Dear Editor,

We greatly appreciate your concern for our paper submitted for Seismica. We addressed all reviewers' comments, which were very helpful and constructive, improving the revised version of our manuscript. The replies to reviewers' remarks are shown below in black font, and our answers in blue italic font.

Pavlos Bonatis, on behalf of all authors

Reviewer #A:

General Comments:

Throughout the text there are references to M_L, Mw and M. Whereas the first two scales are well-known to the community, scale M, probably needs some explanation, especially since it refers to a wide range of magnitudes (e.g., M>4.1, Figure 1; M<2, M≥5.0, Figure 3; M=7.5, line 114). Please be careful and consistent when referring to diverse magnitude scales throughout the manuscript.

We understand the confusion on the original manuscript, thus we tried to make it more clear in the updated version. More specifically, we erased all references to M_L, and used only M and M_w. M refers to the equivalent moment magnitudes (either directly computed or appropriately converted from other magnitude scales; Papazachos et al., 1997; doi.org/10.1785/BSSA0870020474). On the other hand, all mentions to M_w correspond to the moment magnitude estimated from the moment tensor inversions of the present study.

Please refer to acronyms the first time they are mentioned (e.g., NAT and NAF, in the discussion, Line 384; Also ETAS)

Thank you for pointing this out. All references to acronyms are now corrected throughout the revised version of the manuscript.

Fault type inconsistencies: Table 2 – Focal mechanisms in table 2 indicate a rather significant vertical component (e.g. events 3-5 and 7 – which is the Mw5.4 event); then in Lines 424-427 it is mentioned that "the activated structures hosting the 2020 Mw5.3 and 2022 Mw 5.4 main shocks can be attributed to conjugate faults at the termination of the main strike-slip zone, having a small vertical component consistent with the extensional character of the NAF termination"; Lines 439-440: " 2020 Mw5.3 strike-slip main shock are found in deeper parts compared to those from the 2022 Mw5.4 oblique (normal with a strike-slip component)"; Lines 459-460: "with some oblique cases (strike-slip with a considerable normal component) being also present". Lines 280-282: "The complex faulting patterns coming into play are revealed by the presence of either pure normal (No 5, Table 2; Figure 3a) or oblique (normal with a strike-slip component)" – there is also 1 pure normal. The rake of the Mw5.4 event (-43°) is not small, it is rather significant/ considerable. Please be consistent with your statements.

Agreeing with this comment, all remarks to focal mechanisms are now corrected throughout the revised manuscript.

Please be consistent with the terms “strike-slip” or “strike slip”. Also “mainshock” and “main shock”.

Thank you for pointing out the inconsistency. All terms are now homogenized to strike-slip and main shock. The only exception is the Mainshock-Aftershock type of sequence (MSAS) terminology, which we decided to keep given its more widespread usage.

It would be great if the authors provided the relocated catalog as well as the focal mechanism solutions that they derived, in a public repository, or as supporting data to the manuscript.

Thank you for the advice. In the new version, we added the relocated catalog and the focal mechanisms as supporting data.

Specific Comments:

Line 106: I am not aware of a strong earthquake in NAF in 1969. Do the authors refer to the 1968 event instead? (e.g., Drakopoulos and Economides [1971], Aftershocks of February 19, 1968 Earthquake in Northern Aegean Sea and Related Problems, *Pure Appl. Geophys,* 95, 100-115). If so, it is stated as (M7.1) in line 106; and as M=7.5 in line 114.

Thank you for pointing this out. The earthquake we are referring to is indeed the 1968 main shock. We modified the sentence accordingly and added the reference (lines 108-109 of the annotated copy).

Line 116: Please refer to the symbols used for friction coefficient, μ, and Skempton coefficient, B.

Corrected (lines 120-121 of the annotated copy).

Lines 110-119: I like the content of this paragraph! It justifies all the work done.

We appreciate the reviewer’s assessment.

Line 121: It is definitely feasible to install instruments close to offshore faults, e.g., ocean bottom seismometers. I wouldn’t use the phrase “not feasible” here.

Thank you for this comment. Of course, we recognize the possibility of installing these kinds of instruments. The sentence was modified as “near-fault instruments are not in operation” (line 126 of the annotated copy).

Lines 147-148: Is there adequate discussion on “evaluating seismic hazard connected with activated structures?”. Seismic hazard is not really discussed, just mentioned in the last paragraph of the manuscript (conclusions). This is fine, since seismic hazard is not studied in the paper, but Lines 147-148 should be probably removed from the introduction as they don’t refer to a major finding of the study.

We agree with the comment and erased this sentence.

Figure 1:
Where did M≥4.1 come from? Does it correspond to the completeness magnitude for the period shown? Is it a part of the present analysis or adopted from previous studies? The same with the depth range (0 to 40 km). Where do these data come from? Is 40km real depth? As mentioned earlier in the text (Lines 125–126), Konstantinou (2017), found a maximum depth of ~20km.

Following the reviewer's remarks, we attempt to clarify several issues concerning Figure 1. Firstly, Figure 1 serves as an introductory one, aiming at the brief visualization of the seismotectonic setting of North Aegean area and it is not part of the current study (expect for the focal mechanisms of the 2020 and 2022 earthquake). All the data portrayed in Figure 1 are adopted from previous studies, cited in the corresponding caption and the reference list of the revised manuscript.

Regarding the seismicity of North Aegean area, we used all the crustal earthquakes included in the regional parametric earthquake catalog of the Seismological Station of Geophysics Department of the Aristotle University of Thessaloniki (http://geophysics.geo.auth.gr/ss/catalogs_en.html; Aristotle University of Thessaloniki, 1981) for a sufficient time span, ranging from 1975 up to 2022. The selection of the magnitude threshold of M≥4.1 above which the seismicity is depicted in Figure 1, corresponds to the completeness magnitude of North Aegean as defined by Leptokaropoulos et al. (2012) (doi:10.1111/j.1365-246X.2011.05337.x). This information is now cited in the Figure 1. In continuation of the reviewer's general comment about the magnitude scales, we also want to clarify that all magnitudes are equivalent moment magnitudes (either directly computed or appropriately converted from other magnitude scales; Papazachos et al., 1997). Additionally, the depth range up to 40 km, which refers to the formal definition of the depth range of crustal earthquakes, is now corrected given that in the study area the maximum depth of the seismogenic layer does not exceed 20 km. In this respect and agreeing with the reviewer's remark, we correct the crustal earthquakes depth range of the study area as 0≤h≤20 km (line 156 of the annotated copy).

Is it necessary to draw so many arrows? It seems that only one pair (or two) is enough since there is no change of the motion along strike. This will make the other features clearer.

Following the reviewer's suggestion, an effort was made for drawing less arrows depicting the right-lateral strike-slip motion of North Aegean Trough Fault Zone in Figure 1.

The term NATFZ is not mentioned elsewhere in the text. Please remove.

Agreeing with the reviewer's remark the term NATFZ is now removed from the revised version of the manuscript.

Lines 155 and 159. Use of M_L and M. Please see also the corresponding "general comment".

Following the reviewer's remark, this is now corrected throughout the revised version of the manuscript.

Lines 169–172: I am curious about the uncertainties of the 1.74 value? For example, Andinisari present a value of 1.72+-0.0061. Are the uncertainties of the authors' calculations of similar order? By the way, please correct the reference of "Karamanos et al., 2018", to "Karamanos et al., 2007", both in lines 171 and 543 (references).

Yes, the uncertainty of our calculated V_p/V_s value is indeed very close to those from Andinisari et al., being equal to 0.007 (through the Least Median of Squares method). We included this information in the updated manuscript (line 176 of the annotated copy).. Also, the reference of "Karamanos et al., 2007 is now corrected. Thank you for pointing that out.

Lines 175 – 178. Why not using the updated models that the authors derived? (The authors made the effort to do so, anyway!). Moreover, the authors' models should be more appropriate since they fit exactly to the study area. Akyol et al. study is focused on a nearby area (Longitude 26 – 30 degrees west; Latitude ~37 – 40 degrees north).

The reviewer's comment revealed an important mix-up in the manuscript. Instead of Karabulut et al. (2006) (doi:10.1016/j.tecto.2005.09.009) we accidentally mentioned the model of Akyol et al. (2006) (doi.org/10.1111/j.1365-246X.2006.03053.x) in the text and in table 1. The model of Karabulut et al. (2006) was the closest one compared to the model derived from VELEST using our data. We added a figure in the Supplementary Information to show the two models combined. Moreover, we repeated the relocation procedure using our model, following the reviewer's comment. As a result, Figures 2-5 in the updated version of the manuscript were also modified. We also added a figure in the Supplementary Information (Figure S4) showing the horizontal and depth error histograms.

Table 2:
The mean depth of the (strongest) events is 14+-2.5km, thus deeper than the average of ~8km (with the exception of the last event). I find this result very interesting! (also, the depth distribution in Figures 4 and 5). Maybe the authors could comment/discuss more on that finding in their paper. What are the uncertainties of the depth values (also rake?)?

Regarding the depth and rake uncertainties, we added them into the supplementary Information (Table S1).

How do the authors decide the fault plane? Some mechanisms have strike >300^{o} and some others < 185^{o}.

In Table 2 of the manuscript, we show the nodal plane which is in agreement with the epicentral alignment of the aftershocks. However, following the reviewer's remark, we updated the supplementary information to provide both nodal planes (Table S1).

Lines 225 – 226: "Large α values indicate that large magnitude earthquakes trigger a large number of aftershocks". What do small α values indicate, in terms of the parent-event magnitude and its aftershock offspring?

Productivity parameter α describes the efficiency of earthquakes to trigger their own offsprings as a function of their magnitude. In other words, α weights the mainshock magnitude (Seif et al., 2017; doi:10.1002/2016JB012809). Large α values mean that large earthquakes trigger many aftershocks relative to small earthquakes. Small α implies relatively higher triggering capabilities of small earthquakes (e.g., swarm-like activity). This point is now included in the revised version of the manuscript (lines 237-241 of the annotated copy).

Lines 239 – 249. This is a bit confusing, especially the KS1 test (lines 245 – 247). The transformed times express the number of events within a time interval and thus, comprise a discrete variable. The exponential distribution can be applied for continuous variables (e.g., the inter–event times between subsequent earthquakes). If the inter-event times are drawn from an exponential distribution, then the transformed times are drawn from a Poisson distribution (e.g., LLenos and Michael, 2013; doi: 10.1785/0120130017). Please clarify this point.

Following with the reviewer's comment, this issue is now clarified (lines 258-266 of the annotated copy of the manuscript).

Lines 263–264: When does the relocated catalog start and how many events does it contain? Here it is stated that it starts on 1 January 2020; Lines 188 – 191: "The final catalog contains 735 relocated … 434 before the 2022 mainshock and 301 aftershocks of the second sequence." Lines 250 – 255: "434 belonging to the period starting from the initiation of the seismic excitation (26

SEP 2020) until … 16 JAN 2022" and "the rest 301 consisting the second aftershock sequence". Figure 2: There are events before 1 JAN 2020. Figure 3: Seismicity "from the starts of 2019 until the end of 2022" (1st January 2019, 1st January 2023, as indicated by the color scale). Line 323: "Entire initial earthquake catalog (2019–2023)". Then Line 328: "The model was first applied in the entire initial earthquake catalog of the period 2019–2022, and then in two additional and distinctive sub–periods, namely from January 2020 to June 2021 and from July 2021 to December 2022".

Thank you for pointing out all these inconsistencies. We attempted to make it more clear in the updated version of the manuscript.

Moreover, there are several flags in figure 2:
Y-axis (a). Please explicitly state what does this "distance" refer to.
X-axis (a). Same as Y-axis. What does it refer to? Time since what/when?
(a) The time scale is probably not accurate. The time difference between the first two triangles (1January – 26 September) is 270 days, however the distance in the x-axis is shown to be ~400 days.
(b) (See also comment above for Table 2). Depth distribution is very interesting! mean depth is ~16km (from FPSs it is ~14), deeper than the Konstantinou, 2017 (8km)
All events are here referred to as Mw. See also general comment on magnitude scales. What is the Mc in this figure?

We modified the figure according to the reviewer's comments. We added more informative X- and Y- axis captions and corrected the misplaced triangle in the timescale. Also, all mentions to M_w in the figure caption were corrected to M.

Figure 3, 4 and 5: please remove from all frames the events with M<2.0, they are below Mc, and thus only mislead the reader by adding confusion and incomplete data to the plots.

We respectfully disagree with this comment. The data contained in Figures 3, 4 and 5 are not used to perform any quantitative analysis, thus we believe that their inclusion provides information on the recorded and relocated seismicity, it is therefore useful to the reader. In contrast, in the ETAS section we are using only the complete catalog.

Line 291 – 293: It is stated that "The total length of the activated area is approximately 10 km, which is in compliance with frequently used empirical relationships between main shock magnitude and the causative fault length (Papazachos et al., 2004; Wells & Coppersmith, 1994)". However, in line 271, it is stated that "… aftershock activity followed the main shock on 26 September 2020, extended in an area more than 15 km long, much larger than the rupture length of an Mw5.3 main shock, that equals to 7 km as prescribed by scaling laws". It seems from figures 2 and 3 that a length of 10-15 km is plausible. *(Personally, I think it is rather ~10km)*. Please provide justification and consistency.

Thank you for pointing out these inconsistencies in the manuscript. First of all, we erased the reference to Papazachos et al. (2004) because their empirical relationship is not applicable to our case (valid for 6.0≤M≤8.0). Instead, we used the scaling law of Thingbaijam et al. (2017) (doi.org/10.1785/0120170017) which is more appropriate (valid for 5.38≤M≤8.7) (line 291 of the annotated copy).

Line 294: How "short" is defined? What are some typical values for similar magnitude events?

We did not attempt to seek for typical aftershock duration values of similar magnitude earthquakes because aftershock duration does not scale with mainshock magnitude (e.g., Ziv et al., 2006; doi:10.1029/2006GL027141; Toda & Stein, 2018; doi:10.1785/0120170270). Aftershock duration has been proven to vary among different tectonic settings and other region-specific characteristics (Valerio et al., 2017; doi:10.1038/s41598-017-14550-2; Bonatis et al., 2022; doi:10.3390/geosciences12090328). In any case, in this section we are not referring to the total duration of the sequence rather than the duration of "intense" aftershock activity.

Lines 296-299: What is interesting? The occurrence of foreshocks itself? The occurrence of few foreshocks, or their distance from the main shock? Again, what is the typical behaviour to compare with, making this observation interesting?

To our knowledge, the presence of foreshocks (the strongest among them with M_w=4.4) is not a common feature for the study area, for either large (e.g., Drakopoulos & Economides, 1972; doi:10.1007/BF00878858; Konca et al., 2018; doi:10.1093/gji/ggy049) or moderate main shocks (e.g., Karakostas et al., 2014). This is why we considered this observation worth mentioning, along with their location close to the impending main shock.

Line 300: This is very interesting in my opinion! (again: what are the vertical errors of the events?)

Following this remark, we added a figure in the Supplementary Information (Figure S4) showing the vertical error histogram of the events.

Figure 4, 5: last two lines of the caption: That's a bit confusing. What is the difference between b and c (and d)? It is not well described in the captions.

We modified Figures 4 & 5 captions to make the distinction between b, c and d more clear.

Line 325 and Figure S2: I would never go to Mc = 1.9, even if GFT suggest show, since there is a clear violation of the GR law: equal number of events (non cumulative) for M = 1.9 and M = 2.0. This is possibly not very crucial for the results of the paper, since b-value analysis is not carried out here. I would suggest though repeating the calculations for Mc = 2.0 and see how the b-value and the overall results (ETAS modelling) are affected by Mc selection. The b-value (0.78) is most probably an underestimation. The authors may consider for future studies, applying the Amorese 2007 method (repeated medians, https://doi.org/10.1111/j.1365-246X.2009.04414.x), as more efficient method for small datasets.

We would like to clarify that the equal number of earthquakes for the magnitude bins of M = 1.9 and M = 2.0 is likely to be related with the formal error of magnitude estimation (~0.1). This fact means that one M = 2.0 earthquake could be probable equal to either M = 1.9 or M = 2.1, considering its magnitude uncertainty. The same pattern is also observed for larger magnitude bins included in our dataset (e.g., M = 3.7 & M = 3.8). Regarding this fact, we decided to select as completeness magnitude, the magnitude bin above which the percentage of the residuals are lower than the fixed value of 5%, as Wiemer & Wyss (2000) (doi:10.1785/0119990114) and Mignan & Woessner (2012) (doi: doi:10.5078/corssa-00180805) suggested, aiming at including as many as possible earthquakes in our statistical analysis above a robust and acceptable magnitude threshold.

However, we repeated our analysis for the entire time interval (2019-2022) by using as completeness threshold the Mc = 2.0. The b-value of this dataset is equal to b=0.80, which is very close to the previously estimated one (b=0.78 for Mc = 1.9). Table R1 shows the results of the ETAS application with Mc = 2.0, *in comparison with the results using the Mc = 1.9 threshold. As it is observed, the values for all the 5 ETAS parameters are quite similar for both the magnitude thresholds, indicating that the Mc selection does not affect the final results.*

Table R1. Temporal ETAS parameters estimates (μ, K, α, c and p) for the period 2019-2022 by using as magnitude of completeness the thresholds Mc=1.9 and Mc=2.0, along with the respective number of observations.

Period: 2019-2022	μ	K	α	c	p	Obs.
Mc=1.9	0.046	0.015	1.78	0.030	1.23	470
Mc=2.0	0.042	0.014	1.80	0.026	1.20	401

Additionally, we would like to thank the reviewer for his suggestion regarding the repeated medians method of b-value estimation (Amorese, 2007). Following this suggestion, we estimated the b-value of the 2019-2022 earthquake catalog and with Mc = 1.9. The repeated medians method results in a b-value equal to b=0.80, which is again very close to the estimated one by the MLE method. This slight difference between the two methods indicate that in our case the sample size does not significantly affect the b-value estimation.

Line 343-347: "In more detail, the observed seismicity during the period from 2019 to 2022 is characterized by a very low background rate equal to μ=0.046 event/day. This means that almost 98% of the total number of earthquakes are aftershocks of the 2020 Mw5.3 and 2022 Mw=5.4 main shocks, and only 67 out of the 470 are assumed to be independent." How did these numbers come out? 98%? And how are we sure that they are aftershocks of these two events and not of some others? The Epidemic Type model represents a branching process, that each event can induce aftershocks. And how 67/470 relates to 98%. Please clarify.

We would like to clarify the issue concerning the number of background and triggered earthquakes during 2019-2022, as derived by considering the estimated ETAS model. As already stated, the background rate of earthquakes above Mc during the period 2019 – 2022 is estimated equal to μ=0.046 event/day. This corresponds to 68 background earthquakes for the four years of our learning period (4 years x 365.25 days/year x 0.046 event/day). Given that, the typo associated with the number of background earthquakes (68 instead of 67) is now corrected (line 373 of the annotated version of the manuscript). Subsequently, the ratio background earthquakes over the total number of earthquakes for the learning period catalog corresponds to ~0.12 or 12%, while the ratio of the triggered over the total number of earthquakes is equal to 88%. In this respect, the typos regarding the percentage of the triggered earthquakes (88% instead of 98%) is now corrected (line 371 of the annotated version of the manuscript). Additionally, we agree with the reviewer's remark associated with the association of the triggered earthquakes with the two mainshocks. In this respect, we modify this part of paragraph by assigning these earthquakes to the two sequences and not to the mainshocks (line 373 of the annotated copy).

Lines 375-380 (also 466-470): if "98% of the total number of earthquakes are aftershocks of 2020 and 2022 main shocks" (as stated in Line 345, see comment above), this means that the selected time duration cannot capture the entire sequence(s). e.g., the smaller p parameter (Lines 378-379) indicates a shorter sequence.

The authors would like to notify that the smaller values of p parameter the larger the duration of the sequence (or in other words the aftershock decay becomes faster as p increases; Seif et al., 2017, among others). Regarding this fact, our applications indicate that the 2020 sequence had a larger duration than the 2022 one.

Also, if so, in figure 6e, it is shown that part of the 2020 sequence is included in the 2022 sequence. Shall we expect that this would affect the results? Can we also consider the fact that the two sequence might have mixed with each other.

Focusing on the duration of each period associated with the estimation of ETAS parameters of the 2 sequences, the selection is made by considering the temporal distribution of the complete dataset. As it is derived from Figure 6a, the majority of earthquakes associated with the 2020 earthquake occurred between the 630th and 850th days (from 01/01/2019) of the complete earthquake catalog, whereas the second period is selected to started from the 900th day of the complete earthquake catalog (01/07/2021), aiming at to avoid the possible overlapping and the mixed up of the sequences.

Or that the second sequence continues for longer than the selected data covers. Please clarify.

As already said, the second sequence (2022) is characterized by faster aftershock decay rate, since its p value is found to be larger than the 2020's one (which lasted ~200 days). One the other hand, our complete dataset covers the next 11 months after the 2022 mainshock, indicating low seismicity occurrence rates. By combining the aforementioned remarks, it could be state that the second sequence ceased quite earlier from the end 2022.

Figure 7. In the schematic illustration the conjugate faults are indicated as dextral.

Thank you for pointing out this mistake. We modified the figure accordingly.

Lines 435-445: I really like that paragraph!

We appreciate the reviewer's assessment.

Lines 446-451: The entire statistical analysis part is only discussed in 6 lines. Maybe add a comparison with values from other studies?

Please, see the answer to the next comment.

Lines 461-470: this is rather discussion than conclusion

Agreeing with the reviewer's comment, we moved this paragraph from the conclusion to the discussion section.

Last paragraph, good closing!

We appreciate the reviewer's assessment.

Minor Comments:

Lines 100-102. There is no such info shown in the inset of figure 1. Reference to the inset is most suitable earlier in the introduction.

Following the reviewer's suggestion, we moved the reference to the inset of Figure 1 earlier in the section (lines 82-83 of the annotated copy).

Line 130, 142. Maybe refer to the 2 mainshocks as "right magenta circle" and "left magenta circle", respectively.

We followed reviewer's suggestion and added "right" and "left" to be more specific (lines 135 & 148 of the annotated copy).

Line 135, please explain the abbreviation gcmt (as stated in figure 1 caption).

The abbreviation is now explained (line 140 of the annotated copy).

Lines 145 – 147: "We further study the spatiotemporal evolution of the activity, the temporal characteristic through the application of the ETAS model". Please rephrase this sentence.

Following the reviewer's comment, this is now corrected

Line 161. Please make the "o" symbol a superscript

Following the reviewer's suggestion, this is now corrected

Line 190: typo shcok -> shock;

Following the reviewer's remark, this is now corrected

Line 194: I suggest not to mention "with ML≥3.7". Since Mws are calculated and shown in Table 2, reference to ML is only confusing at this point.

Following the reviewer's suggestion, this is now corrected

Lines 225-226: "the efficiency of earthquakes in triggering its own aftershocks" to "the efficiency of an earthquake in triggering its own aftershocks".

The sentence was modified as "the efficiency of earthquakes triggering their own aftershocks"

Line 262: correct "migrate" to "migrates" or "migrated"

The term was modified to "migrated".

Line: 269: "close distance with (Table 2)", something is missing here

We added "…the main shock". (lines 287-288 of the annotated copy).

Line 272: The correct reference is Wells and Coppersmith, 1994 (not 1984)

Following the reviewer's comment, this is now corrected.

Symbols consistency: mc Line 322; m_c line 325

All mentions to m_c are now consistent throughout the text.

Line 334: delete repetition: “The 2019-2022 ETAS application for 2019-2022”

The sentence was modified to avoid repetition (line 360 of the annotated copy).

Lines 336, 346: notation consistency: “2020 Mw5.3 and 2022 Mw=5.4”

Following the reviewer’s comment, this is now corrected.

Line 338-340: Is visual inspection a robust criterion? The KS-test result provides the real evidence!

We would like to clarify that residual analysis is a widely acceptable and powerful method for assessing the fit of a particular model to a set of occurrence times, unless it is a qualitative one (Zhuang et al., 2012; doi:10.5078/corssa-79905851). However, a quantitative statistical evaluation of a certain model must be applied for drawing robust conclusions. This is the reason why the KS goodness-of-fit test is applied in the present study. As it is derived in the manuscript, the visual inspection is used for the qualitative description of Figure 6 (subplots 6b, 6d, 6f), whereas the final conclusions regarding the applications of ETAS model are made by considering the results of KS test.

Line 341: delete duplicate p-value statement.

Following the reviewer’s comment, this is now corrected.

Figure 6: there are some confusing ticks in the right y-axis in all frames (a-f)

Following the reviewer’s comment, the figure is now modified.

Lines 363-365: be careful with the Figure 6 frames, only 6d is stated in the text.

Following the reviewer’s comment, this is now corrected.

Line 418: “A typical example are the fault segments”, probably the verb should be “is”, not “are”

Following the reviewer’s comment, the sentence is now modified.

Figure 7. It would help the reader if small titles were added above each frame (or a more detailed description in the caption).

Following the reviewer’s comment, the figure is now modified.

Line 427: “The primary indications leading to this presumption is”, probably this should be “are”.

Following the reviewer’s comment, this is now corrected.

Line 430: “right”, should be “left”

Following the reviewer’s comment, this is now corrected.

Line 451: There is no reference to Kourouklas et al., 2021 in the list.

We added the reference in the list.

Reviewer #B:

GENERAL COMMENT

This manuscript describes an analysis of two mainshock-aftershock sequences that took place in the North Aegean using a variety of techniques for earthquake location and for highlighting the earthquake statistics. In general, I found the manuscript well written, logically arranged and interesting from the point of view of enhancing our knowledge of seismic hazard in this area. However, I have also noticed several methodological problems that need to be rectified in order for the manuscript to be published in Seismica. I therefore recommend major revision according to the comments I list in detail below.

SPECIFIC COMMENTS

1. Line 79: Please add references to support the statement about N-S deformation in the Aegean crust. Two papers that have determined the stress field in detail are Konstantinou et al. (2017) & Kapetanidis and Kassaras (2019).

We added the references proposed by the reviewer in the updated manuscript (lines 79-80 of the annotated copy).

2. Line 82: I think the reference to McKenzie (1972) is quite old, there have been much more recent references to support the statement about plate kinematics, a few examples: Hollenstein et al. (2008), Floyd et al. (2010), England et al. (2016).

Following the reviewer's comment, we added two more recent studies (England et al., 2016 and Bitharis et al., 2024) to support our statement (lines 87-88 of the annotated copy).

3. Line 170: Please delete "North", the references you cite also cover other parts of the Aegean not only the north part.

Following the reviewer's comment, this is now corrected.

4. Lines 174-178: I don't understand the reasoning behind choosing the velocity model of Akyol et al. (2006). This is a model that has been derived for the western part of Anatolia using very few raypaths that cross the north Aegean. It would be more reasonable to use the velocity model derived by Konstantinou (2018) which mostly uses raypaths that traverse the Aegean and much less the Anatolian crust. In addition to this, the paper by Konstantinou (2018) compares the two models and finds that there is a significant difference for depths greater than 10 km. This is also the depth where most of the seismicity is clustered according to the results of the present manuscript, which makes me wonder whether this clustering is real or a result of the velocity model that was used for location. I would recommend that the authors repeat the location with the model of Konstantinou (2018) in order to make sure that the depth distribution is not an artifact.

Both reviewer's comments revealed an important mix-up in the manuscript. Instead of Karabulut et al. (2006) we mistakenly mention the model of Akyol et al. (2006) in the text and in table 1. The model of Karabulut et al. (2006) was the closest one compared to the model derived from VELEST using our data. We added a figure in the Supplementary Information to show the two models combined. Considering both reviewers' comments, we repeated the relocation procedure using our model, which uses data from our study area. As a result, Figures 2-5 in the updated version of the manuscript were also modified.

5. The manuscript does not include anywhere formal uncertainties both for the absolute locations obtained by HYPOINVERSE and the relative ones from HypoDD. HYPOINVERSE provides such uncertainties for each located event (horizontal & vertical) and histograms of these uncertainties should be included in the manuscript (or supplement) along with a brief summary of the average horizontal and vertical uncertainty. For the relative locations I would strongly suggest to obtain uncertainties by relocating particular clusters using the SVD option of the HypoDD package (see Waldhauser, 2001).

Following the reviewer's comment, we added a figure in the Supplementary Information (Figure S4) showing the horizontal and depth error histograms.

6. Continuing from the previous point, there is also no information about the robustness of the inverted focal mechanisms. I think it is necessary that the authors show (for each of the events in Table 2) the fit they obtained between synthetic and observed waveforms and include the corresponding Figures in their supplement. Additionally, they should also include in Table 2 some metric (fit, or misfit) that summarizes the overall goodness of fit for each event. I am not familiar with the Grond package, hence I cannot be more specific about which metric should be used here.

Following the reviewer's comment, we added additional figures in the Supplementary Information (Supplementary_2) showing the waveform fits for every focal mechanism. Also, a Table was added showing the best fit parameters along with their uncertainties (Table S1)

7. In Table 2 there is a column "depth", however, it is not clear what this represents: is it the depth obtained from HYPOINVERSE? HypoDD? or the depth after waveform inversion for the focal mechanism? Please clarify this in the caption. If the authors have determined depth from the focal mechanism inversion it would be a good idea to compare this with the depth obtained from HYPOINVERSE/HYPODD, as a cross-check of depth reliability.

In Table 2 the focal depths obtained from the earthquake relocation are reported. In the newly added Table in the Supplementary Information, we included the centroid depths obtained from the inverted focal mechanisms (Table S1).

8. In Figure 2 the depth distribution shows very few events located at depths less than 8 km. I think that authors have to provide an explanation for this; could it be an artifact of the velocity model used (see Point 4)? or is this related to the sediment-filled basins that can be found in North Aegean?

In the updated version of the manuscript, after performing the relocation procedure again using our model, there are much more events located at depths < 8 km (Figure 2b).

9. Lines 300-301: Please use the percentile difference P95-P5 in order to estimate the seismogenic layer thickness, as this the procedure most widely accepted (and results with other such estimates can be performed directly).

Thank you very much for this suggestion. We included the 5th and 95th percentiles in the depth histogram in Figure 2b and used this information throughout the text.

10. Line 372: Please change to "On the contrary..."

Following the reviewer's suggestion, this is now corrected.

11. Line 427: Please change "presumtpion" to "suggestion"

Following the reviewer's comment, this is now corrected.

12. Lines 430-431: Right-lateral motion is the same as dextral motion, please rephrase your sentence.

Following the reviewer's comment, this is now corrected.

13. Last (but not least!) I would encourage the authors to upload in the supplement, or in an external database all the results of their work (absolute, relative location catalogs) so that other researchers can use them or try to reproduce them.

Thank you for the advice. In the new version, we added the relocated catalog and the focal mechanisms as supporting data.