Dear Anne Paul,

we created a file with our comments on both reviews as uploading two files as author's response was not possible…

The original reviewer text is in black, our comments in blue.

Kind regards,

Marcel Paffrath

Review #1:

Review of manuscript se-2021-58

Dear Editor,

The manuscript entitled "Imaging structure and geometry of slabs in the greater Alpine area - A P-wave traveltime tomography using AlpArray Seismic Network data" presents the first P-wave traveltime tomography derived from the entire Alparray experiment dataset. Given the importance of this experiment for the understanding of the Alps, this thing in itself would argue for a prompt publication of this work!

Dear reviewer,

Thank you very much for your detailed review which increases the quality of our manuscript! We tried to response to each issue point by point in this document.

The manuscript is well written, the figures are nice and informative. The structure is ok but I think that the description of the dataset should arrive before the methods.

We prefer not to put the description of the dataset in front of the methods, as we think that the knowledge on how travel times and their errors are calculated (Sect. 2.1) should be provided before quantifying travel time residuals and uncertainties.

The discussion section would also deserve clearly defined subsections.

Okay, done.

I regret that the dataset and its processing is not fully detailed in this document, though I acknowledge that the processing of such a large dataset deserves a dedicated publication. I thus propose (below) to the authors to add some supplementary information to increase the self-consistency of the article. In particular, I would like that the authors describe the quality check performed on the seismograms.

We will add some information on the process of event selection and quality control of the calculated travel time residuals.

The method used is classical (ray theory traveltime tomography) and generally well described. The problem of the crustal model is however addressed in a novel way by integrating a 3D model based on local earthquakes tomography as a starting model for the 80 first kilometres of the model.
The consequence of this approach (but I am not sure of it) however caused me some confusion. I have concerns about checkerboard test vertical cross-section shown in Fɪɢ. 10. The observed smearing in the first 80 km of the output solution is dramatic and, I think, prevent from any interpretation of this part of the final model. The authors state in the text that this is caused by the geometry of ray coverage (coarser in the lithosphere) but another resolution test in Fɪɢ. 15 seems to rule out this proposition (no smearing at all...).
It seems to me that the use of a more or less fixed solution in the lithosphere can explain this discrepancy.

As stated in a short comment to this review already, we are aware that the process of integrating the crustal model as part of the starting model has caused some confusion and should be described more clearly especially with regard to the checkerboard test. There is **no checkerboard perturbation** applied in the first ~100 km of the model below the surface! The anomalies in the first 100 km depth in Fig. 10 are simply the (nearly unchanged) a priori crust + uppermost mantle model. We included the a priori crustal model into the test model to examine its effect on resolution in the mantle below, and to study how possible smearing of anomalies in the uppermost mantle may leak into the crustal domain.
We will extend Fig. 10 by a subfigure showing the (input) test model and one showing the anomalies appearing in the crustal domain in the output model after inversion. Moreover, we will modify the text in the revised manuscript to avoid further misunderstandings.

Maybe there is something I miss but, given the fact that some important geological/geodynamical implications lie in the upper-most 100 km of the model, I think that the authors must clarify this point.

See above

Given the fact that this point can take some time to be addressed I propose a moderate to major revision of the manuscript. You will find below my detailed comments and questions sequentially organised (not ordered by priority). I would like the authors to address those points in a revised version of the manuscript.

Best regards

Questions and comments

1. l. 86: "the AASN constitutes a massive improvement of observational coverage." Can the authors quantify this improvement ?
Yes. Compared to the work by e.g., Lippitsch, our dataset has 40 times as many travel time measurements! We will add a statement in the manuscript.

2. l. 146: Add the reference just after "FMTOMO".
Okay, done

3. Given the fact that there are OBS at approximately 4 km beneath the sea level and stations at more than 2 km elevation I imagine that the topography is taken into account. Can the authors say few words about that?
The model grid is extended by 15 km above sea-level. The topography is considered implicitly by correctly positioning the receivers in the 3D model grid according to their elevation.

4. section 3: I would suggest to present the dataset before the methods.
See above

5. section 3: I know that data selection and processing is detailed in an other article, but I think that the present manuscript has to be self-consistent. Can the authors gather in a single paragraph the selection criterions the have used to request the seismograms; for instance and for now, chosen minimum magnitude is indicated at l. 245 and epicentral distances range at l. 251. Can they indicate if the dataset has passed a quality check (and which one, in particular for OBS), as I imagine that some Mw=5.5 quakes can have low signal to noise ratio ? I also would like the authors to justify their choice of filtering (Butterworth lowpass I guess, poles ?) the seismograms at 0.5 Hz. Given the fact that the majority of the stations are temporary stations, I suspect long-period noise level to be quite high at some sites. A bandpass filtering would have thus appeared more suitable.
Regarding quality check we will add a paragraph as stated above. We also note here that we did not clearly say that we use a Butterworth bandpass filter between 0.03 and 0.5 Hz to exclude long-period noise and will also add this in the revised manuscript.

6. l. 260, FIG. 6: Use the term "mode" instead of "maximum" to define the value that appears the most in the distribution.
Okay, done

7. Can the authors (quickly) explain how they estimate the pick uncertainties.
We explain this shortly at line 123 and below and by equation (4).

8. FIG. 8: Can the authors add the colorbar for panels a-f ?
We will split the figure differently over the two pages and add a second colour bar. Adding one to the current figure would consume to much space vertically.

9. FIG. 9: I do not see the dashed and dotted contour lines indicated in the caption. Indicate or remove the solid and dashed curved that correspond to "faults". Also to be consistent with FIG. 8, it could be

more comfortable for the reader to place slices that cut through the center of the anomalies to the left panels and the one that cut through unperturbed zones on the right panels. Is it possible to add the colorbar on the first part of the FIG. 9 ?

The dashed and dotted contour lines were removed as they overloaded the figures. Text relict in caption was removed.
The problem with the slice depths is that in one case the checkerboard domain (below 100 km) starts with a perturbed layer and in the other one with an unperturbed one. Switching sides would lead to an irregular ordering of slices with depth which would be a lot more confusing.

10. FIG. 10: If I understand well, the upper part of the initial model is composed by a linear combination of two 3D models (Diehl et al. (2009) and Tesauro et al. (2008)) and a 1D model. The +/-5% checkerboard anomalies are thus imposed to this 3D model. Am I ok ?

    No, there is no checkerboard perturbation in the upper 100 km.

    Again, if I understand well, the "crustal" model is designed to be possibly (slightly ? l. 215) modified during the inversion.

    Yes. "Slightly" means that the variations in the crustal domain are forced to remain very small by means of the regularization term.

    I thus wonder if the smearing that we see in the resolution test is not mostly caused by this parametrization and not by the ray coverage. What makes me suspicious is that the zone with the strongest smearing appears to be close to 80 km thick, i.e., close to the 77.5 km thickness of the initial "crustal" model. It is also to note that despite the change in size of the anomalies, the recovered geometry is almost exactly the same above 100 km in both cases. Can the authors comment on this point ? Also, the authors explain that the smearing is mostly due to the fact that ray paths are mostly vertical beneath the array. If this is the case I would have expected the smearing in the lithosphere to be mostly vertical, which is not the case. Can they add an E/W vertical cross-section to see if this pattern is also present in this direction?

    There is no notable smearing in the crustal domain as there is no checkerboard pattern in the upper 100 km! As the reviewer observes correctly, the inversion almost exactly retains the a priori crustal model which was included into the input test model. See also the comments above.

11. FIG. 13–14 and text: If the resolution test shown in FIG. 10 is indeed strongly affected by smearing above 100 km depth, I think that the slice at 90 km cannot be interpreted and should thus be removed. This is particularly critical for the cross-sections in FIG. 14 as the presence the "W" and "D" anomalies are mainly visible above 100 km depth.

    There is no notable smearing above 100 km because the inversion almost exactly retains the initial model there. The slice at 90km depth mostly reflects the (crust and) uppermost mantle model by T. Diehl. Also see above.

12. l. 440: It is, I think, reasonable to interpret high-velocity anomalies as lithospheric slabs on the base of petrophysical and geodynamical arguments but not because of "tradition"!

    That was just meant as a phrase, but we will change this!

13. l. 453–463: Low velocity anomalies are often observed around subducting slabs. Faccenda and Capitanio (2013) also propose that such anomalies are caused by strong seismic anisotropy in the vicinity of the slab. The low velocity pattern in the Western Alps do follow the trend that can be seen in Barruol et al. (2011). This can be discussed by the authors.

    Thank you for the suggestion. We also considered a possible effect of mantle anisotropy by horizontal mantle flow perpendicular to the slab dip direction which is indicated by e.g., SKS splitting. We will include this aspect in the discussion.

14. FIG. 15 and l. 493: I am confused by the results shown in this figure. FIGURE 10 shows that in the upper-most part of the model, there is a strong smearing to the South along line R1-R1'. The profile in FIGURE 15 is, I think, line R2 (on FIG. 13) – E1-E1' is not defined in map view–, I.E. about 1_ to the east of R1-R1'. On FIGURE 15, anomalies in the top 80 km are fully recovered, without smearing and without a significant loss in amplitude. How can the authors explain this contradiction ?

Thank you. The profile annotation will be corrected to R1-R1'. This synthetic test was performed in the same way as the checkerboard test. There is simply no perturbation applied in the uppermost ~100 km for the checkerboard test.

Typos

1. l. 208, 216, 218: Add the years to "Diehl et al." and "Tesauro et al.".
Okay

2. FIG. 15: Center and bottom panels are both noted (d), (e) and (f).
Okay

References

Barruol, G., M. Bonnin, H. Pedersen, G. H. R. Bokelmann, and C. Tiberi (2011). Belt-parallel mantle flow beneath a halted continental collision: The Western Alps. Earth planet. Sci. Lett., 302 (3–4), pp. 429–438. DOI: 10.1016/j.epsl.2010.12.040.
Diehl, T., S. Husen, E. Kissling, and N. Deichmann (2009). High-resolution 3-D P-wave model of the Alpine crust. Geophys. J. Int., 179 (2), pp. 1133–1147. DOI: 10.1111/j.1365- 246X.2009. 04331.x.
Faccenda, M. and F. A. Capitanio (2013). Seismic anisotropy around subduction zones: Insights from threedimensional
modeling of upper mantle deformation and SKS splitting calculation. Geochemistry Geophysics Geosystem, 14 (1), pp. 243–262. DOI: 10.1002/ggge.20055.
Tesauro, M., M. K. Kaban, and S. A. P. L. Cloetingh (2008). EuCRUST-07: A new reference model for the European crust. Geophys. Res. Lett.s, 35 (L05313). DOI: 10.1029/2007GL032244.

Review #2:

Dear Editor,

The manuscript entitled "***Imaging structure and geometry of slabs in the greater Alpine area - A P-wave traveltime tomography using AlpArray Seismic Network data*** " presents P -wave traveltime tomography using the data acquired from the temporary network of AlpArray and permanent broadband stations in the study area. The manuscript is well written and well organized. The network and the analysis of the data set are presented on a related paper in sufficient detail. I think the manuscript has the quality to be published in Solid Earth but at least a moderate revision would be necessary.

Dear reviewer,

Thank you very much for your detailed analysis of our manuscript. Your comments and suggestion will help us increase the quality of our work. In this document we try to respond to each point mentioned one by one.

Below are the details of my comments.

The number of stations used are quite large providing spatial coherence among the stations over the network. On the other hand, the number of events are relatively low (331) and some are lower magnitude events. What is the criteria for using Mw5.5 as threshold magnitude ? How many stations recorded the Mw5.5 events ? There is no information on the S/N of the low magnitude events. If the low magnitude events are recorded by fewer stations, would that create any bias as the average of traveltimes of each event is removed traveltime residuals?

We will write a new paragraph in which we describe the selection criteria for each event (and also for an onset on each station) in more detail. We chose a lower boundary of Mw5.5 because we experienced that the number of available picks for distances $> 35°$ is not satisfactory anymore. However, one could try to lower the magnitude limit in a future work to possibly increase the number of events from the poorly covered azimuths especially when using waveform correlation for travel time determination.
We set a lower limit (100) for the number of onsets required for each event to avoid a bias of the array average. We rather experienced variations in the number of travel times per event related to the number of available stations due to delays in the deployment of the AlpArray stations.

The methodology is explained well and the parameters for the inversion are chosen appropriately. The misfit in Figure 12 drops fast during the first 3-4 iterations while stay almost flat for the rest. But the authors prefer to use the results after 12 iterations for very small improvements on the misfit. It would be valuable to present the results after fast drop in misfit (3 iterations) and after 12 iterations. Is there any overfitting the data by increasing the number of iterations ?

We chose to continue the inversion iterations until the roughness (smoothing norm) of the model saturates. This happens after 12 iterations. Thus, up to the 12th iteration we can still get more detail into the model. After step 12 the smoothing norm stagnates and only the damping norm rises indicating that only the amplitude of the anomalies increases.

We also experienced the exact same behaviour in various synthetic tests, where the resulting model further approached the test model when continuing inversion iterations even if the misfit reduction was small.
There should not be any risk of overfitting of the data, as the final misfit of ~3.2 has not yet reached the limit of 1. Below this limit, we would start fitting the uncertainties of the travel time residuals into the model (overfitting).

The checkerboard tests are done but requires more, to present vertical and horizontal smearing in more detail. The authors state that the checkerboard anomalies are smeared at least 20 km at shallow depths below the crust. But they do not provide any information over rest of the domain. The initial checkerboard depth models should be provided together with the recovered patterns (Figure 10). Spike tests would be valuable to monitor smearing over the solution space.

We describe resolution capabilities of the model with depth when we evaluate the checkerboard results for each depth slice in section 4. We also show representative vertical slices though the recovered checkerboard (Fig. 10).
The used model is a combination of a classical checkerboard test and a spike test as there is unperturbed space in between the checkerboard tiles (which can be seen as kind of spikes). Therefore, we can monitor smearing effects quite well with this model and decided not to do additional spike tests.

The incorporation of the crustal model and upper mantle (<100km) into the inversion is a nice idea although more tests are needed to understand the influence of strong constraints on the velocity perturbations. The results after inversion should be compared to the initial model for depths < 100 km. Is there any bias on the transition from strongly constrained upper part (>100km) to unconstrained lower part ?

We will try to highlight the influences on the uppermost ~100 km when we update the vertical sections through the checkerboard to clarify the way we created the checkerboard model. There it also becomes visible in which way artifacts can occur within the weakly constrained parts of the crustal model.
We do not understand the last question regarding the bias on transition from strongly to unconstrained parts.

The authors compares their results to the previous teleseismic tomographic works. It appears that Zhao et al 2016 used lesser amount of data and attained similar resolutions. It would be nice if the authors display vertical cross sections of few profiles from the tomographic images of the previous works (e.g, Koulavov et al, 2009; Zhao et al 2016 ..) crossing the same structures.

We will add a comparison to other models that are digitally available in the same spatial domain.

The presentation of the 3D model in Figure 16 is the least satisfactory part of the paper. The Figure 16 does not make any impression neither as geology nor a velocity model. Depending on the level of velocity contours a different image with different size of low velocity holes and slab thicknesses would appear.

The main idea behind Fig. 16 is to demonstrate the complexity of the model results and to get an idea of the three-dimensionality of the different features. There is of course always some

subjectivity involved, as what we see in the figure depends on the threshold value of the shown iso-surfaces. On the other hand, we learned that there is also a strong bias involved when 2-D slices through a 3-D model (or slabs) are presented, as perceived features such as a dipping direction or size will always depend on the exact profile positions and a natural slab does never follow only one direction along a single (or multiple selected) profiles from start to end as an idealized one would do. We experienced that a repositioning of a profile by less than ~50 km at only one end may significantly change the perceived geometric structures.

We should maybe clarify in the manuscript that the decision for a threshold value for the iso-surfaces is subjective and that the 3-D geometry changes for different values. Still, one should keep in mind that when trying to interpret idealized, discrete slab "boundaries" from a seismic velocity model, these are also often based on interpreting isolines (or iso-surfaces).

It appears that some of the research questions the authors posed in the introduction such as "how thick and how long are the descending slabs ? " remained unanswered.

We discussed the penetration depth of different slabs (and with that their lengths). We will try to answer questions regarding slab thickness in the discussion as well. However, it is difficult to make more than educated guesses on the slab thickness, as there are no discrete boundaries of a subducted (heated up) slab anymore, the further it is subducted, as its thermal signature blurs with time. Also, the discussed smearing effects, make such guesses difficult.

A geology based velocity model derived from the tomography can be used as synthetic test to better constrain the slab thicknesses, the extent of the low velocity zones.

The 3-D structure we show here is directly derived from the data and only depends on the chosen value of the iso-surfaces. For a geological interpretation of the velocity model (positive and negative anomalies) we want to refer here to the paper of Handy et al. (2021).

References

Handy, M., Schmid, S., Paffrath, M., Friederich, W., and the AlpArray Working Group: European tectosphere and slabs beneath the greater Alpine area – Interpretation of mantle structure in the Alps-Apennines-Pannonian region from teleseismic Vp studies, Solid Earth Discuss. [preprint], https://doi.org/10.5194/se-2021-49, in review, 2021.