content on the crust).
Only after removing the low frequencies, the actual reflections are visible without low-frequency interference. Thus, also the Moho is shown as a positive pulse (red) with small sidelobes (blue).

R2.5 - Phase at 12s for the southern stations (l177-180): I don't understand the argument bring by the author to consider this phase as an artefact for the southern stations but not for the northern stations (this phase is also seen for stations 17 to 30). If it is a source-side reverberation it should appear for all the stations (except if the authors selected some events only for the southern stations : : : which should be avoided).

--We agree with you and remove this text, and we stated that only the Moho is interpreted.

R2.6 - Standard deviation (Fig. 3 and S9): It is difficult to compare both Fig. 3 and S9 as the scale is different. Personally, I don't see a major difference between the images : : : Moreover, the higher std's are observed for time ranges where Moho reflected phases are expected. Therefore, can we really interpret the reflectivity images in this time range? (the authors states at lines 185-186: "The time location of larger standard deviation [: : :] should not be interpreted geometrically" !)

--As there is little difference between figure 2 and S8, there is little difference between figure 3 and S9. We have used the colorscale that maximises the differences in each single plot. The standard deviation is now expressed in % with respect to the maximum amplitudes in the respective panel in figure 2 and S8. The main difference to be noted in these figures is between the northern and southern parts of the profile. The northern part has lower std than the southern part. The northern part has relative higher std for time arrivals (˜8 s) earlier than the times arrivals of the moho signal (about 10 to 12 s). This reinforces our observations that the constrains that we give to the southern part of the profile are poor. We have modified the text.

R2.7 - Finally, the authors say that "it is quite likely that the entire [: : :] signal below the Alps is dominated by artefacts" (l200-201) but later they often interpret several features (Moho, intra-crustal structure) in this area (cf. l206, l215, l222-223 + Discussion part and Figure 5).

We agree that this was an awkward statement. Also below the Alps, a large part of the imaged features is likely real. However, the further in depth one goes, the bigger the chance that imaged reflectivity is spurious, due to imaging remaining complex reverberations instead of primary reflections. These complex reverberations are more likely to exist in the Alps than at other parts of the transect.
We rewrote this part of the text and are now more specific on what could be possible artifacts.

Other comments:
(l144) The Fresnel extension should depend on the frequency used and the depth
--frequency and depth are added

(l145-147) Unclear sentence. Please rephrase
--it is rephrased

(l146) Why do you say that "only the Moho fulfills these requirements"? What about a continuous intra-crustal or upper-mantle reflector?
--We wanted to highlight that the Moho is the strongest first order signature, and it is a global feature [while other reflectors (in the crust or upper mantle) might have a local or regional nature]. Since the sentence seems to be confusing, we rephrase it.

(l156) The fact that the authors have "more" phases available than Ruigrok and Wapenaar (2012) is due to different selection criteria (RuigrokandWapenaar used only M>6 events and PKiKP and PKIKP)

--Indeed, this is what we want to highlight. The interferometry in Ruigrok and Wapenaar could give a solid image with 17 earthquakes only, therefore our application (in both examples 27 and 64) delivers a solid image

(l162) In Figures 2a, S7a, S8a (BAR images), the Moho rather corresponds to the limit between positive signals and negative signals (although I don't really understand why) rather than "a strong positive signal".
(l165-169) Explain the reason why the Moho should correspond to a blue-red-blue feature + avoid the repetitions among the various sentences
(l169-l171)Give a physical reason why the crustal "features" should be positive (red) and the mantle "features" should be negative (blue) in the BAR images.

--We have addressed these three remarks already in R1.11 and R2.4

R2.8 (l187-l191) Various part of this paragraph are unclear ("source of wave energy are the selected earthquake", "the subset 27 events that are closer clustered in space", "the second source denotes the targeted structure", : : :) and should be rephrased.
--We have rephrased this

(l204) Why do the authors choose to "pick the lower zero crossing (within the blue-red-blue triplet)" as the Moho and not the central positive pick?
--We have chosen it simply because it is clearer to see, the error introduced in the Moho depth by picking the zero-crossing or the max amplitude of the red phase is within the errors given by the migration with a velocity model rather than another one. Anyways, in order to be in line with previous studies we now mark the max amplitude of the positive signal.

(Figure 4) what type of interpolation is used? What is the unit of the amplitude color scale? Why is it 5 times higher than on figure 2?
--A bilinear interpolation has been done. We substitute the colorscale with a normalized colorscale.

Discussion and conclusion
R2.9 Main comments: - The authors compare they Moho depth estimates to other studies and challenge these previous results (especially the ones from Hetenyi et al. (2018)). But how sensitive is their migrated image (and corresponding estimation of Moho depth / topography) to the uncertainties on the velocity model they use for the migration?
In the conclusion, they mention a potential anomalously high-velocity lowermost crust beneath the Bohemian massif but it is unclear to me if they have it (or Hetenyi et al. (2018)) in their velocity model and what would be the effect to include/remove it.

--Both this study and Hetenyi et al. (2018) are not including for the depth migration the high velocity layer regionally identified by Hrubcova et al (2005). Hetenyi uses the velocity models shown in Figure S10 (EP crust from Molinari and Morelli 2011). We have added in the discussion some considerations about including this layer in the depth migration.

R2.10 - Both in the introduction and the conclusion the authors mention "opposing" views and geodynamic interpretations of the seismic profile in the area in the literature. In this section it would be good to better indicate which of these previous views are supported (and which ones are not) by the results of their study.

--Unfortunately, the place where the previous Moho estimates are differing the most is where our interferometric image is losing resolution, therefore we cannot support one or another view. All previous models are very precious in order to shed light on the study area, and are adding complementary information about the structures at depth, and we conclude that only a comprehensive 3D reconstruction might shed light on the crust-mantle boundary in the area.

Other comments:
(l260) Change "not supported" by "not seen"
--changed

(l280-ll282) Based on this sentence it is still unclear to me why the GloPSI results differ from the RF results from Hetenyi et al. (2018) between 150 and 300km. Both are based on the same EASI stations. Do you mean that the "lateral velocity variations in the crust" (l282) are not properly taken into account in Hetenyi et al. (2018)'s velocity model used in their migration (as stated before l256-l257)? (l303) replace "Europe" by "European plate".
-- Hetenyi et al (2018) uses the regional Vp and Vs model from Molinari and Morelli (2011) for migrating the Moho Ps conversions. We instead, use a transect specific Vp model (Brueckl et al, 2007) which is much more detailed and reliable for the EASI transect. The RF image of Hetenyi et al (2018) is based on velocity models with a much poorer resolution and the Vs model has much higher uncertainty than the Vp model that we could take advantage of. We have added this sentence in the text.

(Figure 5) Legend of Figure 5 indicates that "the Moho signal disappears where the Moho steeply dips beneath the central part of Eastern Alps" => But if the Moho disappears you cannot say that it is steeply dipping!! Please rephrase.
--Rephrased