Interactive comment on “Relocation of earthquakes in the Southern and Eastern Alps (Austria, Italy) recorded by the dense, temporary SWATH–D network using a Markov chain Monte Carlo inversion” by Azam Jozi Najafabadi et al.

Edi Kissling (Referee)
kiss@tomo.ig.erdw.ethz.ch
Received and published: 20 January 2021

Dear Prof. Kissling,

We would like to appreciate the time you have invested and your insightful comments on the manuscript. We take your comments very seriously and fully appreciate them during the revision of the manuscript. In the following, please find our response to your thoughtful remarks as blue texts.

The manuscript regards the compilation of a local earthquakes catalog of 16 months period with the application of a few modifications and improvements to standard location procedure using the dense AlpArray and SWATH-D temprary station network. The study comprises different topics-procedural steps and results- each of potential interest to a wide range of readership.
Interesting enough the first such topic addressed in the abstract is the description and attempted correlation of the seismicity with the regional geology and tectonics. While certainly precisely relocated, the 344 local earthquakes of a 16 months period by no means could be taken as representative for the seismicity in the region and it should not come as a surprise - and not be seen as regional "characteristics“- that it appears in clusters. For a seismotectonic interpretation linking such observed clusters with tectonic faults to conclude, f.e., that "the general pattern of seismicity reflects head-on convergence of the Adriatic indenter with the Alpine orogenic crust.“ one would have hoped the authors to take advantage of the great data set with on average 36 P observation per event to complement the hypocenter locations with focal mechanisms at least for the larger magnitude events.
We fully agree that for an in-depth seismotectonic study, focal mechanisms are essential. This issue was covered in other projects within the SPP. A manuscript dealing with focal mechanism was submitted to Solid Earth (Petersen et al., 2021), which we now reference in our manuscript. We also added some more information about the mechanisms in some sub-regions. Please see our response below.
Furthermore, a comparison and thourough discussion of the relation of the presented high-precision short-period seismicity with the long-term seismicity pattern reveiled by the official catalogs over the past 30 years is not only possible but necessary.
The general pattern of the seismicity agrees with previous studies, e.g., Reiter et al., 2018, which we mentioned in the section “Discussion” of the manuscript. Furthermore, in the discussion section we compare the seismicity distribution of our study with those from previous studies in several sub-regions (based on local networks or national agency data). In the updated manuscript we added some sentences regarding the similarity of the seismicity pattern to long-term seismic catalogs such as the SHARE catalog as well.
The main topic and work of the study regards the successful application of a Markov chain Monte Carlo inversion of the 12,534 P and 7,258 S observations from 344 local earthquakes to

obtain a 1D velocity model and station delays for the region that allows high-precision hypocenter locations. The derivation of the model is well explained and complemented with the description of a synthetic test to provide a statistical estimate of the location uncertainties. In addition, a "ground truth“ location experiment with quarry blasts is presented and discussed. This part of the manuscript is very clearly presented and contains a lot of technical details that allow the interested specialist to follow most steps. Considering the readership that might be interested in the seismicity and their tectonic interpretation though, I suggest to most of the chapter 5 could be moved to the supplementary material. What is missing, however, is a critical discussion of the results and, in particular, their relevance and meaning for the seismic catalog of 344 events presented. Considering that with the Markov chain MC inversion the authors address the coupled hypocenter-velocity model problem for the complet (very high-quality in terms of number of observations per event) 344 event data set, I do not understand why there is no mentioning about the internal consistency of the hypocenter solution or about the great potential of these results as initial data (hypocenters and model alike) for 3D seismic tomography. Rather, the list of relocated earthquakes is presented simply as a higher-precision- "than INGV/ZAMG“ catalog for the region.

We agree with your comment regarding the internal consistency of our hypocenter solution and the ultimate goal of creating this dataset and therefore, we added the following text to the section “introduction” of the updated manuscript:

“The dense, high-quality travel-time picks created in this study potentially lead to constrained hypocenter solutions with high internal consistency. This will enable us to identify the general pattern of seismicity on the surface and at depth throughout the region and contribute to the understanding of active tectonic processes. A further aim of the study is to derive a high-quality dataset suitable to be used in Local Earthquake Tomography (LET).”

A similar text is in the updated section “conclusion”.

We totally agree that the study provides a very well-suited dataset for further studies such as local earthquake tomography. We emphasize this aspect now more clearly throughout the text. In chapter 5 we intended to conduct a synthetic recovery test. With this test, we can study how (well) the synthetic hypocenters are recovered by the inversion routine and how the 1-D velocity (output) looks like the 3-D model (input). Also, how shallow velocity anomalies are “mapped” by the station-corrections and whether the input random noise is recovered by the McMC method. We think that this is an interesting test for showing the performance of the method, especially when keeping in mind that we apply a relatively novel McMC search. We report the outcome of this test (please see also our response to your point number 17). Similar tests are regularly employed in the inversions for (3-D) subsurface structure (i.e., tomography). Therefore, we decided to keep chapter 5 in the main text of the manuscript. Nevertheless, we modified the text in the updated manuscript and hope that the reason behind this test is now better clarified.

Finally, the study also contains a section interesting for seismologists (observatory tasks and seismic tomography) about the semi-automated picking of such a large data set. Much of the description of this work part is already allocated in the supplementary material and it should remain there. It would be logic though if the reference to this work and the presentation and discussion of the results would be appearing before and not after the Markov chain MC inversion of the data set (see Figure 7). Furthermore, some details important for the specialist are missing (see individual points 10 to 15 below).

In section 3.2 of the updated manuscript, we added a table and histograms (table 1 and figure 4) regarding the results (travel-time dataset) of the semi-automatic picking procedure. Further detail about the event list from national catalogs is now added to the updated manuscript as a new appendix A and it is referenced in section 3.2.

The chapters 6 (Results in discussion) and 7 (conclusions) read like they were written by different people with totally different interests and perspectives and the only connection between the two parts are the 344 precisely located hypocenters. There is no geologic-lithologic interpretation or at least comment about correlation presented between the other results (notabene of great importance for the claimed reliability and accuracy of the hypocenters) of the coupled problem, the velocity model and station delays. The tectonic interpretation of the seismicity presented in chapter 6.3 (pages 21 to 25) is missing taking explicitly into account (and explaining to the reader how and why this is used as arguments for the interpretation) the great advantage of this study (having an event data set of high internal consistency and high precision hypocenter location of quantitatively known uncertainties) and the significant limitations (pre-selected events of unspecified magnitude of completeness and only 16 months of observation period).
Our intention is to keep the conclusion (and in general the manuscript) more technical than interpretational (in terms of seismotectonic). However, we rephrased the whole conclusion in order to have a reasonable balance between all chapters of the manuscript. In this regard, we rephrased the abstract as well so that our main aim of the manuscript is better conceived.
In consideration of the above general remarks on quality and deficiancies , I suggest moderate revision of the manuscript before publication.

Specific points:
Below are our responses to your specific points:

(1) Line 12. Replace "accuracy“ with "precision estimate“. Note that in line 13 you correctly assess the "accuracy“ with the blast location test.
Agree – changed

(2) Line 15. Delete the rest of the sentence after:“ ..1.7 km in depth.“
Done

(3) Line 27. Replace "accuracy“ with "precision“
Agree – changed

(4) Line 48. "... has the advantage of being“ largely "independent …“
Done

(5) Line 51. "best model(s).“ a note on ambiguity would be useful
This part is rephrased to:
“Moreover, the results can be statistically analyzed, and thus errors and ambiguities can be estimated. The method extends the probabilistic relocation approaches (Lomax et al., 2000) by inverting for a set of velocity models well explaining the data. “
However, more details are provided in the method chapter of the manuscript

(6) Line 64. Please outline Adriatic microplate in one of the Figures.
Done

(7) Lines 67-75. Needs a figure to show the strain if you keep the introduction as is and the chapter 6.3.
We think that we have already provided all references to our statements regarding the strain (mainly in section 2). Nevertheless, now the companion paper by Verwater et al. 2021 is

submitted (and accessible) in which this topic (including a Figure) is covered in more detail. Therefore, we prefer not to add an additional Figure to our paper. We add a reference to Verwater et al. 2021 to Section 2.

(8) Figure 1. See point 6 above. Red box in bottom figure does not correspond with bounds of upper figure. This figure is not providing all necessary tectonic information mentioned in the manuscript. You should note that the seismic catalog presented by ISC is by far not complete down to magnitude 2. If you want to show the big picture use either EMSC catalog likely complete to M3 or ISC likely complete to M3.5. Otherwise you could use a composite of the various national catalogs that probably are complete to M2.5.
The figure bounds are corrected.
All the information which is mentioned in the text is now added to figure 1.
The big picture of seismicity (Figure 1) is now taken from EMSC with M3.

(9) Line 85. Actually there are earlier catalogs that were compiled: European Geotraverse Blundell et al. 1992, Solarino et al. 1997
Agree – two references are inserted in the text.

(10) Lines 115 to 119. You need to elaborate in detail (this can be done in supplementary material but it is absolutely necessary to have this information) how you identified the "common“ events and how in the end you established the event list of the 2619 events.
An appendix A with the information regarding how the event list is formed is added to the updated manuscript.

(11) Lines 120 – 126. The discussion of the results of this semi-automated picking (that is well described in suppl.) needs to be more extensive and detailed. On what basis did you define the selection criteria (gap<200 , why not <180') (RMS <1s), why no mentioning of number of P obs? Did you check all 12534 P obs manually?
We used the origin-time of 2,639 local events from national catalogs (now in appendix A) as initial data to start the automatic picking with. However, we don't have any information about the precision of these events beforehand. Therefore, in the beginning, we decided to select as many events as possible from the national catalogs, apply an automatic picking procedure for these events, and then make a further selection based on our own picks and location information. After the automatic procedure (2,639 events, 68,099 P- and 17,151 S-Picks), we noticed that many of the events are either on the periphery of the network, too weak, or too noisy to be detected by more than 5 stations. Moreover, we believe that the automatic picker missed some of the good picks or introduced some suspected picks. Therefore, we had an early selection criterion (not too conservative) in order to do a manual/visual inspection of the picks. These selection criteria (gap<200 and RMS <1s which gives 384 events with 18,390 P- and 7,762 S-picks) were applied to the automatic picking results. We manually/visually inspected the P and S picks of these 384 events, with careful consideration for picks in the distances of the triplication zone and farther than that (we inspected a very large amount of the picks from these 384 events, maybe 90% of them). We explain later in text that from these 384 events, some are blasts, and some are unclear to us. Later on, this dataset was further selected for simultaneous inversion (301 local earthquakes with gap<180° and minimum 10 P-picks and 5 S-picks).
We added the statistics about the number of picks and their quality classes of 2,639 events after automatic picking to Appendix B (Table B1) of the updated manuscript. Detailed

information of the selected dataset for the simultaneous inversion is now given in the main text (section 3.2; table 1 and figure 4) of the updated manuscript.

(12) Line 134-6 and Figure 3. The Wadati diagram shows significant numbers of observations with +/-3s residual relative to constant Vp/Vs ratio. Note the the Vp/Vs ratio varies within the crust and at Moho. You may see this in the Figure as the straight line is systematically shifted onto the side of the highest point density after about 25s P travel time. What residual range do you define as corresponding to the sum of 3D, lithologies and regular Gaussian observation uncertainty effects and what value denotes an outlier?
We agree that the points are systematically shifted to above the straight line indicating that the Vp/Vs ratio (slightly) increases at larger distances. This could be related to different (average) Vp/Vs ratios in the crust and upper mantle (which is e.g., also suggested by global models such as ak135). We mention this now in the text: "We notice that at P travel-times larger than ~25 s the observations tend to slightly larger S-P travel-time differences, potentially indicating a higher Vp/Vs ratio at larger depth, i.e., in the upper mantle."
According to own synthetic tests (using the source and receiver geometry of this study and including (1) a Moho topography based on a simplified Moho from Spada et al., 2013 (see section 5), (2) assumed (moderate) differences in average Vp/Vs ratio in the crust and upper mantle, (3) crustal Vp/Vs ratio anomalies which we expect for the region (e.g., Vigano et al.,2015) and generally from other regions of the world, and (4) picking uncertainties) we expect a scatter of no more than ± 2 to 3s around the linear trend of the points in the Wadati diagram. We would like to emphasize that we only considered values with $t_s - t_p > 0.72\ t_p \pm 4$s as "outliers" and removed them. These outliers were only 0.3% of all observations (which should not make a difference even if they were incorrectly flagged as "outliers"). We state this now more clearly in the text. After this removal, only 2% of the whole observations have ts – tp > 0.72 tp±3 s.

(13) Line 144 and Figure 4. "Checking the phase-type is extremly important." I fully agree and record section display is a good zero-order approach. However, I do not think your figure 4 is of help for doing this. How realistic is the ak135 global model for P phase identification considering the Moho topography by Spada et al. 2013 (your Fig.5c) or the 3D LET model by Diehl et al. 2009 but most important the literally more than a dozen refraction seismic lines that have been published (for a review see Kissling et al. 2006).
We agree. The comparison of the cumulated picks (from earthquakes with a variety of different earthquake depths) with the bounds of the synthetic traveltime curves (from a global model for different earthquake depths) was too coarse, too unspecific, and obviously not helpful. As we mentioned in the text (both in section 3.2 and in the appendix), we visually/manually inspected the waveforms of all earthquakes and their automatically determined arrival-times, and added, deleted, or modified picks where applicable. This inspection was done mainly on individual traces and was most helpful for identifying wrongly picked noise burst etc. Arrivals of stronger events, for which we expect XmX or Xn arrivals, were additionally assessed with the help of record sections (epidistance plots of individual earthquakes; based on preliminary locations). So, we are quite confident, that we identified most of the phases correctly. Nevertheless, we acknowledge - and are very much aware of - that even these detailed visual/manual checks cannot prevent some amount of wrongly identified picks in our dataset (a general difficulty for manual picks e.g., already pointed out by Diehl et al., 2009). We state this now more clearly in the text. In this context, we think that Figure 4 of the original manuscript is indeed not helpful for the manuscript and thus we decided to delete it, as we deleted parts of the text of section 3.3 in the original manuscript (some parts were moved to the newly formed section 3.2). In any

case, we would like to emphasize that we never used Figure 4 for identifying and removing individual picks (potentially misidentified).

(14) Line 152. Please provide clear evidence and explain in detail strategy to identify PmP phase by using a totally inadequat 1D model.
We agree, based on this plot we are not able to do this. We deleted this statement (together with the whole paragraph) – please see our comment above.

(15) Lines 152/3. "the number of outliers, ..., is not significant to the total number of picks." What value do you define as being an outlier? (analog question to point 12) Note there are dozens of observations +/-2s from the main intensity of data points (that by the way is totally off your Pg line) and that individual hypocenter location precision (and even more important for accuracy) is in truth measured by the fit of just those observations that refer to the specific event.
Yes, we agree on this point. Our discussion of "outliers" was not specific enough. Similar to the phase-identification issue (see our comment above), we are very much aware that our dataset probably still contains some number of "outliers" (howsoever defined) and we stated this now more clearly. However, as we also state above, we did not use Figure 4 or any related (automatic) mechanism for identifying and removing individual picks (potentially misidentified). We have just used it as a – probably too simple – statement regarding the existence or non-existence of "outliers".
As stated above, we deleted the whole corresponding paragraph (and Figure 4) and modified the text.

(16) Line 218. "... does not depend on initial hypocenters, ..." I seriously doubt this (does not depend) and suggest to phrase it differently. Consider how you would identify an outlier with Figure 4 if you do not have a rather good idea the initial hypocenter! Furthermore, consider that you were using a priori information from existing catalogs for your semi-automated picking and that even with all this information you apparently found mispicks and had to select the 384 best events.
We only partly agree. We are still convinced, that our inversion method itself does not depend on the initial hypocenters (as is the case for the traditional inversion routines using a linearized approach, DLSQ inversion, etc.), see Fig. 8 in Ryberg & Haberland, 2019. However, we agree that some dependency might be introduced through the selection of picks (e.g., "outlier removal"), especially in the case of automatic picking (involving e.g., sequential intermediate location steps based on initial velocity models). Nevertheless, we think that even this influence is minimized because we are using a visually inspected pick-dataset.
We modified the part in the following way:
"Therefore, the McMC method only uses the travel times and does not directly depend on initial hypocenters, origin times, velocity models, or even the model parametrization (e.g., grid node spacing). Nevertheless, because in the (semi-automatic) picking procedure the selection of picks is involving the comparison with travel times based on preliminary hypocenters (which depend on initial velocity models), the results of the inversion depend *stricto sensu* to some degree on initial values."

(17) Line 226. This does not provide an "accuracy" estimate! May be internal consistency, precision.
Agree - It is changed to consistency.
And you need to provide reasons why this should be expected to be of relevance for the real individual

event locations.
The event locations (and the velocity model) derived by the inversion procedure depend on a large number of parameters such as the quantity and spatial distribution of earthquakes and receivers, noise, quality of travel-time readings, the class of the model used (in the inversion) and so forth. Furthermore, the test is not only interesting in respect to the hypocenters but also to the derived velocity model. We think that it is appropriate to test the recovery by synthetic tests as it is standard for the recovery of subsurface structure (e.g., in LET studies). Since the McMC approach for solving the coupled hypocenter-(1-D) velocity problem is quite novel we think that this kind of test is particularly interesting. We already described our reasoning for conducting the test in section 5, however, we added/modified the text slightly. Furthermore, we added some sentences to the discussion of the results of the test.
"Comparing input event locations (synthetic) and the inverted (output) ones allows us to study the recovery of the hypocenters, location consistency, and potential systematic errors related to the use of a 1-D model, which we can generally expect for the derived real hypocenters. For example, it can be studied whether events at the periphery or in certain parts of the model have systematically larger uncertainties (e.g., due to their location and/or spatial distribution of picks). Furthermore, we can study how the (output) 1-D model looks in comparison to the (input) 3-D model, how large the derived noise is in relation to the synthetic input noise and how the pattern of station-corrections corresponds to the shallow velocity anomalies. Similar tests are standard in structural studies (i.e., LET) to study the recovery of certain features. "
By the way, your blast test shows otherwise!
We think that the simultaneous/joint hypocenter relocation is a very powerful concept. We agree that our blast test indeed shows very good performance of this test.

(18) Line 230. This statement about Moho velocities is simply wrong. No 7km/s velocity has been reported in the Alps. Please check the literature.
While the reviewer is certainly right regarding the reported seismic velocity values in the Alps, we would like to point out that this is a synthetic test with the main aim of checking whether and to which extent a 3-D velocity variation (i.e., Moho topography and shallow velocity variations) influences the recovery of the hypocenters (and which 1-D model is derived). We think that the exact values of the velocities are not that important for this test as long as they are in the "usual" range. We used also for example an angular shape of the Tauern Window and the sedimentary basins which is also not close to reality. We never claimed that these are reported values, it's a description of our synthetic model. Therefore, we prefer to leave the synthetic model as it is. Nevertheless (and in response to issues 17 – 21), we modified the whole paragraph describing the synthetic model.

(19) Line 232. The 19km are just along the flank of the Ivrea body and not relevant for the Moho topography beneath the Po plain.
Agree - changed

(20) Line 233. -5km is much to high and you are doing something wrong if you need to avoid rays through air by such model top elevation. Note that the increase in pressure within the earth causes a velocity increase with depth and the seismic waves to show a downward curvature You should use the average station elevation for ray tracing.
During the McMC inversion many different – also weird and unrealistic - velocity values, i.e., also models with a velocity increase with elevation, are (randomly) tested. This is totally different e.g., compared to velest. We are very much aware of the fact that usually on Earth there is a velocity increase with depth. These unrealistic models obviously have to be suppressed

because otherwise, we obtain a somehow “mirrored” (artificial) velocity model above the surface (in our case, only models without low-velocity zones are accepted). Furthermore, the method does not use rays (or raytracing) at all but instead calculates the travel time field with an FD Eikonal solver. More information can be found in Ryberg & Haberland (2019) (exact reference in the bibliography of the manuscript).

(21) Line 245. Choosing a constant Vp/Vs ratio of SQR(3) is very problematic as we know it is wrong because it varies and likely the average is different.
This is a synthetic test, and the goal is mainly to assess the performance of the method to recover the earthquake locations when using a simplified 1-D velocity model in the inversion. We agree that the reality is more complex and maybe our test is not complete and might not cover all aspects which should be checked for. In Ryberg & Haberland (2019) a much simpler recovery test with a 1-D synthetic model was performed, now we wanted to add some more complexity (although not claiming that this is the ultimate test). We modified the section describing the synthetic model (please see also our response to issues 17 – 20 above and 24 below).

(22) Line 250/1. Not only refer to table but provide correct value here.
Done

(23) Figure 5. Your model extent in Figure 5a does not correspond with your map extent in Figure 5c
AND either of these extents differ from your study region shown in Figure 1 AND all of these are different than your Figure 2. Make certain you everywhere show the same study region extent, if you want to show more area around, then mark the study region.
Done

(24) Figure 5b. The Moho topography is wrong. There is a Moho offset across the plate boundary but you show a vertical Moho interface!
Yes, the reviewer is right, the Moho shown in the Figure and used in our synthetic test is only a very simplified version of the Spada et al. 2013 Moho (Moho offset, double Moho, etc. are not reproduced). Furthermore, the crustal structure is very much simplified. We state this now more clearly in the text. However, we still think that simplifications for the synthetic test model are acceptable, and a simplified synthetic model is still useful for studying the recovery of some features (hypocenters, 1-D model).
Please see also our comments on similar issues 18 and 21.

(25) Lines 259/60. "average uncertainty of 240m in longitude, 270m in latitude ....“ How did you determine that? In such way this information is not usefull. With what probability do we have what location uncertainty for any single hypocenter?
In the previous paragraph, it is explained that the model parameters are defined based on the average and standard deviation of the final models. In the old version of the manuscript, we showed in figure 6b the histograms of the uncertainties (1σ) of all earthquakes. However, for avoiding any misunderstanding, we decided to not mention these uncertainties here and only talk about the misfits which are important for recovery assessment.

(26) Figure 7. Move to supplementary material.
We think figure 7a of the original manuscript is quite important to display in the main text because it is the best information showing the recovery of the synthetic test. Therefore, we

removed figure 6b (not as important as Figure 7) and merged figures 6a and 7 into one figure (in the updated manuscript it is figure 6).

(27) Figure 8. Again a DIFFERENT STUDY REGION SHOWN!.
The figure frame is now set correctly.
What about the stations to the West of the Tauern window (as example, there are other regions with no visible symbols)? Do they have all zero station delay values or did you not obtain any values for them?
There are stations that have very small delays and it's probably hard to see them (they are like a point in the figure). The stations with zero delay are indicated now with a different symbol in the updated manuscript. There are also some stations without any delays, which were considered in the early part of the study but not included in the final inversion (e.g., no high-quality picks, station problems, etc.). However, all the stations to the west of Tauern Window have symbols (comparison between figures 2 and 7b of the updated manuscript).
Please explain in more detail what the velocity-depth function shows. In may view, it documents the data set is not capable to resolve the velocity structure below 30km depth and certainly not the Moho. This does not come as a surprise as it is well known that you loose vertical resolution below your deepest hypocenters.
We totally agree that resolution typically degrades below the (deepest) earthquakes. However, we would like to point out that the synthetic (3-D) velocity model in the depth range between ~30 and ~55 km is characterized by a strong Moho topography, which *per se* cannot be exactly recovered by a 1-D model (used in our inversion). We think that is not (only) an effect of the data (less ray coverage and resolution below the deepest earthquakes) but also of the inability of a 1-D model to capture a 3-D variation. So, the input model is indeed not well resolved in this depth range, but it is most likely an effect of the data and the much simpler model class used in the inversion. Within the model complexity (in each depth level) the input model is recovered in average (with some variability indicated by the standard deviations). Of course, upper mantle velocities are reached at a large depth. We think this is a nice result of our test which also sheds light on the interpretability of such 1-D models. We modified the text and hope that this issue is somewhat clearer.

(28) Lines 297 to 304. This model discussion is inadequate with regards to the previously published information about the crustal structure. If your model does not allow to resolve it, then say so and it is OK. But do not claim it is in agreement with prior independent knowledge if it is obviously not.
We agree. For example, our model is relatively close to the Diehl et al. (2009) model in the upper part, however, at larger depth it is different. We describe this now better. Furthermore, the resolution of our model is only good down to about 45km depth, below that the standard deviations (1 sigma) are getting significantly larger indicating fading resolution. Also, this aspect is now described in more detail in the updated manuscript.

(29) Line 320. "it proved to be useful for accurately localizing earthquakes ..". I am missing the prove. Please explain how this was proved.
Our phrasing was obviously misleading. We wanted to add a more general statement that - although it might be difficult to interpret the station corrections in a straightforward ("physical") way - the whole concept of the simultaneous /joint inversion (including the inversion for these station corrections) is very powerful. We reformulated this part in the text to:

“Nevertheless, the general concept of simultaneously inverting local earthquake datasets for a (simplified) 1-D velocity model, hypocenter positions/origin times, and station corrections proved to be very powerful for accurately localizing earthquakes (see e.g., Kissling,1988).”

(30) Figure 10. Figure 10b I would again derive the conclusion from this figure that you are lacking resolution power below 30km for Vp and below 20km for Vp/Vs ratio.
The resolution for Vp is starting to fade below 30km depth, however, standard deviations of around 0.02 in the depth range between 30 and 40 km (yellow ranges in Figure 9 of the updated manuscript) still indicate fair resolution. For Vp/Vs, we see this limit at around 30 km. We write this now in the text.
Figure 10c. Now this looks like the study region. Why not always marking this extent where you do have data from?
We homogenized the lat/long in all the maps.
How do you interprete the distribution of the station delays?
We discussed the pattern of station corrections and possible qualitative relations with geological features in lines 313 – 318 of the original manuscript.
Note that there is a single station delay strongly different from all others within its vicinity located near 11.7E/47N. If I obtained such result I would check if it is real or caused by bad data of some sort. Note that otherwise your largest station delays are all within the periphery of the study region as this is well known from minimum 1Dmodel applications.
Yes, the reviewer points out a totally valid point, thanks. We checked the data and identified this station as having wrong picks. Since we generally have to expect influences of these wrong picks on all other parameters in our simultaneous inversion (velocity model, other station corrections, and hypocenters), we removed all picks of this station, reran the simultaneous inversion, and located again all events with the updated velocity model. So, Figure 9 (of the updated manuscript) shows now the updated data. Please note that the changes in the hypocenters and station corrections were very marginal, and all inferences in the paper are not changed.

(31) Lines 325 to 335 and Figure 11. What hypocenter depth did you test these blast locations with?
How do you define the mis-location vector? Relative to the center of the quarries or do you know the precise location within the quarry for each blast? This accuracy test shows that your previous precision estimates were a bit to optimistic but the accuracy is still very good.
We appreciate this question. It motivated us to calculate the misfit vectors for the blasts similar to Husen et al. (1999) using the centers of the quarry areas (since we do not have the exact blast locations). We, therefore, added the following text to the manuscript:
“Based on the average mislocations of the blasts (relative to the centers of the quarry areas) and also their location uncertainties, we estimate the absolute location errors in the range of 1 km horizontally and 500 m vertically.”

(32) Line 331. Please explain in theory why you suggest that including S observations improve the hypocenter location solution?
According to Gomberg et al. (1990) S-picks provide powerful additional constraints on hypocentre location, especially regarding the focal depth. They demonstrated that the partial derivatives of S-travel-times are larger than for P-waves and that they act as a unique constraint especially in cases of S-picks recorded within 1.4 focal depth's distance. However, precisely measuring the S phases is absolutely necessary and we carefully and conservatively identified the first arriving S-picks on the 3C high-quality recordings. Nevertheless, we think that our

statement in Line 331 (of original manuscript) was not really precise and we decided to delete the statement regarding the S-picks. We replaced it by:
"It is expected that the errors of earthquakes (which are potentially deeper than blasts) are smaller than those estimated from this test because they are less affected by the heterogeneous shallow structure which was poorly accounted for by the model (Kissling 1988 & Husen et al., 1999)."

(33) Lines 340/1. Your absolute depth uncertainty has been documented by your ground truth accuracy test (Fig.11) to be a few km and the epicenter location uncertainty is about 1km. Please correct your numbers.
The location uncertainties that we report here are the standard deviations from the mean value of all the models of McMC. We provided now new estimates for the mislocation (based on quarry blasts test; see point 31) that we can now mention here as well. We rephrased the sentence to:
"Based on statistical analyses of the McMC inversion results, the average epicentral and depth uncertainties (1σ) are ~500 m and ~1.7 km which are compatible with precision estimates by the synthetic test (Sect. 5). However, the absolute location errors estimated by quarry blasts test (Sect. 6.2), are 1 km horizontally and 500 km vertically. "

(34) Lines 343/4. These differences are indeed significant. However, as it just regards a selected best event data set with on average 36 P observations this is comparing apples and grapes. Your data set is excellent for seismic tomography but absolutely not representative for a seismicity catalog (magnitude of completeness? Just 16 months). On the other hand, the national seismic catalogs contain many poorly locatable –or you could also say difficult to locate- events that need extra processing time to obtain a complete catalog and there is the significant difference in number of stations. I believe it would be useful to discuss thourougly these difference in addition to presenting just the numbers.
Our intention was not to show poor hypocenter solutions from the national agencies. Most of the large differences between location of the blasts reported by the agencies and our solutions (and the mislocations) are indeed related to the amount of data (stations). We added the following paraphrase to the text:
"Since the national agencies are (probably) using much less data for the location (smaller number of stations used, larger inter-station distances), a significant difference between their hypocenter solutions and those obtained in this study is expected (average of 2.4 km in epicenter and 3.7 km in depth). The earthquake depths calculated by McMC are systematically shallower than those by national agencies (by an average of 1.1 km). The maximum and minimum differences in the epicenters and depths (between McMC and national catalogs) are seen for the earthquakes from the INGV and SED, respectively.
The derived hypocenters in this study do not represent a representative seismicity catalog of the region (the national catalogs contain also many small, poorly constrained events in a much longer period) but form excellent data for further seismological studies e.g., Local Earthquake Tomography (LET). Moreover, this highly precise hypocentral data allows further tectonic inferences."

(35) Figure 12. Figure 12b is not needed, just provide the uncertainty estimates. Note that for cluster interpretation relative hypocenter location uncertainty estimates (and that is what you obtain with your Markov chain MC inversion of the coupled problem that includes a joint hypocenter determination approach) are most important while for absolute location obviously the accuracy is key.

We agree and delete this part of the figure. We now mention the uncertainty values in the text.

(36) Figure 13. For seismotectonic interpretation of clusters along a fault system, you should definitely
employ focal mechanisms.
We totally agree. In the discussion, we wanted to compare our derived seismicity, which is now captured in a consistent way throughout the study area with a dense network, with the seismicity known from other studies and datasets (e.g., from permanent, though coarse, networks running for many years). In this context we also discussed some occurrences along fault zones, however, we did not go into too much detail here. We are aware that for a conclusive investigation, focal mechanisms as well as a more complete catalog is necessary. Within the research program, more studies are underway focusing on exactly these issues. In the meanwhile, a manuscript (Petersen et al., 2021) was submitted (to SE) showing focal mechanisms from the AlpArray and SWATH-D networks. We refer to this study now in the text, however, results were not available at the time of initial submission of our manuscript.

On behalf of the authors,
Azam Jozi Najafabadi