Dear Dr Skoumal,

Please find attached the revised version of our manuscript entitled “Enhanced view of the Mw 6.4 Petrinja (Croatia) earthquake sequence (2020–2022) using deep learning”. We sincerely apologise for the long delay in submitting the revised manuscript and greatly appreciate the patience of both the editor and reviewers during this process.

The reviewer comments prompted us to substantially re-evaluate and improve the entire workflow presented in the manuscript. In response, we repeated the complete detection, phase association, relocation, and magnitude estimation procedures in order to address the issues identified during review. This included implementing a revised relocation strategy using NonLinLoc SSST corrections, re-evaluating magnitude estimation procedures, revising the frequency-magnitude and magnitude of completeness analyses, and introducing additional quality-control procedures for newly identified events. We also included a new discussion point on temporal aftershock decay. As a result, the revised manuscript differs from the original submission and we believe it has been significantly strengthened.

We have incorporated most of the suggestions made by the reviewers, and all major comments have been carefully addressed. The changes are highlighted within the manuscript. Please see below, in blue, for our point-by-point responses to the reviewers’ comments and concerns. All line numbers refer to the revised manuscript file (without tracked changes).

Reviewers’ comments to the authors:

Reviewer A:

We thank Reviewer A for the thorough and constructive review of our manuscript and for the positive assessment of the study. We appreciate the reviewer’s recognition of the value of the automated workflow and the systematic comparison between the PESMLC catalogue and manually curated catalogues. We also thank the reviewer for the insightful suggestions regarding event locations, magnitude estimation, false detections, and the broader implications and applications of deep-learning enhanced catalogues. We have carefully considered all comments and have revised the manuscript accordingly. Detailed responses to each point are provided below.

<u>Key suggestions</u>

1. Need more discussion of event location discrepancies – especially depths and possible artifacts

We thank the reviewer for this detailed and insightful comment regarding the event location discrepancies and potential depth artefacts in the original manuscript. In response, the event locations in the revised manuscript were recomputed using the NonLinLoc SSST (Source Specific Station Term) method. This approach successfully removed the horizontal depth artefacts previously visible at approximately 2 km and 5 km depth in the original figures. While a small number of depth artefacts are still present in the complete catalogue, they are primarily associated with lower-quality events. Importantly, these artefacts are not observed in the newly identified events after applying the quality-filtering criteria described in the

revised manuscript. We have now explicitly discussed this issue in the manuscript and clarified that the artefacts are related to limitations in the initial location procedure and station correction treatment.

We also clarified the differences between the PESMLC, HHC, and CEC catalogues regarding station corrections and location procedures. The HHC catalogue applies a different station correction methodology, whereas the CEC catalogue was produced without source-specific station corrections. These methodological differences contribute to the observed discrepancies in depth estimates between the catalogues.

Additionally, we expanded the discussion regarding the systematic depth differences between PESMLC and the manual catalogues. Although the same general 1D velocity model framework as Herak and Herak (2023) was used, the implemented velocity model is not identical. Furthermore, as now discussed in the manuscript (line 249), the PESMLC contains a substantially larger proportion of S-wave arrivals compared to the manual catalogues. Specifically, PESMLC contains 259,494 P-wave and 257,923 S-wave picks, whereas the manual catalogue (HHC) contains 174,526 P-wave and 79,419 S-wave picks. This more balanced phase distribution better constrains hypocentral depth and contributes to the observed differences in depth estimation.

We further clarified that the final PESMLC locations presented in the revised manuscript are based solely on the NonLinLoc SSST relocation results. Although hypoDD relocation was initially explored, it did not produce additional relocations after the SSST corrections were applied and was therefore not used in the final catalogue generation or analysis. The manuscript has been revised accordingly to avoid ambiguity regarding the relocation workflow.

Finally, the suggested plotting improvement was implemented by increasing the vertical separation between panels in the corresponding figures.

2. Need more analysis of magnitude estimation, comparisons, discrepancies, and their effect on magnitude-frequency distribution (MFD) and Mc

We thank the reviewer for this comprehensive and constructive comment regarding the magnitude estimation procedure and its implications for the frequency-magnitude distributions and magnitude of completeness analyses. In response to these comments, we substantially revised both the magnitude calculation procedure and the associated discussion in the manuscript.

Most importantly, we corrected the S-wave amplitude time window used for local magnitude estimation to match the procedure applied in the HHC and CEC catalogues. This revision significantly reduced the previously observed magnitude discrepancies between PESMLC and the manual catalogues. In the

revised manuscript, the mean magnitude difference between PESMLC and the manual catalogues is now approximately 0.15 magnitude units. The remaining differences are likely related to factors such as differences in the set of stations contributing to individual magnitude estimates, the larger number of automatically detected S-wave arrivals in PESMLC, and differences in waveform quality and phase availability between the catalogues.

Following the reviewer's suggestion, we also expanded the discussion regarding the origin and implications of magnitude discrepancies. The previous interpretation linking higher b-values and higher Mc directly to the ML-based detection approach was overly simplified and has now been substantially revised. In particular, we removed statements implying that increased detectability necessarily results in higher b-values or higher Mc values.

The revised discussion now focuses on the differing behaviour of the Mc estimation methods themselves, particularly for large catalogues. We clarify that the discrepancy between the Mc estimates obtained using MAXC and those obtained using MBS and K-S methods primarily reflects the differing statistical sensitivities of these methods to catalogue size. Specifically, the MBS and K-S methods become increasingly sensitive to small deviations from the idealised Gutenberg-Richter distribution as the number of events increases, whereas the MAXC method remains comparatively stable. We therefore interpret the higher Mc estimates obtained for the PESMLC using MBS and K-S methods as a methodological effect related to the very large catalogue size, rather than as evidence of poorer detectability or lower catalogue quality.

We additionally revised the manuscript to clarify that manually curated catalogues may indeed achieve lower Mc values than fully automated catalogues, particularly when analysts manually identify very small events recorded on only a limited number of stations. This clarification was added to avoid implying that lower Mc values are necessarily expected for ML-generated catalogues.

Furthermore, the figures presenting the frequency-magnitude distributions were revised to use consistent magnitude-axis limits across all catalogues, improving visual comparability between the datasets.

Finally, the concluding statement regarding future work was revised following the reviewer's suggestion. Rather than implying that ML catalogues require new Mc estimation methods, the revised text now emphasises that further analyses of long-duration ML-derived catalogues could help quantify the extent to which completeness thresholds can be lowered in fully automated workflows.

3. Need to verify a minimal rate of false detections

We performed additional analyses of the temporal distribution and location quality of the newly identified events.

Inspection of the event occurrence times showed no clear concentration during local daytime hours, indicating that the newly identified events are unlikely to be dominated by anthropogenic noise sources. In addition, we introduced a quality-filtering procedure for newly identified events based on the NonLinLoc location uncertainties. Specifically, events were required to have an azimuthal gap smaller than 180°, depths shallower than 20 km, and a major axis length of the 3D confidence ellipsoid smaller than 10 km. After applying these criteria, 14,274 out of 19,436 newly identified events were retained.

The newly detected seismicity is spatially consistent with the overall aftershock sequence rather than representing randomly distributed false detections or anthropogenic noise. The revised manuscript now includes this additional quality assessment and discussion of the newly identified events.

4. Can benefit from expanding on potential uses going forward, and added value, for a deep-learning enhanced catalog such as PESMLC

What will the enhanced PESMLC catalog be used for in the future? Does it find new information about fault structure or earthquake behavior that we didn't know with the HHC or CEC? Or is the main advantage the speed at which the catalog is generated (lines 341-342) – and what need does this rapid availability fulfill?

Lines 369-370: "rapid, detailed, and comprehensive seismic catalogue essential for seismic hazard assessment and mitigation strategies" is a good start

See line 494-511. We included broader implications of our study.

Lines 341-347: I think the "reduction in runtimes" paragraph should have its own subsection title, since it doesn't fall under the "Magnitude of completeness" heading category.

Line 383: Added subsection title "Processing time"

Minor edit suggestions:

Line 3 (Title): Perhaps add "Croatia" after "Petrinja" in the title, for added context to readers?

Added "Croatia" in the title

Line 10-11: "The abilities and performance of machine learning for creating seismic catalogues are still questioned and being tested" – Perhaps rephrase as "Machine learning approaches have shown much potential to automate the creation of seismic catalogues, but their abilities and performance are still questioned and being tested"?

Line 10-11 Suggested rephrasing accepted

Line 22: suggest adding "in time" after "closely spaced events"

Line 23: Suggestion accepted

Line 23: perhaps replace “Overall, PESMLC” with “Nevertheless, PESMLC overall” – since the last sentence is describing a positive view of machine learning pickers, in contrast with the limitations described in the previous sentence?

Line 25: Suggestion accepted

Line 28: instead of “an exponential growth” which is too specific and may not be technically correct, I suggest saying “rapid growth” or “accelerating growth”

Line 35: Suggestion “accelerating growth” accepted

Line 35: suggest adding “automatically” before “detect and classify”

Line 42: Suggestion accepted

Lines 35-37: suggest replacing “with higher accuracy and efficiency than traditional methods (e.g. Ross … 2020) often comparable to manual pickings” with “with high accuracy often comparable to manual picking performance (e.g. Ross … 2020)” to be more concise. [Comparison to traditional methods such as STA/LTA and template matching are discussed later in this paragraph]

Line 42-43: Suggestion accepted

Lines 39-42: The sentence “Both manual and semi-automatic … calibrated for.” is an incorrect statement, as it mixes too many concepts together. I suggest splitting it into 2 sentences, so that each sentence is correct: “In contrast, traditional automated approaches to earthquake phase-picking, such as the amplitude-based short-term average to long-term average (STA/LTA) method (Allen, 1978), are unreliable in low signal-to-noise ratio (SNR) environments. Existing semi-automatic event detection approaches designed to detect smaller events, such as template matching, often only work for the specific network they have been calibrated for.”

Line 46-50: Suggestion accepted

Line 49: Suggest rewriting “However, ML methods are often employed as ‘black-boxes’ approach and it is crucial” as “ML models are often employed as ‘black-boxes’, but it is crucial”

Line 57: Suggestion accepted

Lines 59 – 65: Perhaps include some references (either publications or webpages) for the 2020 M6.4 Petrinja earthquake (e.g. Herak and Herak, 2023)? If you cite the USGS event webpage at https://earthquake.usgs.gov/earthquakes/eventpage/us6000d3zh/executive, use this reference: S. Geological Survey, Earthquake Hazards Program, 2017, Advanced National Seismic System (ANSS) Comprehensive Catalog of Earthquake Events and Products: Various, https://doi.org/10.5066/F7MS3QZH.

Line 67: Added reference for Petrinja earthquake (Herak and Herak, 2023)

Line 102: suggest writing “show before” as “show that before”

Line 109: Suggestion added

Figure 1 (right): foreshock and mainshock (stars) are difficult to see, since they are obscured by an overlapping station (blue triangle). Perhaps plot the foreshock/mainshock layer on top of (after) the stations? It would be nice to have a scale bar in [km].

Figure 1 replotted so mainshock and foreshock are plotted as layer atop of stations. Scale bar added.

Figure 1 caption, also line 204, line 223, Table 1: "04/01/2021 and 05/05/2021" – perhaps write as "4 and 5 January 2021" (to be consistent with line 96: "January 4 2021"). Otherwise, some readers might think OP-Net was installed on April 1 and May 5 2021.

Figure 1 caption: Date formatting corrected in the Figure caption

Line 116: After "manual catalogue", I suggest adding "for a months- to years-long active earthquake sequence". After "ML catalogue", I suggest adding "for the same earthquake sequence". [Just to be clear – since during a seismically quiet time, it doesn't take as long to create a manual catalogue. ?]

Line 123-124: Suggestions accepted

Line 121: before "data", suggest adding "continuous" to be explicit.

Line 130: Suggestion accepted

Lines 121-123: Is there a website or FDSN citation for the Croatian Seismograph Network, OP-Net, and PM-Net, which could be included in the references?

Line 130: Reference for CR network added. There are no references for PM-Net. OP-Net is referened earlier in the text Line 106 as Stipcevic et al (2021)

Line 123-124: "total number of stations reached 55" – adding up number of stations from previous lines: 35 permanent + 5 temporary short-period + 16 temporary broadband = 56. (Was one permanent station offline during the time the temporary stations were in place, to get 56-1 = 55?)

Line 133: Typo corrected

Line 131-132: "The probability thresholds for P and S waves were set to 0.05" – can you also include the probability threshold used for event detection?

Line 139-140: Information for detection probability threshold added "The probability thresholds for earthquake detection, P, and S waves were set ...."

Line 154: Perhaps add a reference to Figure 1 which shows the study region?

Line 170: Added reference for Figure 1 when mentioning study region

Figure 2 caption: Suggest describing which axis goes with which quantity in the caption: replace "blue" with "blue, left vertical axis"; replace "black" with "black, right vertical axis"; replace "orange" with "orange, right vertical axis"

Figure 2 caption: Suggestions to improve Figure 2 captions were accepted

Line 171-172: Suggest removing “The majority of the seismic events are clustered around the primary Petrinja fault plane” since this sentence is repeated in line 173 (except “plane” replaced with “zone”)

Line 190: The redundant sentence “The majority of the seismic events are clustered around the primary Petrinja fault plane” has been removed

Line 175: Do you have a reference for the “March 22, 2020 Zagreb earthquake”? What was its magnitude?

Line 189: Reference and local magnitude for Zagreb earthquake has been added

Line 186: Do you have a reference for this magnitude equation, since it is “standardly used to estimate local magnitude in the Croatian Seismic Catalogue”?

Line 199: Reference for magnitude equation has been added

Figure 4 caption: I don’t see any orange curves in this figure? Did you mean “blue curves, right and bottom” for the kernel density estimation?

This is now Figure 3: Typo “orange” – “blue” has been corrected

Line 199: “matched event as one for which the time difference” – can you clarify if this is the time difference is for event origin times? (or alternatively… all event phase arrival times?)

Line 226-227: Clarification added that we mean event’s origin time. Sentence rephrased to: “In our comparison, we define a matched event as one for which the time difference between PESMLC and the manually detected event's origin time is less than 2 seconds.”

Figure 5 caption: should “PESMLC (grey)” in last sentence, describing the KDE, be “PESMLC (blue)” instead?

Now Figure 4 : Figure caption corrected

Line 222: “Table ??” missing reference – should this be “Table 1”?

Line 244: Latex compile error corrected. It now says “Table 1”

Line 227: Was the comparison of PESMLC and CEC (for matched, missed, new) also limited to a specific geographic region? If so, please specify the region boundaries. (Even if the region boundaries were different from those used in the HHC comparison, you may only wish to count events near your study area, e.g. within the Figure 1 (right) boundaries)

Line 248: added boundaries for the CEC comparison. These are the actual boundaries of the study as shown in Figure 1. Comparison with HHC is on a smaller area as mentioned in the manuscript

Lines 267-310 (“Missed events” section, but also throughout the paper): please be careful about using acronyms “ML” = “machine learning”, and then “ML” = “local magnitude” nearby, as they look alike but mean different things. Perhaps use “M” (as in lines 262-263) or “MCR,L” (as in line 186) to mean “local magnitude” instead?

Line 329-335: ML=local magnitude has been corrected here and throughout the text to be MCR,L to avoid confusion

Line 268: “624 missed events from HHC” – lines 202-203 say 620 missed events. Which number is correct? Please be consistent. (The numbers measured in lines 269-270 add up to 620)

Line 326: We re-ran the detections and association so we now have different numbers. We keep it consistent throughout the text

Line 270-271: “322 events for which HHC gives magnitude 0 (i.e. could not be estimated)” – can you elaborate more on why HHC could not estimate magnitudes for these 322 events?

Line 295: Unfortunately, none of the authors were involved in the process of creating HHC, so we don’t know why exactly some events did not have magnitudes estimated.

Line 293: “minium” should be spelled “minimum”

Line 319: Typo corrected

Figure 8c, d (horizontal axis labels): Perhaps write “CEC” instead of “CSS” to be clear?

Line 347: “manually verifying events above a specific threshold” – what kind of threshold? Magnitude threshold?

Line 466: Added “magnitude and quality threshold”

Line 351: suggest adding “respectively” after “NonLinLoc”

Line 394: Suggestion accepted

Line 370: suggest replacing “noway” with “not” or “rather than”

Line 508: Suggestion accepted

Line 371: “refining parameters” – can you elaborate in a bit more detail on which parameters?

Line 509: Clarified that we mean association parameters

Reviewer B:

We thank Reviewer B for the detailed evaluation of the manuscript and for recognising both the strengths and current limitations of machine learning approaches in seismic catalogue generation. We appreciate the reviewer’s constructive comments regarding magnitude estimation, event location uncertainties, closely spaced events, magnitude of completeness analyses, and the broader implications of ML-based seismic monitoring.

Many of the issues raised by the reviewer B overlap with the detailed comments provided by Reviewer A. In response to these comments, the manuscript has been substantially revised. In particular, we improved the event relocation procedure using the NonLinLoc SSST method, which substantially reduced the

previously observed depth artefacts and improved overall location consistency. We also revised the local magnitude estimation procedure by correcting the S-wave amplitude measurement window to match the methodology used in the manual catalogues, significantly reducing the previously observed magnitude discrepancies.

The discussion of magnitude of completeness Mc and frequency-magnitude distributions was substantially expanded and revised. We clarified that the higher Mc values obtained for the PESMLC using the MBS and K-S methods primarily reflect the increased statistical sensitivity of these approaches for very large catalogues, rather than poorer detectability of the ML-generated catalogue. We additionally clarified that manually curated catalogues may achieve lower Mc values because analysts can identify and associate very small events recorded on only a limited number of stations.

To further assess the robustness of newly identified events, we introduced additional quality-filtering criteria based on azimuthal gap, depth, and location uncertainty, and examined the temporal distribution of events to assess possible anthropogenic contamination. The filtered newly identified events show spatial and temporal behaviour consistent with the overall aftershock sequence.

Specific comments:

- the paper needs to be proofread in English – there are a number of unclear sentences and typos;

Manuscript has been proofread

- some sentences are too general and lead to wrong conclusions (e.g., lines 51-52, lines 162-164);

We tried to avoid too general sentences in the revised manuscript

- references are missing in several places (e.g., lines 60-65, lines 185-188);

References added where needed (e.g. Line 67, Line 189, and Line 201)

- Fig. 4 – briefly state in the figure caption what the areas labelled I, II, III and IV are. Where is the orange-coloured curve, mentioned in the figure caption? Indicate that in c and d subfigures there are cumulative intresections;

Figure 3 caption corrected

- when comparing catalogues (matching), only a temporal criterion is defined, and not a spatial one;

We clearly state that we only define temporal criterion

- line 222: after Table should be 1, instead of ??;

Latex compilation has been corrected

- at the beginning of section 4.3. it should be stated that the spatial window is larger than in the previous analysis (with HHC);

Line 248: we added the boundaries of the study region for comparison with CEC

- line 262: did PESMLC and HHC use the same relation to calculate the magnitudes (important for comparison)?;

Yes, as stated in the manuscript Line 202

- line 307: the sentence implies that the CSS analysts have no experience in their job, which is (actually very) unlikely;

Line 332: We clarified that CSS analysts had different levels of experience, as the data were picked by recent graduates and experienced analysts.

- Fig. 10: the last figure in the series in the second row shows a strange anomaly that requires further investigations and better explainations. Indeed, a larger number of weaker earthquakes in the catalogue must indicate smaller Mc and b values. In addition, the value of a is most likely also incorrect. All this affects the statistical analysis of seismicity, which is a prerequisite for SSH;

The discussion about Mc has been revised and extended including with re-plotting of Figure 9

- line 357: questionable claim;

- lines 368-370: the results show that this statement is also incorrect.

"Overall, our results strongly support the integration of machine learning methods into routine seismic monitoring, as it provides rapid, detailed, and comprehensive seismic catalogue essential for seismic hazard assessment and mitigation strategies."

As we managed to detect 95% of the seismicity shown in manual catalogues, missing only few events events with M>2 (most of them due to association parameter thresholds) we believe that our results do support the integration of machine learning methods into routine seismic monitoring

Dear Dr. Skoumal,

Thank you for your time and effort in reviewing our revised manuscript, and for the encouraging assessment of our work. We are pleased that the revisions were well received, and we are grateful for the additional minor suggestions from both the Associate Editor and Reviewer A, which have further improved the manuscript. We have carefully addressed all comments raised in this round of review, and believe the manuscript is stronger as a result. Below, we provide a point-by-point response to each comment. All line numbers refer to the revised manuscript file (with track changes).

Associate Editor

Figures:

> *Expanding on Reviewer A's request, please add axes labels to all figures (e.g., all of Figure 5, not just the Figure 5b mentioned by Reviewer A.)*

We thank the Associate Editor for this suggestion. We have added axis labels to all panels across all figures. In particular, Figure 5 now has "Longitude" and "Latitude" axis labels on all spatial panels (a, c, d), and "Date (day/month)" and "Frequency" labels on panel (b), consistent with the labelling style applied to Figures 3, 4, and 8.

Line 10:

> *"much" — Qualitative words like this can be removed without impacting the message you're conveying.*

Line 10: Agreed. We have removed "much" from the abstract. The sentence now reads: "Machine learning approaches have shown potential to automate the creation of seismic catalogues..."

Line 25:

> *"ML" — please define the acronym "machine learning" on first use. (Better yet, because you only use 'ML' one other time in the abstract, you could probably just write both words out?)*

We have followed the Associate Editor's suggestion and written "machine learning" in full throughout the abstract, removing the acronym entirely. In the introduction, "machine learning (ML)" is now defined on first use at line 40.

Line 51:

> *"template matching, often only work for the specific network they have been calibrated for." When following template matching best practices (e.g., statistical detection thresholding), it does not require any "calibrations." Rather, the main limitation of template matching is that it requires signals of interest to be identified a priori. If an undetected earthquake does not correlate well with any of the considered signals of interest, it will likely remain undetected.*

Line 50: We thank the Associate Editor for this important correction. The sentence has been revised to: "Existing semi-automatic event detection approaches designed to detect smaller events, such as template matching, require signals of interest to be identified a priori, meaning that earthquakes with waveforms dissimilar to any template will likely remain undetected." This more accurately reflects the fundamental limitation of template matching without implying it requires calibration in the traditional sense.

Line 74:

> *"Zagreb, nation's" — Add the word 'the'*

Line 73: Corrected. The sentence now reads: "...about 48 km south-southeast of Zagreb, the nation's capital."

Line 117:

> *"two, very" — Unnecessary comma*

Line 116: The comma has been removed. The sentence now reads: "This study benefits considerably from two very detailed, manually picked catalogues..."

Line 150:

> *"phase associator based on space-time partitioning" — Are there other types?*

Yes, there are several other types of phase associators employing fundamentally different approaches. For example, GaMMA (Zhu et al., 2022) uses a Bayesian Gaussian mixture model with expectation-maximisation;

REAL (Zhang et al., 2019) uses grid-search and pick counting; PhaseLink (Ross et al., 2019) uses a recurrent neural network on sequences of picks; and GENIE (McBrearty and Beroza, 2023) uses a graph neural network. Specifying "space-time partitioning" therefore meaningfully distinguishes PyOcto from these alternative approaches. We have retained the description in the text, as we feel it is a useful and accurate characterisation of PyOcto's core methodology.

Line 170:

"PESMLC" — Please move the acronym definition on Line 182 to this first use.

Line 168: We have moved the PESMLC acronym definition to its first use. The acronym is now defined at line 178 where we write "...the Petrinja Earthquake Sequence Machine Learning Catalogue (PESMLC) was made consisting of 41,989 events." The redundant definition that previously appeared later has been removed.

Line 216:

"catalogue are microseismic, with magnitudes below 1" — The term 'microseismic' can mean different things to different people. I'd suggest replacing this text with something like "catalog have magnitudes below 1"

Line 206: Agreed. We have replaced "are microseismic, with" with "have", so the sentence now reads: "...suggesting that the majority of events in the catalogue have magnitudes below 1."

Line 211:

How did you determine your S-wave window size used for the amplitudes?

Local magnitudes were calculated using the empirical relationship of Herak (2020), which specifies measuring the maximum S-wave amplitude of ground velocity within a standard time window following the S-wave arrival. The window is defined by the magnitude formula itself rather than chosen independently by us so it follows the standard practice used in the Croatian Seismic Catalogue.

Line 276:

"Spatial pattern...in latitude show excellent alignment...while differences in longitude are more subtle" — I would recommend removing this text or being clear about the motivation of this statement. I'm not seeing in Figure 5 what you're describing here; the patterns look quite similar to me in plot form. Later, when you quantify the differences (Line ~294), you state the latitude has nearly twice the difference (0.85 km) than the longitude (0.4 km), which seems to contradict your statement here.

Line 256: The Associate Editor is correct that this statement is both visually unsubstantiated and contradicted by the quantitative results that follow. We have removed the qualitative sentence

Line 358:

"improving network density is an effective solution" — if there are additional stations, surely the manual processing would also improve with those additional data. This would still result in ML missing events that manual processing would identify.

Line 318: The Associate Editor makes a valid point. A denser network would improve both ML and manual processing. The key advantage of a denser network for ML processing specifically is that it allows the phase associator's minimum pick requirements to be met for events that would otherwise be excluded due to insufficient phases, regardless of whether ML or manual methods are used for phase picking. We have reworded the sentence to: "improving network density is an effective solution, as it allows more events to meet the minimum phase association requirements regardless of whether picks are generated automatically or manually."

Reviewer A:

We thank Reviewer A for their thorough and encouraging assessment, and for the kind words regarding our revisions. We address each minor comment below.

Line 29:

remove "Nevertheless,"

Done. "Nevertheless," has been removed from the start of that sentence in the conclusion.

Figures 3b, 8b:

suggest writing "Time (month/year)" on the x-axis label to be clear

We have updated the x-axis labels of Figures 3b and 8b to read "Date (month/year)".

Figures 4b, 5b:

x-axis labels are unclear and incomplete. Consider writing the date in smaller font and/or vertically oriented so it fits, and add the label "Date" or "Time".

We have updated both figures. The x-axis tick labels in Figures 4b and 5b are now written in a smaller font and oriented vertically to improve legibility, and the axis label now reads "Date (day/month)" for Figure 4b and "Date (day/month)" for Figure 5b.

Line 249:

"3469 missed events" — check your numbers? Figure 5b implies 22963 CEC events minus 19489 matched events = 3474 missed events?

Line 251: The reviewer is correct — there is a discrepancy arising from an update to the CEC event count between manuscript versions. The CEC catalogue total is 22,963 events, giving 22,963 − 19,489 = 3,474 missed events. We have corrected "3469" to "3474" throughout the text and figure to ensure internal consistency.

Line 317:

"out of 136 missed events with magnitudes greater than M_{CR,L} 1 in HHC" — check your numbers? From lines 294-295 for missed events in HHC: 63 events ($1 \leq M \leq 2$) + 6 events ($2 \leq M < 3$) + 2 events ($M > 3$) = 71 events?

Line 320: The reviewer is correct. The number of missed HHC events with $M \geq 1$ is 63 + 6 + 2 = 71, not 136 (which incorrectly included events with $0 < M < 1$). We have corrected "136" to "71" in the text. We have also updated the corresponding number of events that occurred after OP-Net became operational from 25 to 21, which is the corrected count for the $M \geq 1$ subset.

Line 404:

"carefully selecting the estimation method": suggest writing as "carefully selecting the Mc estimation method" just to be clear

Line 408: Agreed. We have added "Mc" so the sentence now reads: "...carefully selecting the Mc estimation method when analysing large, ML-generated catalogues..."

Table 2, second row:

what does "t_cut (d)" and "c(d)" mean? Does this mean "t_cut (days)" and "c (days)"? Please clarify in caption.

Yes, "(d)" denotes days. We have updated the column headers to read "t_cut (days)" and "c (days)" throughout Table 2 to make the units unambiguous.

Table 2 caption:

After "Rows marked † are not reported as reliable estimates due to non-convergence of c within the tested parameter range", perhaps add something like ", where c was nonzero when the parameter fits were attempted with a lower Mc of 1.1" to be explicit.

We have incorporated the reviewer's suggested clarification. The caption now reads: "Rows marked † are not reported as reliable estimates due to non-convergence of c within the tested parameter range, where c remained non-zero across all tested t_cut values when fitting at Mc = 1.1."

We hope that these revisions fully address all remaining comments from both the Associate Editor and Reviewer A, and we look forward to hearing the outcome of this final review.

Sincerely,

Dinko Sindija