Dear Reviewer 2,

Thank you for your comment on our manuscript. We report all your comments in italic below (preceded by "RC2"), followed by our respective answers. The main manuscript is revised accordingly, some of the modified sentences are include below (underlined text).

Reply to RC2

Introduction:

RC2: The introduction is surprisingly short. I miss a proper description of the state of the art (what is known), of controversial issues (what is debated) and of how the manuscript contributes to answering some of these open questions. The introduction does not pose any research questions and thus entirely lacks a motivation for the work presented. I also miss a proper description of previous work done with EASI data and its relevance for the current work. At the end, a brief description of the contents of the paper would be welcome. Instead, the last paragraph rather resembles an abstract.

Reply: We added sentences on actual stage of debates on the E. Alps slab (scientific questions), our motivation for the work and reformulate description of the content of the paper as requested. With that, we consider our description of the classical view and state-of-art of the tomography images of the Alps in the first two paragraphs of the Introduction as representative and sufficient.

While resolution of early tomography of the Alpine region allowed researchers to deal only with the most distinct heterogeneities in the upper mantle, accumulation of high-quality data from the dense AlpArray network motivated us to search finer images of the upper mantle and answers on segmentation of the Alps, dip directions of the subductions, their relevance to European or Adriatic plate, extent of slab delamination and particularly, to elucidate the smaller size heterogeneity beneath the Bohemian Massif north of the E. Alps subduction.

RC2: The meaning of the lines in the tectonic map in Fig 1 needs to be explained in the caption.

Reply: Captions complemented.

Data:

RC2: I assume that also data of the complementary EASI experiment are used. This should be made clear in the text.

Reply: Usage of the AlpArray-EASI data is presented in the 1st sentence of the abstract and in the 3rd par. of the Introduction of the original submission, as well as in the Section Data. Nevertheless we modified the original lines 69-72 of Section 2. Data. It sounds in the revised version as follows:

We collected recordings from stations of the AlpArray Seismic Network (AASN, doi.org/10.12686 /alparray/z3_2015) and AlpArray-EASI network (doi.org/10.12686 /alparray/xt_2014). All the AASN

stations installed, in a 200km-wide band (Fig. 1) along the densely spaced stations of the AlpArray-EASI complementary experiment (Hetényi et al., 2018b), were selected for this study.

RC2: Which procedures for checking data quality have been applied? The paper should be self-contained to some extent.

Reply: We checked data for its completeness and correct timing (uncorrected leap second, failures in clock synchronizations) and station metadata to fix several mistakes (e.g., wrong channel name, station mislocation, poles and zeros given in Hz instead of radians/s).

We have applied several procedures to check the data quality (Vecsey et al., 2017), particularly data completeness and correct timing, to eliminate periods with the uncorrected leap seconds or failures of lock synchronizations. In station metadata files we fixed for example wrong channel naming, station mislocation, and incorrect units for poles and zeros (mixing Hz and radians/s units).

RC2: The description of the picking scheme and the event distributions should be put into main paper.

Reply: We consider this part as valuable for deep specialists and we prefer keeping it in the supplementary materials, which are easily accessible, for those who are interested. We do not want to extend further the main text due to an earlier experience with Solid Earth about manuscript length.

RC2: Which trace is used as reference for cross-correlation and beam forming?

Reply: explanation added

<u>A low-noise beam trace created from stacked cross-correlated and shifted traces of an event</u> serves as a reference in the second cross-correlation step and beam forming in the P-arrival time picking. This means, there is no subjective a priori selected single reference trace.

RC2: Quantitative statements regarding error estimation of the picks and their probabilistic combinations should be made.

Reply: We modified the text, added a new part to Fig.S2 (part S2b, Histograms of uncertainties, means, medians) and extended caption of Figure S2:

The final time of each extreme (green P1, P2 in Fig. S2a) and its error estimate is computed from the normal distribution which approximates a mixture of normal distributions of partial picks.

Figure S2b shows uncertainties of the measured P-wave arrivals, means and medians for both the complete dataset as well as for events selected for tomography (see below).

Each of the red, black and blue picks is complemented by its error estimate defined as a standard error of the normal distribution. Time error of the red extreme depends on a signal noise level (see cyan basins, their height is given by a noise magnitude - red dashed lines), errors of the black and blue correlation picks come from coherence of the signal with the beam. The final time of each extreme (green P1, P2) and its error estimate is computed from the normal distribution which approximates a mixture of normal distributions of partial picks (red, black and blue P1, P2),

RC2: The description of the enhanced data set is confusing. Is the enhanced dataset just the original one with 201 events (in Figure S1 209 events) plus another 43 events within the 60 degree back-azimuth cone? Or is it a completely new selection from the overall available events containing only events within the 60 degree cone? This should be clarified in the manuscript. Regarding the figures S1, **the latter seems to be the case**.

Reply: We stated on lines 86-88 of the original submission "To enhance the resolution in direction of the subducted plates, we selected further rays coming from the northern and southern 60° wide azimuth bins to be included in the tomographic inversion."

Of course, the reason to take only rays from the northern and southern 60° wide azimuth bins was to minimize effects of heterogeneities east and west outside the elongated model. Thus you are right: "the latter is the case". Besides showing the event distributions in Fig. S1, we also include Fig. S5 (original submission, and text there on L90-91), which shows the enhanced resolution along the vertical N-S cross-sections, if all rays are included (left, rays as in FigS1a) and if rays from sides of the model are excluded (right, rays as in Fig. S1b). This means, the original submission has already contained the above recommendation of the RC2. We modify the text and caption of Fig. S5 to prevent misunderstanding:

To eliminate mapping effects of heterogeneities aside the model into its internal part and to enhance the resolution in direction of the subductions, we selected additional rays coming from the northern and southern 60° wide azimuth bins. Only rays propagating through the model within the two azimuthal fans (see Fig. S2) have been included in the final tomographic inversions, tested, discussed and interpreted further. This data comes from 244 earthquakes, each of them being recorded by 120 stations on average, i.e., by 50% of stations in the region.

Figure S5: Velocity perturbations along Profile EASI through models calculated for rays from all directions (left) and for rays from the northern and southern 60° wide azimuth bins (right, see also Fig. S1). Relatively less-well resolved regions along the profile are shaded.

RC2: There are two figures S1 in the supplement. Renumber to S1 and S2.

Reply: Figure S1 has parts a) and b). The second label was missing in the second part of the figure of the original submission, now complemented. Caption and numbering of following figures remain correct.

RC2: Information on crustal structure is taken from several sources. How is the 3D crustal model put together? How are transitions between subregions treated? How are crustal discontinuities treated, as transition regions or as real discontinuities? At least two horizontal sections through the 3D crustal velocity model should be shown. To which depth does the crustal model reach? Does it also contain parts of the uppermost mantle?

How is the correction for crustal structure done? One correction for all events assuming vertical incidence or event-wise corrections taking correct incidence angles into account?

Reply: we compile information for each station and its surrounding, there are no transitions between subregions, individual crust parameters go to Moho depth, i.e., it does not contain the uppermost mantle. The absolute residuals relative to the IASP91 model are corrected for the real crust deviations from the model. For each ray, refractions in the crust according to incidence angle and azimuth are considered. We never applied one correction for all events assuming vertical incidence, but event-wise corrections, which take correct incidence angles into account.

RC2: The meaning of the end of the following sentence (line 100) ".... proper tomographic inversion in our target region to resolve structures, including the spatial limits of our images" is unclear? What are the spatial limits of the images? Do the authors want to say that resolution is also good at the model boundaries?

Reply: Of course, we did not want to say anything like that. We simplify the sentence.

With this approach we gathered a high-quality and uniform dataset of travel time residuals for a proper tomographic inversion to resolve structures in the target region.

RC2: Estimated uncertainties of the travel time residuals should be documented in the manuscript.Show a histogram of the uncertainty distribution and give average and median values.Reply: As above, new Figure S2b with histograms of uncertainties, mean and medianhas been included in the revised version.

METHOD

RC2: It is stated that the matrix W_m in equation (4) provides horizontal smoothing. What about vertical smoothing?

REPLY: There is only horizontal smoothing. There is no option for vertical smoothing in the original code for teleseismic tomography that we use. Both the Telinv and the AniTomo codes have been created without such a possibility in order not to enhance sub-vertical smearing, which is given by the ray geometry.

RC2: Is there also a damping term included in W_m, and if yes, what is the weight of smoothing relative to damping?

Reply: Smoothing matrix forces the model parameters at each inverted node to be close to an average of the model parameters in the surrounding inverted nodes. Such a form of the smoothing matrix is fixed in AniTomo and in Telinv and there is no option to change any weight there. Damping factor epsilon**2 is just a constant multiplying every element of the smoothing matrix.

RC2: With 13 cells in vertical direction, I calculate a model depth of 13*30=390 km and not 435 km as stated in the text. Is there a specific reason for choosing 390 km (or 435 km) as the bottom of the model. Is this depth still warranted by intersection of rays given the length of the EASI profile?

REPLY: The number of 13 refers only to layers of grid nodes, in which the inversion for velocity perturbations is allowed, i.e., at the 60 km, 90 km, ..., 390 km and 420 km. The depth of 435 km corresponds to the bottom of the grid cells that belong to the nodes at 420 km depth. We clarified this in the text. There are still a lot of rays at the depth of 420 km, crossing mainly in the central part of the array. In the northern and the southern parts of the array, the ray crossing is less good compared to the upper parts, of course. See also the new Fig. S7 showing the whole model from the checkerboard test.

The area of about 400 000 km2, centred at 13.3°E 48.5°N, is approximated by 30-by-30 km cell size, horizontally, and with 30 km spacing, vertically. The images are calculated down to 435 km depth on a vertical grid of 30 km spacing. To minimize creating false perturbations, we invert for the velocity perturbations only in the central 5 x 25 x 13 cells, which are well-sampled by criss-crossing rays (Fig. S1), i.e., in nodes between depths of 60 km and 420 km. The model covers the Eastern Alps and a core of the BM, an area of ca. 140 400 km2 in total. Data variance reduction of the final model for the chosen damping parameter attains 66% (Fig. S4).

RC2: What is the damping factor in Fig. S4? Is it epsilon**2 from equation (4)?

Reply: Yes, damping factor is epsilon **2 from equation (4).

Why do the authors plot data variance instead of misfit which is normalized to the data uncertainties. What is model variance? Is it the squared norm of the velocity perturbations? A definition of this quantity should be provided. The authors should provide a value of misfit normalized to the picking uncertainties to get an impression whether there is overfitting or maybe even severe underfitting.

Reply: Evaluation of data variance in tomographic codes AniTomo and Telinv includes normalization to the data uncertainties. We modified caption of Fig. S4. Reduction of data variance of 66 % has been mentioned at the end of Section 3.

Figure S4: Data and model variance trade-off curve evaluated for various values of damping of the isotropic-velocity perturbations and numbers of iterations. The data variance and model variance are squared norms of the time residuals and velocity perturbations, respectively. The data uncertainties are included in the evaluation of the data variance.

RC2: Regarding Fig. S4. I wonder that the authors get increasing data variance and decreasing model variance for decreasing damping factor. I would expect exactly the opposite. This strange behaviour needs to be explained by the authors.

Reply: The strange behaviour of the data and model variances in Fig. S4 is caused by a mistake in manual (re-)colouring of the points of the trade-off curve. The model variance should really decrease

with an increasing damping factor. Thank you for noticing that. We corrected the plotting mistake in Fig. S4.

RC2: Why do the authors only consider 2 iterations? Did they try more iterations and how does the inversion behave then?

We performed one more iteration, but overall the imaged perturbations remain similar, thus we decided to stop calculations after the second iteration to save the time.

Results:

RC2: In the "Data" section two dataset were presented. Which one was used to create the model shown in the figures? Fig. S5 which compares vertical sections obtained from the 2 data sets is only mentioned once. The issue is never discussed in the results or the interpretative section.

Reply: The main features of the perturbations are the same in both models (Fig. S5), but models containing rays from the west and east could be potentially biased from heterogeneities outside the model. Reasoning for limiting rays into the northern and southern fans is expressed in section Data. All results presented, tested, discussed or interpreted are based on this data set. Modification of the text included in Section Data.

RC2: Regarding the vertical sections in fig. 3, I recognize significant perturbations in the gray-shaded upper 50 km of the model which appears to be the crustal domain. As the effect of the crust was subtracted from the travel-time residuals, why do the authors still allow perturbations in the upper 50 km? Theoretically, after crustal correction, the travel-time residuals should represent pure mantle structure and perturbations in the crust should be suppressed. How would the model change if perturbations in the upper 50 km were forced to zero? In particular, what would happen with the model between 50 km and 100 km depth, just beneath the crustal domain? The interpretation of a northward dip strongly depends on the velocity perturbations between 50 km and 100 km depth. Below the high velocity anomalies are rather vertical. Whether HV-EA is delaminated or not, also depends on the anomalies in this depth range. All the conclusions following in the paper about polarity flip, northward dip and detachment of the Eastern Alpine slab depend on this issue.

Reply: Relative residuals corrected for the crustal deviations (including sediments, Moho depth, velocity) relative to the IASP91 represent our tomography input. For each rays, refractions in the crust according to incidence angle and azimuth are considered. First, absolute residuals relative to the IASP91 are corrected for the real crust deviations from the models, then the array-average residual, calculated from the crust corrected residuals of an event, is subtracted from the residuals at each station which recorded the event. The significant perturbations in cross-sections come from our plotting error, which we correct now in all the figures. Because the residuals are corrected for the deviations in the crust, they are assumed to represent pure mantle structure, under the condition that the crustal model is correct. Insufficient corrections map perturbations into the upper 100km of the mantle. On the other hand, an overcorrection can erase or substantially reduce positive perturbations in the upper 100 km (see the text modification in the last part of Section 6). We invert at node depths of 60-420 km (13 grid levels). We do not invert at 30km, neither at the deepest 450 a 480 km. Perturbations related to grid nodes at 60 km correspond to depth range of 45-75km. Therefore, our new images are plotted from depth of 45 km. We do not invert for the crust. The two deepest layers of nodes together with the nodes at shallower depths, where we also do not invert for velocity perturbations, surround the volume studied to stabilize the tomography.

RC2: For a connection to crustal levels, one could plot the perturbations of the crustal model relative to a 1D-model of the crust into the vertical cross sections. But then, what to do with the anomalies that are already there?

Reply: The original submission contains Fig. S3 with information on the crust. We added the Moho depth considered in the crustal corrections into cross-section in Fig. 3. There are no anomalies in the crust, as explained at several places and in our answer to the previous question, we corrected the plotting error.

RC2: What is the criterion for gray-shading in the model domain? Is it derived from the resolution matrix?

Reply: Boundary of the gray-shaded area follows a smoothed contour of diagonal elements of resolution matrix (RDE) equal to 0.15. The contour approximately separate the well and less well-resolved regions in the synthetic tests. This has been specified in the caption.

Resolution Tests

RC2: I would also like to see the results from a checkerboard test (with gaps) which nicely shows lateral and vertical smearing. In particular, vertical smearing at shallow levels should be investigated because it may hide a detachment of the slab or falsely connect the high vp anomaly below 100 km with a shallower one further to the south giving the impression of a significant northward dip.

Reply: We are aware of both the advantages and the disadvantages of the checkerboard test. We complement the Section 5 with the checkerboard test and include in the supplements a new Figure, S7a,b, as requested.

Besides the specific tests described above we also performed standard checkerboard tests to assess resolution capability of the network (Fig. S7- new). The checkerboard test confirms the positive and negative perturbations are retrieved well down at least to 240 km with a weak vertical smearing (Fig. S7a – horizontal slices). Also the vertical cross-section through the central part of the model (Fig.S7b) images the synthetic perturbations reliably.

RC2: The remark that the polarity flip is more or less accepted is certainly a misconception. If I read Paffrath et al. (2021) correctly, they see a rather vertical, detached eastern Alpine slab and favour the interpretation of European provenance because its down-dip length can only be explained by the Tertiary shortening in the Eastern Alps accommodated by south-dipping subduction of European lithosphere. Basically, what is seen in the tomography is only the slab dip but tomography does not tell us the provenance of the slab. Even if it were dipping clearly northwards (in the tomographies by Mitterbauer (P4), Zhao and Paffrath, it is nearly vertical) it could still be overturned European lithosphere. Independent data are needed to decide this issue. The use of the term "polarity flip", however, already implies the interpretation of Adriatic provenance of the slab.

Reply: Tomography images as referenced in our original ms. show steep northward-dipping highvelocity heterogeneity beneath the E. Alps within the upper mantle, down to ~250-300km. We interpret the images, especially in section of synthetic tests, not any concept, and we test the resolution of tomography images we get from available rays. In Section 6, we compare different tomography images, showing the northward dip, though they are interpreted differently. We agree that additional information are needed to decide the issue. We modify our formulations from polarity flip/or reversal to a change of a slab dip direction (as also recommended by RC1) to avoid misunderstanding, though both these terms as used in literature in a broader sense.

RC2: Resolution test 1 also mimics the real pattern quite well below 100 km depth while test 3 misses the increasing dip angle of the real pattern with depth. My remark regarding the treatment of crustal structure also applies to the resolution test. Since crustal corrections were subtracted before inversion of the real data, resolution test data should also be free of crustal contributions and anomalies in the crustal domain should be forced to zero.

Reply: We see substantial differences between the real pattern and that of Test 1 and disagree that the Test 1 mimics the real pattern. Moreover, the test is designed to show that the rays are able to distinguish between one or two heterogeneities (lines 179-180 of the original submission). Resolution test are free of crustal contributions, anomalies in the crustal domain are zero. We correct the previous error in plotting and provide the corrected figures.

Imaging the high-velocity perturbations in different tomography model:

RC2: 2nd paragraph: If I read Paffrath et al (2021) correctly they do not postulate a polarity flip for the Central Alpine slab. They state a steep SE dip of the slab and also do not associate it with Adria. This should be corrected.

Reply: This question touches the same topic as the one raised by C1 concerning lines 222-224. Please see our answer there. We have never associated the Central Alps with Adria.

RC2: In general, I find this section not well structured. The polarity flip issue is discussed in several paragraphs interrupted by a discussion of the 2_{nd} high velocity anomaly underneath the BM. There seems to be no real ordering of thoughts and arguments. This part should be streamlined and rewritten in a concise and non-repetitive way with clear order and structure of arguments.

Reply: We modify this section related to images of the E. Alps slab in different recognized tomography models. The main concerns are depths of the positive heterogeneities beneath the E. Alps and Bohemian Massif, dip directions and their provenance. These aspects are hard to separate completely, which can lead to a "no ordering of thoughts and arguments" opinion. We have improved the structure of this section.

RC2: I am strongly worried by the resolution tests shown in Fig. S7. For each one of the detached slab test models there are strong smearing artifacts reaching up to the surface. The artifacts are strongest in the crustal layer and they mimic an either northward or southward dipping continuous slab. I cannot see that this resolution test proves that a detachment can be resolved by the inversion. It rather suggests the opposite. I wonder how the result of these tests would look like if crustal perturbations were forced to zero. Possibly, all artificial positive anomalies between 50 km and 100 km could become much stronger pretending a dipping high velocity feature.

Reply: We apologize for the earlier plotting mistake, which we have explained in our answer above. The profiles are redrawn to correct the mistake in plotting, which created false perturbations above 45 km, and which we have not considered during the interpretation. There are no crustal perturbations in any figure. Moreover, we include better tests (new Fig.S7) to show whether our tomography is able or not to image the gap above the delaminated (European) slab beneath the E. Alps.

We also add into the main text a new Fig7, in which we mimic the observed perturbations with two (3% and 5%) heterogeneities, the northern one beneath the BM and southern one beneath the E. Alps. In all our interpretations we respect and follow the body-wave tomography principle that perturbations refer to model velocities within each "layer" (grid plane).



P velocity perturbation [%]

East of EASI







Figure 7: Velocity perturbations along the five vertical along-longitude cross-sections calculated from real data (upper row) and those (middle row) calculated from two synthetic 3% and 5% heterogeneities (bottom), plotted over the retrieved perturbations. Green dots mark Moho depths in the model used for calculation of crustal corrections (see also Fig. S3). The along-longitude cross-sections run from 51.65N to 45.35N.



Figure S2b: Histograms of uncertainties of the full dataset (left panel) and of events selected for tomography (right panel). Mean and median of uncertainties, width of quality bins as well as uncertainty of the maximum of picks are indicated by vertical lines. The uncertainties are categorized into five quality classes ranging from 1 (the best) to 5 (the worst). Only arrival times measured with quality 1-3, with the mean 0.08s and median 0.07s, input the tomography.



Figure S7: Velocity perturbations along the N-S cross-section in the center of the array from real data (upper row) and from synthetic data (middle row) calculated for models of the steep detached slab beneath the E. Alps (lower row). The top of the heterogeneity migrates upward (bottom row) from 150 km (detachment as in Paffrath et al., 2021) to 60 km (representing no detachment). The slab detachment larger than the 30km grid would be revealed in the upper 200 km of the EASI-AA model. A potential leakage does not overprint the images.



20 km

45

20

Figure S8: Pairs of horizontal slices (a) and vertical cross-section (b) through the checkerboard model (right images in the pairs) and retrieved perturbations (left images in the pairs), plotted for all inverted node levels. The same mask as in Fig. 3 is used for shading in part (b).



Figure 8: continuation, part (b)